

# NMR-STUDY OF DYNAMIC STRUCTURAL TRANSITIONS IN RNA-MOLECULES

Dissertation  
zur Erlangung des Doktorgrades  
der Naturwissenschaften

vorgelegt beim Fachbereich Biochemie, Chemie und Pharmazie der  
**Johann Wolfgang Goethe-Universität**  
in Frankfurt am Main

von  
**BORIS FÜRTIG**  
aus Bad Nauheim

Frankfurt 2007

vom Fachbereich Biochemie, Chemie und Pharmazie der  
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

DEKAN: Prof. Dr. Harald Schwalbe

GUTACHTER: Prof. Dr. Harald Schwalbe  
Prof. Dr. Clemens Glaubitz

DATUM DER DISPUTATION: 2007

***„Freiheit und Leben kann man uns nehmen - die Ehre nicht.“***

*Otto Wels*





*Meiner Frau Julia*

This thesis was prepared under the supervision of Prof. Dr. Harald Schwalbe between June 2003 and May 2007 at the Institute for Organic Chemistry and Chemical Biology of the Johann-Wolfgang Goethe-University Frankfurt.

CHAPTER I

SUMMARY AND OVERVIEW

1

CHAPTER II

RNA STRUCTURE IS ORDERED HIERARCHICALLY AND CAN BE DESCRIBED BY  
NMR

PRIMARY STRUCTURE	5
SECONDARY STRUCTURE	10
A-FORM HELIX	10
BULGES	12
INTERNAL LOOPS	13
HAIRPIN LOOPS	14
BISTABLE SYSTEMS	18
Natural 2 <sup>nd</sup> structure selection	24
TERTIARY STRUCTURE	26

CHAPTER III

RNA FOLDING

TYPES OF RNA FOLDING	31
MODEL MOLECULES TO STUDY RNA FOLDING	
RIBOZYMES	33
RIBOSWITCHES	34
RNA FOLDING IN THE CELLULAR CONTEXT	36
MONITORING RNA FOLDING	38

## CHAPTER IV

**NMR methods to quantify dynamics**

	41
<b>SOLVENT EXCHANGE RATES</b>	45
METHODS FOR THE MEASUREMENT OF SOLVENT EXCHANGE RATES BY NMR	48
Hydrogen-Deuterium Exchange	48
Two-dimensional NOESY	48
Water-Inversion Experiments	52
<i>Comparison of exchange rates</i>	
<i>determined by different methods</i>	56
<i>Rates from the NOESY</i>	58
Lineshape analysis	58
 SOLVENT EXCHANGE RATES AS MEASURES OF	
BASEPAIRING DYNAMICS AND STABILITY	59
 <b>REALTIME NMR</b>	61
TECHNIQUES FOR REACTION INITIATION	61
Photo-caged compounds	61
<i>Caging the RNA molecule itself</i>	61
<i>Caging an RNA - interacting ligand</i>	65
<i>Caging divalent ions that are required for folding</i>	66
<i>Laser setup</i>	69
 Mixing techniques	69
Temperature jump methods	70
 TECHNIQUES TO RECORD RT-NMR SPECTRA	71
1D Methods	71
<i>Monitoring RNA folding kinetics on</i>	
<i>resonances of imino protons</i>	72
 nD Methods	75
<i>Accordion/ /modulated 2D</i>	76
<i>Hadamard</i>	76
<i>SOFAST</i>	79
<i>Ultrafast</i>	80
 FUTURE DIRECTIONS	81

---

## CHAPTER V

### NMR PARAMETERS AND STRUCTURE CALCULATION

NOE AND DISTANCES	83
SCALAR COUPLINGS AND DIHEDRAL ANGLES	85
SCALAR COUPLINGS AND HYDROGEN BONDS	87
RDCs AND BOND ORIENTATIONS	88
RESTRAINED MD AND SIMULATED ANNEALING	90

## CHAPTER VI

### NMR of RNA

REVIEW ARTICLE: NMR OF RNA	95
----------------------------	----

## CHAPTER VII

### RESONANCE ASSIGNMENT AND DETERMINATION OF LOCAL RNA CONFORMATION

	125
RESEARCH ARTICLE: NEW NMR EXPERIMENTS FOR RNA NUCLEOBASE RESONANCE ASSIGNMENT AND CHEMICAL SHIFT ANALYSIS OF AN RNA UUCG TETRALOOP	127
RESEARCH ARTICLE: CONFORMATIONAL DEPENDENCE OF $^1J$ SCALAR COUPLING CONSTANTS IN RNA: ANALYSIS FOR AN RNA CUUCGG-TETRALOOP	139

## CHAPTER VIII

### RNA-FOLDING I: CONFORMATIONAL SWITCHING OF RNA SECONDARY STRUCTURE ELEMENTS

	157
RESEARCH ARTICLE: KINETICS OF PHOTOINDUCED RNA REFOLDING BY REAL-TIME NMR SPECTROSCOPY	159

<b>RESEARCH ARTICLE:</b> A CAGED URIDINE FOR THE SELECTIVE PREPARATION OF AN RNA FOLD AND DETERMINATION OF ITS REFOLDING KINETICS BY REAL-TIME NMR	163
--	-----

<b>RESEARCH ARTICLE:</b> KINETICS AND THERMODYNAMICS OF RNA CONFORMATIONAL SWITCHING	167
--	-----

<b>REVIEW ARTICLE:</b> TIME-RESOLVED NMR STUDIES OF RNA FOLDING	185
---	-----

## CHAPTER IX

### RNA-FOLDING II: CONFORMATIONAL SWITCHING OF RNA TERTIARY STRUCTURE ELEMENTS

187

<b>REVIEW ARTICLE:</b> STRUCTURES OF RNA SWITCHES: INSIGHT INTO MOLECULAR RECOGNITION AND TERTIARY STRUCTURE	189
--	-----

<b>RESEARCH ARTICLE:</b> RESOLVING LIGAND-INDUCED RNA FOLDING OF A RIBOSWITCH APTAMER DOMAIN AT ATOMIC RESOLUTION	197
---	-----

<b>RESEARCH ARTICLE:</b> MONITORING THE MINIMAL HAMMERHEAD RIBOZYME REACTIVITY BY NMR	227
---	-----

## APPENDIX I

### NMR PULSE SEQUENCES AND NMR EXPERIMENTAL PARAMETERS

<b>EXPERIMENTS FOR RT-NMR</b>	
PSEUDO 3D-1H-NMR EXPERIMENTS	245
1D- <sup>1</sup> H	
1D- <sup>1</sup> H {15N/14N} with Ernst-angle excitation	
HADAMARD <sup>15</sup> N-HSQC	247
<b>EXPERIMENTS FOR NMR OF RNA</b>	
LONG RANGE HMBc WITH C2'-DECOUPLING	248
<sup>1</sup> H- <sup>31</sup> P-HEHAHA	249

APPENDIX II	
PHOTO-CIDNP EXPERIMENTS ON RNA	251
GERMAN SUMMARY	
NMR-SPEKTROSKOPISCHE UNTERSUCHUNGEN DYNAMISCHER STRUKTURELLER ÜBERGÄNGE IN RNA-MOLEKÜLEN	259
CURRICULUM VITAE	265
ACKNOWLEDGEMENT	267





## SUMMARY AND OVERVIEW

The following thesis is concerned with the elucidation of structural changes of RNA molecules during the time course of dynamic processes that are commonly denoted as folding reactions. In contrast to the field of protein folding, the concept of RNA folding comprises not only folding reactions itself but also refolding- or conformational switching- and assembly-processes (see chapter III). The method in this thesis to monitor these diverse processes is high resolution liquid-state NMR spectroscopy. To understand the reactions is of considerable interest, because most biological active RNA molecules function by changing their conformation (reviewed e.g. by Al-Hashimi<sup>2</sup>). This can be either an intrinsic property of their respective sequence<sup>3</sup> or may happen in response to a cellular signal such as small molecular ligand binding (like in the aptamer<sup>4</sup> and riboswitch<sup>5</sup> case), protein<sup>6</sup> or metal binding<sup>7</sup>.

The first part of the thesis (chapters II & III) provides a general overview over the field of RNA structure and RNA folding. The two chapters aim at introducing the reader into the current status of research in the field. Chapter II is structured such that primary structure is first described then secondary and tertiary structure elements of RNA structure. A special emphasis is given to bistable RNA systems that are functionally important and represent models to understand fundamental questions of RNA conformational switching. RNA folding *in vitro* as well as *in vivo* situations is discussed in Chapter III. The following chapters IV and V also belong to the introduction part and review critically the NMR methods that were used to understand the nature and the dynamics of the conformational/structural transitions in RNA. A general overview of NMR methods quantifying dynamics of biomolecules is provided in chapter IV. A detailed discussion of solvent exchange rates and time-resolved NMR, as the two major techniques used, follows. In the final chapter V of the first part the NMR parameters used in structure calculation and structure calculation itself are conferred.

The second part of the thesis, which is the cumulative part, encompasses the conducted original work. Chapter VI reviews the general NMR techniques applied and explains their applicability in the field of RNA structural and biochemical studies in several model cases.

Chapter VII describes the achievement of a complete resonance assignment of an RNA model molecule (14mer cUUCGg tetral-loop RNA) and introduces a new technique to assign quaternary carbon resonances of the nucleobases. Furthermore, it reports on a conformational analysis of the sugar backbone in this RNA hairpin molecule in conjunction with a parameterization of <sup>1</sup>J scalar couplings.

---

<sup>1</sup>O. Wilde, *The Chameleon*, 1894, **1**, "Phrases and philosophies for the use of the young"

<sup>2</sup>H. Al-Hashimi, *ChemBiochem*, 2005, **6**, 1506-1519

<sup>3</sup>E.A. Schultes, D.P. Bartel, *Science*, 2000, **289**, 448-452

<sup>4</sup>S. Muller, *ChemBiochem*, 2003, **4**, 991-997

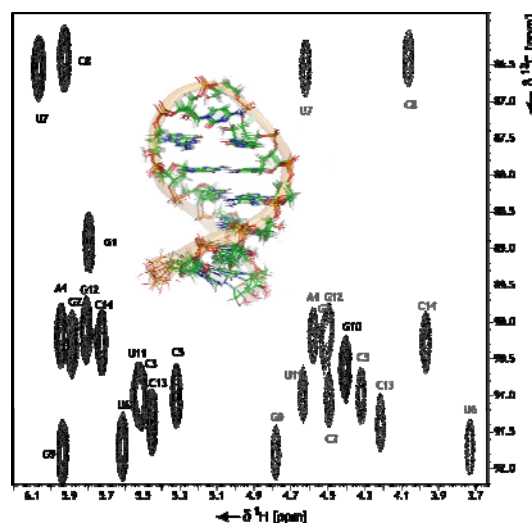
<sup>5</sup>J. Miranda-Rios, M. Navarro, M. Soberon, *Proc. Natl. Acad. Sci.*, 2001, **98**, 9736-9741

<sup>6</sup>J.R. Williamson, *Nat. Struct. Biol.*, 2000, **7**, 834-837

<sup>7</sup>D.E. Draper, D. Grilley, A.M. Soto, *Annu. Rev. Biophys. Biomol. Struct.*, 2005, **34**, 221-243

Achievements:

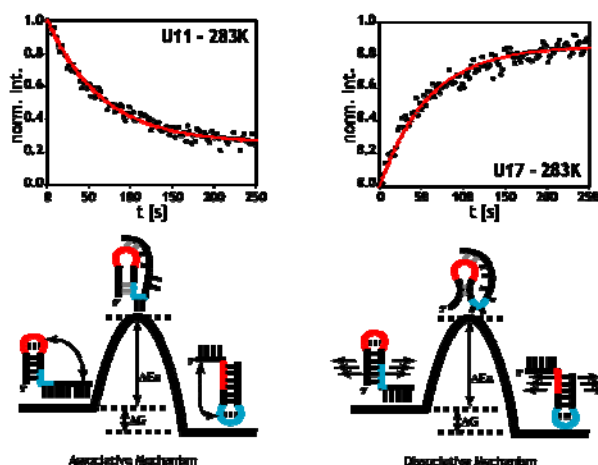
- Establishment of two new NMR pulse-sequences facilitating the assignment of quaternary carbons in RNA nucleobases
- First complete (99.5%) NMR resonance assignment of an RNA molecule (14mer) including  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{31}\text{P}$  resonances
- Description of RNA backbone conformation by a complete set of NMR parameters
- Description of the backbone conformational dependence in RNA of new NMR parameters ( $^1J$  scalar couplings)



Chapters VII & VIII summarize the real-NMR studies that were conducted to elucidate the conformational switching events of several RNA systems. Chapter VIII gives an overview on the experiments that were accomplished on three different bistable RNAs. These molecules were chosen to be good model systems for RNA refolding reactions and so consequently served as reporters of conformational switching events of RNA secondary structure elements.

Achievements:

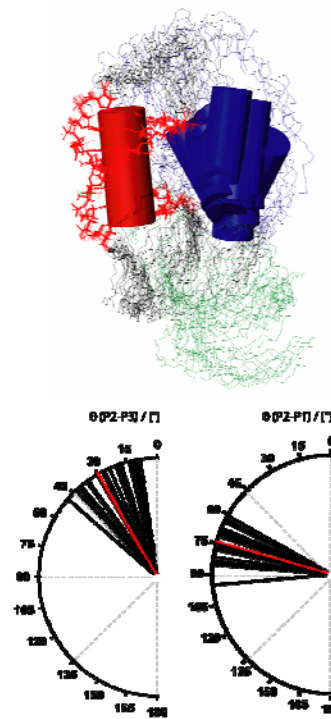
- First kinetic studies of RNA refolding reactions with atomic resolution by NMR
- Application of [new] RT-NMR techniques
  - either regarding the photolytic initiation of the reaction
  - or regarding the readout of the reaction
- Discovery of different RNA refolding mechanisms for different RNA molecules
- Deciphering of a general rule for RNA refolding



Finally, chapter IX deals with the application of the above-established methodology to conformational switching processes of RNA tertiary structure elements. The models for these processes were a) the guanine-dependent riboswitch RNA and b) the minimal hammerhead ribozyme.

Achievements:

- NMR spectroscopic assignment of imino-resonances of the hypoxanthine bound guanine-dependent riboswitch RNA
- Application of RT-NMR techniques to monitor the ligand induced conformational switch of the aptamer domain of the guanine-dependent riboswitch RNA at atomic resolution
- Translation of kinetic information into structural information
- Deciphering a folding mechanism for the guanine riboswitch aptamer domain
- Application of RT-NMR techniques to monitor the reaction of the catalytically active mHHR RNA at atomic resolution



In the appendices the new NMR pulse-sequences and the experimental parameters are described, which are not explicitly treated in the respective manuscripts.

#### LIST OF PUBLICATIONS:

- NMR of RNA; Boris Fürtig, Christian Richter, Jens Wöhnert, Harald Schwalbe, *ChemBiochem*, 2003, **4**, 936-962
- New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop, Boris Fürtig, Christian Richter, Wolfgang Bermel, Harald Schwalbe, *J. Biomol NMR*, 2004, **28**, 69-79
- Conformational dependence of  $1J$  scalar coupling constants in RNA: Analysis for an UUCG tetraloop; Boris Fürtig, Christian Richter, Harald Schwalbe, 2007, in preparation
- Kinetics of photoinduced RNA refolding by real-time NMR spectroscopy, Philipp Wenter, Boris Fürtig, Alexandre Hainard, Harald Schwalbe, Stefan Pitsch, *Angew. Chem. Int. Ed.*, 2005, **44**, 2600-2603
- A caged uridine for the selective preparation of an RNA fold and determination of its refolding kinetic by real time NMR, Philipp Wenter, Boris Fürtig, Alexandre Hainard, Harald Schwalbe, Stefan Pitsch, *ChemBiochem*, 2006, **7**, 417-420

- **Kinetic and thermodynamics of RNA conformational switching**, Boris Fürtig, Philipp Wenter, Christian Richter, Alexandre Hainard, Stefan Pitsch, Harald Schwalbe, 2007, submitted
- **Time-resolved NMR studies of RNA folding**, Biopolymers, Boris Fürtig, Janina Buck, Vijayalaxmi Manoharan, Andres Jäschke, Philipp Wenter, Stefan Pitsch, Harald Schwalbe, 2007, *Biopolymers*, in print
- **Structures of RNA switches: Insight into molecular recognition and tertiary structure**, Harald Schwalbe, Janina Buck, Boris Fürtig, Jonas Noeske, Jens Wöhnert, *Angew. Chem. Int. Ed.*, 2007, **46**, 1212-1219
- **Interplay of ‘induced fit’ and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch**, Jonas Noeske, Janina Buck, Boris Fürtig, Hamid Nasiri, Harald Schwalbe, Jens Wöhnert, *Nucl. Acids. Res.*, 2007, **35**, 572-583
- **Resolving ligand-induced RNA folding of a riboswitch aptamer domain at atomic resolution**, Janina Buck, Boris Fürtig, Jonas Noeske, Jens Wöhnert, Harald Schwalbe, 2007, *PNAS*, accepted with minor revisions
- **NMR Characterization of Phosphodiester Bond Cleavage catalysed by the minimal Hammerhead Ribozyme**, Boris Fürtig, Christian Richter, Peter Schell, Philipp Wenter, Stefan Pitsch, Harald Schwalbe, 2007, submitted
- C. Schmuck, H. Wennemers (Ed.), *Highlights in Bioorganic Chemistry*, Wiley-VCH, Weinheim 2004, Nucleic Acids from A to Z, Co-author.

## RNA STRUCTURE IS ORDERED HIERARCHICALLY AND CAN BE DESCRIBED BY NMR

The structure of RNA is ordered hierarchically; analogous to proteins it is dissected in primary, secondary and tertiary structure. Based on this order NMR experiments are developed that describe the structure on the respective level. Therefore, NMR resonance assignment forms the basis for the determination of structural parameters such as dihedral angles and atomic distances. The NMR experiments for the determination of RNA structure have been reviewed in Fürtig et al. 2003 (see chapter VI).

### PRIMARY STRUCTURE

The primary structure of RNA is represented by the sequences of the four natural monomeric units of nucleic acids - the nucleotides (Figure 1). Nucleotides consist of a phosphodiester backbone, a ribofuranosyl sugar moiety and a purine or pyrimidine nucleobase. The ribose sugar has the D-stereoisomeric configuration. The bases are planar hetero-aromatic structures without chiral atoms. The only exceptions are unusual bases that sometimes have asymmetric side chains attached<sup>9</sup>. The four canonical nucleobases are the purine analogues adenine and guanine as well as the pyrimidine analogues cytosine and uracil. In contrast to DNA, all of the major types of cellular RNA do not follow Chargaff's rules<sup>10</sup> –  $\chi_A = \chi_{T/(U)}$  and  $\chi_G = \chi_C$  –, this may imply that RNA is not primarily organised in canonical double helices. In fact, RNA adopts more complex structural elements.

Naturally occurring RNA-molecules contain a large number of modifications of the nucleobase and sugar moieties which highly diversify the basic set of the four nucleosides<sup>11,12</sup>. The most frequently encountered modifications are methylations of the nucleobase and of the 2'-OH group of the ribose moiety. Some of these methylated nucleotides have been identified to play an important role in controlling the correct secondary structure formation of biologically active RNAs, by suppressing alternative conformations which would be thermodynamically favoured in the absence of the modification<sup>13</sup>.

---

<sup>8</sup>O. Wilde, *The Chameleon*, 1894, **1**, "Phrases and philosophies for the use of the young"

<sup>9</sup>M. Helm, *Nucleic Acid Res.*, 2006, **34**, 721-33

<sup>10</sup>S. Zamenhof, G. Brawerman, E. Chargaff, *Biochem. et Biophys. Acta*, 1952, **9**, 402-405

<sup>11</sup>P.A. Limbach, P.F. Crain, J.A. McCloskey, *Nucleic Acid Res.*, 1994, **22**, 2183-2196

<sup>12</sup>P.F. Agris, *Nucleic Acid Res. Mol.Biol.*, 1996, **53**, 79-129

<sup>13</sup>C. Hörbartner, M.O. Ebert, B. Jaun, R. Micura, *Angew. Chem. Int. Ed.*, 2002, **41**, 605-609;

M. Helm, R. Giegé, C. Florentz, *Biochemistry*, 1999, **38**, 13338-13346;

R. Micura, W. Pils, C. Hörbartner, K. Grubmayr, M.O. Ebert, B. Jaun, *Nucleic Acid Res.*, 2001, **29**, 3997-4005

```

1  GGAUCCGCCU ACCUUUCAGC AGUUGCGCAG UUGUCUGCA AGACUCUAUG AGAAGCAGAU
61  AAGCGAUAAG UUGGCUCAAC AUCUUCUCGG GCADAAGUCG GACACCAUUG CAUCACAGUA
121 UCGUGAUGAC AGAGGCAGGG AGUGGGACAA AAUUGAAADC AAAUAADGAU UUAUUAUUGA
181 CUGADAGUGA CCUGUUCGUU GCAACAAAUU GAUAAGCAAU GCUUUUUAU AADGCCAACU
241 UAGUAUAAA UAGCCAACCU GUUCGACAAA GAUUAUGAGA CAGUCUAUGG CUACCAAAACU
301 GCAGGACGGG AAUACACCUU GUCUGGCAGC UACACCUUCU GAACCAAGUC CCACCGGUCU
361 GUGUUAUGAC UCGGCGGUCG GUGGUUUGUC GGUUUAUGAC GAGAUCCGGC AUCUCUUAAC
421 GGAUCCGCAU UACAGGAGAG CUUGCGAUA CGUCGCGGCG CGGUAUGGGC AAAAAGGCA
481 AGCGUUUAU GUUGAGCGAG UGGUGCAAU UGDCACCGCG GUGCAAGAAC GUUAUCCCU
541 UGCGCGGCU GUUGGCGCUU GCAACACUGC CAGUACCGU UCAUUCUCG CAUUAAGCGA
601 AAAGUUCGAC UUCGCGGUG UGGUGUCU GCGGCGGUAU AAACUUGCG CACGUCUGAC
661 GGCAAUUGC AUUGCGGAT UACUGGCAC CCUGCGGACA GUUAUAGCUU CUUAUACUCA
721 UGAGCGAUC GCGCGUUCG CUAUGAAG CGAGUAGAA AGGCUUGGCU CGGAGAGAG
781 GGUUGAGUG GCGAGAGCG AGCUACAUUG CGAAGAGGU UCUCUGAGG CACUAAAACG

```

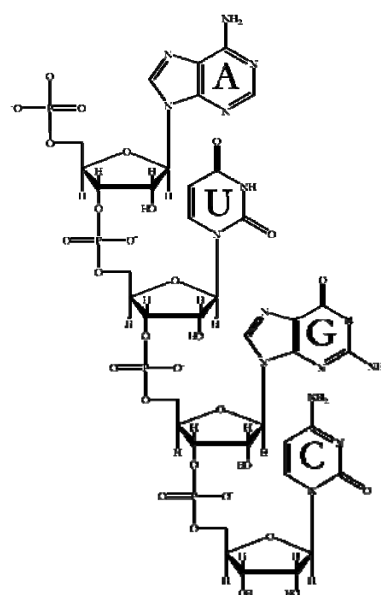


Figure 1: left: Part of the primary structure of the 16S RNA from *E.coli*<sup>14</sup>; right: constitution of an exemplary RNA strand displaying the four nucleotide building blocks.

The 1D  $^1\text{H}$  NMR and 2D  $^{13}\text{C}$ -HSQC spectra of an unlabelled nucleotide (Figure 2) reveal several distinct signals in spectral regions that are restored in the spectrum of RNA molecules. The signals shown in figure 2 stem from non-exchanging protons. These protons are covalently attached to carbon atoms and therefore cannot exchange with solvent protons. The exchangeable protons of the mononucleotides e.g. protons from imino- and amino-groups cannot be monitored in deuterated protic solvents, because they are not protected against chemical exchange with the solvent protons. The spectral regions of all protons found in RNA molecules are listed in the following table (Table 1):

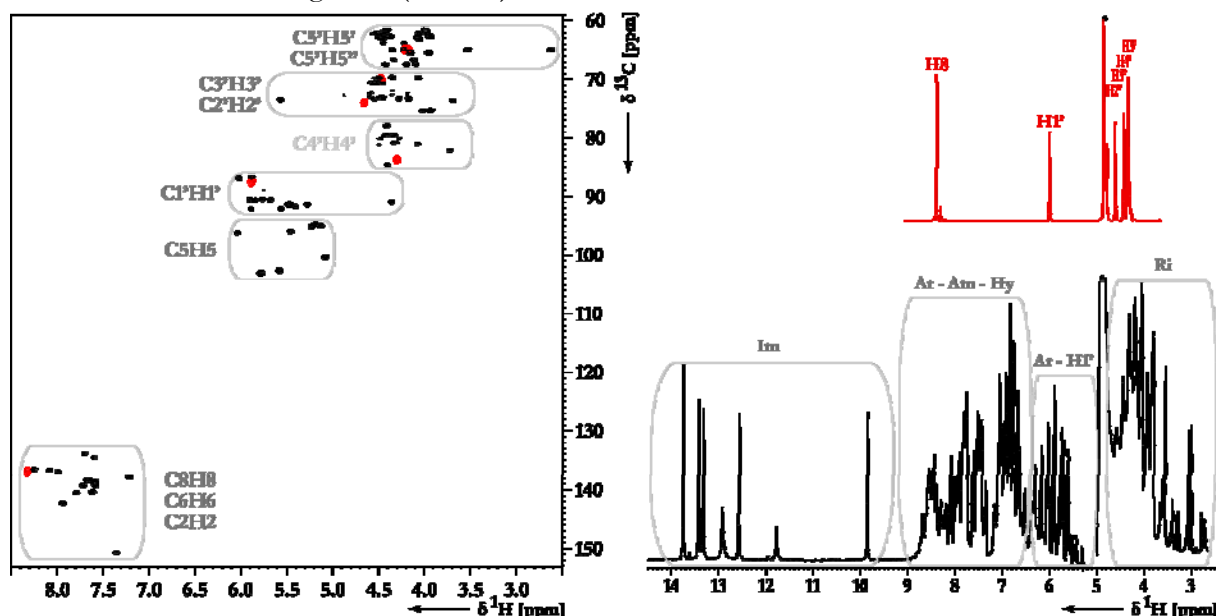
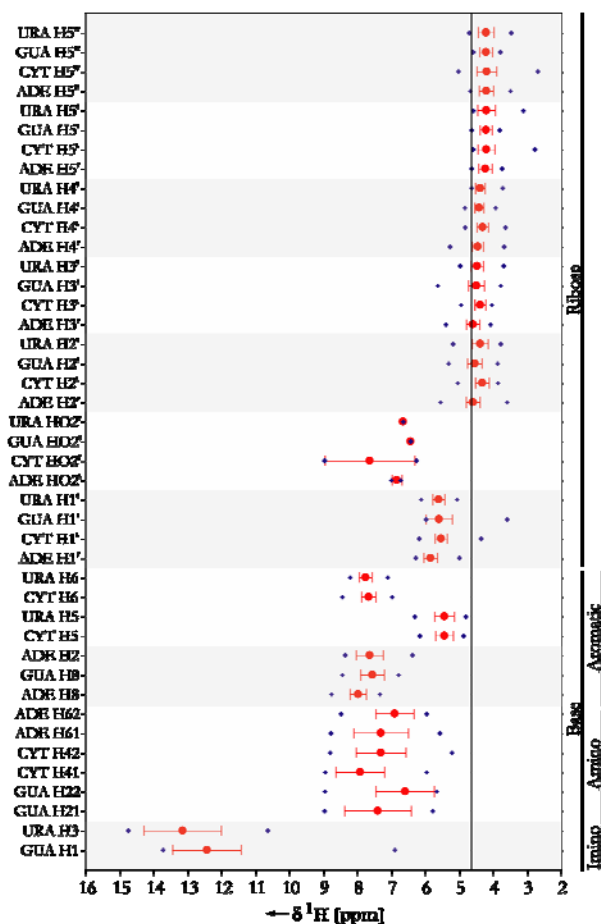


Figure 2: 1D  $^1\text{H}$ - and 2D  $^{13}\text{C}$ -HCQC spectra of the nucleotide-triphosphate guanine (red) and of the RNA 14mer cUUCGg hairpin molecule (black). The typical spectral regions are assigned by grey boxes (Im: imino-protons, Ar: aromatic protons, Am: amino protons, Ri: ribose protons); in the 1D  $^1\text{H}$  spectrum of guanine resonance assignment is annotated.

<sup>14</sup>J. Brosius, T.J. Dull, D.D. Sleeter, H.F. Noller, 1981, *J. Mol. Biol.*, **148**, 107-127

Table 1: Spectral regions of all protons in RNA molecules. In the table the chemical shift values of 82 RNA molecules are summarized as deposited in the BMRB database<sup>15</sup>; red dots indicate the mean value of the respective chemical shift, error bars represent the standard mean deviation, blue crosses indicate the extreme values measured for the respective chemical shift. The black line indicates the chemical shift value of the water resonance.



In a RNA molecule, phosphodiester bonds between the 5' and the 3' ends of the ribose moieties connect the nucleotides. The backbone of an RNA molecule is therefore composed of the ribose units that are intervened by the highly acidic (the  $pK_A$  of the single ionisable proton is around 1) phosphate groups that are negatively charged at natural pH-values.

This connectivity creates an RNA molecule in which six torsion angles (see table 2: Torsion angle of the RNA backbone for definition of the torsions) define the conformation of the backbone of the monomeric unit and incrementally influence the shape of the whole molecule, as discussed in the following chapter. The torsion angle  $\chi$  determines the orientation of the nucleobase relative to the ribosyl moiety.

<sup>15</sup>Status quo 02-11-2007: Biological Magnetic Resonance Data Bank: [www.bmrb.wisc.edu](http://www.bmrb.wisc.edu); B.R. Seavey, E.A. Farr, W.M. Westler, J.L. Markley, *J. Biomol. NMR*, 1991, **1**, 217-236

Table 2: Torsion Angles of the RNA backbone: exocyclic torsion angles, per definition the direction for the definition goes along 5'-end to 3'-end; the glycoside torsion angle is  $\chi$ . Histograms show occurrence of specific torsion in the 23S RNA (2904 nucleotides).

Name	Atoms	A-form	B-form	23S RNA
$\alpha$	$O3'_{i-1}-P_i-O5'_i-C5'_i$	$-68^\circ$	$-46^\circ$	
$\beta$	$P_i-O5'_i-C5'_i-C4'_i$	$178^\circ$	$-147^\circ$	
$\gamma$	$O5'_i-C5'_i-C4'_i-C3'_i$	$54^\circ$	$36^\circ$	
$\delta$	$C5'_i-C4'_i-C3'_i-O3'_i$	$82^\circ$	$157^\circ$	
$\epsilon$	$C4'_i-C3'_i-O3'_i-P_{i+1}$	$-153^\circ$	$155^\circ$	
$\zeta$	$C3'_i-O3'_i-P_{i+1}-O5'_{i+1}$	$-71^\circ$	$-96^\circ$	
$\chi$	Pyr.: $C2'_i-C1'_i-N1_i-C2_i$	$-158^\circ$	$-98^\circ$	
	Pur.: $C2'_i-C1'_i-N9_i-C4_i$			

A frequency of occurrence analysis -as depicted in the right column of table 2- of the 23S RNA structure reveals that the torsion angle values populate very narrow bands of the entire conformational space.



In a further analysis (Figure 3) pairs of subsequent backbone angles are plotted against each other and reveal that the torsions in RNA are highly correlated.

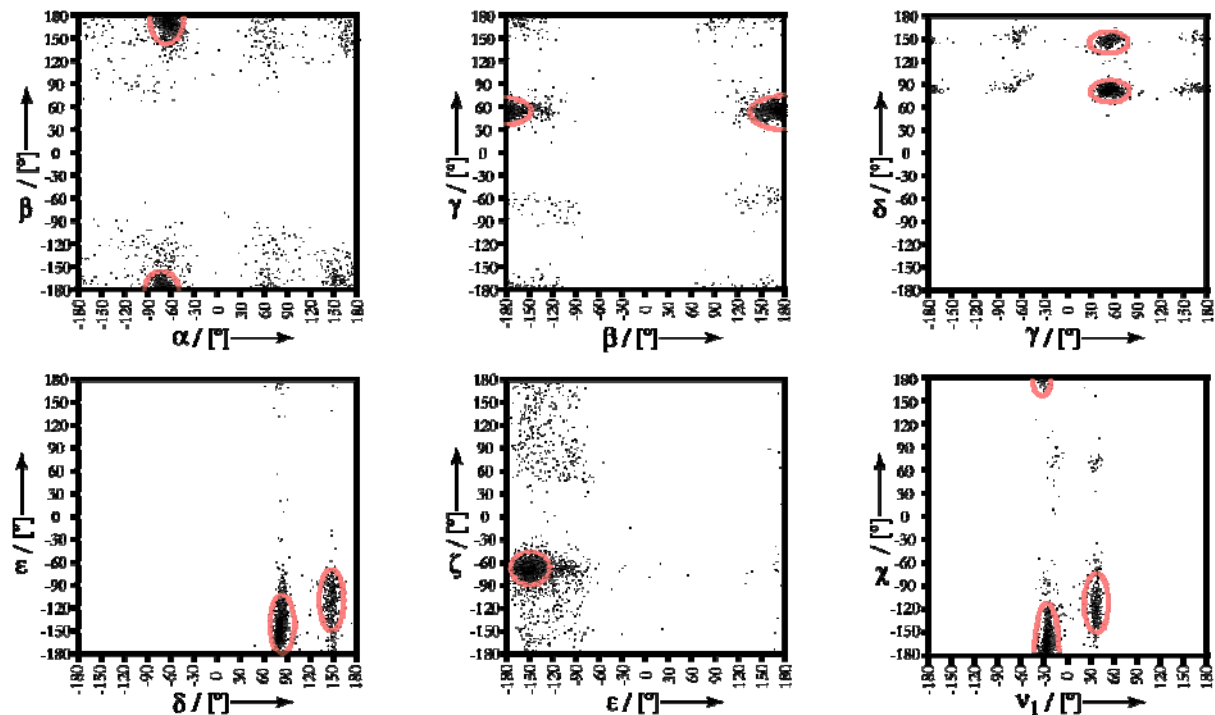


Figure 3: Correlation plots of subsequent backbone torsion angles in RNA; from left to right the combinations  $\alpha/\beta$ ;  $\beta/\gamma$ ;  $\gamma/\delta$ ;  $\delta/\epsilon$ ;  $\epsilon/\zeta$  and  $\nu_1/\chi$  (correlates the sugar pucker with the glycosidic angle) are depicted as found in the crystal structure of the ribosomal 23S RNA. Red circles indicate the most populated areas of angle combinations.

The different correlations can be dissected in two classes: first those that show a single major area of angle combinations that is populated ( $\alpha/\beta$ ;  $\beta/\gamma$ ;  $\epsilon/\zeta$ ) and second those displaying two regions that are most frequently populated ( $\gamma/\delta$ ;  $\delta/\epsilon$ ;  $\nu_1/\chi$ ).

As discussed in the following sections the changes in the torsions and their plasticity give rise to the structural diversity of RNA.

## SECONDARY STRUCTURE

Folding of the polynucleotide backbone enables the nucleobases to interact with each other and to form elementary secondary structural motifs (Figure 4). The dominating motif is the A-form helix with complementary hydrogen-bonded Watson-Crick pairs of nucleobases. These helical elements are often separated by “single-stranded” regions that are categorized in hairpin-loops, bulges and internal or multiple loops.<sup>16</sup> The latter mentioned motifs are considered to be single-stranded, but indeed, they are mostly paired in non-canonical arrangements<sup>17</sup>, building up interactions of higher order within the RNA molecule.

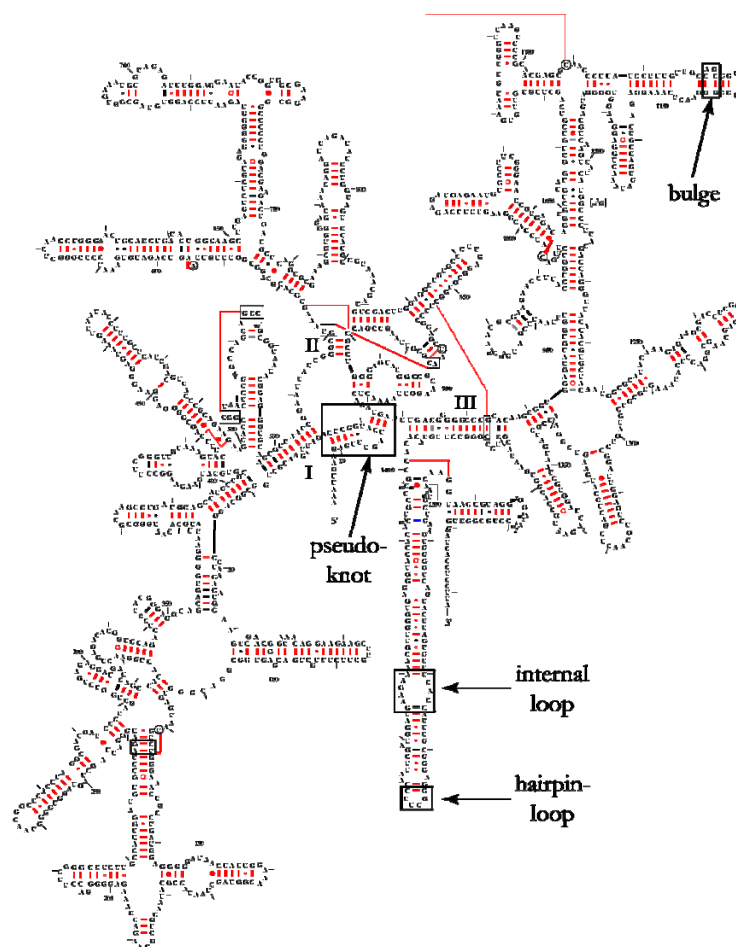


Figure 4: Representation of the secondary structure of the 16S RNA of *E. coli*<sup>18</sup>. Important secondary structure elements are highlighted by boxes, basepairing is indicated by bars (Watson-Crick) and dots (G-U; G-A and U-U). Tertiary interactions with strong comparative data are connected by solid lines. Although pseudoknots are indicated in the figure, they will be discussed in the section tertiary structure.

## A-FORM HELIX

Around Watson-Crick basepairs many different helical geometrics can be built. Whereas in DNA the B-form helix is dominant<sup>19</sup>, the canonical conformation in double stranded regions of RNA is the A-form helix<sup>20</sup> (Figure 5). In this helical arrangement, two anti-parallel strands<sup>21</sup> form a right handed double helix with rise of the helix of 2.55Å/bp that corresponds to a packing of 11bp per helical turn. The phosphate-to-phosphate distance is 5.9Å and therefore shorter than in the

<sup>16</sup>L.X. Shen, Z. Cai, I. Tinoco Jr., *FASEB J.*, 1995, **9**, 1023-1033

J.R. Wyatt, I. Tinoco Jr., *RNA world*, Cold Spring Harbor Lab. Press, Cold Spring Harbor NY, 1993, 465-497

<sup>17</sup>R. Schroeder, A. Barta, K. Semrad, *Nature Reviews Mol. Cell Biol.*, 2004, **5**, 908-919

<sup>18</sup>J.J. Cannone, S. Subramanian, M.N. Schnare, J.R. Collett, L.M. D'Souza, Y. Du, B. Feng, N. Lin, L.V. Madabusi, K.M. Muller, N. Pande, Z. Shang, N. Yu, R.R. Gutell, *BioMed Central Bioinformatics*, 2002, **3**, 2

<sup>19</sup>J.D. Watson, F.H. Crick, *Nature*, 1953, **171**, 737

<sup>20</sup>R.O. Day, N.C. Seeman, J.M. Rosenberg, A. Rich, 1973, *Proc. Natl. Acad. Sci.*, **70**, 849-853

J.M. Rosenberg, N.C. Seeman, J.J.P. Kim, F.L. Suddath, H.B. Nicholas, A. Rich, 1973, *Nature*, **243**, 150-154

<sup>21</sup>A. Kornberg, *Harvey Lectures*, 1958, **53**, 83-112

canonical B-form ( $7.0\text{\AA}$ ). The planes of the nucleobases are not perpendicular to the helix axes but are tilted against it at an angle of  $70^\circ$ . There is also a slight dihedral angle between the base planes of  $\sim 16^\circ$ . These whole parameters therefore describe a double helix that is shorter and wider than the B-form helix. It displays a deep narrow major groove –that is, if thought in the light of RNA-protein interactions, not easily accessible to proteins– and also a wide but shallow minor groove –that is easily accessible but has a lower content of information regarding the sequence–.

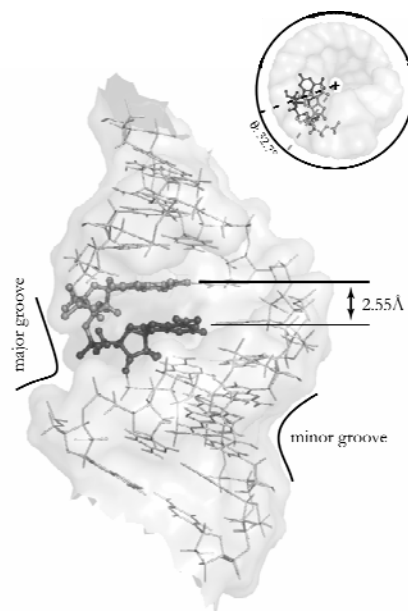


Figure 5: 3D model (modelled with InsightII) of a canonical A-form helix; a) side view where pitch, major and minor groove are highlighted; b) top view indicating the tilt angle between two following nucleotides. The 2.55 Å are describing the value for the rise of the helix per basepair, this is shorter than the actual distance of atoms in the bases ( $\sim 3.33\text{\AA}$ ) due to the fact that the ortho-normal of the baseplanes are not parallel to the helix axis.

Association of two RNA strands to form a double helix is entropically unfavourable because of loss in translational entropy and the whole RNA molecules rigidifies. Furthermore, the accumulation of negatively charged phosphates leads to charge repulsion. However, as also observed in nucleosides and bases –as planar aromatic molecules– that tend to stack upon each other, a snug fit of bases in a helix lowers the enthalpy by favourable London/van-der-Waals interactions. Additionally, the Watson-Crick complementarities of the bases energetically favour the double strand formation. As a consequence, the thermodynamics of strand association are treated as a two step process:

- |                     |                                 |
|---------------------|---------------------------------|
| A) Initiation step: | Formation of the first basepair |
| B) Propagation:     | Addition of new basepairs       |

If melting experiments are analysed according to a nearest neighbour model, then the following energy parameters for propagation of the double helix are obtained<sup>22</sup>:

$$\begin{aligned}
 \Delta H^0 &= -6.6 \quad \text{to} \quad 14.2 \text{ kcal mol}^{-1} \\
 \Delta S^0 &= -18.4 \quad \text{to} \quad -34.9 \text{ cal mol}^{-1} \cdot \text{K}^{-1} \\
 \Delta G^0 &= -0.9 \quad \text{to} \quad -3.4 \text{ kcal mol}^{-1}
 \end{aligned}$$

<sup>22</sup>S.M. Freier, R. Kierzek, J.A. Jaeger, N. Sugimoto, M.H. Caruthers, T. Neilson, D.H. Turner, *Proc. Natl. Acad. Sci.*, 1986, **83**, 9373-9377

Since the terminal basepair possesses solely a single stacked basepair the pair is less constrained and therefore more dynamic – this is considered as ‘breathing’ of the RNA terminus.

Regarding this in dynamic studies of RNA, a decrease of flexibility is observed when going from the terminal ends of the helix towards its middle.<sup>23</sup>

### BULGES

Bulges are defined as double stranded nucleic acid molecules that exhibit unpaired nucleotides on only one strand. Whereas a single pyrimidine bulge is mostly extrahelical, single purine bulges have a strong tendency to stack in the double helix. This finding<sup>24</sup> is consistent with the stacking properties of single nucleobases. Experimental equilibrium constants at 25°C are reported to be around 1M<sup>-1</sup> for purines and 5-10M<sup>-1</sup> for pyrimidines.<sup>25</sup> Thermodynamic data for the stacking event determined in optical spectroscopy studies on di-nucleotides reveals the following energy parameters<sup>26</sup>:

$$\begin{aligned}\Delta H^0 &= -5\text{kcal mol}^{-1} \text{ to } 8\text{kcal mol}^{-1} \\ \Delta S^0 &= -25\text{cal mol}^{-1} \cdot \text{K}^{-1} \\ \Delta G^0 &= -0.9 \text{ to } -1.7\text{kcal mol}^{-1}\end{aligned}$$

The propensity of stacking for the single nucleobases provides the following order G>A>C>>U. Bulges with more than a single unpaired nucleotide bend the helix. The exact bend angle is a function of the exact sequence as well as of the effect of divalent ions such as Mg<sup>2+</sup> that compensate repulsion forces between the negatively charged backbone phosphates.<sup>27</sup> The TAR-element of HIV represents one of the most prominent bulge RNA structures. Here, the bulge consisting of two or three nucleotides (depending on the HIV-type) plays a critical role in the essential interaction with a single arginine residue of the TAT-protein. In the free form the bulge nucleotides of TAR are stacked between the two helices and have no tertiary interactions with stem nucleotides. This conformation changes in the complexed state of TAT-TAR, in which the bulge nucleotides are no longer stacked. A base triple is formed between a bulge U and an A-U basepair in the upper stem of TAR to stabilize the interaction of the TAT arginine residue with a G-C pair and the phosphodiester backbone along the major groove.<sup>28</sup>

<sup>23</sup>M. Akke, R. Fiala, F. Jiang, D. Patel, A.G. Palmer<sup>3rd</sup>, *RNA*, 1997, **3**, 702-709; E. Duchardt, H. Schwalbe, *J. Biomol. NMR*, 2005, **32**, 295-308

<sup>24</sup>P.N. Borer, Y. Lin, S. Wang, M.W. Roggenbuch, J.M. Gott, O.C. Uhlenbeck, I. Pelczer, *Biochemistry*, 1995, **34**, 6488-6503; Y.T. van den Hoogen, A.A. van Beuzekom, E. de Vrom, G.A. van der Marel, J.H. van Boom, C. Altona, *Nucl. Acids. Res.*, 1988, **16**, 5013-5030

<sup>25</sup>V.A. Bloomfield, D.M. Crothers, I. Tinoco Jr., *Nucleic Acids: Structures, properties and functions*, 1998, University Science Books, Mill Valley CA

<sup>26</sup>R.C. Davis, I. Tinoco Jr., *Biopolymers*, 1968, **6**, 223-242

<sup>27</sup>M. Zacharias, P.J. Hagermann, *J. Mol. Biol.*, 1995, **247**, 486-500

<sup>28</sup>J.D. Puglisi, R. Tan, B.J. Calnan, J.R. Williamson, *Science*, 1992, **257**, 76-80

## INTERNAL LOOPS

By definition, internal loops are formed if internal nucleotides on both sides of the double helix cannot pair to form a Watson-Crick or G·U basepair. By introducing distortions in the helical geometry, the RNA can adopt novel shapes, which are key features in molecular recognition processes. Furthermore, internal bulges increase the inherent flexibility of domains in the RNA molecule by decreasing the stability of long helical stretches.

As a general common feature in internal loops with equal number of nucleotides, each pair exhibits non Watson-Crick hydrogen bonds that lead to a distortion of the A-form helix. The distortion can be expressed as a difference in the P-P distance, a deviation from canonical helical twist and as a bending in the helix axis.<sup>29,30</sup>

The decrease in stability is intimately linked to the structural features of the loops. This could be shown by comparison of the duplexes (5'-NNNAGNNN-3')<sub>2</sub> and (5'-NNNACNNN-3')<sub>2</sub>: the standard free energy was  $\Delta G^0_{37} = -6 \text{ kcal/mol}$  versus  $\Delta G^0_{37} = -3 \text{ kcal/mol}$ , respectively. This data show that a duplex with two consecutive G·A pairs is more stable than the one that has to form C·C and A·A mismatches. This is due to the fact that in the first case the single hydrogen-bonded G·A pairs fit more efficiently into the A-form helix than the double hydrogen bonded A·A and C·C counterparts. There is not only a dependence of the thermodynamics and conformation on the loop nucleotides itself but also on the sequence context: In the case of three duplexes that had the same pairing geometry of 2 consecutive G·A mismatches as nucleotides forming the internal loop plus the same number of G·C pairs in the two apical stems, the structures deviate from each other quite significantly depending on the sequence context. Whereas the two duplexes with loop sequences 5'-GGAC-3' or 5'-CAGG-3' possess an expansion of the backbone, the RNA internal loop with 5'-CGAG-3' is characterized by a contraction of the backbone geometry<sup>31</sup>. Unequal numbers of bases on both sides of internal loops lead, depending on the nucleotides involved, to unstacking of single nucleotides – often found are unstacked Us (found in internal loops in the P4-6 domain of the group I ribozyme<sup>32</sup> or in the RRE element of HIV<sup>33</sup>) – that consequently point towards the solvent. These loops may act as sites for tertiary interaction or for RNA-protein recognition. In some cases, they may just form if such a tertiary interaction is present – as demonstrated for the internal loop J6a/6b in the group I intron ribozyme that adopt its native shape just within the presence of tertiary interaction with consecutive A nucleotides several positions downstream the sequence in the same RNA strand (formation of an A-platform)<sup>32</sup> and that is structured differently in the absence of these tertiary interaction partners<sup>34</sup>.

As a conclusion one can summarise that in internal loops the nucleotides involved tend to form cross-strand hydrogen bonds and to compatible fit to adjacent pairs to optimize stacking.

<sup>29</sup>K.J. Baeyens, H.L. DeBondt, A. Pardi, S.R. Holbrook, *Proc. Natl. Acad. Sci.*, 1996, **93**, 12851-12855

<sup>30</sup>C.C. Correll, B. Freeborn, P.B. Moore, T.A. Steitz, *Cell*, 1997, **91**, 705-712

<sup>31</sup>M. Wu, D.H. Turner, *Biochemistry*, 1996, **35**, 9677-9689

J. SantaLucia Jr., D.H. Turner, *Biochemistry*, 1993, **32**, 12612-12623

M. Wu, J. SantaLucia Jr., D.H. Turner, *Biochemistry*, 1997, **36**, 4449-4460

<sup>32</sup>J.H. Cate, A.R. Gooding, C. Podell, B.L. Golden, C.E. Kundrot, T.R. Cech, J. Doudna, *Science*, 1996, **273**, 1696-99

<sup>33</sup>R.D. Peterson, J. Feigon, *J. Mol. Biol.*, 1996, **264**, 863-877

<sup>34</sup>S. Butcher, T. Dieckmann, J. Feigon, *EMBO J.*, 1997, **16**, 7490-7499

## HAIRPIN-LOOPS

Hairpinloops are composed of an anti-parallel helical stem and “single stranded” nucleotides that form the closing loop. Their inherent attribute is to reverse the direction of an RNA chain. Among all secondary structural elements –except for canonical helical stems– they represent a dominating motif. In principle, the loop part can be made up of any number of unpaired nucleotides. The smallest loop is therefore composed of two nucleotides and referred to as the biloop. For example, biloops of the category cUNg are highly abundant in ribosomal RNA<sup>35</sup>. NMR spectroscopic investigations of a biloop of the type cUUG confirm that the two uracile nucleotides are single stranded (or unpaired) and that a basepair between the C and the G nucleotide is closing and connecting the loop to the stem.<sup>36</sup> This conformation is stabilized by the fact that the nucleobase of the first unpaired uracile residue interacts with the minor groove of the stem.

Triloops of the sequence cUUUG have also been studied by NMR spectroscopic methods<sup>37,38</sup>. Remarkably in these cases the closing C-G basepair is preceded by three highly dynamic residues, that point uracile nucleobases directly into the solvent. This is probably due to the fact that no stabilising interactions by base stacking are present – this is in line with former studies of single stranded polyU and UU where the lack of base stacking has also been described<sup>39</sup>. As described in the literature loops, consisting of five nucleotides are also possible. Puglisi<sup>40</sup> et al. describe a loop structure of uUUCUGa, but this structure could possibly also be understood as a triloop that is extended by a G·U basepair. This kind of loop that is followed by a polyU sequence in the mRNA is responsible for the termination of transcription<sup>41</sup>.

Similar to the above-mentioned pentaloop, the structural description of hexaloops as tetraloops extended by non-canonical basepairs is valid<sup>42,43</sup>. The conserved sequence cGUAAUAg is primarily found in rRNA, and it behaves like an UAAU tetraloop closed by a G·A basepair.

Among all hairpin-loops the most prominent one is surely the anticodon-loop of the tRNA that is composed of seven nucleotides. In all characterized tRNA molecules, it is mainly composed of conserved and semi-conserved nucleotides –exchanges of those mostly occur for isosteric nucleobases. These conservations are necessary for the establishment of functional canonical anticodon hairpin structures that include the first motif that has been characterized in RNA molecules, which is called U-turn motif.<sup>44</sup>

Loops with eight or more unpaired nucleotides are at least identified on the basis of biochemical experiments, it is probable that those are stabilised by non-canonical basepairs or by intracatenaric hydrogen bonds.

<sup>35</sup>C.R. Woese, S. Winker, R.R. Gutell, *Proc. Natl. Acad. Sci.*, 1990, **87**, 8467-8471

<sup>36</sup>F.M. Jucker, A. Pardi, *Biochemistry*, 1995, **34**, 14416-14427

<sup>37</sup>P.W. Davis, W. Thurmes, I. Tinoco Jr., *Nucleic Acid Res.*, 1993, **21**, 537-545

<sup>38</sup>C. Sich, O. Ohlenschläger, R. Ramachandran, M. Görlich, L.R. Brown, *Biochemistry*, 1997, **36**, 13989-14002

<sup>39</sup>C.R. Cantor, P.R. Schimmel, *Biophysical Chemistry part I*, 1980, W.H. Freeman and Co., New York

<sup>40</sup>J.D. Puglisi, J.R. Wyatt, I. Tinoco Jr., *Biochemistry*, 1990, **29**, 4215-4226

<sup>41</sup>M.L. Birnstiel, M. Busslinger, K. Strub, *Cell*, 1985, **41**, 349-359








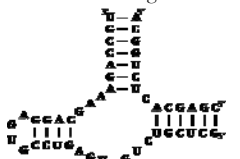
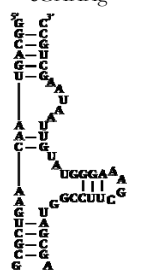

<sup>42</sup>M.A. Fountain, M.J. Serra, T.R. Krugh, D.H. Turner, *Biochemistry*, 1996, **35**, 6539-6548

<sup>43</sup>S. Huang, Y.X. Wang, D.E. Draper, *J. Mol. Biol.*, 1996, **258**, 308-321

<sup>44</sup>G.C. Quigley, A. Rich, *Science*, 1976, **194**, 796-806; P. Auffinger, E. Westhof, *RNA*, 2001, **7**, 334-341

The most abundant hairpin-loops are those composed of four unpaired nucleotides called tetraloops. It was shown that in a database (this database contains the sequences of 214 different 16S and 23S rRNA molecules originating from 133 different organisms) containing 7484 hairpin-loops 51% of all loops are tetraloops<sup>45</sup>. Tetraloops are found to have various biological relevant functions. They are found to act as nucleation sites for RNA folding<sup>46</sup> and they are also important in RNA-RNA<sup>47</sup> as well as in RNA-protein<sup>48</sup> recognition. Several RNA molecules that contain RNA (tetra-)loops are also characterized in this thesis. The molecules are therefore classified and shown in the following table:

Table 3: Loops as abundant secondary structure elements in the RNA molecules under investigation

	<i>tetraloops</i>		<i>different loops</i>
molecule	GNRA	YNMG	
14mer RNA		cUUCGg 	
bistable 20mer (A)	cGGAAG 	gUCCGc 	
bistable 20mer (B)	gGCAAc 	cUUCGg 	
bistable 34mer		gUUCGc & cUGCGg 	cUGUGUUg 
Hammerhead Ribozyme	cGUGAg 		
P5abc	cGAAAG 		
G-Riboswitch			gGAUAUGGc & aCCGUAAAu 

<sup>45</sup>D.J. Proctor, J.E. Schaak, J.M. Bevilacqua, C.J. Falzone, P.C. Bevilacqua, *Biochemistry*, 2002, **41**, 12062-12075

<sup>46</sup>O.C. Uhlenbeck, *Nature*, 1990, **346**, 613-614

<sup>47</sup>E. Westhof, V. Fritsch, *Structure*, 2000, **8**, R55-R65

<sup>48</sup>J.R. Williamson, *Nat. Struct. Biol.*, 2000, **7**, 834-837

Tetraloops are classified according to their nucleotide sequence inherently representing a distinct conformation. As apparent in the literature of 2002<sup>49</sup>, three main classes of tetraloops that are characterized by an increased stability compared to those of random sequences, are given by the sequences GNRA, UNCG and CUUG. As investigated furthermore (see following table), tetraloops referring to the sequence cYNMGg reveal the greatest thermodynamic stability. The relative low stability of the UUUU loop motif is concerted with the highest flexibility observed in this structure element<sup>50</sup>.

Table 4: Thermodynamic parameters for RNA tetraloops exhibiting different loop sequences: in each case the remaining nucleotides are: 5'-GGACLLLLGUCC-3' (L: are the loop nucleotides listed in the table). Values are determined in 1M sodium chloride, 0.01M sodium phosphate, 0.1mM EDTA at pH7.<sup>51</sup>

Sequence		T <sub>M</sub> [°C]	ΔH [kcal/mol]	ΔG [kcal/mol]
YNMG	UUCG	76.2	-55.9	-6.3
	UACG	73.8	-53.6	-5.7
non-YNMG	GCUU	70.9	-45	-4.4
	UUUG	70.3	-44	-4.2
	UUUU	69.9	-44.3	-4.2

A representative of the YNMG motif that is also included in the class of the UNCG tetraloops is given by the cUUCGg tetraloop (Figure 6). It is a structurally well-characterized system that was intensively investigated both by NMR-spectroscopy<sup>52</sup> as well as X-ray crystallography<sup>53</sup>. As a structural element in the large ribosomal subunit -namely of the 23S ribosomal RNA in *Haloarcula marismortui*<sup>54</sup> - it became clear that this kind of tetraloop is involved in long-range tertiary RNA-RNA interactions via its second unpaired uridine, which is able to bind an unpaired adenosine nucleotide remote in sequence. In the light of the three dimensional structure the remarkable stability can be explained as follows:

- Nucleotides uracile 1 and guanine 4 are forming a mismatch-pair because guanine 4 adopts the syn-conformation around the glycosidic bond. This pairing is stabilized by hydrogen bonds between the imino- and amino-group of guanine and the oxygen atom O2 of the uracile-base {imino-group: G4(N1) ···H ···(O2)U1, amino-group and G4(N2) ···H ···(O2)U1}. Furthermore, the pairing is stabilized by an unusual hydrogen bond between the hydroxyl-moiety of the ribose ring of uracile 1 and the base oxygen O6 of guanine 4. Astonishingly although the pair is classified as a mismatch, the Watson Crick side of guanine 4 is completely hydrogen bonded to its counterpart uracile 1.

<sup>49</sup>D.J. Proctor, J.E. Schaak, J.M. Bevilacqua, C.J. Falzone, P.C. Bevilacqua, *Biochemistry*, 2002, **41**, 12062-12075

<sup>50</sup>J. Koplin, Y. Mu, C. Richter, H. Schwalbe, G. Stock, *Structure*, 2005, **13**, 1255-1267

<sup>51</sup>V.P. Antao, I. Tinoco Jr., *Nucl. Acids Res.*, 1992, **20**, 819-824

<sup>52</sup>C. Cheong, G. Varani, I. Tinoco Jr., *Nature*, 1990, **346**, 680-682

F.H.-T Allain, G. Varani, *J. Mol. Biol.*, 1995, **250**, 333-353

<sup>53</sup>E. Ennifar, A. Niklin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, P. Dumas, *J. Mol. Biol.*, 2000, **304**, 35-42

<sup>54</sup>N. Ban, P. Nissen, J. Hansen, P.B. Moore, T.A. Steitz, *Science*, 2000, **289**, 905-920



- The base of the second nucleotide uracile 2 is flipped out and points towards the solvent. The ribose moiety of this nucleotide and of the following nucleotide (cytosine 3) adopts the C2'-endo conformation and thereby facilitates the 180° turn of the nucleic acid backbone.
- Between the nucleobases of cytosine 3 and uracile 1 there is an additional basestacking interaction that might be extended in the case of the cUUCGg tetraloops by additional stacking to the cytosine closing-base. Furthermore, a side-on hydrogen bond between the amino group and one of the phosphate oxygens  $\{C3(N4) \cdots H \cdots (^{proR}O)C3\}$  and one between the hydroxyl group and the base  $\{C3(O2') \cdots H \cdots (O2)C3\}$  are stabilising the structural arrangement.

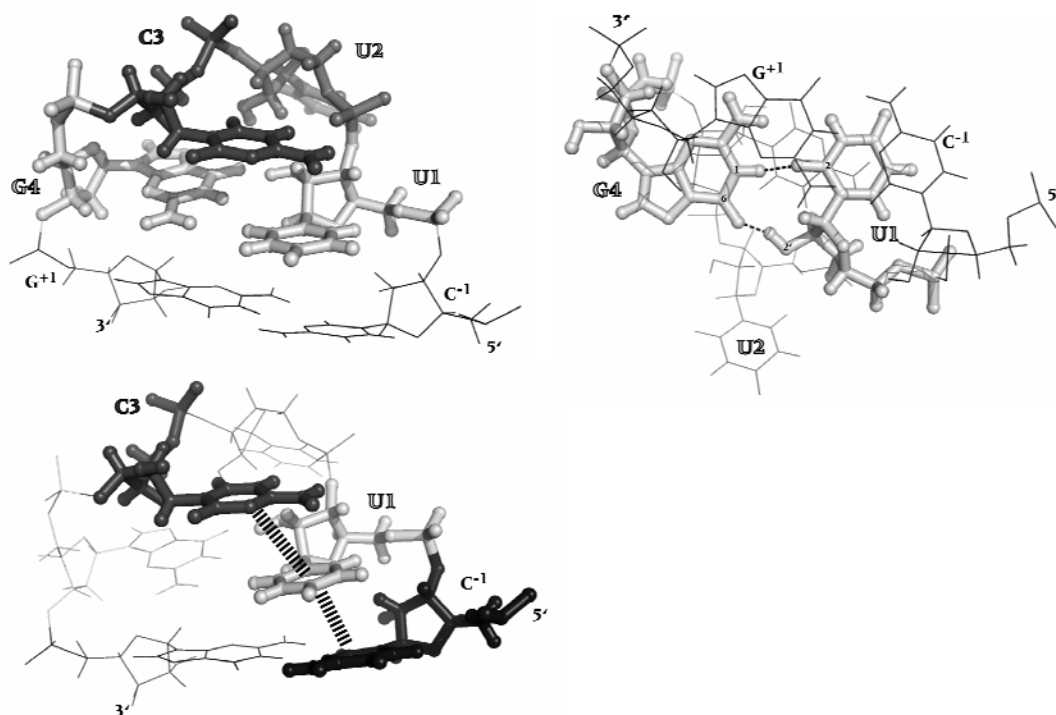


Figure 6: Three dimensional structure of an cUUCGg tetraloop as revealed by X-ray crystallography (PDB-Id.: 1F7Y): upper left: arrangement of the loop nucleotides, where the nucleotides U1 and G4 form a mismatch GU basepair (light grey), the bulged out nucleotide U2 pointing into the solvent is coloured in dark grey, the nucleotide C3 positioned above the mismatch basepair is shown in black grey, the closing basepair named C<sup>-1</sup> and G<sup>+1</sup> are indicated as black lines; upper right: bottom view of the cUUCGg tetraloop highlighting the mismatch between G4 and U1 that is stabilized by hydrogen bonds between  $G4(N1) \cdots H \cdots (O2)U1$  and  $C3(O2') \cdots H \cdots (O2)C3$ , furthermore the bulged out position of U2 becomes obvious; lower left: the favourable stacking interaction between the nucleotides C3(black grey), U1(light grey) and C<sup>-1</sup> is highlighted as indicated by the connecting lines.

The stability of the other stable family of tetraloops described as containing the consensus sequence GNRA of nucleotides can also be explained by the three dimensional conformation of the loops that facilitates enthalpic favourable interactions between the nucleotides involved in this structural arrangement (Figure 7).

In GNRA tetraloops, a network of heterogeneous hydrogen bonds is found as the central structural feature of these secondary structure elements. An extensive stacking of nucleobases within the loop furthermore complements these intramolecular interactions between several hydrogen bond donor and acceptor sides. As revealed by high resolution NMR structures of three different GNRA tetraloops, the major change in direction of the molecule is made between the first and the second nucleotide and therefore results in an asymmetric loop structure where

the first nucleotide -namely the nucleotide G- stacks on the 5'-side of the stem and the nucleotides two to four -N, R and A- stack onto the 3'-side of the helical stem. All of the nucleotides are found in the anti conformation and so a sheared anti-anti G-A base pair between the first and the last loop nucleotide is facilitated. In these non-canonical base pair, the minor groove side of the nucleotide G1 is paired via hydrogen bonds to the major groove side of the nucleotide A4. Interestingly, this base pair continues the stacking of the helix because A4 is moved towards the minor groove, shortening the phosphate-to-phosphate distance by 6.2Å compared to a canonical Watson-Crick G-C base pair.

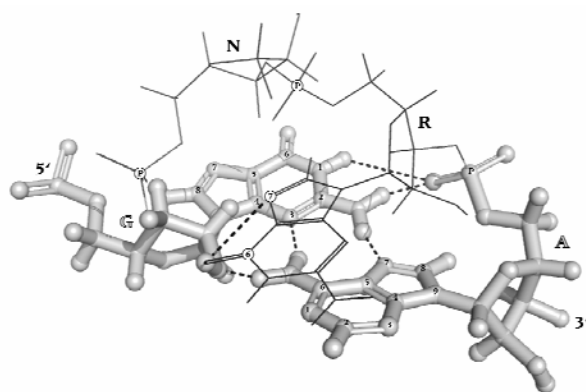


Figure 7: Three dimensional structure of an GNRA tetraloop as revealed by NMR spectroscopy (PDB-Id.: 1ZIF, 1ZIG, 1ZIH): the hydrogen bonds that are formed by the nucleotides G1, R3 (in the depicted case G3) and A4 are indicated as dashed black lines; the nucleotides forming the base pair are shown in stick-ball representation, whereas the nucleotides N/R are shown as black lines; numbers are indicating atom nomenclature; P indicate phosphorous atoms in the backbone, the base of nucleotide N2 is omitted for clarity.

A network of up to 7 hydrogen bonds stabilizing the loop surrounds the described central G-A pair. It should be mentioned that the loop hydrogen-bonding network is dynamic, it differs between individual conformations out of the structure ensemble generated by the NMR structure calculation and bonding pattern also varies among the loop structures studied. The variation may arise from limitations of the original NMR data, but may also just reflect a conformational fluctuation that is best described by a network of heterogeneous hydrogen bonds. This stacking of nucleobases plays a stabilizing role and is reflected not only in the extension of the helix by the G-A pair but also by the fact that with decreasing ability to perform stacking the order of the tetraloop is decreased.

### BISTABLE SEQUENCES

To reveal the structure-function relationship of RNA, it is likely that knowledge about the static structure, analogous to its minimum energy structure, is mostly not satisfactory. Dynamic aspects, including especially transitions at the level of 2<sup>nd</sup> structure, can be tremendously important to understand the native function of the RNA. Involvement of secondary structure transitions are crucial for adopting multiple stable conformations because the energies involved in 2<sup>nd</sup> structure formation are large compared to those of tertiary contacts and therefore dominating the minimum free energy of the structures involved. This is mainly due to the fact that breaking of basepairs -as major determinants of secondary structure- is costly in terms of energy. Consequently, the alternative conformations are separated by high energy-barriers but have

comparable energies. On the folding pathway of RNA molecules, the formation of non-native like secondary structure is also possible. If this happens, then major structural reorganisation of the folding chain becomes necessary in order to reach the ground state with only native contacts present.

The capability of natural RNA molecules to form multiple (meta)-stable conformations with different functions is used in nature to implement so called molecular switches, which regulate and control a number of biological processes. Some selected examples from the literature are listed in the following table.

Table 5: Biological examples of bistable RNA molecules

Function	Description	Citation
<b>Viroid replication</b>	During viroid replication and processing single RNA strands (either + or -) of a given sequence adopt different conformations and it is shown that energetically favored double-strand formation may at least partially be prohibited by stable secondary structures of the single strands.	R. Hecker, Z.M. Wang, G. Steger, D. Riesner, <i>Gene</i> , 1988, <b>72</b> , 59-74
	The formation of a thermodynamically metastable hairpin structure during synthesis of viroid replication intermediates is crucial for the infectivity of potato spindle tuber viroid RNA.	P. Loss, M. Schmitz, G. Steger, D. Riesner, <i>The EMBO Journal</i> , 1991, <b>10</b> , 719-727
	Computer simulations reveal that the formation of viroid RNA structure strongly depends on dynamics of competition between alternative RNA structures. Inter alia simulations of the PSTVd minus strand folding during transcription reveal a metastable hairpin, that is necessary to guide the folding towards a functional minus strand.	A.P. Gultyaev, F.H. van Batenburg, C.W. Pleij, <i>J. Mol. Biol.</i> , 1998 <b>276</b> , 43-55
<b>Gene regulation by attenuation</b>	The transcript of <i>E. coli</i> tRNA <sup>Phe</sup> synthetase can be folded into three alternative secondary structures, one of which is a site of transcription termination. This is a mechanism of gene attenuation.	G. Fayat, F.J. Mayaux, M. Springer, M. Grunberg-Manago, S. Blanquet, <i>J. Mol. Biol.</i> , 1983, <b>171</b> , 239-261
	Coordination of the expression of two tRNA <sup>Thr</sup> synthetase genes in <i>B. subtilis</i> is facilitated by alternative RNA structures.	H. Putzer, N. Gendron, M. Grunberg-Manago, <i>The EMBO Journal</i> , 1992, <b>11</b> , 3117-3127
	In <i>B. subtilis</i> attenuation of a trp gene is facilitated by a protein called TRAP (trp RNA-binding attenuation protein). It acts by binding to the nascent transcript and prevents formation of an alternative RNA antiterminator structure.	P. Babitzke, C. Yanofsky, <i>Proc. Natl. Acad. Sci.</i> , 1993, <b>90</b> , 133-137
<b>Protein synthesis at the ribosome</b>	The ribosomal domain in the 28S rRNA, in which the ribotoxins alpha-sarcin and ricin catalyse covalent modifications in adjacent nucleotides, has to adopt alternate conformations (which are a seven-base-pair helical stem and a 17-member single-stranded loop) because the RNA elements that are recognized by both toxins are not only different but incompatible. The alternative conformations are attributed to differential binding either to EF-Tu or EF-G.	I.G. Wool, A. Glück, Y. Endo, <i>Trends Biochem. Sci.</i> , 1992, <b>17</b> , 266-269
	A conformational switch in 16S ribosomal RNA affects tRNA binding to the ribosome and decoding of messenger RNA.	J.S. Lodmell, A.E. Dahlberg, <i>Science</i> , 1997, <b>277</b> , 1262-1267
	A central region in small ribosomal RNA switches between two helices in translation to maintain translational fidelity	U. von Ahsen, <i>Chem. Biol.</i> , 1998, <b>5</b> , R3-6

<b>Ribozymes</b>	A single RNA sequence can adopt two different conformations that either act as a ribozyme catalysing a ligation or cleavage reaction.	E.A. Schultes, D.P. Bartel, <i>Science</i> , 2000, <b>289</b> , 448-452
<b>Small ligand binding</b>	Engineered ligand-specific molecular switches composed of RNA act upon binding of the ligand to the receptor by triggering a conformational change within a bridging structural element and this structural reorganization dictates the activity of an adjoining ribozyme.	G.A. Soukup, R.R. Breaker, <i>Proc. Natl. Acad. Sci.</i> , 1999, <b>96</b> , 3584-3589
	Natural occurring Riboswitches (as regulators of gene expression upon ligand binding), detailed description in following chapter.	J. Miranda-Rios, M. Navarro, M. Soberon, <i>Proc. Natl. Acad. Sci.</i> , 2001, <b>98</b> , 9736-9741
<b>RNA thermometer</b>	RNA sequences known as RNA-thermometers adopt alternative conformations in response to temperature and acting thereby as transcriptional regulators.	F. Narberhaus, T. Waldminghaus, S. Chowdhury, <i>FEMS Microbiol. Rev.</i> , 2006, <b>30</b> , 3-16

As an exemple one of the bistable sequences used in the context of this thesis is described in further detail: In solution the sequence 5'-GAC CGG AAG GUC CGC CUU CC-3' captures two different conformations (Figure 8). One conformation (conformation B) adopts a hairpin structure, containing a six-basepair-helix and a UCCG-tetraloop, with a four nucleotide single-stranded overhang at the 5'-side of the RNA. In contrast the second conformation (conformation A) adopts a slightly shorter hairpin structure, containing a four base-pair-helix and a GGAA-tetraloop, with an eight nucleotide single-stranded overhang at the 3'-site of the molecule. The existence of the sequence in either of these two respective conformations can be explained by the small free-energy difference between them, which is predicted to be  $\Delta G=0.6\text{kcal mol}^{-1}$ . This difference<sup>55</sup> in free energy relates to a population relation of 3:1 for conformation B versus conformation A at 37°C.

As analysis with bioinformatics tools reveals<sup>56</sup>, the difference in the free energy is an inherent feature of the respective conformation - which is also consequentially logical because it cannot be a feature of the sequence itself: If calculation of the free energy decomposition is conducted, it becomes evident that the main difference in free energy of the two conformation arises from the different contributions of the stem-loop regions. The difference in the single strand helical regions is rather small but the main difference is observed for the base-paired regions where the difference is up to  $4\text{kcal mol}^{-1}$  (as side note it should be mentioned that the free energies that are predicted for the tetraloops are contradictory to all experimental data that is available for tetraloop).

<sup>55</sup>According to  $\Delta G = -RT \ln(K)$

<sup>56</sup>M. Zuker, *Nucl. Acids. Res.*, 2003, **31**, 3406-3415 (mfold program v3.2)

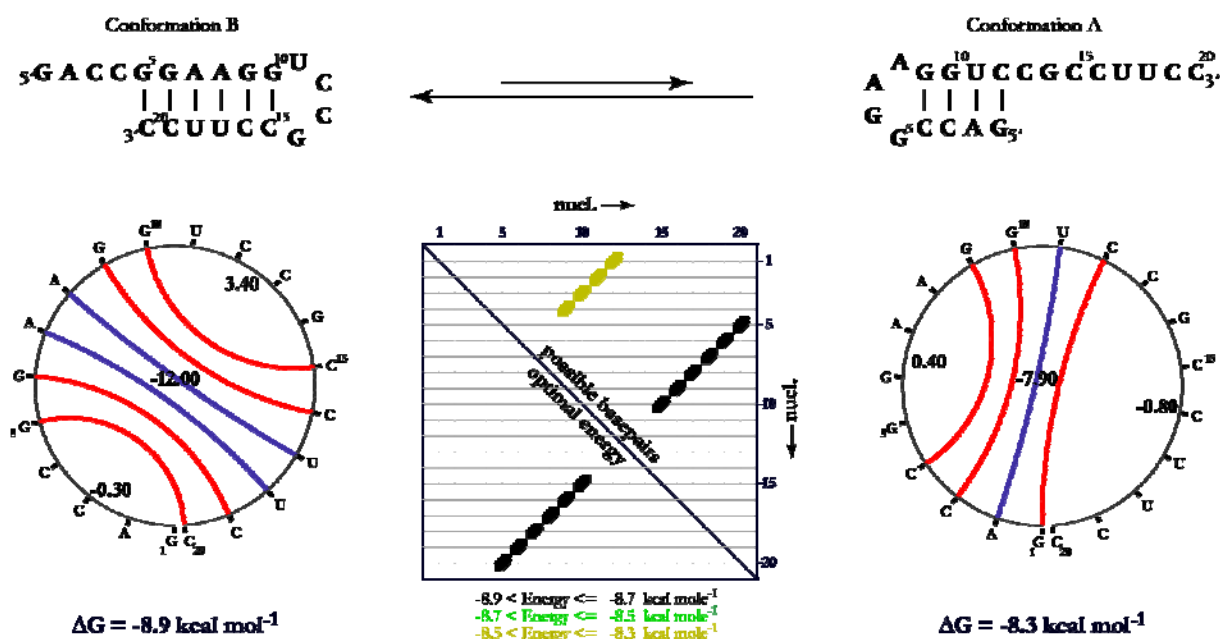


Figure 8: Representation of the secondary structure inversion for a given bistable RNA sequence. Upper panel: Secondary structures of the two populated RNA conformations. Lower panel left and right: Circular plots of the two conformations; red and blue lines indicating GC and AU basepairs, respectively, numbers are indicating the values of the energy contribution of the respective structural element (ss-region; hairpin-loop and regular A-helix) according to a loop-free energy decomposition. Lower panel mid: Dot-plot indicating the possible (upper triangle) and the energetically favoured basepairs (lower triangle) that can be adopted by the respective sequence.

In order to analyse and characterize the equilibrium of bistable RNAs, comparative studies can be undertaken that make use of truncated forms of the sequences.<sup>57</sup> These truncated forms can then just adopt one of the conformations. Experiments like UV-melting and NMR spectroscopy can be applied to the truncated and the full length constructs and the differences can be evaluated.<sup>58</sup> As an example, the <sup>31</sup>P 1D-spectrum and the imino region of the <sup>1</sup>H 1D-spectrum of both the full length and the two truncated hairpins are shown (Figure 9). The truncated forms with the sequences 5'-CGG AAG GUC CGC CUU CC-3' and 5'-GAC CGG AAG GUC C-3' represent either the stem-loop of conformation B or conformation A, respectively. Comparison of the spectra reveals that in the equilibrium form all structural features of the truncated forms have to be preserved, because the signals of the truncated and the full length form match. Secondly, the ratio of the two forms can be determined by a weighted deconvolution of the equilibrium spectrum by the spectra of the truncated hairpins. The imino-proton spectrum is of great advantage because it displays just protons involved in hydrogen bonds (basepairs). This advantageously represents long-range structural restraints helping to evaluate the structural elements and additionally the major difference between the truncated and the full length form - namely the single stranded overhang- remains silent (due to exchange -see following chapter).

<sup>57</sup>C. Höbartner, R. Micura, *J. Mol. Biol.*, 2003, **325**, 421-431

<sup>58</sup>R. Micura, W. Pils, C. Höbartner, K. Grubmayr, M.O. Ebert, B. Jaun, *Nucl. Acids Res.*, 2001, **29**, 3997-4005

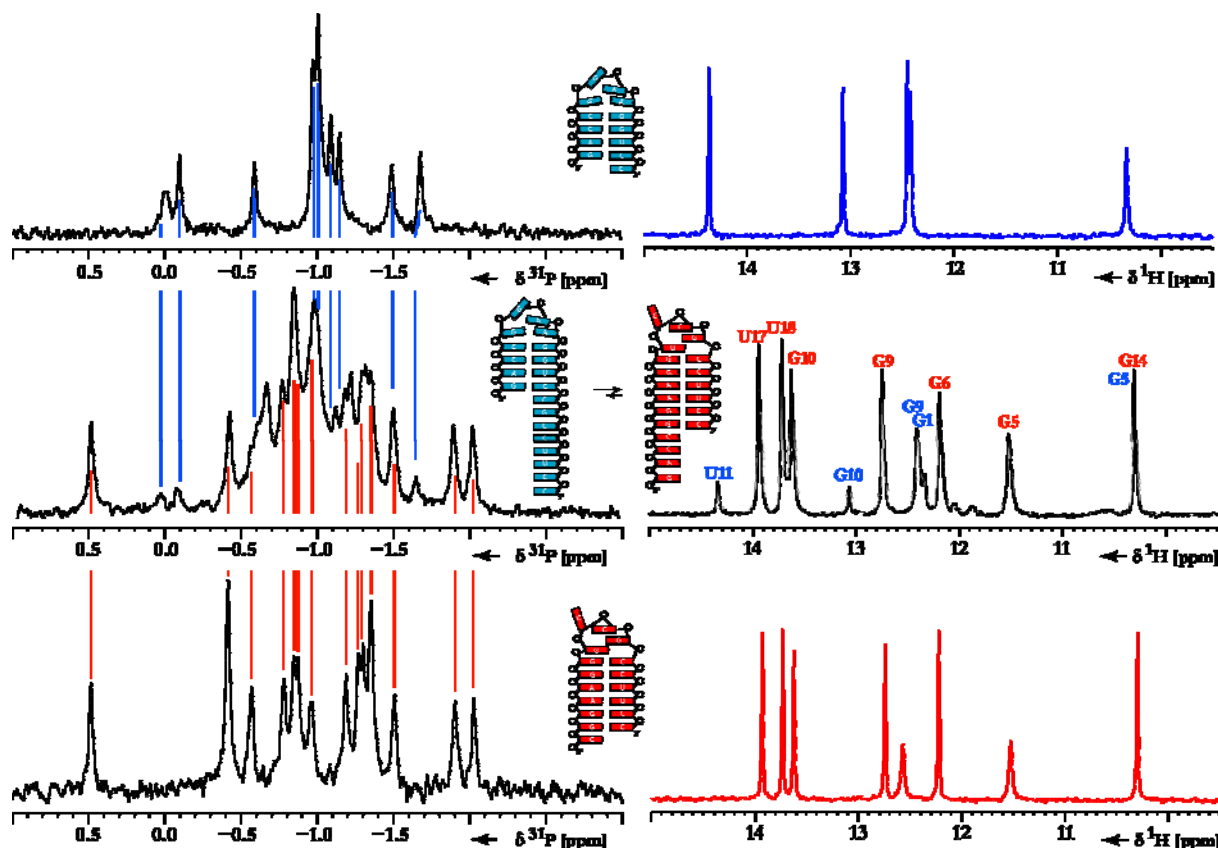


Figure 9: Comparative NMR data to characterize bistable RNAs: middle panel:  $^{31}\text{P}$  and  $^1\text{H}$  1D NMR spectra (imino region) of bistable 20mer RNA 5'-GAC CGG AAG GUC CGC CUU CC-3' that adopts the two conformations B (red) and A (blue); upper and lower panel:  $^{31}\text{P}$  and  $^1\text{H}$  1D NMR spectra (imino region) of truncated reference hairpins that adopt exclusively the respective hairpin-loop conformation and are not able to adopt the respective other conformation: 13mer (Conf. A; blue upper panel): 5'-GAC CGG AAG GUC C-3' and 17mer (Conf. B; red lower panel): 5'-CGG AAG GUC CGC CUU CC-3'; obviously the spectra of the truncated hairpins are restored in the full-length construct; the advantage of the proton spectra also emerges indicating that here just information of the base-paired region is visible and spectral crowding as in the case of the  $^{31}\text{P}$  spectra is avoided.  $^1\text{H}$ -1D spectra were recorded on a Bruker AV800MHz spectrometer equipped with a HCN- $\pi$ grad cryoprobe at 288K. Proton pulses were applied on resonance at the water frequency, hard proton pulses had a field strength of 26.3kHz. Spectra were recorded with 128scans and 4k complex points.  $^{31}\text{P}$ -1D spectra were recorded at 298K on a Bruker DRX600MHz spectrometer equipped with HCP- $\pi$ grad probe. The transmitter frequency for  $^{31}\text{P}$  was set to 0ppm. Proton pulses were applied on resonance at the water frequency. The decoupling sequence  $\text{waltz16}$  was applied with 3,1kHz. Spectra were recorded with 1024 scans and 32k complex points; the relaxation delay was set to 1.5s resulting in an experimental time of 38min.

An alternative approach to study the structural elements of bistable RNA molecules is the classical way of approaching RNA structures by NMR including resonance assignment and the analysis of spectral parameters. An example is given in the following figure (Figure 10) where hetero-nuclear NMR methods are used in conjunction with selective labelling of the RNA molecule. The full length 20mer-RNA sequence contains two  $^{15}\text{N}$ -labels at position N(3) of nucleotides C3 and position N(1) of nucleotide G10. Therefore two N1/H1 cross-peaks are expected in the  $^{15}\text{N}$  HSQC spectrum arising from the two basepairs in which G10 is involved (C3-G10: Conf. B and C15-G10: Conf. A). In contrast, the HNN-COSY spectrum contains two diagonal peaks at the chemical shift values as determined in the HSQC plus one additional cross-peak at the cross section for N1/H1 for G10 in conformation A. This arises solely because just in this respective conformation C3 as the binding partner for G10 contains a  $^{15}\text{N}$  label to entail a  $^2\text{hJ}_{\text{NN}}$  coupling over which magnetisation transfer is possible.

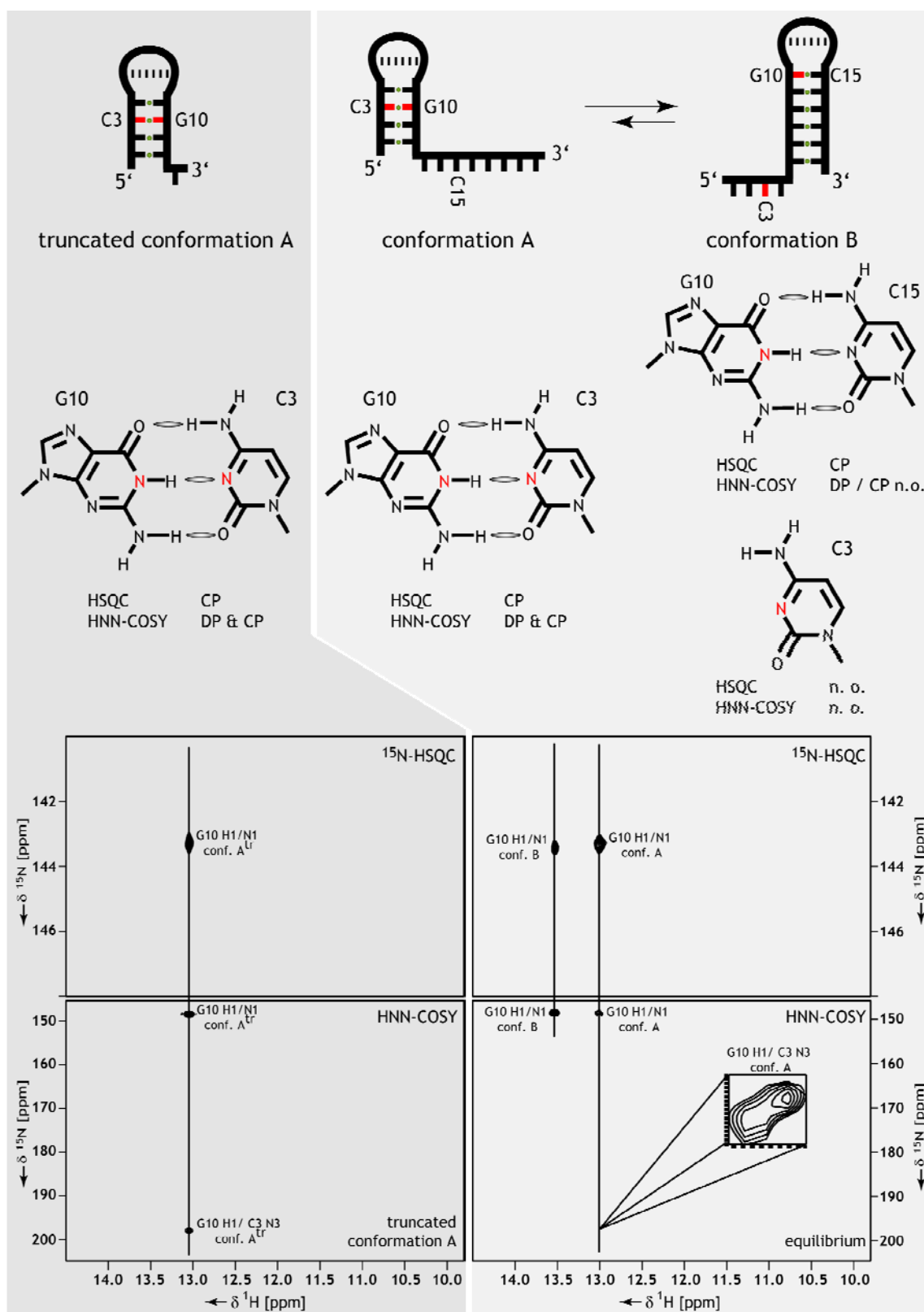


Figure 10:  $^{15}\text{N}$ -HSQC and HNN-COSY spectra of the imino proton region of the 20mer-RNA sequence containing two  $^{15}\text{N}$ -labels at N(3) of C3 and N(1) of G10 (positions as indicated). Left: truncated sequence adopting conformation A. right: equilibration, adopting conformation B and conformation A in a ratio of 5:1. In the  $^{15}\text{N}$ -HSQC- and the HNN-COSY-spectrum, all base paired  $^{15}\text{N}$ -labelled G10s produce a cross-peak, whereas in the HNN-COSY additional cross-peaks are observed only for the base pairs G10-C3, in which both nucleotides are  $^{15}\text{N}$ -labelled; due to the low equilibrium concentration, the additional cross-peak of is only visible after enhancement (as shown in the enlarged inset).

### Natural 2<sup>nd</sup> Structure Selection

If the existence of multi-stable conformations is an intrinsic feature of RNA and if each conformation is assigned to a respective function, how can structural plasticity be restrained?

As an example for such a natural structure selection loop 45 of the 3'-end of the 16S RNA is chosen. This loop caps the last helix of this rRNA and its sequence is highly conserved among all bacteria and eukaryotes.

Among these species not only sequence is conserved but also the complete primary structure, namely the methylation pattern<sup>59</sup> (Figure 11).

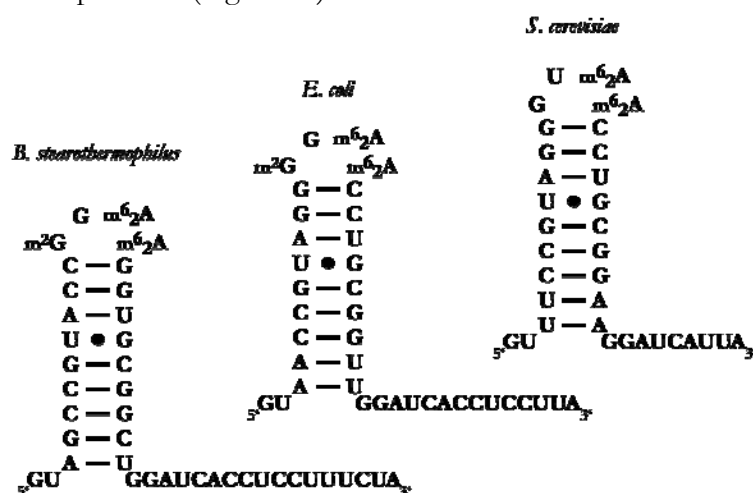


Figure 11: Secondary structure representation of sequences of helix 45 at the 3' end of the rRNA of the small subunit of the ribosome (From left to right for *Bacillus stearothermophilus*, *Escherichia coli* and *Saccharomyces cerevisiae*). The methylation pattern arising in the stem loops structures is annotated. Among the different species; the dimethylation of the nucleotides A is highly conserved.

Although sequence and methylation pattern were known for quite a while, the function of the helix itself and especially the one of the methylation pattern remained unclear.

Biochemical data suggested that the helix plays a critical role in the association process of the subunits and therefore interferes with initiation of protein synthesis. This was mainly due to the fact that an organism lacking adenine dimethylase, which is encoded in the *ksgA* gene and responsible for the transfer of two methyl groups from S-adenosyl-L-methionine to the N6 of A1518 and A1519 of the 16S RNA, shows resistance to the antibiotic kasugamycin.<sup>60</sup> This molecule from *Streptomyces kasugiensis* (this, for western ears, uncommon name of the antibiotic and of the bacterium arises from the Kasuga shrine in Nara City, where the bacterium was detected in the soil) specifically inhibits translation initiation of canonical but not of leaderless mRNA and inhibits formation of the full ribosome.<sup>61</sup>

<sup>59</sup>C. Höbartner, M.O. Ebert, B. Jaun, R. Micura, *Angew. Chem. Int. Ed.*, 2002, **41**, 605-607

<sup>60</sup>T.L. Helser, J.E. Davies, J.E. Dahlberg, *Nat. New. Biol.*, 1971, **233**, 12-14

T.L. Helser, J.E. Davies, J.E. Dahlberg, *Nat. New. Biol.*, 1972, **235**, 6-9

<sup>61</sup>F. Schlutzenzen, C. Takemoto, D.N. Wilson, T. Kaminishi, J.M. Harms, K. Hanagawa-Suetsugu, W. Szaflarski, M. Kawazoe, M. Shirouzu, K.H. Nierhaus, S. Yokoyama, P. Fucini, *Nature Struct. Mol. Biol.*, 2006, **13**, 871-878



As seen in the crystal structure<sup>62</sup> of an activated small ribosomal subunit, the terminal helix45 resides between the two helices H44 and H24 and is guiding to the 3'-end of the rRNA (Figure 12). The 3' segment is known to be highly adaptable and contains the region that pairs with the trigger (Shine-Dalgarno) sequence in the mRNA. The switch helix can undergo changes in its basepairing scheme; these may induce cooperative movements that lead to rearrangements of the platform, the shoulder and the head.

The NMR structure analysis revealed that in contrast to an unmodified GGAA tetraloop the sequence adopts a different conformation, which does not exhibit the hydrogen bonding and stabilization pattern known for the GNRA family.

This can be summarized as nature's attempt to select a certain RNA conformation out of a bistable system by introduction of a chemical modification at the interaction surface of specific nucleotides.

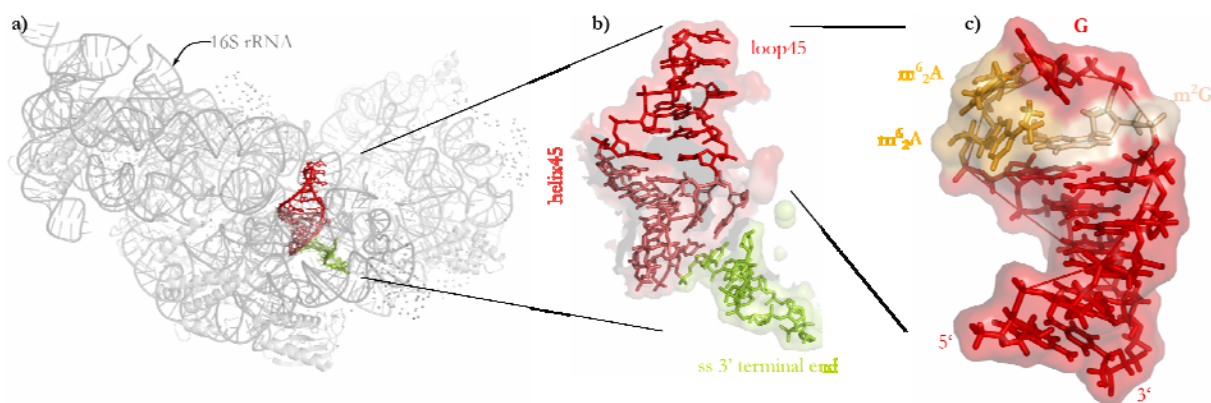


Figure 12: High resolution structures of helix45: a) complete structure of the functionally activated small ribosomal subunit of *T.thermophilus* at 3.3 Å resolution (pdb-code:1FKA<sup>62</sup>), the 16S-rRNA and the S-proteins are represented as cartoons except for the 3'-end of the RNA consisting of helix 45 (shown in red and orange) and the 3'-single stranded end (lime); b) enlargement of helix 45 (shown in red and orange) and the 3'-single stranded end (lime); c) methylated helix45 from *B.stearothermophilus* as solved by NMR (pdb-code: 1WTS), methylated residues are color coded (N2-methylguanosine shown in wheat; N6,N6-dimethyladenosine. shown in orange).

<sup>62</sup>F. Schlutzenzen, A. Tocij, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath, *Cell*, 2000, **102**, 615-623

### TERTIARY STRUCTURE

A higher level of structural order is described by the tertiary structure elements. Here the association of the secondary structural elements towards complex shaped molecules is abstracted. The interaction between the different secondary structural elements is thereby facilitated by various van der Waals contacts or by hydrogen bonds origination from additional Watson-Crick and unusual pairs of nucleobases.

“The parsing of energy levels between secondary and tertiary structures is reasonable in large RNAs, considering the relative energies and the clear identification of the secondary structure elements. In some cases, it is even possible to cut RNA molecules into modular domains which can reassociate only through tertiary contacts.”<sup>63</sup>

The tertiary structure elements are characterized by those interactions that comprise either two helices or two unpaired regions or one unpaired region and a double-stranded helix.

Interaction between two helices is provided by contiguous stacking of the two strands or by fitting the respective shallow groves into another. If two unpaired regions (which can be true single stranded regions or nucleotides originating from loops or bulges) interact, this can lead to the formation of pseudoknots (single loop is involved) or to loop-loop (multiple loops are involved) motifs. The interaction between unpaired regions and double-stranded helices may lead to various tertiary structure motifs such as triple helices or the A-minor motif.

Although, a comprehensive overview about RNA tertiary structure elements in the riboswitch field is given in the later part of this thesis (Schwalbe et al., 2007), the best-studied and most important elements are discussed in the following section:

### PSEUDOKNOT

Although frequently defined as secondary structure elements pseudoknots are typical representatives of tertiary structure elements (as defined above). A pseudoknot involves intramolecular pairing of bases in a hairpin loop with a few bases outside the stem of the loop to form a further stem and loop region. Consequently a true knot is formed, if each stem contains a full helical turn.<sup>64</sup>

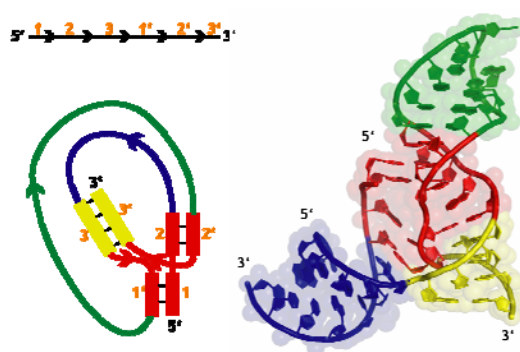


Figure 13: The left panel shows a schematic drawing of the RNA pseudoknot topology of the Diels-Alderase ribozyme, which adopts a 1-2-3-1'-2'-3' topology. The right panel shows the x-ray structure of the Diels-Alderase. In the same colouring scheme as the schematic drawing (pdb-code: 1YLS).

<sup>63</sup>E. Westhof, P. Auffinger, in *Encyclopedia of Analytical Chemistry*, 2000, 5222-5232, John Wiley & Sons Ltd, Chichester

<sup>64</sup>J.D. Puglisi, J.R. Wyatt, I. Tinoco Jr., *Nature*, 1988, **331**, 283-286

As example the structure of the Diels-Alderas ribozyme<sup>65</sup> (Figure 13) is highlighted that forms a double nested pseudoknot with an 1-2-3-1'-2'-3' topology (equal numbers indicate here double helical stretches; unprimed numbers indicating the primary structure sequential occurrence of the 5' parts of the stretches and primed numbers indicating the structure sequential occurrence of the 3' parts).

### U-TURN

A well-characterized and common RNA motif involved in tertiary interactions<sup>66</sup> is the so-called U-turn. It was first seen in the structure of tRNA<sup>Phe67</sup>, and is also prominent in the crystal structures of the hammerhead ribozyme active site<sup>68</sup>.

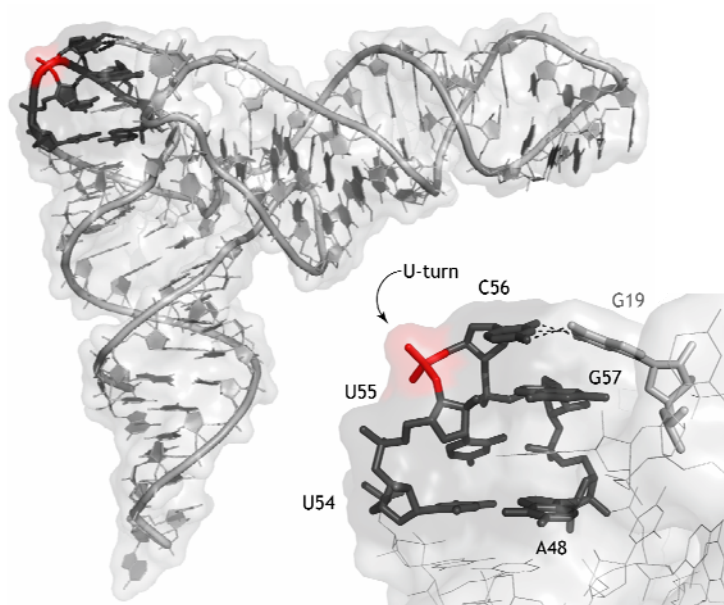


Figure 14: U-turn motif as seen in the tRNA<sup>Phe</sup> (4TRA); The nucleotides forming the U-turn motif are coloured in black, in a stick model, the turn in strand direction that is made around a backbone phosphate group is indicated by red colour. Left panel displays the whole molecule, the anticodon-loop pointing right, the CCA-end down; Right panel is a close-up of the U-turn motif, in the context of the tRNA<sup>Phe</sup> molecule it facilitates a long-range interaction between the nucleotide C56 directly 3' of the turn and the nucleotide G19 (also shown in stick representations, grey)

This stable structure element facilitates as the name is indicating a sharp change in the strand direction, and can be thereby understood as a way to close a RNA hairpin (Figure 14). Due to the fact that the motif exposes the nucleotides located at the 3' side of the turn towards the solvent, it creates an anchor for long-range tertiary interactions. Typically the sequential signature of a U-turn motif is 5'-UNR-3' and it is closed by Y-Y, Y-A or G-A pairs. As revealed by comparative sequence analysis the U-turn is a very abundant motif; it is identified 33 times in a large set of ribosomal RNAs (library of more than 7000 rRNA of different species)<sup>69</sup>.

<sup>65</sup>A. Serganov, S. Keiper, L. Malinina, V. Tereshko, E. Skripin, C. Höbartner, A. Polonskaia, A.T. Phan, R. Wombacher, R. Micura, Z. Dauter, A. Jäschke, D.J. Patel, *Nat. Struct. Mol. Biol.*, 2005, **12**, 218-224

<sup>66</sup>J. Doudna, *Structure*, 1995, **3**, 747-750

<sup>67</sup>J.L. Sussman, S.R. Holbrook, R.W. Warrant, G.M. Church, S.H. Kim, *J. Mol. Biol.*, 1978, **123**, 607-630

<sup>68</sup>H.W. Pley, K.M. Flaherty, D.B. McKay, *Nature*, 1994, **372**, 111-113, W.G. Scott, J.T. Finch, A. Klug, *Cell*, 1995, **81**, 991-1002

<sup>69</sup>R.R. Gutell, J.J. Cannone, D. Konings, D. Gautheret, *J. Mol. Biol.*, 2000, **300**, 791-803

## BASE TRIPLE

In the world of RNA structures diverse schemes of basepairing are found. This is due to the fact that a single nucleobase exhibits three main edges that are prone for interactions. Those are the already mentioned Watson-Crick edge (defined by the functional groups at atom positions R: 2, 1, 6 or Y: 2, 3, 4), the Hoogsteen edge (R: 6, 5, 7, 8 or Y: 4, 5) and the shallow groove edge (R: 2, 3, 4, 2' or Y: 1, 2, 2'). Additionally pairing diversity is increased by the possibility to form cis- and trans-basepairs, where the bases are at the same or the opposite side of a line median to the hydrogen-bonds. This structural plasticity of the basepairing behaviour is exploited in RNA-RNA recognition or tertiary interactions.

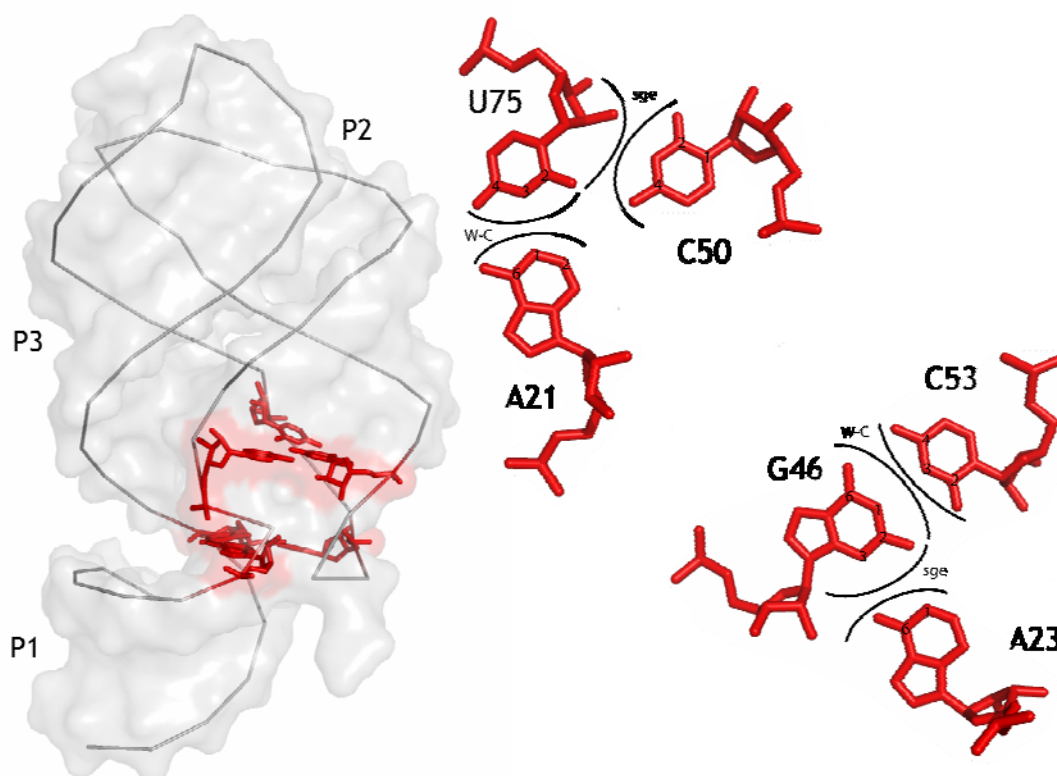


Figure 15: Schematic representation of base triples found in the structure of the riboswitch RNA (left panel, pdb-code: 1U8D), where they facilitate tertiary interactions. Two base triples are displayed in stick representation showing that these interactions are taking place as well at the Watson-Crick- (W-C) as on the shallow groove edge (sge).

Consequently arrangements of base triples are common features in RNA structures. Examples are the base triples G1091-C1100•G1071 and C1092-G1099•C1072 that bend a nine nucleotide loop in the L11 binding region of the 23S RNA – same tertiary motifs are found in the U2 snRNA<sup>70</sup> and in the loop region 715 of the 23S RNA. Frequently those base triples are found in the above mentioned RNA pseudoknots – for example in the ScYLV RNA facilitating the fold back of the knotted strands<sup>71</sup>. As shown in the following figure, also in case of the riboswitch RNA base triples are involved in tertiary interactions (Figure 15).

<sup>70</sup>C.S. Stallings, P.B. Moore, *Structure*, 1997, **5**, 1173-1185

<sup>71</sup>P.V. Cornish, D.P. Giedroc, M. Hennig, *J. Biomol. NMR*, 2006, **35**, 209-223

## A-MINOR MOTIF

One of the most interesting tertiary structure motifs in RNA is the A-minor motif<sup>72</sup> that is widely found for example in the hammerhead ribozyme<sup>68</sup>, in the P4-P6 domain of group I intron<sup>73</sup> and of course in the rRNA<sup>74</sup> (Figure 16). It describes a nucleotide-nucleotide interaction (stressing the fact that not only the bases, but also the ribose moieties are involved in this interaction).

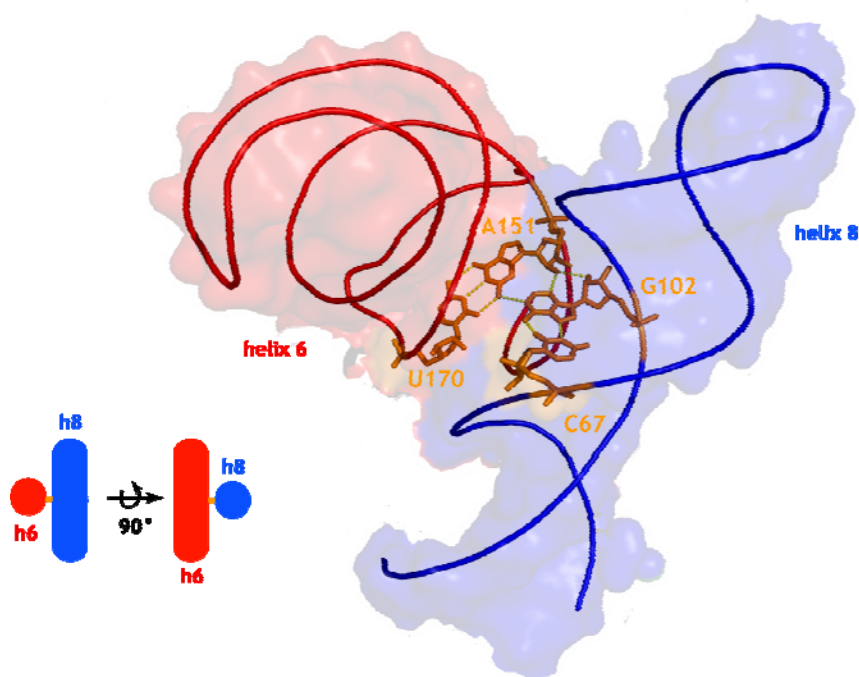


Figure 16: A-minor motif in the 16S ribosomal RNA; Structural representation of how a type II A-minor motif facilitates helix packing of the helices h8 (in blue colour) and h6 (in red colour); the nucleotides involved in the tertiary interaction are shown as sticks in orange, nucleotide A151 forms a type II A-minor motif by interacting with the basepair of G102-C67, A151 itself is additionally involved in a Hoogsteen type of interaction with U170.

The motif is defined as a single-stranded adenine that interacts via hydrogens bonds and van der Waals contacts at the minor groove of a given basepair (can be either AU or GC). The A-minor motif is classified into two types. Type I describes the interaction of the adenosine with both nucleotides involved in the basepair whereas type II represents just the interaction with one out of the two basepair forming nucleotides.

<sup>72</sup>P. Nissen, J.A. Ippolito, N. Ban, P.B. Moore, T.A. Steitz, *Proc. Natl. Acad. Sci.*, 2001, **98**, 4899-4903

<sup>73</sup>J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, J.A. Doudna, *Science*, 1996, **273**, 1678-1685

<sup>74</sup>H.F. Noller, *Science*, 2005, **309**, 1508-1514



## RNA FOLDING

## TYPES OF RNA FOLDING

The wide topic of RNA folding can be dissected into three different types (Figure 17): The first would include everything that is understood as *folding* itself, i.e. the mechanism of how a one dimensional sequence of nucleotides adopt its native three dimensional structure. The second type would regard the *refolding* of RNA structures, describing how a particular RNA conformation switches into another, yet different conformation, including changes on the level of secondary, and tertiary structure as well as a change in the function or functional state of the molecule. The third type is considering the *assembly* of supra-molecular RNA-structures or RNA-protein complexes as found in the native cellular context as for example ribosomes or spliceosomes.

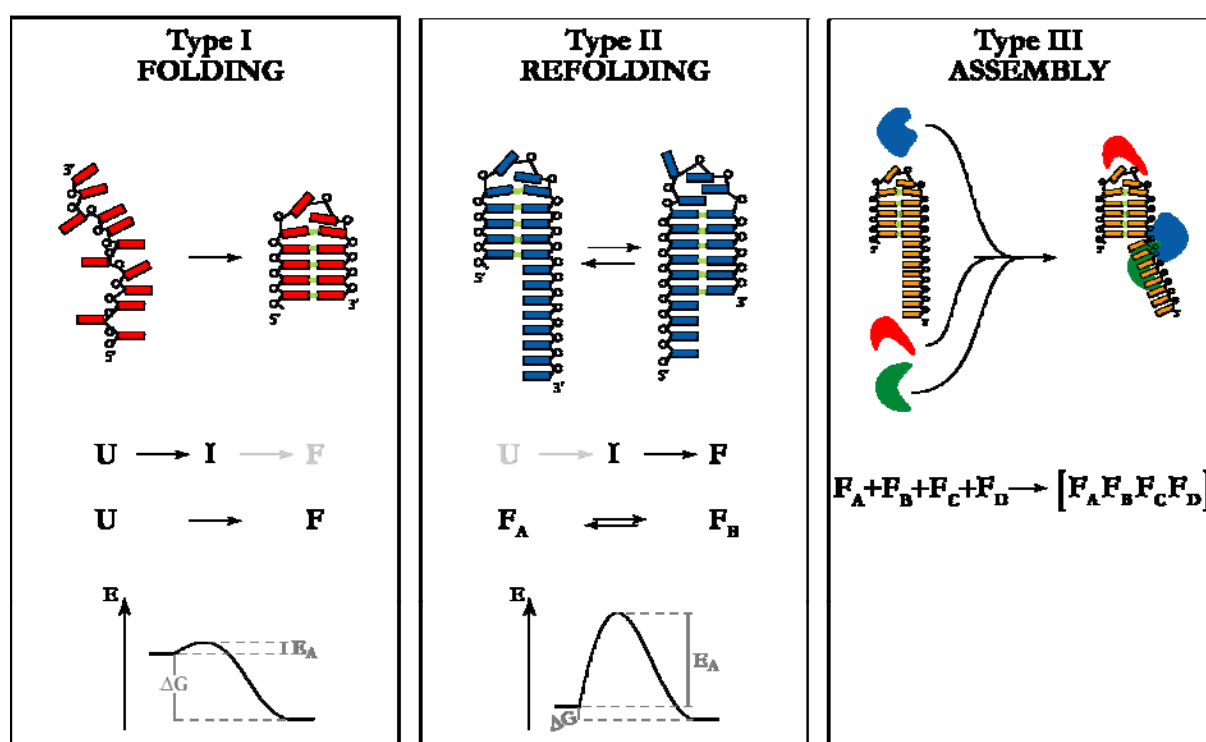


Figure 17: Schematic representation of the three types of RNA folding; left side: type I representing the folding events in their signification, starting at an unfolded state (U) and going either towards the folded/ native (F) state or towards a folded intermediate (I), these processes are characterized as fast and with low activation energy; middle: type II representing conformational switching events within a given RNA sequence that can be monitored either during the interconversion of two populated native conformations (F<sub>A</sub>, F<sub>B</sub>) or as the folding pathway from folded non-native intermediates towards the folded native conformation, such processes are characterized as slow and exhibit high activation energies; right side: type III representing the assembly of supra-molecular complexes of different RNAs, or RNA and protein molecules, this can be associated with conformational change, respectively folding, of either a single or all components, these processes display a wide range of kinetics and also energetic profiles. Notably the reactions of type III are of higher order than those of the first two types.

<sup>75</sup>O. Wilde, *The Chameleon*, 1894, 1, “Phrases and philosophies for the use of the young”



All these types of RNA folding events occur on a wide range of timescales. The fastest processes are those that can be categorised into the type of real *folding* events starting at an unfolded, respectively randomized nucleotide-chain. Under this precondition the folding of secondary structure elements occurs between  $\tau \sim 10\mu\text{s}$  and  $\tau \sim 100\mu\text{s}$ . The folding of tertiary structure elements is a slightly slower route and it is monitored so far on a timescale of milliseconds ( $\tau \sim 10\text{--}100\text{ms}$ ). Generally, the kinetics of helix formation seems to be independent of the exact nucleotide sequence and is fairly temperature dependent. The activation energy barrier is low. The very early findings were made on tRNA molecules. In thermal denaturation studies, the fastest process monitored was the folding-unfolding transition of the T-stem with  $\tau \sim 10\mu\text{s}$ , followed by the folding of the anticodon helix taking place with  $\tau \sim 100\mu\text{s}$ . The formation of the tertiary contact of the D and the T loop occurred on a timescale of a different order of magnitude and was with  $\tau \sim 10\text{ms}$  the slowest event. On an intermediate time range of  $\tau \sim 1\text{ms}$  the dissociation of the acceptor stem was reported. Though this event happens on the secondary structure level it nevertheless represents an unfolding event and therefore has to be categorized differently<sup>76</sup>.

In contrast, RNA-refolding events are monitored to be slower and characterized by a strong temperature dependence, reflecting a high energy-barrier between the two folded states. The first kinetic studies on such processes were conducted on the interconversion processes of folding intermediates –also referred to as folding traps– into the fully folded forms of the RNA molecules. This was for instance described for a partially folded tRNA<sup>lle</sup> molecule that folds upon  $\text{Mg}^{2+}$  towards the native conformation via a folded but non-native intermediate ( $F_{\text{part}} \rightarrow F_{\text{inter}} \rightarrow F_{\text{native}}$ ). The two processes are described by time constants (measured at  $10^\circ\text{C}$ ) of  $\tau(F_{\text{part}} \rightarrow F_{\text{inter}}) = 15\text{s}$  and  $\tau(F_{\text{inter}} \rightarrow F_{\text{native}}) = 150\text{s}$  with the respective activation energies of  $E_A(F_{\text{part}} \rightarrow F_{\text{inter}}) = 27\text{kcal mol}^{-1}$  and  $E_A(F_{\text{inter}} \rightarrow F_{\text{native}}) = 33\text{kcal mol}^{-1}$ <sup>77</sup>. Similarly the structural transition of the equilibrium between two alternative pseudo-knotted forms of a sequence from the *E.coli*  $\alpha$  mRNA ( $F_A \rightarrow F_B$ ) takes place with time constants of  $\tau(F_A \rightarrow F_B) = 4\text{s}$  and  $\tau(F_B \rightarrow F_A) = 8\text{s}$  with an activation barrier of  $E_A(F_A \rightarrow F_B) = 34\text{kcal mol}^{-1}$  and  $E_A(F_B \rightarrow F_A) = 46\text{kcal mol}^{-1}$  (measured at  $41^\circ\text{C}$ )<sup>78</sup>.

Likewise, the folding of the RNase P (which is a multiple turnover ribozyme that cleaves off an extra or precursor sequence on tRNA molecules) follows a similar folding route via a folded intermediate conformation ( $U \rightarrow F_{\text{inter}} \rightarrow F_{\text{native}}$ ). Here the transition between the intermediate and the native conformation is described by time constants ranging from  $\tau(F_{\text{inter}} \rightarrow F_{\text{native}}) = 0.1\text{s}$  at  $37^\circ\text{C}$  up to  $\tau(F_{\text{inter}} \rightarrow F_{\text{native}}) = 63\text{s}$  at  $3^\circ\text{C}$ <sup>79</sup>. The energy barrier between the folded intermediate and the native structure that has to be overcome during the transition is determined to be  $E_A(F_{\text{inter}} \rightarrow F_{\text{native}}) = 32\text{--}36\text{kcal mol}^{-1}$ .

<sup>76</sup>R. T. Batey, R. P. Rambo, J. A. Doudna, *Angew. Chem. Int. Ed.*, 1999, **38**, 2327–2343

<sup>77</sup>D. C. Lynch, P. R. Schimmel, *Biochemistry*, 1974, **13**, 1841–1852

<sup>78</sup>T. C. Gluick, R. B. Gerstner, D. E. Draper, *Journal of Molecular Biology*, 1997, **270**, 451–463

<sup>79</sup>P. P. Zarrinkar, J. R. Williamson, *Science*, 1994, **265**, 918–924



## MODEL MOLECULES TO STUDY RNA FOLDING

### RIBOZYMES

Up to now, ribozymes were used as model molecules to examine RNA folding reactions. Procedures like strand association, conformational reorientation of structural elements or whole RNA domains, chemical reactions and strand dissociation are underlying each ribozyme conversion. Most of these events can contribute to the understanding of how RNA folds and refolds and what is influencing its stability and dynamics. The advantage of ribozymes is that the educt and product state are distinguishable and that the rate of the ribozyme reaction under the influence of external stimuli and characteristic inhibitors can be easily measured. This is corresponding to the field of protein folding, where the early studies were also carried out on enzyme molecules, because the read out of the effect of different folding procedures was easy to establish by monitoring the catalytic reaction facilitated by the enzyme molecules.<sup>80</sup>

The close analogy of ribozyme to protein-enzyme reactions can be understood by comparing their properties:

- Ribozymes exhibit comparable catalytic rates as protein-enzymes for analogous reactions. For example the catalytic rate of the self-cleaving Hepatitis Delta Virus Ribozyme is determined to be  $k_{\text{cat}}=10^2\text{-}10^4\text{s}^{-1}$ , which is comparable to the rate of the protein-enzyme RNaseA displaying a rate of  $k_{\text{cat}}=1.4\cdot 10^3\text{s}^{-1}$ .
- The stability of the enzyme-substrate complexes is considerably large. For a hammerhead ribozyme the dissociation constant for establishing such a complex is in the nanomolar range  $K_D=10^{-9}\text{M}$ .<sup>81</sup>
- As first demonstrated for the RNase P, ribozyme molecules exist not only as single turnover but also as multiple-turnover catalysts.<sup>82</sup>
- As a central characteristic, ribozymes catalyse their respective reaction with high selectivity. The hammerhead, the HDV and the hairpin ribozyme cleave all a phospho-diester bond of the nucleic acid backbone, but act at specific points in a defined sequence context. This is due to several stabilizing interactions like basepairs and others, which facilitate a selective interaction between substrate sequence and ribozyme.
- To be able to form those mentioned interactions, ribozymes exhibit highly ordered tertiary structures.
- Ribozymes are able to facilitate catalysis by the usage of different co-factors like imidazole or several amino acids.<sup>83</sup>
- On ribozymes small molecules can act as specific allosteric effectors, in some cases it is also possible that these small molecules can switch the ribozyme molecules on and off (similar to riboswitches as delineated later).<sup>84</sup>

---

<sup>80</sup>M. Sela, F.H. White Jr., C.B. Anfinsen, *Science*, 1957, **125**, 691-692

<sup>81</sup>K.J. Hertel, D. Herschlag, O. Uhlenbeck, *Biochemistry*, 1994, **33**, 3374-3385

<sup>82</sup>C. Guerrier-Takada, K. Gradiner, T. Marsh, N. Pace, S. Altman, *Cell*, 1983, 147-157

<sup>83</sup>A.T. Perrotta, I. Shih, M.D. Been, *Science*, 1999, **286**, 123-126

<sup>84</sup>J. Tang, R.R. Breaker, *Chem. Biol.*, 1997, **4**, 453-459

G.A. Soukup, R.R. Breaker, *Proc. Natl. Acad. Sci.*, 1999, **96**, 3584-3589

## RIBOSWITCHES

Riboswitches serve as exemplary molecules where refolding of the RNA is directly linked to a change in the functional state.<sup>85</sup> Riboswitches are highly structured RNA genetic control elements found in the 5'untranslated region of messenger RNA that influence transcription termination or translation initiation by a conformational rearrangement of the RNA in response to direct metabolite binding. The riboswitch molecules are conceptually partitioned into two domains, the so-called aptamer and expression platforms. The aptamer domain directly binds the small molecule and thereafter undergoes a structural change. Since it is responsible for sensing the external stimulus with high affinity and selectivity it is the most conserved region in riboswitches.<sup>86</sup> These structural changes also affect the expression platform, which is then the mechanism by which the gene expression is regulated. Compared to the aptamer domains the expression platforms vary widely in sequence and structure.<sup>87</sup> Ligand binding leads therefore either to transcription termination by the formation of transcription terminating hairpins, abolishing ribosome binding by sequestering the ribosome binding site or by self-cleavage (=metabolite-responsive ribozyme<sup>88</sup>).

A wide range of riboswitches is described so far, regulating a great variety of genes and itself being regulated by diverse ligands (Table 6). They are found in bacteria, fungi and eukaryotes. The following table is by far not complete but aims to give a flavour how diverse the ligands, genes and organisms are, and to point out, that riboswitches, as bifunctional RNA molecules, are a versatile and evolutionary old mechanism of gene regulation.

The guanine riboswitch binds with high affinity and specificity to guanine ( $K_D=0.004\mu\text{M}\pm 0.003\mu\text{M}$ ) and hypoxanthine ( $K_D=0.759\mu\text{M}\pm 0.066\mu\text{M}$ ) and is among the smallest riboswitches described so far.<sup>89</sup> The X-ray-structure of its aptamer domain in complex with guanine/hypoxanthine reveals an intricate RNA-fold consisting of a three-helix junction stabilized by long-range base pairing interactions.<sup>90</sup> Binding of the ligands guanine and hypoxanthine is facilitated in a narrow binding pocket in the core of the riboswitch by a network of hydrogen bonds. These hydrogen bonds arising from conserved residues of the RNA molecule screen the Watson-Crick site and the N3/N9 edge of the ligand ensuring for selectivity for the right purine type nucleobase and against nucleotides.<sup>91</sup>

The conformational transitions of the aptamer domain induced by binding of hypoxanthine can be analyzed by high-resolution NMR-spectroscopy.<sup>92</sup> The long-range base pairing interactions are already present in the free RNA and preorganize its global fold. The ligand binding core region is lacking hydrogen bonding interactions and is therefore likely to be unstructured in the absence of

<sup>85</sup>J. Miranda-Rios, M. Navarro, M. Soberon, *Proc. Natl. Acad. Sci.*, 2001, **98**, 9736-9741

<sup>86</sup>M. Mandal, R.R. Breaker, *Nat. Struct. Mol. Biol.*, 2004, **11**, 29-35

<sup>87</sup>J.K. Soukup, G.A. Soukup, *Curr. Opin. Struct. Biol.*, 2004, **14**, 344-349

<sup>88</sup>W.C. Winkler, A. Nahvi, A. Roth, J.A. Collins, R.R. Breaker, *Nature*, 2004, **428**, 281-286

<sup>89</sup>S.D. Gilbert, C.D. Stoddard, S.J. Wise, R.T. Batey, *J. Mol. Biol.*, 2006, **359**, 754-768

<sup>90</sup>R.T. Batey, S.D. Gilbert, R.K. Montange, *Nature*, 2004, **432**, 411-415

<sup>91</sup>J. Noeske, C. Richter, M.A. Grundl, H.R. Nasiri, H. Schwalbe, J. Wöhnert, *Proc. Natl. Acad. Sci.*, 2005, **102**, 1372-1377

<sup>92</sup>J. Noeske, J. Buck, B. Fürtig, H.R. Nasiri, H. Schwalbe, J. Wöhnert, *Nucl. Acids Res.*, 2007, **35**, 572-583

ligand. The following statement summarizes best the binding of the ligand to the riboswitch RNA: purine ligands replace tertiary contacts and thereby force the RNA-molecule to fold in a compact structure.<sup>93</sup>  $Mg^{2+}$ -ions are not essential for ligand binding and do not change the structure of the RNA-ligand complex but stabilize the structure at elevated temperatures. A mutant RNA was identified where the long-range base pairing interactions are disrupted in the free form of the RNA but form upon ligand binding in a  $Mg^{2+}$ -dependent fashion. The tertiary interaction motif is stable outside the riboswitch context.

Table 6: Table of selected riboswitches representing different ligands, organisms and regulation mechanisms; \*the theophylline riboswitch is an engineered riboswitch that can be used as a biochemical tool to regulate the expression of different genes in *B. subtilis*

Ligand	Gene & Organism	Regulation mechanism	Literature
Coenzyme B12	<i>btuB</i> in <i>Escherichia coli</i>	ribosome binding	A. Nahvi et al., <i>Chem. &amp; Biol.</i> , 2002, <b>9</b> , 1043-1049
FMN (Flavin mononucleotide)	<i>ribDEAHT</i> in <i>Bacillus subtilis</i>	transcription termination	W.C. Winkler et al., <i>Proc. Natl. Acad. Sci.</i> , 2002, <b>99</b> , 15908-15913
	<i>ypaA</i> in <i>Bacillus subtilis</i>	ribosome binding (sequestration of Shine Dalgarno sequence)	
Glycine	<i>gcvT</i> in <i>Vibrio cholerae</i>	transcription control	M. Mandal et al., <i>Science</i> , 2004, <b>306</b> , 275-279
Theophylline	any* in <i>Bacillus subtilis</i>	ribosome binding	B. Suess et al., <i>Nucl. Acids Res.</i> , 2004, <b>32</b> , 1610-1614
Guanine	<i>xpt-pbuX</i> in <i>Bacillus subtilis</i>	transcription control	M. Mandal et al., <i>Cell</i> , 2003, <b>113</b> , 577-586
Thiamine pyrophosphate	<i>thiM</i> in <i>Escherichia coli</i>	ribosome binding	W.C. Winkler et al., <i>Nature</i> , 2002, <b>419</b> , 952-956
	<i>thiC</i> in <i>Arabidopsis thaliana</i>	mRNA stability	N. Sundarasan et al., <i>RNA</i> , 2003, <b>9</b> , 644-647
S-adenosyl methionine	<i>metF-metH2</i> in <i>Thermoanaerobacter tengcongensis</i>	transcription control	W.C. Winkler et al., <i>Nat. Struct. Biol.</i> , 2003, <b>10</b> , 701-707
Glucosamine-6-phosphate	<i>glmS</i> in <i>Thermoanaerobacter tengcongensis</i>	RNA cleavage (ribozyme)	W.C. Winkler et al., <i>Nature</i> , 2004, <b>428</b> , 281-286

Experiments on ligand induced folding of the structurally kindred adenine riboswitch aptamer domain (just a single nucleotide is replaced) reveal that folding in the presence of 10mM  $Mg^{2+}$  (representing a  $10^5$  excess compared to RNA concentration) occurs with half-life times between ~5s (ligand/RNA=5/1) and ~16s (ligand/RNA=0.25/1). The riboswitch folding reaction is determined to be exergonic ( $\Delta G = -10.73 \text{ kcal mol}^{-1}$ ) and exothermal ( $\Delta H = -40.3 \text{ kcal mol}^{-1}$ ) with a considerable loss of entropy ( $\Delta S = -97.6 \text{ cal mol}^{-1}$ ).

As further revealed by FRET studies, the control of gene expression depends on the rates of the ligand-assisted folding and on the rate of RNA synthesis as an external cellular factor.<sup>94</sup>

<sup>93</sup>A. Lescoute, E. Westhof, *Chemistry & Biology*, 2005, **12**, 10-13

<sup>94</sup>J.-F. Lemay, J.C. Penedo, R. Tremblay, D.M.J. Lilley, D.A. Fontaine, *Chem. & Biol.*, 2006, **13**, 857-868

### RNA FOLDING IN THE CELLULAR CONTEXT

Viewing the topic of RNA folding in the cellular or biological context, respectively, two considerations emerge; (1) the time window for RNA-folding, -refolding and -assembly is mainly restrained by the time (duration) of transcription and (2) the major task during RNA folding is to avoid non-native structures formed from sequentially nearby nucleotide stretches (Figure 18).

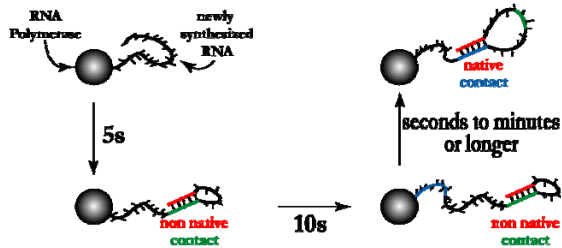


Figure 18: Schematic representation of RNA synthesis at the RNA polymerase demonstrating that the time window for RNA folding is generated by the speed of the transcription reaction and according to the ability of RNA to fold and form basepairs even before complete synthesis of the whole RNA molecule. One major task is to avoid non-native contacts or allowing the system to fold properly on longer timescales.

The most important step considering RNA folding is elongation, during which the emerging RNA strand has to fold and to undergo structural changes in order to adopt its native conformation. The elongation rate (the major determinant of the speed of transcription) is therefore also an affector of RNA folding.

The event of transcription is basically divided into three distinct steps. Initiation is regarded as the step in which the polymerase binds to the promoter region of the DNA template strand. During the elongation the RNA gets synthesised in a processive manner. The third and last step is termination where synthesis is ended, transcription is halted and the newly synthesised RNA is released after dissociation of the transcription-complex from the DNA template.

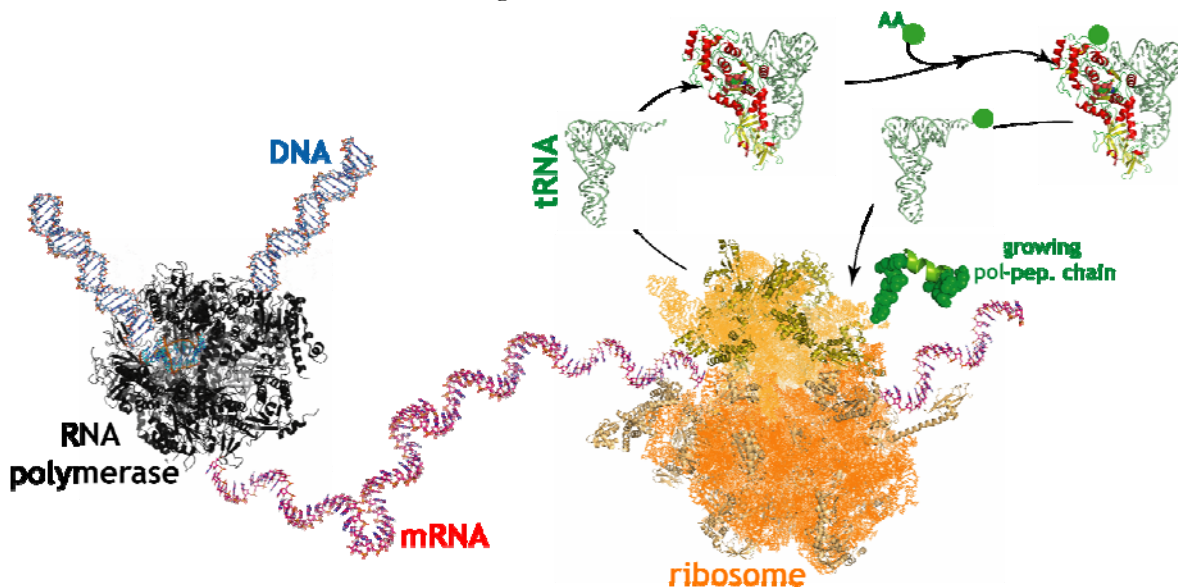


Figure 19: Opening a time window for (m)RNA folding I: Representation of the protein synthesis cycle in the cell. DNA gets transcribed into RNA by the polymerase, RNA is further translated into protein by the ribosome that in turn is being loaded by tRNA molecules that get amino-acylated by the respective enzyme. The different reaction rates of these key players determine the amount of 'free' RNA present in the cell and being able to fold into complex structures.

The transcription rate determines the window in a concerted interplay with the rate of translation (Figure 19). The reason causing this relation between transcription and translation is the fact that the nascent RNA chain is directly screened and processed by ribosomes (as evident by EM

pictures of RNA strands that are fully loaded by stretches of ribosomes, and are called polysomes).<sup>92</sup> Assuming a rate of transcription between 20-80nt s<sup>-1</sup> (for bacterial polymerases, and ~50nt s<sup>-1</sup> for *E.coli*)<sup>93</sup> and the ribosomal velocity to be 10pep-bnd s<sup>-1</sup><sup>94</sup> (concerning the rate of translation, the dissociation of EF-Tu GTP is determined to be the rate limiting step of protein synthesis at the ribosome<sup>95</sup>). This opens the route to have up to several hundred nucleotides of RNA being transcribed but being still untranslated, which then can possibly fold into different native conformations (Figure 20).

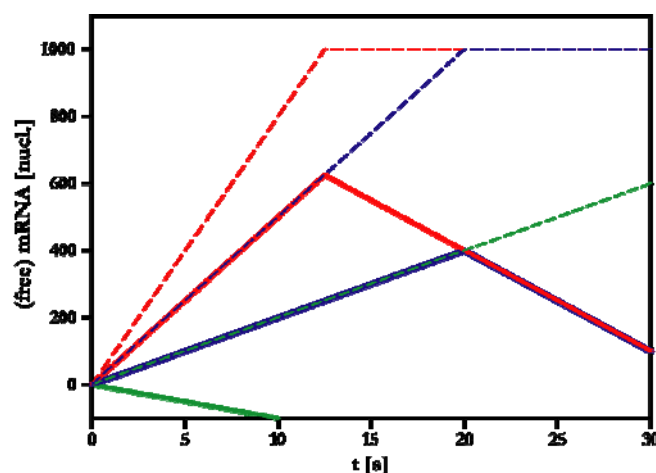


Figure 20: Opening a time window for (m)RNA folding II: Graphics showing the synthesis rate of RNA (dashed lines) and the amount of free RNA (solid lines) assuming a gene of 1000nucleotides in total and a transcription speed of 20nt s<sup>-1</sup> (green), 50nt s<sup>-1</sup> (blue) and 80nt s<sup>-1</sup> (red), the velocity of translation at the ribosome is set to 10pep-bnds s<sup>-1</sup> (=30nt s<sup>-1</sup>). The negative values correspond to the case where the newly synthesized RNA gets directly translated into peptides and therefore does not exhibit free nucleotides.

Not only the actual rates of translation and transcription are important for the generation of a suitable time window for RNA folding, but also transcriptional pausing sites are elongating the folding time frame and open the route for interactions of the nascent RNA chain with folding influencing proteins such as RNA-chaperones or the polymerase itself.

How pausing sites and proteins can alter the folding pathway of RNAs is demonstrated for RNase P folding in conjunction with the NusA protein (Figure 21). The folding of this RNA molecule that exhibits two domains (domain 1 is the specificity domain consisting of 154 nucleotides and domain 2 is the catalytic domain with 255 nucleotides) in the absence of the protein follows a two step process where first an intermediate is formed that consists of a natively folded catalytic domain and a non-native, respectively unfolded, specificity domain and then proceeds towards the native completely folded form. The folding event is then completed after ~400s, and the intermediate is maximal populated at ~100s. The folding-rate is independent of the general transcription speed that was varied from 3-200nt s<sup>-1</sup>. Upon addition of NusA not only the total velocity of the reaction increases and is completed after ~200s but also the folding pathway is altered. It is still represented by a two step process but the intermediate is different since now the specificity domain adopts its native conformation first, while the catalytic domain

<sup>92</sup>L. Stryer, *Biochemistry*, 1988, W.H. Freeman and Company, New York

<sup>93</sup>T. Pan, T. Sosnick, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 161-175

<sup>94</sup>M.A. Sorensen, S. Pedersen, *J. Mol. Biol.*, 1991, **222**, 265-280

<sup>95</sup>T. Pape, W. Wintermeyer, M.V. Rodnina, *EMBO J.*, 1998, **17**, 7490-7497

folds hereafter. This can be explained by the effects of NusA that alters the retention at the pausing sites and thereby decreasing the transcription speed but now at distinct positions on the synthetic pathway. So to say the NusA-induced pausing at a site in the catalytic domain prevented the formation of this non-native interaction, allowing more rapid folding of the specificity domain.

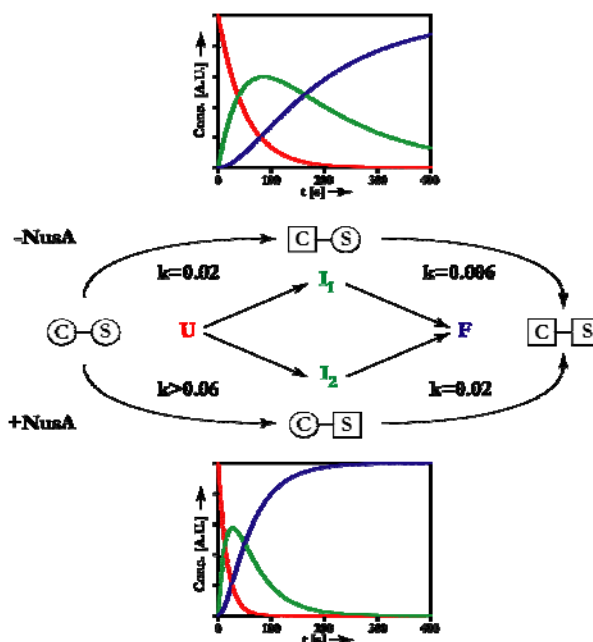


Figure 21: Schematic representation of the folding pathways of RNase P in the absence (upper panel) and presence (lower panel) of the NusA protein. The domain (C: catalytic domain, S: specificity domain) architecture of the RNase P is symbolically represented with circles and squares, where circles represent unfolded and squares folded domains. The time course of the population for each conformation (U: unfolded—,  $I_{1,2}$ : intermediates—, F: native/folded—) is given according to a simple two step process and calculated for the experimentally derived rates as indicated.

Furthermore, this and also mutational studies on the *Tetrahymena* groupI intron, where circular permutations of RNA structural elements were applied, indicate that the order of synthesis is also a determinant of the RNA folding velocity and mechanism.<sup>96</sup>

### MONITORING RNA FOLDING

As emphasised above, RNA folding occurs on a wide time-range and therefore also a wide range of methods have to be applied in order to study these processes (Figure 22).

For the initiation of folding reaction mainly three methods are applied: pH-<sup>97</sup>, temperature-<sup>98</sup> and ionic-strength changes.

<sup>96</sup>S.L. Heilman-Miller, S.A. Woodson, *RNA*, 2003, **9**, 722-733

<sup>97</sup>M. Bina-Stein, D.M. Crothers, *Biochemistry*, 1975, **14**, 4185-4191; J.H.A. Nagel, A.P. Gulyaev, K.J. Öistämö, K. Gerdes, C.W.A. Pleij, *Nucleic Acids Res.*, 2002, **30**, e63

<sup>98</sup>M. Menger, F. Eckstein, D. Porschke, *Biochemistry*, 2000, **39**, 4500-4507; D. Porschke, M. Eigen, *J. Mol. Biol.*, 1971, **62**, 361-381

For the very fast processes like the hairpin-formation that happen on  $\mu\text{s}$ -scale, only the T-jump method is fast enough to allow deciphering the course of reaction. Folding events happening on slower timescales of milli-seconds to seconds can be initiated by salt- or pH-changes in conjunction with the application of a stopped-flow apparatus or manually.

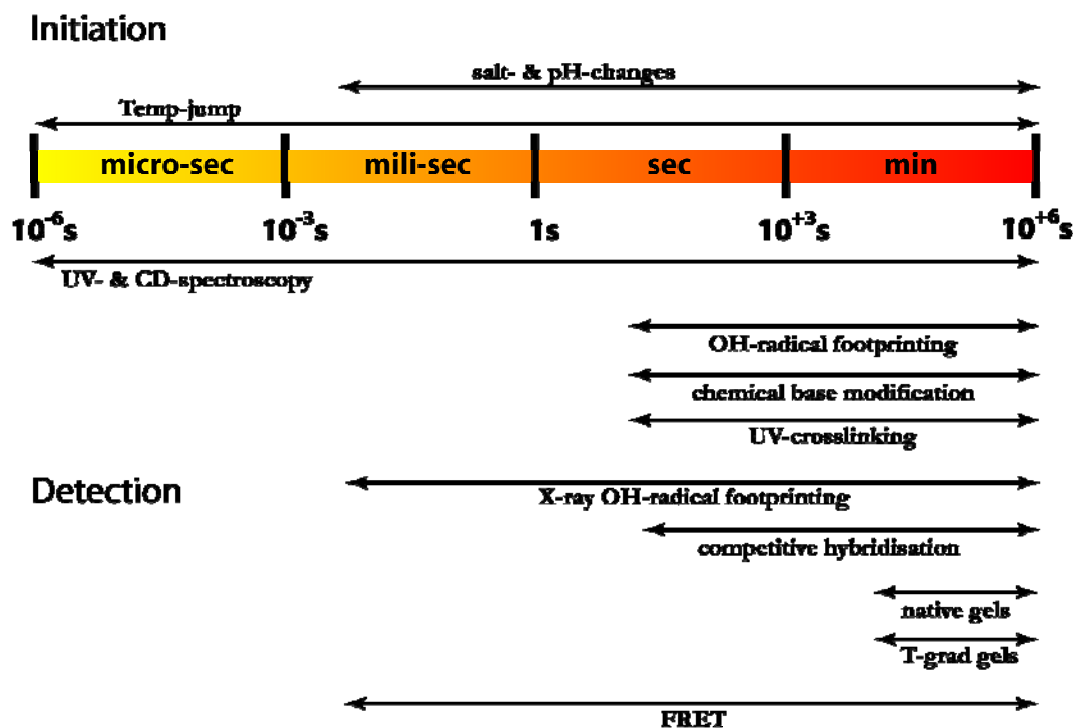


Figure 22: Initiation and detection techniques that cover (partially) the range of the RNA folding timescales

Methods for detection RNA folding are UV-spectroscopy, which is also applicable to fast events (for reactions with rates up to  $10^{-6}\text{s}^{-1}$ ), and CD-spectroscopy.<sup>99</sup> For slower reactions on the range of seconds to minutes three major biochemical methods are used to analyse the structural transitions and to decode the folding pathway; these are TR-OH-radical footprinting<sup>100</sup> and chemical base modification<sup>101</sup> -both methods probe the solvent accessibility of nucleotides- and UV-crosslinking<sup>102</sup> -this method allows the tracing of structural packing events-. All three methods need high resolution gel electrophoresis to analyse the fragmentation or modification pattern.

Higher temporal resolution of up to 10ms can be reached by X-ray OH-radical footprinting providing “snapshots” of RNA tertiary interactions. Here, as in the above mentioned footprinting method the solvent accessibility of the RNA backbone is determined from its susceptibility to cleavage in the presence of hydroxyl radical. The hydroxyl radicals are generated by 10-30 ms exposures to a synchrotron X-ray beam. The fragmentation pattern of the RNA is

<sup>99</sup>T. Pan, T.R. Sosnick, *Nat. Struct. Biol.*, 1997, **4**, 931-938

<sup>100</sup>D.W. Celander, T.R. Cech, *Science*, 1991, **251**, 401-407; S.G. Chaulk, A.M. MacMillan, *Biochemistry*, 2000, **39**, 2-8

<sup>101</sup>A.R. Banerjee, D.H. Turner, *Biochemistry*, 1995, **34**, 6504-6512

<sup>102</sup>M.E. Harris, J.M. Nolan, A. Malhotra, J.W. Brown, S.C. Harvey, N.R. Pace, *EMBO J.*, 1994, **13**, 3953-3963

determined by sequencing. This provides information about the parts of the RNA that have been folded in the elapsed interval since the start of the folding reaction.

With a temporal resolution of 0.5 to 1min competitive hybridisation with subsequent RNase H digestion allows also monitoring of structural transitions during RNA folding.<sup>79</sup>

If the structural transition is as slow, that it is completed not until several minutes or hours, native gel electrophoresis is applicable -here, the second precondition would be a sufficient shift of the distinct conformations. Nowadays, temperature gradient gels represent a further development of this technique.<sup>103</sup>

FRET is a further assay to measure the structural dynamics of RNA during folding in real-time; it is possible to get temporal resolution sufficient to monitor events occurring on a milli-second timescale. In these kinds of assays, a pair of fluorescent donor and acceptor is attached to the host RNA. The energy transfer efficiency between the donor and acceptor depends on their intermolecular distance, making FRET sensitive to conformational changes of the RNA molecule.<sup>104</sup>

---

<sup>103</sup>M. Nakano, E.M. Moody, J. Liang, P.C. Bevilacqua, *Biochemistry*, 2002, **41**, 14281-14292

<sup>104</sup>X. Zhuang, L. Bartley, H. Babcock, R. Russell, T. Ha, D. Herschlag, S. Chu, *Science*, 2000, **288**, 2048-2051



RELIGIONS DIE WHEN THEY ARE PROVED TO BE TRUE.  
SCIENCE IS A RECORD OF DEAD RELIGIONS.<sup>105</sup>

## NMR METHODS TO QUANTIFY DYNAMICS

NMR spectroscopy is the only experimental technique available that provides dynamic information of bio-macromolecules at atomic resolution. A wide range of experimental parameters is available describing the dynamic events at various timescales covering in total five orders of magnitude (Figure 23).

According to the timescale covered, the techniques can be dissected into four major classes:

- A) Spin relaxation
- B) Structural parameters (averaging)
- C) Chemical exchange
- D) Real time NMR

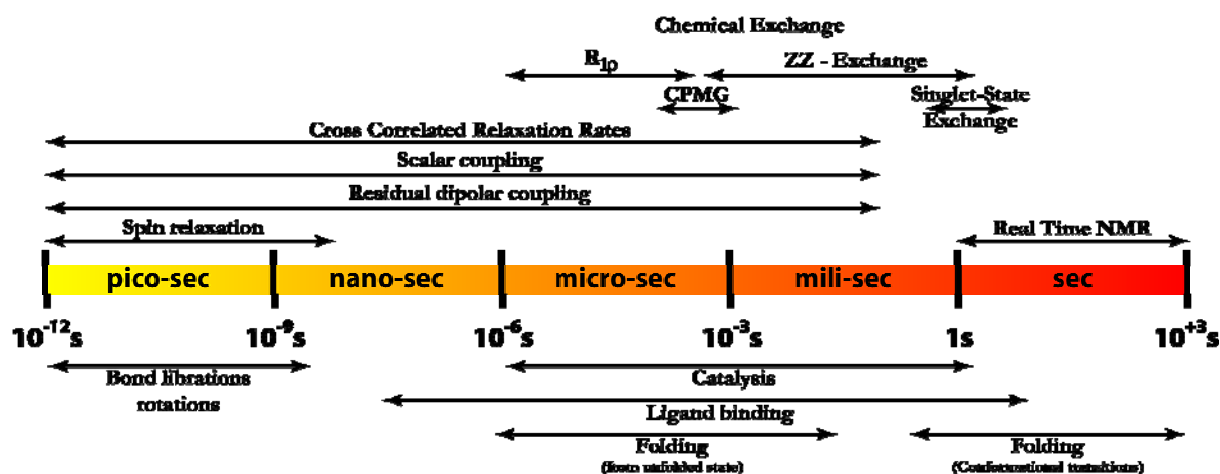


Figure 23: Graphical representation of a timeline that juxtapose for the sake of comparison different NMR methods that allow the description of different dynamic events occurring in biomacromolecules - and especially in RNA.

**A)** For the analysis of the fast pico- to nano-second processes in RNA such as bond librations, the measurement of auto-correlated transverse and longitudinal spin relaxation rates is used. Spin relaxation is measured for all different nuclei  $^{15}\text{N}^{106}$ ,  $^{13}\text{C}^{107}$ ,  $^{31}\text{P}^{108}$  - present in labelled RNA. Although those nuclei can serve as respective reporters for the fast dynamic processes in the nucleobase, the nucleobase/sugar-moiety and in the backbone, in each case the analysis is hampered by some difficulties that are exclusive for RNA molecules and do not exist in that extend in proteins.

<sup>105</sup>O. Wilde, *The Chameleon*, 1894, **1**, "Phrases and philosophies for the use of the young"

<sup>106</sup>M. Akke, R. Fiala, F. Jiang, D. Patel, A. Palmer 3rd, *RNA*, 1997, **3**, 702-709

<sup>107</sup>E. Duchardt, H. Schwalbe, *J. Biomol. NMR*, 2005, **32**, 295-308; Z. Shajani, G. Varani, *J. Mol. Biol.*, 2005, **349**, 699-715

<sup>108</sup>L. J. Catoire, *J. Biomol. NMR*, 2004, **28**, 179-184

In the case of  $^{15}\text{N}$  relaxation studies the problem arises that the number of detectable N-H spin pairs is reduced due to their intrinsic higher solvent exchange rates compared to proteins. Mostly those spin pairs are observed as one per base-pair in canonical helices that are of minor interest, and in regions of high structural plasticity just few N-H groups are detect- and analyzable.

Although spin pairs of the type C-H are much more abundant, their analysis is also by far straightforward: large  $^1\text{J}_{\text{CC}}$  couplings render the measurements of relaxation rates not easy. The other problem is the  $^{13}\text{C}$ -CSA that may be not collinear with the C-H bond and hampers analysis, or the exact value of the tensor is dependent on the conformation generating a hen-egg problem (for the knowledge of the right CSA value, the knowledge of the structural variation is necessary and vice versa).

The analysis of  $^{31}\text{P}$  relaxation rates theoretically represents a good tool to study the dynamics of the phospho-diester bond because there is a single  $^{31}\text{P}$  site per residue and it is positioned at each nucleotide connection. Nevertheless, due to the low dispersion of the signals the analysis requires either the labelling with  $^{13}\text{C}$  to get well resolved spectra (HCP<sup>109</sup> or HCP-CCH-TOCSY<sup>110</sup>) or is restricted to small RNA systems that can be analysed with 1D  $^{31}\text{P}$  experiments. An additional problem arises due to the different relaxation mechanisms that are active. In contrast to the cases for  $^{13}\text{C}$  or  $^{15}\text{N}$  relaxation analysis where dipolar relaxation is predominant and merely determined by the strong influence of the directly bound proton, the contribution of dipolar relaxation in the  $^{31}\text{P}$  case is conformation dependent. The distance of relaxation active protons depends on the geometry of the backbone. Unfortunately the second mechanism of relaxation the  $^{31}\text{P}$ -CSA also seems to be dependent on the exact conformation of the phosphate<sup>111</sup>, so that an analysis of the relaxation rates is only applicable if high resolution structural data is available.

Analogous to the protein case analysis of the relaxation rates is performed by application of the Lipari-Szabo model-free approach. The assumption underlying this method is that the internal and overall motions of the molecule are decoupled. Because in often cases this is not true for RNA molecules the measurement of relaxation in modified/elongated molecules is necessary.<sup>112</sup>

**B)** Another method to obtain dynamic information about the RNA molecules is to analyse structural parameters in order to gain knowledge of their averaging behaviour.

There are two major advantages of this procedure. The timescale covered ranges from pico- to mili-seconds, and the dynamic information are directly linked to structural information.

Among the structural dependent NMR parameters, scalar couplings where the first ones used as reporters of dynamics and interpreted considering the effect of motional averaging. Already in 1985 Hoch et al. stated: “Accurately measured coupling constants are shown to provide information concerning the magnitude of dihedral fluctuations [in proteins].”<sup>113</sup>. Generally the dynamic analysis is maintained by comparison of the experimentally derived couplings to either a

---

<sup>109</sup>J.P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D.M. Crothers, C. Griesinger, *J. Am. Chem. Soc.*, 1994, **116**, 6472-6473

<sup>110</sup>J.P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D.M. Crothers, C. Griesinger, *J. Biomol. NMR*, 1995, **5**, 87-92

<sup>111</sup>J. Precechtelova, M.L. Munzerov, P. Novak, V. Sklenar, *J. Phys. Chem. B.*, 2007, **111**, 2658-2667

<sup>112</sup>Q. Zhang, X. Sun, E.D. Watt, H.M. Al-Hashimi, *Science*, 2006, **311**, 653-656

<sup>113</sup>J.C. Hoch, C.M. Dobson, M. Karplus, *Biochemistry*, 1985, **14**, 3831-3841

reference structure<sup>114</sup> or to trajectories of molecular dynamics simulations<sup>115</sup>. Hereby the measured experimental value is interpreted as a best fit of different weighted populations of distinct conformation. More recently Lindorff-Larsen et al.<sup>116</sup> proposed the interpretation of dynamically averaged scalar couplings in proteins through an ensemble of conformations that is determined through dynamic ensemble refinement (DER<sup>117</sup>).

Subsequently the derived rotamer populations can be converted into generalized order parameters  $S_j^2$ . This new parameter provides then inside into the side chain dynamics and can be compared to the order parameters derived from spin relaxation experiments. Generally an agreement is observed but in several cases the order parameters from scalar couplings are smaller than those derived from relaxation rates  $S_j^2 < S_R^2$ . This is indicating that some rotameric averaging occurs on a time scale too slow to be observed with spin relaxation experiments<sup>118</sup>.

Another NMR spectroscopic tool to decipher molecular motions are RDC measurements. They provide an inside on the amplitude and the direction<sup>119</sup> of global motions of domains<sup>120</sup>. A prerequisite of such an analysis is that the local structure has to be known. If so, the RDCs can be interpreted to determine a domain specific order tensor describing the average domain alignment relative to the static magnetic field. The subsequent super-imposing of the order tensor frames results in a determination of the relative domain orientation and the comparison of domain specific principal order parameters reveals the amplitudes and directions of internal motions<sup>121</sup>. Difficulties of such an analysis can arise if internal motions do affect the global alignment (this is equivalent to the break down of the decoupling approximation made in a similar way by the Lipari-Szabo model-free approach). For RNA molecules such a failure can occur if exhibiting equal domains. An opposite problem arises if one domain dominates the overall alignment and basically masks the fraction of alignment of the others.

In order to improve the quality of such a dynamic analysis by RDCs the amount of measured couplings has to be as high as possible<sup>122</sup>, which then directly lead to an decrease in the size of the molecular fragments between relative motions can be measured - in best cases this lead to a dynamic description of base-base orientations<sup>123</sup>.

**C)** The theoretical treatment of chemical exchange<sup>124</sup> phenomena in RNA and the derivation of information about RNA dynamics are presented in the following section exemplary for solvent

<sup>114</sup>Y. Karim-Nejad, J.M. Schmidt, H. Rüterjans, H. Schwalbe, C. Griesinger, *Biochemistry*, 1994, **33**, 5481-5492

<sup>115</sup>J. Graf, P.H. Nguyen, G. Stock, H. Schwalbe, *J. Am. Chem. Soc.*, 2007, **129**, 1179-1189

<sup>116</sup>K. Lindorff-Larsen, R.B. Best, Vendruscolo, *J. Biomol. NMR*, 2005, **32**, 273-280

<sup>117</sup>K. Lindorff-Larsen, R.B. Best, M.A. Depristo, C.M. Dobson, M. Vendruscolo, *Nature*, 2005, **433**, 128-132

<sup>118</sup>J. Chou, D.A. Case, A. Bax, *J. Am. Chem. Soc.*, 2003, **125**, 8959-8966

<sup>119</sup>H. Al-Hashimi, *Chembiotech*, 2005, **6**, 1506-1519; H. Al-Hashimi, Y. Gossler, A. Gorin, W. Hu, A. Majumdar, D.J. Patel, *J. Mol. Biol.*, 2002, **315**, 95-102

<sup>120</sup>J.R. Tolman, H. Al-Hashimi, *Annu. Rep. NMR Spectrosc.*, 2003, **51**, 105-166

<sup>121</sup>J.A. Losonczi, M. Andrec, M.W.F. Fischer, J.H. Prestegard, *J. Magn. Reson.*, 1999, **138**, 334-342; E.T. Molloy, M.R. Hansen, A. Pardi, *J. Am. Chem. Soc.*, 2000, **122**, 11561-11562

<sup>122</sup>E. O'Neil-Cabello, D.L. Bryce, E.P. Nikonowicz, A. Bax, *J. Am. Chem. Soc.*, 2004, **126**, 66-67

<sup>123</sup>L. Trantirek, M. Urbasek, R. Stefl, J. Feigon, V. Sklenar, *J. Am. Chem. Soc.*, 2000, **122**, 10454-10455

<sup>124</sup>Reviewing chemical exchange methods:

exchange events. Here just singlet-state exchange is presented as a new emerging technique that will allow following the equilibrium interconversion of two RNA conformations in an exchange experiments.

Generically, the longitudinal relaxation time  $T_1$  is the limiting factor in EXSY<sup>125</sup> experiments. It defines the upper limit of the accessible time scale. Because many RNA folding processes exhibit a half-life time that exceeds the respective  $T_1$  multiple times, application of those methods is not suitable (Attempts that make use of the longer longitudinal relaxation time of  $^{15}\text{N}$  spins just seem to be applicable for the fastest processes of RNA refolding<sup>126</sup>). A possibility to circumvent this is probably given by two-dimensional singlet-state exchange spectroscopy (SS-EXSY)<sup>127</sup>. This is because in cases where two scalar coupled spins can be excited into singlet-states their lifetime  $T_s$  is approximately an order of magnitude longer than the respective longitudinal relaxation times<sup>128</sup>.

D) Real-Time NMR methodology that covers the time range of hundreds of milliseconds up to hours seems to be the most suitable technique to monitor RNA refolding reactions and is discussed in great detail in the following chapter.

---

<sup>125</sup>J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546-4553

<sup>126</sup>P. Wenter, G. Bodenhausen, J. Dittmer, S. Pitsch, *J. Am. Chem. Soc.*, 2006, **128**, 7579-7587

<sup>127</sup>R. Sarkar, P.R. Vasos, G. Bodenhausen, *J. Am. Chem. Soc.*, 2007, **129**, 328-334

<sup>128</sup>M. Caravetta, M.H. Levitt, *J. Am. Chem. Soc.*, 2004, **126**, 6228-6229

### SOLVENT EXCHANGE RATES

Cellular processes like the formation of macromolecular assemblies, ligand binding and enzymatic catalysis can be understood as direct pathways of energetic coupling between individual sites. To probe the energetics of molecular dynamics with site specificity, which is needed to decipher such a coupling, solvent exchange rates provide a valuable tool<sup>129</sup>. They report on the accessibility to an exchangeable hydrogen atom and the modulation of this exchange due to encounter with other molecules.

In RNA molecules, the protons that are directly bound to nitrogen or oxygen such as the imino-protons H1 and H3 in guanine and uracile, respectively, as well as the 2'-hydroxyl protons or the amino-protons, undergo chemical exchange with the protons of the solvent water molecule (H<sub>2</sub>O) at a pH of ~7.

If the nucleotide is not involved in any interactions with another nucleotide in the molecule the solvent exchange rates for such protons resemble those of single nucleotides which are  $k_{\text{ex}}=1,7 \cdot 10^5 \text{ s}^{-1}$  (as determined for a single thymidine nucleotide<sup>130</sup> - for single RNA nucleobases values of H<sup>+</sup> and OH<sup>-</sup> catalysed exchange rates are measured<sup>131</sup> ranging from  $k_{\text{H}^+}^{[\text{U-H3}]}=1,3 \cdot 10^4 \text{ M}^{-1} \text{ s}^{-1}$ ,  $k_{\text{OH}^-}^{[\text{G-H1}]}=5 \cdot 10^8 \text{ M}^{-1} \text{ s}^{-1}$  up to  $k_{\text{OH}^-}^{[\text{U-H3}]}=1,5 \cdot 10^{10} \text{ M}^{-1} \text{ s}^{-1}$ ). In the context of highly structured RNA molecules, however, there are inter- or intramolecular interactions like base-pairing or other molecular architectures that allow formation of hydrogen bonds involving these protons or protect them from direct access of the solvent molecule (Figure 24 depicts three examples by whereby exchangeable protons can be protected against solvent exchange) the rate may drop dramatically down to values with a maximum of  $k_{\text{ex}}=10^{-2} \text{ s}^{-1}$  down to values in the range of less than  $k_{\text{ex}}=10^{-8} \text{ s}^{-1}$ . This corresponds to lifetimes of less than a second (in the millisecond timescale) up to several hours and in extreme cases up to a year.

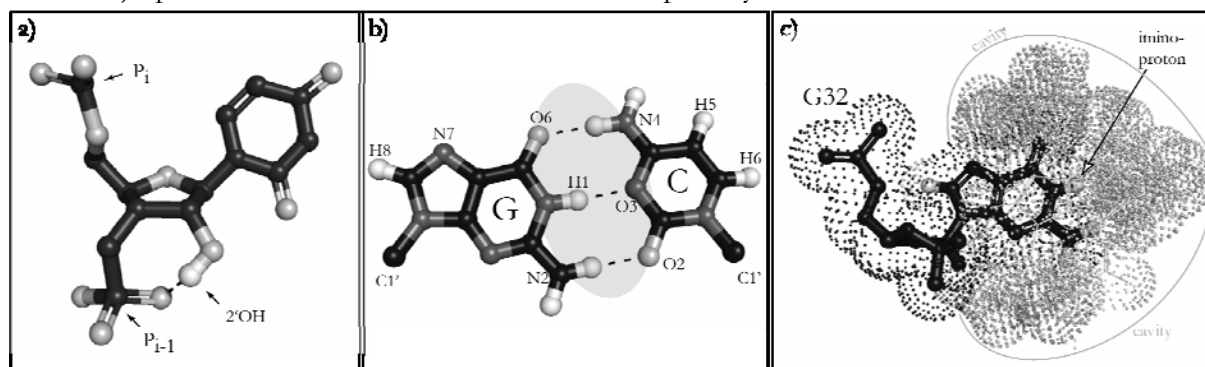


Figure 24: Examples representing mechanisms whereby the exchangeable protons get protected from exchange with solvent: a) Formation of a hydrogen bond of the 2'-hydroxyl group of the ribose moiety with the phosphate group of the 3'-adjacent nucleotide –example taken from a hydrogenated and energy minimized structure of the hypoxanthine-bound riboswitch RNA, ball-and-stick-representation, nitrogens, phosphorous and carbons are shown in black, oxygens shown in grey, the 2'OH is shown in white; b) Formation of hydrogen bonds between nucleotides guanine and cytosine in order to form a Watson-Crick-basepair; c) Formation of a molecular cavity that surrounds the base completely, the exchange protection results here from a steric exclusion of the water molecules, the imino-proton of the residue is not hydrogen bonded; the nucleotide is shown in black, cavity forming atoms are represented as dots at the size of van-der-Waals-radii – the examples taken again from a hydrogenated and energy minimized structure of the hypoxanthine bound riboswitch RNA (original pdb-code: 1U8D)

<sup>129</sup>I. M. Russu, *Methods in Enzymology*, 2004, **379**, 152-175

<sup>130</sup>G.M. Dhavan, J. Lapham, S. Yang, D.M. Crothers, *J. Mol. Biol.*, 1999, **288**, 659-671

<sup>131</sup>H. Fritzsche, L.-S. Kann, P.O.P. Ts'o, *Biochemistry*, 1981, **20**, 6118-6122

For nucleotides not protected against solvent exchange, no signal can be detected at ambient temperature. Nevertheless at very low temperature (either in supercooled water or in unnatural solvent mixtures) also solvent unprotected imino-signals are observable.<sup>132</sup>

If no or a decreased exchange is monitored, the situation is called a closed state wherein the respective proton participates in a hydrogen bond, or lacks accessibility to water and exchange catalyst. In the context of regular nucleic acid structures such as A-form helices, the transition yielding exchange-competent states for imino-hydrogens is generally equated to the opening of a basepair.

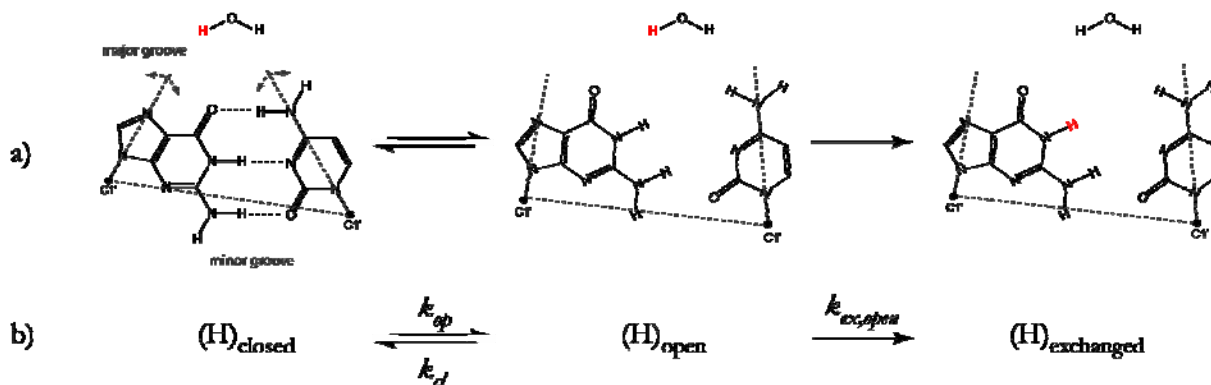


Figure 25: Mechanism of exchange: a) structural representation of b) kinetic formal representation. In order to perform exchange with solvent a given basepair has to open up (rate  $k_{op}$ ) thereby presenting the exchangeable imino proton towards the water molecule. In the next step the exchange can happen (rate  $k_{ex,open}$ ) or the basepair can be re-established in its closed form (rate  $k_{cl}$ ).

Generally, the exchange process is described in the following two state model (Figure 25), in which the proton may be involved in a closed conformation e.g. canonical Watson-Crick-basepair and is not exchange competent in this state. With a distinct opening rate ( $k_{op}$ ), the closed conformation changes towards a second one e.g. by a rotation of the bases out of the helix core by  $\geq 50^\circ$  and in this second conformation the proton is able to be exchanged with a proton of the water molecule (described by a rate  $k_{ex,open}$ ) or to reside in its chemical position and relax back into the first conformation (with rate  $k_{cl}$ ). In native biomolecules  $k_{op} \ll k_{cl}$  the monitored exchange rate is defined as:

$$k_{ex} = \frac{k_{op} \cdot k_{ex,open}}{k_{op} + k_{cl} + k_{ex,open}}$$

As described by the formula two exchange regimes are possible:

1) Exchange from the open state is rate limiting (EX2):

$$\text{if } k_{ex,open} \ll k_{cl} \text{ then } k_{ex} = K_{op} \cdot k_{ex,open} \text{ with } K_{op} = k_{op}/k_{cl}$$

2) Exchange occurs at every opening event (EX1):

$$\text{if } k_{ex,open} \gg k_{cl} \text{ then } k_{ex} = k_{op}$$

<sup>132</sup>J. Skalicky, D.K. Sukumaran, J.L. Mills, T. Szyperki, *J. Am. Chem. Soc.*, 2000, **122**, 3230–3231; K. Weisz, J. Jähnchen, H.H. Limbach, *J. Am. Chem. Soc.*, 1997, **119**, 6436–6437; K.T. Schroeder, J.J. Skalicky, N.L. Greenbaum, *RNA*, 2005, **11**, 1012–1016

It is assumed that the exchange of the open state is generally independent from the structural opening reaction and occurs in a similar manner as the exchange of mononucleotides under the same conditions. The direct exchange of the imino-proton with its water counterpart is catalysed by  $H^+$ ,  $OH^-$  and by hydrogen acceptors present in the solvent. Therefore, differentiation between the two regimes can be made by increase of the catalyst concentration where the first regime is sensitive to increasing catalyst concentration, while the second regime is independent of catalyst concentration. In nucleic acids, a lack of dependence on catalyst concentration can be observed additionally when the exchange is catalysed efficiently by an internal catalysis of neighbouring adenines and cytidines.

As known from early studies of Pörschke<sup>133</sup> and Craig<sup>134</sup>, the rate for closing a basepair is determined to be  $k_{cl} > 10^7 s^{-1}$  (this data is extracted from relaxation kinetic measurements of dimer formation by self-complementary DNA-oligonucleotides). If this were true for all nucleotides in RNA, exchange would be always in the EX2 regime. However, in several studies it is found that exchange occurs in both regimes.<sup>135</sup>

To illustrate the range of exchange rates displayed by protons in RNA molecule the following table (Table 7) summarizes early and recent studies in the field (does not claim completeness):

Table 7: Exchanges rates of exchangeable protons as found in the literature (w.i.: water inversion)

Lowest rate [s <sup>-1</sup> ]	Highest rate [s <sup>-1</sup> ]	Method	Conditions	Molecule	Citation
1,9*10 <sup>-6</sup>	5,8*10 <sup>-4</sup>	deuterium exchange	14mM NaCacodylate; 0,1M NaCl; 15mM MgCl <sub>2</sub> ; pD6,2; 291K	tRNA	Figueroa et al., <i>PNAS</i> , 1983, <b>80</b> , 4330-4333
5,8*10 <sup>-5</sup>	1,2*10 <sup>-3</sup>	deuterium exchange	14mM NaCacodylate; 0,1M NaCl; pD6,2; 291K	tRNA	Figueroa et al., <i>PNAS</i> , 1983, <b>80</b> , 4330-4333
1,8*10 <sup>-5</sup>	1,2*10 <sup>-3</sup>	deuterium exchange	14mM NaCacodylate; 0,1M NaCl; 15mM MgCl <sub>2</sub> ; pD6,2; 301K	tRNA	Figueroa et al., <i>PNAS</i> , 1983, <b>80</b> , 4330-4333
0,3	70	w.i.	10mM KP; 50mM KCl; 15mM NaCl; 0,5mM EDTA; pH8,3, 283K	Sarcin Ricin Domain	Chen et al., <i>Biochemistry</i> , 2006, <b>45</b> , 13606-13613
1	25	w.i.	20mM Tris-HCl; 45mM NaCl; 0,5mM EDTA; 303K	IHF, DNA	Crothers et al., <i>J. Mol. Biol.</i> , 1999, <b>288</b> , 659- 671
1	80	w.i.	75mM Tris-HCl, 45mM NaCl, 0,5mM EDTA, 303K	IHF, DNA	Crothers et al., <i>J. Mol. Biol.</i> , 1999, <b>288</b> , 659- 671
9*10 <sup>-4</sup>	45	w.i.	10mM NaCl; 5mM MgCl <sub>2</sub> ; pH4,6; 278K	RNA triplex	Russu <i>Method. Enzymology</i> , 2003, <b>379</b> , 152-175
9	15	w.i.	0,1M NaCl; 0,01NaP; pH7; 293K	<i>E.coli</i> tRNA <sup>Val</sup>	Tropp et al. , <i>NAR</i> , 1983, <b>11</b> , 2121-2134
19	120	w.i.	0,1M NaCl; 0,01NaP; pH7; 313K	<i>E.coli</i> tRNA <sup>Val</sup>	Tropp et al. , <i>NAR</i> , 1983, <b>11</b> , 2121-2134
9	25	w.i.	0,1M NaCl 0,01NaP pH7 293K 8eq Sperm	<i>E.coli</i> tRNA <sup>Val</sup>	Tropp et al. , <i>NAR</i> , 1983, <b>11</b> , 2121-2134
8	9	w.i.	0,1M NaCl; 0,01NaP; pH7; 293K; 4eq Mg <sup>2+</sup>	<i>E.coli</i> tRNA <sup>Val</sup>	Tropp et al. , <i>NAR</i> , 1983, <b>11</b> , 2121-2134
104	518	line width $k_{ex} = \pi \Delta \nu^{1/2}$	0,01M phosphate; pH6,9; 278K	RNA miniduplex	Fritzsche et al., <i>Biochemistry</i> , 1983, <b>22</b> , 277-280

<sup>133</sup>D. Pörschke & M. Eigen, *J. Mol. Biol.*, 1971, **62**, 361-381

<sup>134</sup>M.E. Craig, D.M. Crothers, P. Doty, *J. Mol. Biol.*, 1971, **62**, 383-401

<sup>135</sup>C. Chen, L. Jiang, R. Michalczyk, I.M. Russu, *Biochemistry*, 2006, **45**, 13606-13613 and references herein

As seen, the time range that is covered by those rates demands several different time regimes to be analysed by NMR:

- 1) Rates slower than  $k_{\text{ex}}=10^{-2}\text{-}10^{-3}\text{s}^{-1}$  can be analyzed via hydrogen-deuterium exchange measurements.
- 2) Rates in the range of  $k_{\text{ex}}=10^{-1}\text{s}^{-1}$  to  $k_{\text{ex}}=10^{+2}\text{s}^{-1}$  can be determined by experiments that transfer magnetisation from water to the respective site –such as selective water inversion experiments as well as EXSY-experiments
- 3) Rates up to  $k_{\text{ex}}=10^{+3}\text{s}^{-1}$  are too fast for the previously mentioned methods and can be assessed by indirect NMR methods such as line width measurements.

#### METHODS FOR THE MEASUREMENT OF SOLVENT EXCHANGE RATES BY NMR-SPECTROSCOPY

##### Hydrogen-Deuterium Exchange

Akin to a classical pulse-chase experiment, hydrogen-deuterium exchanges is performed by incubating the RNA molecule first in  $\text{H}_2\text{O}$ , then freeze-drying the sample and finally dissolve it in  $\text{D}_2\text{O}$ . This time point is considered as  $\tau=0$ , then subsequent NMR experiments are performed on a pretuned spectrometer. The decay of signal intensity of the exchangeable proton is described by monoexponential behaviour

$$I(\tau) = [I(0) - I(\infty)] \cdot e^{-k_{\text{ex}} \cdot \tau} + I(\infty)$$

Here,  $I(\tau)$  is the measured signal intensity at timepoint  $\tau$ ,  $I(0)$  is the initial signal intensity in 100%  $\text{H}_2\text{O}$  and  $I(\infty)$  differs from zero if residual  $\text{H}_2\text{O}$  is in the chase sample. The drawback of the method is that mixing has to be performed outside the NMR spectrometer and so the deadtime is approximately  $t_{\text{dead}}=5\text{min}$ . The time resolution of the exchange measurement itself is determined by the repetition rate of the NMR experiment.

However, only a few sites reveal exchange rates slow enough to allow for application of pulse-chase experiments. This is different to the situation in proteins due to the higher intrinsic exchange rate of imino hydrogen atoms compared to amide hydrogens in proteins.<sup>136</sup>

Also it should be mentioned that the number of exchangeable protons in RNA showing such a slow exchange rate is not very high and the method has -in the field of RNA NMR research- a more historical relevance.

##### Two-dimensional NOESY-experiment

The NOESY experiment is a different method that allows the measurement of solvent exchange rates<sup>137</sup>. It is best to use if the experimental conditions are chosen in a way that the exchange rate is fast compared to the longitudinal relaxation and slow compared to spectral parameters that are

<sup>136</sup>G. Wagner, C.I. Stassinopoulou, K. Wüthrich, *Eur. J. Biochem.*, 1984, **145**, 431-436; K. Kuwajima, R.L. Baldwin, *J. Mol. Biol.*, 1983, **169**, 299-323

<sup>137</sup>J. Jeener, B.H. Meier, P. Bachmann, R.R. Ernst, *J. Chem. Phys.*, 1979, **71**, 4546-4553  
G. Bodenhausen, R.R. Ernst, *J. Am. Chem. Soc.*, 1979, **101**, 6441-6442



affected by the exchange process. In the standard experiment -that is now discussed in further detail- frequency labelling is maintained before exchange occurs. By this methodology it is possible to trace back the pathway of magnetization by utilizing the fact that magnetization is in non-equilibrium during the experiment while the chemical species remain in equilibrium. The standard NOESY experiment is comprised by the pulse-sequence:  $(\pi/2)_{\pm x} - t_1 - (\pi/2)_{\pm y} - \tau_m - (\pi/2)_{+y} - (t_2)_{+}$ . In this type of experiment the transfer of magnetisation from one spin to another is obtained by a mixing period  $\tau_m$  during which the magnetisation is along the z-direction. During this mixing time dipolar coupling can evolve and chemical exchange occur.

Modified classical Bloch equations<sup>138</sup> describe for a system of N-sites, which are in chemical exchange and where each site containing n magnetically equivalent nuclei, the dynamic behaviour of the system during the experiment.

Initially all magnetization is along z and can be described by a vector  $\vec{M}_0$  with the elements  $M_{i0} = n_i x_i M_0$ ; that is converted into transversal magnetization by the first  $(\pi/2)$ -pulse, the magnetisation is given by:

$$\vec{M}^+(t_1 = 0) = \vec{M}_x(t_1 = 0) + i\vec{M}_y(t_1 = 0) = \vec{M}_0 \text{ where } \vec{M}_y(t_1 = 0) = 0^{139}$$

The evolution of magnetization during  $t_1$  is given by

$$\frac{d}{dt} \vec{M}^+ = \vec{L}^+ \vec{M}^+.$$

where  $\vec{L}^+$  comprises the effect of chemical shift, of transverse relaxation and the kinetic rate of the exchange process:

$$\vec{L}^+ = i\vec{\Omega} - \vec{\Lambda} - \vec{K} \quad (\vec{\Omega} \text{ contains } \omega_i; \vec{\Lambda} \text{ contains } \lambda_i = \frac{1}{T_2^i}).$$

In the slow exchange regime, the exchange part during  $t_1$  is small and so  $\vec{L}^+$  reduces to

$$\vec{L}^+ = i\vec{\Omega} - \vec{\Lambda}.$$

So the magnetization during  $t_1$  can be described by the following solution:

$$\vec{M}^+(t_1) = \exp(\vec{L}^+ t_1) \vec{M}^+(t_1 = 0).$$

The second  $(\pi/2)$ -pulse converts this now into z-magnetization with

$$\vec{M}_z(\tau_m = 0) = -\text{Re}(\vec{M}^+(t_1))$$

<sup>138</sup>F. Bloch, *Phys. Rev.*, 1946, **70**, 460-474

<sup>139</sup>complex detection assumed

The evolution of the longitudinal z-magnetization during the mixing time is governed by  $\vec{L}$  representing a dynamic matrix of

$$\vec{L} = -\vec{R} + \vec{K}$$

where  $\vec{R}$  describes the relaxation behaviour of the system and  $\vec{K}$  the chemical exchange processes affecting the magnetization. The change of magnetization with time is therefore governed by

$$\frac{d}{d\tau_m} \overrightarrow{\Delta M_z}(\tau_m) = \vec{L} \overrightarrow{\Delta M_z}(\tau_m).$$

This general treatment is possible because cross-relaxation and exchange are formally similar effects. The solution of the differential equation above is given by

$$\overrightarrow{M_z}(\tau_m) = \overrightarrow{M_0} \exp(\vec{L}\tau_m) \overrightarrow{\Delta M_z}(\tau_m = 0).$$

The third ( $\pi/2$ )-pulse then transfers magnetization along the transversal plane

$$\overrightarrow{M^+}(t_2 = 0) = \overrightarrow{M_z}(\tau_m)$$

and evolves in an analogous manner as in  $t_1$ .

The evolution of magnetization during the three time-delays  $t_1$ - $\tau_m$ - $t_2$  can then be summarized, and, if phase cycling is applied in order to get real data, simplified to

$$\overrightarrow{M^+}(t_1; \tau_m; t_2) = -\exp(\vec{L}^+ t_2) \exp(\vec{L} \tau_m) \exp(\vec{L}^+ t_1) \overrightarrow{M_0}$$

The dynamic matrix describing the change in the system during the mixing time is spanned as

$$\vec{L} = \begin{pmatrix} -k_{BA} - R_1^A & k_{AB} - R^{AB} \\ k_{BA} - R^{BA} & -k_{AB} - R_1^B \end{pmatrix} \text{ whereas } R^{AB}=0 \text{ if no cross-relaxation is present.}$$

This matrix describes the amplitude of the peaks in the final 2D spectrum – in contrast  $\vec{L}^+$  describes the position and shape of the respective signal. Solving the equations for the amplitudes of the peaks it becomes evident that the diagonal peak is described by /Lösung Eigenwert angeben/

$$I_{AA} = I_{BB} = \cosh(k\tau_m) \exp(-(k_{ex} + R_1)\tau_m)$$

and the cross peak by

$$I_{AB} = I_{BA} = \sinh(k\tau_m) \exp(-(k_{ex} + R_1)\tau_m)$$

For the case of a symmetrical exchange without cross-relaxation (this is the normal case as the distance between imino- and water-protons is regularly  $>5\text{\AA}$ , rare exceptions are those imino-protons that are involved in water mediated hydrogen bonds, such as H1 protons of

pseudouridine<sup>140</sup>) and with the assumption made, that the spin-lattice relaxation rates of spin A and spin B are the same, the rate  $k_{AB}=k_{BA}=k_{ex}$  can be determined from the intensities of the cross and the diagonal peak (Figure 26)

$$\frac{I_{AB}}{I_{AA}} = \frac{1 + \exp(-2k_{ex}\tau_m)}{1 - \exp(-2k_{ex}\tau_m)} \quad (\text{Formula } \alpha)$$

In the initial rate regime -at short mixing time intervals  $k\tau_m < 1$ - the ratio (represented by the formula above) is linear and the behaviour of the ratio can be approximated by

$$\frac{I_{AB}}{I_{AA}} \cong k_{ex} \tau_m \quad (\text{Formula } \beta)$$

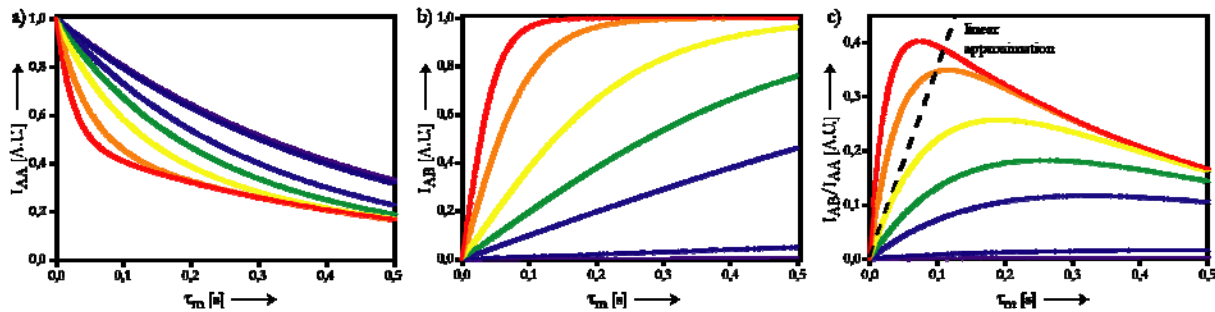


Figure 26: Simulation of a) diagonal peak amplitude ( $I_{AA}$ ) b) cross peak amplitude and ( $I_{AB}$ ) c) ratio of amplitudes of cross to diagonal peak ( $I_{AB}/I_{AA}$ ). Simulation is taking into account an equally populated two site exchange model with equal relaxation rates of  $R_1^A=R_1^B=2.2s^{-1}$ . Shown are curves for different exchange rates  $k_{ex}$  [— 0.01  $s^{-1}$ ; — 0.1  $s^{-1}$ ; — 1  $s^{-1}$ ; — 2  $s^{-1}$ ; — 4  $s^{-1}$ ; — 10  $s^{-1}$ ; — 20  $s^{-1}$ ]. In c) a dashed line indicates the linear approximation of the initial rate regime on the curve for 4  $s^{-1}$ . It shows that the linear approximation is valid for  $\tau_m < 0.04s$ .

In RNA-molecules the exchange rates of the imino-protons that are meant to be a measure for the stability of a given basepair can be measured by recording a 2D  $^1H$ - $^1H$  NOESY or if overlap of peaks is an issue –as it is in sizeable and uniformly labelled RNA molecules- in an 3D  $^{15}N$ -NOESY-HSQC experiment. The separated peaks –the diagonal as well as all cross peaks- along one of the proton chemical shift axes are integrated and reveal the exchange rate due to the formula –where corrections for the different dipolar contributions also observable are added.<sup>141</sup> As stated above, the treatment of dipolar relaxation and chemical exchange is identical and therefore, the correction due to dipolar couplings of spins is given by:

$$k_{ex} = \frac{CP_{HDO}}{(CP_{HDO} + DP_{Imino} + \sum CP_{dipolar}) * \tau_m}$$

In principle, overlapping signals at the resonance frequency of  $^1H_{(H_2O)}$  could influence the intensity of the exchange-induced water–imino cross-peaks in the 2D NOESY experiment;

<sup>140</sup>M.I. Newby, N.L. Greenbaum, *Proc. Natl. Acad. Sci.*, 2002, **99**, 12697-12702

<sup>141</sup>C. Richter, *PhD thesis*, 1999, University Frankfurt/Main

however, the ribose protons are too far away from the imino protons. Hydroxyl group protons are in exchange with water and could potentially contribute to the cross-peak intensity; again, the effect is made negligible by long distance. Base protons resonate upfield and cannot contribute to the cross-peak intensity.

Measurements of imino exchange rates using NOESY experiments were performed using Watergate<sup>142</sup> or Jump-Return-NOESY<sup>143</sup> pulse sequences in order to obtain optimal water suppression and maximum S/N (Figure 27). The peaks are then modulated differently in the two frequency domains of the experiment. It is therefore necessary to analyse the crosspeaks of the imino to water signal in the  $\omega_1$  direction. No differences in the determined exchange rates are observed for the two different pulse sequences. For clarity, an example of a 2D NOESY spectrum is shown in which the exchange rate is measured by the above described formalism.

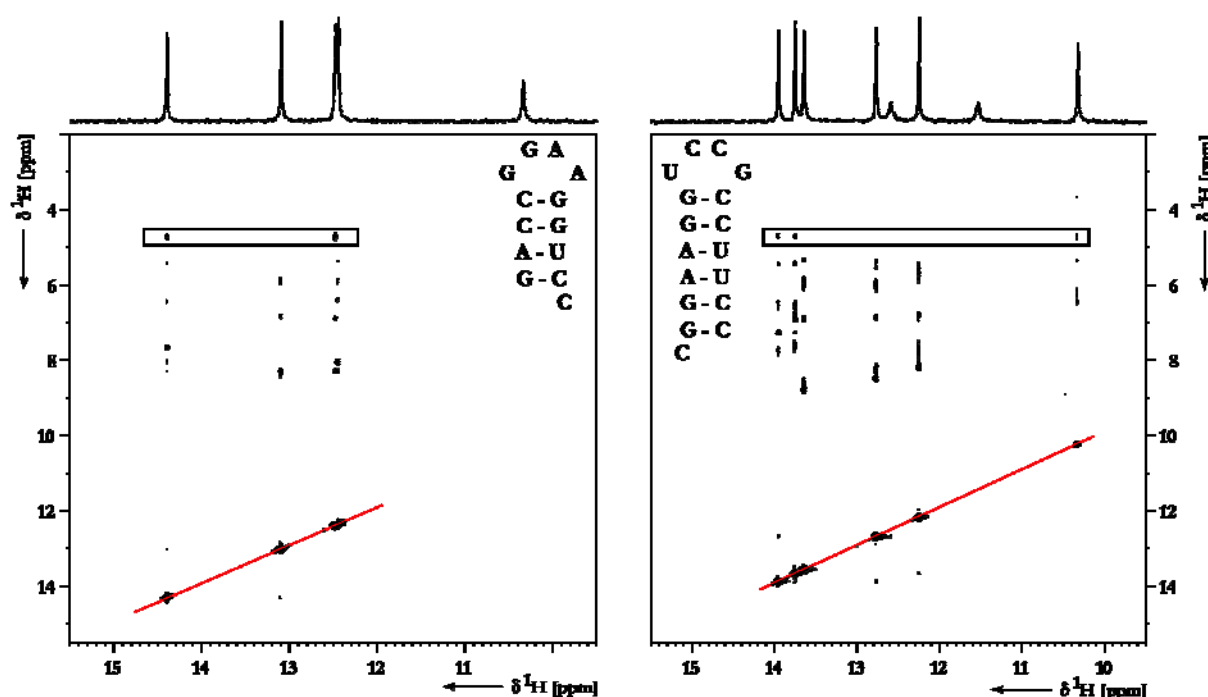


Figure 27: Imino-region of two 2D-NOESY spectra of different RNA hairpin-loops. The imino-water cross peaks are indicated by a box; the diagonal peaks are highlighted by a red line. Both spectra were recorded at 283K at 600MHz. Proton pulses were applied on resonance of the HDO frequency. Water suppression was achieved by the jump-return-echo method. Hard proton pulses were applied with field strength of 29.8kHz. The relaxation delay was set to 2.0s and the mixing time was 150ms. The experiment was recorded with 4096pts and 1024pts along  $t_2$  and  $t_1$ , respectively. The number of scans was set to 96 for an optimal S/N-ratio. For processing the spectrum was zero-filled to a size of 2048pts for along  $t_1$ . For apodization a squared sine function was applied to the FID.

### Water-Inversion Experiments

A third method to detect exchange processes is basically the 1D version of the 2D EXSY experiment. Here, one selectively modifies (magnetic labelling) the spin state of one site that is involved in chemical exchange, then wait for a given time and finally reads out the spin state of the second site. By variation of the time between preparation and read-out the spectrum will be

<sup>142</sup>M. Piotto, V. Saudek, V. Sklenar, *J. Biomol. NMR*, 1992, **2**, 661 - 666

<sup>143</sup>V. Sklenar, A. Bax, *J. Magn. Res.*, 1987, **74**, 469-479

modulated according to the rate of exchange. Analogous to the case of the 2D NOESY the exchange matrix is spanned by elements that describe the longitudinal relaxation of inverted magnetization in the water and as exchange proceeds also in the imino site and of course by elements that take the exchange rate into account.

$$\vec{L} = \begin{pmatrix} -k_{ex} - R_1^{Im} & k_{ex} \\ k_{ex} & -k_{ex} - R_1^{HDO} \end{pmatrix}$$

To gain the exchange rate the imino cross peaks have to be integrated along  $t_1 = \tau_m$  and the trace can then be fitted according to

$$I^{Im} = 1 - 2 * k_{ex} \left( \frac{e^{-R_1^{Im} * \tau_m} - e^{-R_1^{HDO} * \tau_m}}{R_1^{HDO} - R_1^{Im}} \right)$$

The shape of the curve is determined by 3 different parameters: The exchange rate  $k_{ex}$ , the longitudinal relaxation rate for the proton in the water environment  $R_1^{HDO}$  and for the proton in the imino environment  $R_1^{Im}$ .

The similarity of the NOESY and the selective inversion method becomes evident by comparing the basal pulse sequences of the two experiments (Figure 28). Chemical exchange is evolved simultaneously during a period  $\tau_m$  in which the magnetization is along z (longitudinal).

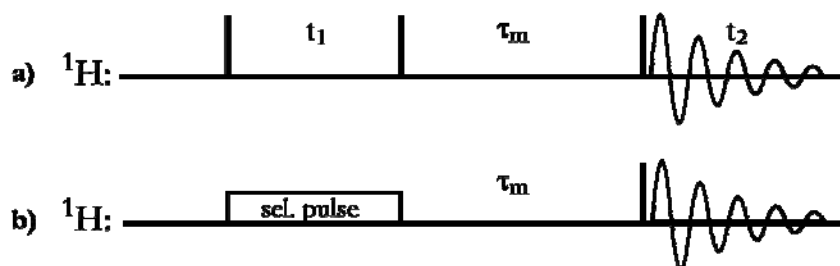


Figure 28: Schematic representation of a) NOESY pulse sequence b) selective inversion experiment

The main difference between the NOESY experiment and the water inversion method is therefore that a second chemical shift period proceeds in the first case. The advantage of the NOESY method is truly, that studying a network involving a large number of sites is possible in only a single experiment while the selective inversion experiment has to be executed for each site separately. Overlapping peaks are in neither technique a problem, in the NOESY these peaks could be intrinsically resolved, and in the selective inversion experiment one can replace the readout pulse by a filter or editing experiment (x-filter<sup>144</sup>, HSQC<sup>145</sup> etc.). The advantage of the selective inversion method is that it can be done at many different mixing times, whereas in the

<sup>144</sup>G. Otting, K. Wuethrich, *J. Magn. Reson.*, 1989, **85**, 586-594

<sup>145</sup>G. Bodenhausen, D.J. Ruben, *Chem. Phys. Lett.*, 1980, **69**, 185

case of the NOESY this is rather time consuming. However, it is also possible if  $t_1$  and  $\tau_m$  are varied simultaneously.<sup>146</sup>

The experimental setup for an edited selective inversion experiment is realized with the following sequence (Figure 29):

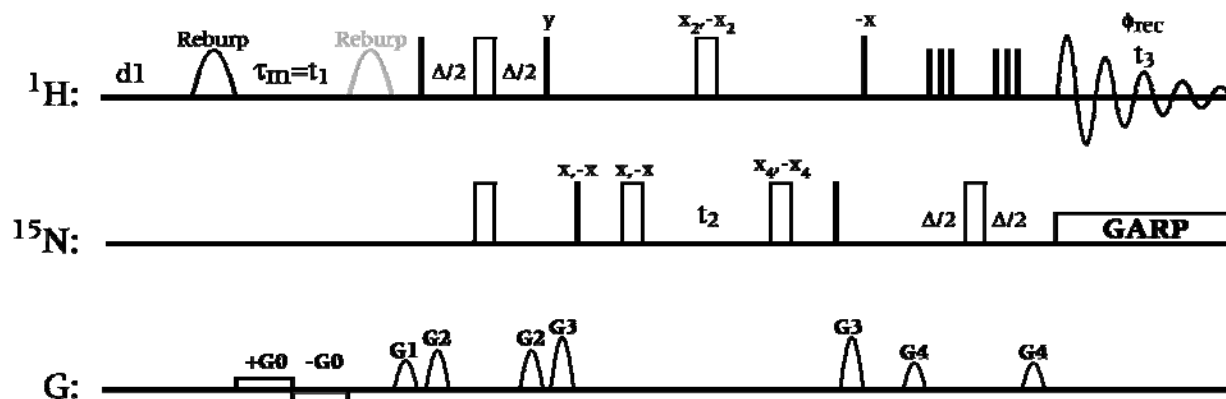


Figure 29: Pulse sequence of the pseudo-3D-water-inversion- $^{15}\text{N}$ -HSQC experiment. Except for the rectangular gradients  $G_0$  all gradients have a SINE.100 shape and are applied with strength of  $gp_0$ :5%,  $gp_1$ :40%,  $gp_2$ :50%,  $gp_3$ : 80% and  $gp_4$ : 30%. The selective  $180^\circ$  pulses with a Reburp.100 profile applied for water inversion have a field strength of 400Hz. Nitrogen decoupling is realized by a GARP decoupling sequence during acquisition. The mixing time  $\tau_m$  is varied between 4 $\mu$ s and 4s. All pulses if not stated explicitly are applied with phase  $\alpha$ , the receiver phase is  $\phi_{\text{rec}} = (\alpha, -\alpha, \alpha, -\alpha, -\alpha, \alpha, -\alpha, \alpha)$ . The second selective Reburp pulse (shown in grey) is only applied for short  $\tau_m < 500\text{ms}$  where the water magnetisation is along  $-\alpha_z$  in order to obtain proper water suppression.

Therefore initially the waterline is inverted by a shaped  $180^\circ$  pulse that covers a range of 400Hz and is on resonance with the HDO signal. After the inversion, an incremented period  $\tau_m$  follows during which the water magnetisation is along  $-z$  whereas all other magnetisation is along  $+z$ . To avoid spectral imperfections due to water induced radiation damping all time during  $\tau_m$  a dephasing  $z$ -Gradient with a sign change at  $\tau_m/2$  is applied. Facilitating the read out of the effect of chemical exchange a  $^{15}\text{N}$  HSQC spectrum is subsequently recorded. To minimize disturbances from the outside, the experiment is recorded in a double interleaved manner. Accordingly, for every increment in  $\tau_m$  and  $t_2$  all scans are recorded subsequently.

<sup>146</sup>G. Bodenhausen, R.R. Ernst, *J. Am. Chem. Soc.*, 1982, **104**, 1304-1309; G. Bodenhausen, R.R. Ernst, *J. Magn. Res.*, 1981, **45**, 367-373

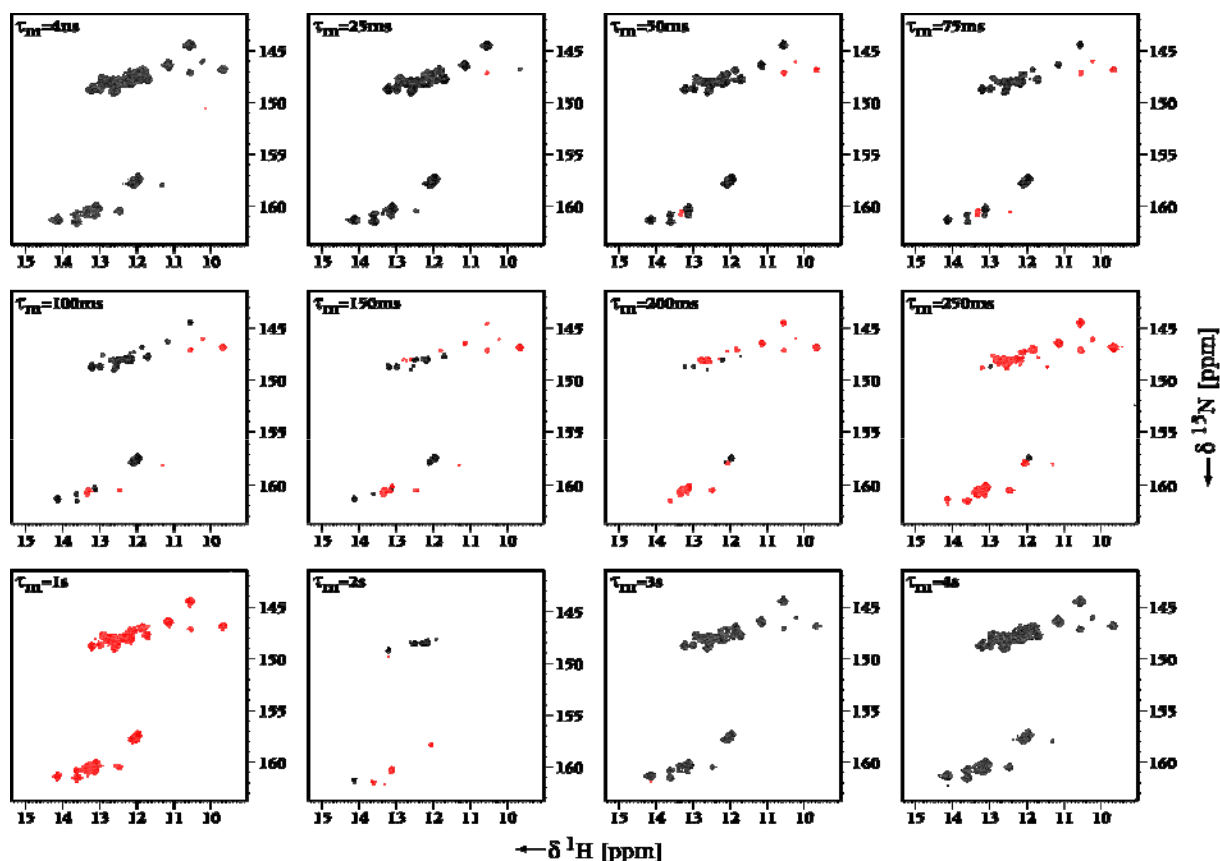


Figure 30: Series of  $^1\text{H}$ ,  $^{15}\text{N}$ -planes of the pseudo-3D-water-inversion- $^{15}\text{N}$ -HSQC experiment at 600MHz and 283K. The transmitter frequency was set to 153ppm for  $^{15}\text{N}$ , proton pulses were applied on resonance of the HDO frequency. Hard proton and nitrogen pulses were applied with field strength of 23.5kHz and 6.5kHz, respectively. The GARP decoupling during acquisition was applied with 0.8kHz. The HSQC planes were recorded phase sensitive according to the States-TPPI procedure. The relaxation delay was set to 5.5s. The experiment was recorded with 2048pts, 128pts and 17pts along  $t_3$ ,  $t_2$  and  $t_1(=\tau_m)$ , respectively. The number of scans was set to 16 for an optimal S/N-ratio. For processing the spectrum was zero-filled to a size of 256pts for along  $t_2$ . For apodization a squared sine function was applied to the FID.

The resulting experiment is a pseudo-3D spectrum (Figure 30) with a proton chemical shift evolution period along  $t_3$ , a  $^{15}\text{N}$  chemical shift evolution period along  $t_2$  and the mixing time  $\tau_m$  along  $t_1$ . Integration of the respective HSQC imino crosspeak (Figure 31) along  $t_1$  results than in the wanted curve that can be analysed according to the above derived equation.

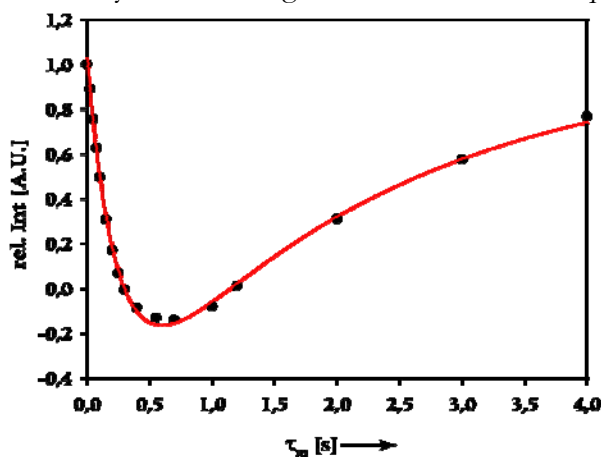


Figure 31: Integration of a single imino crosspeak along  $t_1$ .

### Comparison of exchange rates determined by different methods

If one compares the solvent exchange rates determined by NOESY experiments with those measured by selective water inversion (Figure 32) it becomes obvious that the absolute values differ by a factor of  $\sim 20$ , despite the identical relation in-between the respective set of rates. Meaning that for a comparison of rates determined for different experimental environments or at different RNA molecules either one of the methods has to be used and simultaneous use of both has to be avoided.

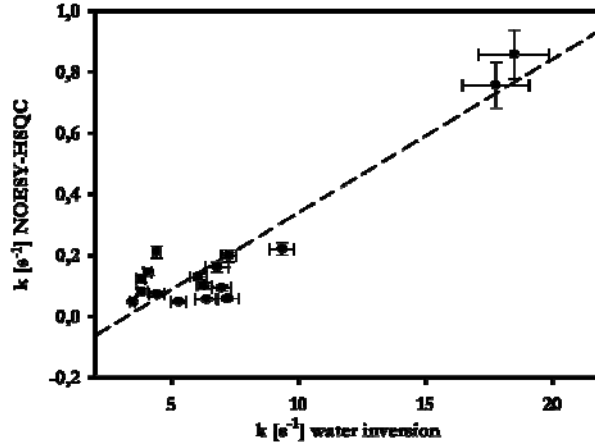


Figure 32: Scatter-plot of solvent exchange rates measured by the water-inversion method versus the rates extracted from the NOESY according to the so far described formalism (formula  $\beta$ ). The dashed linear regression line indicates that the relation of the rates within one set is identical. Nevertheless deviation of the slope from a value of one is indicating that the absolute values are different.

The remaining questions are why and which of the absolute values are correct.

Returning back to the theory of the NOESY experiment we get –as seen above– as exchange matrix

$$\vec{L} = \begin{pmatrix} -k_{ex} - R_1^A & k_{ex} - R_{cross}^{A,B} \\ k_{ex} - R_{cross}^{B,A} & -k_{ex} - R_1^B \end{pmatrix}$$

So far we considered the cross relaxation rates to be zero (what is true) and the longitudinal relaxation rate to be equal  $R_1^A = R_1^B$  (what is an assumption that is just fairly true in the case of RNA and water).

Taking into account that the longitudinal relaxation rates are different, but still considering a symmetrical two side exchange (which is again a approximation that is hardly acceptable for the RNA-water case), the solution of the exchange matrix gets more complicated and the intensities of diagonal and cross-peaks are described by

$$I_{AA}(\tau_m) = \frac{1}{2} \left[ \begin{aligned} & \left[ 1 - \frac{\delta}{D} \right] \exp \left( \left( - \left( \frac{1}{2} (2k_{ex} + R_1^A + R_1^B) \right) + D \right) \tau_m \right) \\ & + \left[ 1 + \frac{\delta}{D} \right] \exp \left( \left( - \left( \frac{1}{2} (2k_{ex} + R_1^A + R_1^B) \right) - D \right) \tau_m \right) \end{aligned} \right] M_{A0}$$

(Formula  $\gamma'$ )



and

$$I_{AB}(\tau_m) = \frac{1}{2} \left( \exp \left( \left( -\left( \frac{1}{2} (2k_{ex} + R_1^A + R_1^B) \right) + D \right) \tau_m \right) - \exp \left( \left( -\left( \frac{1}{2} (2k_{ex} + R_1^A + R_1^B) \right) - D \right) \tau_m \right) \right) M_{A0} \frac{k_{ex}}{D}$$

(Formula  $\gamma$ )

where

$$D = \sqrt{\left( \frac{1}{2} (R_1^A - R_1^B) \right)^2 + k_{ex}^2} \quad \text{and} \quad \delta = \frac{1}{2} (R_1^A - R_1^B)$$

If one now takes the rates measured from water inversion measurements, in whereby also  $R_1^{HDO}$  and  $R_1^{lm}$  are determined as variables, the dependence of the ratio of cross to diagonal peak versus the mixing time is fairly different (Figure 33).

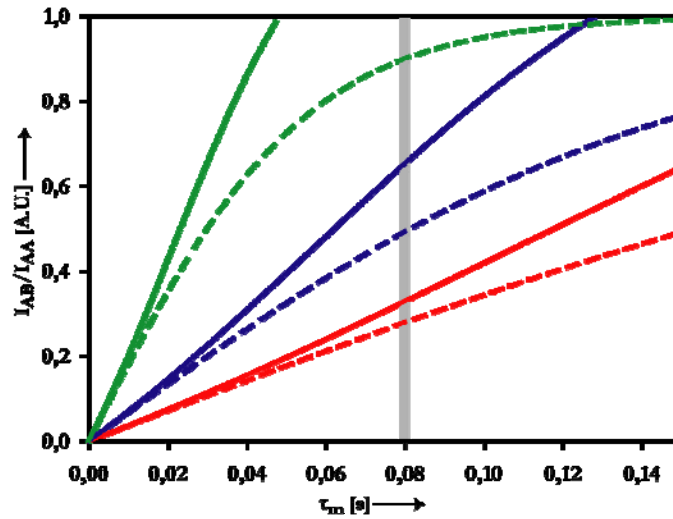


Figure 33: Simulation of the ratio of amplitudes of cross to diagonal peak ( $I_{AB}/I_{AA}$ ). The solid lines are representing the 'exact' solution of the exchange matrix (formulae  $\gamma$ ), whereas the dashed lines are representing the approximate solution of the exchange matrix (formula  $\alpha$  – taking the longitudinal relaxation rates as identical). The three scenarios are calculated according to the values of  $k_{ex}$ ,  $R_1^{lm}$ ,  $R_1^{HDO}$  for the nucleotides U20 (–):  $3.59s^{-1}$ ,  $4.78s^{-1}$ ,  $0.42s^{-1}$ ; U26 (–):  $6.75s^{-1}$ ,  $8.98s^{-1}$ ,  $0.42s^{-1}$ ; G32 (–):  $18.41s^{-1}$ ,  $22.37s^{-1}$ ,  $0.42s^{-1}$ ; respectively. The grey bar indicates a mixing time of 80ms, the value where the actual rates were measured.

Extracting the values at the mixing time of  $\tau_m=80ms$  (as in the NOESY experiment the case) and back-calculating the exchange rate with the approximated formula derived before (formula  $\beta$ ) one gets rates that are in the order of those determined directly in the experiment. And exactly as seen for the experimental rates their relative ratio is the same (Figure 34).

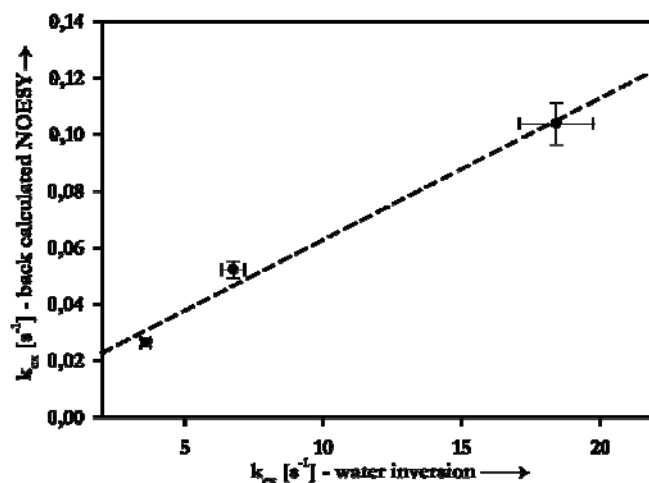


Figure 34: Scatter-plot of solvent exchange rates measured by the water-inversion method versus the rates where the ratio of intensities is back-calculated by the 'exact' formalism and rates are extracted by the approximate formula (formula  $\beta$ ).

### Rates from the NOESY

That the absolute values of the exchange rates as extracted by the approximate formalism cannot be true is obvious alone from the consideration that at short mixing times,  $\tau_m < k_{ex}^{-1}$ , the cross peaks are absent, and the spectrum contains only diagonal peaks. Just by increasing the mixing time, the intensities of the cross peaks increase, whereas the diagonal peaks weaken.

Although, to stress this point again, if a comparison study is needed the NOESY exchange rates are highly sufficient – because nearly the same approximations are made in each case (the ratio of  $R_1^{Im}$  and  $R_1^{HDO}$  are basically constant for each site).

### Lineshape analysis

Exchange rates of  $k_{ex} > XX$  affect the lineshape of the respective signals. Therefore the rates can be extracted by an analysis of the shape of the resonance line. For the case that the exchange rate is smaller than half of the difference in the chemical shift of the respective nuclei (this is known as the slow intermediate regime of chemical exchange) the form of the signal is defined as

$$S(\Omega) = \frac{1}{2} \left( 1 - \frac{ik_{ex}}{R} \right) L(\Omega; \bar{\Omega} + R, \lambda + k) + \frac{1}{2} \left( 1 + \frac{ik_{ex}}{R} \right) L(\Omega; \bar{\Omega} - R, \lambda + k) \quad ^{147}$$

where

$$R = \sqrt{k_{ex}^2 - (\Omega_A/2)^2} \quad \text{and} \quad \bar{\Omega} = \frac{1}{2}(\Omega_A^0 + \Omega_B^0)$$

So the three parameters influencing the signal are the chemical shifts, the relaxation rates and the exchange rate. To demonstrate the effect on the line shape a simulation is shown in the following figure. The simulation was set up for a  $B_0$  field of 300MHz and the input parameters were chosen as if one had an exchange between water (index b) and an imino proton (index a) in a RNA molecule.

<sup>147</sup> $L$  defines here the complex Lorentzian function (as defined in the appendix)

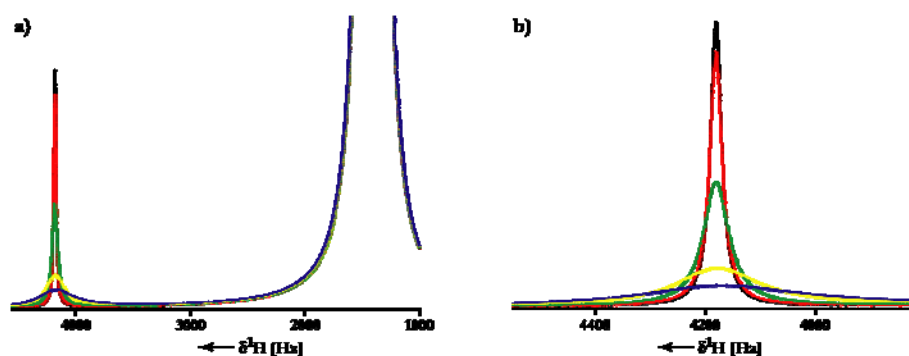


Figure 35: Simulated lineshapes for a two site exchange model based on input parameters derived from RNA spectra. a) shows both the HDO and the imino signal; b) represents a zoom in of the imino signal. Lineshapes are shown for different exchange rates  $k_{ex}$ :  $10^0 \text{ Hz}$ ,  $10^1 \text{ Hz}$ ,  $10^2 \text{ Hz}$ ,  $5 \cdot 10^2 \text{ Hz}$  and  $10^3 \text{ Hz}$ . [ $\Omega_A = 4180 \text{ Hz}$ ;  $\Omega_B = 1410 \text{ Hz}$ ;  $\lambda_A = 24 \text{ Hz}$ ;  $\lambda_B = 7 \text{ Hz}$ ; B:A=0,1%]. Simulation was performed with the software windnmr7.1.12<sup>148</sup>

It is obvious that the higher the exchange rate, the more the signal gets broadened (Figure 35). If in a RNA molecule the exchange has rates suitable for such an analysis the value can be extracted from the half width at half height that is –as demonstrated above– a function of  $k_{ex}$ .

$$\Delta \nu_{\frac{1}{2}}^{ex} = \Delta \nu_{\frac{1}{2}}^{im} - \Delta \nu_{\frac{1}{2}}^{non-ex} = \frac{1-p}{\pi} (k_{ex})$$

Since also the relaxation behaviour of the site influences the broadness of the line a correction by the half width of a non-exchangeable proton of the same site has to be done.<sup>131</sup>

#### SOLVENT EXCHANGE RATES AS MEASURES OF BASE-PARING DYNAMICS AND STABILITY

As already outlined in the beginning of the section solvent exchange rates measured for imino-protons are directly linked to basepair opening events. This is because just in an open geometry the site of this particular hydrogen atom is solvent accessible. Obviously a stronger, respectively more stable basepair displays a longer lifetime and therefore the probability of an exchange event is reduced. The actual probability for the opening of a basepair is indeed very low and determined to be  $10^{-4}$  to  $10^{-7}$  at  $15^\circ\text{C}$ .<sup>149</sup> Based on MD-simulations and kinetic measurements the lifetime of GC- and AU-basepairs is determined to be 26-55ms and 5-10ms, respectively. Open basepair lifetimes are therefore in the range of 10ns up to hundreds of ns. The energetic difference related to this opening event is expressed in the relative free energy of an open basepair and accounts for 7-10kcal mol<sup>-1</sup> in WC basepairs and for 3-6kcal mol<sup>-1</sup> in a wobble basepair.<sup>150</sup>

Interestingly, simulation reveal that the opening of the basepair is not restricted to the major groove side of the RNA helix, but occurs in nearly symmetrical towards either one of the sides. Although thought to be hindered by a steric clash of exocyclic groups and because of the proximity of the sugar backbone at the minor groove side the energetic preference for opening to the major groove side is just preferred by  $\sim 1 \text{ kcal mol}^{-1}$ .<sup>151</sup>

<sup>148</sup>H. Reich, *J. Chem. Ed. Software*, 1996, **3D**, 2

<sup>149</sup>P. Várnai, M. Canalia, J.-L. Leroy, *J. Am. Chem. Soc.*, 2004, **126**, 14659-14667

<sup>150</sup>*NAR*, 2003, **31**, 1434-1443

<sup>151</sup>E. Giudice, R. Lavery, *J. Am. Chem. Soc.*, 2003, **125**, 4998-4999

In several studies performed on all different kind of oligonucleotides (B-DNA<sup>152</sup>, Z-DNA<sup>153</sup>, A-RNA homopolymers<sup>154</sup>, tRNA<sup>155</sup>) it became clear that basepair kinetics and therefore exchange rates are dependent on the structural context.

Carried out under different temperature conditions solvent exchange rates are able to draw a picture of how energetic contributions of specific basepairs in a specific sequence or structural context contribute to the overall stability and dynamics of an oligonucleotide/RNA-structure. As a result of such a study the activation energies for basepairing opening are determined from exchange rates to be 12-26kcal mol<sup>-1</sup> for GC- and 11-18kcal mol<sup>-1</sup> for AU-basepairs. Not very surprisingly the activation barrier that has to be passed on the opening route increases going from terminus to the middle of a RNA-helix – as example from 68-78kcal mol<sup>-1</sup> for a i+1 basepair to 80-110kcal mol<sup>-1</sup> for an i+2 basepair.<sup>156</sup>

---

<sup>152</sup>J.-L. Leroy, E. Charretier, M. Kochoyan, M. Guéron, *Biochemistry*, 1988, **27**, 8894-8898

<sup>153</sup>M. Kochoyan, J.-L. Leroy, M. Guéron, *Biochemistry*, 1990, **29**, 4799-4805

<sup>154</sup>J.-L. Leroy, D. Broseta, M. Guéron, *J. Mol. Biol.*, 1985, **184**, 165-178

<sup>155</sup>J.-L. Leroy, N. Bolo, N. Figueroa, P. Plateau, M. Guéron, *J. Biomol. Struct. Dyn.*, 1985, **2**, 915-939

<sup>156</sup>K. Snoussi, J.-L. Leroy, *Biochemistry*, 2001, **40**, 8898-8904

## REAL TIME NMR

Real time NMR techniques are used to study dynamic events on a timescale of several milliseconds up to several hours, or ranging from  $k=10\text{s}^{-1}$  to  $k=0,1\text{s}^{-1}$  if expressed in terms of rates. The combination of high resolution spatial data with kinetic experiments is not achieved by any other biophysical method.

Dynamic events of biomacromolecules in this time range are large scale conformational transitions, folding events (for proteins) and macromolecular assemblies.

In order to monitor these events two preconditions have to be fulfilled, first the initiation of the reaction should be performed fast as compared to the velocity of the monitored process and preferentially in the NMR spectrometer to avoid long dead times, second the NMR techniques should be conducted in a manner that the temporal resolution is optimal.

*All kind of techniques that are introduced here, are discussed focusing on their application to nucleic acids or their inherent potential to be used in this field. The author is aware that some of these techniques are introduced for different purposes, but wants also to demonstrate their possible value for the field of RNA folding.*

## TECHNIQUES FOR REACTION INITIATION

### Photo-caged compounds

The idea of photo-caged compounds is that a particular state of a molecular system is stabilised over a second state -that is in equilibrium with the first one or represents the product state of the folding pathway- by introduction of a photosensitive functional group or molecule. By application of a short light flash, or laser pulse respectively, this photosensitive ‘cage’ gets irreversibly and rapidly destroyed and the system can relax towards its equilibrium. In the case of RNA-folding photo-caged compounds can potentially be:

- $\alpha$  the RNA molecule itself
- $\beta$  an RNA - interacting ligand
- $\gamma$  divalent ions that are required for folding

***$\alpha$  Caging the RNA molecule itself*** can be realized by the introduction of a photo-labile functional group at sequence positions that are critically involved in the stabilization of the RNA fold.

In a first example the method is in close analogy to a biological example of structure-selection (see chapter 2). We substituted selected nucleobases with 1-(2-nitrophenyl)ethyl-substituted (= NPE-substituted) nucleobases, which are designed to impair the formation of selected base-pairs.

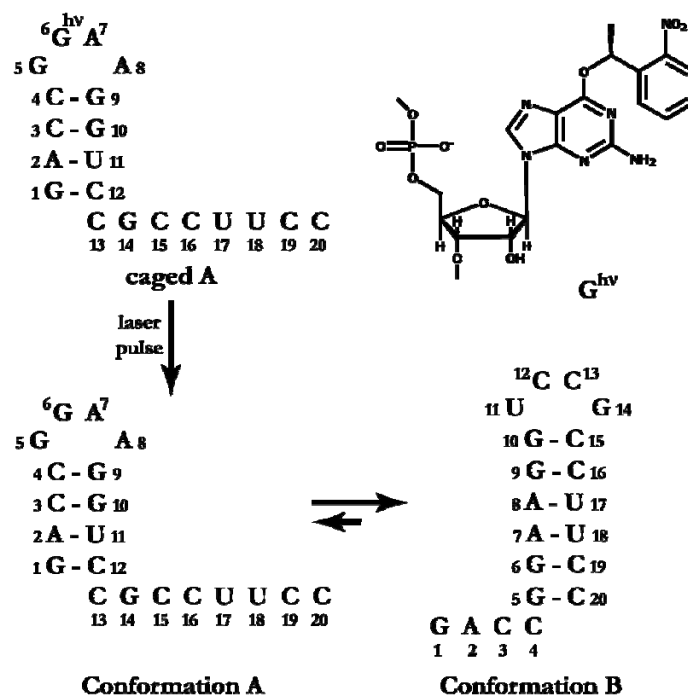


Figure 36: Example of a bistable RNA-sequence that exhibits equilibrium of two conformations. By introduction of a photoprotecting group ( $G^{hv}$ ; upper right) at proper sequence position exclusively conformation A can be stabilized and released towards relaxation into equilibrium by a short laser pulse.

When introduced at proper sequence position within the oligonucleotide fold, these photocleavable modifications lead to a specific destabilization of preselected folds (Figure 36). The bulky and photolabile (2-nitrophenyl)ethyl group (=NPE group) was expected to disrupt Watson-Crick base pairs and thereby to destabilize associated secondary structures. This is mainly due to steric effects because the bulky group camouflages the Watson-Crick side completely and does not allow establishment of the normal WC-basepairing geometry any longer and secondly completely shifts the hydrogen-bond donor- and acceptor-pattern (Figure 37).

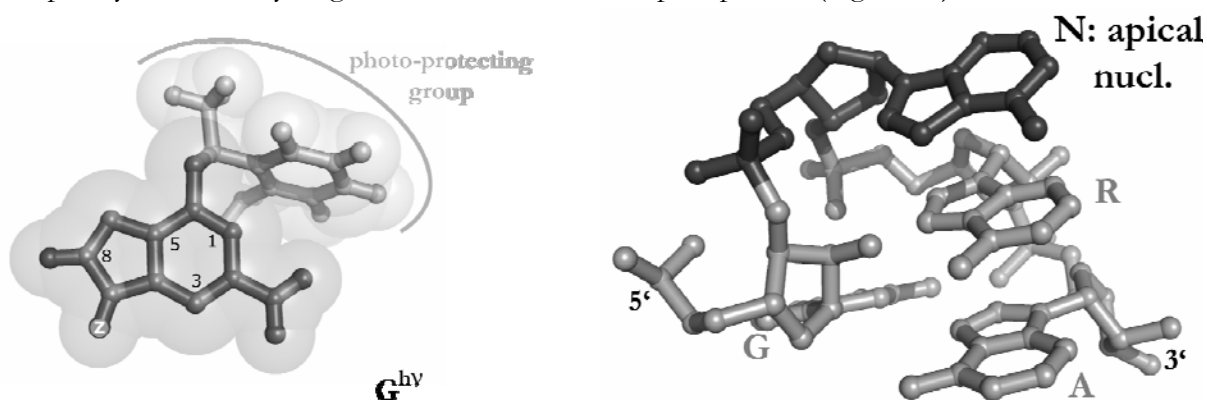


Figure 37: Structural representation of (upper panel): guanine-nucleobase carrying the NPE-photoprotecting group that masks its Watson-Crick site; (lower panel): loop nucleotides of GNRA tetraloop, imposing that at the position of the apical nucleotide N the introduction of the protecting group does not perturb the three-dimensional structure of the loop.

Upon subsequent photolytic cleavage, however, the parent base-pairing property would be fully restored and refolding could occur. A rapid and traceless removal of the photolabile group then allows the refolding process to occur.

Revisiting from the point of energetics the difference in free energy ( $\Delta G$  that directly correlated to the population of the two inter-converting conformations by  $\Delta G = -RT \ln K = -RT \ln \frac{[A]}{[B]}$ ) of both conformations in the equilibrium of the free sequence has to be much smaller than in the case of the photo-protected sequence. In this latter case the introduction of the photoprotecting group has to increase the free energy of conformation B. So is in the depicted case (Figure 38) the free energy difference calculated to be  $\Delta G = -0.6 \text{ kcal mol}^{-1}$ , and the energy penalty for a non-formable Watson-Crick G-C basepair in the case of the photoprotected form of the sequence (the photoprotecting group prevents the formation of G6-C19 in conformation B) is at least in the range of  $\Delta G \approx -3.4 \text{ kcal mol}^{-1}$ .

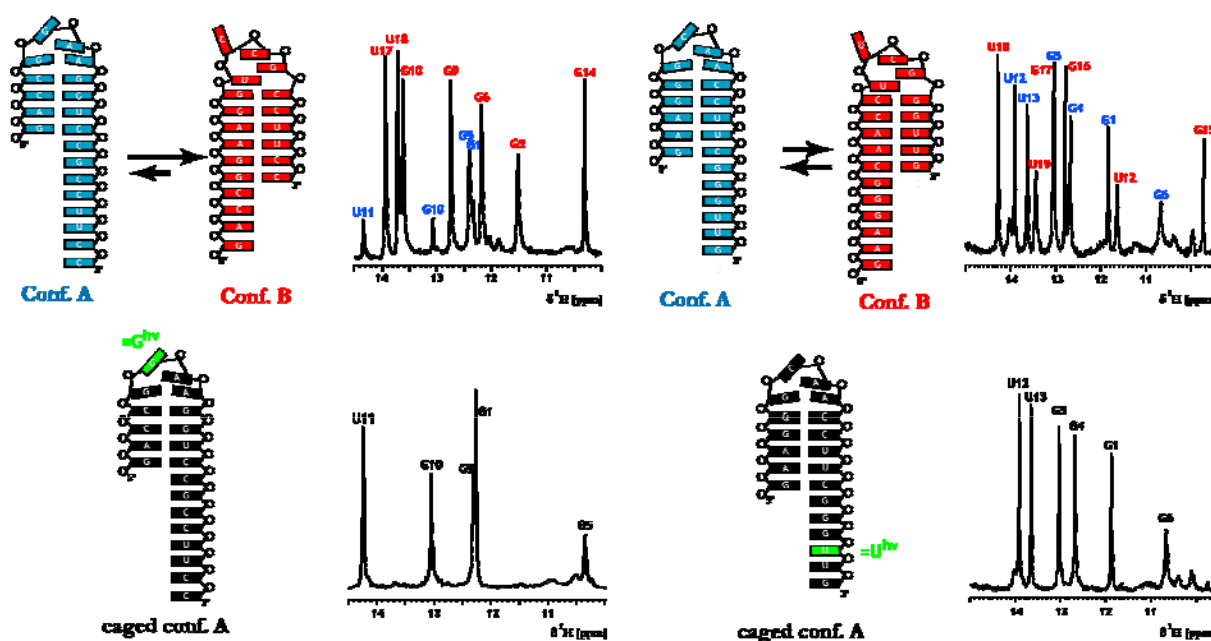


Figure 38: Structure selection by introduction of a photolabile-protecting group monitored by 1D  $^1\text{H}$  NMR spectra: The imino region of the bistable 20mer RNAs exhibit the signals of both conformations (upper panel), whereas upon introduction of a photolabile-protecting group at proper sequence position the sequence is exclusively stabilized in single conformation A as reported by the NMR spectrum that solely display signals of the respective conformation. All spectra were recorded at a Bruker AV800MHz spectrometer equipped with a cryogenic probe at 288K. 128 transients and 4096 complex points were recorded. The relaxation delay was set to 1.5s water suppression was achieved by hard WATERGATE sequence; proton pulses were applied with a field strength of 26.32kHz at the water resonance frequency.

In a second example one can take again advantage of the steric amount of the photolabile-protecting group. Here we tried to mask a magnesium-binding site in the P5abc system. Since the NMR studies<sup>157</sup> reveal that folding just occurs in the presence of 5mM magnesium and the crystal structure<sup>158</sup> identifies the exact binding sites, introduction of the bulky photolabile-protecting group at the sequence position that is involved in the magnesium binding and additionally undergoes a structural transition upon metal binding should inhibit the folding of the RNA (Figure 39).

<sup>157</sup>M. Wu, I. Tinoco Jr., *Proc. Natl. Acad. Sci.*, 1998, **95**, 11555-11560

<sup>158</sup>J.H. Cate, A.R. Gooding, E. Podell, K. Zhou, B.L. Golden, C.E. Kundrot, T.R. Cech, J.A. Doudna, *Science*, 1996, **273**, 1678-1685

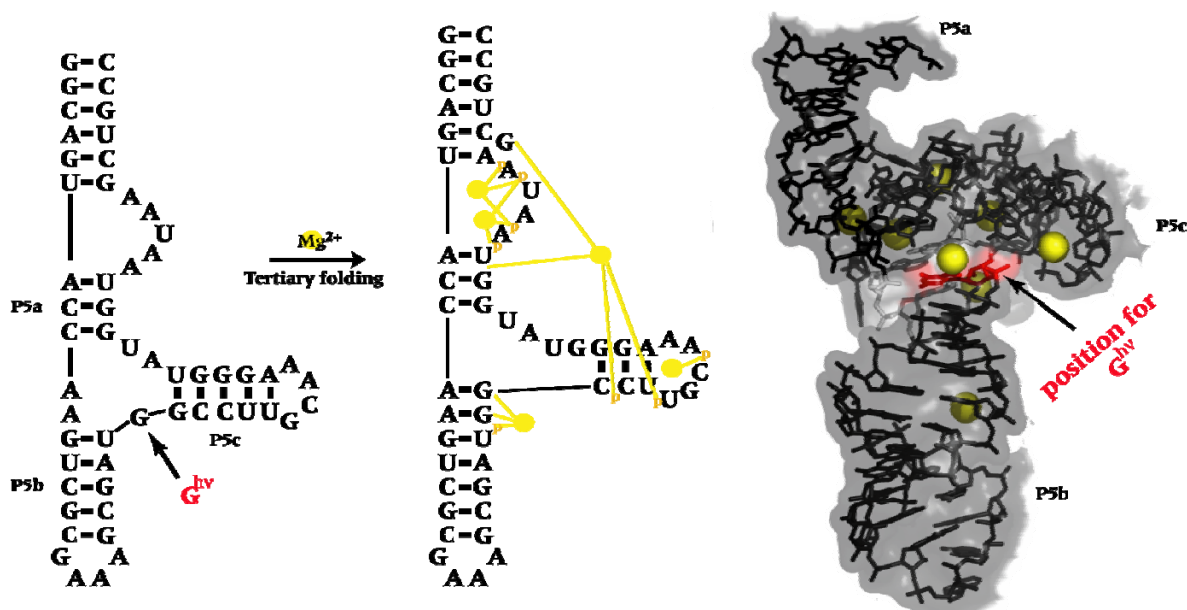


Figure 39: Left panel: secondary structure representation of the magnesium induced folding event of the P5abc RNA molecule. Addition of  $Mg^{2+}$  (yellow balls) leads to a change in tertiary as well as secondary structure; introduction of a photolabile protected nucleotide ( $G^{hv}$ ; shown in red) at a position nearby of a magnesium binding site (see right panel: crystal structure of the folded P5abc element; again yellow balls are  $Mg^{2+}$ ; red nucleotide G at position where  $G^{hv}$  was introduced) should arrest the molecule in the ‘unfolded’ state.

Since the magnesium free and bound states are characterized by NMR spectroscopy, comparison of the photo-protected RNA sequence and the free form reveal that structural transition cannot be inhibited completely. As apparent in the imino-spectral region, the structural parts in the periphery of the molecule seem to adopt already the ‘folded’ conformation. In contrast the structural parts in the proximity of the photo-protected nucleotides seem to adopt still the ‘unfolded’,  $Mg^{2+}$ -free conformation. Upon photolytic cleavage the system fully relaxes into the metal-bound conformation (Figure 40).

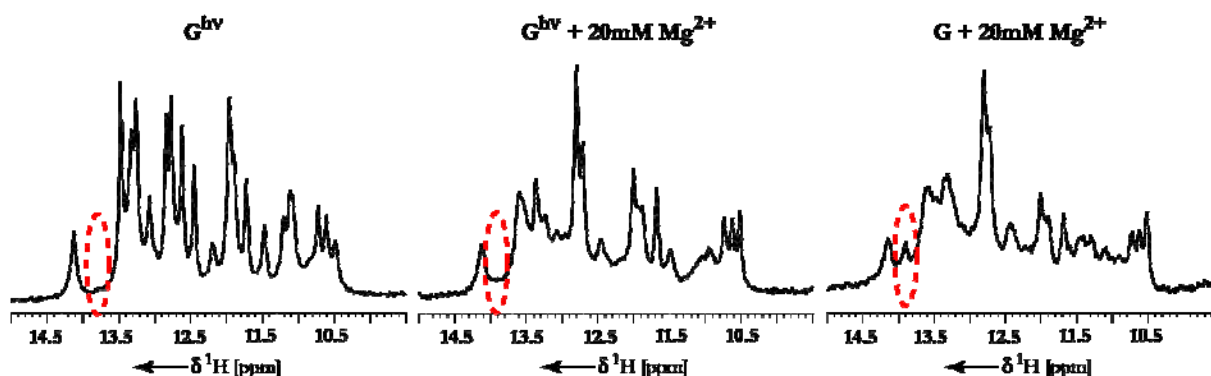


Figure 40: Imino proton region of 1D  $^1H$  spectra of the 56mer P5abc RNA. The spectrum of the RNA containing the photoprotecting group in buffer without  $Mg^{2+}$  (0.3mM RNA in 10mM NaPhosphate pH6.4 at 293K) displays the signature of the ‘unfolded’ conformation (indicative for this conformation is the strong signal of U142 at 14.2ppm) (left panel); after addition of 20mM  $Mg^{2+}$  the spectrum changes, but does not exhibit the typical reporter signals for a complete conformational transition of the ‘folded’ state (empty circle) (mid panel), after photolytic cleavage of the protecting group, the spectrum shows the complete signature of the ‘folded’ conformation (here the indicator is the now shifted signal U142 at 13.9ppm -signal in the red circle) (right panel).



***$\beta$  Caging of a small molecular ligand*** that is needed for the initiation of RNA folding processes can be applied to bimolecular systems like aptamers or riboswitches. The rationale behind this method is to interfere at the binding interface of the small molecular ligand that induces folding or is needed in order to stabilize the RNA molecule in its folded state respectively. Whereas in the case of the P5abc (described in section alpha) the RNA-binding site is masked, the folding event can also be suppressed by masking functional groups of the ligand, that are crucial for binding. This is realized for the guanine-dependent riboswitch. Here small molecules, such as hypoxanthine and guanine, are able to bind the RNA molecule and thereby inducing a major structural reorganisation, both on the level of secondary as well as on the level of tertiary interactions. The recognition of the small metabolite is therefore based on the formation of a base triple (Figure 41) involving the interaction of the Watson-Crick site as well as the N7/N9 edge. As revealed by X-ray crystallographic structural studies and by photo-CIDNP experiments in the RNA bound state the hypoxanthine is furthermore enclosed in a tight and solvent protected binding pocket. The introduction of a similar photo-protecting group at the WC-site of the ligand therefore prevents the binding and the subsequent folding of the RNA molecule by disrupting the formation of the crucial basetriple and also excludes the ligand from the binding pocket by increasing the size of the molecule nearly by a factor of two.

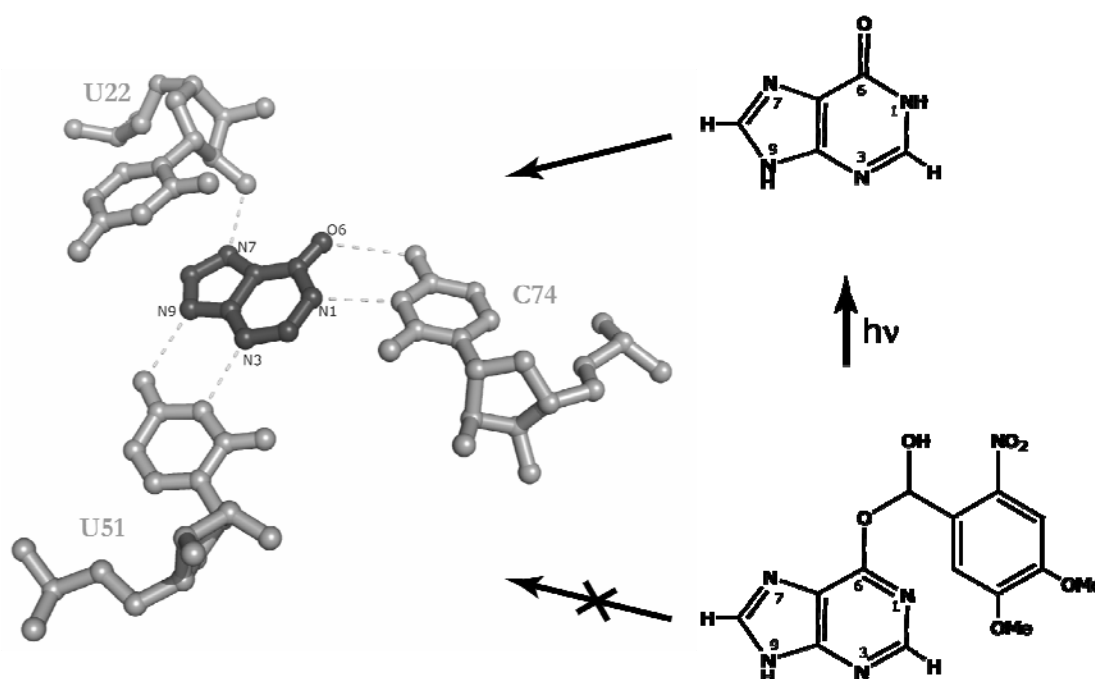


Figure 41: Representation of how to cage a small RNA ligand: left panel: clipping of the X-ray structure of hypoxanthine (dark grey) bound to the guanine riboswitch RNA (light grey), indicating that the ligand is bound into a narrow binding pocket and binding is facilitated by a series of hydrogen bonds spanned by the RNA nucleotides in the proximity of the ligand. The introduction of a bulky photolabile-protecting group therefore inhibits ligand binding by the steric amount of the group (being too large for the binding pocket) and by the reversion of the hydrogen donor and acceptor pattern at the WC-site of the molecule (hydrogen bonds are masked).

*$\gamma$  Caging of divalent metal ions* that are required for the induction of RNA folding may be achieved by photo-labile complex-builders. Like in the case of protein folding studies<sup>159</sup>, the metal ion that is needed for the proper formation of the RNA structure (in the most cases  $\text{Mg}^{2+}$ ) is caged by a complex builder, which has a higher affinity to the metal ion than the bio-macromolecule. Upon photolytic cleavage the caging compound gets decomposed and because the photoproducts exhibit now a lower affinity to the metal ion the ions can be bound by the RNA molecule and thereby inducing the folding event.

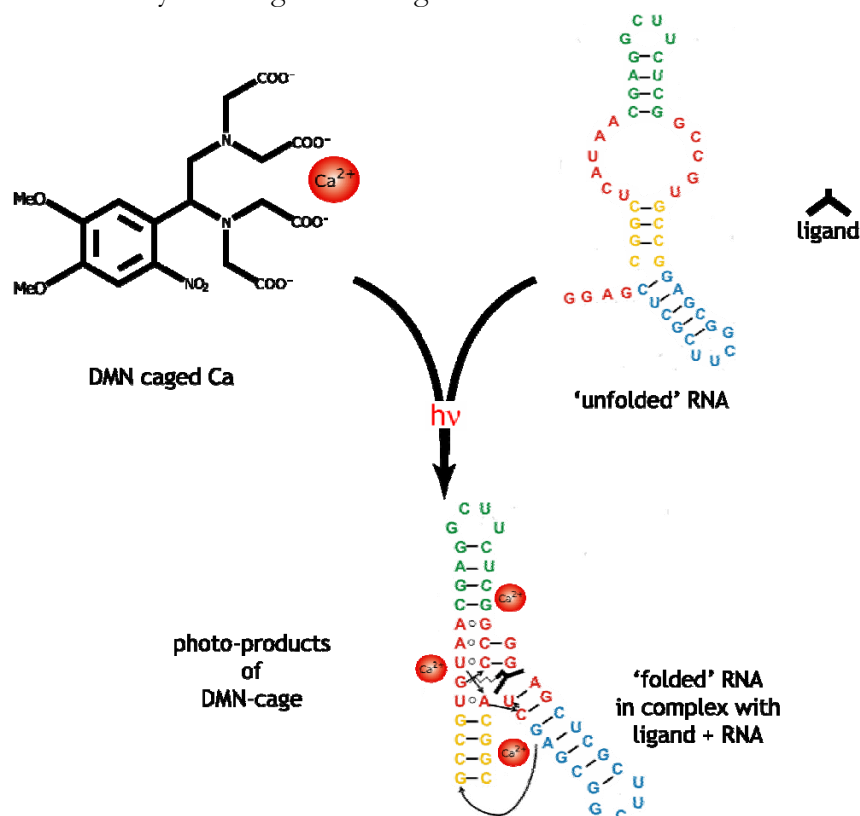


Figure 42: Schematic representation of an example for caging a metal ion to trigger a RNA folding event: the RNA shown here is the Diels-Alderase Ribozyme, that in the presence of a ligand and  $\text{Ca}^{2+}$  (as a substitute for  $\text{Mg}^{2+}$ ) folds out of an extended conformation into a pseudo-knotted RNA molecule with high tertiary structural order. Caging the  $\text{Ca}^{2+}$  with the photo-labile caging compound DMN can prevent this folding step.

As an example the strategy for the investigation of tertiary and secondary structure folding of the Diels-Alderase ribozyme<sup>160</sup> is outlined (Figure 42). This *in-vitro* selected catalytic active RNA molecule folds into its active pseudo-knotted RNA structure<sup>161</sup> upon addition of divalent metal-ions ( $\text{Mg}^{2+}$  or  $\text{Ca}^{2+}$ ) and its small molecular ligand (that may be a educt, product or transition state analogue of the catalyzed Diels-Alder reaction). In order to observe the conformational transitions from its pre-folded form towards the pseudo-knotted structure the molecule is prepared in a solution containing the ligand, and a photo-caged form of  $\text{Ca}^{2+}$  (DMN<sup>162</sup> (Dimethoxy-nitrophen) caged Ca). A short laser pulse that decomposes the caging compound can

<sup>159</sup>T. Kühn, H. Schwalbe, *J. Am. Chem. Soc.*, 2000, **122**, 6169-6174

<sup>160</sup>B. Seelig, A. Jäschke, *Chem. Biol.*, 1999, **6**, 167-176

<sup>161</sup>A. Serganov, S. Keiper, L. Malinina, V. Tereshko, E. Skripkin, C. Höbartner, A. Polonskaia, A.T. Phan, R. Wombacher, R. Micura, Z. Dauter, A. Jäschke, D. Patel, *Nat. Struct. Mol. Biol.*, 2005, **12**, 218-224

<sup>162</sup>G.C.R. Ellies-Davis, J.H. Kaplan, *J. Org. Chem.*, 1988, **53**, 1966-1969

then release the metal ions (the binding affinity of the DMN photoproducts to the Ca-ions is lower [ $K_D=3 \cdot 10^{-3}M$ ] compared to the binding affinity of the RNA to  $Ca^{2+}$  and much lower as the binding affinity of the DMN itself to the Ca-ions [ $K_D=5 \cdot 10^{-9}M$ ]). Upon binding of the freshly released divalent ions the RNA molecules fold into their active conformations, this transition can thereby be followed by RT-NMR methods (Figure 43).

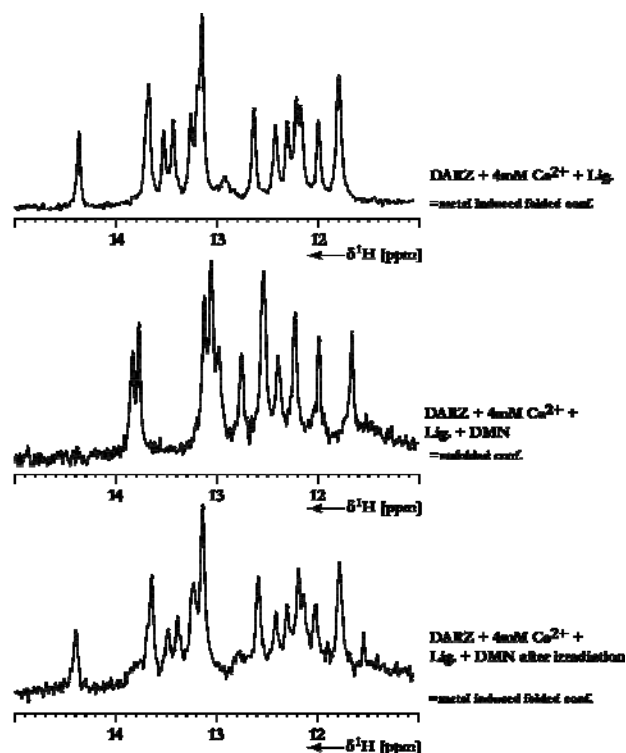


Figure 43: Imino region of 1D  $^1H$  spectra recorded for the Diels-Alderase ribozyme (DARZ) in the presence of metal ion and ligand (folded conformation -upper panel), after addition of photo-labile metal ion caging compound (DMN) that binds the metal ion with higher affinity than the RNA molecule and leads to unfolding of the DARZ (unfolded conformation -middle panel) and after irradiation with laser light to release the metal ions and induce folding of the RNA molecule again (folded conformation -lower panel).

**In order to perform kinetic studies** of folding processes with one of the caging strategies presented above, the preceding cleavage reaction has to occur considerably faster than the refolding, thereby separating the two events in time.

As shown exemplary for the first case ( $\alpha$ ) the parent sequence would be liberated in a defined state by only one short laser pulse. Single scan imino proton spectra of the S-NPE-modified RNA sequence before and 1s after photolysis by one laser-pulse were nearly identical, indicating that the sequence was released in its parent folding structure. However, in order to directly monitor the course of the cleavage reaction, we carried out identical experiments with a 1:1 mixture of the two diastereoisomeric RNAs, obtained by introduction of a R/S-NPE protected guanosine. Upon photolysis by one laser pulse, the two sets of imino proton signals, caused by the two configurations on the photolabile group, collapsed within  $\leq 1$  sec to one set, belonging to the parent RNA-sequence in fold A. These experiments unambiguously showed that the folding structure initially imposed by the photolabile group was completely conserved

immediately after its photolytic removal. With one laser-pulse, almost complete deprotection of RNA samples  $c \leq 0.1\text{mM}$  could be established.

The advantages and disadvantages of the different possible techniques are summarized in the following table:

Table 8: Advantages and disadvantages of different caging techniques

	RNA	Caging the Small ligand	Metal ion
Advantages	Study of equilibrium and non-equilibrium processes possible	Different heteronuclear labelling schemes possible - easy -inexpensive	Different heteronuclear labelling schemes possible - easy -inexpensive
	Total concentration of the photo-labile group is low = good deprotection yield	Total concentration of the photo-labile group is low = good deprotection yield	Synthesis of metal ion caging compounds is easy and well established
	Study of folding processes at constant ionic strength and metabolite / buffer concentration		
Disadvantages	Solid phase synthesis of RNA needed	Synthesis of small ligand can be difficult	Total concentration of the photo-labile group is high = bad deprotection yield
	$^{15}\text{N}$ or/and $^{13}\text{C}$ labelling is difficult and expensive	Solubility of (caged) ligand can be very poor	High concentration of photo-by products may interfere with folding

Caging the RNA directly is the only way to monitor the kinetics of equilibrium processes at a slow timescale. The advantage of this concept is furthermore the high photolytic yield because the photosensitive compound has the same (low) concentration as the RNA molecule. The main disadvantage of the concept is that synthesis of such modified RNA sequences has to be done by solid phase chemistry and is therefore limited in size and labelling potential. This particular disadvantage is overcome in the two other cases because here the photolabile compound and the monitor molecule (= the RNA) are decoupled so that labelling and synthesis can be achieved by *in vitro* transcription. The synthesis of the photolabile compound itself is well established for the case of the metal caging compound but is hardly a literature known strategy for the plethora of molecules acting as ligands for RNA. So in the second case for each new ligand a new synthetic strategy has to be developed. Also the solubility of the photocaged ligand has to be taken into account since most of the protecting groups are aromatic systems with poor solubility in water but with a molecular mass that may be equal to the one of the compound that has to be caged. Nevertheless these three techniques are complementary and facilitate research on RNA folding with focus on different questions.

**Photolytic deprotection is achieved in the NMR spectrometer** using a setup<sup>159</sup> that was initially developed for protein folding studies (see Figure 44). The slightly modified system consists of a CW-Ar-ion laser {BeamLok 2060 (Spectra Physics) 9.7W power operating at 340-365nm} whose beam is conducted to the sample tube via an optical fibre {LWL-Kabel UV1000/1100N (Ceram Optec)}. A high power resistant lens {lens: U13X (Newport) mounted in a combined lens and fibre holder: F-91-C1-T, FPR1-C1A (Newport)} obtains coupling of the laser beam into the fibre {other optical components used -beamsplitter 200-0 (CVI), -mirror 200-10 (CVI)}. Irradiation of the complete sample is maintained using pencil shaped Shigemi tips<sup>163</sup>. A shutter {shutter: uniblitz electronic VS14S2ZM1 (Vincent Associates); shutter-driver: uniblitz VCM-D1 (Vincent Associates)}, connected to the spectrometer blocks the laser beam in front of the fibre and can trigger the laser irradiation.

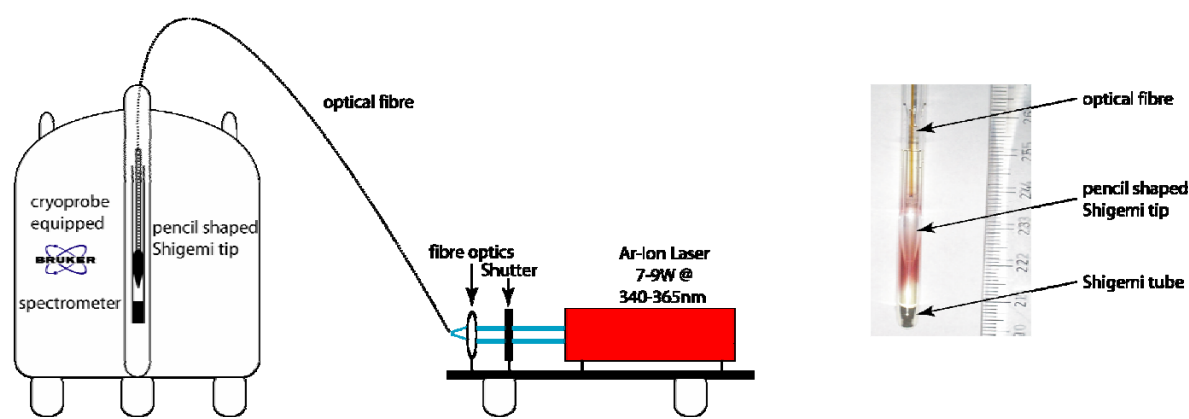


Figure 44: Left panel: Schematic representation of laser coupling to the NMR-sample: The laser beam is transmitted via an optical fibre into the sample residing in the active site of the NMR spectrometer; right panel: a pencil shaped shigemi tip connected to the optical fibre and sitting in a sample of reduced volume achieves uniform sample illumination.

## Mixing techniques

As first described for organic reactions in the early 1970s rapid mixing of reactants and monitoring the reaction by NMR-spectroscopy is feasible and opened the route for monitoring non-equilibrium reactions between different states in real-time.<sup>164</sup>

Essentially two strategies for concentration-jump-, respectively mixing-, techniques are possible: the first represent mixing of the solutions outside the NMR probe (as equivalent to optical stopped flow spectrometers) and then rapid delivery of the mixture into the radio frequency coil-region<sup>165</sup>; the second uses the NMR-tube itself as the mixing chamber<sup>166</sup>. Here also two possibilities are amenable, where in the first method one solution is injected into another inside

<sup>163</sup>Although Kuprov et al. report that usage of a tapered fibre allows more efficient and uniformly irradiation of the sample with a tapered fibre, we couldn't reproduce this better irradiation by a self made tapered fibre prepared according to their protocol: I. Kuprov, P.J. Hore, *J. Magn. Reson.*, 2004, **171**, 171-175

<sup>164</sup>J. Grimaldi, J. Baldo, C. McMurray, B.D. Sykes, *J. Am. Chem. Soc.*, 1972, **94**, 7641-7645

<sup>165</sup>R.O. Kühne, T. Schaffhauser, A. Wokaun, R.R. Ernst, *J. Magn. Reson.*, 1979, **35**, 39-67

<sup>166</sup>D.B. Green, J. Lane, R. Wing, *Applied Spectroscopy*, 1987, **41**, 847-851

the NMR-tube, so that after the injection the turbulences cause homogeneous and rapid mixing<sup>167</sup>. The second method uses mobile separators of the two solutions that can be removed rapidly and thereby facilitates mixing<sup>168</sup>.

For the very early mixing instruments a spectrum with sufficient resolution could be recorded already after a dead time of  $\tau_{\text{dead}}=0.2\text{s}$ , and the subsequent spectra with repetition rates of 0.6s per spectra.<sup>169</sup> One of the first kinetic studies conducted on biomacromolecules using a NMR mixing device was a study of the alkaline denaturation of hemoglobine where the NMR deadtime was already reduced to  $\tau_{\text{dead}}=0.06\text{s}$  and where a detection of bulk protein peaks was possible.<sup>170</sup>

Nowadays the typical deadtime is in this range and after approx. 50ms mixing is completed and a sufficient field-homogeneity is achieved. If applying such a methodology to RNA folding experiments the effect of shearing forces to larger RNA-molecules has to be considered – but normally the NMR size limitation should be under the critical cut-off that would make usage of rapid mixing devices impossible.

### Temperature-jump methods

A possible method to induce conformational changes with different amplitudes in RNA molecules is to increase the temperature. If the increase of temperature is significantly faster than the response of the system, kinetics of the establishment of the new conformation can be extracted. Currently, there are three different strategies published to perform fast temperature-jump experiments in high-resolution NMR spectroscopy: (a) CO<sub>2</sub> laser heating applied in solid state MAS NMR spectroscopy<sup>171</sup>, (b) microwave-assisted heating<sup>172</sup> and (c) conductive heating using probe inserts applying rf-heating in the interior of the NMR-tube<sup>173</sup>. Most successful and easiest to implement is rf-heating since the design can be coupled to any probe including cryoprobes. The performance of the system is reported as follows: 0.15-0.4K/Ws and powers of 200W are applied for less than 100ms. The rate of temperature jump can be improved by increasing the rf-power of the amplifier with reasonable fast decay of temperature gradients after the application of the temperature pulse. As the downside of the published solution (in c) so far, the field homogeneity is not optimal and the filling factor is compromised by having the additional rf-circuit in the interior of the main rf-coil.

<sup>167</sup>K. H. Mok, T. Nagashima, I.J. Day, J.A. Jones, C.J.V. Jones, C.M. Dobson, P.J. Hore, *J. Am. Chem. Soc.*, 2003, **125**, 12484-12492

J. Balbach, V. Forge, N.A.J. van Nuland, S.L. Winder, P.J. Hore, C.M. Dobson, *Nature Struct. Biol.*, 1995, **2**, 868-870

<sup>168</sup>M. Spraul, M. Hofmann, H. Schwalbe, *U.S. Patent 5,726,570*, 1998

<sup>169</sup>D.A. Couch, O.W. Howarth, P. Moore, *Journal of Physics E*, 1975, **8**, 831-833

<sup>170</sup>W.A. McGee, L.J. Parkhurst, *Anal. Biochem.*, 1990, **189**, 267-273

<sup>171</sup>D.B. Ferguson, T.R. Krawietz, J.F. Haw, *J. Magn. Reson.*, 1994, **109A**, 273-275

H. Ernst, D. Freude, T. Mildner, I. Wolf, *Anal. Chem.*, 1995, **67**, 3342-3348

<sup>172</sup>A. Naito, A. Nakatani, M. Imanari, K. Akasaka, *J. Magn. Reson.*, 1990, **87**, 429-432

M. Kawakami, K. Akasaka, *Rev. Sci. Instr.*, 1998, **69**, 3365-3369

<sup>173</sup>P.V. Yushmanov, I. Furó, *J. Magn. Reson.*, 2006, **181**, 148-153

## TECHNIQUES TO RECORD RT-NMR SPECTRA

## 1D-Methods

The fundamental RT-NMR technique is to monitor the dynamic process by subsequent recording of 1D spectra during the time course of the reaction. In the present work this approach is applied in NMR experiments that are constructed as pseudo 2D or 3D experiments.

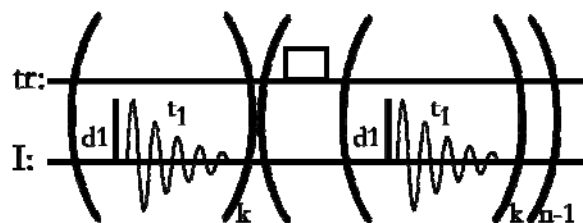


Figure 45: General scheme of kinetic pseudo-3D experiment (pulse sequence codes are found in appendix). Before the first trigger (*tr*) is applied a complete pseudo 2D plane of *k* 1D spectra is recorded, equally after all *n-1* trigger events such a plane is recorded.

So the dimensions of the experiment are the chemical shift ( $\delta^1\text{H}$  along  $\omega_2/\omega_3$ ) and the real time (RT along  $t_1/t_1$  and  $t_2$ ). The first plane in a pseudo 3D is recorded before application of the first trigger, whereas all following planes are recorded after a respective trigger event. All  $1+n$  planes comprise therefore a single kinetic trace (Figure 46). If a single trigger event is sufficient to transfer the system from non-equilibrium to full equilibrium the pseudo-dimensionality can be reduced to two and the trigger is then located in the midpoint of the pseudo-second dimension.

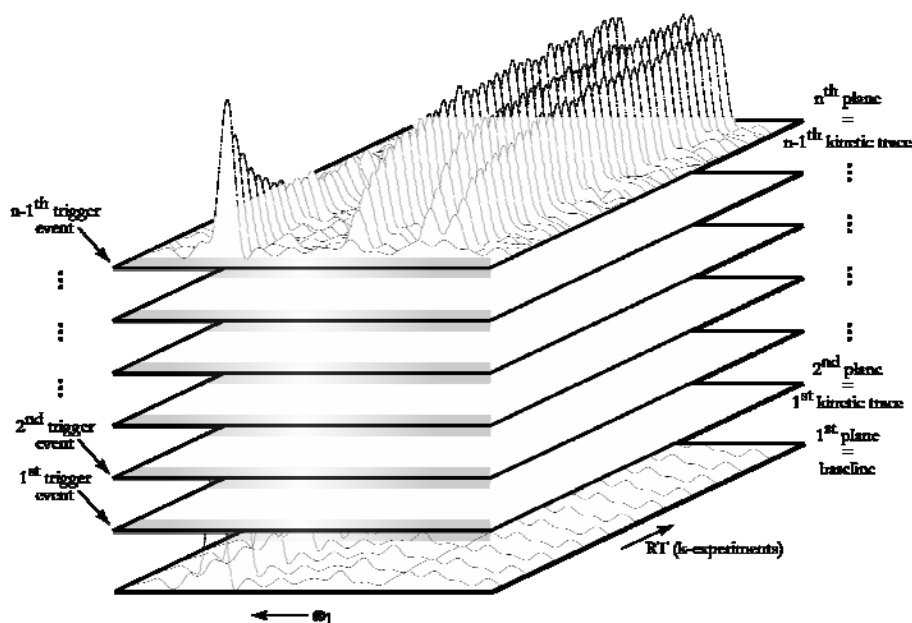


Figure 46: Schematic representation of pseudo-3D as realized in the kinetic studies reported within this thesis. The first plane is used to establish a baseline for the following kinetic planes. In all following planes kinetic traces that monitor the dynamic reaction initiated by the trigger are recorded.

The advantage of the 1D method is its intrinsic high sensitivity at highest temporal resolution. Normally  $^1\text{H}$ ,  $^{19}\text{F}^{(174)}$  and  $^{31}\text{P}^{(175)}$  are the nuclei on which the experiments are conducted regarding

<sup>174</sup>S.D. Hoeltzli, C. Frieden, *Proc. Natl. Acad. Sci.*, 1995, **92**, 9318-9322

their high natural abundance (99,9885%, 100%, 100%, respectively) and their high sensitivity (2,79, 2,62, 1,13, relative value of z-component of nuclear magnetic moment in units of nuclear magneton, respectively). Despite its intrinsically high sensitivity the spectral resolution of the 1D method is limited. It is best applied to spectral regions of high intrinsic resolution as the imino signal region for RNA. Furthermore this limitation can be overcome by using selectively labelled compounds and application of filter experiments. As an example in the case of kinetic studies on RNA it is possible to increase resolution by using a selective  $^{15}\text{N}$ -uridine-labelled RNA which allows varying detection of either the  $^{15}\text{N}$ -bound protons, or the  $^{14}\text{N}$ -bound protons, meaning that either only uridine or the guanosine residues are visible. In order to facilitate detection of  $^{15}\text{N}$ -bound signals as well as  $^{14}\text{N}$ -bound signals during a single kinetic experiment, application of selective filter techniques<sup>176</sup> in HMQC-fashion is necessary (Figure 47). In such an experiment, the phase of the first  $^{15}\text{N}$   $90^\circ$ -pulse, that acts as the discriminator, is incremented interleaved in  $m$ , where  $m$  is the number of subsequent 1D-spectra. This setup results in inverted signs of the coupled ( $^1\text{H}$ - $^{15}\text{N}$ ) and uncoupled ( $^1\text{H}$ - $^{14}\text{N}$ ) spin systems in the first data point, while the signs of both are the same in the second one. By adding the two, one can extract the kinetic traces of the  $^{14}\text{N}$ -bound proton signals while subtraction leads to analysis of the  $^{15}\text{N}$ -bound protons.

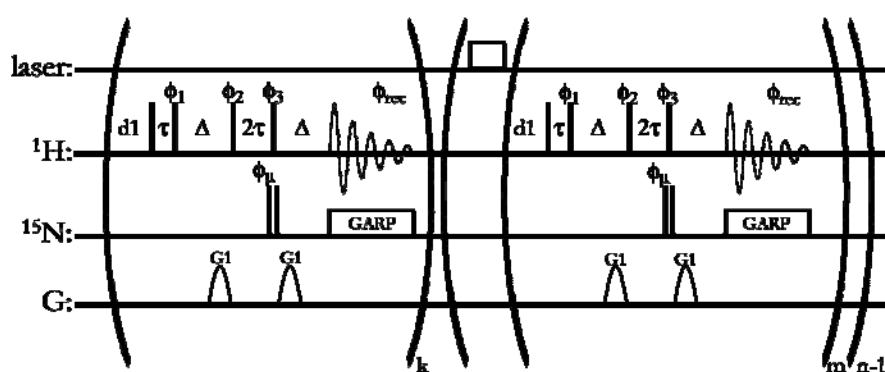


Figure 47: Pulse sequence for a pseudo 3D  $x$ -filtered/edited jump-return echo experiment with laser irradiation as trigger for the kinetic event. Gradients have a SINE,100 shape and are applied with strength of  $g\beta 1:90\%$ . Nitrogen decoupling is realized by a GARP decoupling sequence during acquisition. All pulses if not stated explicitly are applied with phase  $x$ ,  $\phi_1=(x, x, y, y, -x, -x, -y, -y)$ ,  $\phi_2=(-x, -x, -y, -y, x, x, y, y)$ ,  $\phi_\mu=(x, -x)$  is incremented interleaved in the cycle  $m=2*k$ , the receiver phase is  $\phi_{rec}=(x, x, -x, -x)$ . The delay  $\Delta$  is set to match the one bond scalar coupling  $-\Delta=1/(2*\gamma J(\text{NH}))$ , and  $\tau$  is the jump return delay.

### Monitoring RNA folding kinetics on resonances of imino protons

The folding of RNA molecules is monitored primarily on the imino proton resonances. The advantage of that attempt is the highest intrinsic spectral resolution obtained. Comparison of signals per spectral region for a given RNA molecule in non deuterated solvent indicates that no other spectral region omits a better resolution (see the following table). Furthermore the imino resonances monitor the basepairing behaviour of the nucleotides in the RNA molecule which is major determinant of the structural elements and the stability of the whole RNA. Therefore the

<sup>175</sup>K. Werner, PhD Thesis, 2007, Frankfurt, Germany

<sup>176</sup>G. Otting, H. Senn, G. Wagner, K. Wüthrich, *J. Magn. Reson.*, 1986, **70**, 500-505



presence of an imino-signal can be directly understood as a long-range restrain that represents a basepair between two distinct nucleotides of the RNA molecule.

Table 9: Spectral regions and amount of signals per spectral region (theoretical maximum number; calculated for 14mer RNA)

spectral region	atom	number	ppm / signal
15 – 9.5	Imino	8	0.7
	Amino	22	
8.5 – 6.5	H8	6	0.05
	H6	8	
	H2	1	
6.5 – 4.7	H5	8	0.08
	H1'	14	
4.7 – 2.5	H2' - H5''	84	0.03

But since the imino protons are labile with respect to exchange it is important to know if there may be a difference in the exchange rate of the two (or more) states in-between the folding is characterized. If there would be a major difference in the exchange rate, the apparent rates of folding would be influenced by that and could not directly be interpreted as rates of folding. This problem was addressed as an example in the case of the guanine-dependent riboswitch. Here the solvent exchange rates were determined as described above for the ligand free (unfolded) and for the ligand bound (folded) state (Figure 48).

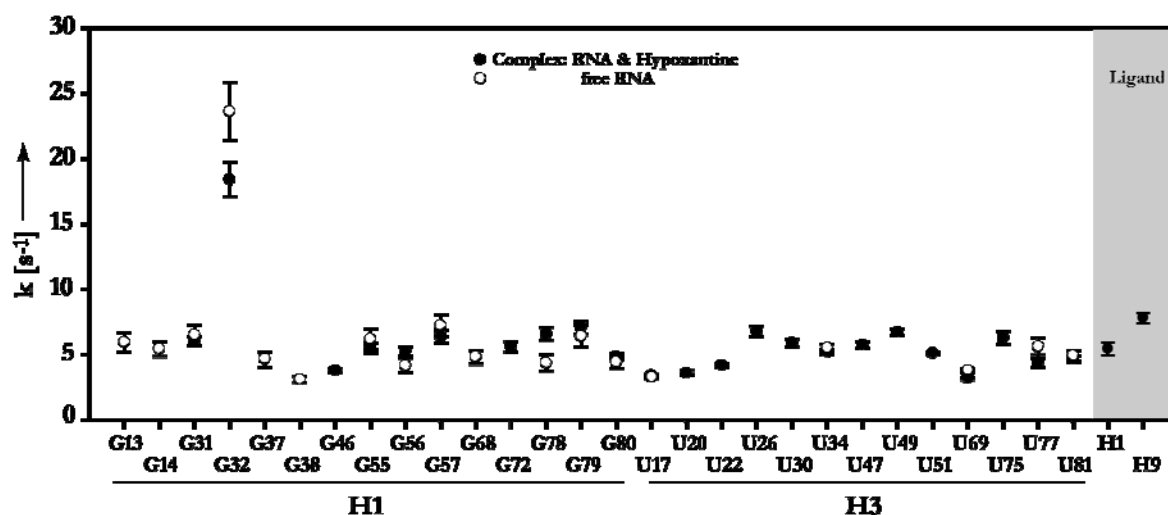


Figure 48: Solvent exchange rates as measured by selective inversion experiment for the free guanine riboswitch (free RNA: open circles) and for the ligand induced folded guanine riboswitch (complex: RNA & Hypoxanthine: closed circles).

As seen in figure 12 the water-imino exchange rates ( $k$  [ $s^{-1}$ ]) show very similar values (varying between  $3.1s^{-1}$  and  $6.7s^{-1}$ ) for both states of the RNA with the only exception is residue G32 with an exchange rate of  $23.6s^{-1} \pm 2.2s^{-1}$  in the free and  $18.4s^{-1} \pm 1.3s^{-1}$  in the ligand-complexed form of the RNA. Interestingly, the crystal structures of the RNA-ligand complexes<sup>177</sup> reveal that G32 stacks between the loop L2 nucleotide A33 and the loop L2 closing base pair. The high exchange

<sup>177</sup>R.T. Batey, S.D. Gilbert, R.K. Montange, *Nature*, 2004, **432**, 411-415

rate indicates a high flexibility for the tertiary interaction in which G32 is involved and allows no interpretation of the obtained kinetics. However, the very similar values of our iminoproton exchange rates allow analysis of the folding rates reflecting only the global folding event.

### Optimized 1D method

Although 1D methods represent the intrinsically highest temporal resolution they are limited in this respect by the  $d1$  relaxation delay that is needed to start every experiment with a magnetically relaxed system. The parameter that describes the behaviour of the system depending on the relaxation delay is the signal-to-noise ratio per experimental time and therefore this has to be optimized. The methodology of choice to tweak this parameter is Ernst-angle excitation<sup>178</sup>; meaning that for fast repetitive pulse experiments, small pulse rotation angles are required to achieve maximum signal amplitude.

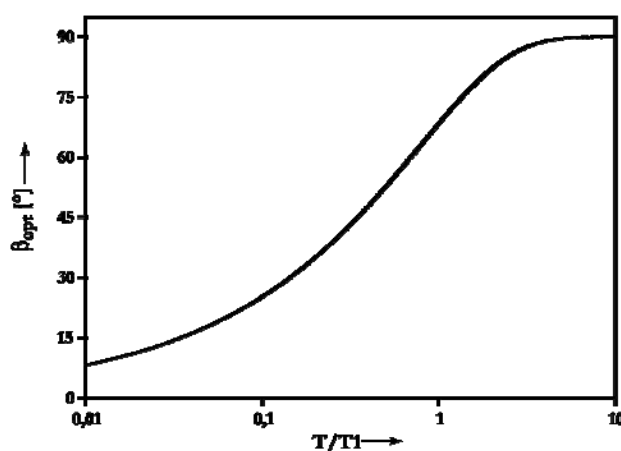


Figure 49: Profile of the optimal pulse flip angle per ratio of experimental time over longitudinal relaxation time. The grey box indicates the value for an optimal S/N of subsequent recorded 1D experiments (jump-return echo pulse sequence, with an acquisition time of  $aq \sim 110\text{ms}$  and a recycle delay of  $d1 \sim 100\text{ms}$ ) to follow the folding kinetics of a 49nt long RNA (DARZ-ribozyme) with a relaxation time of  $T1 \sim 277\text{ms}$ . For this scenario the optimal flip angle is set to  $\beta_{opt} \sim 62^\circ$ .

This can be derived as follows: The amount of z-magnetization after a pulse ( $M_z^{ap}$ ) is described as the z-magnetisation before the pulse ( $M_z^{bp}$ ) multiplied by the cosine of the pulse flip-angle  $\beta$ :

$$M_z^{ap} = M_z^{bp} \cdot \cos \beta$$

The longitudinal magnetization at the end of a complete scan ( $T$  = duration of pulsetrain + acquisition time + recycle delay) is given by

$$M_z(T) = M_z^{ap} \cdot \exp\left(-\frac{T}{T_1}\right) + M_0 \left(1 - \exp\left(-\frac{T}{T_1}\right)\right)$$

<sup>178</sup>R.R. Ernst, G. Bodenhausen, A. Wokaun, *Principles of nuclear magnetic resonance in one and two dimensions*, International series of monographs on chemistry, 1986, **14**, Oxford Science Publications

If after a complete scan the magnetization is set equal to the magnetization that is wanted before the next pulse  $M_z(T) = M_z^{bp}$  it is found:

$$M_z^{bp} = M_0 \frac{1-E}{1-E \cos \beta} \text{ with } E = \exp\left(-\frac{T}{T_1}\right)$$

This results in an initial amplitude of the FID of

$$M_x^{ap} = M_0 \frac{1-E}{1-E \cos \beta} \sin \beta.$$

To define the maximum of magnetization the first derivative of this term is set to 0 and solved for  $\beta$ :

$$\frac{dM_x^{ap}}{d\beta} = 0 = \frac{\cos \beta(1-E) + (-E + E^2)}{1 - 2E \cos \beta + E^2 \cos^2 \beta} \text{ then } \cos \beta = E$$

and therefore the optimal flip angle is determined to

$$\beta_{opt} = \arccos\left(\exp\left(-\frac{T}{T_1}\right)\right)$$

As depicted in Figure 49 the shorter an experiment (T) gets at a constant  $T_1$ —which means that the recycle delay is decreased— the smaller the flip angle has to be in order to get a maximum of signal.

### Multidimensional (nD)-Methods

The intrinsic highest temporal resolution of one dimensional RT-NMR techniques is paid by their limited spectral resolution. To circumvent this limitation switching to nD-NMR methods is necessary. These multidimensional methods are providing a higher spectral resolution for resolving nearly all nuclear sites in a RNA molecule by correlating frequencies of different nuclei and spreading these correlations over several frequency axes. Regrettably, because classical multidimensional NMR methods can be basically understood as modulated series of 1D spectra, these methods are time consuming and the temporal resolution is decreased. This long total-acquisition time hampers the applicability of those experiments to conduct real-time studies of slow dynamic processes. Generally the sensitivity of the experiment determines the time required for its acquisition. Nowadays recent advances in spectrometer- and especially probe-technology (implementation of cryoprobes<sup>179</sup>) have increased the general sensitivity and therefore decreased the amount of time needed. Therefore the time, which is obligatory for most experiments, is determined by the scanning rate that in turn is defined by the phase-cycle and the frequency-

<sup>179</sup>P. Styles, N.F. Soffe, C.A. Scott, D.A. Cragg, F. Row, D.J. White, P.C.J. White, *J. Magn. Reson.*, 1984, **60**, 397-404

sampling. In order to facilitate kinetic studies with nD experiment this problem has to be overcome.

### *Accordion//modulated 2D*

Accordion spectroscopy and modulated 2D experiments are analogue techniques to reduce the dimensionality of spectra that are used to analyse dynamic processes. Accordion was developed in order to measure build up curves of exchange peaks, where normally a 3D experiment, composed of 2D-EXSY spectra at multiple mixing times, is required. The resulting signals dependent on three variables  $S(\omega_1, \tau_m, \omega_2)$ , and a third FT leads subsequently to a new frequency domain. To simplify the experiment and reducing the dimensionality one can couple  $t_1$  and  $\tau_m$  in a skew projection. Therefore incrementation of  $\tau_m$  is done concerted as a fraction of  $t_1$  ( $\tau_m = \chi t_1$ ). Instead of a 3D spectrum now a 2D spectrum is obtained where the lineshape of the signal along  $\omega_1$  is modulated by the rate of the dynamic process.

In close analogy to this methodology 2D techniques are developed that report on a dynamic process by recording a spectrum while the process proceeds resulting in a modulation of the lineshapes. In order to obtain such a spectrum the time of the process and the overall time of the experiment have to match. The reaction is initiated as recording of the experiment starts. The spectrum then contains the resonances of the different states modulated by their respective kinetics<sup>180</sup>. For faster processes the recording of a sensitive experiment with high chemical shift dispersion along the indirect dimension such as a HSQC is possible<sup>181</sup>, for slow processes also the application of experiments that need longer total experimental time such as NOESY spectra are suitable<sup>182</sup>.

The resulting lineshapes are modulated by the result of the FT of the kinetic process along the indirect dimension.

### *Hadamard*

In conventional multidimensional NMR experiments frequencies are explored indirectly by allowing them to process freely during one or more successive evolution periods after an excitation that was carried out by hard pulses. This method reflects an extensive and uniform sampling, which is not dedicated purely by the signal-to-noise ratio. The large number of increments of the indirect evolution periods is mainly needed to achieve satisfactory definition and spectral width in the indirect dimension. In order to achieve pure in-phase spectra a minimal phase cycle has to be applied that also elongates the experimental time. And although all frequencies are monitored, yet most of them carry only noise. Just a small subset contains the resonances that are wanted to be observed.

---

<sup>180</sup>J. Balbach, V. Forge, W.S. Lau, N.A.J. van Nuland, K. Brew, C.M. Dobson, *Science*, 1996, **274**, 1161-1163

<sup>181</sup>M. Zeeb, J. Balbach, *Methods*, 2004, **34**, 65-74

<sup>182</sup>J. Balbach, C. Steegborn, T. Schindler, F.X. Schmidt, *J. Mol. Biol.*, 1999, **285**, 829-842

If instead the actual frequencies are known, regular sampling is no longer necessary. Approaches to reduce the experimental time are then Hadamard-NMR experiments<sup>183</sup> in which frequency-domain excitation -using polychromatic pulses<sup>184</sup> on multi channels- with a Hadamard encoding scheme is applied. In the processing of those experiments the resonances get disentangled by reference to the encoding scheme.

The encoding of the rf-channels is realized by the signs of a Hadamard matrix<sup>185</sup>. A Hadamard matrix of grade  $n$  is a  $n \times n$  matrix, solely composed of -1 and +1. All rows and columns are orthogonal to each other. A Hadamard matrix is therefore defined by

$$\vec{H} \cdot \vec{H}^t = \vec{H}^t \cdot \vec{H} = n \cdot \vec{I}$$

(H: Hadamard matrix;  $H^t$ : transposed Hadamard matrix; I: identity matrix). It is believed that Hadamard matrices exist for all  $n = 4k$  if  $k$  is an integer. The largest matrix found so far is for  $n=668$ .

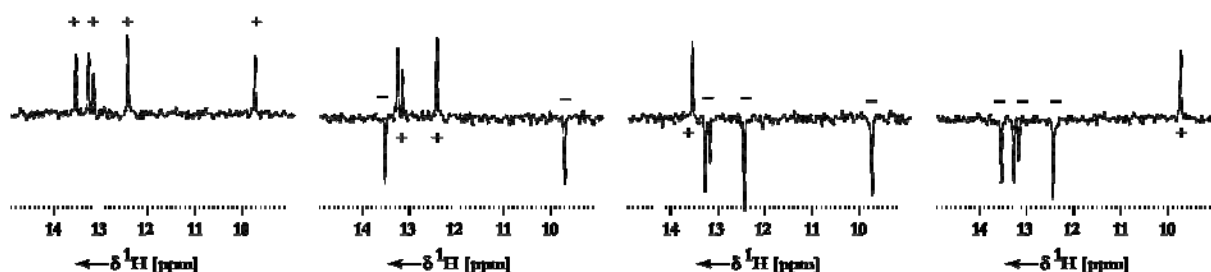


Figure 50: Processed subsequent FIDs of a Hadamard HSQC experiment recorded on the 14mer RNA. The signals are modulated by the signs of the Hadamard matrix (as indicated by the signs in the grey circles). Disentanglement results in a Hadamard 2D spectrum that is shown in figure 52.

A pulse sequence (Figure 51) for such an experiment is very similar to the conventional type of pulse sequence. In the case of a HSQC experiment a list of shaped  $180^\circ$  pulses is applied to  $^{15}\text{N}$  simultaneously but modulated with the encoding matrix instead of the evolution period  $t_1$ . This enables the wanted polychromatic excitation and decoding of the signals according to the frequency of the shaped pulses. The sign of the resulting signals in each FID is therefore identical to the sign of the respective position in the Hadamard matrix (as illustrated in Figure 50).

The construction of the 2D spectrum is then just the combination of the FT of the direct dimension plus the disentanglement according to the matrix form and the chemical shift information given by the respective selective pulses. As shown in Figure 52 a Hadamard spectrum contains only the resonances that have been irradiated by the selective pulses. Therefore the spectrum contains as many peak-carrying-lines along the indirect dimension as selective pulses (=number of rows/columns in the matrix) were applied. The linewidth in the indirect dimension is identical to the spectral width of a single selective pulse.

<sup>183</sup>E. Kupce, R. Freeman, *J. Magn. Reson.*, 2003, **162**, 300-310

E. Kupce, R. Freeman, *J. Magn. Reson.*, 2003, **163**, 56-63

<sup>184</sup>E. Kupce, R. Freeman, *J. Magn. Reson. A*, 1994, **108**, 268-273

<sup>185</sup>J. Hadamard, *Bull. Sci. Math.*, 1893, **17**, 240-248

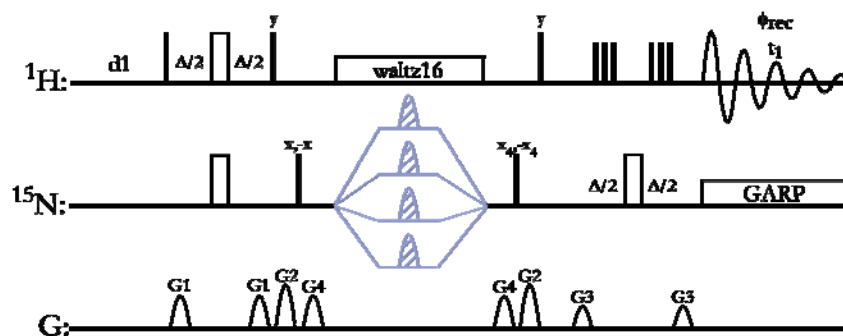


Figure 51: Pulse sequence of a Hadamard  $^{15}\text{N}$ -HSQC; the indirect time period is replaced by an array of selective pulses that are modulated by a Hadamard matrix. All gradients have a SINE.100 shape and are applied with strength of gp1:50%, gp2:80%, gp3:30% and gp4 40%. The  $^{15}\text{N}$  selective  $180^\circ$  pulses with a Gaussp.100<sup>186</sup> profile are applied simultaneously at the respective frequency of the wanted resonances. Nitrogen decoupling is realized by a GARP decoupling sequence during acquisition. Proton decoupling achieved in the first 'dimension' is realized by a waltz16 decoupling sequence during application of the polychromatic pulses. If not stated explicitly all pulses are applied with phase  $\times$ , the receiver phase is  $\phi_{\text{rec}} = (\times, -\times, \times, -\times, -\times, \times, -\times, \times)$ .

As illustrated on the spectrum of the 14mer RNA, Hadamard methods allow the recording of a 2D HSQC in several seconds rather than several minutes. A drawback might lie in the case of overlapping signals along the indirect dimension. Those resonances get excited similarly by the band selective pulse and give rise to signals that are at the identical shift value. So one of the signals is slightly shifted in the Hadamard spectrum (as this is the case for resonance of G10 in Figure , that gets excited by the band selective pulse which should ideally just act on resonance G12).

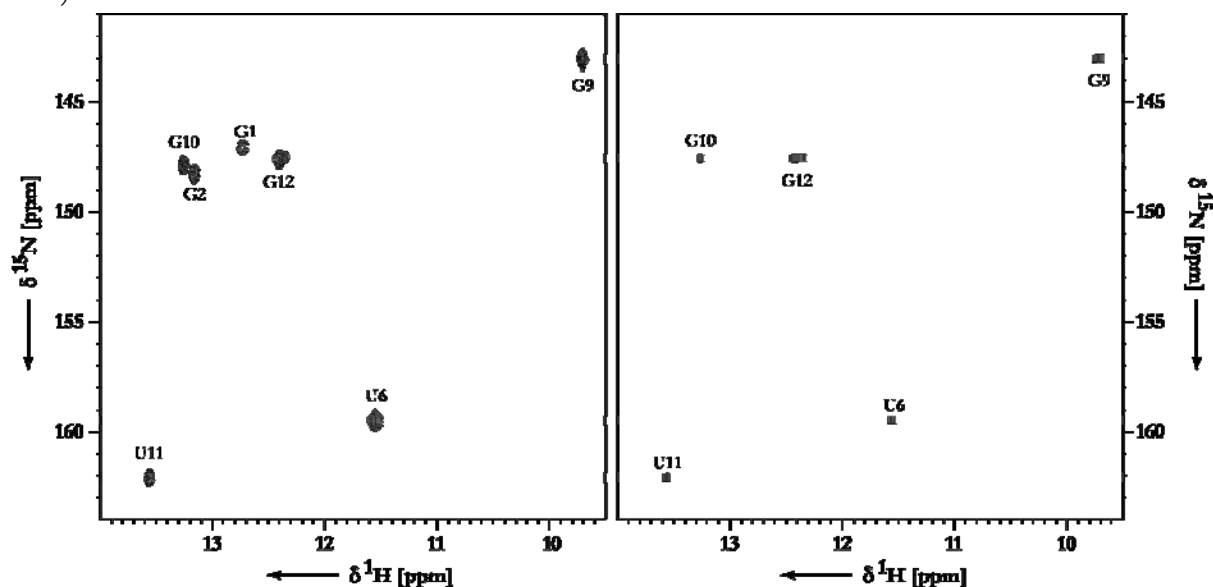


Figure 52: Comparison of conventional FT 2D  $^{15}\text{N}$ -HSQC (left) and Hadamard 2D  $^{15}\text{N}$ -HSQC spectrum (right). Both spectra were recorded at 293K on a Bruker AV600MHz NMR system. The FT spectrum was recorded within 2291s (2048pts  $\times$  256pts along F2 and F1, respectively; 8scans and 8 dummy scans, relaxation delay 1s); whereas the Hadamard spectrum was recorded in 72s (2048pts along F2; Hadamard matrix of  $n=8$ ; 6scans and 8 dummy scans, relaxation delay 1s). All proton pulses were applied with a field strength of 24.39kHz on resonance with the HDO signal, water suppression was achieved with a hard WATERGATE sequence. The selective  $^{15}\text{N}$  pulses have a field strength of 100Hz and a Gauss.1000 shape (the bandwidth of the selective pulses is indicated by the grey lines crossing both spectra).

<sup>186</sup>C.J. Bauer, R. Freeman, T. Frenkiel, J. Keeler, A.J. Shaka, J. Magn. Reson., 1984, **58**, 422-457

As depicted for the case of the 14mer RNA, a subset of the FT HSQC spectrum containing selected resonances can be recorded with the Hadamard methodology in a time fraction of 3/100 without compromising in S/N.

In order to extract kinetic information a series of 2D Hadamard type of experiments is recorded during the conformational switching reaction of a singly  $^{15}\text{N}$ -labelled bistable 20mer RNA. The decay of signal is monitored as the system relaxes towards equilibrium. The kinetic trace (Figure 53) of this process is extracted by integration of the peak in every recorded spectrum. The rate constant extracted is equivalent as recorded by the earlier described 1D method and therefore the technique is proven to be utilizable for the investigation of RNA folding processes.

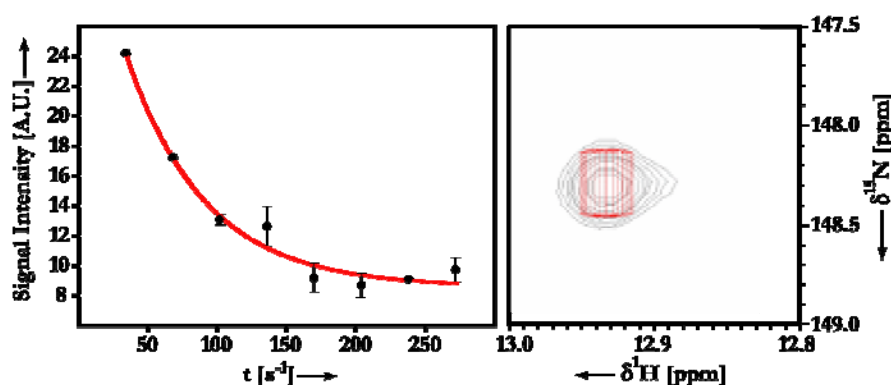


Figure 53: Kinetic trace (left panel) extracted of an array of 2D Hadamard  $^{15}\text{N}$ -HSQC spectra recorded during the conformational switching reaction of a single G10  $^{15}\text{N}$ -labelled bistable 20mer (as introduced earlier) induced by laser irradiation of a photo-protected RNA sequence. Overlay of a regular FT- $^{15}\text{N}$ -HSQC (grey) and Had- $^{15}\text{N}$ -HSQC spectrum displaying the peak of G10 (H1/N1) in the first experiment recorded

### SOFAST

In 2005<sup>187</sup> the idea of reducing the total amount of experimental time by shortening the recycle delay<sup>188</sup> was resuscitated and could be used to record a series of 2D  $^1\text{H}$ - $^{15}\text{N}$ -HMQC experiments that report on hydrogen-exchange processes, now with a temporal resolution of seconds compared to the former resolution of minutes.

The idea behind the techniques that is called SOFAST-NMR (for band-selective optimized flip-angle short transient NMR) is to take advantage of a shortened  $T_1$ -relaxation time (and therefore a short recycle delay in the ms-range) by applying a HMQC-pulse-sequence that uses solely band-selective excitation for protons and has an optimized flip-angle according to the principles of Ernst-Angle excitation.

The usage of band selective pulses is required on the first hand to facilitate water suppression by not exciting the water-resonance and on the second hand in order to shorten the longitudinal relaxation time by reducing the amount of total excited protons. This is in accordance to the

<sup>187</sup>P. Schanda, B. Brutscher, *J. Am. Chem. Soc.*, 2005, **127**, 8014-8015

<sup>188</sup>A. Ross, M. Salzmann, H. Senn, *J. Biomol NMR*, 1997, **10**, 389-396

finding that the selective manipulation of proton spins in a macromolecule, while leaving all other protons unperturbed, yields significantly enhanced R1 relaxation rates.<sup>189</sup>

The HMQC pulse-sequence is therefore advantageous because it has a limited number of proton-pulses compared to the HSQC-experiment that contains the same information content. It is of advantage that, besides the shape of the band selective pulses (in the case of the initial publication a PC9 shape<sup>190</sup>), no additional technical feature is needed and so this technique can be applied on any spectrometer and the data can be processed by regular FT-techniques.

### Ultrafast

Frydman and co-workers recently introduced the technique of ultrafast-NMR<sup>191</sup>, where they could demonstrate the possibility of recording of a multidimensional spectrum in a single.

The basic idea of the concept is to replace the incremental evolution of the indirect time-domain, as used in standard NMR experiments that follow the PEMD<sup>192</sup> scheme, by a spatial encoding procedure. Instead that the whole sample as an entire ensemble of spins evolves chemical shift along the indirect dimension in parallel, the sample is split up into independent sub-ensembles that each for its own evolves chemical shifts in a sub-ensemble dependent manner. After an independent mixing procedure (so far TOCSY, COSY, HSQC among others are successfully applied<sup>193</sup>) an again sub-ensemble specific acquisition is conducted, now in the physical, second time domain.

The partitioning of the sample into sub-ensembles is carried out by a spatial encoding scheme that uses gradients - the spins are discriminated by their position (spatial coordinate  $r$ ) according to the additional applied field-gradient  $G = \partial B_0 / \partial r$ . In other words the indirect time is now expressed as a function of the spatial position of the spin  $-t_1(r)-$ . The resulting signal is then no longer a function of two different times  $-S(t_1, t_2)-$ , but dependent on a variable  $k$  and the acquisition time  $t_2$   $-S(k, t_2)-$ . The variable  $k$  is thereby encoding the spatial distribution of the respective spin along the  $z$ -axis of the gradient  $-k = \int G(t) dt-$ .<sup>194</sup> Subsequent Fourier Transformation leads to the final 2D spectrum.

The capability of the technique to monitor dynamic processes in bio-macromolecules is already demonstrated for an ubiquitin sample where H/D exchange processes were monitored with a resolution of 4s per 2D <sup>15</sup>N-HSQC spectrum.<sup>195</sup>

An inherent drawback of this technique is that in contrast to the conventional 2D methods, the signal of a single scan is spread over the whole two dimensional space, so that the resulting S/N ratio in the spectrum is very poor. The ubiquitin sample used in the above mentioned study had a concentration of 3.2mM and is of low molecular weight. This high concentration might be suitable for this case but is normally above the preparation yields and precipitation borders for more sophisticated and larger bio-macromolecules. Moreover, the enhanced relaxation of larger

---

<sup>189</sup>K. Pervushin, B. Vögeli, A. Eletsky, *J. Am. Chem. Soc.*, 2002, **124**, 12898-12902

<sup>190</sup>E. Kupce, R. Freeman, *J. Magn. Reson.*, 1993, **102A**, 122-126

<sup>191</sup>L. Frydman, T. Scherf, A. Lupulescu, *Proc. Natl. Acad. Soc.*, 2002, **99**, 15858-15862

<sup>192</sup>J. Jeener, *International Ampere Summer School*, 1971, Basko Polje, Yugoslavia

<sup>193</sup>L. Frydman, A. Lupulescu, T. Scherf, *J. Am. Chem. Soc.*, 2003, **125**, 9204-9217

<sup>194</sup>L. Frydman, *C. R. Chimie*, 2006, **9**, 336-345

<sup>195</sup>M. Gal, M. Mishkovsky, L. Frydman, *J. Am. Chem. Soc.*, 2006, **128**, 951-956



macromolecules may also hamper the application of this technique, regarding the decreased signal intensity.

Very recently a combination of the ultrafast- and SOFAST-NMR approaches was published as ultraSOFAST-NMR.<sup>196</sup> The authors argue that the complementary techniques -the first accelerates acquisition by reducing the number of scans and the second reduces the delay between consecutive scans- enable in their combination the acquisition of 2D spectra at Hz rates. The example given is again the unpretentious H/D exchange of ubiquitin.

#### FUTURE DIRECTIONS

Up to now the only observable in RT-experiments is the behaviour of certain resonances arising from certain atoms in the molecule. Or expressed in other words, the only parameter observed by RT-NMR methods is the chemical shift.

Whereas in the case of monitoring imino-resonances in RNA molecules the chemical shift information represents a defined long-range restraint, in other spectral regions such an analysis fails. It can just be complemented by a detailed characterization of the product and educt state. This is different for proteins and peptides, because here the chemical shifts are highly correlated to secondary structure elements.<sup>197</sup>

Possible additional parameters that could be measured are of course scalar couplings but the disadvantage is that those need to be measured with high spectral resolution (need of a well resolved 2D spectrum) to determine conformational changes. This is due to the size of the most relevant  $^3J_{XY}$  scalar couplings being below 10Hz. Therefore the resolution has to be at 0.5Hz, which is at the edge of resolution obtained in RT-experiments. The second disadvantage is that this parameter only reveals information on local conformation although it is obligatory to measure a complete set of coupling constants to describe the global conformational changes of the RNA molecule.

Hence possible parameters to describe the local as well as the global changes of the molecular conformation during the reaction course monitored by RT-NMR are RDCs. The advantage is hereby that the resolution of the recorded spectra is sufficient to be  $\leq 3$ Hz in order to get insightful parameters. The biggest disadvantage is of course that the RT-experiment has to be conducted twice -once under isotropic conditions and once under anisotropic conditions in an alignment medium such as phage or bicelles. An interaction between all conformational states populated transiently during the reaction course and the alignment medium have to be excluded, because this would otherwise lead to a distortion of the reaction path and a kinetic analysis would fail.

---

<sup>196</sup>M. Gal, P. Schanda, B. Brutscher, L. Frydman, *J. Am. Chem. Soc.*, 2007, **129**, 1372-1377

<sup>197</sup>D.S. Wishart, B.D. Sykes, F.M. Richards, *J. Mol. Biol.*, 1991, **222**, 311-333; example given in: C. Schlörb, K. Ackermann, C. Richter, J. Wirmer, H. Schwalbe, *J. Biomol. NMR*, 2005, **33**, 95-104

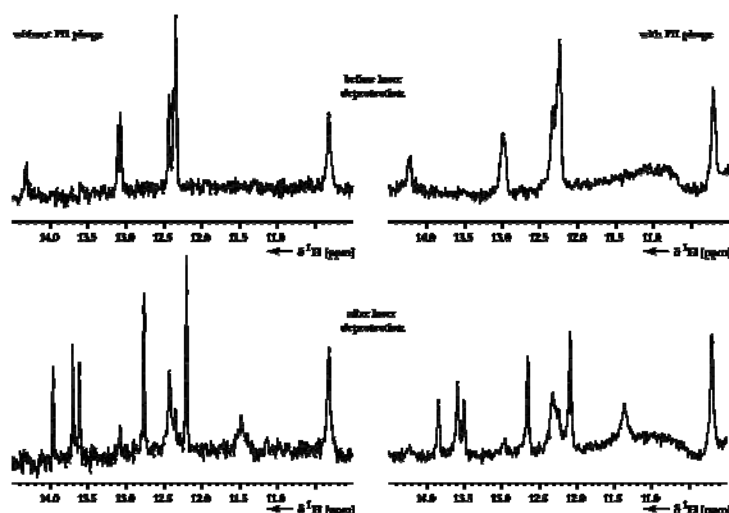


Figure 54: Imino proton region of 1D  $^1\text{H}$  spectra recorded on photo-caged RNA system (upper panel) that relaxes into its bistable equilibrium form after laser induced deprotection (lower panel). The reaction and the reaction path are equal in solution without (left panel) and with Pfl phage (right panel) as cosolvent.

Orienting experiments conducted on solutions containing phages as alignment medium show the following results (Figure 54):

- A) Kinetic refolding experiments can be undertaken in alignment media such as phages.
- B) The reaction displays the same kinetics as without alignment medium.
- C) As a useful side effect the deprotection yield of the photo-removable protection group is higher.

The route for extracting RDCs in RT-experiments of RNA folding reactions is now open.

A REALLY WELL-MADED BUTTONHOLE IS THE  
ONLY LINK BETWEEN ART AND NATURE.<sup>198</sup>

## NMR PARAMETERS AND STRUCTURE CALCULATION

A wide range of NMR parameters can be measured and translated into structural information. The following section will give a brief introduction into the origins of the respective NMR parameters and their utilization in the structural refinement of RNA molecules.

### NOE AND DISTANCES

In a rigid two-spin system the NOE (nuclear Overhauser effect) represents the cross-relaxation rate between these two spins. It is directly proportional to the inverse sixth power of distance and to the overall tumbling of the molecule:

$$\sigma_{ij} \propto r_{ij}^{-6} f(\tau_c)$$

Generally, due to this behaviour the NOE effect is only observed between nuclei that are less than 5 Å apart (Figure 55). But nevertheless being a short-range effect, the NOE can be directly translated into a distance restraint. To obtain an exact measure of the NOE and therefore of the interatomic distance a NOESY-experiment (Figure 56) can be measured at several mixing times and the build-up curve can be analysed (here the best method of analysis would be using a relaxation matrix that is back-calculated from a structural model, that is obtained during the structural refinement). In a different approach a good approximation of the distance is made if the volume/intensity of the signal originating from the NOE is calibrated on a known distance.

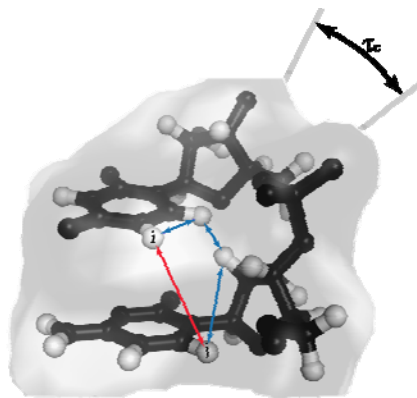


Figure 55: Clipping of a RNA structure (ball and stick model; surface of the molecule is shaded; hetero-atoms are coloured in black, protons in white) that should illustrate the pathway for spin diffusion (blue) and the direct NOE effect between the two nuclei *i* and *j* (red). The distance between the indicated protons is always less than 5 Å. The surrounding circle should indicate that the size of the NOE depends on the molecular tumbling behaviour that is described by the time  $\tau_c$ .  $\tau_c$  is defined as the time the molecule needs to tumble one degree solid angle.

<sup>198</sup>O. Wilde, *The Chameleon*, 1894, **1**, “Phrases and philosophies for the use of the young”

The distance restraint ( $r_{ij}$ ) then used in structure calculation is derived from the signal intensity of the crosspeak ( $I^{NOE}$ ) between the two protons ( $i, j$ ) and is given by

$$r_{ij} \approx \left( c_{cal} I_{ij}^{NOE} \right)^{-\frac{1}{6}}.$$

The calibration constant is defined from the above mentioned reference distance that should be a covalently fixed and not dynamically modulated distance e.g.  $r_{H5,H6}$  in pyrimidine nucleobases:

$$c_{cal} = \frac{r_{ref}^{-6}}{I_{ref}^{NOE}}$$

A major source of errors in NOE distances is the effect of spin diffusion, when magnetization between the two spins is transferred via indirect paths between them. Unfortunately, depending on the molecular geometry these indirect paths may be more efficient (if the indirect path includes a small distance between the atoms due to the  $r^{-6}$ -dependence of the NOE effect) than the direct one. Spin diffusion thereby leads to an apparent increase in the signal intensity of the NOE cross-peak and the distance between the nuclei is underestimated.

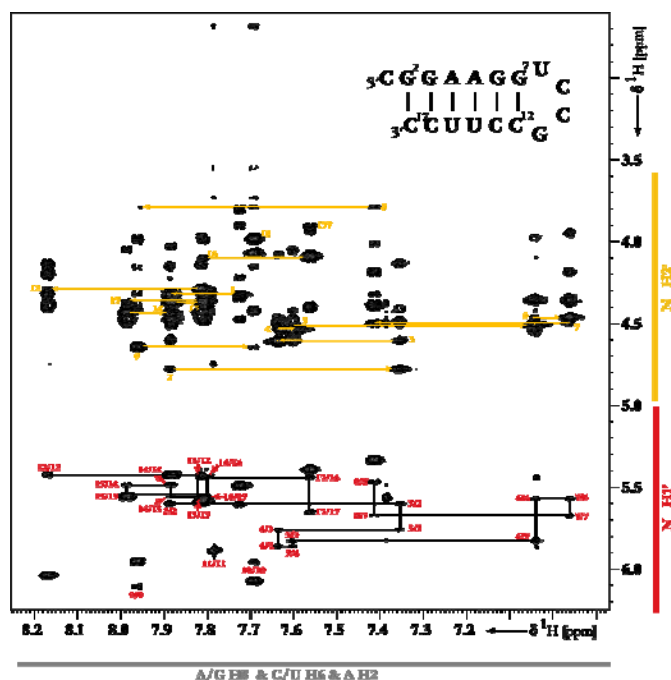


Figure 56: 2D  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum of a RNA hairpin structure recorded at 298K and 600MHz on a  $\text{D}_2\text{O}$  sample. The connectivity walk that provides first structural information for  $\text{H}^{\text{ar}}_i/\text{H1}'_{i-1} \gg \text{H}^{\text{ar}}_i/\text{H1}'_i \gg \text{H}^{\text{ar}}_{i+1}/\text{H1}'_i$  is shown by the black lines, for  $\text{H}^{\text{ar}}_i/\text{H2}'_{i-1} \gg \text{H}^{\text{ar}}_i/\text{H2}'_i \gg \text{H}^{\text{ar}}_{i+1}/\text{H2}'_i$  it is indicated by the yellow lines. Obviously a lot of further distance information can be extracted additional to this sequential connectivity information.

The second influence on the NOE analysis that may affect the structural interpretation of the effect is that all NOE data is modulated by internal dynamics.

Assuming that the molecule is not totally rigid the cross-relaxation rate is also a function of the order parameter and reflects therefore not only the distance of the protons but also their internal mobility. It is notable that non-linear averaging over distance increases the measured rate, whereas angular fluctuations lead to a reduction of the rate.

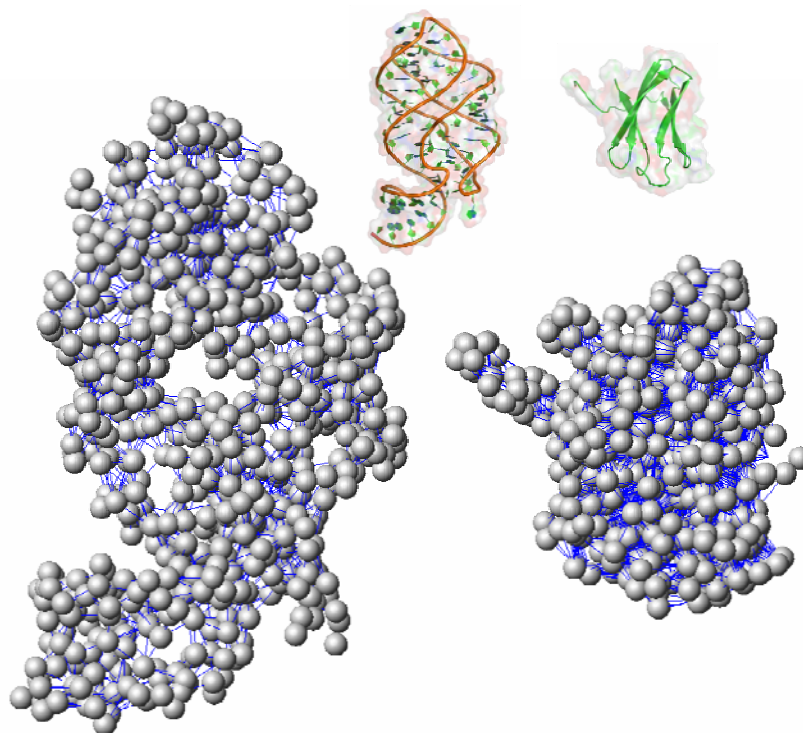


Figure 57: Illustration of the proton density in (left side) RNA (Riboswitch RNA; X-ray structure; pdb-code: 1U8D hydrogens added and minimized) and (right side) protein molecules (Tendamistat; NMR structure; pdb-code: 3AIT) – both molecules have an equal number of respective monomeric building blocks: RNA consists of 66 nucleotides; protein consists of 74aa. The upper panel shows cartoon and surface representation of the molecules to underline the structural elements involved in the respective structure architecture and the overall size of the biomacromolecules. In the main panel just the hydrogen atoms (protons) of each molecule are represented as spheres of their van der Waals-radii; the blue lines indicate all proton-proton distances that are less than 5 Å and would lead to a measurable NOE. Obviously the density of protons per volume unit of the structure is higher in the protein case and therefore the number of measurable distances via NOESY spectroscopy is higher.

An additional problem arises when using NOEs to elucidate RNA structures compared to protein structures: the proton density and therefore the amount of theoretically measurable interatomic distances is low (Figure 57).

Notwithstanding all difficulties mentioned above, distances derived from NOE measurements are fundamental to all structural calculations. The basis of the structures determined is mostly a network of distances that is determined to completeness.

### SCALAR COUPLINGS AND DIHEDRAL ANGLES

Hetero- and homonuclear coupling constants provide inside into the RNA backbone conformation and its conformational dynamics. The measurement of scalar couplings is especially important for RNA because of the low proton density (as explained above) and the large number of free torsion angles (see chapter 2). Additionally, in non-canonical regions of RNA such as bulges and loops (see also chapter 2) conformational heterogeneity is often observed, therefore the use of NOE distance restraints is often insufficient to determine such conformational dynamics.

Especially the vicinal  $^3J_{\text{XY}}$  scalar couplings are used to determine torsion angles, because they directly report on the underlying torsion. But also  $^2J_{\text{XY}}$  and  $^1J_{\text{XY}}$  couplings depend on the local conformation and can be translated into structural restraints.

The dependence of a certain coupling on a certain torsion angle is mainly derived from a parameterisation of the respective J-coupling on locked nucleotides with known conformation. These ‘known’ conformations are derived from independent NMR measurements or mostly from X-ray crystallographic data.<sup>199</sup> The parameterisation leads generally to a Karplus-equation (Figure 58). As an inherent feature the curves display an ambiguity for certain J-coupling values. Therefore the measurement of more than a single coupling constant is necessary to unambiguously determine the torsion angle.

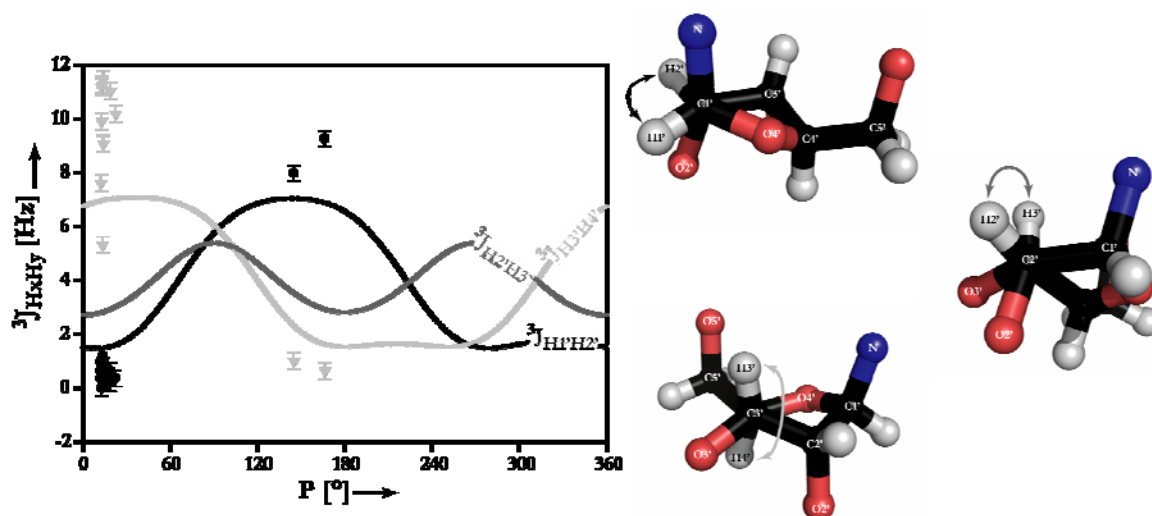


Figure 58: Karplus curve for three  $^3J_{\text{HxHy}}$  couplings (parametrization is following Haasnoot et al. 1981) describing the dependence on the conformation of the ribose-ring in RNA (the conformation is described as the pseudo-rotation phase  $P/ [^\circ]$ <sup>200</sup> that is defined as presented in chapter one. The angles described by the displayed scalar couplings are indicated on the right side of the figure – for a ribose unit in stick-ball-representation populating the C3'-endo conformational space (which is the canonical conformation for A-RNA; for B-DNA the C2'-endo conformation is the canonical).

For the determination of the sugar ring conformation  $^3J_{\text{HH}}$  couplings can be measured.<sup>201</sup> Also  $^3J_{\text{HC}}$  couplings can be used as reporters on the conformation of the five membered furanose ring.<sup>202</sup> Information about the phosphodiester backbone conformation can be obtained by  $^3J_{\text{HP}}$  and  $^3J_{\text{CP}}$ .<sup>203</sup> The value of the torsion angle  $\chi$  that describes the conformation of the glycosidic bond can be measured by  $^3J_{\text{H1}'\text{C8/6}}$  coupling constants.<sup>204</sup>

<sup>199</sup>summarized in: J.H. Ippel, S.S. Wijmenga, R. de Jong, H.A. Heus, C.W. Hilbers, E. de Vrom, G.A. van der Marel, J.H. van Boom, *Magn. Reson. Chem.*, 1996, **34**, 156-176

<sup>200</sup>C. Altona, M. Sundaralingam, *J. Am Chem. Soc.*, 1972, **94**, 8205-8212

<sup>201</sup>J. van Wijk, B.D. Huckriede, H.J. Ippel, C. Altona, *Methods Enzymol.*, **211**, 286-306

<sup>202</sup>C.A.G. Haasnoot, F.A.A.M. de Leeuw, C. Altona, *Tetrahedron*, 1980, **36**, 2783-2792; C.A.G. Haasnoot, F.A.A.M. de Leeuw, H.P.M. de Leeuw, C. Altona, *Biopolymers*, 1981, **20**, 1211-1245; E. Duchardt, C. Richter, B. Reif, S.J. Glaser, J.W. Engels, C. Griesinger, H. Schwalbe, *J. Biomol. NMR*, 2001, **21**, 117-126

<sup>203</sup>H. Schwalbe, J.P. Marino, G.C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR*, 1994, **4**, 631-644

<sup>204</sup>R.U. Lamieux, L. Nagabushan, B. Paul, *Can. J. Chem.*, 1978, **50**, 773-776

## SCALAR COUPLINGS AND HYDROGEN BONDS

As lined out in the previous chapters hydrogen bonding plays the key role in folding and stabilisation of the structure of a nucleic acid. Up to the late nineties only indirect observables reported on the presence of a hydrogen bond between nucleotides. As nowadays still used as indicators, the reduced hydrogen exchange rate with the solvent<sup>205</sup> and the isotropic proton<sup>206</sup> or nitrogen<sup>207</sup> chemical shift were the primary source of information about the hydrogen bonding pattern in RNA molecules. However in 1998 Dingley and Grzesiek<sup>208</sup> showed that the electron density between hydrogen bonded nitrogen atoms in Watson Crick base-pairs is high enough to give rise to  $^2J_{\text{NN}}$  couplings in the range up to 7Hz.

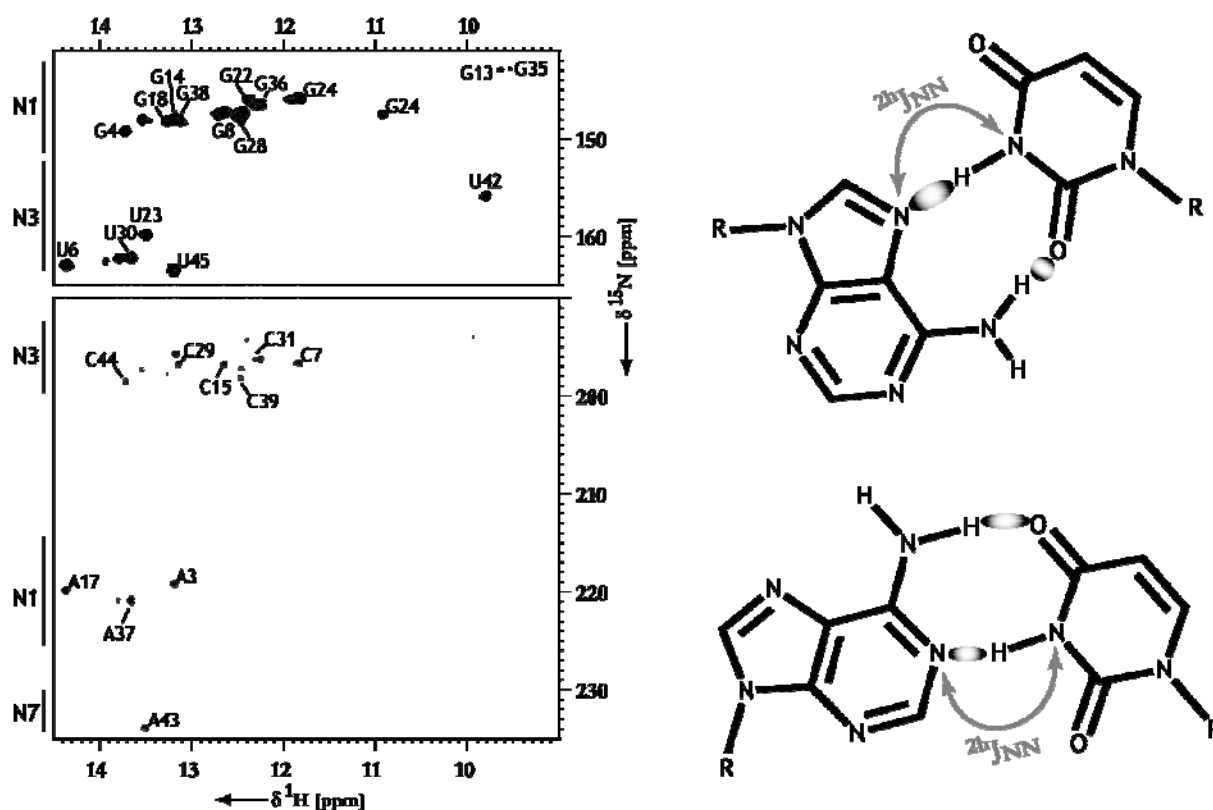


Figure 59: Utilisation of the scalar  $^2J_{\text{NN}}$  couplings to determine the base-pairing pattern in a RNA molecule (Diels-Alderase ribozyme in 4mM  $\text{Ca}^{2+}$  and in the presence of an inhibitor). Left panel shows the diagonal and cross-peaks of an HN(CO)N-COSY experiment recorded at 800MHz and 298K. The upper part of the spectrum displays the peaks that arise from imino-groups, the peaks for guanine and uracil are well separated due to the different chemical shift of N1 and N3, respectively. Also the cross-peaks are well dispersed again due to the chemical shift differences for the nitrogen atoms in different nucleobases. Not only the canonical WC-base-pairs (A-U Watson-Crick-base-pair illustrated in the lower part of the right panel) can be identified but again due to the diverse chemical shifts also non-canonical base-pairs as for example reverse Hoogsteen-base-pairs (as in the example between A43 and U23 (geometry of this base-pair is displayed in the upper half of the right panel; the ellipsoids represent the electron density that procures the scalar couplings).

<sup>205</sup>G. Wagner, *Q. Rev. Biophys.*, 1998, **16**, 1-57

<sup>206</sup>R.R. Shoup, H.T. Miles, E.D. Becker, *Biochem. Biophys. Res. Commun.*, 1966, **23**, 194-201

<sup>207</sup>V. Markowski, G.R. Sullivan, J.D. Roberts, *J. Am. Chem. Soc.*, 1977, **99**, 714-718

<sup>208</sup>A.J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.*, 1998, **120**, 8293-8297

However, as has been shown in several following works also for non-canonical base-pairs the HNN-COSY experiment is applicable, and leads to information about the base pairing pattern.<sup>209</sup>

Taking into account the low proton density in RNA that hampers structure determination based solely on NOEs and the structural importance of base pairing in nucleic acids, the information gained by the HNN-COSY experiments cannot be overestimated and must be considered as a long range constraint.

### RDCs AND BOND ORIENTATIONS

Besides scalar couplings across hydrogen bonds and in contrast to NOEs, scalar couplings for the determination of torsion angles and cross-correlated relaxation rates, residual dipolar couplings (RDC) represent the second NMR parameter that is capable of the describing long range interactions. As depicted earlier in this thesis (chapter 4) RDCs are additionally also a probe of dynamics.

RDCs originate from an anisotropic spin interaction that depends on the orientation of the molecule to the external magnetic field and that can further be described as a second rank tensor in the laboratory frame. The source of dipolar couplings is the affection of the local field at a given nucleus by the magnetic flux originating from a second neighbouring nucleus. The magnitude of this field perturbation is a function of inter-nuclear distance and orientation of the inter-nuclear vector to the magnetic field ( $B_0$ ). The angular and distance dependence is described by

$$D \propto \langle 3\cos^2 \theta - 1 \rangle \cdot \frac{1}{r_{ij}^3}$$

where the angular dependence represents a time average. Assuming perfectly isotropic sampling of the oriental space the interaction becomes averaged out. This happens because among the anisotropic interactions the dipolar interaction (and the quadrupolar interaction) possesses a traceless Hamiltonian, so that dynamic averaging conditions alter the effective interaction to be zero. Obviously this becomes evident by looking at the angular term that is reduced to zero in the isotropic case:

$$\langle 3\cos^2 \theta - 1 \rangle = \int_0^\pi \frac{\sin \theta}{2} \cdot (3\cos^2 \theta - 1) d\theta = 0$$

Here  $\frac{\sin \theta}{2}$  is the distribution function, which itself is the ratio of the probability of finding a vector along  $\theta$  ( $2\pi \sin \theta$ ) and the total solid angles in a unit sphere ( $4\pi$ ).

A divergence from the isotropic tumbling renders the anisotropic interaction to be effective again and residual dipolar couplings can be measured. This can be achieved by the addition of a co-solvent that hinders sterically the isotropic reorientation of the molecule. The degree of

---

<sup>209</sup>J. Wöhnert, A.J. Dingley, M. Stoldt, M. Görlach, S. Grzesiek, L.R. Brown, *Nucl. Acids Res.*, 1999, **27**, 3104-3110



alignment -respectively the size of the residual dipolar coupling- should not be smaller than  $10^{-5}$  because otherwise the couplings would be not measurable at needed precision, or larger than  $10^{-2}$  because this would cause spectral crowding and lead to severe problems with overlapping signals. A degree of alignment in the order of  $10^{-3}$  would be ideal (and since this is by far not the full dipolar coupling the name residual dipolar coupling is now self-explanatory).

Up to now several chemically very diverse co-solvents are known that get ordered by the external magnetic field and are able to propagate some of their order to the molecule. The mechanism of order-transmission from the alignment medium to the bio-macromolecular solute is believed to involve a combination of steric obstruction and charge-charge interactions.<sup>210</sup> The first alignment medium introduced by Tjandra et al.<sup>211</sup> were liquid crystalline disc shaped phospholipids (called “bicelles” and formerly used as mimics of membrane bilayers<sup>212</sup>), which undergo a phase transition from an isotropic phase to an anisotropic phase at  $\sim 303\text{K}$ . Therefore, a set of anisotropic (scalar couplings plus residual dipolar couplings) and isotropic (scalar couplings) can be measured at the same sample at different temperature conditions. However, unfortunately the medium is not very stable and precipitates after short periods<sup>213</sup>. The widely used alignment medium for nucleic acids is represented by the filamentous bacteriophage Pf1<sup>214</sup> that has a high tolerance to the ionic strength of the solute and is very robust over a wide range of temperature. Pf1 is a rod like, parallel to the external magnetic field oriented object that is substantially negatively charged and therefore induces alignment of the likewise negatively charged nucleic acids through a mechanism that is familiar to a steric obstruction though electrostatic repulsion can modify this behaviour.

An alignment medium that is aligned perpendicular to the external magnetic field, is a lamellar phase formed by alkyl-polyethylene glycol and n-hexanol. It is robust to a wide range of different ionic strength conditions, temperatures and pH<sup>215</sup>.

A new method to align molecules is the usage of stretched and compressed gels<sup>216</sup> as a matrix in which the solution is placed into the magnetic field- this is also possible with RNA. The most recent advance is to incorporate phages into gel matrices and to obtain alignment of the molecules, where the direction of the medium can be manipulated independently of the static magnetic field<sup>217</sup>.

The RDCs are measured as additional splittings in the anisotropic medium (J+D) and can be extracted by subtracting the assessed value in the isotropic case (J). Because RDCs are normally

---

<sup>210</sup>H. al-Hashimi, lecture notes, 2004, University Frankfurt, Germany

<sup>211</sup>N. Tjandra, A. Bax, *Science*, 1997, **10**, 289-292

<sup>212</sup>C.R. Sanders 2<sup>nd</sup>, J.P. Schwonek, *Biophys. J.*, 1993, **65**, 1460-1469

<sup>213</sup>M. Ottiger, A. Bax, *J. Biomol. NMR*, 1998, **12**, 361-372

<sup>214</sup>M.R. Hansen, L. Mueller, A. Pardi, *Nat. Struct. Biol.*, 1998, **5**, 1065-1074; G.M. Clore, M.R. Starich, A.M. Gronenborn, *J. Am. Chem. Soc.*, 1998, **120**, 10571-10572; M.R. Hansen, P. Hanson, A. Pardi, *Methods. Enzymol.*, 2000, **317**, 220-240

<sup>215</sup>M. Rückert, G. Otting, *J. Am. Chem. Soc.*, 2000, **122**, 7793-7797

<sup>216</sup>H.J. Sass, G. Musco, S.J. Stahl, P.T. Wingfield, S. Grzesiek, *J. Biomol. NMR*, 2000, **18**, 303-309; R. Tycko, F.J. Blanco, Y.J. Ishii, *J. Am. Chem. Soc.*, 2000, **122**, 9340-9341; Y.J. Ishii, M.A. Markus, R. Tycko, *J. Biomol. NMR*, 2001, **21**, 141-151

<sup>217</sup>J.F. Trimpe, F.G. Morin, Z. Xia, R.H. Marchessault, K. Gehring, *J. Biomol. NMR*, 2002, **22**, 83-87; J.R. Tolman, lecture at the XXII<sup>nd</sup> ICMRBS, 2006, Göttingen, Germany

measured on covalently linked, mostly directly bonded nuclei their distance is ( $r_{ij}$ ) is defined. Therefore, the dipolar couplings describe the orientation of a bond-vector with respect to the alignment tensor.

### RESTRAINED MD AND SIMULATED ANNEALING

Structure calculation is defined as a minimization problem for a target function that measures the agreement between a conformation of a macromolecule and the set of restraints measured by NMR. Two major techniques for solving this minimization problem are developed represented by (a) metric matrix distance geometry (DG) and (b) Cartesian or torsion angle restrained molecular dynamics simulation (rMD). The first technique is realized in programs as e.g. DIG-II<sup>218</sup> and DIANA<sup>219</sup>, rMD in programs like AMBER<sup>220</sup>, CHARM<sup>221</sup>, XPLOR<sup>222</sup> and CNX/CNS<sup>223</sup>. In both cases the aim is to generate an ensemble of molecular structures that fits with the experimentally derived restraints.

In the case of rMD the structures are calculated using the NMR restraints and an energy minimization with a potential energy function -also called target function- that is composed of classical terms ( $V_{\text{classic}}$ : describing the energy of the molecule itself) and of terms describing the NMR parameters measured ( $V_{\text{NMR}}$ ). Compared to the DG-methods the rMD is less sensitive to the conformation of the initial structure.<sup>224</sup>

$$V_{\text{tot}} = V_{\text{classic}} + V_{\text{NMR}}$$

with

$$V_{\text{classic}} = \omega_{\text{bond}} V_{\text{bond}} + \omega_{\text{angle}} V_{\text{angle}} + \omega_{\text{dihedral}} V_{\text{dihedral}} + \omega_{\text{improper}} V_{\text{improper}} + \omega_{\text{van der waals}} V_{\text{van der waals}} + \omega_{\text{electrostatics}} V_{\text{electrostatic}}$$

and

$$V_{\text{NMR}} = \omega_{\text{NOE}} V_{\text{NOE}} + \omega_{\text{J-coupling}} V_{\text{J-coupling}} + \omega_{\text{H-bond}} V_{\text{H-bond}} + \omega_{\text{RDC}} V_{\text{RDC}}$$

The terms  $w$  hereby describe the force constants that scale the each energy contribution independently. In the case of the classical terms they are non-modifiable parts of the respectively used force-field, in case of the NMR-terms they are heuristically chosen.

The potential used for the NMR terms are introduced here exemplary for the NOE:

Among several different functions used to characterize the potential energy contribution of the NOE-effect, a flat-well (or flat-bottom) potential is frequently used, which consists of the energy

<sup>218</sup>T.F. Havel, *Prog. Biophys. Mol. Biol.*, 1991, **56**, 43-78

<sup>219</sup>P. Güntert, K. Wüthrich, *J. Biomol. NMR*, 1991, **1**, 447-456

<sup>220</sup>D.A. Pearlman, D.A. Case, J.C. Caldwell, G.L. Seibel, U.C. Singh, P. Weiner et. al., *Amber v. 4.0*, UCSF, 1991, S.F.

<sup>221</sup>B.R. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.*, 1983, **4**, 187-217

<sup>222</sup>A.T. Brünger, *X-PLOR v3.1*, Yale University Press, 1992, New Haven

<sup>223</sup>A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve et. al., *Acta. Cryst.*, 1998, **D54**, 905-921

<sup>224</sup>M.P. Allen, D.J. Tildesley, *Computer Simulation of Liquids*, Clarendon, 1987, Oxford

contributions of NOE distance violations relative to the lower (L) and upper (U) distance bounds (Figure 60). The parameter  $r$  describes the inter-proton distance for each proton pair from the generated structures for which a NOE is measured:

$$E_{NOE} = \begin{cases} (r-U)^2 & \text{if } r > U \\ (r-L)^2 & \text{if } r < L \\ 0 & \text{if } L < r < U \end{cases}$$

If instead of a range of distances from a single experiment, a precise inter-proton distance is measured from NOESY build-up curves out of multiple experiments with different mixing times, a biharmonic potential (Figure 60) can be used in order to describe the NOE-potential function:

$$E_{NOE} = \begin{cases} (r-r_0)^2 & \text{if } r > r_0 \\ (r-L)^2 & \text{if } r < r_0 \end{cases}$$

To reduce possible artefacts stemming from spin diffusion or conformational averaging, so called soft potentials (Figure 60) can be used instead that restrain the upper bonds more gently:

$$E_{NOE} = \begin{cases} (r-L)^2 & \text{if } r < L \\ 0 & \text{if } L < r < U \\ (U-r)^2 & \text{if } U < r < S \\ A(r-U)^{-1} + B(r-U) + C & \text{if } r > S \end{cases}$$

The potential functions for the different NMR parameters are given in a similar manner.

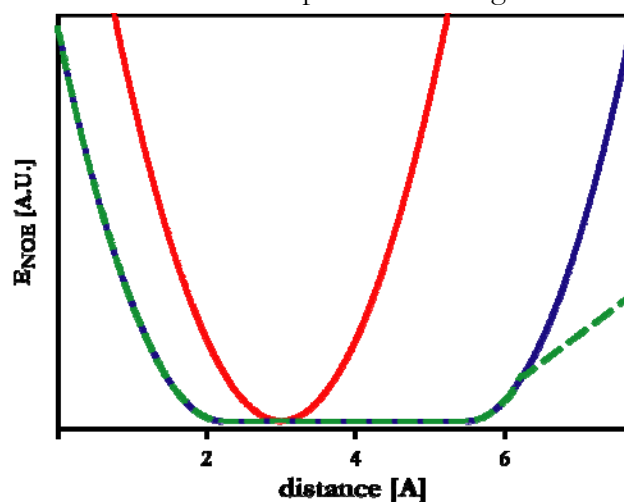


Figure 60: Graphical representation of the three main classes of potentials used to restrain NOE-distances in structure calculations. In blue the flat bottom potential, the biharmonic potential shown in red and the soft potential in dashed green.

The MD simulation is conducted by solving the Newton equation of motion using the forces generated by varying the potential energies of the structure. The minimum energy structure is

hereby obtained by solving the first derivative of potential energy with respect to the coordinates of each atom assuming that the derivative is zero. Passing the energy barriers is maintained by raising the temperature during simulation (Figure 61). The temperature is controlled and varied according to the minimization algorithm of simulated annealing.<sup>225</sup>

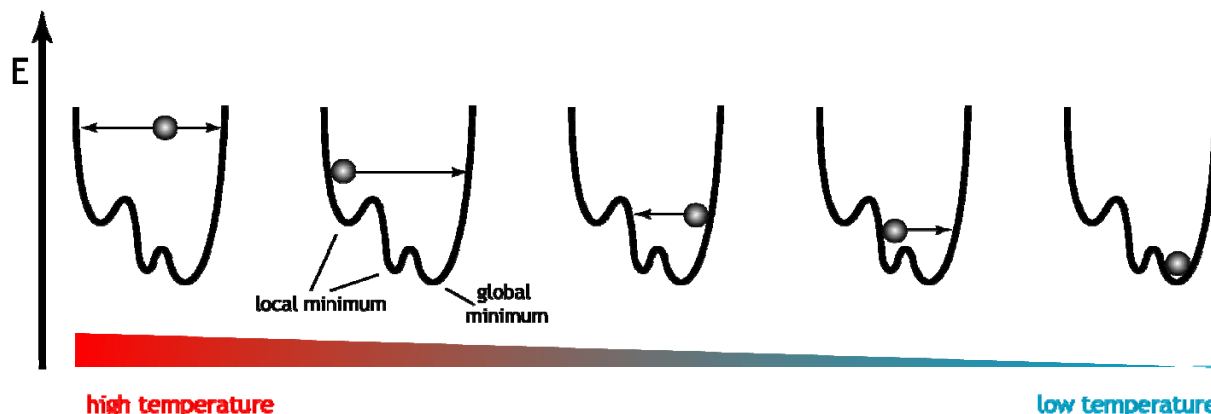


Figure 61: Representation of the simulated annealing protocol: the conformational space (represented by the black rough surface) is sampled by going from high to low temperature in order to find the global minimum. In order to avoid trapping in a local minima the SA protocol can include loops that repeat some higher temperature steps and preceding then to the lower temperatures.

In practice the protocol consists of the following simulation steps: TAD-MD (torsion angle dynamics-molecular dynamics) with

- first a high temperature phase starting at  $T_{\text{max}}=20000$ , with 4000 steps and time steps of 0.01ps
- second a first slow cool annealing stage starting at  $T_{\text{max}}=20000$  with 20000 steps of 0.01ps
- third a second slow cool annealing stage starting from  $T_{\text{max}}=2000$  with 3000 steps of 0.005ps
- and finally a subsequent minimization of 8 cycles each with 1000 minimization steps.

As starting structures, completely randomized RNA chains are used. Calculations were performed with CNX2005 from Accelrys Inc..

As an example the structure calculation of the 14mer cUUCGg tetraloop RNA is illustrated. Starting the simulated annealing protocol from a randomized RNA chain and with random variation of the initial conditions (=velocities), 100 structures are calculated and sorted afterwards by the total energy of the respective structure (Figure 62). Then the ten structures displaying the lowest energies are selected and combined to a bundle that represents the structure best according to the input NMR structural parameter.

<sup>225</sup>S. Kirkpatrick, C.D. Gelatt, M.P. Vecchi, *Science*, 1983, **220**, 671-680

Different indicators can critically evaluate the quality of a NMR structure:

- Convergence:** How many of the 100 calculated structures exhibit a total energy in the same order as the best structures? This does not mean that an energetically converged structure calculation must have a lower structural RMSD value compared to a calculation where a steep increase in energy is over the calculated structures is monitored.
- Precision:** I) How big is the RMSD value between the structures of the ensemble? This might be problematic to evaluate the quality of the structure since it depends on empirical parameters as the bounds of the input NMR parameters and the weights of the different energy terms. And if just a few restraints are added or deleted the result may change dramatically.  
II) How big is the RMSD value between the experimentally measured input parameters and the same parameters back calculated from the structure ensemble?
- Accuracy:** Describes the differences between the calculated and the “true”/gold standard structure, which might be considered to be the X-ray crystal structure.

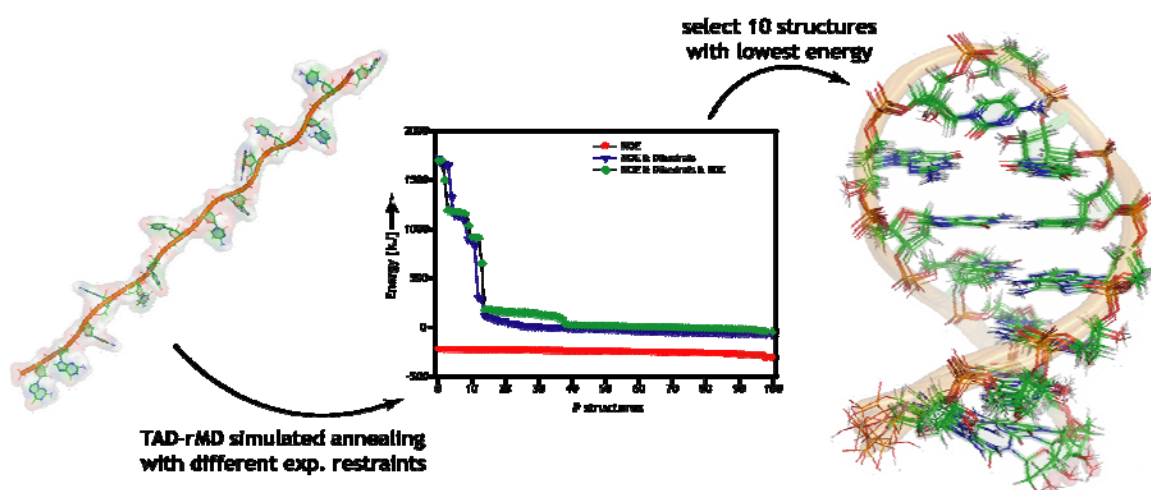


Figure 62: Schematic description of the structure calculation process: Starting from a random elongated RNA structure the TAD-rMD is conducted and yields 100 different structures that are subsequently sorted by their final energy after minimization. The 10 best structures are then assembled in a bundle of structural models representing the molecular structure best according to the input parameters. As seen in the energy diagram in any case convergence is good, but as lesser restraints are used as better is convergence and as lower is the total energy.



---

## NMR of RNA

**REVIEW ARTICLE:** NMR SPECTROSCOPY OF RNA  
Boris Fürtig, Christian Richter, Jens Wöhnert and Harald Schwalbe,  
*ChemBioChem*, 2003, **4**, 936-962

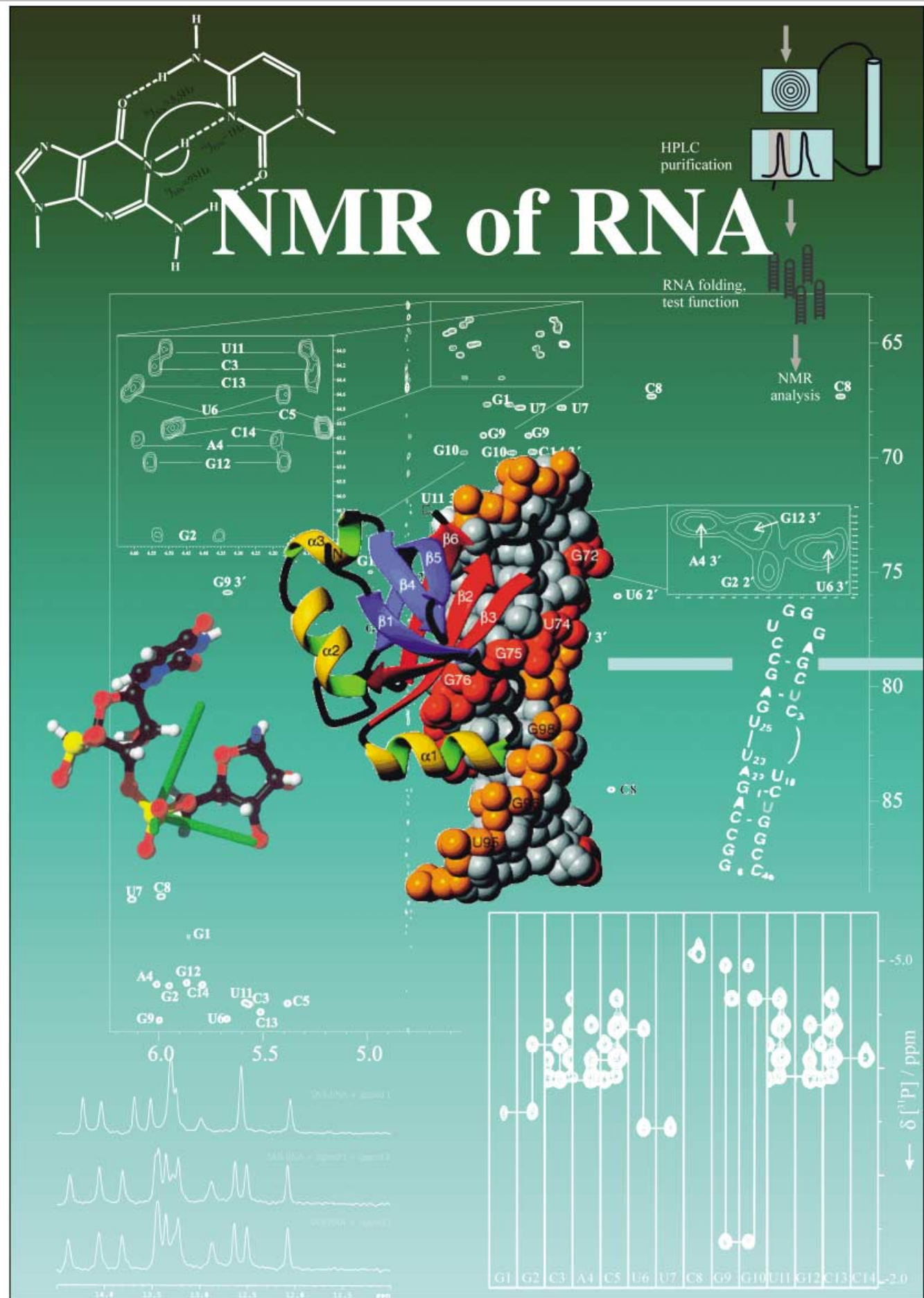
In this article state-of-the-art NMR technology for the investigation of RNA structure and dynamics has been reviewed. Special emphasis was placed on resonance assignment methods that are the first and crucial step of all NMR studies, on the determination of base-pair geometry, on the examination of local and global RNA conformation, and on the detection of interaction sites of RNA.

The NMR techniques were introduced using examples of NMR investigations on several different RNA molecules -mostly of the 14mer RNA cUUCGg tetraloop and the 5S RNA. Hereby it was possible to illustrate the information content accessible by NMR spectroscopy and the applicability of NMR methods to a wide range of biologically appealing RNA molecules.

The author of this thesis contributed to all aspects of writing the manuscript and performing the experiments.







# NMR Spectroscopy of RNA

Boris Fürtig, Christian Richter, Jens Wöhnert, and Harald Schwalbe<sup>\*[a]</sup>

*NMR spectroscopy is a powerful tool for studying proteins and nucleic acids in solution. This is illustrated by the fact that nearly half of all current RNA structures were determined by using NMR techniques. Information about the structure, dynamics, and interactions with other RNA molecules, proteins, ions, and small ligands can be obtained for RNA molecules up to 100 nucleotides. This review provides insight into the resonance assignment methods that are the first and crucial step of all NMR studies,*

*into the determination of base-pair geometry, into the examination of local and global RNA conformation, and into the detection of interaction sites of RNA. Examples of NMR investigations of RNA are given by using several different RNA molecules to illustrate the information content obtainable by NMR spectroscopy and the applicability of NMR techniques to a wide range of biologically interesting RNA molecules.*

## 1. Introduction

NMR spectroscopy is a powerful tool for studying the structure and dynamics of RNA molecules in solution and their interactions with ligands such as proteins, other nucleic acids, molecules of low molecular weight, ions, and solvent molecules. Up to now, nearly half of all three-dimensional structures of nucleic acids have been solved by NMR spectroscopy.

Within the current size limits of NMR measurements on RNA (about 100 nucleotides for information with intermediate resolution and 50 nucleotides for high-resolution structure characterization), there is a plethora of information that can be derived from NMR spectroscopic studies:

- The base-pairing pattern. This includes standard and non-standard Watson–Crick-type base pairs and allows verification and prediction of the secondary structure elements of RNA (discussed in Section 3) and determination of the base-pair dynamics.
- Information about conformational equilibria, such as those between hairpin and duplex structures (Section 4).
- Site-specific information about ion binding to RNA (Section 5).
- NMR spectroscopy resonance assignment of RNA (Section 6) and analysis of chemical shifts (Section 7).
- Delineation of secondary structure motifs, such as hairpins and bulges (Section 8).
- The local structure and dynamics (Section 9) and global structure of RNA derived from residual dipolar couplings (Section 10).
- Mapping of the interaction surfaces of RNA with small ligands, other RNA molecules, or proteins (Section 11).

The introduction of methods for the preparation of milligram quantities of RNA in an isotope-labeled ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ) form has been a prerequisite for all of these NMR studies and biochemical methods. The preparation of milligram quantities of RNA is discussed in Section 2.

With isotope-labeled RNA at hand, the first step is to assign every NMR-active atom ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ,  $^{31}\text{P}$ ) in the molecule to its

respective resonance in the NMR spectra; this is followed by interpretation of NMR parameters such as NOE contacts,  $J$  couplings, residual dipolar couplings, and cross-correlated relaxation rates for the determination of a three-dimensional structure. Resonance assignment in NMR spectroscopy is more difficult for RNA than for proteins. In comparison to proteins, the chemical-shift dispersion in the spectra of RNA, a biopolymer made out of only four different nucleotides, is significantly reduced. In addition, the A-form helix is the sole dominating secondary structure element found in RNA. Therefore, many nucleotides experience a similar chemical environment in helical secondary structures which, as a consequence, causes similar chemical shifts. Chemical-shift dispersion is often only observed in noncanonical structural elements such as hairpins, bulges, or internal loops.

The limited chemical-shift dispersion can be overcome by the examination of appropriately sized modular secondary structure elements (see Section 8) that are often involved in molecular recognition, whereas the canonical A-form helical elements frequently serve as a scaffold. In addition, the application of multidimensional heteronuclear NMR experiments (for example, as reviewed by Varani et al.<sup>[1]</sup> and by Wijmenga and van Buuren<sup>[2]</sup>) to completely or selectively isotope-labeled RNA molecules<sup>[3–5]</sup> increases the resolution observed in NMR experiments by the combination of a proton chemical-shift dimension with one or two heteronuclear chemical-shift dimensions.

In this review, we introduce NMR experiments that use a 14-mer RNA cUUCGg tetraloop as a model system with high spectral resolution (Figure 1).<sup>[6]</sup> The cUUCGg tetraloop is well character-

[a] H. Schwalbe, B. Fürtig, C. Richter, J. Wöhnert  
Institute for Organic Chemistry and Chemical Biology  
Center for Biomolecular Magnetic Resonance  
Johann Wolfgang Goethe University  
Marie-Curie-Strasse 11, 60439 Frankfurt am Main (Germany)  
Fax: (+49) 69-798-29515  
E-mail: schwalbe@nmr.uni-frankfurt.de

Harald Schwalbe, born in 1966, studied chemistry in Frankfurt at the Johann Wolfgang Goethe University and received his diploma and PhD thesis in the group of Prof. Griesinger. He was a postdoctoral fellow in the group of Prof. Dobson at the University of Oxford, UK, and then worked at the University of Frankfurt as a Habilitand from 1996–1999. In 1999, he moved to the Massachusetts Institute of Technology (MIT), Cambridge, USA, where he was first Assistant Professor and then Associate Professor in the Department of Chemistry. In 2001, he became a full professor for organic chemistry in Frankfurt. His research interests focus on the determination of structure, dynamics, and functions of proteins and RNA by using high-resolution NMR spectroscopy.



Left to right: B. Fürtig, H. Schwalbe, J. Wöhnert, C. Richter.

Jens Wöhnert, born in 1970, studied biochemistry at the Martin Luther University, Halle/Saale, and graduated in 1996. He obtained his PhD at the Institute of Molecular Biotechnology, Jena, where he applied NMR spectroscopy to study the structure of RNA and RNA–protein complexes. In 2000, he joined the group of Prof. Schwalbe at the Francis Bitter Magnet Laboratory at MIT. Since 2003, he has been a group leader in the SFB 579 “RNA–Ligand Interactions” at the Johann Wolfgang Goethe University, Frankfurt.

Christian Richter, born in 1970, studied chemistry at the Johann Wolfgang Goethe University, Frankfurt, where he graduated in 1996. He obtained his PhD in Frankfurt in the research group of Prof. Griesinger where he developed new NMR pulse sequences for the structure calculation of RNA. In 1999, he joined the application laboratory at Bruker BioSpin, first in Rheinstetten and then in 2000 at Fällanden, Switzerland. During this time, the main focus of his work was participation in the development of the CryoProbe. Since 2002, he has held a permanent position and is responsible for the NMR spectrometers and the NMR training of the researchers in the group of Prof. Schwalbe at the Johann Wolfgang Goethe University.

Boris Fürtig, born in 1978, studied biochemistry in Frankfurt and received his diploma in the group of Prof. Schwalbe in spring 2003. He is now working towards a PhD in the group of Prof. Schwalbe. His work is focused on the investigation of RNA by NMR spectroscopic methods.

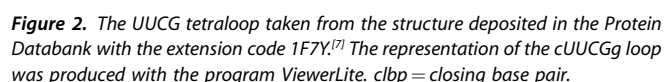
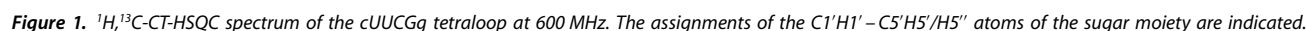
ized both by NMR spectroscopy and by X-ray crystallography.<sup>[7, 8]</sup> It is extremely stable<sup>[9]</sup> and frequently found in nature. It can also be considered as a secondary structure building block incorporated into many larger RNA structures (see Section 8). It consists of a five-base-pair A-form helix closed by the four loop nucleotides UUCG (Figure 2). In addition, spectra for a number of different RNA systems, including a 10-mer RNA containing a dynamic cUUUUG tetraloop, a double-stranded 16-mer RNA containing a tandem G:A mismatch, a secondary structure element of the 5S rRNA containing a cUUCGg tetraloop, and a 30-mer RNA derived from the coxsackie virus are also shown for comparison.

## 2. Preparation of Isotope-Labeled RNA

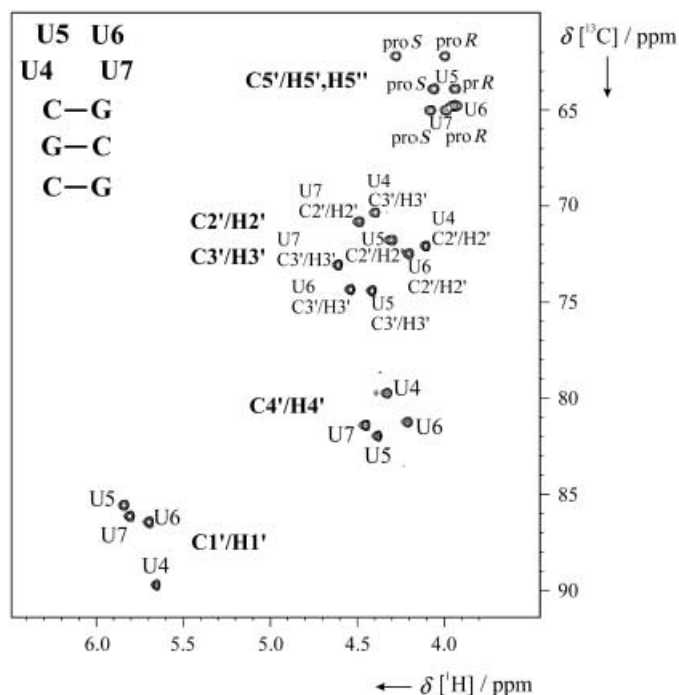
The restricted resolution of NMR spectra of RNA makes the introduction of stable  $^{13}\text{C}$ ,  $^{15}\text{N}$  isotopes an attractive tool for improving the quality of RNA structures determined by NMR spectroscopy. There are currently both biochemical and chemical methods (reviewed by Kojima et al.<sup>[10]</sup> and by Lagoja and

Herdewijn<sup>[11]</sup>) for the synthesis of isotope-labeled RNA. The main advantages of chemical synthesis and subsequent phosphoramidite chemistry for the synthesis of the oligonucleotides are the selective incorporation of labeled nucleotides at specific positions of interest (see Figure 3 as an example<sup>[12]</sup>) and the possibility to incorporate nonstandard nucleobases, which are quite often found in RNA.

However,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled precursors for the chemical synthesis of completely labeled RNA molecules are expensive and difficult to synthesize. Therefore, enzymatic in vitro transcription with DNA-dependent RNA-polymerases such as T7-, T3-, or SP6-RNA-polymerase<sup>[13–17]</sup> has become the method of choice in many groups for the synthesis of  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled RNA molecules. The most widely used polymerase is the T7-RNA-polymerase. Besides nucleotide triphosphates in an appropriately labeled form, the polymerase requires a DNA template. Linearized plasmids, synthetic double-stranded DNA, or single-stranded DNA with a double-stranded promoter region can serve as DNA templates (Figure 4). In general, linearized plasmids or double-stranded DNA appear to be more effective templates. There are certain

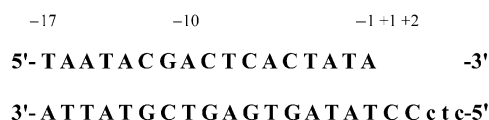


For the purification of the desired RNA from nonincorporated nucleotides and abortive transcription products, preparative



**Figure 3.** 2D  $^1\text{H}$ ,  $^{13}\text{C}$  correlation spectrum of the selectively  $^{13}\text{C}$ -labeled 3'-CGCUUUUGCG-5' RNA. In this sample, only the ribose rings of four uracils are  $^{13}\text{C}$ -labeled. This results in a considerable reduction of resonance overlap. If this spectrum were recorded on a larger RNA, for example, a 30-mer, and only four nucleotides were labeled, the chemical-shift resolution of all isotope-labeled residues would be retained.

polyacrylamide gel electrophoresis under denaturing conditions with gels containing 8 M urea is widely used. This method has the advantage that single nucleotide resolution can be achieved and the above-mentioned  $n+1$  and  $n+2$  products of the polymerase reaction can be separated. An alternative approach is the so-called 'ion-pair'-reversed phase HPLC.<sup>[21]</sup> Here, RNA complexed with the hydrophobic counterion tetrabutylammonium hydrogen sulphate is fractionated by increasing hydrophobicity by using an acetonitrile gradient.<sup>[22]</sup> The HPLC approach is especially efficient when preceded by an anion-exchange column step in which most of the unincorporated nucleotides and short abortive products can first be separated from the desired RNA product.<sup>[23]</sup> However, 'ion-pair'-reversed phase HPLC achieves single-nucleotide resolution only for oligonucleotides that have less than 20 residues. For longer RNA molecules, it is therefore advantageous to employ hammerhead ribozyme fusions since then essentially only two larger RNA molecules of different length need to be separated. Following the HPLC step, the fractions containing the desired RNA molecules are lyophilized and desalted on gel-filtration columns. The remaining tetrabutylammonium hydrogen sulphate can be separated from the RNA by precipi-



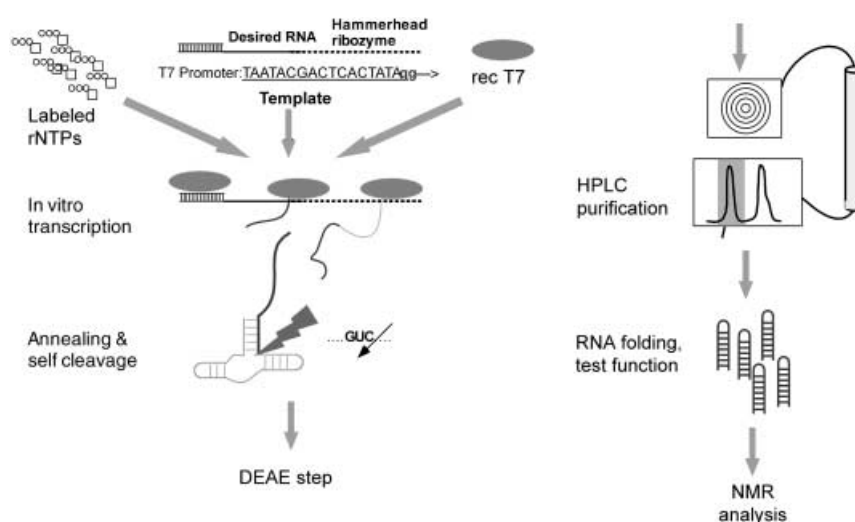
**Figure 4.** Consensus sequence of the T7-RNA-polymerase promoter. Nucleotides in positions +3 to +5 (lower case letters) can be chosen according to the sequence of the target RNA. Positions +1 and +2 should be guanine for efficient *in vitro* transcription.

tation with acetone/ $\text{LiClO}_4$ . Finally, the RNA has to be folded into a homogeneous form and exchanged into the buffer that will be used for NMR spectroscopy. Conditions for the correct folding of RNA have to be established individually for every new RNA molecule under study, therefore no general procedure can be given—especially for RNA molecules with more complex folds. However, simple hairpins can normally be obtained by a fast cooling step following a heat denaturation under conditions of low RNA and salt concentrations. High salt and RNA concentrations and slow cooling, on the other hand, favor the formation of duplex structures. A flow scheme for RNA preparation, as described in Stoldt et al.<sup>[23]</sup> and routinely used in our lab, is shown in Figure 5.

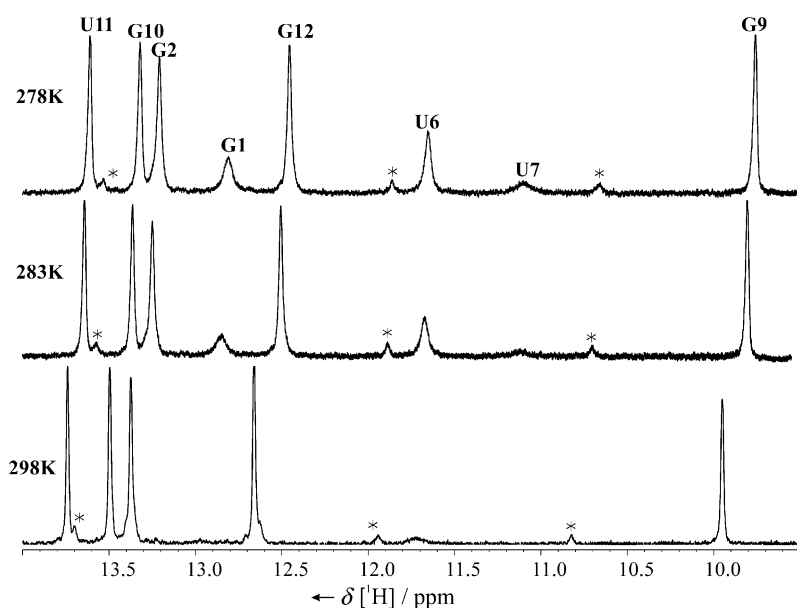
### 3. Base-Pairing Pattern

#### 3.1. Information about the base-pairing pattern from 1D NMR spectroscopy

Even in the early NMR studies of RNA molecules in the 1970s it was clear that the region of the imino proton resonances of the guanines and uracils between 10–15 ppm contained valuable information about base pairing in the RNA molecule. These signals are only observable when the imino protons are protected from exchange with the bulk solvent water and are therefore involved in hydrogen bonding. By counting the number of imino proton resonances, it is essentially possible



**Figure 5.** Flow scheme of the synthesis of isotope-labeled RNA molecules by *in vitro* transcription. For further explanations, see the text. DEAE = diethylaminoethyl, rec T7 = recombinant T7 polymerase, rNTPs = ribonucleotide triphosphates.

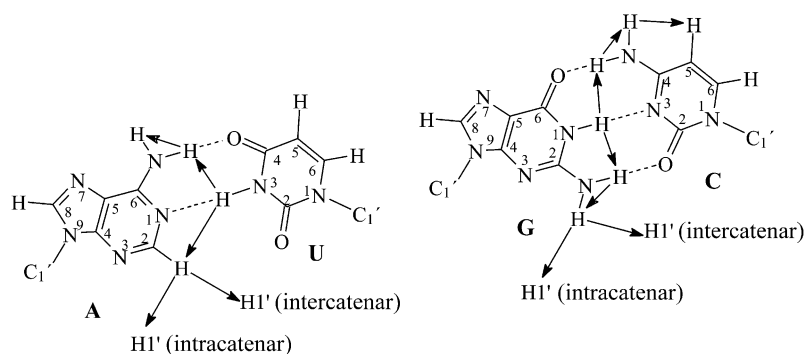


**Figure 6.** Imino region of the 1D  $^1\text{H}$  spectra of the 14-mer cUUCGg tetraloop RNA recorded at 600 MHz and three different temperatures (278, 283, and 298 K). Assignments are annotated for the imino proton resonances. Resonances stemming from the duplex form are indicated with \*.

to count the number of base pairs. In addition, imino protons of Watson–Crick base pairs tend to be found in the region of 12–15 ppm, whereas imino protons of noncanonical base pairs often have upfield chemical shifts (Figure 6). The stability of RNA molecules can be investigated by following temperature-induced changes in the 1D imino spectra.

### 3.2. Information about base pairing from homonuclear 2D NOESY experiments: Sequential assignment of imino proton resonances

The starting point for the determination of the hydrogen-bonding pattern is to sequentially assign the well-resolved resonances of the imino protons in a 2D NOESY experiment. As depicted later, the NOESY experiment



**Figure 7.** Schematic representation of the Watson–Crick base pairs A:U and G:C. The arrows indicate possible NOE contacts.

correlates all protons within a distance of 5 Å. The chemical shifts of the imino proton resonances in RNA depend strongly on the chemical environment, on base stacking, on ring effects, and on hydrogen bonding to solvent molecules. The imino proton of guanine bases resonates at 12–13.5 ppm if it is involved in a G:C base pair and at 10–12 ppm if it is part of a G:U base pair (Table 1).<sup>[24, 25]</sup>

The traditional assignment strategy for base pairs is based on the observation of NOE contacts (as indicated in Figure 7). A:U base pairs are easily identified by a strong NOE cross-peak between the H2 proton of adenine and the uracil H3 imino proton. In G:C base pairs, the H1 imino proton of guanine shows a strong NOE contact to the amino protons of the base-pairing cytosine. For this reason, the amino protons of cytosines involved in base pairing are easy to identify in comparison to the guanine and adenine amino protons. In  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled RNA molecules, identification of the cytosine amino protons can

**Table 1.** Chemical shifts ( $\delta$ ) of imino and amino resonances in Watson–Crick base pairs and wobble base pairs.

Base pair	Base-pair atoms	Chemical-shift regions [ppm]	Distance [Å]
G:C	G N1–H1...N3 C	12–13.5	1.89
	G N2–H2...O2 C	8–9	2.08
	G O2...H4–N4 C	8–9	1.71
	G N2–H2	6.5–7	–
	C N4–H4	6.5–7	–
U:A	U N3–H3...N1 A	13–15	1.93
	U O4...H6–N6 A	7.5–8.5	1.82
	A N6–H6	6.5–7	–
G:U	G N1–H1...O2(O4) U	10–12	1.76
	G O2...H3–N3 U	11–12	1.96

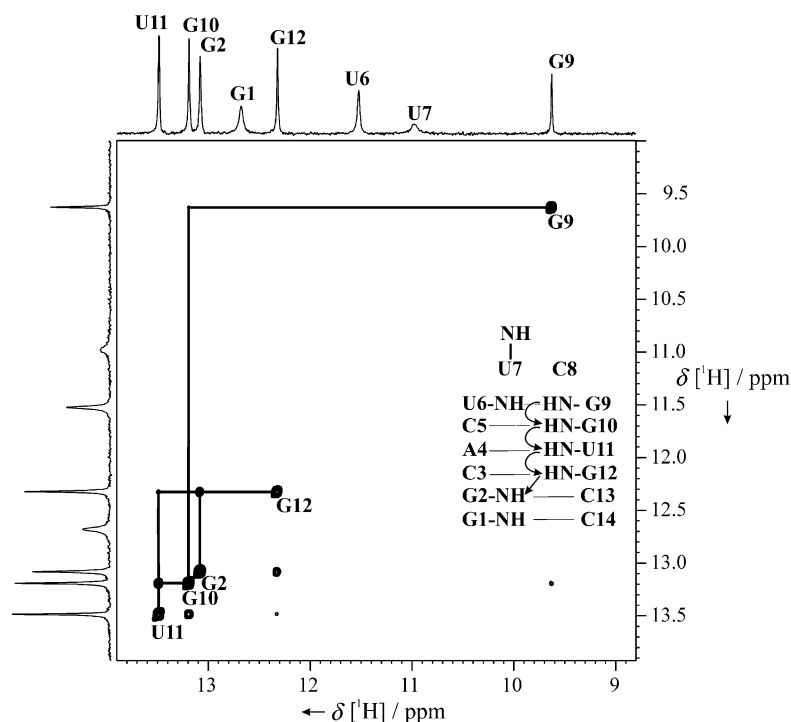
be used for further correlation to the H5 and H6 aromatic protons by direct correlation experiments (see Section 6). In helical RNA, NOESY cross-peaks can be observed between guanine H1 imino protons and cytosine H5 protons due to spin diffusion. Typical NOE interactions are shown in Table 2.

Imino–imino NOE cross-signals occur sequentially and between strands. As a consequence, there is no differentiation between sequential intra- and intercatenar NOE contacts. Chemical-shift information about imino protons can be helpful

**Table 2.** NOE cross-peaks that are usually observed in a helical region of RNA.

NOE correlation	Sequential	Intrabase pair
NH–NH	weak	strong (only G:U)
NH–H <sub>2</sub>	weak	strong
NH–NH <sub>2</sub>	weak	strong
NH <sub>2</sub> –NH <sub>2</sub>	weak	–





**Figure 8.** The imino–imino proton cross-peak region of the cUUCGg tetraloop in a 2D NOESY spectrum at 700 MHz with excitation sculpting for water suppression. The lines indicate the imino proton connectivity. The insert shows the sequence with the imino protons and the arrows indicate the sequential assignment walk. The NOESY mixing time was 300ms at 298K. As a result of exchange with the solvent, both cross-peaks and diagonal peaks of guanine 1, uridine 6, and uridine 7 are absent.

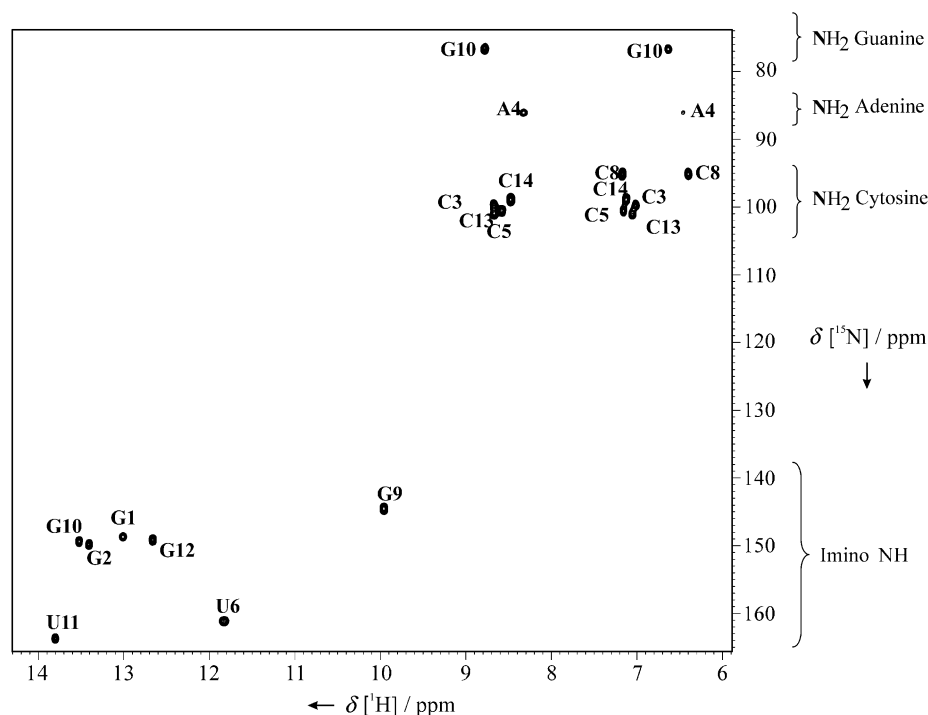
in the assignment of the NOESY signals (Table 1). In the cU<sup>6</sup>U<sup>7</sup>C<sup>8</sup>G<sup>9</sup>g tetraloop, the imino proton of G9, which resonates at 9.6 ppm, serves as a starting point for the assignment of the imino proton region of the NOESY experiment (Figure 8). The sequential NOE contacts can easily be assigned following the primary sequence of the cUUCGg tetraloop. In the tetraloop, the stability of the G:U base pair differs from that of the G:C base pairs. The expected NOE cross-signal to U6, which is involved in a U6:G9 base pair, is not visible due to fast water exchange of the imino proton of U6; this is also the reason for the absence of diagonal peaks for the nucleotides G1 and U7. The water exchange is also responsible for the apparent asymmetry of the cross-peak intensities of the NOESY spectrum.

By using NOE information, it is therefore possible to distinguish Watson–Crick G:C and A:U base pairs. Stable G:U wobble base pairs can be identified by a strong NOE contact between the guanine imino and uracil imino protons.

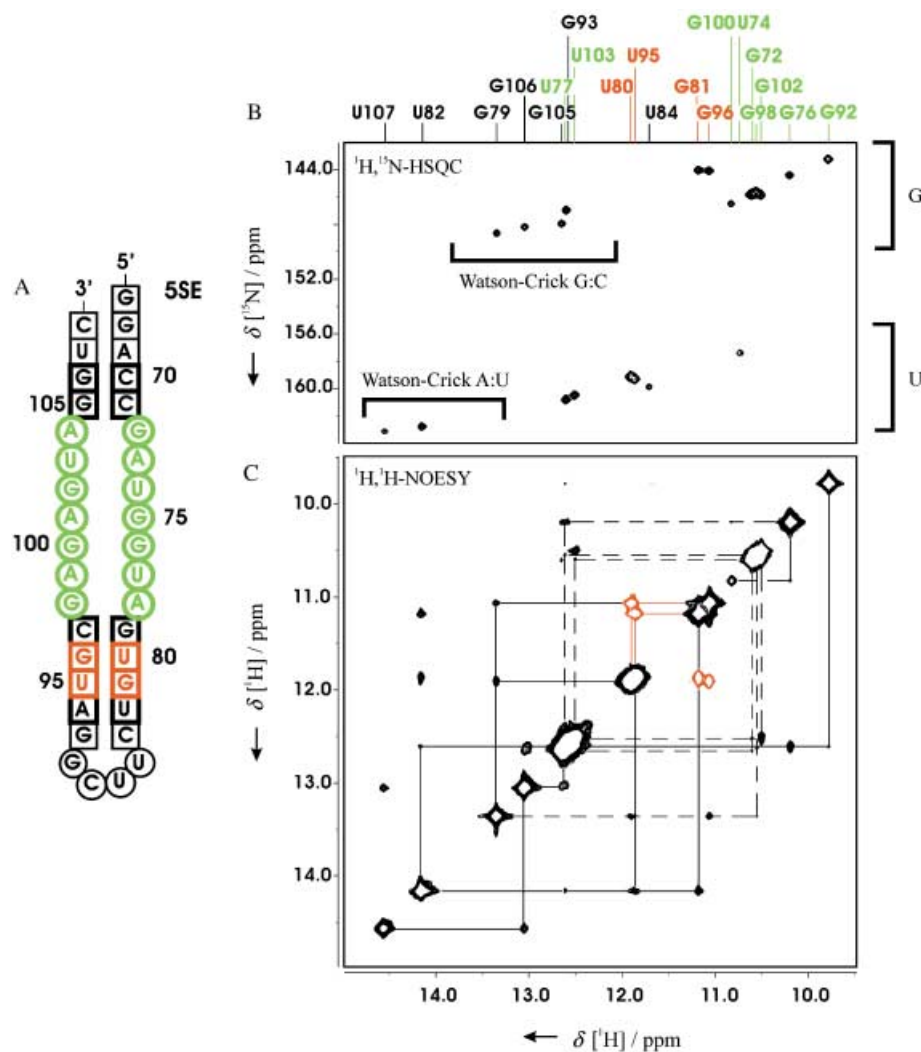
Furthermore, NOE contacts between the imino protons of neighboring base pairs are observable and facilitate the sequential assignment of the imino proton signals (Table 2). With this information, it has been possible to derive secondary structure models of tRNA molecules<sup>[26–28]</sup> and 5S rRNA.<sup>[29–31]</sup>

### 3.3. Information from heteronuclear 2D HSQC and HNN-COSY experiments: Elucidation of base pairing and secondary structure

With the introduction of <sup>15</sup>N-labeling and heteronuclear correlation experiments it becomes easier to distinguish between uracil and guanine imino resonances since the resonance frequencies of the imino nitrogen atoms are separated by ~10 ppm (Table 3; Figures 9 and 10). However, a major breakthrough for the elucidation of base-pairing and more complex hydrogen-bonding patterns in RNA by NMR spectroscopy was the discovery that sizeable scalar couplings across an N–H...N-type hydrogen bond can be observed between the nitrogen atom and proton of the hydrogen-bond donor and the nitrogen atom of the hydrogen-bond acceptor<sup>[32, 33]</sup> in RNA and DNA through the use of a so-called HNN-COSY experiment.



**Figure 9.** <sup>1</sup>H/<sup>15</sup>N-HSQC spectrum of the cUUCGg tetraloop at 700 MHz recorded at 283 K. The assignment is indicated.

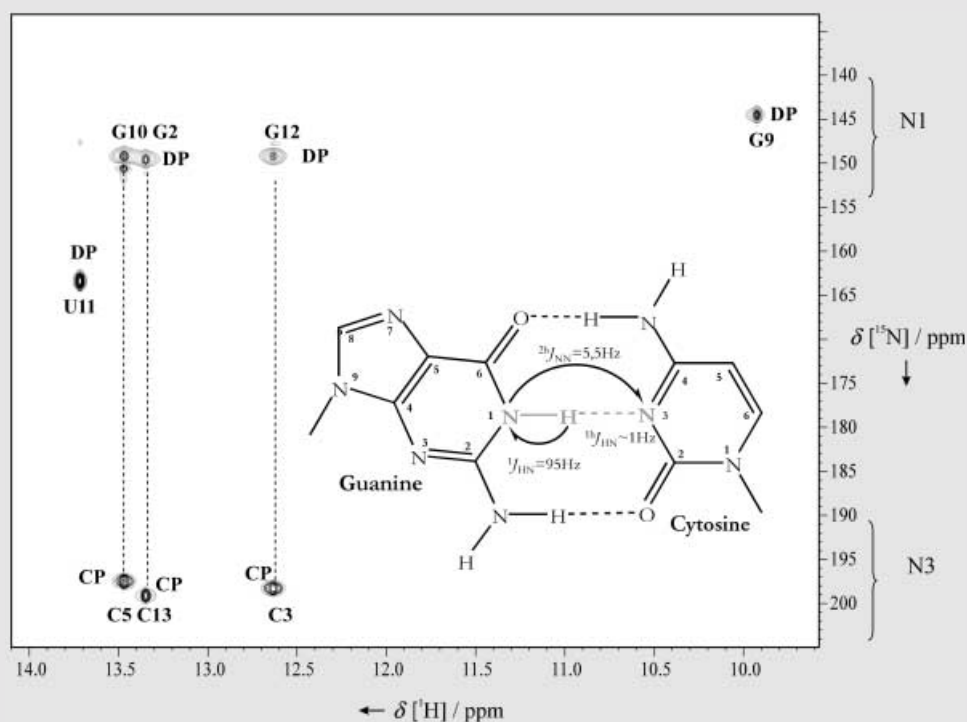


**Figure 10.** Example of secondary structure determination for longer RNA molecules based on imino group signals. A) Secondary structure of an RNA hairpin containing nucleotides 70–82 and 94–106 of *E. coli* 5S rRNA, the so-called E-loop. B) Imino group region of the  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum with the assignments of the proton resonances given at the top. Imino groups of noncanonical base pairs are shown in green, imino groups of wobble G:U base pairs are indicated in red, and imino groups of Watson–Crick base pairs are depicted in black. The  $^{15}\text{N}$ -chemical-shift ranges for G and U imino groups, respectively, are indicated on the right of the spectrum. The typical  $^1\text{H}$ -chemical-shift ranges for imino groups in Watson–Crick base pairs are indicated as well. C) Sequential assignments of imino resonances by sequential NOE contacts. Solid and dashed lines indicate sequential NOE interactions for the 5' and the 3' halves of the molecule, respectively. Intra-base-pair NOE contacts between G and U imino resonances typical for stable wobble G:U base pairs are marked with red lines.

Atoms	$\delta^1\text{H}$ [ppm]	$\delta^{15}\text{N}$ [ppm]	$\delta^{13}\text{C}$ [ppm]
C1'H1'	4.4–6.5	–	89–95
C2'H2'	4–5	–	70–80
C3'H3'	3.8–5	–	70–80
C4'H4'	3.8–4.8	–	81–86
C5'H5'	2.5–5	–	62–70
C2H2	6.5–8.5	–	150–154
C5H5	5.0–6.3	–	167–170
C6H6	7.0–7.7	–	137–140
C8H8	7.0–8.0	–	133–140
NH <sub>2</sub> (amino)	6.5–9.0	74–76 G 80–82 A 96–98 C	–
NH (imino)	9.0–15	145–148 G 157–162 U	–

The HN $\alpha$ -COSY experiment utilizes the through-space scalar  $^2\text{h}J(\text{N},\text{N})$  coupling constant in base pairs of RNA and DNA, which monitors the interaction from  $^{15}\text{N}$  imino donor nuclei and the corresponding  $^{15}\text{N}$  acceptor nuclei of the complementary bases. The  $^2\text{h}J(\text{N},\text{N})$  coupling constants between N3 of uracil and N1 of adenine as well as between N1 of guanine and N3 of cytosine are of the order of 5–7 Hz. During the pulse sequence, which is a straightforward extension of the HNHA experiment by Vuister and Bax,<sup>[34]</sup> one part of the originally generated magnetization remains on the starting nitrogen atom (annotated DP in Figure 11), whereas the other part is transferred to the hydrogen-bonded nitrogen atom (annotated CP in Figure 11). This gives rise to a diagonal peak and a cross-peak at the  $^{15}\text{N}$  frequency of N1 of guanine and N3 of cytosine. Due to a modulation of both peaks by either  $\sin^2(^2\text{h}J(\text{N},\text{N})T)$  or  $\cos^2(^2\text{h}J(\text{N},\text{N})T)$ , the size of the coupling constant can be obtained by comparing the peak volumes  $I^{\text{cross}}$  and  $I^{\text{diag}}$  of the cross-signal to the diagonal signal ( $I^{\text{cross}}/I^{\text{diag}} = \tan^2(2\pi \cdot ^2\text{h}J(\text{N},\text{N})T)$ ).





**Figure 11.** HNN-COSY experiment at 700 MHz and 298 K. On the right side, a Watson–Crick G:C base pair is depicted. The coupling constants are annotated. In this experiment, N1 of guanine can be correlated with the quaternary nitrogen atom of the cytosine residue.

It quickly became clear that similar scalar couplings could be observed for N–H...N-type hydrogen bonds in noncanonical base pairs as well. Remarkably, since the chemical shifts of the different possible nitrogen donor and acceptor groups are well separated, it is often possible to directly derive the base-pairing pattern. Thus, by using the cross-hydrogen-bond scalar couplings it was possible to characterize base pairing in A:A, G:A (see Figure 12), and G:G mismatches (for examples, see refs. [35–38]), reversed Hoogsteen A:U base pairs,<sup>[38]</sup> Watson–Crick-type U:C base pairs,<sup>[39]</sup> and base-triples<sup>[40, 41]</sup> in DNA, RNA, and RNA–protein complexes. Even intermolecular hydrogen bonds between guanine N7 nitrogen atoms and arginine side-chain guanidinium groups in an RNA–peptide complex could be detected.<sup>[42]</sup> This experiment can also be applied to A:U base pairs when D<sub>2</sub>O is the solvent.<sup>[40, 43, 44]</sup> The correlation starts on the H2 proton of adenine and the magnetization is transferred to N1 of adenine ( $^2J(\text{H2}, \text{N1}) = 14.5 \text{ Hz}$ ) and to the hydrogen-bonded N3 of uracil. By direct observation of the hydrogen bond in D<sub>2</sub>O, it is possible to differentiate between Watson–Crick and Hoogsteen base pairs.

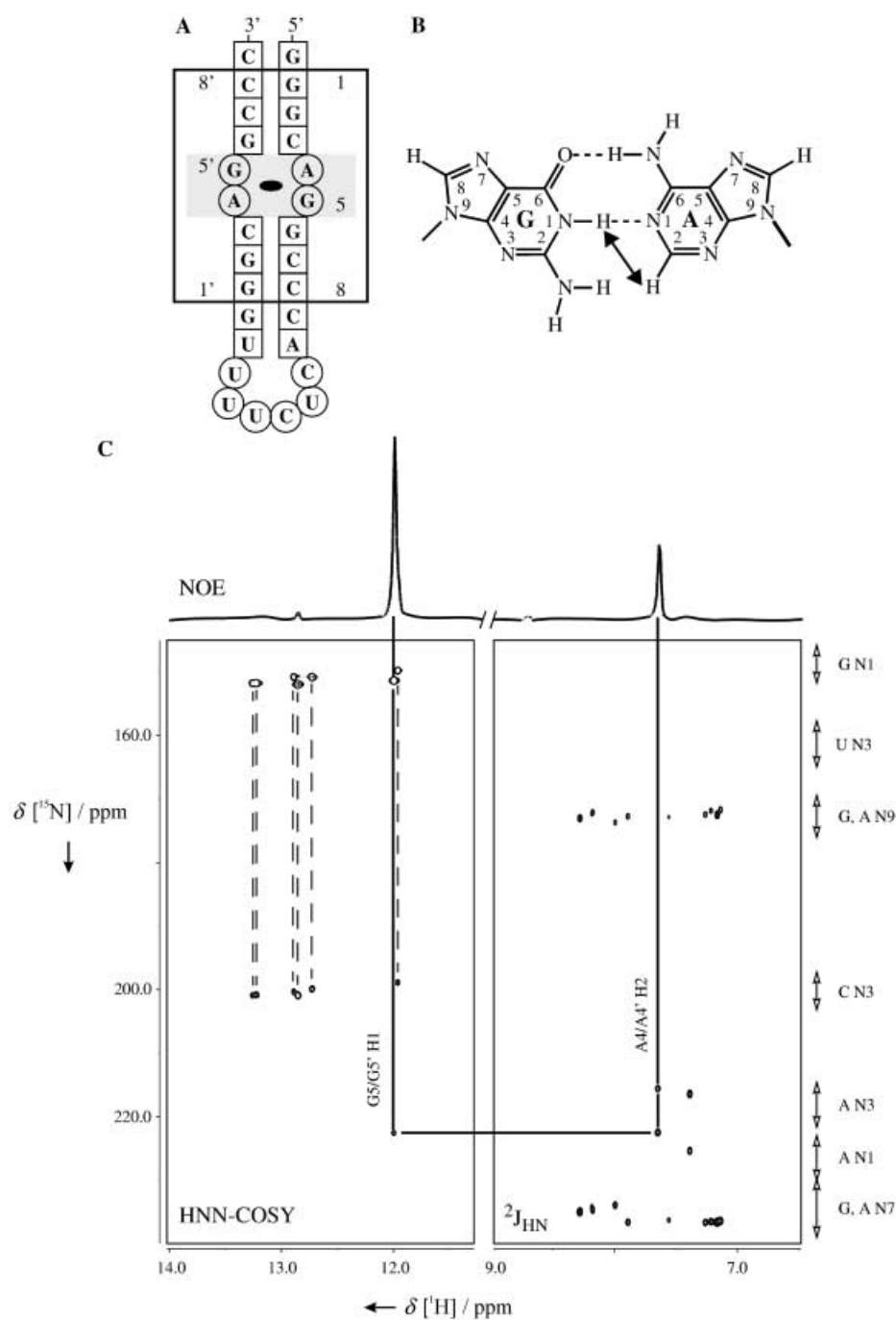
Scalar couplings across hydrogen bonds can be observed for N–H...O=C-type hydrogen bonds as well, at least in small nucleic acids.<sup>[45][42]</sup> However, there the coupling constants are much smaller than that of the N–H...N-type hydrogen bonds, which makes their direct detection by NMR spectroscopy much harder. Only recently, it was found that tertiary hydrogen bonds between a 2'-hydroxy group as the hydrogen-bond donor and a nitrogen atom as the acceptor also give rise to measurable scalar couplings.<sup>[46]</sup>

The elucidation of hydrogen-bonding patterns involving carbonyl groups as acceptors might also benefit from chemical-shift analysis. In particular, the chemical shifts of the C2 and C4 carbonyl groups of uracils that can be measured in <sup>13</sup>C-labeled RNA molecules are influenced by their involvement in hydrogen bonds. Thus, when comparing the chemical shifts of uracil C4 and C2 groups in Watson–Crick A:U base pairs, in which C4 is a hydrogen-bond acceptor and C2 is not hydrogen bonded, with those in a wobble G:U base pair, where C2 is hydrogen bonded and C4 is not, the signal for C4 in the A:U base pair is shifted downfield with regard to that for C4 in the G:U base pair, whereas the signal for C2 in the A:U base pair is shifted upfield. Interestingly, in reversed Hoogsteen A:U base pairs, where C2 of uracil is the hydrogen-bond acceptor in contrast to C4 in a Watson–Crick A:U

base pair, the C2 chemical shift is comparable with that in a wobble G:U base pair (Wöhnert, unpublished results). This relationship holds true for U:U base pairs as well. For instance, in asymmetric U:U base pairs, two N–H...O=C hydrogen bonds are observed, one with C2 as the acceptor and one with C4 as the acceptor. Accordingly, one C4 group and the C2 group of the other uracil are not involved in a hydrogen bond. For one uracil in the base pair, a downfield-shifted C4 signal and an upfield-shifted C2 signal are observed, whereas for the other uracil the C4 signal is shifted upfield and the C2 signal is shifted downfield (Figure 13).<sup>[47][39]</sup> However, these chemical-shift signatures of carbonyl groups must be further explored in other base-pairing geometries, before they will become a general tool for the assignment of hydrogen-bonding patterns.

#### 4. Description of a Hairpin–Duplex Equilibrium with NMR Spectroscopy

Secondary structure formation in RNA molecules depends on many factors, for example, sample concentration, salt concentration, pH value, and temperature. In the optimization of the formation of one specific conformation, monitoring of the imino region in a proton 1D NMR experiment proves helpful. As an example, we show the results for a cUUUUg tetraloop (5'-CGCUUUUGCG-3'; Figure 14). Similarly to UUU loops<sup>[22]</sup> the cUUUUg tetraloops can form a hairpin and a duplex; at elevated temperature, the hairpin melts to form single-stranded RNA (Figure 14E).<sup>[48]</sup> We start with low sample concentration (0.15 mM) and change both the phosphate and NaCl concen-

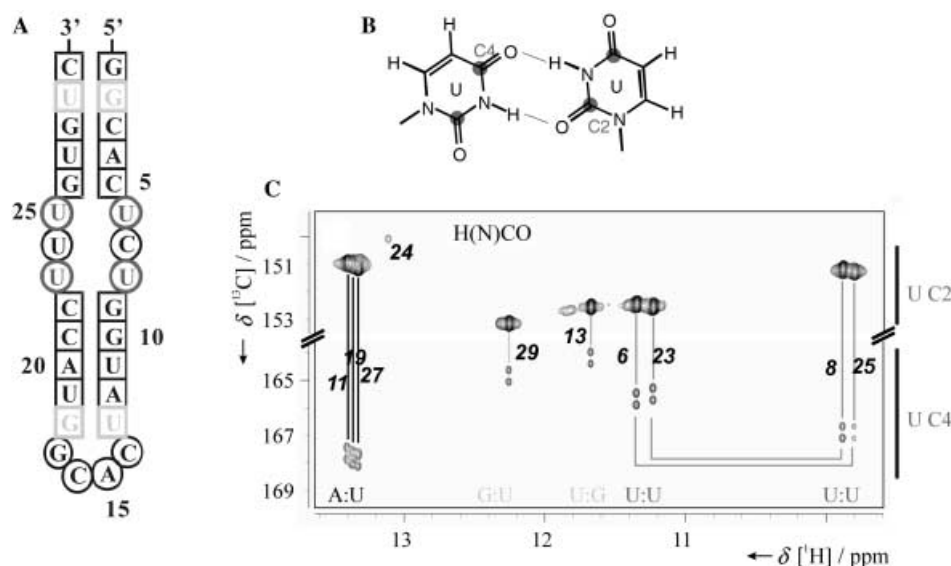


**Figure 12.** Identification of an imino-hydrogen-bonded G:A base pair with a GN1H1–AN1 hydrogen bond in the HNN-COSY experiment. A) Secondary structure of an RNA hairpin with a tandem G:A mismatch with a pseudo twofold symmetry. B) Geometry of an imino-hydrogen-bonded G:A base pair. An expected strong NOE contact between the G imino proton and the A H2 proton typical for this base-pairing geometry is indicated by an arrow. C) Identification of the donor and acceptor groups involved in the hydrogen bond by a combination of the HNN-COSY,  $^2J_{\text{HN}}\text{-}^1\text{H},^{15}\text{N}$ -HSQC, and NOESY spectra. The HNN-COSY spectrum shows cross-correlations between G H1 hydrogen atoms and G N1 and C N3 nitrogen atoms (dashed lines) typical for Watson–Crick G:C base pairs and between the G5/G5' H1 hydrogen atom and the G5/G5' N1 and A4/A4' N1 nitrogen atoms (solid line). The assignment of the nitrogen atom as an adenine N1 is supported by the  $^2J_{\text{HN}}\text{-}^1\text{H},^{15}\text{N}$ -HSQC spectrum, which provides a correlation from the A4/A4' H2 hydrogen atom to the A4/A4' N1 and N3 nitrogen atoms (solid line). Typical chemical-shift ranges of the relevant nitrogen atoms are indicated at the right side of the spectrum. A one-dimensional cross-section from a 2D  $^1\text{H},^1\text{H}$ -NOESY spectrum taken at the chemical shift of the G5/G5' H1 hydrogen atom shows the expected strong NOE cross-peak to the A4/A4' H2 hydrogen atom.

trations (Figure 14A). The phosphate concentration has only a small effect on the line shape of the resonances. In contrast, increasing the NaCl concentration has a pronounced effect on the line shape and the chemical shift of resonances in the spectra. The signals at 12.5–13 ppm are guanine imino protons and the signals at 10.5–11.5 ppm are uracil imino protons. The upfield shift of the uracil resonances is unusual and results from U:U base pairs. When the concentrations of RNA and NaCl are increased (120 mM), four sharp lines appear, which arise from the formation of two U:U base pairs in the duplex conformation (Figure 14B). The temperature dependence of the NMR spectra reveals that the duplex conformation is only stable at low temperature (3 °C; Figure 14C). At 27 °C, only the hairpin conformation is visible, and at even higher temperatures, the hairpin melts, the exchangeable resonances disappear, and the aromatic signals resonate at positions characteristic for single-stranded RNA. The temperature transitions are reversible and are also evidenced in native gels (Figure 14F, G). The 2D NOESY spectra of the uracil imino signals provide clear evidence that all four uracils form base pairs to each other (Figure 14D).

Such transitions of secondary structure are also observed for other RNA molecules.<sup>[49]</sup> The D-loop of *Escherichia coli* 5S rRNA, for example, can exist as both a stem-loop and as a duplex with two U:U base pairs in the bulge region. Additionally, only the dimeric form can be observed in crystals of the RNA molecule.

To distinguish between either duplex RNA or RNA hairpins, pulse-field-gradient NMR experiments<sup>[50, 51]</sup> can be used to measure diffusion con-



**Figure 13.** Chemical-shift dependence of the uracil C2 and C4 chemical shifts on base pairing. A) Secondary structure of an RNA hairpin containing A:U, wobble G:U, and asymmetric U:U base pairs. B) Geometry of an asymmetric U:U base pair. For one of the uracils C2 is the hydrogen-bond acceptor as in a wobble G:U base pair, whereas in the other one C4 is the hydrogen-bond acceptor similar to an A:U base pair. C) H(N)CO spectrum of the hairpin RNA. The chemical-shift ranges for the C2 and C4 carbons are indicated.

stants.<sup>[52]</sup> Due to the dependence on both the length and the shape of the RNA molecule, the diffusion constant proves to be a sensitive measure of both forms of an RNA molecule.

## 5. Elucidation of RNA – Metal-Ion Interactions

It has long been known that for many RNA molecules, metal ions are integral parts of their tertiary structure. This idea has recently been extended towards the notion that complex RNA tertiary structures possess a metal-ion core analogous to the hydrophobic core in protein structures.<sup>[53]</sup> Thus, it is not surprising that many RNA molecules contain specific binding sites for monovalent and divalent metal ions. NMR spectroscopic approaches to characterize metal-ion binding are based either on chemical-shift changes, paramagnetic line broadening, or intermolecular NOE contacts.

Chemical-shift changes can be measured with any ion. The chemical-shift changes induced by ion binding can be either directly transmitted due to the deshielding effect of the ion or due to an ion-induced structural transition of the RNA. For instance, chemical-shift changes induced by high concentrations of sodium, magnesium, and lead ions have been used to map out metal binding sites for the lead-dependent ribozyme.<sup>[54]</sup> Butcher et al.<sup>[55]</sup> used magnesium ions and cobalt hexamine ions to identify metal binding sites in the two domains of the hairpin ribozyme. In both ribozymes, it was found that the metal binding site is preformed in the RNA structure. Cobalt hexamine is an analogue of a hexahydrated magnesium ion but causes larger chemical-shift effects due to its higher charge. It has been introduced by Kieft and Tinoco as a mimic of outer-sphere complexation of hydrated magnesium ions by RNA.<sup>[56]</sup> Sensitive

reporters of chemical-shift changes induced by metal-ion binding are not only the resonances of the proton, nitrogen, and carbon nuclei of RNA but also the phosphorus resonances as part of the negatively charged RNA backbone. By using <sup>31</sup>P chemical-shift changes upon addition of magnesium, Hansen et al.<sup>[57]</sup> identified a high-affinity Mg<sup>2+</sup> binding site in the catalytic core of the hammerhead ribozyme that was not found in the published crystal structure. Butcher and co-workers<sup>[58]</sup> added an interesting new twist to this approach. In their investigation of a core component of the spliceosome, the U6 RNA intramolecular stem-loop, they inserted a single thiophosphate group into the backbone of the molecule and measured the chemical-shift changes induced by the thiophilic cadmium ion. Interestingly, they could show that cadmi-

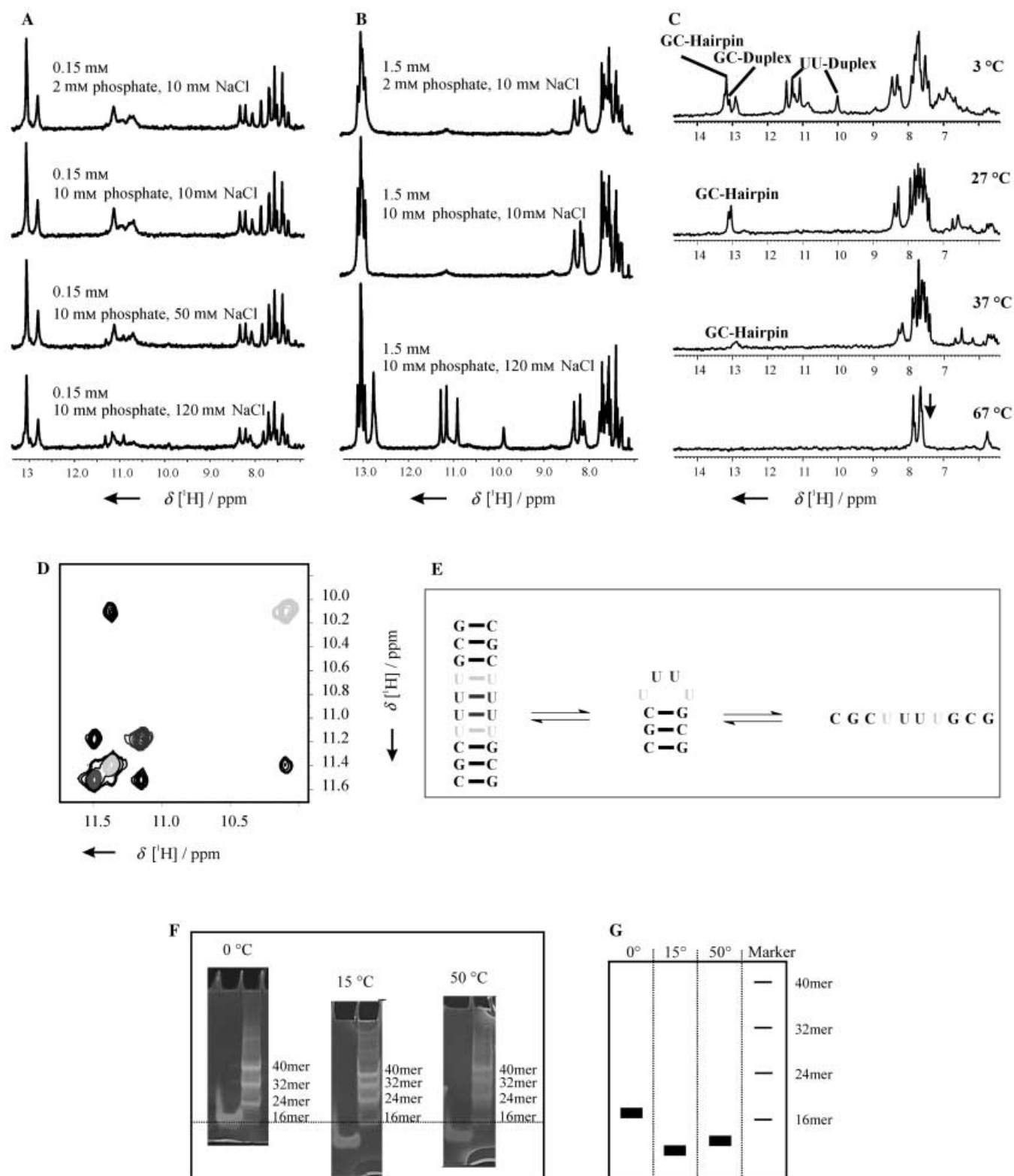
um bound with high diastereoselectivity when the pro *S* but not the pro *R* oxygen atom was replaced with a sulfur atom.

Chemical-shift mapping as a probe for magnesium binding sites is often complemented by paramagnetic line-broadening studies, where the paramagnetic manganese ion is used to replace magnesium.<sup>[59]</sup> The presence of the paramagnetic ion results in an enhanced relaxation of the nuclei in its vicinity and in turn leads to the disappearance of their NMR signals. The strength of the effect is dependent on the distance and can therefore be used to extract distance information that can be incorporated in the structure calculation of the ion–RNA complex.<sup>[55]</sup>

However, a more direct way to obtain distance information is to use intermolecular NOE contacts. These become available when cobalt hexamine is used as a mimic of magnesium<sup>[56]</sup> or ammonium ions are used to mimic monovalent cations such as potassium or sodium.<sup>[55]</sup> By using intermolecular NOE contacts between the amino groups of the complex ion and the RNA protons, the Tinoco group solved the structures of cobalt hexamine bound to domains of the group 1 intron,<sup>[56, 60]</sup> a frame-shifting pseudoknot of the mouse mammary tumor virus,<sup>[61]</sup> the GAAA stable tetraloop,<sup>[62]</sup> and the P4 element of the RNase P ribozyme.<sup>[63]</sup> In many cases, the metal binding pocket was formed by tandem mismatch base pairs, such as tandem G:U base pairs<sup>[60]</sup> and tandem G:A base pairs.<sup>[62]</sup>

## 6. NMR Resonance Assignment of RNA

The first step for a complete resonance assignment is the identification of nucleobase spin systems (<sup>1</sup>H<sup>N</sup>, NH<sub>2</sub>, H2, H5, H6, and H8). Secondly, the protons of the sugar moiety (H1', H2', H3',



**Figure 14.** 1D JR-Echo spectra at 600 MHz on the cUUUg loop RNA. A) With different salt concentrations and 0.15 mM RNA at 7 °C. B) With different salt concentrations and 1.5 mM RNA at 7 °C. C) With a 1.5 mM RNA sample at different temperatures. D) The 2D imino proton NOESY spectrum for the U duplex cross-peaks. The indicated black and grey cross-signals belong together to one base pair. E) Schematic representation of the temperature-induced conformational change. F) The native gels of the cUUUg tetraloop RNA at three different temperatures with a DNA marker. The gels were calibrated on the lane of the DNA 16 mer for better comparison. G) Schematic representation of the native polyacrylamide gels.

H4', H5', H5'') are assigned. The spin systems of the nucleobases are correlated with the sugar spin system either by observation of NOE contacts between aromatic and H1' protons or by direct correlation of the H1' resonances with the resonances of the glycosidic nitrogen atoms of the nucleobases in HCN-type experiments. As a third step, the assignments for the ribose spin systems are completed by optimized HCCH-TOCSY experiments. Then, sequential links between the nucleotide spin systems are established by sequential NOE experiments (H6/H8 to H1') or by HCP and HCP-TOCSY experiments. In these four steps, a complete assignment of all atoms can be established.

### 6.1 Assignment of the aromatic protons of the nucleobases: Correlation of imino to aromatic protons

Traditionally, aromatic protons in pyrimidines of RNA are assigned by examining NOE cross-signals from imino protons to the H5 proton of cytosine and uracil. By using long mixing times, cross-peaks to H6 are also observed due to spin diffusion. Unfortunately, there is no direct solution for assigning H8 of guanine and adenine. In uniformly  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled RNA molecules, a direct correlation of H8 resonances with N7 and N9 nitrogen by using  $^nJ$  coupling constants is possible. The different transfer pathways of the available pulse sequences are summarized in Figure 15.

The first experiments to correlate hydrogen atoms from the Watson–Crick site of the nucleobase to the hydrogen atoms H8 (purines) and H6/H5 (pyrimidines) were developed by Simorre et al.,<sup>[64, 65]</sup> Sklenar et al.,<sup>[66]</sup> and Fiala et al.<sup>[67]</sup> Simorre et al.<sup>[65]</sup> (Fig-

ure 15A) introduced a pulse sequence that was specific for guanine nucleotides. From the imino proton, the magnetization is transferred through an INEPT step to N1. By using CN-TOCSY transfer (DIPS-3, field strength 1.9 kHz), the magnetization is transferred to the base carbon C6. The  $^{15}\text{N}$  and  $^{13}\text{C}$  carriers are set to 146 ppm and 161 ppm. The homonuclear TOCSY-period (FLOPSY-8) with a duration of 38 ms and an rf field strength of 5 kHz transfers the magnetization directly to C8 or through C6 and C4 to C8. For this step, the carrier is set to 145 ppm. Subsequently, the in-phase C8 coherence is transferred by reverse refocused INEPT steps to H8.

In Figure 15D, E, the transfer pathways exploited in the experiment reported by Sklenar et al.<sup>[66]</sup> are depicted: the magnetization is transferred in a reverse manner from H8/H6 to the imino proton. For the CC transfer and the CN transfer the same rf field strength of 2.9 kHz is applied. Mixing times for CC-TOCSY and CN-TOCSY transfer are 19 ms and 58 ms, respectively. The carrier is set to 150 ppm for  $^{13}\text{C}$  and 153 ppm for  $^{15}\text{N}$ . This pulse sequence is also applicable to guanine and uracil bases.

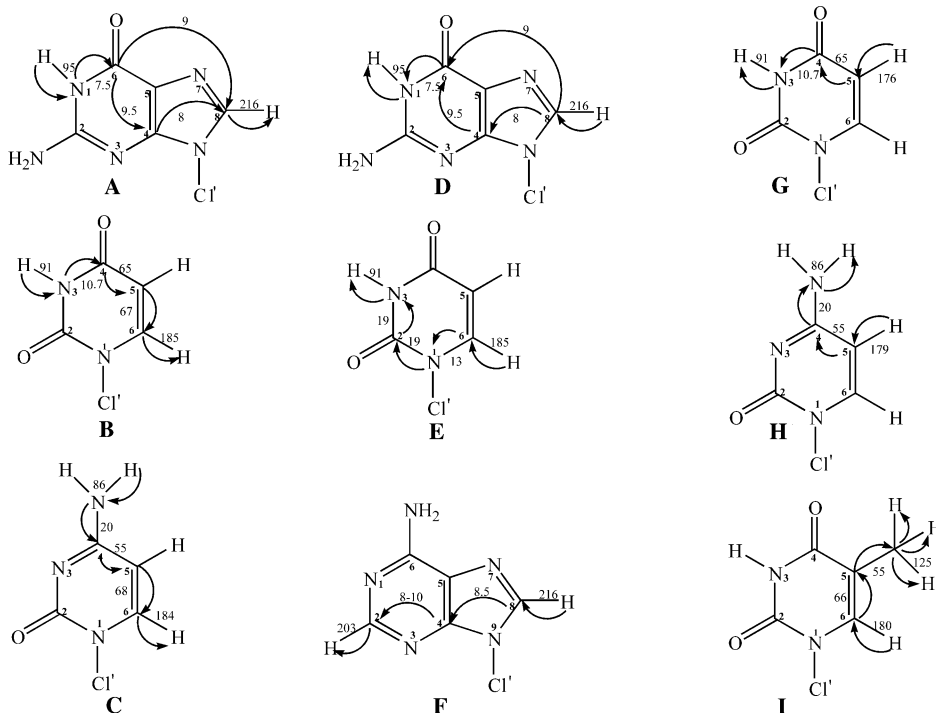
The pulse sequence by Fiala et al.<sup>[67]</sup> (transfer pathway in Figure 15D) correlates the nonexchangeable protons with the exchangeable protons in guanine bases. The CN transfer is achieved by INEPT steps with a selective C6 pulse. The carrier for  $^{15}\text{N}$  is set to 120 ppm, which corresponds to the middle of the resonance frequency of N1 and  $\text{NH}_2$ . The CC-TOCSY mixing time is 55 ms with a field strength of 3 kHz on carrier at 150 ppm.

All three experiments are quite similar; the main difference is the TOCSY transfer step and the chosen carrier frequency. Also the last experiment uses a CN-INEPT instead of CN-TOCSY. Wijmenga and van Buren<sup>[2]</sup> simulated the transfer efficiency of these sequences. The efficiency depends only on the CC-TOCSY step. CN-INEPT and CN-TOCSY transfer steps are equal in their transfer efficiencies. The simulation shows that the main transfer is supplied by the C8–C6

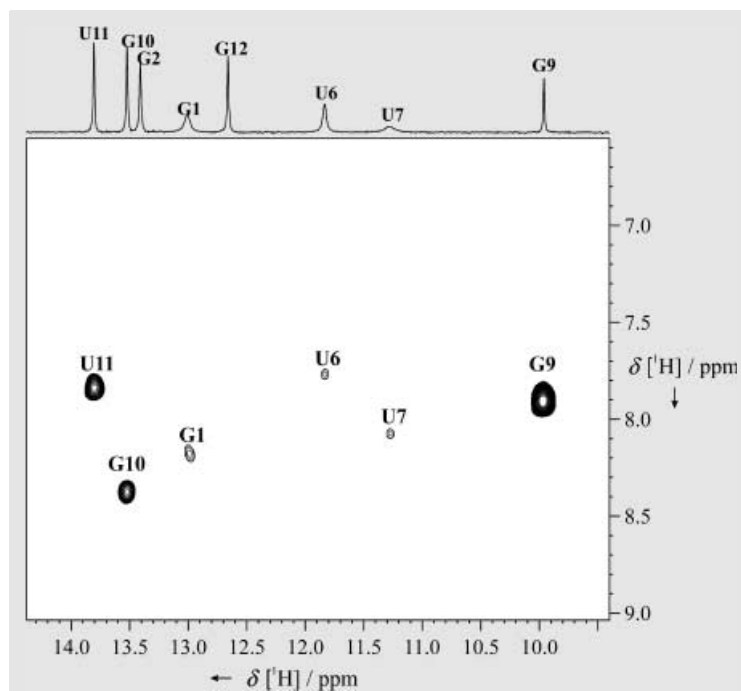
route and indicates that the sequence by Simorre et al. is the most efficient for larger systems. The sequence by Sklenar et al. has the advantage that it correlates imino protons in guanines and in uracils with nonexchangeable protons in a single experiment; the disadvantage is that it shows the worst transfer efficiency as compared to the other sequences. However, for RNA molecules which give rise to high-quality NMR spectra, the desired correlation peaks can be observed, as shown in Figure 16 where the sequence was applied to the cUUCGg tetraloop RNA.

A second pulse sequence by Simorre et al.<sup>[64]</sup> (Figure 15B, C) correlates the exchangeable protons in uracil (H3) and guanine ( $\text{NH}_2$ ) with the H6 protons. In this experiment, the magnetization is transferred from C4 to C6 not by a CC-TOCSY step but by using two INEPT transfers because of the different chemical shifts and coupling constants of C4, C5, and C6.

The correlation of the H2 protons with the H8 protons can be achieved in an HCCH-TOCSY experiment<sup>[68, 69]</sup> (Figure 15F). The carbon network in



**Figure 15.** The transfer ways for the NMR experiments correlating nuclei in the nucleobases: A) HNC-TOCSY-CH, B) and C) HNCCCH, D) and E) HCCNH TOCSY, F) HCCH TOCSY, G) and H) H5(C5C4N)H, and I) H6C6CCH(Me). References are given in the text, numbers on the arrows indicate the size of scalar couplings between the connected atoms in Hz.



**Figure 16.** 2D HCCNH experiment (Sklénar et al.<sup>[66]</sup>) on the cUUCGg tetraloop RNA at 700 MHz.

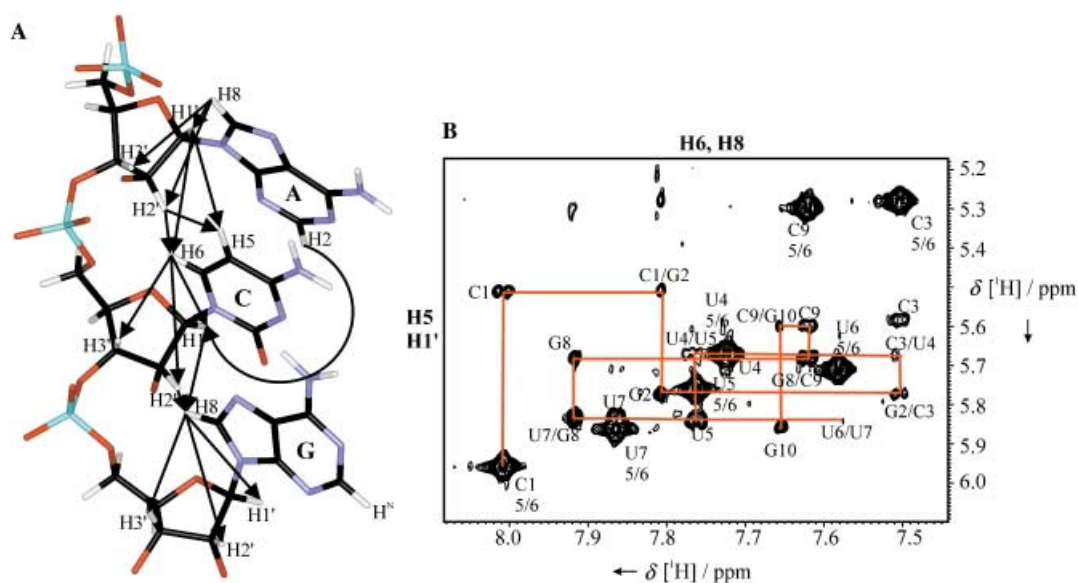
adenine is relatively complex and with a CC-TOCSY experiment a considerable part of the magnetization is lost through unwanted coherence transfer pathways (for example, involving the spin of C5). An alternative way to achieve this correlation is the HMBC experiment.<sup>[70]</sup>

Wöhnert et al.<sup>[71]</sup> (Figure 15 G, H) correlated the H5 proton in uracil and cytosine with the exchangeable protons by applying sequential relay steps with selective pulses. For <sup>13</sup>C-labeled DNA, it is possible to correlate the H6 proton with the methyl group in thymine nucleotides<sup>[72]</sup> (Figure 15 I).

Recently, a modified pulse sequence has been developed by Wöhnert et al.<sup>[73]</sup> for the simultaneous correlation of either H6/H5 or C5/C6 to exchangeable protons in pyrimidines. The experiment uses the same transfer pathway depicted in Figure 15 B, C but in the opposite direction, starting from the nonexchangeable protons. This experiment has a high sensitivity.

## 6.2 Assignment of protons of the sugar moiety: Correlation of aromatic protons to sugar protons

Interresidual sequential assignment in helical A-form RNA is obtained by observation of NOE contacts between aromatic and sugar protons (Figure 17). All aromatic protons show NOE cross-peaks to the H1' protons of their own ribose and to the preceding nucleotide in the 5' direction. Although the sequential H1' aromatic proton distance is larger than 4 Å in helical A-form RNA, a cross-peak can be observed due to spin diffusion through the H2' proton.<sup>[74]</sup> The major problem in applying this NOE-based assignment procedure is to find suitable starting or anchoring points for the assignment walk. One possibility is to use the H2 protons of adenine. The distance between H2 and the sequential H1' proton in helical A-form RNA is shorter than 4 Å. The H2 proton can be identified by NOE contacts to the uracil imino proton in H<sub>2</sub>O or in a <sup>1</sup>H,<sup>13</sup>C-HSQC, because the C2H2 region (150 ppm) is well resolved and no signals arising from other resonances are observed.



**Figure 17.** A) Schematic representation of the sequential assignment strategy in helical A-form RNA for nonexchangeable protons. The arrows show the intraresidual NOE connectivities between the aromatic and the sugar protons H1' – H3' and the sequential NOE correlation between the H3' – H6, H8 protons and the H5 – H1' protons. The sequential assignment of the helical A-form conformation is possible by determination of these NOE cross-peaks. In addition to the exchangeable protons, only the intercatenar NOE interactions between the adenine H2 and H1' of the corresponding RNA strand give information about the helical conformation. B) An example for the NOESY assignment procedure shown for the cUUUUG loop RNA. The NOESY spectrum was recorded in D<sub>2</sub>O at 600 MHz and the mixing time was 300 ms. Annotation by using two residues indicates connectivities due to sequential NOE contacts and annotation with one nucleotide indicates intraresidual NOE interactions.



The NOE cross-peaks of 5'- and/or 3'-terminal nucleotides can serve as alternative starting points. The chemical shifts of the protons in terminal nucleotides are shifted, for example, the chemical shifts of H2' and H3' protons at the 3' end are moved lower field because of the missing phosphate group.

In A-form helices, three relatively strong NOE contacts and one weak NOE are observed for every aromatic proton in the aromatic-sugar region in the NOESY spectra recorded with short mixing times (50–100 ms; Table 4). Strong NOE contacts reveal inter- and intranucleotide connectivities to the H3' proton (distances smaller than about 3 Å) and a very strong NOE exists to the sequential H2' proton (distance about 2 Å). The intra-residual aromatic/H2' cross-peak is weak. These strong NOE contacts to the neighboring nucleotides arise due to intracatenar base stacking.

**Table 4.** Typical NOE interactions observed in helical A-form conformations.<sup>[a]</sup>

NOE interaction	Sequential	Intraresidual
H1' – H8/H6	w	w
H2' – H8/H6	s	w
H3' – H8/H6	m	m
H2' – H1'	w	s
H6/H5	w	s

[a] w = weak (4–6 Å), m = medium (2.5–4 Å), s = strong (1.5–2.5 Å).

By using long mixing times (> 300 ms), spin-diffusion-mediated NOE contacts to the sugar protons H4', H5', and H5'' can be observed. However, in most RNA sequences, their assignment is difficult because of the low chemical-shift dispersion in this region. The aromatic-aromatic interresidual NOE contacts in NOESY experiments with long mixing times are very helpful for sequential assignment. In summary, for the NOE-based sequential assignment procedure, a careful analysis of the cross-peaks observed in different experiments yields the desired information.

The pyrimidine H5 and H6 resonances are easy to identify because of the strong NOE cross-signals observed in NOESY experiments with short mixing times. In addition, the signals are split by 7 Hz due to the homonuclear  $^3J(\text{H5}, \text{H6})$  coupling. In  $^{13}\text{C}$ -labeled RNA, the identical chemical-shift resonances of C6 and C8 can be distinguished by the  $^1J(\text{C5}, \text{C6})$  coupling constant of the C6 resonances. Cytosine and uracil can be distinguished by the difference between the chemical shifts of the C5 resonances in both nucleotides.

The assignment of resonances in single-stranded RNA is very difficult; base stacking is missing which leads to a much reduced chemical-shift dispersion. By using NOESY experiments with different mixing times, it is possible to assign the aromatic resonances (H2, H5, H6, and H8) and the sugar resonances (H1' – H3'). Only in exceptional cases is the assignment of the rest of the sugar protons possible. A sequential assignment breaks at nonpaired nucleotides, for example, bulges. For these regions, isotope labeling of RNA molecules has proven indispensable.

In addition, in  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled RNA, heteronuclear-edited 3D NOESY-HSQC experiments, in which the chemical shifts of the

heteronuclei  $^{13}\text{C}$  or  $^{15}\text{N}$  or both are evolved, can be used for assignment. Some useful experiments are:

- 2D ( $^1\text{H}, ^1\text{H}$ )-NOESY in  $\text{H}_2\text{O}$  with different mixing times
- 2D ( $^1\text{H}, ^1\text{H}$ )-CPMG-NOESY in  $\text{H}_2\text{O}$ <sup>[75]</sup>
- 3D ( $^1\text{H}, ^1\text{H}, ^{15}\text{N}$ )-NOESY-HSQC in  $\text{H}_2\text{O}$
- 3D ( $^1\text{H}, ^1\text{H}, ^{13}\text{C}$ )-NOESY-HSQC in  $\text{D}_2\text{O}$
- 3D ( $^1\text{H}, ^1\text{H}, ^{13}\text{C}$ )-NOESY-HSQC selective for C2, C6, and C8 in  $\text{H}_2\text{O}/\text{D}_2\text{O}$
- 3D ( $^1\text{H}, ^1\text{H}, ^{13}\text{C}$ )-NOESY-HSQC selective for  $\text{C}_{\text{ribose}}$  in  $\text{D}_2\text{O}$
- For base-specifically labeled RNA molecules or RNA complexes,  $\omega_1$ - $^{13}\text{C}$ -filtered,  $\omega_3$ - $^{13}\text{C}$ -edited 3D ( $^1\text{H}, ^1\text{H}, ^{13}\text{C}$ )-NOESY-HSQC in  $\text{D}_2\text{O}$ <sup>[76]</sup>

### 6.3 Direct correlation of resonances from the nucleobases with resonance from the ribose sugar moiety

Correlation of the nucleobase protons with sugar protons is obtained in HCN triple-resonance experiments in  $^{13}\text{C}, ^{15}\text{N}$ -labeled RNA.<sup>[77–82]</sup> This experiment is very important for the assignment because of the quite different nitrogen chemical shifts of purines and pyrimidines. In addition, a smaller difference in the  $^{15}\text{N}$  chemical shift allows cytosine and uracil to be distinguished (Figure 18).

The easiest implementation of the HCN experiment uses INEPT steps in an out-and-back experiment.<sup>[77]</sup> The delay for the CH transfer is a compromise for the varying  $^1J(\text{C}, \text{H})$  coupling constants:  $^1J(\text{C1}', \text{H1}') = 168 \text{ Hz}$ ,  $^1J(\text{C6}, \text{H6}) = 185 \text{ Hz}$ , and  $^1J(\text{C8}, \text{H8}) = 216 \text{ Hz}$ . The  $^1J(\text{C1}', \text{C2}')$  coupling constant can be suppressed in the simultaneous  $t_1$  time and CN transfer by either using a constant time delay ( $\text{CT} = 1/{}^1J(\text{C1}', \text{C2}')$ ),<sup>[77]</sup> a C2'-selective refocusing pulse,<sup>[78]</sup> or a C1'-selective inversion pulse.<sup>[80]</sup> The coherence transfer from the aromatic proton to the nitrogen atom is optimized by using a  $^{15}\text{N}$ -selective pulse (150 ppm) to suppress the magnetization transfer to N7 (220 ppm) for purine. The C6/C8 → N1/N9 transfer is optimal for a delay of 40 ms. The transfer efficiency in the HCN experiment was simulated by Wijmenga et al.<sup>[2]</sup> The simulation reveals that implementation of selective C2' decoupling provides the best sensitivity. However, a more important role is played by the decoupling of the aromatic carbon atoms C2, C4, and C5 with a C6/C8-selective pulse.

As originally demonstrated for proteins by Grzesiek and Bax,<sup>[83]</sup> a sensitivity enhancement for larger isotope-labeled proteins or RNA molecules can be obtained in certain cases by evolution of multiquantum (DQ/ZQ coherence, DQ = double quantum, ZQ = zero quantum) instead of single-quantum coherence for HC correlation.<sup>[79]</sup> For DQ/ZQ coherence, relaxation due to CH dipole-dipole interactions is inactive, which results in a considerable gain in sensitivity. More sensitivity enhancement, at least for correlation peaks involving aromatic carbon atoms, is obtained by exploiting the TROSY effect in the HCN experiment.<sup>[81]</sup> The combination of TROSY and multiquantum excitation is therefore optimal for larger RNA molecules.<sup>[82]</sup>

In the HCNCH-type<sup>[84]</sup> experiments, the HCN experiment is expanded by a relay step. The H1' proton magnetization is transferred in a manner analogous to the HCN experiment to N1/N9 and forward through C6/C8 to H6/H8. In this experiment, it is necessary to choose an optimal combination of selective pulses, delays, and carrier frequencies for the transfer. For the purine nucleotides, alternative implementations have been reported<sup>[85]</sup> that use the  $^2J(\text{N9}, \text{H8})$  coupling constant (purine:  $^2J(\text{N9}, \text{H8}) = 8 \text{ Hz}$ ; pyrimidine:  $^2J(\text{N9}, \text{H8}) = 4 \text{ Hz}$ ) for the direct transfer from N9 to H8.

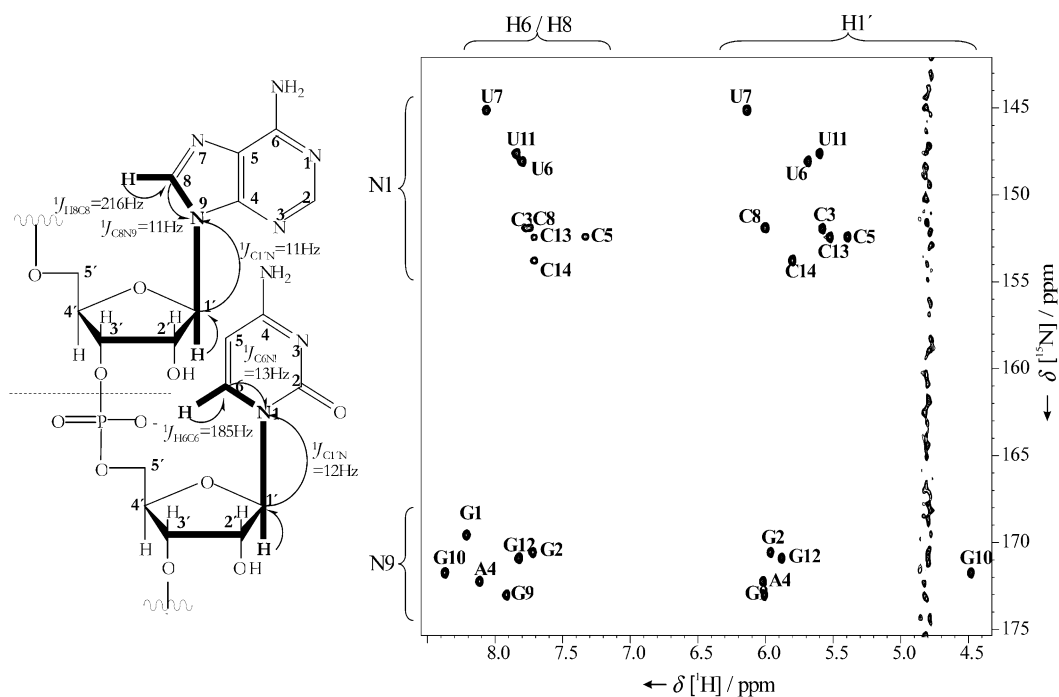


Figure 18. 2D H(C)N experiment at 700 MHz and 298 K for the cUUCGg tetraloop.

#### 6.4 Complete resonance assignment of the nucleobases

As described above, the resonance assignment of the imino and base protons can be derived by using several NOESY experiments with different mixing times. In the next step, the application of standard  $^1\text{H}$ , $^{13}\text{C}$ -HSQC and  $^1\text{H}$ , $^{15}\text{N}$ -HSQC experiments reveal the resonances of the heteroatoms attached to the assigned protons.

For the assignment of the N7 and N9 nitrogen atoms of guanine and adenine and the N1 and N3 atoms of adenine one uses a modified HSQC experiment, the so-called  $^2\text{J}$ - $^{15}\text{N}$ -HSQC.<sup>[86]</sup> In this experiment, the delay of the INEPT step for magnetization transfer from proton to nitrogen is optimized for the  $^2\text{J}(\text{H},\text{N})$  coupling. Therefore, this experiment correlates the aromatic protons H8 and H2 with nitrogens N7/N9 and N1/N3, respectively (Figure 19).

With the application of another standard pulse sequence, the 2D HNCO,<sup>[87]</sup> a correlation between the imino atoms and the C2 and C6 carbon atoms in guanine and C2 and C4 atoms in uracil can be achieved. For the magnetization transfer, subsequent INEPT steps exploit the  $^1\text{J}(\text{H},\text{N})$  scalar coupling between the imino proton and the attached nitrogen atom and also the scalar

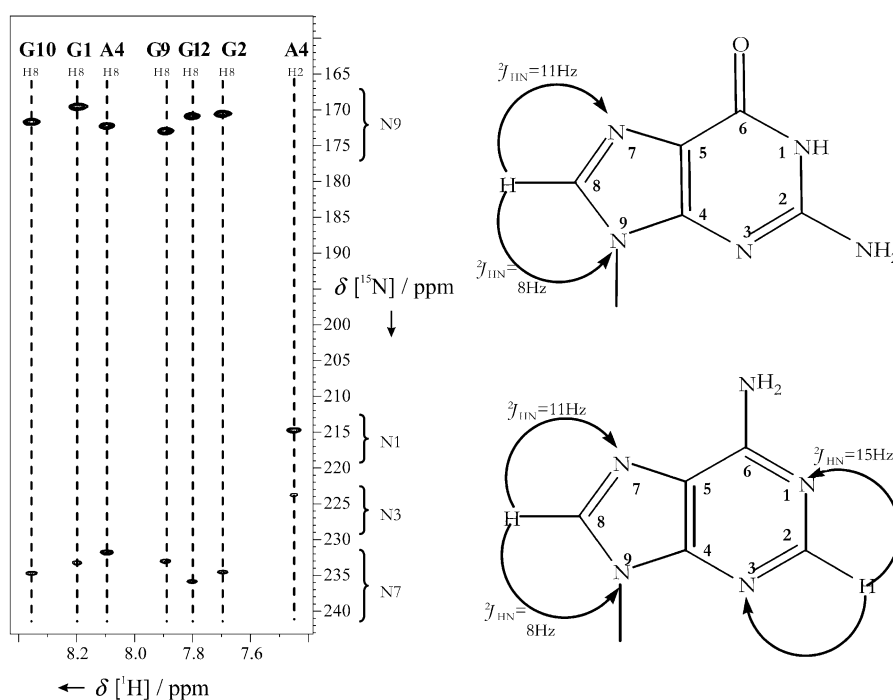
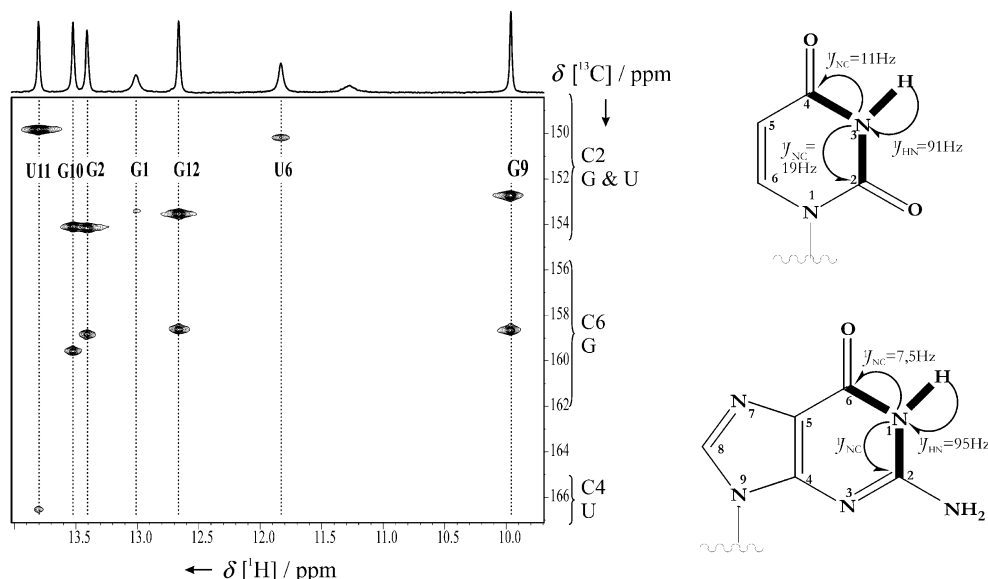


Figure 19. HSQC spectrum at 700 MHz and 298 K with a depiction of the magnetization transfer in adenine and guanine.

coupling  $^1\text{J}(\text{N},\text{C})$  between the nitrogen atom and the attached carbon atom of the carbonyl group (Figure 20). For the correlation of amino protons in cytosine, which are assignable by using an  $^{15}\text{N}$ -HSQC-NOESY experiment, with the carbon atom C4, one applies the same HNCO experiment, the only difference



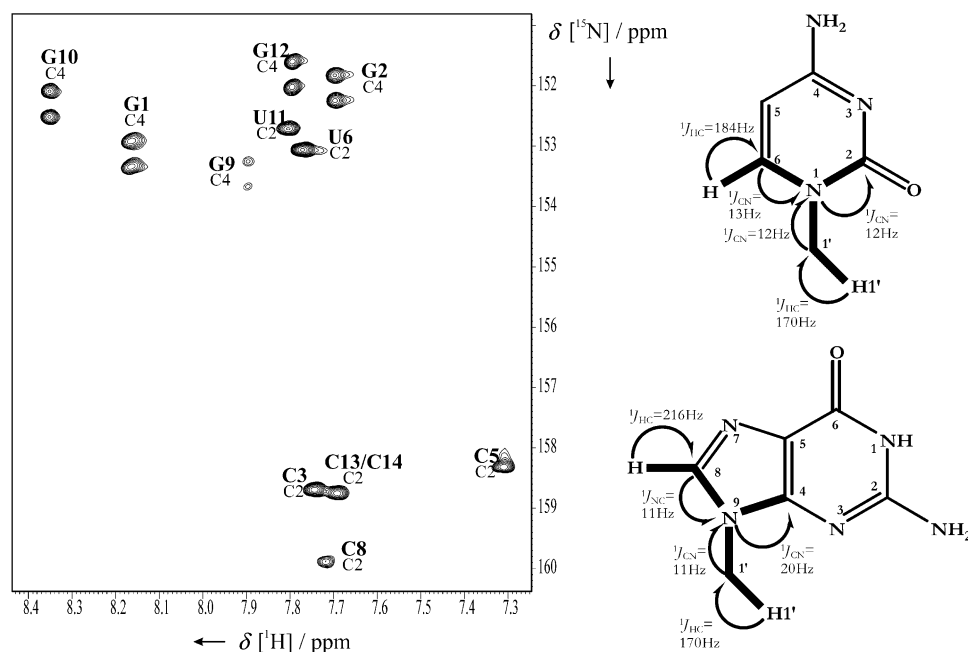


**Figure 20.** 2D H(N)CO spectrum at 700 MHz and 298 K with a depiction of the magnetization transfer in guanine and uracil.

being that the transmitter frequency of nitrogen is set to the resonance frequency of the amino nitrogen atom at 80 ppm.

For the resonance assignment of the remaining quaternary atoms in the nucleobases two new experiments<sup>[6]</sup> have been developed. The first one is the so-called HCNC experiment that correlates the aromatic protons and the sugar proton H1' with the quaternary C4 carbon atom of guanine and C2 of cytosine. The transfer of magnetization is obtained by INEPT steps, which exploit the scalar couplings shown in Figure 21. The advantage of this experiment is that in both cases it starts at protons (either the sugar protons H1' or the aromatic protons H6/H8) that are well resolved and easy to assign.

just a single experiment, the 3D forward-directed HCC-TOCSY-CCH-E.COSY,<sup>[88–90]</sup> to assign all the remaining atoms in the sugar moieties of the RNA molecule. In this experiment, the <sup>1</sup>H and <sup>13</sup>C atoms of the ribose ring are correlated by two sequential transfer steps, a CC-TOCSY and a COSY, by exploiting the large scalar <sup>1</sup>J(C,H) and <sup>1</sup>J(C,C) couplings that depend only to a small degree on the conformation of the nucleotide. This experiment is also advantageous in comparison to the normal HCCH-TOCSY experiments<sup>[91]</sup> because the resolution is much higher. In this optimized 3D experiment, one observes the peaks of the type C*i*H*i*+1' (with *i*=1–4, sugar nomenclature) of every single nucleotide in a frequency plane that is edited with the chemical



**Figure 21.** H6/H8–C6/C8–N1/9–C2/C4 experiment at 700 MHz and 298 K with a depiction of the magnetization transfer during the HCNC experiments.

By using a modified HNCOCOA experiment,<sup>[87]</sup> it is possible to assign the quaternary C5 carbon atom in guanine. The resulting pulse sequence is called HNC6C5, named after the pathway of magnetization. The experiment uses selective INEPT steps to correlate the imino proton with the quaternary carbon atom C5. As depicted in the spectrum (Figure 22), the dispersion of chemical shifts for these atoms is limited, but it is sufficient to obtain a complete resonance assignment of all NMR-active nuclei in the nucleobases.

## 6.5 Complete assignments of sugar atoms

Based on the sequential assignment of the H1' protons, one needs just a single experiment, the 3D forward-directed HCC-TOCSY-CCH-E.COSY,<sup>[88–90]</sup> to assign all the remaining atoms in the sugar moieties of the RNA molecule. In this experiment, the <sup>1</sup>H and <sup>13</sup>C atoms of the ribose ring are correlated by two sequential transfer steps, a CC-TOCSY and a COSY, by exploiting the large scalar <sup>1</sup>J(C,H) and <sup>1</sup>J(C,C) couplings that depend only to a small degree on the conformation of the nucleotide. This experiment is also advantageous in comparison to the normal HCCH-TOCSY experiments<sup>[91]</sup> because the resolution is much higher. In this optimized 3D experiment, one observes the peaks of the type C*i*H*i*+1' (with *i*=1–4, sugar nomenclature) of every single nucleotide in a frequency plane that is edited with the chemical shift of the well-resolved corresponding H1' proton. The resonance assignment is obtained by choosing the distinct <sup>1</sup>H,<sup>13</sup>C frequency planes edited with the corresponding H1' proton chemical shift in the <sup>1</sup>H–<sup>1</sup>H projection. By comparing these planes to the <sup>1</sup>H,<sup>13</sup>C-HSQC spectrum, all signals of one spin system, according to a ribose moiety, can be correlated and assigned (Figure 23).

Sequential assignment of adjacent sugar moieties can be obtained by using the 3D HCP experiment.<sup>[92–94]</sup> In this experiment, H3',C3'<sub>*i*</sub> and H4',C4'<sub>*i*</sub> resonances are correlated with P<sub>*i*+1</sub> on the 3' side and the H5'/H5'',C5'<sub>*i*</sub> and H4',C4'<sub>*i*</sub> resonances are correlated with P<sub>*i*</sub> on the 5' side of the phosphodiester backbone. This allows not only a sequential assign-

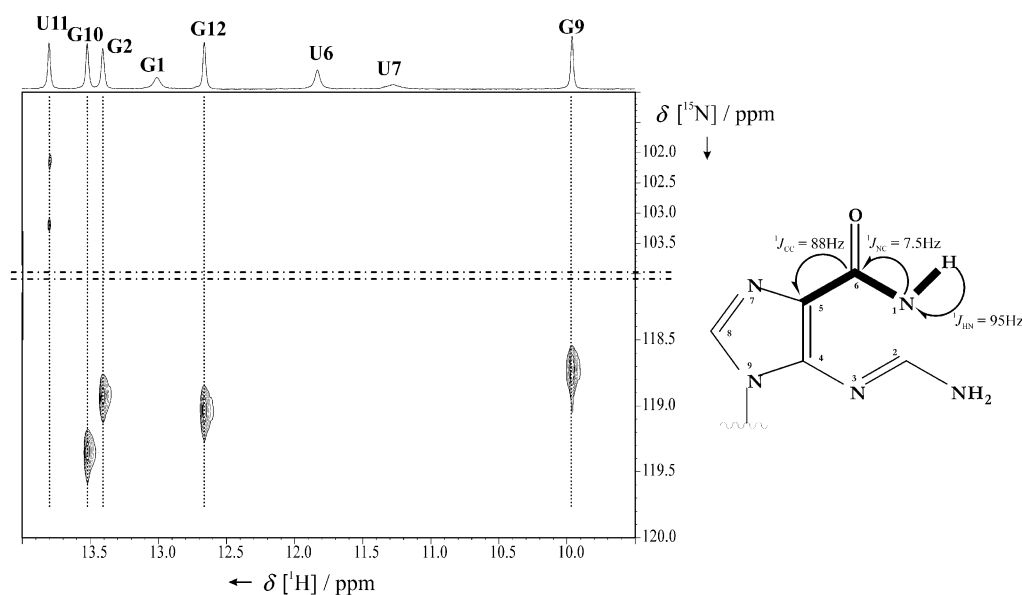


Figure 22. 2D H(NC)C spectrum at 700 MHz and 298 K with schematic presentation of the magnetization transfer in guanine.

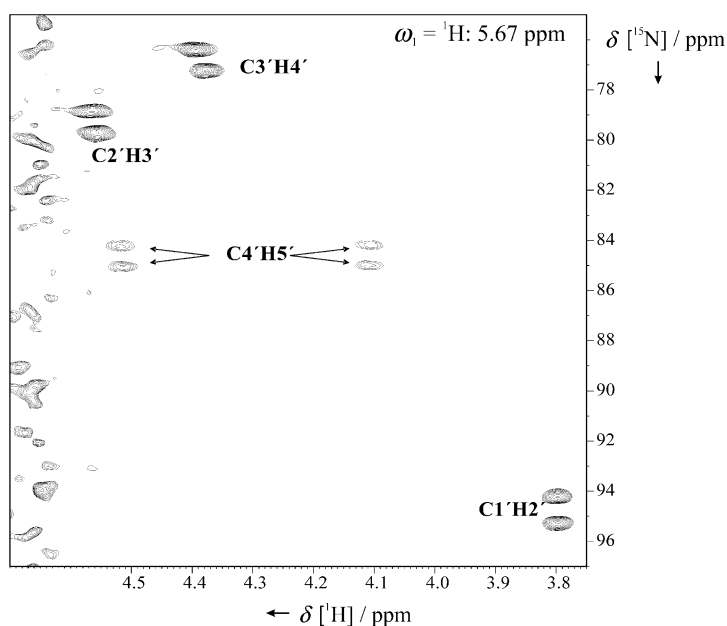


Figure 23.  $^1\text{H}/^{13}\text{C}$  plane at  $\omega_1 = \{^1\text{H}\} = 5.674$  ppm in forward-directed HCC-TOCSY-CCH-COSY experiments at 600 MHz and 298 K for the UUCG tetraloop.

ment of the C4'H4' resonances but also a partial assignment of the resonances in the ribose spin system. The magnetization transfer is similar to those in the CT-HNCO and HCN experiments and follows an out-and-back manner with sequential INEPT steps (Figure 24).

For small RNA molecules, the 3D HCP experiment can be extended by an additional CC-TOCSY transfer step to directly correlate the phosphorous resonances with the well-resolved C1'H1' region of the spectrum.<sup>[95]</sup>

## 7. Conformation Analysis with Chemical Shifts

The chemical shift is a sensitive measure for RNA conformation. Unusual conformations can be detected by comparing the referenced chemical shifts of the examined RNA with the data deposited in the BMRB database. Such a procedure can be applied to all NMR-active nuclei. The dependence of  $^1\text{H}$  chemical shifts on secondary structure has been analyzed in detail for a number of different RNA molecules including the cUUCGg tetraloop<sup>[96]</sup> and shall not be discussed here.

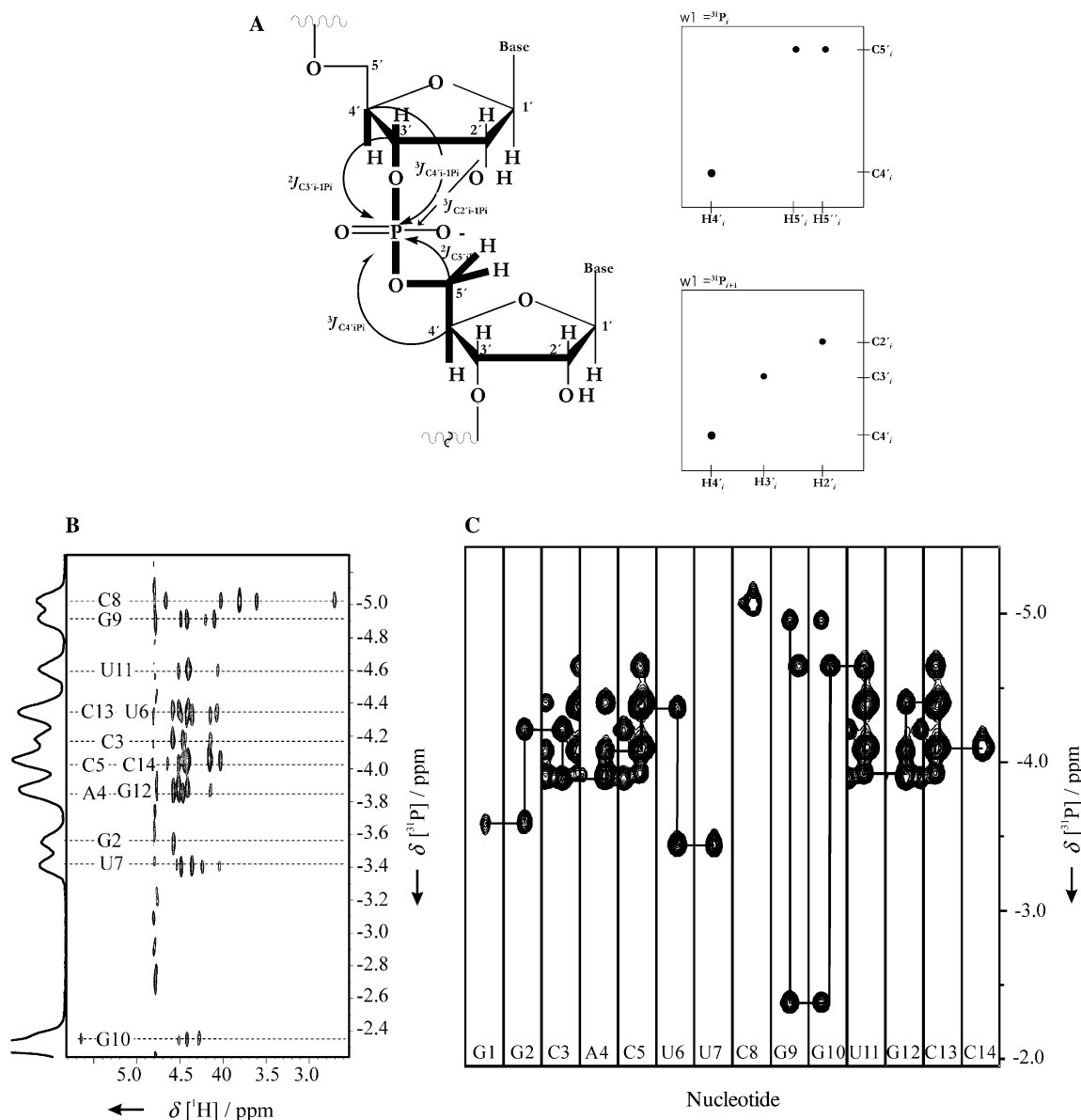
$^{31}\text{P}$  chemical shifts and their deviations from standard chemical shifts in A-form RNA are often taken to be indicative for unusual conformations around the phosphodiester backbone. It is possible to calculate the deviations in the  $^{31}\text{P}$  chemical shift from the mean value of the assigned RNA molecules deposited in the BMRB database. In the cUUCGg tetraloop, significant variations for the phosphorous chemical shift are observed for the loop residues U7, C8, G9, and the nucleotide G10 that is part of the closing base pair (Figure 25C). The chemical shifts of the residues in canonical A-form conformation are not significantly different from the mean value.

$^{13}\text{C}$  chemical-shift data yielding the sugar pucker modes could be obtained following the sophisticated analysis of solid-state NMR data by Ebrahimi et al.<sup>[97]</sup> For this analysis, canonical coordinates were calculated by using the chemical-shift data ( $\delta$ ) according to Equations 1 and 2.

$$\text{can1} = 0.179\delta_{\text{C1}'} - 0.225\delta_{\text{C4}'} - 0.0585\delta_{\text{C5}'} \quad (1)$$

$$\text{can2} = -0.0605(\delta_{\text{C2}'} + \delta_{\text{C3}'} - 0.0556\delta_{\text{C4}'} - 0.0524\delta_{\text{C5}'} \quad (2)$$

The first canonical coordinate *can1* describes the pseudorotation phase of the sugar moiety. For *can1* > -6.25 ppm, the

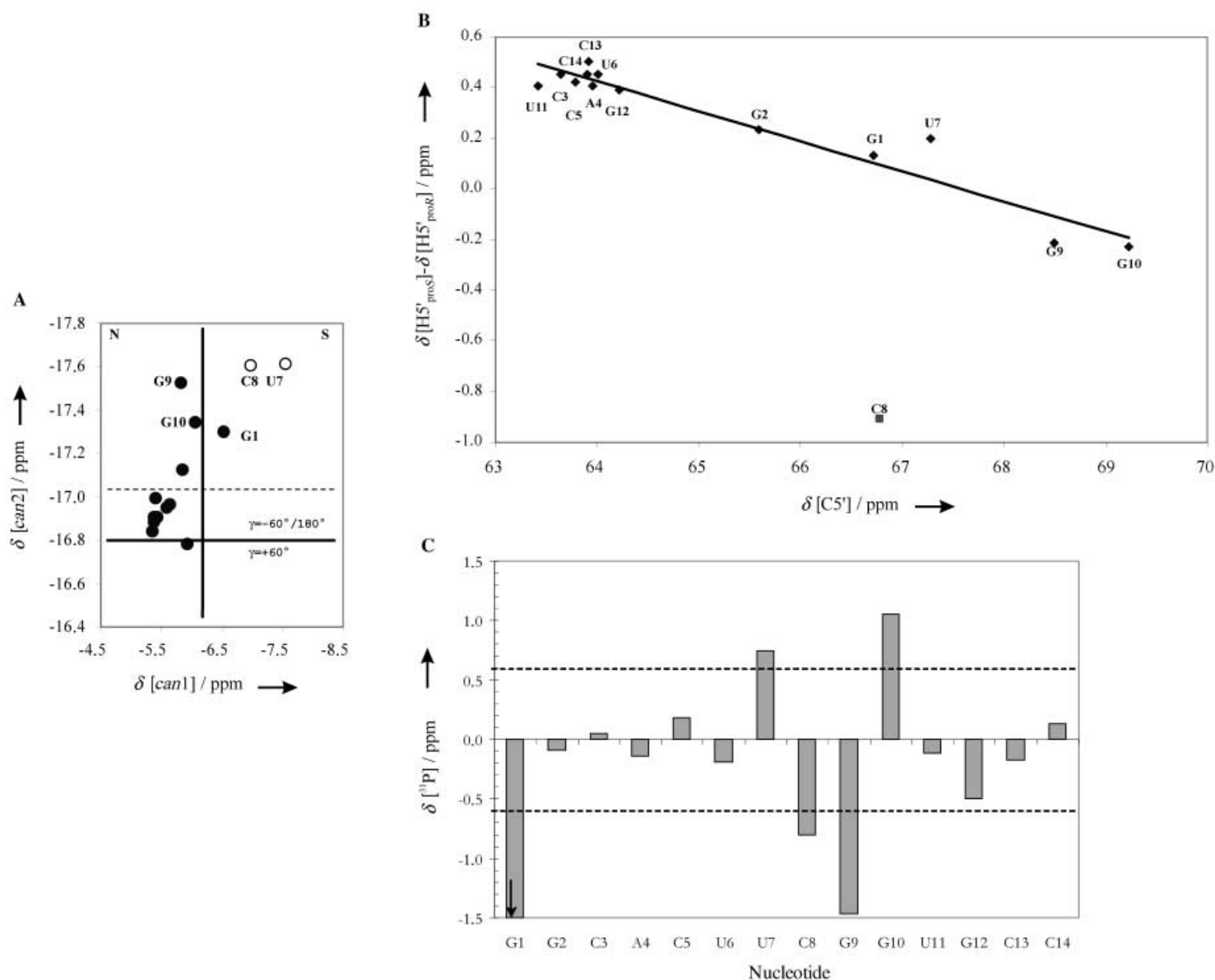


**Figure 24.** A) Representation of the HCP experiment on the left. The correlated atoms in the experiment are indicated through the exploited scalar coupling represented by arrows. The right-hand side shows the expected 2D  $^1H$ - $^{13}C$  plane of the 3D experiment at different phosphorous chemical shifts. The  $C4'H4'$  resonance of one ribose unit is correlated with two phosphorous resonances,  $P_i$  and  $P_{i+1}$ . The  $C2'H2'$  and the  $C3'H3'$  are correlated with  $P_{i+1}$  and  $C5'H5'/H5''$  with  $P_i$ ; the  $C4'H4'$  unit shows cross-peaks to both  $P_i$  and  $P_{i+1}$ . B)  $^1H$ - $^{31}P$  projection of the 3D HCP spectrum of the cUUCG tetraloop. C) The 3D HCP experiment in form of a strip plot. The sequential assignment pathway is indicated.

sugar is in a C3'-endo conformation, for  $can1 < -6.25$  ppm, the sugar adopts a C2'-endo conformation. The second coordinate  $can2$  determines the conformation of the exocyclic torsion angle  $\gamma$ . If the torsion angle  $\gamma$  is in a gauche-gauche conformation ( $\gamma = +60^\circ$ ),  $can2 \geq -16.8$  ppm. Population of either of the two gauche-trans conformations ( $\gamma = -60^\circ, 180^\circ$ ) results in a coordinate  $can2$  smaller than  $-16.8$  ppm. Application of these rules to the chemical-shift data of the cUUCG tetraloop allows clear discrimination between residues with different sugar pucker modes (Figure 25A). According to this analysis, residues C8 and U7 adopt a C2'-endo conformation; while all other residues are in C3'-endo conformation. The coordinate analysis fails for residue G1, maybe due to a higher conformational

flexibility at the 5' end of the stem or due to the additional charge that alters the chemical shifts. These results are in good agreement with the published structures and are also in very good agreement with the analysis of scalar  $^3J(H,H)$  coupling constants and cross-correlated relaxation rates in the ribose ring.

Discrimination between gg and gt conformations and the exocyclic torsion angle  $\gamma$  on the basis of  $can2$  is less convincing. Most of the values of  $can2$  are clustered in the region between  $-16.6$  and  $-17.0$  ppm. Only the nucleotides G1, U7, C8, G9, and G10 are clearly in the gg conformation. The uncertainty in the analysis could be due to the simultaneous use of the chemical shifts of C2' and C3' in the calculation. For the tetraloop, the



**Figure 25.** Chemical-shift analysis for the determination of: A) the sugar pucker mode and the exocyclic torsion angle  $\gamma$  (from the work of Ebrahimi et al.<sup>[97]</sup>), B) the stereospecific assignment of the H5' and H5'' protons in RNA (from the work of Marino et al.<sup>[89]</sup>), and C) the conformation of the phosphodiester backbone (from the work of Fürtig et al.<sup>[6]</sup>).

analysis would predict residues with  $\text{can2} < -17$  ppm to be in a gg conformation around the exocyclic torsion angles  $\gamma$ .

By analyzing the chemical-shift data of the C5'H5'/C5'H5'' resonances, a stereospecific assignment of the prochiral H5'/H5'' protons was obtained. It is observed that in helical RNA structures, the resonance of the H5'(pro-S) proton is shifted up-field with regard to the resonance of the H5'(pro-R) proton.<sup>[98]</sup> This rule is unfortunately not applicable to noncanonical regions of RNA structures. However, a stereospecific assignment is possible by correlation of the difference in the proton chemical shifts of the pro-S and pro-R protons with the respective carbon chemical shift,  $\Delta\delta[\text{H5}(\text{pro-S}) - \text{H5}(\text{pro-R})](\delta^{13}\text{C})$ .<sup>[89]</sup> If one depicts the differences of the proton chemical shifts versus the carbon chemical shift, the stereochemical assignment is revealed (Fig-

ure 25B). A general *anti* correlation could be observed for the carbon chemical shift of these resonances. An exception is residue C8; this is probably the result of the different conformation of  $\gamma$  at this loop position. These results are in good agreement with the data obtained from the analysis of the  $^2J(\text{C4}', \text{H5}'/\text{H5}'')$  and  $^3J(\text{H4}', \text{H5}'/\text{H5}'')$  coupling constants.

## 8. Delineation of Secondary Structure Motifs such as Hairpins and Bulges

Many RNA structures are built up in a modular fashion from smaller structural elements that fold autonomously and are stable on their own. Due to their noncanonical structure, some of these structural elements give rise to resonances in the NMR

spectra with unusual chemical shifts outside of the standard chemical-shift ranges. Furthermore these chemical shifts are often similar when such a modular element is present in different contexts in different RNA molecules. A classical example of this is the UNGC family of stable tetraloops. The tight fold of the loop with an unusual U:G base pair, a guanine residue in the *syn* conformation, and a cross-strand stacking interaction between the first and the third loop-nucleotides<sup>[8]</sup> results in a number of NMR signals far outside the usual chemical-shift ranges<sup>[59]</sup> (Figure 26 A). These chemical-shift patterns do not change when the loop is inserted into other RNA molecules (Figure 26 B) and due to their good separation from the bulk of the other signals they can still be recognized and the presence of a UNGC loop element can be verified. A similar fingerprint of unusually shifted resonances can be found for another family of stable tetraloops, the GNRA tetraloops.<sup>[99–101]</sup> A further example of a structural motif giving rise to a specific chemical-shift pattern might be provided by the loop E-motif. The loop E-motif<sup>[102]</sup> occurs in a variety of internal bulges in different RNA molecules such as 5S rRNA,<sup>[103, 104]</sup> the ribosomal sarcin loop,<sup>[100, 101]</sup> the hairpin ribozyme domain B,<sup>[105]</sup> and viroids.<sup>[106]</sup> However, with the availability of more completely assigned and structurally character-

ized RNA molecules, the connection between chemical-shift patterns and conserved structural elements will become clearer and more motifs might be found.

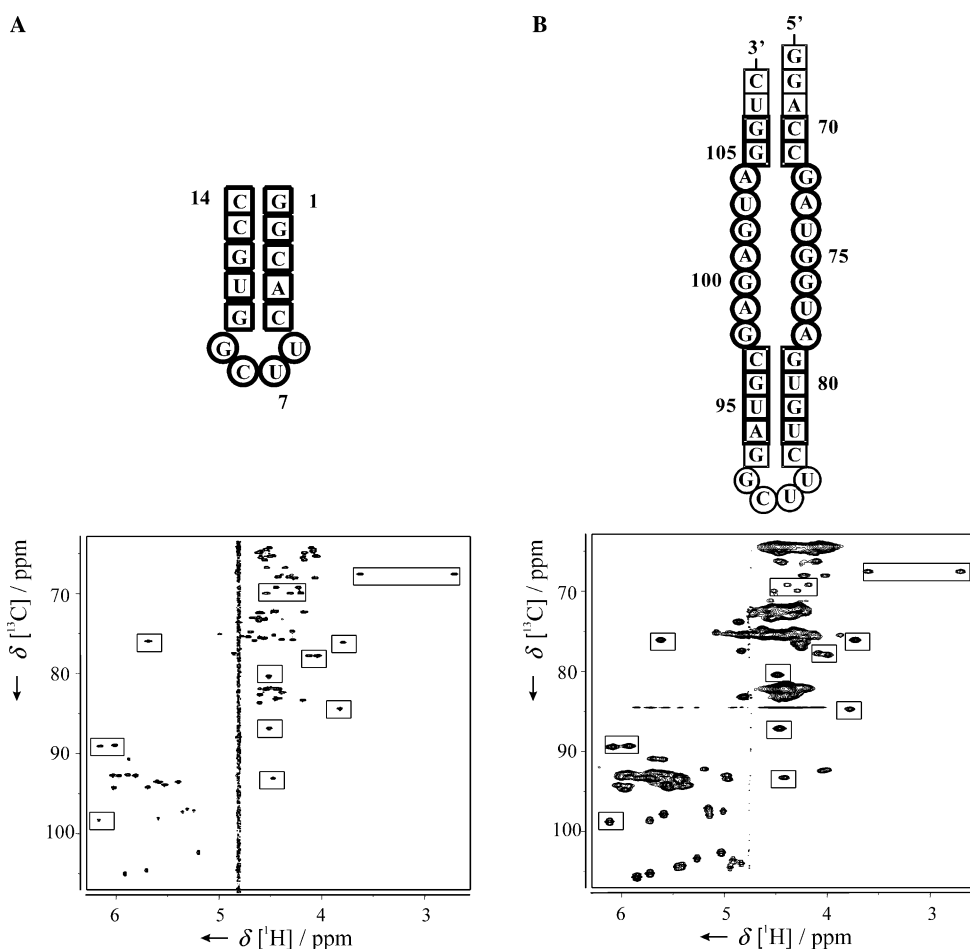
## 9. Determination of Local Conformation and Conformational Dynamics in RNA

The conformation of the phosphodiester backbone and the ribosyl furanose plays a dominant role in preorganizing the overall structure of a nucleic acid. Therefore, the properties and interaction possibilities of an RNA molecule depend on the conformation and the dynamics of the ribose ring and the phosphodiester backbone.

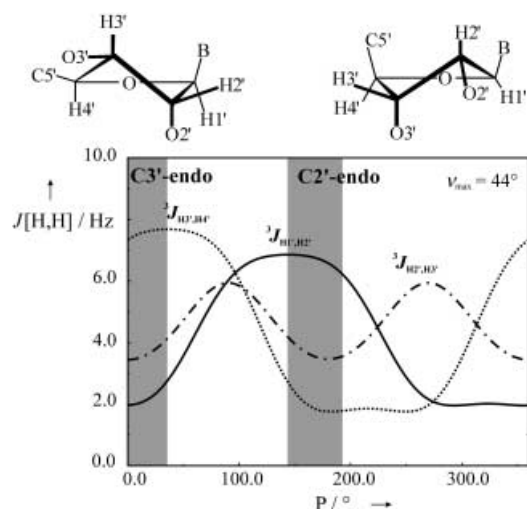
The determination of the pseudorotation phase  $P$  and pseudorotation amplitude  $\nu_{\max}$  of the ribose moiety is based on the interpretation of homonuclear  $^3J(\text{H},\text{H})$  coupling constants. The sugar pucker in canonical A-form RNA is C3'-endo. In this conformation, the orientation of H1' to the H2' proton is synclinal and one observes a small  $^3J(\text{H}1',\text{H}2')$  coupling constant. The H3' and H4' protons are in an antiperiplanar orientation and one observes a large  $^3J(\text{H}3',\text{H}4')$  coupling constant. In contrast, small  $^3J(\text{H}3',\text{H}4')$  and large  $^3J(\text{H}1',\text{H}2')$  coupling constants are the

signature for a ribose sugar moiety in C2'-endo conformation. The  $^3J(\text{H}2',\text{H}3')$  coupling constant cannot be used for differentiation because in both conformations its value is around 4 Hz. The dependence of the  $^3J(\text{H},\text{H})$  coupling constants on the sugar conformation is shown in Figure 27.

The homonuclear  $^3J(\text{H},\text{H})$  coupling constants for two RNA hairpins (the cUUUUg loop and cUUCGg loop) have been determined in the forward-directed HCC-TOCSY-CCH-E.COSY experiment (Figure 28),<sup>[88, 90]</sup> and Table 5 provides a summary of the extracted coupling constants. The pseudorotation phase  $P$  and amplitude  $\nu_{\max}$  can be calculated by using the parameterization of Haasnoot et al.<sup>[107–109]</sup> The results (see Table 5) for the cUUUUg loop show that U4 is in a C3'-endo conformation whereas the ribose rings of U5 and U7 adopt C2'-endo conformations. The coupling constants of U6 cannot be explained with one static conformation of the ribose ring. Therefore, the sugar moiety of this nucleotide must be flexible and can be modeled under the as-



**Figure 26.** Identification of a UUGC stable tetraloop motif in two different RNA molecules by the chemical-shift signature. A)  $^1\text{H},^{13}\text{C}$ -HSQC spectrum of a small hairpin RNA closed by a stable tetraloop with resonances exhibiting unusual chemical shifts due to the tetraloop fold indicated in boxes. B)  $^1\text{H},^{13}\text{C}$ -HSQC spectrum of a larger RNA containing a similar tetraloop. Again the resonances belonging to the loop are marked in boxes.



**Figure 27.** Karplus relation of  $^3J(\text{H1}',\text{H2}')$ ,  $^3J(\text{H2}',\text{H3}')$ , and  $^3J(\text{H3}',\text{H4}')$  coupling constants depending on the pseudorotation phase  $P$  at a pseudorotation amplitude  $\nu_{\text{max}}$  of  $44^\circ$ .

sumption of an equilibrium between the C2'- (37%) and C3'-endo (63%) conformations. This result is consistent with those obtained when measuring a larger set of heteronuclear  $^nJ(\text{C,H})$  coupling constants<sup>[110]</sup> or cross-correlated relaxation rates.<sup>[111]</sup>

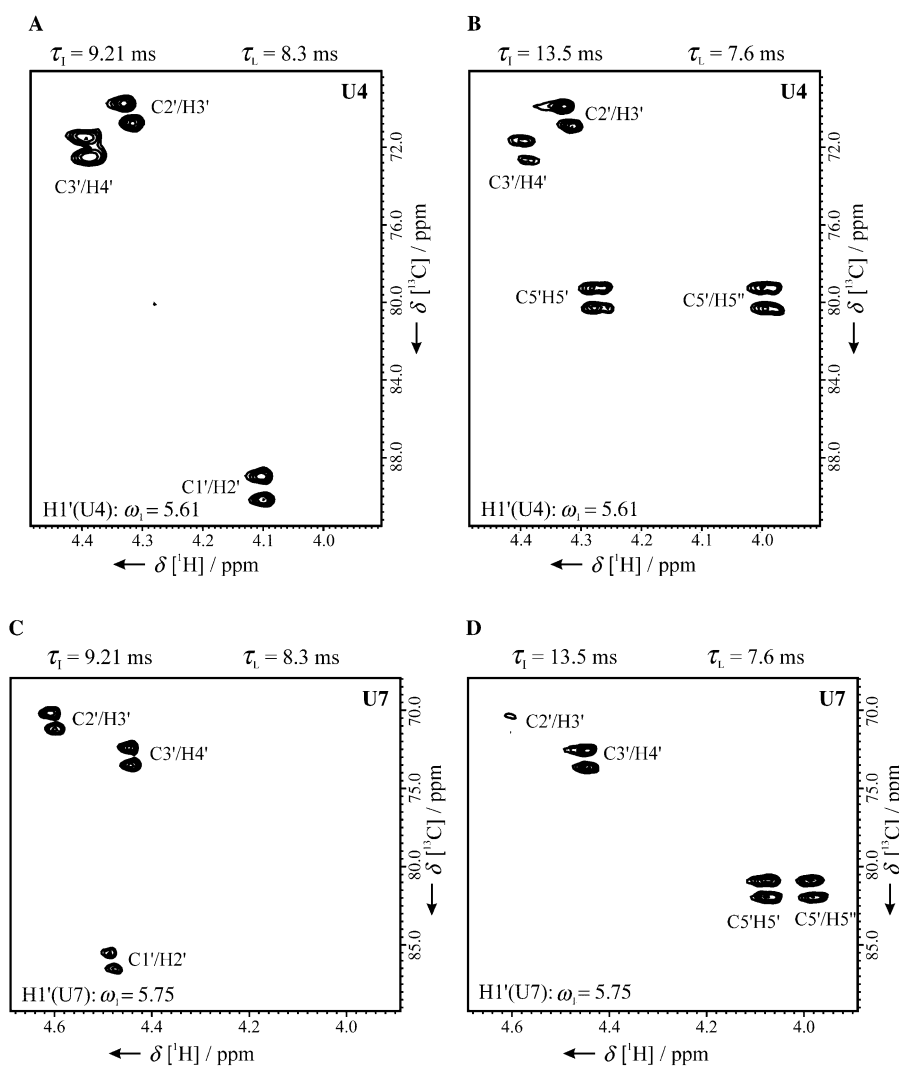
While the cUUUUg loop shows conformational flexibility in at least one of the ribose moieties of the loop nucleotides, the cUUCGg-loop nucleotides adopt more rigid sugar conformations. The ribose rings of U6 and G9 adopt canonical A-form conformations whereas the ribose rings of U7 and C8 adopt B-form conformations. The scalar coupling data (Figure 29) indicate that in the case of the cUUCGg loop, the closing base G10 might adopt an unusual sugar conformation. If one interprets the values of the coupling constants as mean values of two differently populated canonical conformations, the values must be based on an equilibrium between 67% in A- and 33% in B-form conformations for nucleotide G10.

Measurement of  $^3J(\text{H,H})$  coupling constants provides a valuable tool for describing the conformation, including the dynamic equilibria, of the sugar rings that determines the overall

conformation. A further discussion of available NMR parameters for the determination of the phosphodiester backbone has been provided previously (by Schwalbe et al.<sup>[112]</sup> and by Griesinger et al.<sup>[113]</sup>). Table 6 summarizes the NMR parameters available for the determination of dihedral angles in RNA: homo- and heteronuclear vicinal coupling constants ( $^3J$ ) and cross-correlated relaxation rates ( $\Gamma$  rates).

## 10. RNA Global Structure: Residual Dipolar Couplings

A notorious problem in the structure determination of RNA based on NOE contacts and scalar couplings is the absence of long-range NOE contacts due to the often elongated rodlike helical structures of RNA. The short-range nature of the NOE information does not allow a precise definition of the relative spatial orientation of distant parts of the molecule. Therefore, it is difficult to extract information about helical bend and end-to-end orientation or interhelical angles for different helical parts of a molecule. Here residual dipolar couplings (RDCs) present a new source of structural information that can be used to overcome these obstacles. RDCs contain global structural information since they report on the orientation of individual bond vectors with regard to the axes of the alignment tensor of the molecule. They

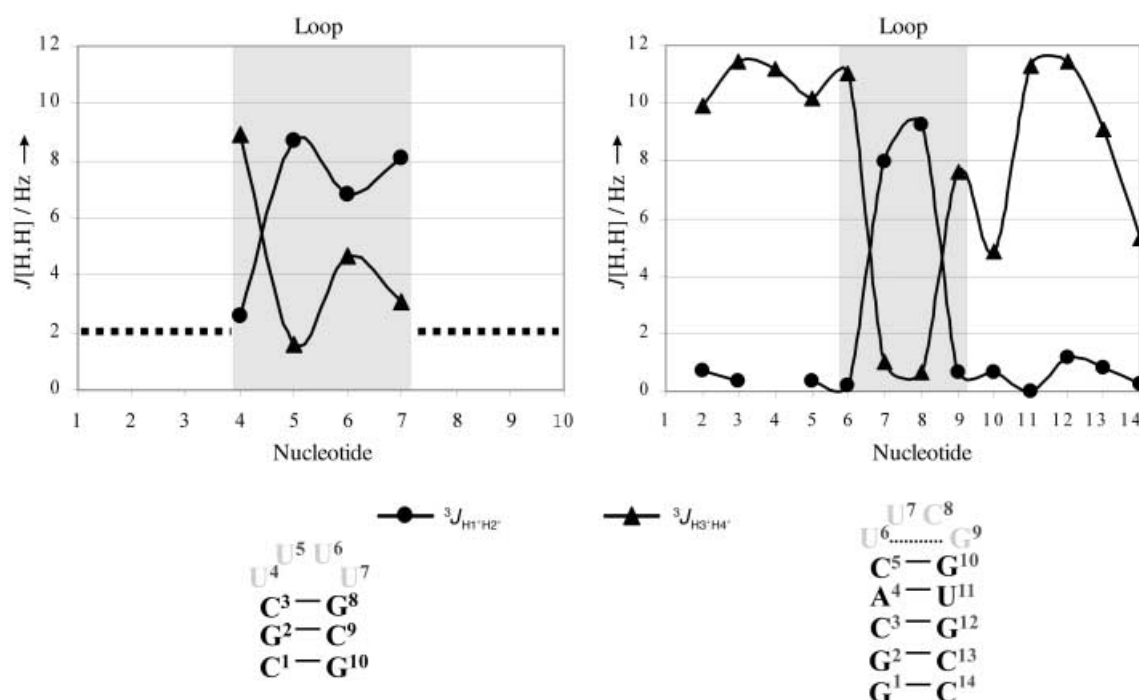


**Figure 28.** Forward-directed HCC-TOCSY-E.COSY spectra for 1.5 mM UUUU tetraloop RNA. A, C) Correlation from C1'–C3'. B, D) Correlation from C4'–C5'.

**Table 5.** Experimental homonuclear proton coupling constants for the cUUUUG and cUUCGg tetraloops extracted from forward-directed HCC-TOCSY-CCH-E.COSY spectra.

	U4	U5	U6	U7	U6	U7	C8	G9
$^3J(\text{H1}',\text{H2}')^{[a]}$	$2.6 \pm 0.3$	$8.7 \pm 0.1$	$6.8 \pm 0.1$	8.1	0.2	8.0	9.3	0.6
$^3J(\text{H2}',\text{H3}')^{[a]}$	$5.2 \pm 0.3$	$5.5 \pm 0.1$	$5.4 \pm 0.1$	$5.4 \pm 0.3$	5.5	5.2	4.9	3.4
$^3J(\text{H3}',\text{H4}')^{[a]}$	$8.9 \pm 0.2$	$1.6 \pm 0.2$	$4.7 \pm 0.1$	$3.1 \pm 0.2$	11.0	1.0	0.6	7.6
$\text{rms}J^{[b]}$	0.42	0.42	1.07	0.24	1.3	0.2	0.7	0.6
Pseudorotation amplitude <sup>[c]</sup> ( $\nu_{\text{max}}$ )	44	43	42	44	43	38	44	44
Pseudorotation phase <sup>[c]</sup> (P)	44	144	123	134	29	160	155	16

[a] Average coupling values; determination of the coupling constant at two different submultiplets. [b]  $\text{rms}J = (\sum (J_{\text{theo}} - J_{\text{exp}})^2)^{-1/2} / (n)^{-1/2}$  measured in Hz, where  $n$  is the number of couplings. [c] Prediction with the parameters described in the work of Haasnoot et al.<sup>[107–109]</sup>

**Figure 29.** Depiction of the scalar couplings  $^3J(\text{H1}'\text{H2}')$  and  $^3J(\text{H3}'\text{H4}')$  versus the nucleotide sequence. In the lower part of the figure, the corresponding secondary structures of the molecules are depicted. Grey numbers of the nucleotide indicate that these nucleotides are  $^{13}\text{C}$ - and  $^{15}\text{N}$ -labeled. Loop nucleotides are depicted in grey like the loop region of the plots.**Table 6.** Experimental parameters available for the determination of the local conformation ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ ,  $\chi$ ,  $P$ , and  $\nu_{\text{max}}$ ) of RNA molecules.

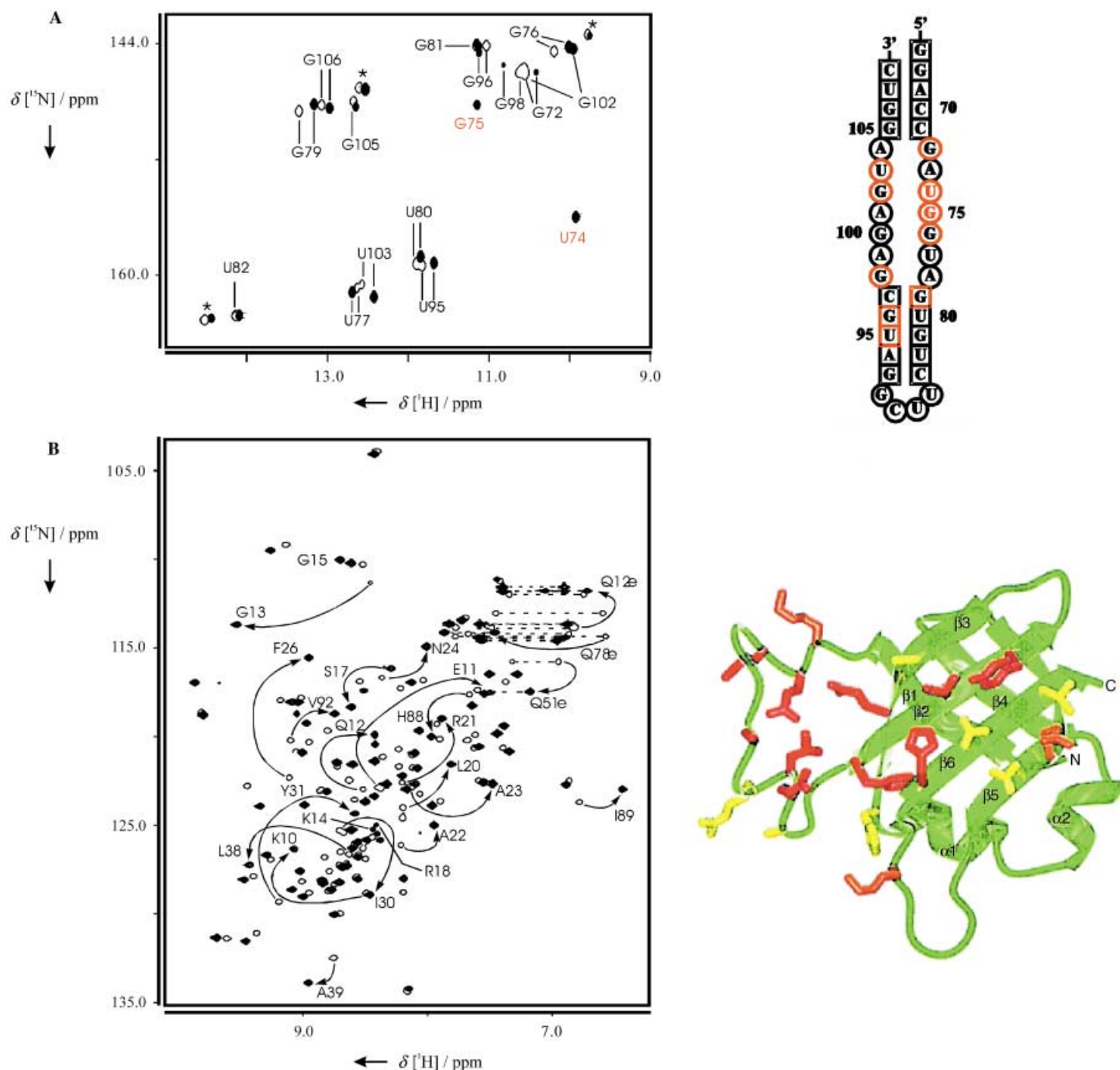
Structural parameter	NMR spectroscopy parameter			Ref.
	Cross-correlated relaxation rates	Homonuclear $^3J$ coupling constants	Heteronuclear $^3J$ coupling constants	
backbone:				
$\alpha$	$\Gamma_{\text{C1H1},\text{P1}}^{\text{DD,CSA}}$			[141]
$\beta$	$\Gamma_{\text{HS1},\text{CS1},\text{HS1},\text{P1}}^{\text{DD,DD}}$			[142–147]
$\gamma$		$^3J(\text{H4}',\text{H5}'), ^3J(\text{H4}',\text{H5}'')$	$^3J(\text{H5}',\text{P}), ^3J(\text{H5}'',\text{P}), ^3J(\text{C4}',\text{P})$	[88, 89, 110, 148]
$\delta$		$^3J(\text{H3}',\text{H4}')$	$^3J(\text{C3}',\text{H5}'), ^3J(\text{C3}',\text{H5}'')$	[88, 89, 110]
$\epsilon$			$^3J(\text{C5}',\text{H3}'), ^3J(\text{C2}',\text{H4}')$	[142–147, 149, 150]
$\zeta$	$\Gamma_{\text{H3},\text{C3},\text{H3},\text{P1+1}}^{\text{DD,DD}}$		$^3J(\text{H3}',\text{P}), ^3J(\text{C2}',\text{P}), ^3J(\text{C4}',\text{P})$	[141]
glycosidic bond: $\chi$	$\Gamma_{\text{C3},\text{H3},\text{P1+1}}^{\text{DD,CSA}}$			[143, 151, 152]
ribose moiety: $\nu_{\text{max}}, P$	$\Gamma_{\text{C1H1},\text{N1}}^{\text{DD,CSA}}, \Gamma_{\text{C2H2},\text{N}}^{\text{DD,CSA}}, \Gamma_{\text{C3H3},\text{C4H4}}^{\text{DD,DD}}, \Gamma_{\text{C3H3},\text{C4H4}}^{\text{CSA,CSA}}, \Gamma_{\text{NH},\text{H}}^{\text{DD,CSA}}$	$^3J(\text{H1}',\text{H2}'), ^3J(\text{H3}',\text{H4}')$	$^3J(\text{C6}/\text{C8},\text{H1}'), ^3J(\text{C4}/\text{C2},\text{H1}')$	[88, 110, 111, 153]
hydrogen bonding			$^hJ(\text{N},\text{N}), ^hJ(\text{H},\text{N})$	[32, 154, 155]

are caused by the presence of an aligning medium that interferes with the isotropic tumbling of a molecule and induces a certain degree of alignment of the molecule with respect to

the magnetic field. Alignment media that have been used for nucleic acids are filamentous phages<sup>[114]</sup> introduced by Pardi and co-workers, *n*-alkyl-polyethylene glycol/*n*-alkyl-alcohol bi-

celles,<sup>[115, 116]</sup> and the magnetic field itself.<sup>[117, 118]</sup> The introduction of orientation restraints derived from residual dipolar couplings leads to a higher precision of the derived structures in standard structure determinations (for examples, see refs. [119–122]), especially in loop regions where NOE contacts tend to be more scarce.<sup>[116]</sup> A further improvement of structure determinations

can be expected from the measurement of more exotic dipolar couplings such as PH couplings between protons in the bases and the phosphorus nuclei in the RNA backbone,<sup>[123]</sup> since they report on the conformation of the backbone, which is notoriously underdefined in classical NMR structure determinations due to the low proton density along the backbone. In addition,



**Figure 30.** Identification of the interaction surfaces in the complex of the *E. coli* 5S rRNA E-loop and ribosomal protein L25 by chemical-shift mapping (from the work of Stoldt et al.<sup>[132]</sup>). A) Left:  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of 5S rRNA in the absence (open contours) and the presence (filled contours) of rL25 with the assignments indicated. Residues labeled in red are only observable in the bound state due to protection from fast exchange with the solvent in the presence of the protein. Right: Mapping of residues with substantial chemical-shift changes (red circles) or protection from solvent exchange (red squares) onto the secondary structure of 5S rRNA to reveal a bipartite protein binding site. B) Left:  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC spectrum of rL25 in the absence (open contours) and the presence (filled contours) of 5S rRNA with the assignments indicated. Right: Mapping of residues with substantial chemical-shift changes onto the structure of rL25 in the RNA-free form (from the work of Stoldt et al.<sup>[132]</sup>).



$^1\text{H}$ ,  $^1\text{H}$  residual dipolar couplings<sup>[124]</sup> can potentially be used as sources for long-range distances and for orientation information.

However, the full potential of RDCs as constraints is shown when in the definition of global structural features such as interhelical angles, for example, in the complex of the U1A protein with its target RNA,<sup>[125]</sup> the theophylline aptamer complex,<sup>[126]</sup> or the domain orientation in tRNA molecules.<sup>[127]</sup> The example of tRNA already shows that it is possible to derive structural information for larger RNA molecules that are normally not amenable to a full structural characterization by NMR spectroscopy. In a study of the hammerhead ribozyme<sup>[128]</sup> it was shown that the relative orientations of the three helical parts in the absence of magnesium in solution is totally different to the crystal structure and major interhelical reorientations have to occur before the ribozyme folds into its catalytically active conformation. Furthermore, it is possible to align ligands and their target RNA with respect to each other without the need for a full structure determination, based on RDC information.<sup>[129]</sup> In addition, Patel's group used residual dipolar couplings to obtain information on RNA dynamics in the case of HIV1-Tar RNA with respect to interhelical motions;<sup>[130]</sup> this indicated that the free RNA was already able to sample conformations of the ligand-bound state. RDCs also appear to have the potential to speed up RNA structure determination in general in structural genomics approaches.<sup>[131]</sup>

## 11. Mapping of Interaction Surfaces of RNA Molecules

Many RNA molecules carry out their function in complex with ligands, such as proteins, other RNA molecules, or small molecules or metal ions. It is often of interest to determine the functional groups of an RNA molecule that interact with these ligands for guiding mutational studies, for interpreting phylogenetic data such as sequence conservation, or for other biological questions. Sometimes the biophysical properties of RNA–ligand complexes are detrimental to solving a complete structure of the complex or even assigning the RNA in its bound state due to, for example, a large size, a limited solubility, or unfavorable kinetic behavior of the complex. However, even in such cases it is possible to obtain information on RNA groups involved in binding. The most straightforward method is the observation of chemical-shift changes of the RNA upon gradual addition of the ligand—a method called chemical-shift mapping. Probably the first application of this method to RNA was reported by Kime and Moore in 1983<sup>[31]</sup> in an investigation of the binding of the ribosomal protein L25 to 5S rRNA from *E. coli*. In their pioneering work, they used changes in the 1D  $^1\text{H}$  spectra of the imino proton region of 5S rRNA upon addition of rL25 to identify nucleotides in the E-loop region of 5S rRNA as the possible protein binding site. It then took sixteen more years to characterize the binding of rL25 to the 5S rRNA E-loop in atomic detail, both by NMR spectroscopy<sup>[132]</sup> and X-ray crystallography.<sup>[133]</sup> The example of the  $\approx 120$  nucleotide 5S rRNA indicates that chemical-shift mapping can be used even for larger RNA molecules that are beyond a complete structural characterization with current NMR methods. Ligand binding often does

not only induce chemical-shift changes but allows the observation of imino proton resonances that are not observable or are broadened in the free RNA due to fast exchange with the solvent (Figure 30); this occurs either by stabilization of the RNA structure (for example, see ref. [134]) or by protection of the imino protons from exchange (for example, see refs. [135, 136]) by direct RNA–protein contacts. Chemical-shift mapping is also applicable in cases when RNA–protein complexes of lower affinity are formed and the components are in fast or intermediate exchange on the NMR timescale (for example, see ref. [137]).

An alternative to the chemical-shift mapping approach is to use cross-saturation experiments.<sup>[138]</sup> In these experiments, resonances of one component of a biomolecular complex are selectively saturated. By spin diffusion, this saturation is transferred to the second component of the complex, thereby leading to a loss of the signal intensities of its NMR spectra resonances. The largest effects are observed for residues close to the interface of the complex, which allows the identification of residues that are part of the interaction surface. In RNA–protein complexes one can take advantage of the fact that there are numerous spectral windows where protein and RNA resonances do not overlap, for example, the region of the RNA imino protons, the RNA H1' protons, and the region of the aliphatic resonances of the protein upfield of 2 ppm. This makes the selective irradiation of one of the binding partners especially straightforward.<sup>[139]</sup> In addition, since the detection is mostly done by using highly sensitive HSQC spectroscopy this method appears to be applicable to larger complexes.<sup>[140]</sup>

## Acknowledgements

The authors would like to thank Matthias Görlach, Matthias Stoldt, Christian Sich, Oliver Ohlenschläger, Ramadurai Ramachandran, Christian Griesinger, Elke Duchardt, Wolfgang Bermel, John Marino, Mirko Hennig, Steffen Glaser, and Bernd Reif for many interesting discussions about all aspects of RNA structure, biochemistry, and NMR spectroscopy. We are grateful to Matthias Stoldt, Matthias Görlach, and Elke Duchardt for help with figures and for sharing data prior to publication. We are grateful to Aphrodite Anastasiadis-Pool and Emily S. Collins for reading the manuscript. The work has been supported by the state of Hesse, the European Large-Scale Facility, the Sonderforschungsbereich 579 "RNA–Ligand Interactions", and the Fonds der Chemischen Industrie.

**Keywords:** conformation analysis • isotope labeling • NMR spectroscopy • resonance assignment • RNA

- [1] G. Varani, F. Aboul-ela, F. H.-T. Allain, *Progress in NMR spectroscopy* **1996**, 29, 51–127.
- [2] S. S. Wijmenga, B. N. M. van Buuren, *Progress in NMR spectroscopy* **1998**, 32, 287–387.
- [3] R. T. Batey, M. Inada, E. Kujawinski, J. D. Puglisi, J. R. Williamson, *Nucleic Acids Res.* **1992**, 20, 4515–4523.
- [4] E. P. Nikonowicz, A. Sirr, P. Legault, F. M. Jucker, L. M. Baer, A. Pardi, *Nucleic Acids Res.* **1992**, 20, 4507–4513.
- [5] P. Büchner, W. Maurer, H. H. Rüterjans, *J. Magn. Reson.* **1978**, 29, 45–63.

- [6] B. Fürtig, C. Richter, W. Bermel, H. Schwalbe, *J. Biomol. NMR* **2003**, in press.
- [7] E. Ennifar, A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, P. Dumas, *J. Mol. Biol.* **2000**, *304*, 35–42.
- [8] F. H.-T. Allain, G. Varani, *J. Mol. Biol.* **1995**, *250*, 333–353.
- [9] D. J. Proctor, J. E. Schaak, J. M. Bevilacqua, C. J. Falzone, P. C. Bevilacqua, *Biochemistry* **2002**, *41*, 12062–12075.
- [10] C. Kojma, M. A. Ono, A. Ono, M. Kainosho, *Meth. Enzymol.* **2001**, *338*, 261–282.
- [11] I. M. Lagoja, P. Herdewijn, *Synthesis* **2002**, *3*, 301–314.
- [12] S. S. P. Quant, R. W. Wechselberger, P. Schell, J. W. Engels, C. Griesinger, H. Schwalbe, *Tetrahedron Lett.* **1994**, *35*, 6649–6652.
- [13] M. Chamberlin, J. Ring, *J. Biol. Chem.* **1973**, *248*, 2235–2250.
- [14] E. T. Butler, M. J. Chamberlin, *J. Biol. Chem.* **1982**, *257*, 5772–5778.
- [15] J. F. Milligan, D. R. Groebe, W. G. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783–8798.
- [16] J. R. Wyatt, M. Chastain, J. D. Puglisi, *Biotechniques* **1991**, *11*, 764–769.
- [17] W. T. Stump, K. B. Hall, *Nucleic Acids Res.* **1993**, *21*, 5480–5484.
- [18] J. F. Milligan, O. C. Uhlenbeck, *Meth. Enzymol.* **1989**, *180*, 51–62.
- [19] C. A. Grosshans, T. R. Cech, *Nucleic Acids Res.* **1991**, *19*, 3875–3880.
- [20] A. R. Ferre-D'Amare, J. A. Doudna, *Nucleic Acids Res.* **1996**, *24*, 977–978.
- [21] A. Pingoud, A. Fliess, V. Pingoud, *HPLC of Macromolecules*, IRL Press, Oxford, **1989**.
- [22] C. Sich, O. Ohlenschlaeger, R. Ramachandran, M. Goerlach, L. R. Brown, *Biochemistry* **1997**, *36*, 13989–14002.
- [23] M. Stoldt, J. Wöhnert, M. Görlach, L. R. Brown, *EMBO J.* **1998**, *17*, 6377–6384.
- [24] G. Varani, I. Tinoco, Jr., *Q. Rev. Biophys.* **1991**, *24*, 479–532.
- [25] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
- [26] P. R. Schimmel, A. G. Redfield, *Ann. Rev. Biophys. Bioeng.* **1980**, *9*, 181–221.
- [27] D. R. Kearns, D. J. Patel, R. G. Schulman, *Nature* **1971**, *229*, 338–339.
- [28] B. R. Reid, N. S. Ribeiro, L. McCollum, J. Abbate, R. E. Hurd, *Biochemistry* **1977**, *16*, 2086–2094.
- [29] D. R. Kearns, Y. P. Wong, *J. Mol. Biol.* **1974**, *87*, 755–774.
- [30] M. J. Kime, P. B. Moore, *Biochemistry* **1983**, *22*, 2615–2622.
- [31] M. J. Kime, P. B. Moore, *Biochemistry* **1983**, *22*, 2622–2629.
- [32] A. J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.* **1998**, *120*, 8293–8297.
- [33] K. Pervushin, A. Ono, C. Fernández, T. Szyperski, M. Kainosho, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14147–14151.
- [34] G. W. Vuister, A. Bax, *J. Am. Chem. Soc.* **1993**, *115*, 7772–7777.
- [35] A. Majumdar, A. Kettani, E. Skripkin, D. J. Patel, *J. Biomol. NMR* **1999**, *15*, 207–211.
- [36] A. Majumdar, A. Kettani, E. Skripkin, *J. Biomol. NMR* **1999**, *14*, 67–70.
- [37] A. Majumdar, Y. Gosser, D. J. Patel, *J. Biomol. NMR* **2001**, *21*, 289–306.
- [38] J. Wöhnert, A. J. Dingley, M. Stoldt, M. Görlach, S. Grzesiek, L. R. Brown, *Nucleic Acids Res.* **1999**, *27*, 3104–3110.
- [39] O. Ohlenschläger, J. Wöhnert, E. Bucci, K. Sidigi, S. Häfner, R. Zell, M. Görlach, unpublished work.
- [40] M. Hennig, J. R. Williamson, *Nucleic Acids Res.* **2000**, *28*, 1585–1593.
- [41] A. J. Dingley, J. E. Masse, R. D. Peterson, M. Barfield, J. Feigon, S. Grzesiek, *J. Am. Chem. Soc.* **1999**, *121*, 6019–6027.
- [42] A. Liu, A. Majumdar, W. Hu, A. Kettani, E. Skripkin, D. J. Patel, *J. Am. Chem. Soc.* **2000**, *122*, 3206–3210.
- [43] C. Richter, PhD thesis, University of Frankfurt, Germany, **1999**.
- [44] B. Luy, J. P. Marino, *J. Am. Chem. Soc.* **2000**, *122*, 8095–8096.
- [45] A. J. Dingley, J. E. Masse, J. Feigon, S. Grzesiek, *J. Biomol. NMR* **2000**, *16*, 279–289.
- [46] D. P. Giedroc, P. V. Cornish, M. Hennig, *J. Am. Chem. Soc.* **2003**, *125*, 4676–4677.
- [47] C. A. Theimer, L. D. Finger, L. Trantirek, J. Feigon, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 449–454.
- [48] C. Richter, PhD thesis, Johann Wolfgang Goethe University, Frankfurt, Germany, **1999**.
- [49] M. Grüne, J. P. Fürste, S. Klusmann, V. A. Erdmann, L. R. Brown, *Nucleic Acids Res.* **1996**, *24*, 2592–2596.
- [50] E. L. Hahn, *Phys. Rev.* **1950**, *80*, 580–594.
- [51] E. O. Stejskal, J. E. Tanner, *J. Chem. Phys.* **1965**, *42*, 288–292.
- [52] J. Lapham, J. P. Rife, P. B. Moore, D. M. Crothers, *J. Biomol. NMR* **1997**, *10*, 255–262.
- [53] J. H. Cate, R. L. Hanna, J. A. Doudna, *Nat. Struct. Biol.* **1997**, *4*, 553–558.
- [54] P. Legault, C. G. Hoogstraten, E. Metlitzky, A. Pardi, *J. Mol. Biol.* **1998**, *284*, 325–335.
- [55] S. E. Butcher, F. H.-T. Allain, J. Feigon, *Biochemistry* **2000**, *39*, 2174–2182.
- [56] J. S. Kieft, I. J. Tinoco, *Structure* **1997**, *5*, 713–721.
- [57] M. R. Hansen, J.-P. Simorre, P. Hanson, V. Mokler, L. Bellon, L. Beigelman, A. Pardi, *RNA* **1999**, *5*, 1099–1104.
- [58] A. Huppler, L. J. Nikstad, A. M. Allmann, D. A. Brow, S. E. Butcher, *Nat. Struct. Biol.* **2002**, *9*, 431–435.
- [59] F. H.-T. Allain, G. Varani, *Nucleic Acids Res.* **1995**, *23*, 341–350.
- [60] G. Colmenajero, I. J. Tinoco, *J. Mol. Biol.* **1999**, *290*, 119–135.
- [61] R. L. J. Gonzalez, I. J. Tinoco, *J. Mol. Biol.* **1999**, *289*, 1267–1282.
- [62] S. Rüdisser, I. J. Tinoco, *J. Mol. Biol.* **2000**, *295*, 1211–1223.
- [63] M. Schmitz, I. J. Tinoco, *RNA* **2000**, *6*, 1212–1225.
- [64] J.-P. Simorre, G. R. Zimmerman, A. Pardi, B. T. I. Farmer, L. Mueller, *J. Biomol. NMR* **1995**, *6*, 427–432.
- [65] J. S. Simorre, G. R. Zimmerman, L. Müller, A. Pardi, *J. Biomol. NMR* **1996**, *7*, 153–156.
- [66] V. Sklenar, T. Dieckmann, S. E. Butcher, J. Feigon, *J. Biomol. NMR* **1996**, *7*, 83–87.
- [67] R. Fiala, F. Jiang, D. J. Patel, *J. Am. Chem. Soc.* **1996**, *118*, 689–690.
- [68] P. Legault, B. T. Farmer II, L. Mueller, A. Pardi, *J. Am. Chem. Soc.* **1994**, *116*, 2203–2204.
- [69] J. P. Marino, J. H. Prestegard, D. M. Crothers, *J. Am. Chem. Soc.* **1994**, *116*, 2205–2206.
- [70] M. v. Dongen, S. Wymengen, R. Eritja, F. Azorin, C. Hilbers, *J. Biomol. NMR* **1996**, *8*, 207.
- [71] J. Wöhnert, R. Ramachandran, M. Görlach, L. R. Brown, *J. Magn. Reson.* **1999**, *139*, 430–433.
- [72] V. Sklenar, J. E. Masse, J. Feigon, *J. Magn. Res.* **1999**, *137*, 345–349.
- [73] J. Wöhnert, M. Görlach, H. Schwalbe, *J. Biomol. NMR* **2003**, *26*, 79–83.
- [74] G. Varani, I. J. Tinoco, *J. Am. Chem. Soc.* **1991**, *113*, 9349–9354.
- [75] L. Mueller, P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1995**, *117*, 11043–11048.
- [76] C. Zwahlen, P. Legault, S. Vincent, J. Greenblatt, R. Kourat, L. E. Kay, *J. Am. Chem. Soc.* **1997**, *119*, 6711–6721.
- [77] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1993**, *3*, 721–727.
- [78] B. T. Farmer II, L. Mueller, E. P. Nikonowicz, A. Pardi, *J. Biomol. NMR* **1994**, *4*, 129–133.
- [79] J. P. Marino, J. L. Diener, P. B. Moore, C. Griesinger, *J. Am. Chem. Soc.* **1997**, *119*, 7361–7366.
- [80] V. Sklenar, T. Dieckmann, S. E. Butcher, J. Feigon, *J. Magn. Reson.* **1998**, *130*, 119–124.
- [81] R. Fiala, J. Czernek, V. Sklenar, *J. Biomol. NMR* **2000**, *16*, 291–302.
- [82] B. Brucher, J.-P. J. Simorre, *J. Biomol. NMR* **2001**, *21*, 367–372.
- [83] S. Grzesiek, A. Bax, *J. Biomol. NMR* **1995**, *4*, 325–332.
- [84] V. Sklenar, M. R. Rejante, R. D. Peterson, E. Wang, J. Feigon, *J. Am. Chem. Soc.* **1993**, *115*, 12181–12182.
- [85] B. T. Farmer II, L. Muller, E. P. Nikonowicz, A. Pardi, *J. Am. Chem. Soc.* **1993**, *115*, 11040–11041.
- [86] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1994**, *4*, 117–122.
- [87] D. R. Muhandiram, L. E. Kay, *J. Magn. Reson. Ser. B* **1994**, *103*, 203–216.
- [88] H. Schwalbe, J. P. Marino, S. J. Glaser, C. Griesinger, *J. Am. Chem. Soc.* **1995**, *117*, 7251–7252.
- [89] J. P. Marino, H. Schwalbe, S. J. Glaser, C. Griesinger, *J. Am. Chem. Soc.* **1996**, *118*, 4388–4395.
- [90] S. J. Glaser, H. Schwalbe, J. P. Marino, C. Griesinger, *J. Magn. Reson.* **1996**, *B112*, 160–180.
- [91] A. Pardi, E. P. Nikonowicz, *J. Am. Chem. Soc.* **1992**, *114*, 9202–9203.
- [92] H. A. Heus, S. S. Wijmenga, F. J. M. van de Ven, C. W. Hilbers, *J. Am. Chem. Soc.* **1994**, *116*, 4983–4984.
- [93] R. Ramachandran, C. Sich, M. Grüne, V. Soskic, L. R. Brown, *J. Biomol. NMR* **1996**, *7*, 251–255.
- [94] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *J. Am. Chem. Soc.* **1994**, *116*, 6472–6473.
- [95] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *J. Biomol. NMR* **1995**, *5*, 87–92.

- [96] J. A. M. T. C. Cromsig, C. W. Hilbers, S. S. Wijmenga, *J. Biomol. NMR* **2001**, *21*, 11–29.
- [97] M. Ebrahimi, P. Rossi, C. Rogers, G. S. Harbison, *J. Magn. Reson.* **2001**, *150*, 1–9.
- [98] M. Remin, D. Shugar, *Biochem. Biophys. Res. Commun.* **1972**, *48*, 636–642.
- [99] F. M. Jucker, H. A. Heus, P. F. Yip, E. H. M. Moors, A. Pardi, *J. Mol. Biol.* **1996**, *264*, 968–980.
- [100] A. A. Szewczak, P. B. Moore, *J. Mol. Biol.* **1995**, *247*, 81–98.
- [101] A. A. Szewczak, P. B. Moore, Y.-L. Chan, I. G. Wool, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 9581–9585.
- [102] B. Wimberly, *Nat. Struct. Biol.* **1994**, *1*, 820–827.
- [103] B. Wimberly, G. Varani, I. J. Tinoco, *Biochemistry* **1993**, *32*, 1078–1087.
- [104] A. Dallas, P. B. Moore, *Structure* **1997**, *5*, 1639–1653.
- [105] S. E. Butcher, J. M. Burke, *Biochemistry* **1994**, *33*, 992–999.
- [106] A. D. Branch, B. J. Benenfeld, H. D. Robertson, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 6590–6594.
- [107] C. A. G. Haasnoot, F. A. A. M. De Leeuw, C. Altona, *Tetrahedron* **1980**, *36*, 2783–2792.
- [108] C. A. G. Haasnoot, F. A. A. M. de Leeuw, H. P. M. de Leeuw, C. Altona, *Org. Magn. Reson.* **1981**, *15*, 43–52.
- [109] C. A. G. Haasnoot, F. A. A. M. de Leeuw, H. P. M. de Leeuw, C. Altona, *Biopolymers* **1981**, *20*, 1211–1216.
- [110] E. Duchardt, C. Richter, B. Reif, S. J. Glaser, J. W. Engels, C. Griesinger, H. Schwalbe, *J. Biomol. NMR* **2001**, *21*, 117–126.
- [111] I. C. Felli, C. Richter, C. Griesinger, H. Schwalbe, *J. Am. Chem. Soc.* **1999**, *121*, 1956–1957.
- [112] H. Schwalbe, T. Carlamango, M. Hennig, J. Junker, B. Reif, C. Richter, C. Griesinger, *Meth. Enzymol.* **2001**, *338*, 35–81.
- [113] C. Griesinger, M. Hennig, J. P. Marino, B. Reif, C. Richter, H. Schwalbe, *Modern Techniques in Protein NMR* **1999**, *16*, 259–367.
- [114] M. R. Hansen, L. Mueller, A. Pardi, *Nat. Struct. Biol.* **1998**, *5*, 1065–1074.
- [115] M. Rückert, G. Otting, *J. Am. Chem. Soc.* **2000**, *122*, 7793–7797.
- [116] T. C. Leeper, M. B. Martin, H. Kim, S. Cox, V. Semenchenko, F. J. Schmidt, S. R. Van Doren, *Nat. Struct. Biol.* **2002**, *9*, 397–403.
- [117] R. D. Beger, V. M. Marathias, B. F. Volkman, P. H. Bolton, *J. Magn. Reson.* **1998**, *135*, 256–259.
- [118] B. N. M. van Buuren, J. Schleucher, V. Wittmann, C. Griesinger, H. Schwalbe, S. S. Wijmenga, unpublished work.
- [119] J. J. Warren, P. B. Moore, *J. Biomol. NMR* **2001**, *20*, 311–323.
- [120] S. A. McCallum, A. Pardi, *J. Mol. Biol.* **2003**, *326*, 1037–1050.
- [121] P. Vallurupalli, P. B. Moore, *J. Mol. Biol.* **2003**, *325*, 843–856.
- [122] D. C. Lawrence, C. C. Stover, J. Noznitsky, Z. Wu, M. F. Summers, *J. Mol. Biol.* **2003**, *326*, 529–542.
- [123] M. Hennig, T. Carlamango, J. R. Williamson, *J. Am. Chem. Soc.* **2001**, *123*, 3395–3396.
- [124] M. R. Hansen, M. Rance, A. Pardi, *J. Am. Chem. Soc.* **1998**, *120*, 11 210–11 211.
- [125] P. Bayer, L. Varani, G. Varani, *J. Biomol. NMR* **1999**, *14*, 149–155.
- [126] N. Sibille, A. Pardi, J.-P. Simorre, M. Blackledge, *J. Am. Chem. Soc.* **2001**, *123*, 12 135–12 146.
- [127] E. T. Molloy, M. R. Hansen, A. Pardi, *J. Am. Chem. Soc.* **2000**, *122*, 11 561–11 562.
- [128] K. Bodensgaard, E. T. Molloy, A. Pardi, *Biochemistry* **2002**, *41*, 11 532–11 542.
- [129] H. M. Al-Hashimi, A. Gorin, A. Majumdar, D. J. Patel, *J. Am. Chem. Soc.* **2001**, *123*, 3179–3180.
- [130] H. M. Al-Hashimi, Y. Gosser, A. Gorin, W. Hu, A. Majumdar, D. J. Patel, *J. Mol. Biol.* **2002**, *315*, 95–102.
- [131] H. M. Al-Hashimi, A. Gorin, A. Majumdar, Y. Gosser, D. J. Patel, *J. Mol. Biol.* **2002**, *318*, 631–649.
- [132] M. Stoldt, J. Wöhnert, O. Ohlenschläger, M. Görlach, L. R. Brown, *EMBO J.* **1999**, *18*, 6508–6521.
- [133] M. Lu, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 2023–2028.
- [134] Y. Yang, N. Declerck, X. Manival, S. Aymerich, M. Kochoyan, *EMBO J.* **2002**, *21*, 1987–1997.
- [135] P. W. Howe, K. Nagai, D. Neuhaus, G. Varani, *EMBO J.* **1994**, *13*, 3873–3881.
- [136] D. Fourmy, E. Guittet, S. Yoshizawa, *J. Mol. Biol.* **2002**, *324*, 137–150.
- [137] I. Lebars, S. Yoshizawa, A. R. Stenholm, E. Guittet, S. Douthwaite, D. Fourmy, *EMBO J.* **2003**, *22*, 183–192.
- [138] H. Takahashi, T. Nakanishi, K. Kami, Y. Arata, I. Shimada, *Nat. Struct. Biol.* **2000**, *7*, 220–223.
- [139] A. Ramos, G. Kelly, D. Hollingworth, A. Pastore, T. A. Frenkiel, *J. Am. Chem. Soc.* **2000**, *122*, 11 311–11 314.
- [140] A. Lane, G. Kelly, A. Ramos, T. A. Frenkiel, *J. Biomol. NMR* **2001**, *21*, 127–139.
- [141] C. Richter, B. Reif, C. Griesinger, H. Schwalbe, *J. Am. Chem. Soc.* **2000**, *122*, 12 728–12 731.
- [142] H. Schwalbe, W. Samstag, J. W. Engels, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1993**, *3*, 479–486.
- [143] H. Schwalbe, J. P. Marino, G. C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 631–644.
- [144] P. Legault, F. M. Jucker, A. Pardi, *FEBS Lett.* **1995**, *362*, 156–160.
- [145] C. Richter, B. Reif, K. Worner, S. Quant, J. P. Marino, J. W. Engels, C. Griesinger, H. Schwalbe, *J. Biomol. NMR* **1998**, *12*, 223–230.
- [146] W. Hu, S. Bouaziz, E. Skripkin, A. Kettani, *J. Magn. Res.* **1999**, *139*, 181–185.
- [147] T. Carlamango, M. Hennig, J. R. Williamson, *J. Biomol. NMR* **2002**, *22*, 65–81.
- [148] J. V. Hines, G. Varani, S. M. Landry, I. Tinoco, Jr., *J. Am. Chem. Soc.* **1993**, *115*, 11 002–11 003.
- [149] V. Sklenar, A. Bax, *J. Am. Chem. Soc.* **1987**, *109*, 7525–7526.
- [150] G. M. Clore, E. C. Murphy, A. M. Gronenborn, A. Bax, *J. Magn. Res.* **1998**, *134*, 164–167.
- [151] L. Trantirek, R. Stefl, J. E. Masse, J. Feigon, V. Sklenar, *J. Biomol. NMR* **2002**, *23*, 1–20.
- [152] E. Duchardt, C. Richter, O. Ohlenschläger, M. Görlach, J. Wöhnert, H. Schwalbe, *J. Am. Chem. Soc.* **2003**, in press.
- [153] C. Richter, C. Griesinger, I. C. Felli, P. T. Cole, G. Varani, H. Schwalbe, *J. Am. Chem. Soc.* **1999**, *121*, 1956–1957.
- [154] R. Riek, *J. Magn. Res.* **2001**, *149*, 149–153.
- [155] E. Chiarparin, S. Rüdiger, G. Bodenhausen, *ChemPhysChem* **2001**, *2*, 41–45.

Received: June 24, 2003 [A700]

## RESONANCE ASSIGNMENT AND DETERMINATION OF LOCAL RNA CONFORMATION

**RESEARCH ARTICLE:** NEW NMR EXPERIMENTS FOR

RNA NUCLEOBASE RESONANCE ASSIGNMENT AND CHEMICAL  
SHIFT ANALYSIS OF AN RNA UUCG TETRALOOP

Boris Fürtig, Christian Richter, Wolfgang Bermel, Harald Schwalbe,  
*J. Biomol. NMR*, 2004, **28**, 69-79

**RESEARCH ARTICLE:** CONFORMATIONAL DEPENDENCE OF

$^1\text{J}$  SCALAR COUPLING CONSTANTS IN RNA: ANALYSIS FOR  
AN RNA cUUCGg-TETRALOOP

Boris Fürtig, Christian Richter, Harald Schwalbe, *submitted*

The two papers present static NMR studies on a 14mer RNA cUUCGg tetraloop molecule. This RNA serves as a model system for the development and validation of new NMR pulse sequences and structural parameters.

The first paper describes the complete resonance assignment for a RNA molecule and further provides two new experiments that allow the determination of chemical shifts of quaternary carbons in RNA.

In the second contribution the impact of  $^1\text{J}$  scalar coupling on the determination of the conformation of the sugar moiety in RNA molecules is emphasized. In order to be able of parameterize the new parameters the torsion angles in the sugar moiety were determined at high accuracy with established methods including cross-correlated relaxation rates and  $^3\text{J}_{\text{HH}}$  scalar couplings.

The author of this thesis has performed all NMR experiments and analysis, in part together with Dr. Christian Richter.





## New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop

Boris Fürtig<sup>a</sup>, Christian Richter<sup>a</sup>, Wolfgang Bermel<sup>b</sup> & Harald Schwalbe<sup>a,\*</sup>

<sup>a</sup>*Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe-University, Marie-Curie-Strasse 11, D-60439 Frankfurt/M., Germany*

<sup>b</sup>*Bruker Biospin GmbH, Silberstreifen 4, D-76287 Rheinstetten, Germany*

Received 22 May 2003; Accepted 22 July 2003

**Key words:** chemical shift analysis, NMR resonance assignment, NMR spectroscopy, RNA, UUCG tetraloop

### Introduction

Hairpins are among the most important secondary structure elements found in RNA. They are involved in a variety of RNA functions ranging from nucleation sites for RNA folding to mediating intermolecular interactions with other RNAs or proteins. Among all hairpin loops, tetraloops are the most abundantly found in RNA. Common stable tetraloop motifs are e.g. UNCG, GNRA, and CUUG loop sequences (N representing a pyrimidine nucleotide). Structure determination of RNA tetraloops has provided significant insight into the structural basis of their stability.

In particular, the structures of the UNCG tetraloops have been solved both by NMR (Cheong et al., 1990; Varani et al., 1991; Allain & Varani, 1995a, b) and by X-ray methods (Nissen et al., 2000; Wimberly et al., 2000; Ennifar et al., 2000). The 14mer RNA UUCG tetraloop discussed here is very stable and a wealth of structural data is available. In addition, the RNA can be purchased from commercial sources.

This paper provides the near to complete resonance assignment of a 14mer hairpin containing a cUUCGg-tetraloop including the resonances of nuclei not directly bonded to hydrogens. To this end, new experiments were developed for the assignment of quaternary carbons in the nucleobases. In addition, the information of chemical shifts regarding conformation is evaluated.

While the newly available chemical shifts of carbon atoms C2 in cytosines and C4 and C5 in purines of the nucleobases differ only marginally along the

sequence and from the chemical shifts found in mononucleotides, <sup>13</sup>C and <sup>31</sup>P chemical shifts in the phosphodiester backbone are sensitive measures of RNA conformation.

In the first part of this paper we report the resonance assignment of <sup>1</sup>H, <sup>15</sup>N, <sup>13</sup>C, and <sup>31</sup>P atoms. In the second part of this report, the information content of the chemical shifts of C1', C2', C3', C4', and C5' to derive the sugar pucker mode and the exocyclic torsion angle  $\gamma$ , the chemical shifts of <sup>31</sup>P to detect non-canonical backbone conformation and the combined chemical shifts of H5'/H5'' and C5' to determine the spectroscopic assignment at the diastereotopic protons H5'/H5'' are analyzed.

### Methods and experiments

The uniformly <sup>13</sup>C,<sup>15</sup>N labeled RNA tetraloop sample with the sequence 5'-PO<sub>3</sub><sup>-</sup>-PO<sub>2</sub><sup>-</sup>-PO<sub>2</sub><sup>-</sup>-GGC-AC(UUCG)GUGCC-OH-3' has been purchased from Silantes GmbH (Munich, Germany). Samples for NMR-spectroscopy contained ~0.7 mM RNA in 20 mM KHPO<sub>4</sub>, pH 6.4, 0.4 mM EDTA and 10% <sup>2</sup>H<sub>2</sub>O. In NMR spectra essentially no duplex of the RNA could be detected. <sup>1</sup>H chemical shifts are referenced directly to 3-[2,2,3,3-<sup>2</sup>H<sub>4</sub>]-trimethylsilylpropionate (TSP) as an external reference. <sup>13</sup>C and <sup>15</sup>N chemical shifts are referenced indirectly to external TSP (Wishart et al., 1995). The shifts of <sup>31</sup>P are referenced according to the recommendations of IUPAC (Harris et al., 2001).

Spectra were acquired at 25 °C, 10 °C and at 5 °C on Bruker DRX600, AV700, and AV800 spectrometers equipped with z-axis gradient <sup>1</sup>H{<sup>13</sup>C,<sup>15</sup>N} or

\*To whom correspondence should be addressed. E-mail: schwalbe@nmr.uni-frankfurt.de

**Table 1.**  $^1\text{H}$ -,  $^{13}\text{C}$ -, and  $^{31}\text{P}$ -chemical shift assignments of the ribose moiety and the phosphodiester backbone. Chemical shifts were referenced to TSP. Superscripts indicate the experiments used for resonance assignment a: NOESY, b: fwd. HCC-TOCSY-CCH-E.COSY, c: HCP, d: stereospecific assignment by use of selective  $\text{C}5'/\text{H}5'$ -HSQC and fwd. HCC-TOCSY-CCH-E.COSY. Signals assigned in different experiments, are annotated with  $\alpha$ : HCCH-COSY,  $\beta$ :  $^{13}\text{C}$  edited NOESY and  $^{13}\text{C}$  HSQC,  $\gamma$ : 2D-HH-NOESY

	C1 <sup>a</sup> (ppm)	C2 <sup>b</sup> (ppm)	C3 <sup>b,c</sup> (ppm)	C4 <sup>b,c</sup> (ppm)	C5 <sup>b,c</sup> (ppm)	H1 <sup>a</sup> (ppm)	H2 <sup>b</sup> (ppm)	H3 <sup>b,c</sup> (ppm)	H4 <sup>b,c</sup> (ppm)	H5 <sup>b,c,d</sup> (ppm)	H5 <sup>b,c,d</sup> (ppm)	p <sup>c</sup> (ppm)
G1	90.9	75.0 <sup><math>\alpha</math></sup>	74.2 <sup><math>\beta</math></sup>	83.7	67.7	5.86	4.98 <sup><math>\alpha</math></sup>	4.46 <sup><math>\beta</math></sup>	4.60	4.43	4.33	P <sub><math>\gamma</math></sub> – 14 P <sub><math>\beta</math></sub> – 25.6 P <sub><math>\alpha</math></sub> – 12.9
G2	93.0	75.5	73.4 <sup><math>\beta</math></sup>	82.7	66.5	5.95	4.6	4.61 <sup><math>\beta</math></sup>	4.60	4.54	4.36	–3.55
C3	93.8	75.3	72.1 <sup><math>\beta</math></sup>	81.8	64.2	5.57	4.71	4.48 <sup><math>\beta</math></sup>	4.49	4.55	4.08	–4.18
A4	93.0	75.8	72.9	81.9	65.2	6.01	4.65	4.67	4.54	4.60	4.19	–3.88
C5	93.8	75.6	72.1	81.9	65.0	5.38	4.39	4.17	4.43	4.47	4.06	–4.05
U6	94.5	76.0	73.1	82.4	64.6	5.67	3.80	4.56	4.39	4.62	4.17	–4.35
							OH2' 6.4 <sup><math>\gamma</math></sup>					
U7	89.3	74.7	77.7	87.0	67.8	6.14	4.04	4.69	4.51	4.27	4.07	–3.41
C8	89.1	77.7	80.3	84.5	67.3	5.99	4.69	4.51	3.83	2.73	3.65	–5.03
G9	94.5	77.4	75.9	83.2	69.0	6.00	4.85	5.67	4.45	4.23	4.44	–4.92
G10	93.3	75.1	74.6	83.1	69.8	4.47	4.47	4.29	4.42	4.53	4.23	–2.4
U11	93.8	75.4	72.2	81.9	64.0	5.59	4.60	4.60	4.43	4.51	4.11	–4.35
G12	92.9	75.5	72.9	81.9	65.5	5.87	4.6	4.62	4.60	4.56	4.17	–3.95
C13	94.1	75.7	72.1	81.9 <sup><math>\beta</math></sup>	64.5	5.51	4.28	4.45	4.17 <sup><math>\beta</math></sup>	4.59	4.09	–4.4
C14	93.0	77.8 <sup><math>\alpha</math></sup>	69.7	83.4	65.1	5.79	4.06 <sup><math>\alpha</math></sup>	4.21	4.19	4.51	4.06	–4.1

**Table 2.**  $^1\text{H}$ -,  $^{13}\text{C}$ -chemical shift assignments of the nucleobases. Entries given in italics were measured at 10 °C, all other chemical shift values were measured at 25 °C. Chemical shifts were referenced to TSP. Superscripts indicate the experiments which were used for chemical shift assignment: a: HNCO, b: HCNC, c: HNCC, d:  $^{13}\text{C}$ -HSQC, e: HCN, f: NOESY, g:  $^{13}\text{C}$  edited NOESY, h: CPMG-NOESY and i: TROSY relayed HCCH-COSY

	C2 (ppm)	C4 (ppm)	C5 (ppm)	C6 (ppm)	C8 (ppm)	H1 (ppm)	H2 (ppm)	H3 (ppm)	H5 (ppm)	H6 (ppm)	H8 (ppm)	NH2 (ppm)
G1	<i>156.3<sup>a</sup></i>	<i>153.1<sup>b</sup></i>	<i>118.8<sup>c</sup></i>	<i>157.1<sup>a</sup></i>	139.0 <sup>de</sup>	<i>13.01<sup>f</sup></i>					8.17 <sup>ge</sup>	
G2	<i>157<sup>a</sup></i>	<i>152<sup>b</sup></i>	<i>118.9<sup>c</sup></i>	<i>161.7<sup>a</sup></i>	137.0 <sup>de</sup>	<i>13.41<sup>f</sup></i>					7.70 <sup>ge</sup>	
C3	<i>158.7<sup>b</sup></i>	<i>168.8<sup>a</sup></i>	<i>97.6<sup>d</sup></i>	<i>140.8<sup>de</sup></i>					5.34 <sup>g</sup>	7.75 <sup>ge</sup>		8.67/7.02 <sup>h</sup>
A4	<i>153.1<sup>d</sup></i>	<i>147.3<sup>i</sup></i>	<i>109.8<sup>i</sup></i>	<i>159.6<sup>i</sup></i>	139.4 <sup>d</sup>		7.46 <sup>gd</sup>				8.09 <sup>ge</sup>	8.33/6.46 <sup>h</sup>
C5	<i>158.3<sup>b</sup></i>	<i>168.5<sup>a</sup></i>	<i>97.4<sup>d</sup></i>	<i>140.3<sup>de</sup></i>					5.23	7.31 <sup>ge</sup>		8.59/7.16 <sup>h</sup>
U6	<i>153.1<sup>a</sup></i>	<i>168.0<sup>a</sup></i>	<i>105.1<sup>d</sup></i>	<i>140.6<sup>de</sup></i>				11.83 <sup>f</sup>	5.69 <sup>g</sup>	7.77 <sup>ge</sup>		
U7	<i>155.1</i>	<i>169.1<sup>a</sup></i>	<i>105.5<sup>d</sup></i>	<i>144.7<sup>de</sup></i>				11.26 <sup>f</sup>	5.89 <sup>g</sup>	8.05 <sup>ge</sup>		
C8	<i>159.9<sup>b</sup></i>	<i>168.0<sup>a</sup></i>	<i>98.6<sup>d</sup></i>	<i>142.8<sup>de</sup></i>					6.14 <sup>g</sup>	7.72 <sup>ge</sup>		7.18/6.4 <sup>h</sup>
G9	<i>155.6<sup>a</sup></i>	<i>153.4<sup>b</sup></i>	<i>118.7<sup>c</sup></i>	<i>161.6<sup>a</sup></i>	142.9 <sup>de</sup>	9.96 <sup>f</sup>					7.90 <sup>ge</sup>	
G10	<i>162.5<sup>a</sup></i>	<i>152.3<sup>b</sup></i>	<i>119.3<sup>c</sup></i>	<i>157.0<sup>a</sup></i>	139.0 <sup>de</sup>	<i>13.52<sup>f</sup></i>					8.35 <sup>ge</sup>	8.78/6.63 <sup>h</sup>
U11	<i>152.7<sup>a</sup></i>	<i>169.4<sup>a</sup></i>	<i>102.8<sup>d</sup></i>	<i>141.6<sup>de</sup></i>				13.80 <sup>f</sup>	5.19 <sup>g</sup>	7.80 <sup>ge</sup>		
G12	<i>156.4<sup>a</sup></i>	<i>151.8<sup>b</sup></i>	<i>119.0<sup>c</sup></i>	<i>161.5<sup>a</sup></i>	136.2 <sup>de</sup>	<i>12.66<sup>f</sup></i>					7.79 <sup>ge</sup>	8.15/6.02 <sup>h</sup>
C13	<i>156.1<sup>b</sup></i>	<i>168.4<sup>a</sup></i>	<i>97.2<sup>d</sup></i>	<i>141.0<sup>de</sup></i>					5.30 <sup>g</sup>	7.69 <sup>ge</sup>		8.67/7.06 <sup>h</sup>
C14	<i>156.1<sup>b</sup></i>	<i>169.1<sup>a</sup></i>	<i>98.4<sup>d</sup></i>	<i>141.7<sup>de</sup></i>					5.57 <sup>g</sup>	7.69 <sup>ge</sup>		8.48/7.13 <sup>h</sup>

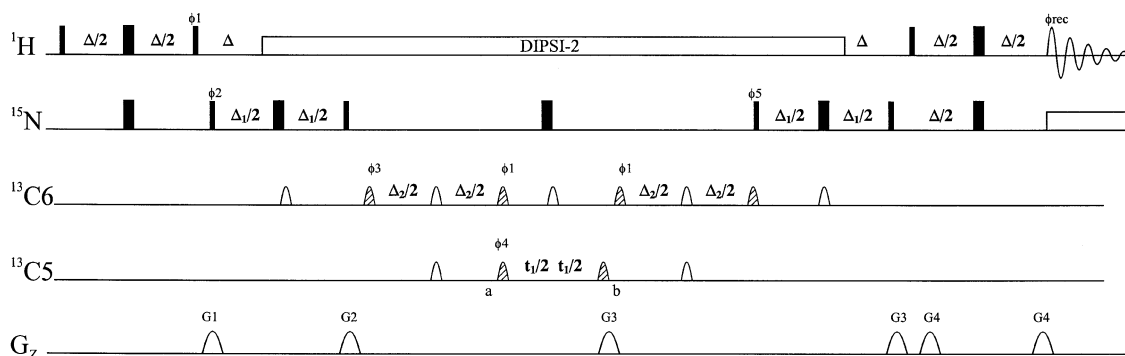


Figure 1. Pulse sequence of the HNC6C5 experiment with phase sensitive detection using States-TPPI (Marion et al. 1989). Q3 and Q5 pulses (Emsley, L. and Bodenhausen, G., 1992) were used as selective  $180^\circ$ - and  $90^\circ$ -pulses (where hatched symbols indicate  $90^\circ$  and empty symbols  $180^\circ$  pulses) with a pulse length of  $500\ \mu\text{s}$  and  $750\ \mu\text{s}$  at 700 MHz, respectively. The delays for the INEPT steps were set to  $\Delta = 1/2J(\text{H},\text{N})$ ,  $\Delta_1 = 1/2J(\text{N},\text{C6})$  and  $\Delta_2 = 1/2J(\text{C6},\text{C5})$ . The DIPSI-2 sequence (Shaka et al., 1988) was applied for decoupling of  $^1\text{H}$ , and the GARP sequence (Shaka et al., 1985) for decoupling of  $^{15}\text{N}$  during acquisition. Sine-shaped gradients were applied with gradient strength of  $G1 = 60\%$ ,  $G2 = 40\%$ ,  $G3 = -10\%$ , and  $G4 = 80\%$ , where 100% is approx. 55 G/cm. Phase cycle:  $\phi_1 = y$ ;  $\phi_2 = x_8, -x_8$ ;  $\phi_3 = x, -x$ ;  $\phi_4 = x_4, -x_4$ ;  $\phi_5 = x_2, -x_2$ ;  $\phi_{\text{rec}} = (x, -x, -x, x)(-x, x, x, -x)_2(x, -x, -x, x)$ .

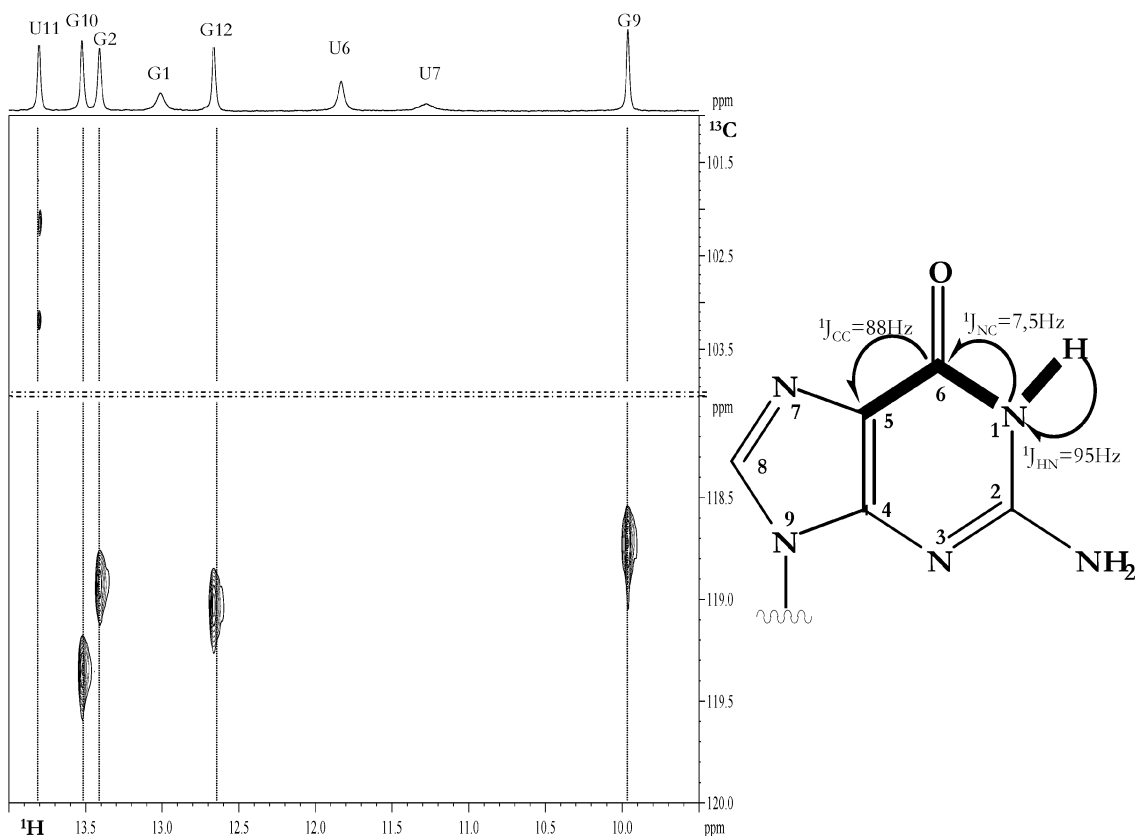


Figure 2. Spectrum of HNC6C5-experiment at 700 MHz and  $10^\circ\text{C}$ . The carrier for  $^{15}\text{N}$  was set to 154 ppm. Hard  $^1\text{H}$  pulses were applied on-resonance at the water frequency. For hard  $^1\text{H}$  and  $^{15}\text{N}$  pulses, field strengths of 27.5 kHz and 7.4 kHz were used, respectively. The GARP-decoupling (Shaka et al., 1985) during acquisition was applied with a field strength of 1.3 kHz. The 2D experiment was recorded for 8.5h with  $t_{\text{max}}$  of 28 ms (100pts, complex), 63 ms (1024 pts, complex) for  $t_1$  and  $t_2$ . The relaxation delay was set to 1.0 s. The dashed lines separate the spectrum in two parts. The upper part shows the region of  $^{13}\text{C}$  chemical shifts between 101 and 104 ppm, whereas the lower one shows the region between 118 and 120 ppm. The pointed line connects the cross peaks with respective imino-proton resonances.



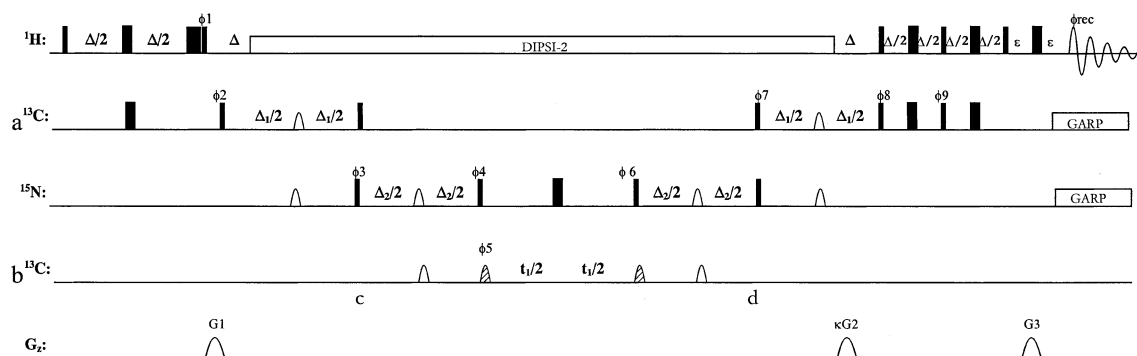


Figure 3. Pulse sequence of the HCNC-experiments with phase sensitive detection according to echo/antiecho modulation (Cavanagh et al., 1991; Palmer et al., 1991). The upper trace for the  $^{13}\text{C}$ -channel (a) shows pulses applied at the frequency of C1' in the H1'-C1'-N9(N1)-C4(C2)-experiment or at the frequency of the aromatic C8(C6) in the H8(H6)-C8(C6)-N9(N1)-C4(C2)-experiment. The lower trace for the  $^{13}\text{C}$ -channel (b) represents in both cases the frequency of C4(C2). As selective  $^{13}\text{C}$ -pulses Q3 ( $180^\circ$ ,  $\tau_P = 2900 \mu\text{s}$ , empty symbols) and Q5 ( $90^\circ$ ,  $\tau_P = 2500 \mu\text{s}$ , hatched symbols) (Emsley and Bodenhausen, 1992) were applied; as selective  $^{15}\text{N}$ -pulses IBURP2 pulses ( $180^\circ$ ,  $\tau_P = 1300 \mu\text{s}$ ) (Geen and Freeman, 1991) were applied. The delays for the inept steps were set to  $\Delta = 1/2J(\text{HC}^a)$ ,  $\Delta_1 = 1/2J(\text{C}^a\text{N})$  and  $\Delta_2 = 1/2J(\text{NC}^b)$ . The default phase was x. Phase cycle:  $\phi_1 = y$ ,  $\phi_2 = x, -x$ ;  $\phi_5 = (x)_{16}, (-x)_{16}$ ;  $\phi_7 = x$ ;  $\phi_8 = x, x, -x, -x$ ;  $\phi_9 = y, y, -y, -y$ ;  $\phi_4 = (y)_8, (-y)_8$ ;  $\phi_3 = (x)_4, (-x)_4$ ;  $\phi_{\text{rec}} = ((x, -x, -x, x, -x, x, x, -x)(-x, x, x, -x, -x, x, x, -x))_2$ . Sine shaped gradients were applied with gradient strength of G1 = 50%, G2 = 80%, and G3 = 20.1%, where 100% is approx. 55 G/cm. With the change of the sign of  $\kappa$ , the phase  $\phi_9$  is shifted by  $180^\circ$  in every second experiment, stored separately and processed according to sensitive-enhanced echo/antiecho sign discrimination.

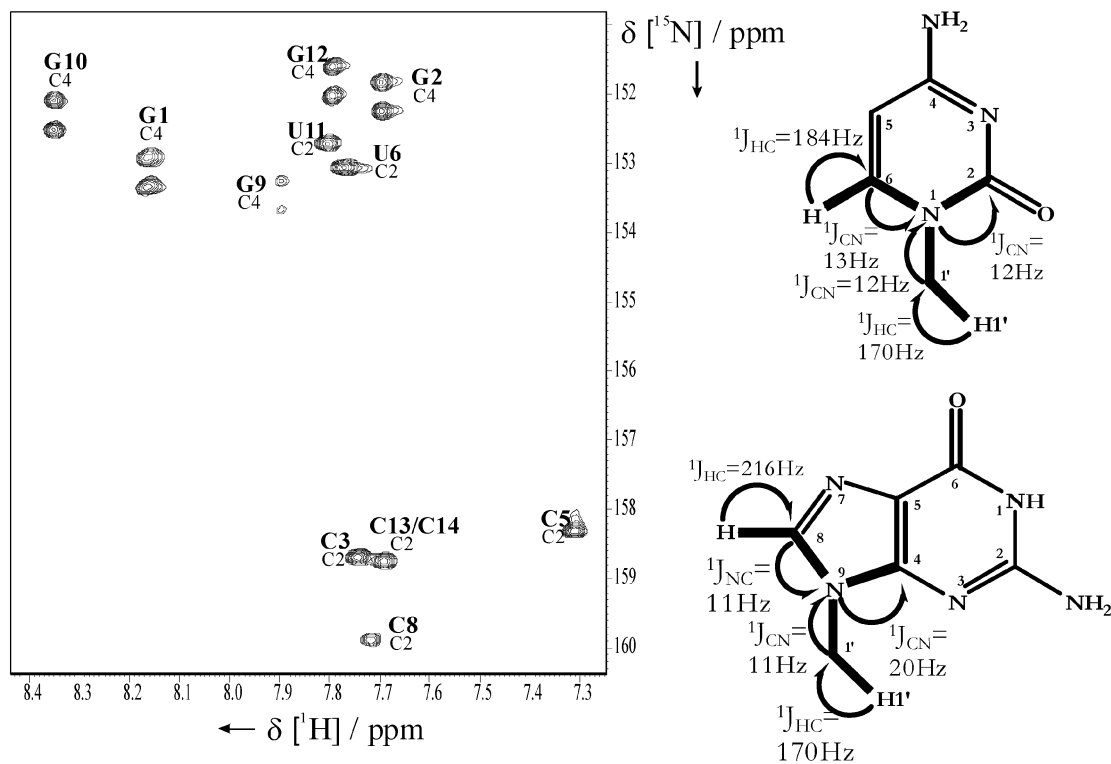


Figure 4. H8(H6)-C8(C6)-N9(N1)-C4(C2) experiment at 600 MHz and  $25^\circ\text{C}$ . The carrier for  $^{15}\text{N}$  was set to 158 ppm. The carrier for  $^{13}\text{C}$  is shifted during the experiments from 138 ppm to 152 ppm and back to 138 ppm (at points c and d annotated in the Figure 3). All  $^1\text{H}$  pulses were applied on-resonance at the resonance frequency of water. For hard  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  pulses, field strengths of 27 kHz, 22.5 kHz and 7.5 kHz were applied, respectively. The GARP-decoupling sequence (Shaka et al., 1985) was applied during acquisition with field strength of 0.8 kHz to  $^{15}\text{N}$  and with 1.8 kHz to  $^{13}\text{C}$ . DIPSI-2 decoupling of  $^1\text{H}$  (Shaka et al., 1988) was applied with a field strength of 3.1 kHz. The 2D experiment was recorded in 16h with  $t_{\text{max}}$  of 72 ms (64 pts, complex), 73 ms (512 pts, complex) for  $t_1$  and  $t_2$ . The relaxation delay was set to 1.5 s.

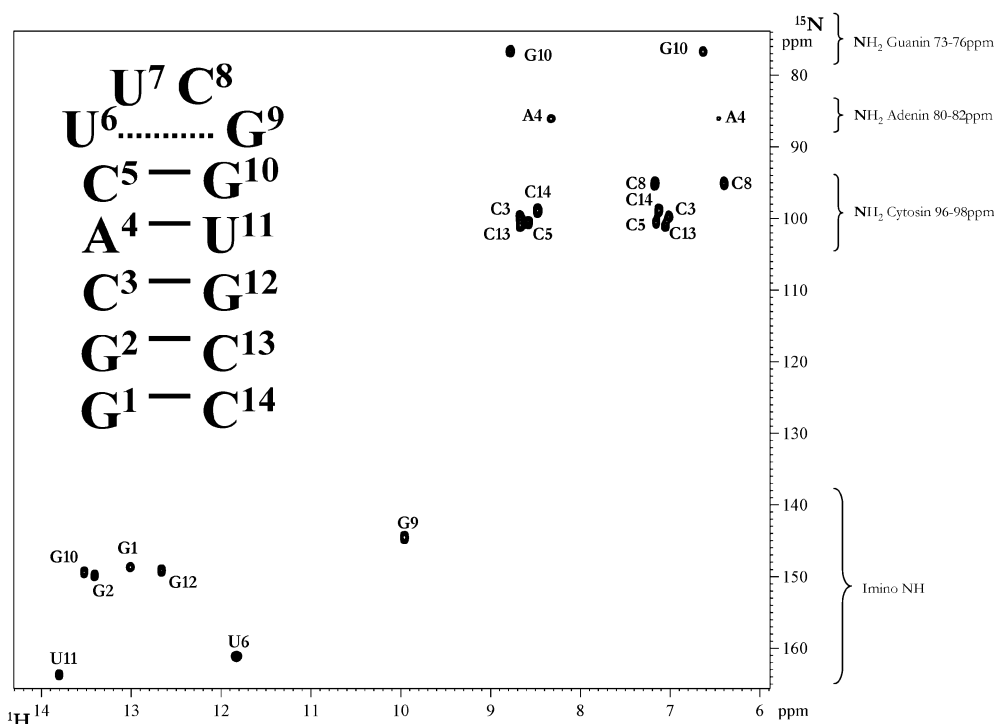


Figure 5.  $^1\text{H},^{15}\text{N}$ -HSQC at 700 MHz with phase sensitive detection using States-TPPI (Marion et al., 1989) recorded at  $10^\circ\text{C}$  with the assignments indicated. In the left part of the spectrum, the secondary structure of the tetraloop is shown. The carrier for  $^{15}\text{N}$  was set to 110 ppm. Hard pulses were applied with field strengths of 27 kHz and 7.5 kHz for  $^1\text{H}$  and  $^{15}\text{N}$ , respectively. During acquisition, GARP decoupling (Shaka et al., 1985) was applied with a field strength of 1.3 kHz. The 2D experiment was recorded in 1.5 h with  $t_{\text{max}}$  of 7.6 ms (64 complex pts, complex), 61 ms (512 complex pts) for  $t_1$  and  $t_2$ . The relaxation delay was set to 1.3 s.

z-axis gradient  $^1\text{H}\{^{13}\text{C},^{31}\text{P}\}$  triple resonance probes. Spectra were processed with XwinNMR 3.5 (Bruker) and analyzed with felix2000 (MSI). Resonance assignments were obtained from triple resonance 2D and 3D NMR experiments such as 2D  $^1\text{H},^{13}\text{C}$  CT-HSQC, 2D  $^1\text{H},^{15}\text{N}$   $^1\text{J}$ -FHSQC, 2D  $^1\text{H},^{15}\text{N}$   $^2\text{J}$ -FHSQC (Mori et al., 1995), 3D HCP (Marino et al., 1994), 3D HCN (Sklénar et al., 1993), 3D HCCH-TOCSY (Kay et al., 1993), 3D forward directed HCC-TOCSY-CCH-COSY (Schwalbe et al., 1995), 2D H(N)CO (Kay et al., 1994), 3D  $^{15}\text{N}$ -HSQC-NOESY (Sklénar et al., 1993),  $^1\text{H},^{15}\text{N}$ -CPMG-NOESY (Mueller et al., 1995) as well as 2D TROSY relayed HCCH-COSY (Simon et al., 2001). Experiments used to assign the nuclei are indicated in the Tables 1–3 by superscripts.

### New NMR pulse sequences

The following new experiments are described: 2D HNC6C5, 2D H8(H6)-C8(C6)-N9(N1)-C4(C2) and 2D H1'-C1'-N9(N1)-C4(C2). The names of these

pulse sequences are derived from the correlated atoms, names given in parenthesis are the atoms in pyrimidine nucleotides, names given without parenthesis are atoms in purine nucleotides.

### The 2D HNC6C5 experiment

For the resonance assignment of the quaternary carbon C5 in guanine, a triple-resonance 2D experiment similar to the HNCOCA sequence (Muhandiram et al., 1994) has been developed (Figure 1). In an out-and-back-manner, magnetization is transferred from the imino proton via the imino nitrogen and the carboxyl carbon atom to the C5 ( $^1\text{J}(\text{N},\text{H}^{\text{N}}) \sim 95\text{ Hz}$ ;  $^1\text{J}(\text{N},\text{C6}) \sim 7.5\text{ Hz}$ ;  $^1\text{J}(\text{C6},\text{C5}) \sim 88\text{ Hz}$ ), whose chemical shift evolves in  $t_1$ . After backtransfer, the imino resonances are detected during  $t_2$ . Coherences of carbons C5 (115–120 ppm) and C6 (160–165 ppm) have been differentiated using shaped pulses. The use of these selective pulses also prevents the magnetization transfer from nitrogen to carbon C2.

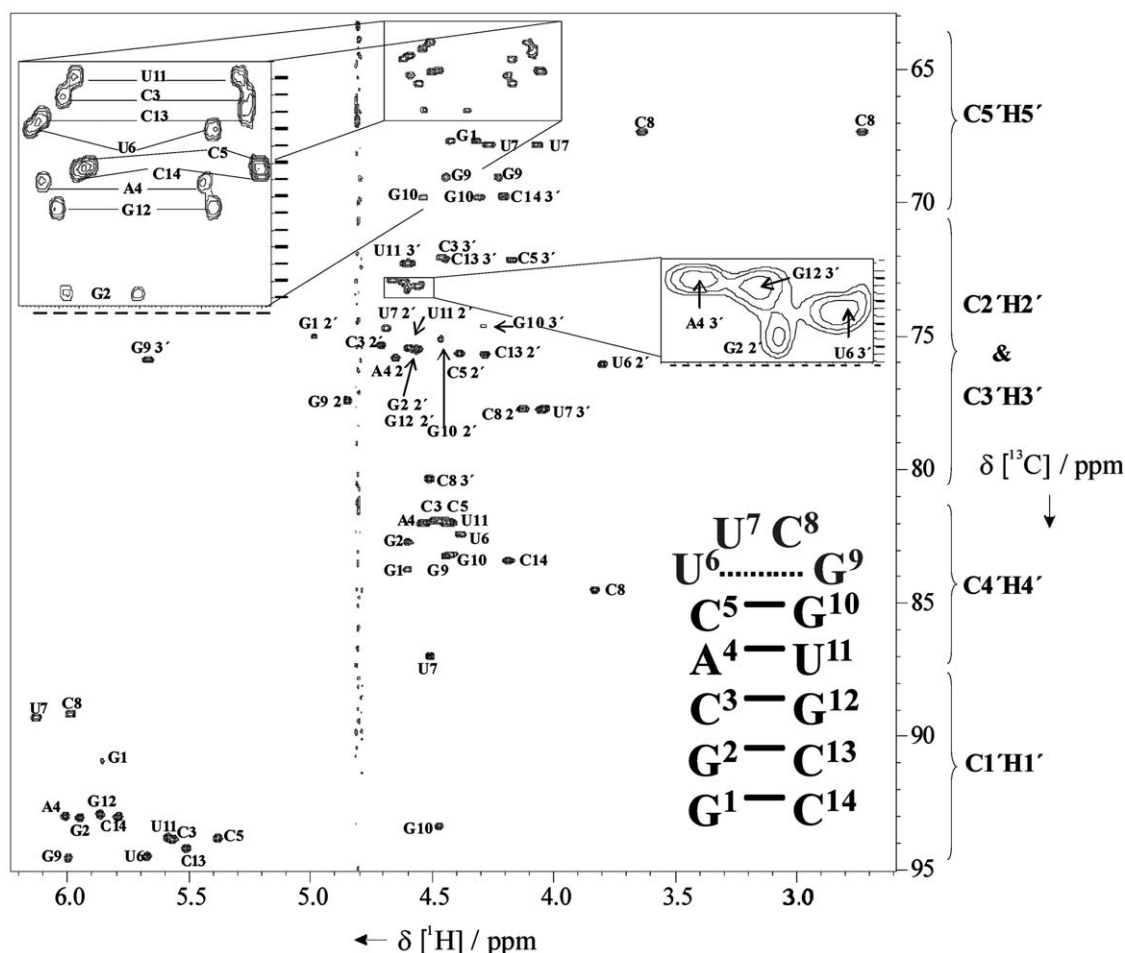


Figure 6.  $^1\text{H}$ ,  $^{13}\text{C}$ -CT-HSQC at 600 MHz with the assignments of the sugar region indicated. The carrier for  $^{13}\text{C}$  was set to 106 ppm. All proton pulses were applied on-resonance at the water frequency. Hard pulses were applied with a field strength of 29 and 19 kHz for proton and carbon, respectively. During acquisition, GARP decoupling was applied with a field strength of 3.8 kHz. The 2D experiment was recorded in 2 h with  $t_{\text{max}}$  of 17.6 ms (256 complex pts), 142 ms (1024 complex) for  $t_1$  and  $t_2$ . The relaxation delay was set to 1.5 s.

In the resulting spectrum (Figure 2), the resonances of the quaternary carbons C5 (as singlet due to selective decoupling of the neighboring carbon atoms) of nucleotides G2, G9, G10, and G12 and, in addition, carbon C6 of the only based paired U nucleotide U11 (as a doublet due to  $^1\text{J}(\text{C5}, \text{C6})$ ) can be observed. Four of the five quaternary C5 carbons could therefore be assigned, the imino resonance of the terminal G1 is too broad to be observed. Nearly uniform chemical shift values are observed for the carbons C5 along the sequence. The average value is similar to the chemical shift of carbon C5 in a GTP mononucleotide dissolved in the same buffer ( $\delta\text{C5}^{\text{GTP}}$ : 119 ppm).

#### The 2D H8(H6)-C8(C6)-N9(N1)-C4(C2) and 2D H1'-C1'-N9(N1)-C4(C2) experiment

For the assignment of the quaternary carbons C2 in pyrimidines and C4 in purines, two new pulse sequences have been developed that either correlate the resonances of the sugar H1' protons or the resonances of the aromatic H8/H6 protons with C4 and C2 in purines and pyrimidines, respectively. The experiment is of the out-and-back-type (Figure 3). The flow of magnetization is briefly discussed for the experiment correlating H1' with N9 and C4 in purines. It consists of three subsequent INEPT-steps exploiting the scalar coupling constants  $^1\text{J}(\text{H1}', \text{C1}') \sim 170$  Hz,  $^1\text{J}(\text{C1}', \text{N9}) \sim 10$ –12 Hz, and  $^1\text{J}(\text{N9}, \text{C4}) \sim 20$  Hz. For the transfer from H8 or H6 to the directly bound carbon, delays

Table 3.  $^{15}\text{N}$ -chemical shift assignments of the nucleobases. Entries given in italics were measured at 10 °C, all other chemical shift values were measured at 25 °C. Chemical shifts were referenced to TSP. Superscripts indicate the experiments which were used for chemical shift assignment: a:  $^{15}\text{N}$ -HSQC, b: HCN, c:  $^2\text{J}$ - $^{15}\text{N}$ -HSQC, d: HNN-COSY, e: HCCN

	N1 (ppm)	N2 (ppm)	N3 (ppm)	N4 (ppm)	N6 (ppm)	N7 (ppm)	N9 (ppm)
G1	<i>148.6<sup>a</sup></i>		<i>161.7<sup>d</sup></i>			233.1 <sup>c</sup>	169.5 <sup>bc</sup>
G2	<i>149.7<sup>a</sup></i>		<i>162.6<sup>d</sup></i>			234.3 <sup>c</sup>	170.5 <sup>bc</sup>
C3	151.9 <sup>b</sup>		198.1 <sup>d</sup>	99.7 <sup>a</sup>			
A4	214.6 <sup>c</sup>		223.6 <sup>c</sup>		86.1 <sup>a</sup>	231.6 <sup>c</sup>	172.2 <sup>bc</sup>
C5	152.4 <sup>b</sup>		197.4 <sup>d</sup>	100.9 <sup>a</sup>			
U6	148.1 <sup>b</sup>		<i>161.0<sup>a</sup></i>				
U7	145.1 <sup>b</sup>		159.5 <sup>a</sup>				
C8	151.9 <sup>b</sup>		204.3 <sup>e</sup>	95.1 <sup>a</sup>			
G9	<i>144.4<sup>a</sup></i>		<i>171.7<sup>d</sup></i>			232.9 <sup>c</sup>	172.9 <sup>bc</sup>
G10	<i>149.3<sup>a</sup></i>	76.7 <sup>a</sup>	<i>165.5<sup>d</sup></i>			234.5 <sup>c</sup>	171.7 <sup>bc</sup>
U11	147.6 <sup>b</sup>		<i>163.6<sup>a</sup></i>				
G12	<i>149.1<sup>a</sup></i>		<i>162.9<sup>d</sup></i>			235.6 <sup>c</sup>	170.8 <sup>bc</sup>
C13	152.4 <sup>b</sup>		199.1 <sup>d</sup>	100.6 <sup>a</sup>			
C14	153.7 <sup>b</sup>		197.8 <sup>d</sup>	98.9 <sup>a</sup>			

Table 4. Canonical coordinates *can1* and *can2* (ppm) as defined in the text

Nuc.	<i>can1</i> (ppm)	<i>can2</i> (ppm)
G1	-6.5	-17.3
G2	-5.9	-17.1
C3	-5.4	-16.9
A4	-5.6	-17.0
C5	-5.5	-16.9
U6	-5.4	-17.0
U7	-7.6	-17.6
C8	-7.0	-17.6
G9	-5.8	-17.5
G10	-6.1	-17.3
U11	-5.4	-16.8
G12	-5.7	-17.0
C13	-5.4	-16.9
C14	-5.9	-16.8

were tuned to  $^1\text{J}(\text{H8},\text{C8}) \sim 220$  Hz and  $^1\text{J}(\text{H6},\text{C6}) \sim 185$  Hz, respectively, while for transferring magnetization from N1 to C2 in cytosine and uracile, the delay was optimized based on a scalar coupling  $^1\text{J}(\text{N1},\text{C2}) \sim 12$  Hz. The transfer between glycosidic nitrogen N1 or N9 and the quaternary carbons C2 and C4 was achieved using selective carbon pulses. During

the first evolution period, no decoupling for carbons was applied.

In the spectrum of such a type of experiment (Figure 4), correlation peaks are observed between H8/H6 and the corresponding C4 and C2 resonances as annotated in the figure. The assignment of the quaternary carbons is possible by relating the spectrum to the assigned regions H6C6 or H8C8 (or H1'/C1' for the other experiment) of a constant-time  $^1\text{H}, ^{13}\text{C}$ -HSQC. C4 resonances of guanines are observed as doublets due to modulation with  $^1\text{J}(\text{C4},\text{C5})$ .

The S/N ratio is better for the experiment starting at the aromatic hydrogens. Nevertheless, both pulse sequences could also be applied to an RNA 30mer (Ohlenschläger et al., 2003) with  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled guanine and cytosine residues; for the 30mer, resonance assignment for 14 out of 16 possible resonances could be achieved (data not shown). According to the analysis using only two different data sets, the chemical shifts of the carbon C4 in guanines seem to vary depending on the extent of base stacking. Although the maximum difference of chemical shift values is just 1.7 ppm, resonances for guanines involved in base stacking are shifted upfield compared to non-base-stacked nucleotides and to mononucleotides ( $\delta\text{C4}^{\text{GTP}}: 154.6$  ppm).

### Extent of NMR resonance assignment

For the 14 residues of the UUCG tetra loop, essentially complete assignments could be obtained (Figures 5 and 6). Of all  $^1\text{H}$  (except OH-resonances),  $^{13}\text{C}$ ,  $^{15}\text{N}$ , and  $^{31}\text{P}$  resonances, 324 resonances could be assigned, corresponding to 97% of all resonances. Not assigned are the resonances of the  $\text{NH}_2$ -groups of guanines G1, G2, and G9 which are broadened by exchange beyond detection.

The assigned  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ ,  $^{31}\text{P}$  chemical shifts and a number of scalar coupling constants of the cUUCGg tetraloop RNA have been deposited in the BioMagResBank under accession number BMRB-5705. The resonance assignment found for nucleotides C5 to G10 forming the apical tetraloop are similar to the resonance assignment by the Varani group (Allain et al., 1995b) for an cUUCGg tetraloop with different stem residues.

### Analysis of chemical shift data

For RNA molecules, chemical shift analysis is espe-

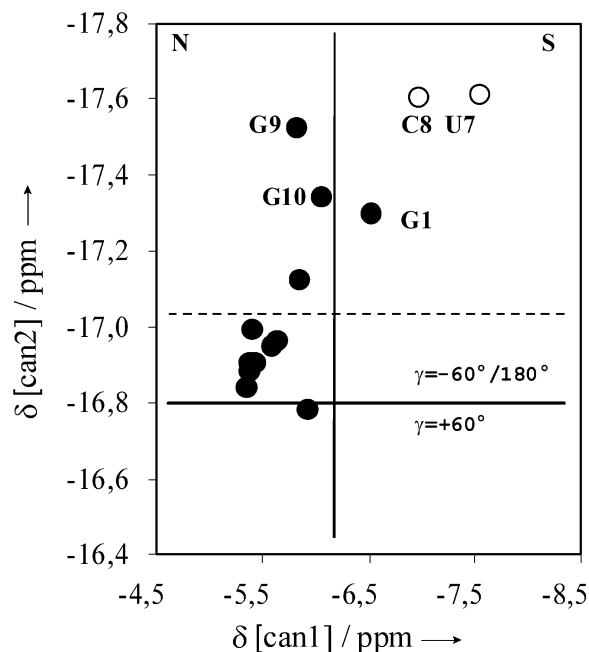


Figure 7. Plot of canonical coordinate *can2* versus *can1*. Canonical coordinates are defined in equation 1 in the text following Ebrahimi et al. (2001). N and S annotate the North and the South conformation of the ribose moiety respectively, and gg and gt give the conformation around the exocyclic torsion angles  $\gamma$  (gg stands for  $\gamma = 60^\circ$ ; gt stands for  $\gamma = -60^\circ$ ).

cially valid, since deviations from average chemical shifts are predominantly observed for residues deviating from canonical A-form helical conformation such as the loop region of the 14mer RNA (Cheong et al., 1990; Varani et al., 1991; Allain & Varani, 1995b). For example, the resonance frequencies of C2'H2' of U6, of C4'H4' of U7, of C5'H5'/C5'H5'' of C8 and of N1H1 of guanine 9 differ from those of the atoms in canonical regions. The observed chemical shift deviations can in part be explained by inspection of the structure: The chemical shifts of C2'H2' of U6 and N1H1 of G9 can be due to the unusual hydrogen bonds between U6(O2) and G9(N2H2) and between U6(O2'H2') and G9(O6). This unusual hydrogen bonding pattern results also in the ability to assign the chemical shift of the hydrogen in the 2'-hydroxy group of U6 in the NOESY spectrum ( $\delta 2'\text{OH}^{\text{U6}}$ : 6.8 ppm). The unusual chemical shifts of C4'H4' of U7 and C5'H5'/C5'H5'' of C8 can be the result of a strong variation in the backbone conformation of the loop. The downfield shift of resonances C3'H3' and C2'H2' of G9 can be caused by base stacking effects.

The dependence of  $^1\text{H}$  chemical shifts on secondary structure has been analyzed in detail for a number of different RNAs including the UUCG tetraloop (Cromsigt et al., 2001). Here, we focus on the analysis of the  $^{13}\text{C}$ - and  $^{31}\text{P}$ -chemical shift data.  $^{13}\text{C}$ -chemical shift data yielding the sugar pucker modes could be obtained following the analysis of solid-state NMR data by Ebrahimi et al. (Ebrahimi et al. 2001), which is based on the measurement and analysis of  $^{13}\text{C}$ -chemical shifts. For this analysis, canonical coordinates were calculated using the chemical shift data  $\delta$ .

$$\begin{aligned} \text{can1} &= 0.179\delta_{\text{C1}'} - 0.225\delta_{\text{C4}'} - 0.0585\delta_{\text{C5}'} \\ \text{can2} &= -0.0605(\delta_{\text{C2}'} + \delta_{\text{C3}'}) - 0.0556\delta_{\text{C4}'} - 0.0524\delta_{\text{C5}'} \end{aligned} \quad (1)$$

The first canonical coordinate *can1* describes the conformation of the sugar. For *can1* > -6.25 ppm, the sugar is in a North-conformation, for *can1* < -6.25 ppm, the sugar adopts a South-conformation. The second coordinate *can2* determines the conformation of the exocyclic torsion angle  $\gamma$ . If the torsion angle  $\gamma$  is in a gauche-gauche conformation ( $\gamma = +60^\circ$ ), *can2* > -16.8 ppm. Population of either of the two gauche-trans conformations ( $\gamma = -60^\circ, 180^\circ$ ) results in a coordinate *can2* larger than -16.8 ppm. Application of these rules to the chemical shift data of the UUCG tetraloop allows determination of the sugar pucker mode of all residues (Table 1, Figure 7). According to this analysis, residues cytosine 8 and uracile 7 adopt a C2'-endo conformation; while all other residues are in a C3'-endo conformation. The coordinate analysis fails for residue guanine 1, may be due to a higher conformational flexibility at the 5'-end of the stem—that can be detected by the analysis of the  $^3\text{J}(\text{H1}'/\text{H2}')$  and  $^3\text{J}(\text{H3}'/\text{H4}')$  coupling constants (to be published elsewhere)- or due to the additional charge of the 5'-terminal phosphate group. These results are in good agreement with the published structures and with the analysis of scalar  $^3\text{J}(\text{H,H})$ -coupling constants and cross-correlated relaxation rates in the ribose ring (to be published elsewhere).

Discrimination between gg- and gt-conformations on the basis of *can2* is less convincing. Most of the values of *can2* are clustered in the region between -16.6 ppm and -17.0 ppm. Only the nucleotides guanine 1, uracile 7, cytosine 8, guanine 9 and guanine 10 are clearly in the gt-conformation. The uncertainty of the analysis can be due to the simultaneous use of the chemical shift of C2' and C3' in the calculation. For the 14mer, the analysis would predict residues

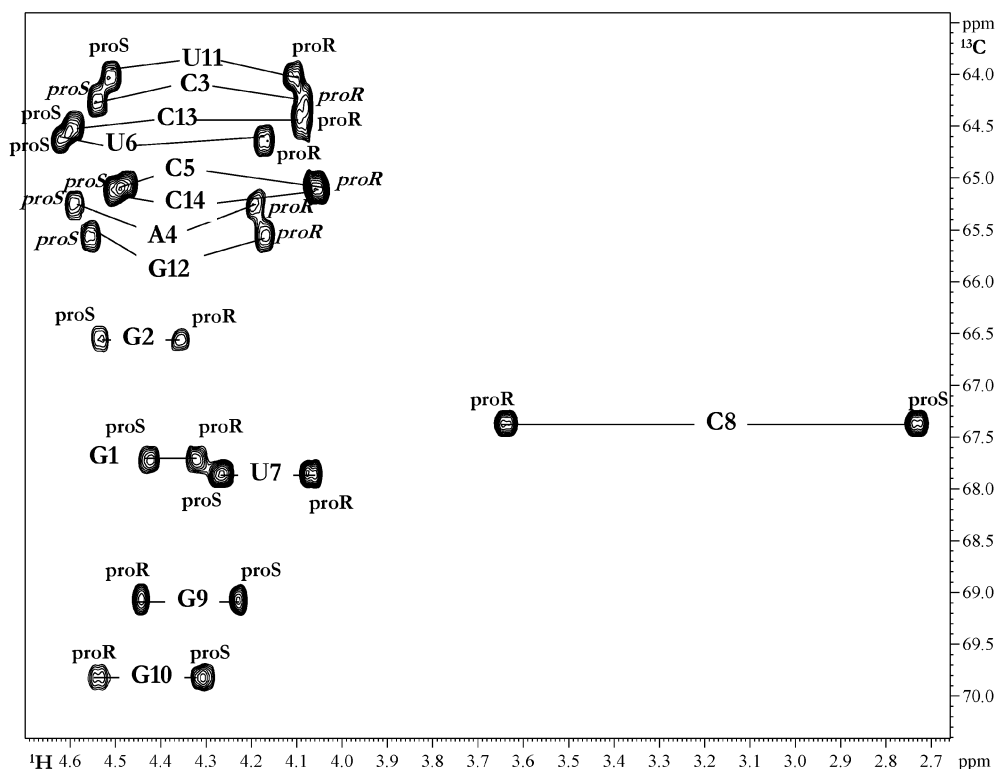


Figure 8. Expanded plot of the C5'/H5' region of a  $^1\text{H},^{13}\text{C}$ -CT-HSQC performed on the UUCG-tetraloop with the assignments annotated. The stereochemical assignment is given in *italics* if derived from chemical shift arguments alone; all other assignments have been verified by analysis of  $^2\text{J}(\text{C4}',\text{H5'/H5''})$ - and  $^3\text{J}(\text{H4}',\text{H5'/H5''})$ -coupling constants.

with  $\text{can2} > -17$  ppm to be in a *gg*-conformation around the exocyclic torsion angles  $\gamma$ .

By analyzing the chemical shift data of the C5'/H5'/C5'/H5'' resonances, a stereo-specific assignment of the prochiral H5'/H5'' protons could be obtained. It is observed that in helical RNA structures the resonance of the H5'(proS) proton is up-field shifted with respect to the resonance of the H5'(proR) proton (Remin et al., 1972). Unfortunately, this rule is not applicable to non-canonical regions of RNA structures. However, a stereo-specific assignment is possible by correlation of the difference in the proton chemical shift of the proS and proR protons with the respective carbon chemical shift  $\Delta\delta[\text{H5}(\text{proS})-\text{H5}(\text{proR})]$  ( $\delta^{13}\text{C}$ ) (Marino et al., 1996). If one depicts the differences of the proton chemical shifts versus the carbon chemical shift, the stereochemical assignment is revealed. The results are in good agreement with the data obtained by the analysis of the  $^2\text{J}(\text{C4}',\text{H5'/H5''})$ - and  $^3\text{J}(\text{H4}',\text{H5'/H5''})$ -coupling constants (to be published elsewhere). As shown in Figure 9, a general anticorrelation could be observed for the carbon chemical

shift of these resonances. An exception is the residue cytosine 8 which is probably a result of the different conformation of  $\gamma$  at this loop position.

$^{31}\text{P}$ -chemical shifts and their deviations from standard chemical shifts in A-form RNA are often taken to be indicative for unusual conformations around the phosphodiester backbone. Significant variations of the phosphorous chemical shift from the mean value of twelve assigned RNA molecules deposited in the BMRB-database are observed for the loop residues uracile 7, cytosine 8, guanine 9 and the nucleotide guanine 10 that is part of the closing base pair (Figure 10). The chemical shifts of the residues in canonical A-form conformation are not significantly different from the mean value.

## Summary

With the application of standard and new NMR-experiments, a complete resonance assignment of an 14mer RNA containing an UUCG-tetraloop was obtained. The analysis of chemical shift data revealed a

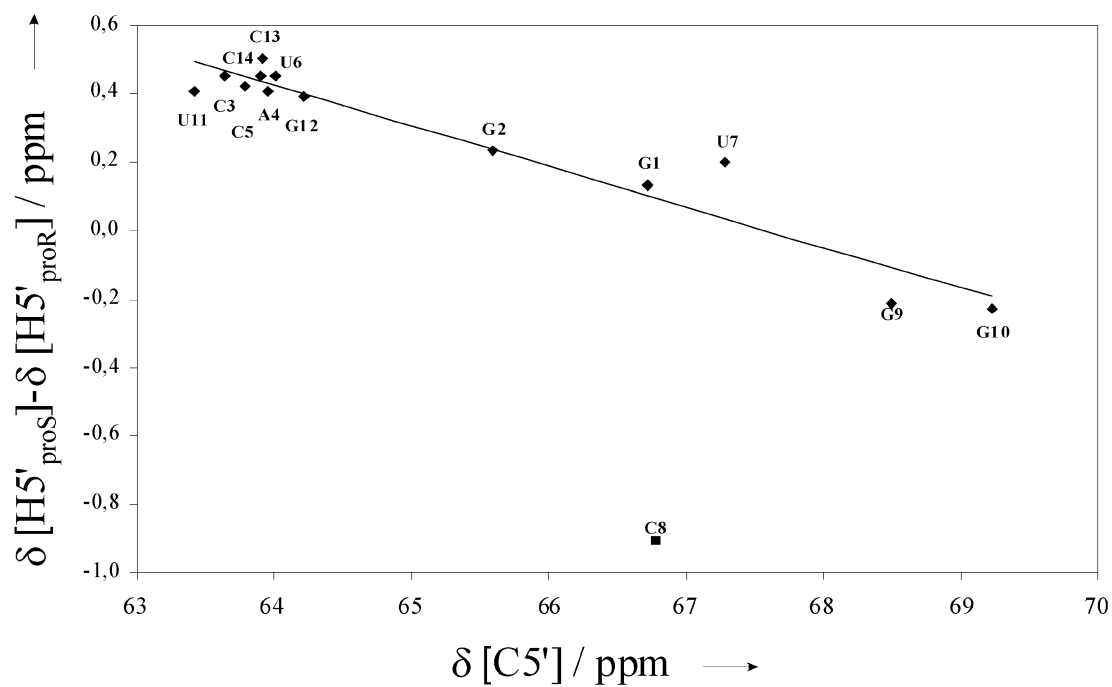


Figure 9. Plot of the difference of the proton chemical shift of the H5'(proS)proton and H5'(proR)proton versus carbon chemical shift of C5'. The fitted line has a correlation coefficient  $R$  of 0.92.

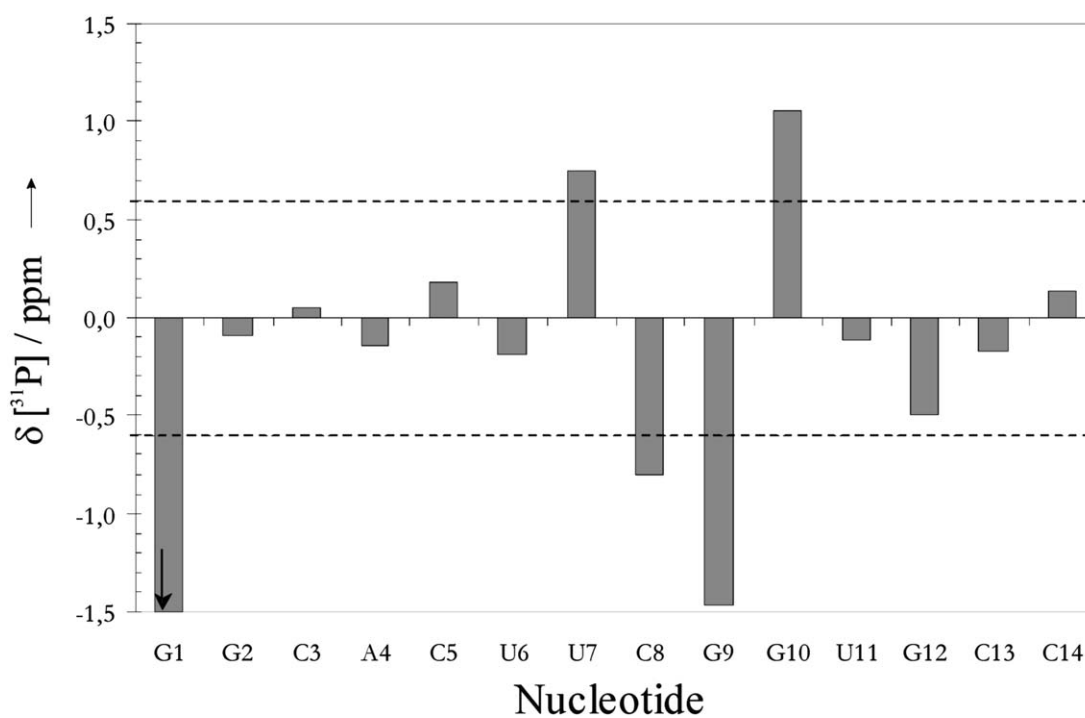


Figure 10. Deviation of the  $^{31}\text{P}$ -chemical shift from the mean value of  $^{31}\text{P}$ -chemical shifts from twelve different RNA-molecules deposited in the BMRB-database. The standard mean deviation is 0.6 ppm not regarding the deviation of the chemical shifts of the phosphorous atoms of the triphosphate of guanine 1.

first insight into the conformation of the tetraloop. In contrast to proteins, the number of published chemical shift assignments for RNAs is scarce. However, due to the smaller number of nucleotide building blocks and the intrinsic correlation of various torsion angles, statistical analysis of chemical shift data for secondary structure prediction is expected to be powerful.

## Acknowledgements

We thank Prof Vladimír Sklenář and Dr Radek Marek for helpful discussions about the initial assignment. We are very thankful to Dr Jens Wöhnert for providing an RNA sample and for helpful discussions. This work was supported by the European Large Scale facility for Biomolecular NMR (HPRI-1999-CT-00014).

## References

- Allain, F.H.-T. and Varani, G. (1995a) *J. Mol. Biol.*, **250**, 333–353.
- Allain, F.H.-T. and Varani, G. (1995b) *Nucl. Acid Res.*, **23**, 341–350.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B. and Steitz, T.A. (2000) *Science*, **289**, 905–920.
- Cavanagh, J., Palmer, A.G., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **91**, 429–436.
- Cheong, C., Varani, G. and Tinoco, I.J. (1990) *Nature*, **346**, 680–682.
- Cromsigt, J.A.M.T.C., Hilbers, C.W. and Wijmenga, S.S. (2001) *J. Biomol. NMR*, **21**, 11–29.
- Ebrahimi, M., Rossi, P., Rogers, C. and Harbison, G.S. (2001) *J. Magn. Reson.*, **150**, 1–9.
- Emsley, L., Bodenhausen, G. (1992) *J. Magn. Reson.*, **97**, 135–148.
- Ennifar, E., Nikulin, A., Tishchenko, S., Serganov, A., Nevskaya, N., Garber, M., Ehresmann, B., Ehresmann, C., Nikonov, S. and Dumas, P. (2000) *J. Mol. Biol.*, **304**, 35–42.
- Geen, H. and Freeman, R. (1991) *J. Magn. Reson.*, **93**, 93–141.
- Harris, K.R., Becker, E.D., Cabral de Menezes, S.M., Goddfellow, R. and Granger, P. (2001) *Pure Appl. Chem.*, **73**, 1795–1818.
- Kay, L.E., Xu, G.X. and Yamazaki, T. (1994) *J. Magn. Reson.*, **A109**, 129–133.
- Kay, L.E., Xu, G.Y., Singer, A.U., Muhandiram, D.R. and Forman-Kay, J.D. (1993) *J. Magn. Reson.*, **B101**, 333–337.
- Lynch, S.R., Pelton J.G. and Tinoco, I.J. (1996) *Magn. Reson. Chem.*, **34**, 11–17.
- Marino, J.P., Schwalbe, H., Anklin, C., Bermel, W., Crothers, D.M. and Griesinger, C. (1994) *J. Am. Chem. Soc.*, **116**, 6472–6473.
- Marino, J.P., Schwalbe, H., Glaser, S.J. and Griesinger, C. (1996) *J. Am. Chem. Soc.*, **118**, 4388–4395.
- Marion, D., Ikura, R., Tschudin, R. and Bax, A. (1989) *J. Magn. Reson.*, **85**, 393–399.
- Mori, S., Abeygunawardana, C., O’Neil-Johnson, M. and van Zijl, P.C.M. (1995) *J. Magn. Reson.*, **B108**, 94–98.
- Mueller, L., Legault, P. and Pardi A. (1995) *J. Am. Chem. Soc.*, **117**, 11043–11048.
- Nissen, P., Hansen, J., Ban, N., Moore, P. B. and Steitz, T.A. (2000) *Science*, **289**, 920–930.
- Ohlenschläger, O., Wöhnert, J., Bucci, E., Sidigi, K., Häfner, S., Zell, R. and Görlach, M. (2003) submitted.
- Palmer, A.G., Cavanagh, J., Wright, P.E. and Rance, M. (1991) *J. Magn. Reson.*, **93**, 151–170.
- Remin, M. and Shugar, D. (1972) *Biochem. Biophys. Res. Commun.*, **48**, 636–642.
- Schwalbe, H., Marino, J.P., Glaser, S.J. and Griesinger, C. (1995) *J. Am. Chem. Soc.*, **117**, 7251–7252.
- Shaka, A.J., Barker, P.B. and Freeman, R. (1985) *J. Magn. Reson.*, **64**, 547–552.
- Shaka, A.J., Lee, C.J. and Pines A. (1988) *J. Magn. Reson.*, **77**, 274–293.
- Simon, B., Zanier, K. and Sattler, M. (2001) *J. Biomol. NMR*, **20**, 173–176.
- Sklenář, V., Peterson, R.D., Rejante, M.R. and Feigon, J. (1993) *J. Biomol. NMR*, **3**, 721–727.
- Sklenář, V., Piotto, M., Leppik, R. and Saudek, V. (1993) *J. Magn. Reson.*, **A102**, 241–245.
- Varani, G., Cheong, C. and Tinoco, I.J. (1991) *Biochemistry*, **30**, 3280–3289.
- Wimberly, B.T., Brodersen, D.E., Clemons, Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vornrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) *Nature*, **407**, 327–339.
- Wishart, D.S., Bigam, C.G., Yao J., Abildgaard F., Dyson H.J., Oldfield E., Markley J.L. and Sykes B.D. (1995) *J. Biomol. NMR*, **6**, 135–140.



**Conformational dependence of  $^1J$  scalar coupling constants in RNA: Analysis for  
an RNA cUUCGg-Tetraloop**

Boris Fürtig, Christian Richter, and Harald Schwalbe\*

Institute for Organic Chemistry and Chemical Biology,  
Center for Biomolecular Magnetic Resonance,  
Johann Wolfgang Goethe-University,  
Marie-Curie-Straße 11, D-60439 Frankfurt/M., Germany,  
Phone: ++49-69-798 29737,  
Fax: ++49-69-798 29515,  
E-mail: [schwalbe@nmr.uni-frankfurt.de](mailto:schwalbe@nmr.uni-frankfurt.de)

[\*] to whom correspondence should be addressed.

**Keywords**

Conformational analysis, cross-correlated relaxation, Karplus parameterization, RNA, Scalar coupling constants

**Character count**

24770

**Summary**

The size of  $^1J(C,C)$  and  $^1J(C,H)$  coupling constants has been determined for a 14mer RNA oligonucleotide. It is shown that  $^1J$  show a considerable spread and vary depending on the sugar pucker mode. The conformation of the sugar moiety of the 14mer RNA is carefully analysed based on the measurement of  $^3J(H,H)$  coupling constants and dipolar cross-correlated relaxation rates  $\Gamma_{HC,HC}^{DD,DD}$  allowing a parameterization of Karplus equation for the newly measured  $^1J$  coupling constants. It can be shown that measurement of  $^{13}C$  chemical shifts and  $^1J$  coupling constants is sufficient for the derivation of sugar pucker mode.

## Introduction

The development of methods for the measurement of homo- and heteronuclear scalar coupling constants has provided important insight into the conformational dynamics of RNA molecules. Measurement of scalar coupling constants is especially important for RNA because of low proton density and a large number of free torsion angles. In addition, conformational heterogeneity is often observed in non-canonical regions of RNA structure such as loops and bulges. Therefore the use of solely NOE distance restraints especially in noncanonical regions of RNA molecules is insufficient to determine such conformational dynamics. In the past, methods for the determination of coupling constants (Furtig et al. 2003; Ippel et al. 1996; Marino et al. 1999; Schwalbe et al. 1994) have focussed on vicinal homo- and heteronuclear  $^3J$  constants due to the existence of Karplus parameterization for the coupling constants. The parameterizations have been derived using mononucleosides and –nucleotides that undergo conformational dynamics e.g. for the ribofuranosyl moiety of the RNA. Recently, methods for the determination of cross-correlated relaxation rates have been developed and good agreement for the prediction of pseudorotation phase P has been observed (Felli et al. 1999 ; Richter et al. 1999; Richter et al. 2000; Schwalbe et al. 2001). Both for vicinal coupling constants and for cross-correlated relaxation rates that involve a number of different carbon centers, the sensitivity of the underlying experiments makes the application to large RNA problematic. Recently, Brutscher and co-workers (Boisbouvier et al. 2000) have introduced a method that correlates  $\Gamma_{HC,C}^{DD,CSA}$  with the sugar conformation and that is more sensitive than the originally measured  $\Gamma_{HC,HC}^{DD,DD}$ .

Here, we report on the measurement of homo- and heteronuclear  $^1J$  scalar couplings and examined their conformational dependence using an RNA cUUCGg-tetraloop (secondary structure is represented in Figure 1a) as a model system, for which chemical shift assignment has been reported previously (Furtig et al. 2004).  $^1J$  coupling constants can be measured reliably and easily even for larger molecules. The current basis for a parameterization of such scalar couplings is very limited given the fact that few data have been reported (Doreleijers et al. 2003) and it will have to be supplemented by data for different RNA model systems and quantum chemical calculations. Nevertheless there are reports present in the literature giving evidence for a dependence of  $^1J$  heteronuclear coupling constants on conformation and structural features of sugars and sugar moieties in nucleic acids (Ippel et al. 1996; Kline et al. 1990). By comparison with  $^3J(H,H)$  coupling constants and  $\Gamma_{HC,HC}^{DD,DD}$  cross-correlated relaxation rates we can show here that homonuclear  $^1J(C,C)$  and heteronuclear  $^1J(C,H)$  coupling constants show considerable variation, derive a Karplus parameterization for those coupling constants allowing the determination of the sugar pucker mode.

## Methods and Experiments

A uniformly  $^{13}C,^{15}N$  labelled RNA tetraloop sample with the sequence 5'-PO<sub>3</sub><sup>-</sup>-PO<sub>2</sub><sup>-</sup>-PO<sub>2</sub><sup>-</sup>-GGCAC(UUCG)GUGCC-OH-3' has been purchased from Silantes GmbH (Munich, Germany). Chemical shift assignment for this RNA has been reported previously (Furtig et al. 2004). Samples for NMR-spectroscopy contained ~0.7mM RNA in 20mM KHPO<sub>4</sub>, pH 6.4, 0.4mM EDTA and 10%  $^2H_2O$  or 99.9%  $^2H_2O$ . In NMR spectra, essentially no duplex of the RNA could be detected.  $^1H$  chemical shifts were referenced directly to 3-[2,2,3,3- $^2H_4$ ]-trimethylsilylpropionate (TSP) as an external reference.  $^{13}C$  and  $^{15}N$  chemical shifts were referenced indirectly to external TSP (Wishart et al. 1995).

Spectra were acquired at 25°C on Bruker DRX600 and AV700 spectrometers equipped with z-axis gradient  $^1\text{H}\{^{13}\text{C},^{15}\text{N}\}$  or z-axis gradient  $^1\text{H}\{^{13}\text{C},^{31}\text{P}\}$  triple resonance probes. Spectra were processed with XwinNMR 3.5 (Bruker) and analyzed with felix2000 (MSI).

#### Measurement of $^3\text{J}(\text{H},\text{H})$ coupling constants in the ribose sugar ring in the forward directed HCC-TOCSY-CCH E.COSY-experiment

$^3\text{J}(\text{H},\text{H})$  couplings constants were measured in a 3D forward directed HCC-TOCSY-CCH E.COSY experiment (Glaser et al. 1996; Schwalbe et al. 1995). The times for the transfer were set to  $\tau_1=14\text{ms}$  for the CC-TOCSY and to  $\tau_L=8\text{ms}$  for the CT-delay to optimize the transfer amplitude as discussed in detail in Glaser et al. (Glaser et al. 1996). The  $^{13}\text{C}$ -carrier was set to 75ppm. All  $^1\text{H}$  pulses were applied at the water frequency. Hard pulses were applied with field strength of 29kHz and 19kHz for  $^1\text{H}$  and  $^{13}\text{C}$ , respectively. GARP decoupling (Shaka et al. 1985) during acquisition was applied with field strength of 3.85kHz; the DIPSI-3 sequence (Shaka et al. 1988) during the CC-TOCSY transfer was applied with field strength of 9.2kHz. The spectrum was recorded on a Bruker DRX600 spectrometer at 298K with 108pts, 88pt and 2048pts for  $t_1$ ,  $t_2$ , and  $t_3$ , respectively, resulting in an experimental duration of 36h (relaxation delay of 1.4s). Processing was conducted with XWINNMR 3.5; the Fourier transformed spectrum was converted to MSI Felix2000 for further analysed. The scalar coupling constants were extracted using the method described (Schwalbe et al. 1995). The resolution in the direct dimension where the  $^3\text{J}(\text{H},\text{H})$  splitting was measured was set to 1.75Hz per complex point and was zero-filled during processing to a resolution of 0.4Hz per complex point.

#### Measurement of $\Gamma_{\text{HC},\text{HC}}^{DD,DD}$ by a 2D quantitative $\Gamma$ -HCCH experiment

$\Gamma_{\text{H}^1\text{C}^1,\text{H}^2\text{C}^2}^{DD,DD}$  and  $\Gamma_{\text{H}^3\text{C}^3,\text{H}^4\text{C}^4}^{DD,DD}$  cross-correlated relaxation rates were measured using the quantitative 2D  $\Gamma$ -HCCH experiment. The  $^{13}\text{C}$  carrier was set to 75ppm. All  $^1\text{H}$  pulses were applied at the water frequency. Hard pulses were applied with a field strength of 26kHz and 22kHz for  $^1\text{H}$  and  $^{13}\text{C}$ , respectively. GARP decoupling during acquisition was applied with field strength of 3.57kHz. The delays  $\Delta'$  for evolution of scalar coupling were set to 0ms and to 3.36ms for the cross and the reference experiment, respectively. The spectra were recorded on a Bruker AV700 spectrometer at 298K with 192pts and 4096pts for  $t_1$  and  $t_2$ , respectively. The relaxation delay was 2s. Processing was conducted with XWINNMR 3.5; the Fourier transformed spectrum was converted to MSI Felix2000 for further analysis.

#### Measurement of $\Gamma_{\text{HC},\text{C}}^{DD,CSA}$ by a 2D (CT)-TROSY-experiment

$\Gamma_{\text{HC},\text{C}}^{DD,CSA}$  cross-correlated relaxation rates were measured using the CT-TROSY experiment as described by Brutscher et al. The  $^{13}\text{C}$  carrier was set to 75ppm. All  $^1\text{H}$  pulses were applied at the water frequency. Hard pulses were applied with field strength of 26kHz and 22kHz for  $^1\text{H}$  and  $^{13}\text{C}$ , respectively. GARP decoupling during acquisition was applied with field strength of 3.57kHz. Coherence selection ( $\text{C}^+\text{H}^\beta$ -TROSY and the  $\text{C}^+\text{H}^\alpha$ -TROSY) was achieved as described in the original paper. The spectra were recorded on a Bruker AV700 spectrometer at 298K with 304pts and 1024pts for  $t_1$  and  $t_2$ , respectively. The relaxation delay was 1.5s. Processing was conducted with XWINNMR 3.5; the Fourier transformed spectrum was converted to MSI Felix2000 for further analysis.

#### Measurement of $^1J(C,H)$ and $^1J(C,C)$ coupling constants by coupled HSQC spectra

$^1J(C,C)$  coupling constants were measured in the indirect evolution period of a real-time  $^1H,^{13}C$ -HSQC experiment.  $^1J(C,H)$ -coupling constants were measured in the direct evolution period of a constant-time  $\omega_2$ -antiphase  $^1H,^{13}C$ -HSQC experiment in which the two  $\pi$ -pulses on  $^1H$  and  $^{13}C$  of the final INEPT were omitted and acquisition was started directly after the two  $90^\circ$  pulses applied to  $^1H,^{13}C$ .

Figure 1

No gradients were applied. The resulting spectrum showed doublet peaks split by the  $^1J(C,H)$  coupling constant in antiphase (Figure 1b). The  $^{13}C$  carrier frequency was set to 74ppm. All proton pulses were applied on resonance at the water frequency. The hard pulses were applied with field strength of 28.4kHz, 19kHz and 6.3kHz for proton, carbon and phosphor respectively. During acquisition, a GARP decoupling for  $^{31}P$  with 0.625 kHz was applied. The experiment was recorded for 7h with 256 and 4096 complex points in  $t_1$  and  $t_2$  resulting in an acquisition time of 142ms and a  $t_{1max}$  of 85ms. The relaxation delay was 1.5s. Processing was conducted with XWINNMR 3.5; the Fourier transformed spectrum was converted to MSI Felix2000 for further analysis. The resolution in the direct dimension where the  $^1J(C,H)$  splitting was measured was 1.75Hz per complex point and was zero-filled during processing and analysis to a resolution of 0.4Hz per complex point.

## Results

### Determination of ribose sugar pucker

#### Analysis of $^3J(H,H)$ coupling constants

The conformation of the furanosidic sugar moiety and the exocyclic torsion angle  $\gamma$  in RNA can be determined by analysis of  $^3J(H,H)$  coupling constants (Altona 1982; Marino et al. 1999). A total of 52 out of 70 possible coupling constants (12  $^3J(H1',H2')$ , 8  $^3J(H2',H3')$ , 12  $^3J(H3',H4')$ , 10  $^3J(H4',H5')$ , 10  $^3J(H4',H5'')$ ) in the sugar moieties could be determined by application of a forward directed HCC-TOCSY-CCH-COSY (see a summary for the relevant parameters in Table 1).

Table 1

Figure 2 and Figure 3

Analysis of the coupling constants (Figure 2a + Figure 3a) revealed that the sugar moiety of the stem nucleotides and of the hydrogen-bonded loop nucleotides U6 and G9 adopt C3'-endo conformation, whereas the two loop nucleotides U7 and C8 are in C2'-endo conformation that is rigid on the NMR time scale. The  $^3J(H3',H4')$  coupling constant of residue G10 is affected by strong coupling, and therefore the value deviates from the expected value for a sugar moiety in C3'-endo conformation.

The  $^3J(H3',H4')$  coupling constants in C3'-endo, but also the  $^3J(H1',H2')$  constants for C2'-endo conformation, exceed the range from the Karplus parameterization. At the same time the corresponding  $^3J(H1',H2')$  for C3'-endo and  $^3J(H3',H4')$  for C2'-endo are too small. This has been observed previously for more sizeable RNA molecules (Richter et al. 1999; Schwalbe et al. 1995) and may be due to the fact, that the parameterization has been derived

from mononucleotides which are intrinsically averaged. This observation is in contrast to the analysis derived from  $\Gamma_{HC,HC}^{DD,DD}$  which is consistent.

#### Analysis of $\Gamma_{HC,HC}^{DD,DD}$ cross-correlated relaxation rates

$\Gamma_{HC,HC}^{DD,DD}$  cross-correlated relaxation rates have been introduced as reporters for the conformation of the RNA sugar moiety. As shown in Figure 2b + Figure 3b, the cross-correlated relaxation rates indicate that the conformation of the furanosidic sugar moiety in the nucleotides U7 and C8 are different with respect to all other nucleotides. Furthermore, it is indicative that these two nucleotides are in a C2'-endo conformation, while all the other nucleotides populate the C3'-endo conformational space. Based on the functional dependence of  $\Gamma_{HC,HC}^{DD,DD}$  on the pseudorotation phase P (Richter et al. 1999) shown in Figure 3b, most of the nucleotides adopting C3'-endo conformation except for the nucleotides U7 and C8 adopting C2'-endo conformation and the pucker mode of all sugar moieties can be extracted.

#### Analysis of $\Gamma_{HC,C}^{DD,CSA}$ cross-correlated relaxation rates

More recently, Boisbouvier et al. suggested the determination of the sugar pucker mode from CSA-dipolar cross-correlated relaxation rates measured in a CT-TROSY experiment. Namely, the  $\Gamma_{H1'C1',C1'}^{DD,CSA}$  and  $\Gamma_{H3'C3',C3'}^{DD,CSA}$  should be sensitive reporters for the sugar conformation. For the 14mer under study, no clear correlation  $\Gamma_{HC,C}^{DD,CSA}$  cross-correlated relaxation rate on the sugar pucker could be observed (Figure 2c). This may reflect the fact that this method relies on relaxation processes optimal for large anisotropic RNA-molecules. Our system with a ratio of  $D_{\parallel}/D_{\perp}=1.35\pm0.02$  and an rotational correlation time of  $\tau_c=2.31\pm0.13$ ns (Duchardt et al. 2005) is possibly just too small to show the effect as pronounced as reported as in the original paper.

#### Analysis of $^1J(C,H)$ and $^1J(C,C)$ coupling constants

Figure 4 shows the size of  $^1J(C,H)$  and  $^1J(C,C)$  for each residue. Especially  $^1J(C1',C2')$  and  $^1J(C2',H2')$  adopt clearly different values for C2'-endo and C3'-endo conformation.

Figure 4

In contrast, additional factors influence the size of  $^1J(C1',H1')$  coupling constants. Based on quantum mechanical calculations, Munzerova et al. (Munzarova et al. 2003) have suggested a dependence of  $^1J(C1',H1')$  coupling constants on the glycosidic torsion angle  $\chi$  that is also influenced by the sugar pucker mode of the ribose unit. Since only the nucleotide G9 adopts a syn-conformation around  $\chi$ , our findings are consistent with the report by Munzerova et al.  $^1J(C1',H1')$  is in fact influenced both by the sugar pucker and by the conformation of the glycosidic torsion angle (annotated in Figure 5a by an arrow). The values for the stem residues are all within a narrow range with variations of approx. 2Hz to 3Hz, while the  $^1J(C1',H1')$  coupling constants for loop nucleotides differ by 9Hz, 12Hz and 12Hz for  $^1J(C1',H1')$ ,  $^1J(C2',H2')$  and  $^1J(C1',C2')$ (respectively) compared to the mean value for these couplings of the stem residues.

Figure 5

Since the conformation of the furanosidic sugar moiety in the backbone of the RNA molecule could be defined by analysis of the earlier described parameters, a new parameterization of the measured  $^1\text{J}$  couplings could be attempted. Therefore, the dihedral angle  $\phi(\text{H1}',\text{H2}')$  representing the conformation of the sugar moiety is determined from  $\Gamma_{\text{HC},\text{HC}}^{\text{DD},\text{DD}}$  and plotted versus the respective scalar coupling. In analogy to the case for the  $^3\text{J}$  scalar couplings a standard Karplus equation (Karplus 1959) was fitted to the data. The obtained values for the three cases depicted in Figure 5 are listed in Table 2. Surprisingly the values for the angle  $\phi(\text{H1}',\text{H2}')$  in the 14mer RNA are representing the distribution found in RNA well, to underline this fact a distribution of the conformational space is given for the rRNA in the Ribosome (calculated with Molmol (Koradi et al. 1996) using the ribosome structure 1JJ2 (Klein et al. 2001)).

Table 2

## Discussion

First evidence for a possible dependence of  $^1\text{J}(\text{C},\text{H})$  scalar coupling constants from the conformation of the sugar phosphate backbone of RNA was noticed in early works of Serianni et al. (Kline et al. 1990; Podlasek et al. 1996; Serianni et al. 1994) who determined several  $^n\text{J}(\text{C},\text{H})$  [ $n=1,2,3$ ] and proposed a dependence on furanose ring conformation for  $^1\text{J}(\text{C1}',\text{C2}')$  and  $^1\text{J}(\text{C2}',\text{H2}')$  and by Ippel et al. (Ippel et al. 1996) who determined a complete set of scalar coupling constants in model substances for nucleic acids and determined an according trend in the  $^1\text{J}$  coupling constants. Furthermore, Serianni and co-workers (Thibaudeau et al. 2004) showed that in complex sugar molecules there is a correlation between the conformation of a  $\text{CH}_2\text{OH}$  element and the value of  $^1\text{J}$  coupling constant. The  $^1\text{J}$  couplings are additionally supposed to be sensitive reporters of the length of CH and CC-bonds (Stenutz et al. 2002) as well as of H-bond donation effects as shown in ethanol (Maiti et al. 2006). It is therefore likely that similar effects are responsible for a conformational dependence of such couplings in the sugar moiety of RNA molecules. Furthermore, Dayie (Dayie 2005) reported, but did not show, a similar correlation of  $^1\text{J}(\text{C},\text{H})$  couplings to the ribose conformation in RNA as presented in this work.

As previously reported by Varani and co-workers (Varani et al. 1991), similar  $^1\text{J}(\text{C},\text{H})$  couplings could be made observed for another tetraloop with a slightly different stem sequence.

Therefore,  $^1\text{J}(\text{C1}',\text{H1}')$ ,  $^1\text{J}(\text{C2}',\text{H2}')$  and  $^1\text{J}(\text{C1}',\text{C2}')$  are valuable tools for the fast detection of noncanonical ribose sugar conformations in RNA. Previously, we could show that the analysis of chemical shift data based on the procedure reported by Ebrahimi et al. (Ebrahimi et al. 2001) provides a means for the distinction of the principle pucker mode.

Figure 6

The comparison of the scalar  $^1\text{J}$ -coupling data with this method and the standard method of analysing scalar  $^3\text{J}$  couplings allows concluding that both methods are useful for a subsequent validation of ribose conformation in RNA molecules directly after the assignment procedure. A three dimensional depiction of the data (Figure 6) shows a clustering of both conformations at different positions in this three dimensional parameter space ( $\square$   $^{13}\text{C}; ^1\text{J}(\text{C},\text{H}); ^3\text{J}(\text{H},\text{H})$ ). The loop nucleotides Uracil7 and Cytosin8 that are in C2' endo conformation are well

separated from nucleotides that are in C3' endo conformation. However, the high sensitivity of experiments to extract one bond scalar couplings - especially the  $^1J(\text{C,H})$  couplings that could also be determined at natural abundance - and the non-prerequisite of a complete resonance assignment opens the route to a fast determination of conformations in large biologically relevant RNA molecules or for extracting conformational information in edited spectra that are recorded in a time resolved fashion (Wenter et al. 2005; Wenter et al. 2006).

## Acknowledgments

We are grateful to Prof. Dr. Jens Wöhnert and to Dr. Elke Duchardt for helpful discussions. The work has been supported by the state of Hesse (BMRZ), the SFB 579 "RNA-Ligand-Wechselwirkung" and the "Studienstiftung des Deutschen Volkes" (B.F.).

## References

- Altona, C. (1982). *Recl. Trav. Chim. Pays-Bas* **101**: 413-433.
- Boisbouvier, J., B. Brutscher, A. Pardi, D. Marion and J. P. Simorre (2000). *Journal of the American Chemical Society* **122**: 6779-6780.
- Dayie, K. T. (2005). *Journal of Biomolecular Nmr* **32**: 129-139.
- Doreleijers, J. F., S. Mading, D. Maziuk, K. Sojourner, L. Yin, J. Zhu, J. L. Markley and E. L. Ulrich (2003). *Journal of Biomolecular Nmr* **26**: 139-146.
- Duchardt, E. and H. Schwalbe (2005). *Journal of Biomolecular Nmr* **32**: 295-308.
- Ebrahimi, M., P. Rossi, C. Rogers and G. S. Harbison (2001). *Journal of Magnetic Resonance* **150**: 1-9.
- Felli, I. C., C. Richter, C. Griesinger and H. Schwalbe (1999). *Journal of the American Chemical Society* **121**: 1956-1957.
- Furtig, B., C. Richter, W. Bermel and H. Schwalbe (2004). *Journal of Biomolecular Nmr* **28**: 69-79.
- Furtig, B., C. Richter, J. Wöhnert and H. Schwalbe (2003). *ChemBiochem* **4**: 936-962.
- Glaser, S. J., H. Schwalbe, J. P. Marino and C. Griesinger (1996). *Journal of Magnetic Resonance Series B* **112**: 160-180.
- Ippel, J. H., S. S. Wijmenga, R. deJong, H. A. Heus, C. W. Hilbers, E. deVroom, G. A. vanderMarel and J. H. vanBoom (1996). *Magnetic Resonance in Chemistry* **34**: S156-S176.
- Karplus, M. (1959). *Journal of Chemical Physics* **30**: 11-15.
- Klein, D. J., T. M. Schmeing, P. B. Moore and T. A. Steitz (2001). *Embo Journal* **20**: 4214-4221.
- Kline, P. C. and A. S. Serianni (1990). *Journal of the American Chemical Society* **112**: 7373-7381.
- Koradi, R., M. Billeter and K. Wuthrich (1996). *Journal of Molecular Graphics* **14**: 51-&.
- Maiti, N. C., Y. P. Zhu, I. Carmichael, A. S. Serianni and V. E. Anderson (2006). *Journal of Organic Chemistry* **71**: 2878-2880.
- Marino, J. P., H. Schwalbe and C. Griesinger (1999). *Accounts of Chemical Research* **32**: 614-623.
- Munzarova, M. L. and V. Sklenar (2003). *Journal of the American Chemical Society* **125**: 3649-3658.
- Podlasek, C. A., W. A. Stripe, I. Carmichael, M. Y. Shang, B. Basu and A. S. Serianni (1996). *Journal of the American Chemical Society* **118**: 1413-1425.
- Richter, C., C. Griesinger, I. Felli, P. T. Cole, G. Varani and H. Schwalbe (1999). *Journal of Biomolecular Nmr* **15**: 241-250.
- Richter, C., B. Reif, C. Griesinger and H. Schwalbe (2000). *Journal of the American Chemical Society* **122**: 12728-12731.

- Schwalbe, H., T. Carlomagno, M. Hennig, J. Junker, B. Reif, C. Richter and C. Griesinger (2001). *Methods Enzymol* **338**: 35-81.
- Schwalbe, H., J. P. Marino, S. J. Glaser and C. Griesinger (1995). *Journal of the American Chemical Society* **117**: 7251-7252.
- Schwalbe, H., J. P. Marino, G. C. King, R. Wechselberger, W. Bermel and C. Griesinger (1994). *Journal of Biomolecular Nmr* **4**: 631-644.
- Serianni, A. S., P. B. Bondon, C. A. Podlasek, W. A. Stripe and T. J. Church (1994). *Biophysical Journal* **66**: A236-A236.
- Shaka, A. J., P. B. Barker and R. Freeman (1985). *Journal of Magnetic Resonance* **64**: 547-552.
- Shaka, A. J., C. J. Lee and A. Pines (1988). *Journal of Magnetic Resonance* **77**: 274-293.
- Stenutz, R., I. Carmichael, G. Widmalm and A. S. Serianni (2002). *Journal of Organic Chemistry* **67**: 949-958.
- Thibaudeau, C., R. Stenutz, B. Hertz, T. Klepach, S. Zhao, Q. Q. Wu, I. Carmichael and A. S. Serianni (2004). *Journal of the American Chemical Society* **126**: 15668-15685.
- Varani, G. and I. Tinoco (1991). *Journal of the American Chemical Society* **113**: 9349-9354.
- Wenter, P., B. Furtig, A. Hainard, H. Schwalbe and S. Pitsch (2005). *Angewandte Chemie-International Edition* **44**: 2600-2603.
- Wenter, P., B. Furtig, A. Hainard, H. Schwalbe and S. Pitsch (2006). *ChemBiochem* **7**: 417-420.
- Wishart, D. S., C. G. Bigam, J. Yao, F. Abildgaard, H. J. Dyson, E. Oldfield, J. L. Markley and B. D. Sykes (1995). *Journal of Biomolecular Nmr* **6**: 135-140.



## Tables

Table 1:  $^3J(\text{H},\text{H})$  coupling constants and cross correlated relaxation rates determined for the 14mer RNA. Values are given in [Hz]. <sup>#</sup>The sugar pucker rotation phase is given for a maximum amplitude of  $\nu^{\text{max}}=40^\circ$ ; \*ratio is defined as  $\Gamma_{H^3C^3',H^4'C^4'}^{DD,DD} / \Gamma_{H^1C^1',H^2'C^2'}^{DD,DD}$ .

Nuc.	P <sup>#</sup>	$\Gamma_{H^1C^1',H^2'C^2'}^{DD,DD}$	$\Gamma_{H^3C^3',H^4'C^4'}^{DD,DD}$	ratio*	$\Gamma_{H^1C^1',C^1'}^{DD,CSA}$	$^3J(\text{H}1',\text{H}2')$	$^3J(\text{H}3',\text{H}4')$
G1							
G2	10,1	-10,1 ± 0,6	13,1 ± 0,1	-0,8 ± 0,04	-2,4 ± 0,2	0,7 ± 0,4	9,9 ± 0,4
C3	12,4	-8,7 ± 0,4	15,3 ± 0,8	-0,6 ± 0,03		0,4 ± 0,4	11,5 ± 0,4
A4	13,6	-9,0 ± 0,4	13,5 ± 0,7	-0,7 ± 0,03	-2,3 ± 0,2		11,2 ± 0,4
C5	22,6	-6,8 ± 0,3			-2,5 ± 0,3	0,4 ± 0,4	10,2 ± 0,4
U6	19,1	-5,0 ± 0,3	14,5 ± 0,7	-0,3 ± 0,02	-2,0 ± 0,2	0,2 ± 0,4	11,0 ± 0,4
U7	167,3	15,8 ± 0,8	-4,5 ± -0,2	-3,5 ± 0,18	-5,0 ± 0,5	8,0 ± 0,4	1,0 ± 0,4
C8	166,4	17,3 ± 0,2	-7,0 ± -0,4	-2,5 ± 0,12	-4,0 ± 0,4	9,3 ± 0,4	0,6 ± 0,4
G9	12,4	-8,3 ± 1,9	12,6 ± 0,6	-0,7 ± 0,03	-3,1 ± 0,3	0,6 ± 0,4	7,6 ± 0,4
G10	19,3		20,7 ± 1,0		-1,8 ± 0,2	0,6 ± 0,4	
U11	12,4	-8,6 ± 0,6	12,0 ± 0,6	-0,7 ± 0,04		0,0 ± 0,4	11,3 ± 0,4
G12	13,6	-9,4 ± 0,2			-1,8 ± 0,2	1,2 ± 0,4	11,5 ± 0,4
C13	13,6	-6,9 ± 0,3			-2,9 ± 0,3	0,8 ± 0,4	9,1 ± 0,4
C14	13,6	-4,7 ± 0,0			-0,7 ± 0,2	0,3 ± 0,4	5,3 ± 0,4

Table 2: Karplus parameters according to the general Karplus formula:  $J=A*\cos^2(\theta)+B*\cos(\theta)+C$

$^1J(X,Y)$	<b>A</b>	<b>B</b>	<b>C</b>
$^1J(C1',H1')$	-13,97 $\pm$ 4,00	-4,33 $\pm$ 3,77	177,04 $\pm$ 0,64
$^1J(C2',H2')$	-6,94 $\pm$ 5,27	5,58 $\pm$ 4,67	160,17 $\pm$ 0,85
$^1J(C1',C2')$	11,40 $\pm$ 4,36	-1,65 $\pm$ 3,91	46,00 $\pm$ 0,71

## Figures

Figure 1: a) representation of the secondary structure of the RNA 14mer used in this study. b) modified 2D HSQC spectrum recorded at 600MHz and 298K.

Figure 2: The values of a)  $^3J(H,H)$  couplings, of b)  $\Gamma_{HC,HC}^{DD,DD}$  relaxation rates, and of c)  $\Gamma_{HC,C}^{DD,CSA}$  relaxation rates vs. nucleotide sequence.

Figure 3: Determination of the conformation of the furanosidic moiety in the RNA backbone a) Karplus parameterization of the  $^3J(H1',H2')$  (solid line) and the  $^3J(H3',H4')$  (dashed line) in a ribose moiety. The experimental determined values for the nucleotides of the 14mer RNA are entered as closed and open circles for  $^3J(H1',H2')$  and the  $^3J(H3',H4')$ , respectively; b) Karplus parameterization of the  $\Gamma_{H1'C1',H2'C2'}^{DD,DD}$  (solid line),  $\Gamma_{H3'C3',H4'C4'}^{DD,DD}$  (pointed line) and the ratio  $\Gamma_{H1'C1',H2'C2'}^{DD,DD} / \Gamma_{H3'C3',H4'C4'}^{DD,DD}$  (dashed line) in a ribose moiety. Experimental determined values for all nucleotides are entered (closed circles representing  $\Gamma_{H1'C1',H2'C2'}^{DD,DD}$ , open circles representing  $\Gamma_{H3'C3',H4'C4'}^{DD,DD}$  and triangles representing the ratio  $\Gamma_{H1'C1',H2'C2'}^{DD,DD} / \Gamma_{H3'C3',H4'C4'}^{DD,DD}$ ). In both figures a) and b) the region for C3'-endo and C2'-endo conformations are highlighted in gray. The two nucleotides U7 and C8 sampling the C2'-endo conformational space are annotated.

Figure 4: a)  $^1J(C,H)$  coupling constants. Closed circles representing  $^1J(C1',H1')$  and open circles  $^1J(C2',H2')$  coupling constants. Dashed lines indicating the mean value for the stem residues, black lines representing the corresponding root mean square deviation; b)  $^1J(C1'C2')$  coupling constants. Dashed line indicates the mean value for the stem residues, black lines representing the corresponding root mean square deviation. The nucleotide G9 adopting syn conformation around the  $\chi$  angle is indicated by an arrow.

Figure 5: Karplus Plot for a)  $^1J(C2',H2')$ , b)  $^1J(C1',H1')$  and c)  $^1J(C1',C2')$  versus the  $\phi(H1',H2')$  dihedral angle. Solid lines representing the fitted Karplus curve, whereas the dashed lines representing the error bonds for the fitting. In a) and b) histogram plot representing the distribution of the  $\phi(H1',H2')$  in the rRNA molecules. (in canonical A-form helices this angle adopts a value of  $96^\circ$  whereas in B-form the angle is  $158,8^\circ$ ) The value of  $^1J(C1',H1')$  for the nucleotide G9 is circled in gray, because it is the nucleotide that shows additional influence of this coupling by the conformation of the  $\chi$  angle.

Figure 6: Comparison of  $^3J(H1'H2')$ ,  $^{13}C$  chemical shift (can1), where can1 is defined as  $0.179 \cdot \delta_{C1'} - 0.225 \cdot \delta_{C4'} - 0.0585 \cdot \delta_{C5'}$  and  $^1J(C2'H2')$  data, as a measure of sugar pucker conformation, crosshairs representing the data triplet for each residue (annotated), red triangles representing the midpoint of each cluster.

Figure 1

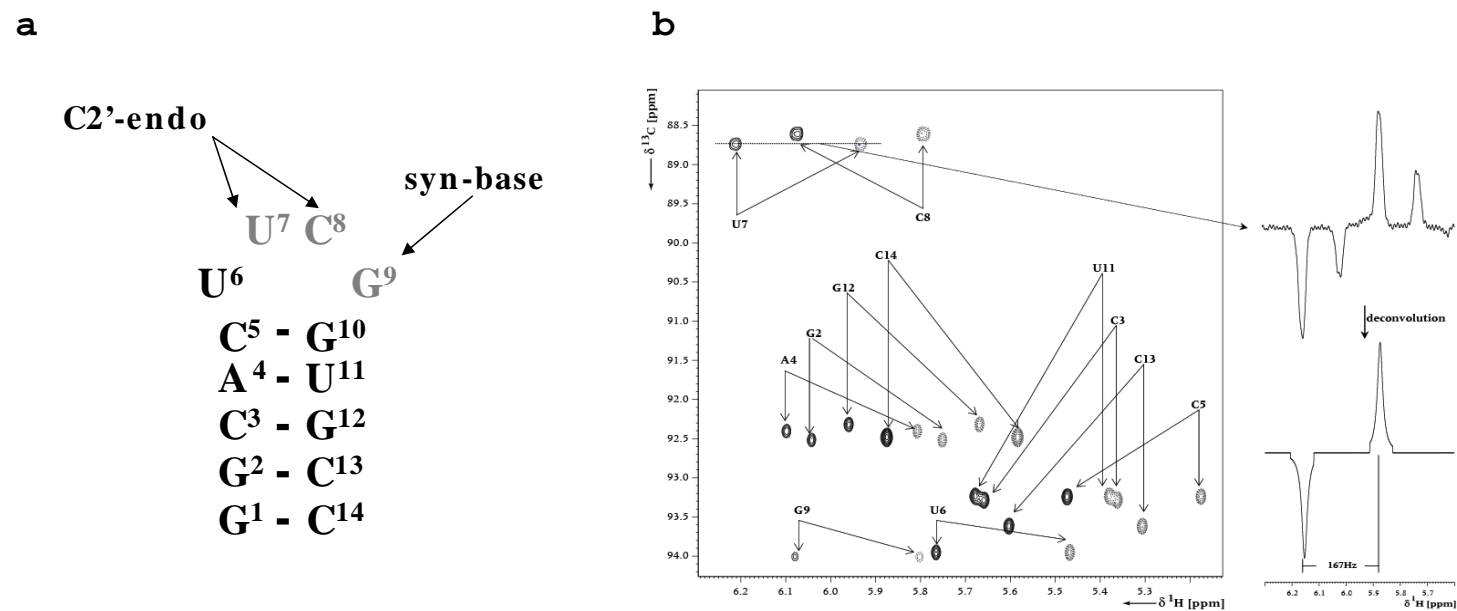


Figure 2

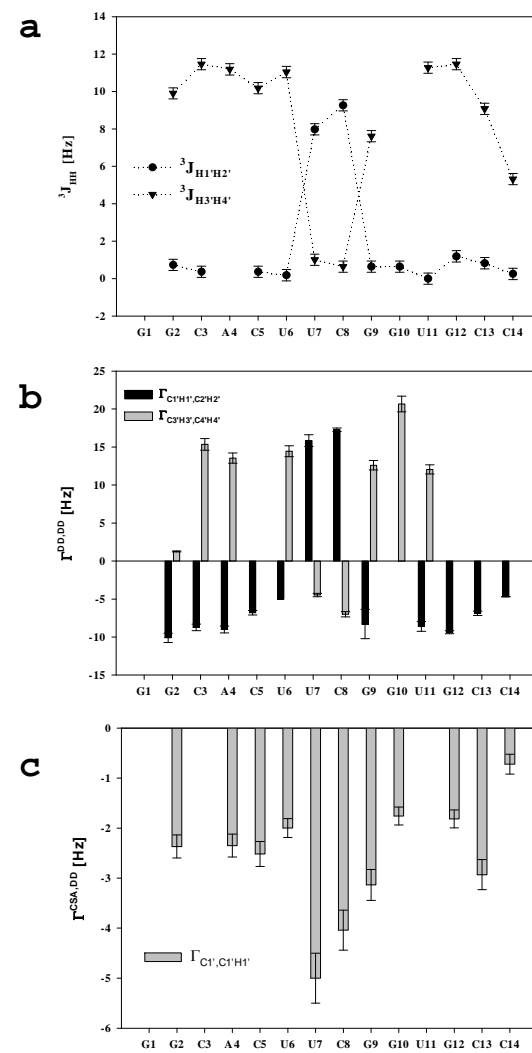


Figure 3

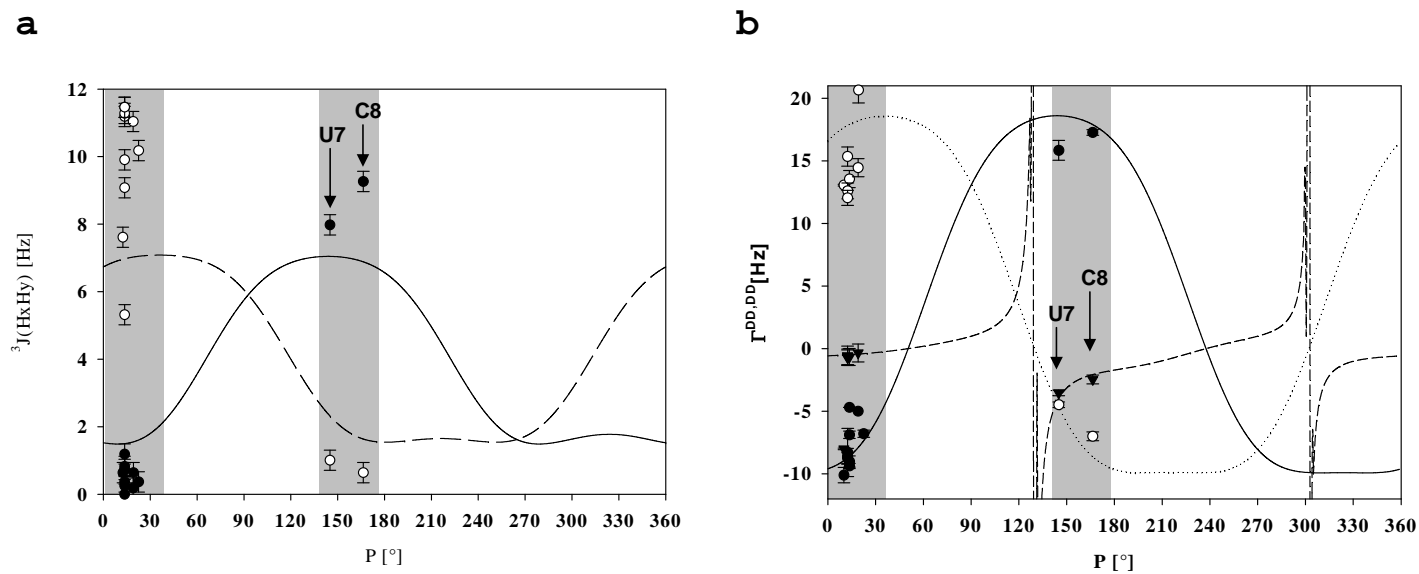


Figure 4

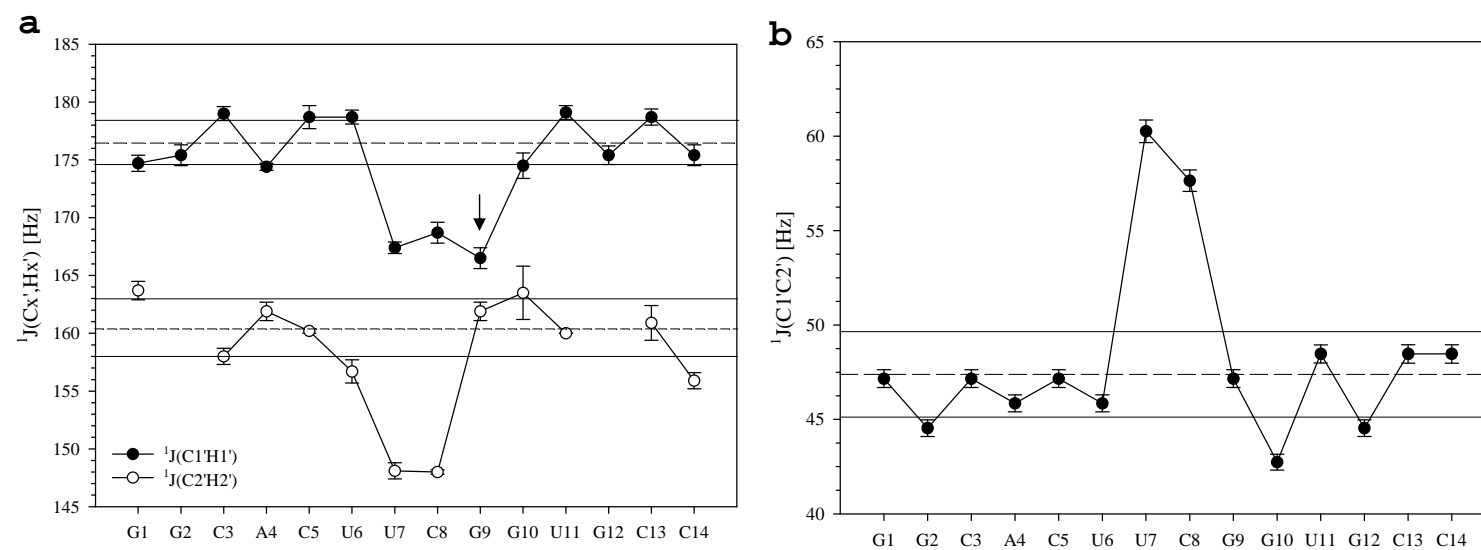


Figure 5

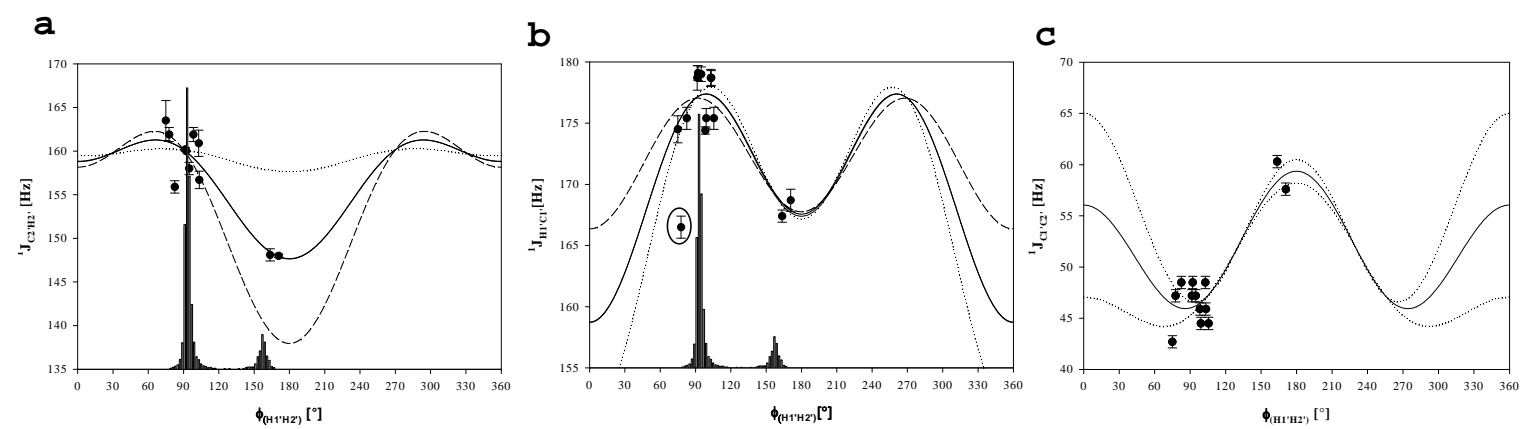
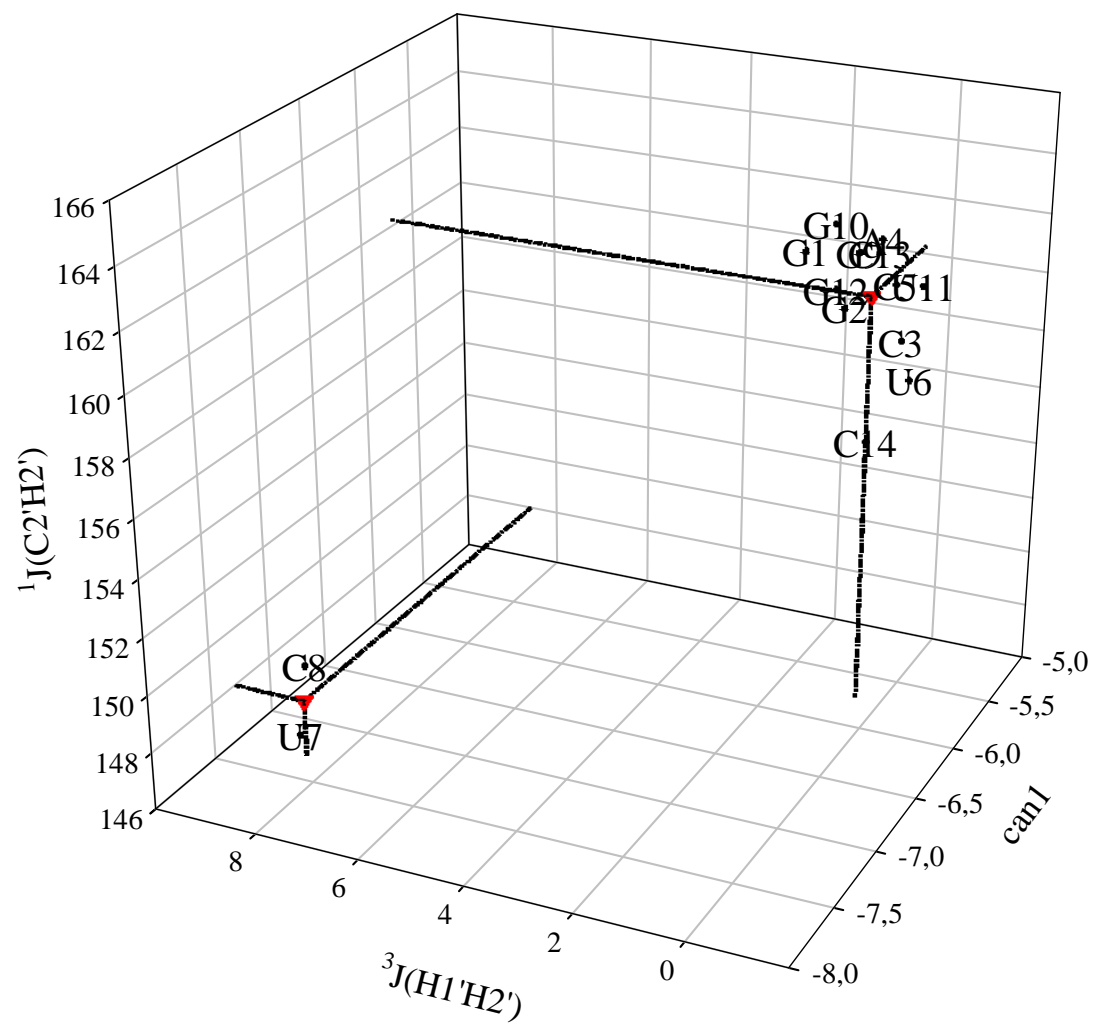




Figure 6



---

## RNA-FOLDING I: CONFORMATIONAL SWITCHING OF RNA SECONDARY STRUCTURE ELEMENTS

**RESEARCH ARTICLE:** KINETICS OF PHOTOINDUCED

RNA REFOLDING BY REAL-TIME NMR SPECTROSCOPY

Philipp Wenter, Boris Fürtig, Alexandre Hainard, Harald Schwalbe, Stefan Pitsch, *Angew. Chem. Int. Ed.*, 2005, **44**, 2600-2603**RESEARCH ARTICLE:** A CAGED URIDINE FOR THE SELECTIVEPREPARATION OF AN RNA FOLD AND DETERMINATION OF  
ITS REFOLDING KINETICS BY REAL-TIME NMRPhilipp Wenter, Boris Fürtig, Alexandre Hainard, Harald Schwalbe, Stefan Pitsch, *ChemBioChem*, 2006, **7**, 417-420**RESEARCH ARTICLE:** KINETICS AND THERMODYNAMICS

OF RNA CONFORMATIONAL SWITCHING

Philipp Wenter, Boris Fürtig, Christian Richter, Alexandre Hainard,  
Harald Schwalbe, Stefan Pitsch, *submitted***REVIEW ARTICLE:** TIME-RESOLVED NMR STUDIES OF RNA FOLDINGBoris Fürtig, Janina Buck, Vijayalaxmi Manoharan, Wolfgang Bermel,  
Andres Jäscke, Philipp Wenter, Stefan Pitsch, Harald Schwalbe,  
*Biopolymers*, 2007, **RNA special issue**, *in print*[IN THE CONTEXT OF THIS THESIS JUST THE ABSTRACT IS LISTED, BECAUSE  
THE ENTIRE ARTICLE IS BASED ON THE CHAPTERS III AND IV]

All three papers deal with conformational switching events of secondary structural elements of RNA analyzed by RT-NMR. The molecules used in the three reports were bistable RNA molecules acting as model systems for RNA refolding events.

The first paper introduces the newly developed technique that combines the application of NMR methodology with the preparation of conformational states of RNA by introduction of a photo-labile, synthetically derived protecting group. Application of this new investigation strategy leads to the first description of an RNA refolding mechanism at atomic resolution.

All synthesis work was done by our co-workers in Lausanne, all NMR experiments, data processing, thermodynamic and mechanistically characterisation was done by the author of the thesis.



# Kinetics of Photoinduced RNA Refolding by Real-Time NMR Spectroscopy\*\*

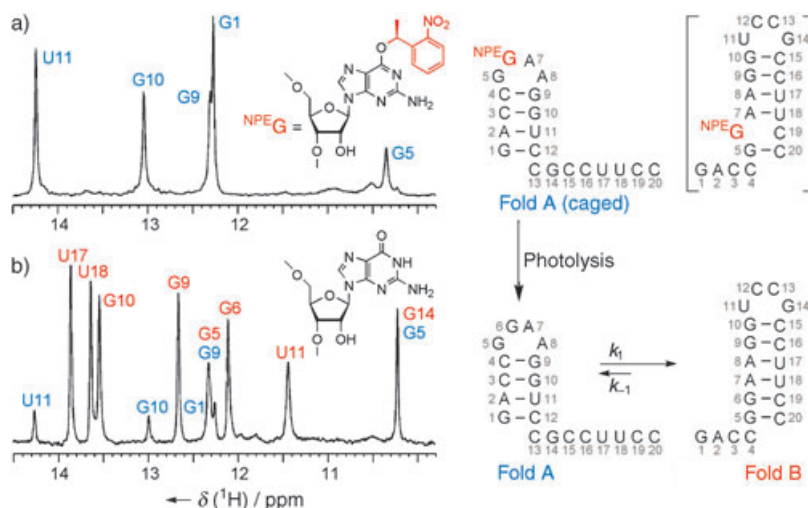
Philipp Wenter, Boris Fürtig, Alexandre Hainard, Harald Schwalbe,\* and Stefan Pitsch\*

The catalytic and regulatory functions of RNA strongly depend on the reorganization of the tertiary structure and the refolding of secondary structures.<sup>[1]</sup> This dynamic diversity is illustrated by the frequent occurrence of multiple folding pathways,<sup>[2]</sup> as well as metastable<sup>[3]</sup> and coexisting (“bistable”) conformations.<sup>[4]</sup> The folding of RNA from a denatured, non-native state has been studied extensively by a variety of techniques,<sup>[5]</sup> which have established a hierarchical folding pathway in which the rapid formation of secondary structural elements precedes the slower formation of tertiary elements.<sup>[6]</sup> In contrast, only a few time-resolved studies of RNA refolding from an unperturbed native state have been described.<sup>[7]</sup>

Herein we report a new method for the investigation of structure- and time-resolved RNA refolding that is based on the photolytic generation of preselected conformations in a nonequilibrium state, followed by the observation of refolding with real-time NMR spectroscopy. This method was applied to the detailed kinetic characterization of a bistable 20-base RNA sequence, which is known to adopt two coexisting hairpin-loop structures.<sup>[8]</sup>

The 20-base RNA sequence 5'-r[GACCGGAAG-GUCCGCCUUCC]-3' forms two interconverting hairpin

structures in a temperature-dependent equilibrium (Folds A and B, Figure 1b). To selectively destabilize the more stable Fold B, the selected guanosine groups that form base pairs in Fold B, but not in Fold A, were replaced by *O*<sup>6</sup>-[(*S*)-1-(2-nitrophenyl)ethyl] guanosine ((*S*)-<sup>NPE</sup>G, Figure 1a).<sup>[9]</sup> The bulky NPE group was expected to disrupt Watson–Crick G–C base-pair formation, thereby destabilizing associated sec-



**Figure 1.** <sup>1</sup>H NMR spectra of imino protons and folding structures of the 20-base RNA sequence (*c* = 0.27 mm, 288 K) with color-coded signal assignments; imino proton NMR spectroscopy reveals individual signals for each proton involved in the central H bond of Watson–Crick base pairs. a) Spectrum of the sequence that is conformationally locked by the *O*<sup>6</sup>-(*S*)-NPE modified guanosine at position 6 (<sup>NPE</sup>G6, structure shown as inset); only the set of signals of Fold A are apparent. b) Spectrum obtained after photolytic removal of the NPE group by three laser pulses and subsequent equilibration for 1 h, which shows the two sets of signals of Folds A and B in a 1:5 ratio.

ondary structures; upon subsequent photolytic cleavage, the base-pairing properties of the parent guanosine would be restored to allow refolding to take place.<sup>[10]</sup>

The RNA sequence 5'-r[GACCG(<sup>(S)</sup>-NPEG)AAG-GUCCGCCUUCC]-3', which contains a single (*S*)-NPE modification on G6, showed the characteristic five imino proton signals exclusively for Fold A (Figure 1a and Figure S1, Supporting Information).<sup>[11]</sup> Further 2D-HSQC and HNN-COSY experiments on the same sequence labeled with <sup>15</sup>N at N(3) of C3 and N(1) of G10 showed that G10 is involved in only one base-pair interaction, and confirmed that its pairing partner is C3 (Figure S2, Supporting Information). These experiments unambiguously show that the introduction of only one photolabile NPE group leads to a complete shift of the conformational equilibrium towards the intrinsically less stable Fold A. The broadened imino proton signal of G9, however, indicates a perturbation of the hairpin-loop structure.

The imino proton spectrum shown in Figure 1b was generated from a sample of the (*S*)-NPEG6-modified, <sup>15</sup>N-labeled RNA sequence (Figure 1a) photolyzed by three laser pulses directed into the NMR tube through a glass fiber and a pencil-like quartz tip,<sup>[12]</sup> and is identical to the spectrum of the unmodified sequence (Figures S1 and S2, Supporting Information); HPLC and MS analyses confirmed the clean and

[\*] P. Wenter, A. Hainard, Prof. Dr. S. Pitsch  
École Polytechnique Fédérale de Lausanne  
Laboratory of Nucleic Acid Chemistry, EPFL-BCH  
1015 Lausanne (Switzerland)  
Fax: (+41) 21-693-9355  
E-mail: stefan.pitsch@epfl.ch

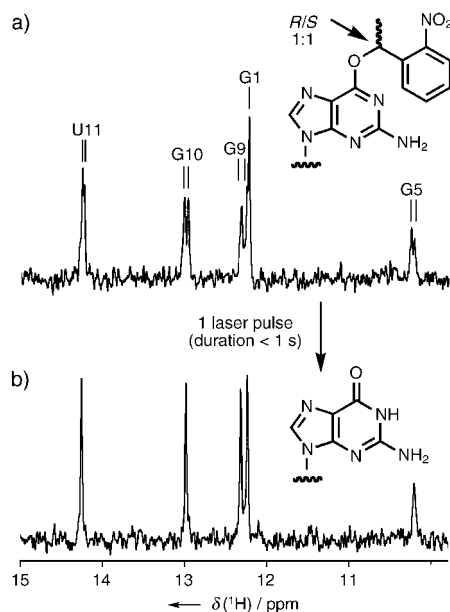
B. Fürtig, Prof. Dr. H. Schwalbe  
Johann Wolfgang Goethe University  
Institute for Organic Chemistry and Chemical Biology  
Center for Biological Magnetic Resonance  
Marie-Curie Strasse 11, 60439 Frankfurt/Main (Germany)  
Fax: (+49) 69-7982-9515  
E-mail: schwalbe@nmr.uni-frankfurt.de

[\*\*] We are grateful to Jens Wöhnert, Julia Wirmer, and Christian Richter for discussions, and to Rita Merz for conducting syntheses. The work was supported by the EPFL, the Swiss National Science Foundation (No. 2000-068090), the SFB 579 “RNA-Ligand-Wechselwirkung”, and the “Studienstiftung des Deutschen Volkes” (B.F.).

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.

traceless removal (>98%) of the NPE group (Figure S3, Supporting Information).

Single-scan imino proton NMR spectra of the (*S*)-NPE-modified RNA sequence both before and 1 s after photolysis by a single laser pulse were nearly identical, which indicates that the sequence retained its preformed folded structure after release of the NPE group. To directly monitor the course of the cleavage reaction, identical experiments were carried out with a 1:1 mixture of two diastereomeric 20-base RNA oligomers that were obtained by the introduction of an (*R/S*)-NPE-protected guanosine residue. The imino proton signal splitting caused by the two configurations of the photolabile group (Figure 2a) collapsed to a single set within 1 s after



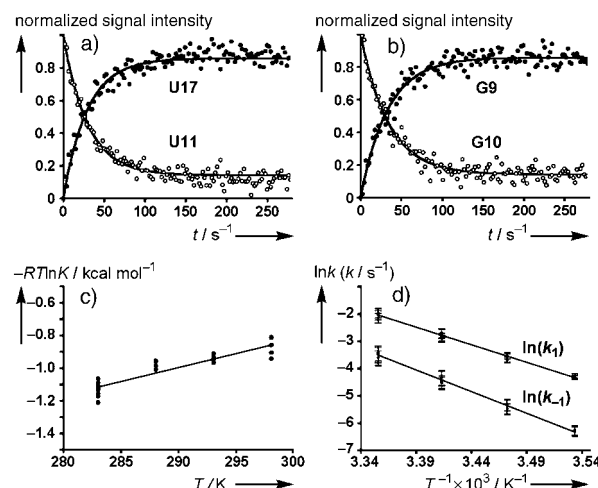
**Figure 2.** The course of the photolysis reaction at  $c=0.27$  mM, as monitored by single-scan  $^1\text{H}$  NMR spectra of the 20-base RNA sequence initially modified at G6 with a 1:1 mixture of (*S*)- and (*R*)-NPE groups. a) Before photolysis; the peak splitting results from the two different NPE configurations. b) Immediately after a single laser pulse (duration < 1 s); the initial peak splitting disappears as a result of the photolytic removal of the NPE group.

photolysis by one laser pulse (Figure 2b). Clearly, the folded conformation initially imposed by the photolabile group was completely conserved immediately after its photolytic removal. Therefore, this method was successful for the selective preparation of one of the two folded structures that normally coexist for this sequence.

For the determination of kinetic parameters, two sets of experiments were carried out: one in which the RNA concentration was varied ( $c=0.17$ ,  $0.26$ , and  $0.34$  mM;  $T=283$  K) and the other in which temperature was the variable ( $T=283$ ,  $288$ ,  $293$ , and  $298$  K;  $c=0.26$  mM). After one laser pulse, 128 single-scan NMR spectra were recorded over a time period of 280 s during which equilibrium was established. The analysis was carried out with the resolved imino proton signals of U11/G10 and U17/G9; the former two

correspond to the disappearing Fold A and the latter correspond to the appearing Fold B.

The equilibrium constants  $K$  were obtained from the relative intensities of the pairs of imino proton signals of U11/U17 and G9/G10. From the temperature dependence of the  $K$  values, which range between  $K\approx 4$  at 298 K and  $K\approx 7$  at 283 K, the values of  $\Delta H^\ddagger = -5.9 \pm 0.5$  kcal mol $^{-1}$  and  $\Delta S^\ddagger = -17.0 \pm 1.9$  cal mol $^{-1}$  K $^{-1}$  were determined within a confidence interval (CI) of 90% (Figure 3c). The refolding rate constants  $k_1$  were determined by fitting the normalized build-up and decay data for U17/G9 and U11/G10 signals, respectively, to the rate law for a reversible unimolecular reaction (Figure 3a,b; Figure S4, Supporting Information).



**Figure 3.** Kinetic analysis of the refolding reaction monitored through time-dependent single-scan NMR spectroscopic experiments with the photolyzed 20-base RNA sequence. a,b) Normalized imino proton signal intensities (U11/G10 from Fold A; U17/G9 from Fold B) as a function of time, recorded at 288 K after photolysis by a single laser pulse. The data curves were fit to the rate law of a reversible, unimolecular reaction (See Supporting Information for detailed procedures and fits at other temperatures). c) Linearized plot of  $K = [\text{Fold B}]/[\text{Fold A}]$  according to  $\Delta G = -RT \ln K = \Delta H - T\Delta S$ . d) Arrhenius plot of  $k_1$  and  $k_{-1}$  according to  $\ln k = -\Delta H^\ddagger/(RT) + \ln A$ ; the error bars (shown at 99% CI) include the  $K$ -standard error of 10% and the curve-fit standard deviation.

The mean rate constants for the forward reaction ranged from  $k_1 = 0.0136 \pm 0.0010$  s $^{-1}$  at 283 K to  $k_1 = 0.131 \pm 0.024$  s $^{-1}$  at 298 K, and those for the reverse reaction from  $k_{-1} = 0.0019 \pm 0.0002$  s $^{-1}$  at 283 K to  $k_{-1} = 0.031 \pm 0.006$  s $^{-1}$  at 298 K (90% CI). These values were fit with the Arrhenius equation to yield the activation enthalpy  $\Delta H^\ddagger = 25.5 \pm 3.1$  kcal mol $^{-1}$  and the frequency factor  $A = 10^{(17.8 \pm 1.2)}$  s $^{-1}$  for the forward reaction within 90% CI. The respective values for the reverse reaction were  $\Delta H^\ddagger = 30.6 \pm 3.1$  kcal mol $^{-1}$  and  $A = 10^{(20.9 \pm 1.4)}$  s $^{-1}$  (Figure 3d). The rate constants determined at different concentrations were identical within 90% CI (for details, see Tables S5 and S6, Supporting Information).

These results suggest a pure first-order refolding mechanism. The activation enthalpy  $\Delta H^\ddagger = 25.5$  kcal mol $^{-1}$  amounts to about half the value for the entire base-pairing enthalpy

$\Delta H = 55 \text{ kcal mol}^{-1}$ , which was determined by thermal denaturation studies.<sup>[8]</sup> This indicates that the rate-limiting step of the refolding mechanism involves disruption of about half the number of base pairs present.

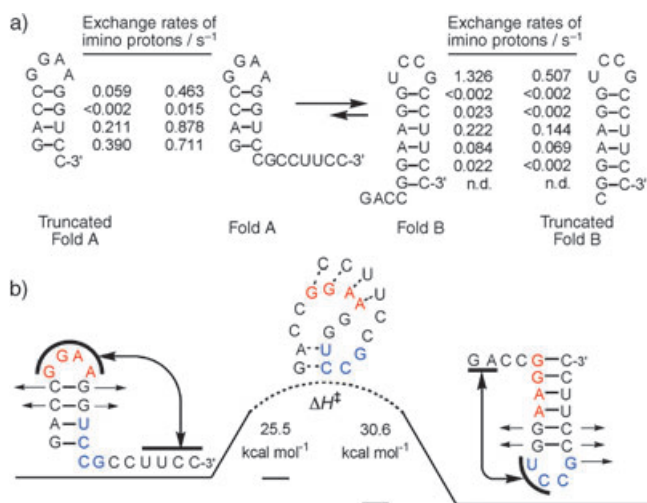
Imino proton exchange rates with water were extracted from NOESY cross-peaks to obtain more information on the influence that global dynamics (interconverting folds) have on the local dynamic properties of individual base pairs.<sup>[13]</sup> These values were compared with those of the two corresponding truncated RNA hairpins that are conformationally locked to represent Fold A and Fold B, as they lack unpaired overhangs (Figure 4). In general, the individual water

refolding of a 20-base bistable RNA sequence. The refolding rate constants are at least four orders of magnitude less than those measured for hairpin formation in similar sequences from an unfolded state ( $k = 10^4\text{--}10^5 \text{ s}^{-1}$ ).<sup>[6]</sup> This new method might be extended toward the targeted preparation and kinetic characterization of distinct intermediates that are present along the complex folding pathways of larger RNA sequences.

Received: November 25, 2004

Published online: March 22, 2005

**Keywords:** kinetics · NMR spectroscopy · oligonucleotides · photolysis · RNA structures



**Figure 4.** a) Exchange rates between base-pair imino protons and water, as determined in both folds of the bistable 20-base RNA sequence and the corresponding truncated, conformationally locked hairpin sequence; the detection level was  $0.002 \text{ s}^{-1}$  and the accuracy was estimated at  $\pm 5\%$ ; (n.d. = not detected).<sup>[13]</sup> b) Proposed associative refolding mechanism consistent with water exchange rates and kinetic parameters. During the course of refolding, which presumably involves several elementary steps,<sup>[14]</sup> transient structures could be formed in which the terminal base pairs of the stems of both Folds A and B are present simultaneously, as indicated schematically. The loop-forming nucleotides are indicated in color for clarity.

exchange rates within the bistable RNA sequence were significantly higher than the rates within the two truncated hairpin sequences (Figure 4a), which demonstrates that base pairs in the interconverting RNA structures are less rigid than the identical base pairs in the static structures. Interestingly, the increase in the water-exchange rates of the 20-base RNA sequence relative to those of the truncated hairpins became more important the closer the base pairs were to the loop region.

This result suggests an associative mechanism of refolding, which is proposed schematically in Figure 4b. In this mechanism, the refolding would be initiated by the formation of new base pairs between complementary nucleotides of the loop and the dangling ends, with a simultaneous detachment of the base pairs close to the loop.

With the method presented herein, we were able to accurately characterize the structure- and time-resolved

- a) N. G. Walter, D. A. Harris, M. J. B. Pereira, D. Rueda, *Biopolymers* **2002**, 61, 224–241; b) W. C. Winkler, R. R. Breaker, *ChemBioChem* **2001**, 2, 1024–1032; c) P. Brion, E. Westhof, *Annu. Rev. Biophys. Biomol. Struct.* **1997**, 26, 113–137.
- J. Pan, D. Thirumalai, S. A. Woodson, *J. Mol. Biol.* **1997**, 273, 7–13.
- D. K. Treiber, M. S. Rook, P. P. Zarrinkar, J. R. Williamson, *Science* **1998**, 279, 1943–1946.
- C. Höbartner, R. Micura, *J. Mol. Biol.* **2003**, 325, 421–431.
- a) P. P. Zarrinkar, J. R. Williamson, *Science* **1994**, 265, 918–924; b) X. Zhuang, L. E. Bartley, H. P. Babcock, R. Russell, T. Ha, D. Herschlag, S. Chu, *Science* **2000**, 288, 2048–2051; c) B. Sclavi, M. Sullivan, M. R. Chance, M. Brenowitz, S. A. Woodson, *Science* **1998**, 279, 1940–1943; d) J. Ravetch, J. Gralla, D. M. Crothers, *Nucleic Acids Res.* **1974**, 1, 109–127; e) M. Menger, F. Eckstein, D. Porschke, *Biochemistry* **2000**, 39, 4500–4507.
- a) D. Porschke, *Mol. Biol. Biochem. Biophys.* **1977**, 24, 191–218; b) D. M. Crothers, P. E. Cole, C. W. Hilbers, P. G. Shulman, *J. Mol. Biol.* **1974**, 87, 63–88; c) D. Thirumalai, S. A. Woodson, *RNA* **2000**, 6, 790–794; d) D. Thirumalai, N. Lee, S. A. Woodson, D. K. Klimov, *Annu. Rev. Phys. Chem.* **2001**, 52, 751–762; e) I. Tinoco, Jr., C. Bustamante, *J. Mol. Biol.* **1999**, 293, 271–281.
- a) T. C. Gluick, R. B. Gerstner, D. E. Draper, *J. Mol. Biol.* **1997**, 270, 451–463; b) D. C. Lynch, P. R. Schimmel, *Biochemistry* **1974**, 13, 1841–1852; c) D. C. Lynch, P. R. Schimmel, *Biochemistry* **1974**, 13, 1852–1861.
- C. Höbartner, M. O. Ebert, B. Jaun, R. Micura, *Angew. Chem.* **2002**, 114, 619–623; *Angew. Chem. Int. Ed.* **2002**, 41, 605–609.
- The synthesis of the <sup>(5)</sup>-NPE-G-phosphoramidite is described in the Supporting Information; the S configuration of the NPE group was chosen arbitrarily. All RNA sequences were prepared from 2'-O-triisopropylsilyloxymethyl-protected phosphoramidites according to a) S. Pitsch, P. A. Weiss, L. Jenny, A. Stutz, X. Wu, *Helv. Chim. Acta* **2001**, 84, 3773–3795; the synthesis of the <sup>15</sup>N-labeled phosphoramidites is reported in b) P. Wenter, S. Pitsch, *Helv. Chim. Acta* **2003**, 86, 3955–3974.
- A similar approach for the selective destabilization of coexisting RNA secondary structures with O<sup>6</sup>-(2,2,2-trichloroethyl)-modified guanosine was reported by C. Höbartner, H. Mittendorfer, K. Breuker, R. Micura, *Angew. Chem.* **2004**, 116, 4012–4015; *Angew. Chem. Int. Ed.* **2004**, 43, 3922–3925; however, removal of the Zn-labile trichloroethyl group is much slower than RNA refolding processes and thus prevents the determination of kinetic parameters.
- All NMR spectroscopic experiments were carried out with a Bruker AV 800 MHz spectrometer, equipped with 5-mm TXI-HCN cryogenic probes and z-axis gradients; detailed NMR spectroscopic methods are described in the Supporting Information; all experiments were carried out in H<sub>2</sub>O/D<sub>2</sub>O (9:1) and

potassium arsenate buffer (25 mM, pH 7.0; identical results were obtained with sodium and potassium phosphate buffers in preliminary experiments). The 2D-HSQC and HNN-COSY spectra were recorded according to a) S. Mori, C. Abeygunawardana, M. O'Neil-Johnson, P. C. M. Vanzijl, *J. Magn. Reson. B* **1995**, *108*, 94–98 and b) A. J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.* **1998**, *120*, 8293–8297.

- [12] T. Kühn, H. Schwalbe, *J. Am. Chem. Soc.* **2000**, *122*, 6169–6174.
- [13] Water exchange rates were extracted according to R. R. Ernst, G. Bodenhausen, A. Wokaun, *International Series of Monographs on Chemistry, Vol. 14*, Oxford University Press, **1987**; further details are given in the Supporting Information. In principle, overlapping signals at the resonance frequency of  $^1\text{H}_{\text{H}_2\text{O}}$  could influence the intensity of the exchange-induced water–imino cross-peaks in the 2D NOESY experiment; however, the ribose protons are too far away from the imino protons. Hydroxy group protons are in exchange with water and could potentially contribute to the cross-peak intensity; again, the effect is made negligible by long distance. Base protons resonate upfield and cannot contribute to the cross-peak intensity.
- [14] a) C. Flamm, W. Fontana, I. L. Hofacker, P. Schuster, *RNA* **2000**, *6*, 325–338; b) W. Zhang, S. J. Chen, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1931–1936.



DOI: 10.1002/cbic.200500468

# A Caged Uridine for the Selective Preparation of an RNA Fold and Determination of its Refolding Kinetics by Real-Time NMR

Philipp Wenter,<sup>[a]</sup> Boris Fürtig,<sup>[b]</sup> Alexandre Hainard,<sup>[a]</sup> Harald Schwalbe,<sup>\*,[b]</sup> and Stefan Pitsch<sup>\*,[a]</sup>

The biological function of many RNAs is linked to a reversible conformational switching between active and inactive folds that differ in their secondary and/or tertiary structures.<sup>[1]</sup> Regulatory RNAs, such as riboswitches,<sup>[2,3]</sup> and catalytic RNAs, such as ribozymes,<sup>[4]</sup> undergo this conformational switching upon binding of metabolites or during catalysis. In general, the energetic differences between the alternative conformations are only small, and therefore the equilibrium distribution is strongly affected by ligand or substrate binding, or by small structural modifications, such as methylation of the nucleobases.<sup>[5]</sup>

Recently, we have introduced a method for studying the kinetics of RNA-conformational switching by time-resolved NMR spectroscopy.<sup>[6]</sup> The method is based on RNA sequences modified with 1-(2-nitrophenyl)ethyl-substituted (NPE-substituted) nucleobases, which are designed to impair the formation of selected base pairs. When introduced at proper sequence position within the oligonucleotide fold, these photocleavable modifications lead to a specific destabilization of preselected folds. Rapid and traceless removal of the photolabile groups, carried out by coupling laser optics into the NMR tube, then allows NMR detection of the refolding process. The induced structural transition can be monitored from the time-dependent increase and decrease in intensity of imino proton resonances, which belong to specific base pairs present in specific folds. For medium-sized oligonucleotides, these resonances often show sufficient resolution in chemical shifts to be resolved in 1D experiments.

In this context, we have already reported the refolding kinetics and a plausible refolding mechanism of the bistable 20-mer RNA sequence 5'-r[GACCGGAAGGUCCGUCCUCC]-3'. For this study, an (S)-NPE-substituted guanosine was introduced at position 6. The activation energy was  $E_a = 26 \text{ kcal mol}^{-1}$  and the

frequency factor was  $A = 10^{18} \text{ s}^{-1}$  for the forward process, or  $31 \text{ kcal mol}^{-1}$  and  $10^{21} \text{ s}^{-1}$  for the backward process, respectively. These refolding rates are surprisingly low compared to the folding of similar systems from an unfolded state<sup>[7]</sup> and are in agreement with a transient disruption of half of the base pairs. Together with an additional dynamical characterization by comparative water-exchange experiments, an associative mechanism was deduced, in which refolding is initiated by base-pair contacts between previously unpaired nucleotides.

In this communication, we extend our real-time-NMR investigation of RNA secondary structural transitions to another bistable, 20-mer RNA sequence 5'-r(GAAGGGCAACCUUCGGGUUG)-3' that was designed by Höbartner et al.<sup>[8]</sup> This RNA sequence adopts two almost equally stable hairpin folds A and B, each consisting of a five-base-pair stem, a stable tetraloop and six unpaired nucleotides (Figure 1C, D). In order to selectively destabilize fold B, we introduced an (R)-NPE-protected uridine at position U18 (Figure 1B).<sup>[9]</sup> Structural and kinetic characterization were carried out by analyzing the imino proton NMR signals, each representing a specific base pair.<sup>[10]</sup>

The unmodified sequence shows 12 signals in the imino-proton region (Figure 1D), which were assigned to the base pairs of the two coexisting folds A and B by  $^1\text{H}$ ,  $^1\text{H}$ -NOESY experiments. The signals assigned to fold A agreed well with the signals of the corresponding truncated 15-mer RNA hairpin (Figure 1A). After introduction of the (R)-NPE-protected U18 into the 20-mer RNA sequence, only the set of six imino proton signals characteristic of fold A were detected (Figure 1B). Importantly, this single NPE-modification completely suppressed the formation of fold B by preventing the formation of the associated U18–A9 base pair. Removal of the NPE group by photolysis and equilibration resulted in a NMR spectrum (Figure 1C) that was practically identical to the spectrum of the parent sequence (Figure 1D).

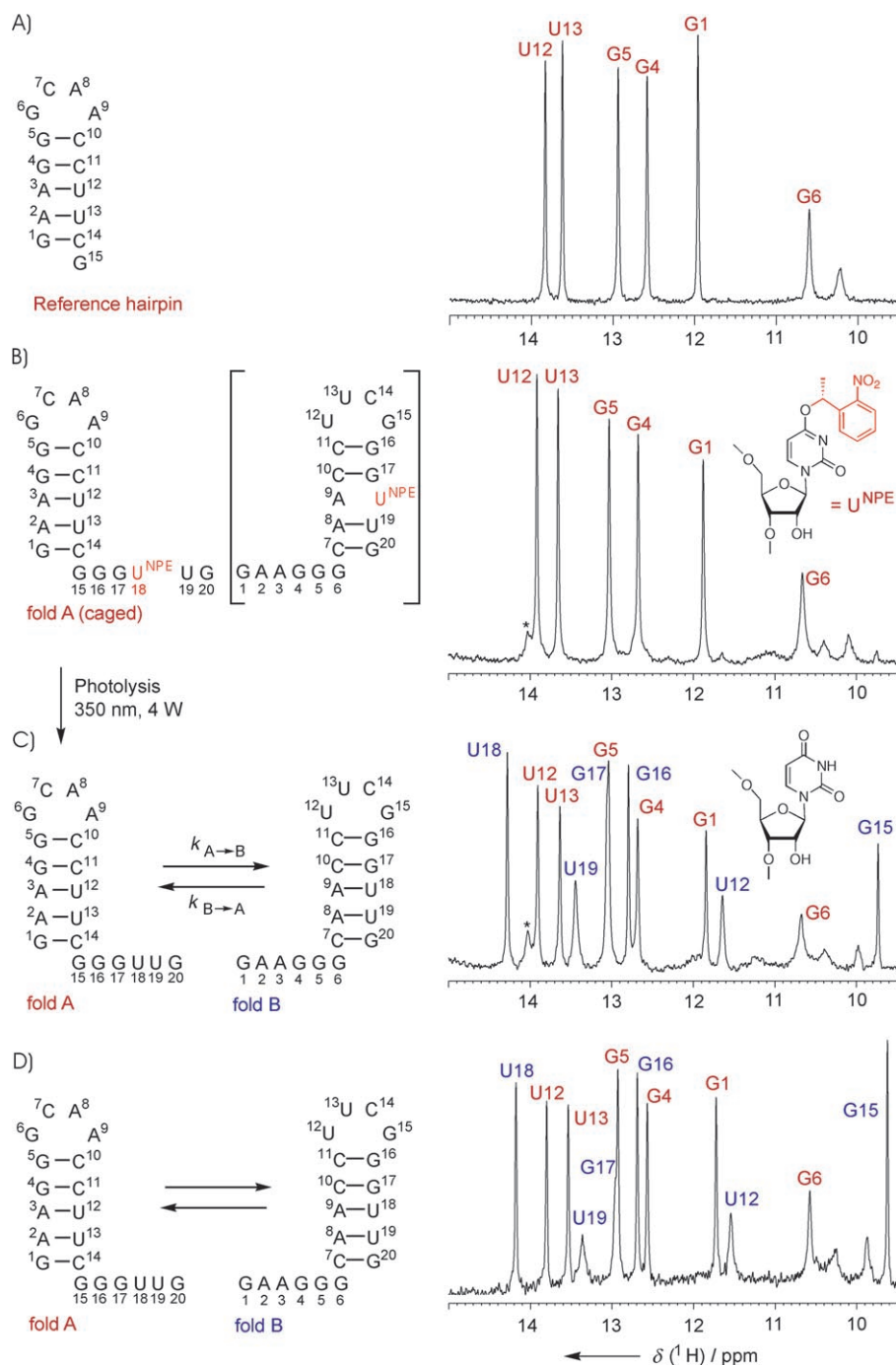
To determine the kinetic parameters for refolding, the NPE group was released by a short continuous-wave laser pulse (duration 1.5 s) guided by an optical fibre into the NMR tube. Relaxation of the system towards equilibrium was monitored by the subsequent recording of 1D  $^1\text{H}$  spectra.<sup>[6,11]</sup> The refolding rate constants were obtained by fitting the time-dependent intensities of two selected pairs of imino proton signals according to the equation for a reversible first-order process (Figure 2A–D). The independently determined equilibrium constant  $K_{A \rightarrow B}$  was employed as an additional fit parameter. The thermodynamic parameters  $\Delta H_{A \rightarrow B} = 1.6 \pm 0.3 \text{ kcal mol}^{-1}$  and  $\Delta S_{A \rightarrow B} = 6.0 \pm 1.0 \text{ cal mol}^{-1} \text{ K}^{-1}$  were determined from the temperature dependence of these  $K$  values (Figure S1 in the Supporting Information). The directly determined rate constants  $k_{A \rightarrow B}$  ranged between  $0.005 \pm 0.001 \text{ s}^{-1}$  at 283 K and  $0.091 \pm 0.020 \text{ s}^{-1}$  at 298 K; the rate constants  $k_{B \rightarrow A}$  were calculated from the  $k_{A \rightarrow B}$  and  $K$  values, and ranged between  $0.005 \pm 0.002 \text{ s}^{-1}$  at 283 K and  $0.076 \pm 0.020 \text{ s}^{-1}$  at 298 K. In each experiment, rates that were identical within the error range were determined for the two independently analyzed pairs of imino proton signals (all rates are given in Figure S2 in the Supporting Information). Arrhenius analysis revealed an activation energy of  $E_{a, A \rightarrow B} = 32.7 \pm 3.0 \text{ kcal mol}^{-1}$  ( $E_{a, B \rightarrow A} = 31.0 \pm$

[a] P. Wenter, A. Hainard, Prof. Dr. S. Pitsch  
Laboratory of Nucleic Acid Chemistry  
Ecole Polytechnique Fédérale de Lausanne (EPFL)  
1015 Lausanne (Switzerland)  
Fax: (+41) 21-693-9355  
E-mail: stefan.pitsch@epfl.ch

[b] B. Fürtig, Prof. Dr. H. Schwalbe  
Johann Wolfgang Goethe University  
Institute for Organic Chemistry and Chemical Biology  
Center for Biomolecular Magnetic Resonance  
Marie-Curie Strasse 11, Frankfurt, 60439 Frankfurt/Main (Germany)  
Fax: (+49) 69-7982-9515  
E-mail: schwalbe@nmr.uni-frankfurt.de

Supporting information for this article is available on the WWW under <http://www.chembiochem.org> or from the author.





**Figure 1.**  $^1\text{H}$  NMR spectra of imino proton signals recorded for the indicated RNA structures at 288 K.<sup>[10]</sup> The colour-coded assignments were obtained from 2D-NOESY spectra. \* denotes a non-RNA impurity. A) Truncated 15-mer RNA hairpin representing fold A. B) Caged bistable 20-mer RNA sequence containing an  $O^4$ -(*R*)-NPE-modified uridine at position 18 (structure shown as inset<sup>[9]</sup>) displaying only the set of signals of fold A. C) Spectrum after photolytic removal of the NPE shown by three laser pulses and subsequent equilibration for 1 h, showing two sets of signals, assigned to fold A and B, in a 1:1.15 ratio. D) Independently synthesized, parent 20-mer RNA sequence.

$3.0 \text{ kcal mol}^{-1}$ ) and a frequency factor of  $A_{A \rightarrow B} = 10^{(22.6 \pm 2.4)} \text{ s}^{-1}$  ( $A_{B \rightarrow A} = 10^{(21.3 \pm 2.4)} \text{ s}^{-1}$ ; Figure 2E, F).<sup>[12]</sup>

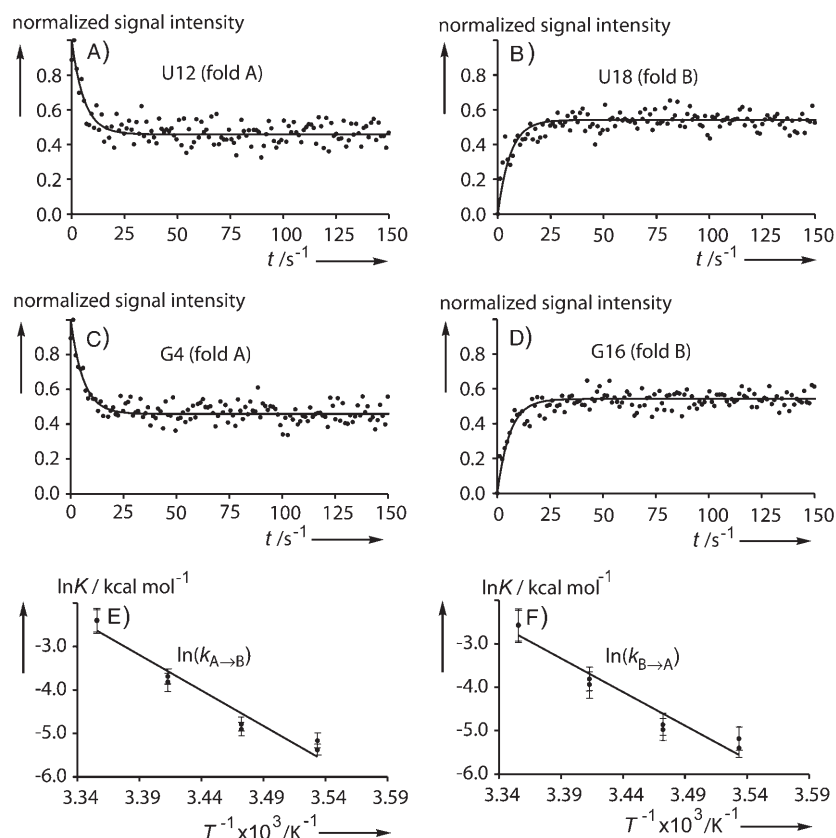
To deduce information about a plausible folding mechanism, imino-proton water-exchange rates were determined for both folds of the parent 20-mer RNA sequence and the two corre-

sponding truncated hairpin sequences, which served as static references (for rates and structures see Figure S3 in the Supporting Information). The water-exchange rates for the corresponding imino protons of the refolding system and the static reference systems were nearly identical. This result suggests that the global dynamic properties of this bistable RNA are not reflected in the local dynamic properties of individual base pairs.

We have shown that NPE-protected uridine is capable of destabilizing secondary structures to the same extent as the previously reported NPE-protected guanosine,<sup>[6]</sup> although A–U base pairs disrupted in this way are weaker than G–C base pairs and contribute less to the total base-pair energy. The photochemical properties of the two protected nucleotides are identical; this allows fast, clean and almost complete removal by one laser pulse ( $\geq 90\%$  deprotection with a 4 W laser pulse for 1.5 s at 350 nm).

This method of temporarily blocking the formation of base pairs with photolabile groups is a precious tool for reversibly controlling the formation of secondary structures. The perturbation is predictable, and, therefore, the site of modification can be chosen rationally. The refolding of the small RNA sequence presented here can be well described in terms of a simple, reversible, two-state process. The activation energy of  $\Delta H^\ddagger \approx +30 \text{ kcal mol}^{-1}$  is large compared to the total base-pairing enthalpy of  $\Delta H \approx 48 \text{ kcal mol}^{-1}$ ,<sup>[13]</sup> this suggests an open transition state with only few base pairs still in place. The global refold-

ing dynamics of this bistable RNA sequence do not affect the local base-pair dynamics in the ground state. Together with the large activation energy, this suggests a dissociative refolding mechanism, in which partial dissociation of the stems has to precede the formation of new base pairs. In contrast, the



**Figure 2.** Kinetic analysis of the refolding reaction, as derived from time-dependent single-scan NMR experiments of the photolyzed 20-mer RNA sequence. A)–D) Normalized imino-proton signal intensities as a function of time, recorded at 298 K after photolysis by one laser pulse. The fitted curves represent a reversible, unimolecular reaction (for the detailed procedure and the fits at other temperatures, see the Supporting Information). E), F) Arrhenius plots of temperature-dependant rate constants  $k_{A \rightarrow B}$  and  $k_{B \rightarrow A}$  according to  $\ln k = -E_a/(RT) + \ln A$ ; the error bars (shown at a 95% confidence interval) include the  $K$  uncertainty of 10% and the standard deviation of the curve fit.

previously described RNA sequence of equal base-pair enthalpy but different structure probably refolded through an associative mechanism, in which some base-pair contacts of the alternative fold were formed before most of the existing ones had broken.<sup>[6]</sup>

The activation energies of  $E_a = 25\text{--}32 \text{ kcal mol}^{-1}$  required for the refolding of these two 20-mer RNA sequences are well within the range of activation energies determined for the refolding of larger sequences that represent the substructures of biologically relevant RNA species ( $E_a = 20\text{--}38 \text{ kcal mol}^{-1}$ ).<sup>[14]</sup> This similarity of kinetic parameters suggests that the structural transitions of complex RNA folds are based on the refolding of minimal secondary structure motifs.

## Acknowledgements

We thank Sébastien Porcher for synthetic contributions, as well as Jens Wöhnert and Christian Richter for helpful discussions. The work was supported by the EPFL, the Swiss National Science Foundation (Grant Nr. 2000-06890), the SFB 579 "RNA-Ligand-

Wechselwirkung" and the "Studienstiftung des Deutschen Volkes" (B.F.).

**Keywords:** kinetics • NMR spectroscopy • oligonucleotides • photolysis • RNA structures

- [1] H. M. Al-Hashimi, *ChemBioChem* **2005**, *6*, 1506–1519.
- [2] B. J. Tucker, R. R. Breaker, *Curr. Opin. Struct. Biol.* **2005**, *15*, 342–348.
- [3] J. Noeske, C. Richter, M. A. Grundl, H. R. Nasiri, H. Schwalbe, J. Wöhnert, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1372–1377.
- [4] D. M. Lilley, *Curr. Opin. Struct. Biol.* **2005**, *15*, 313–323.
- [5] a) C. Höbartner, M.-O. Ebert, B. Jaun, R. Micura, *Angew. Chem. Int. Ed.* **2002**, *41*, 605–609; b) W. C. Winkler, R. R. Breaker, *ChemBioChem* **2003**, *4*, 1024–1032.
- [6] P. Wenter, B. Fürtig, A. Hainard, H. Schwalbe, S. Pitsch, *Angew. Chem. Int. Ed.* **2005**, *44*, 2600–2603.
- [7] D. Thirumalai, N. Lee, S. A. Woodson, D. K. Klimov, *Annu. Rev. Phys. Chem.* **2001**, *52*, 751–762.
- [8] C. Höbartner, H. Mittendorfer, K. Breuker, R. Micura, *Angew. Chem. Int. Ed.* **2004**, *43*, 3922–3925.
- [9] The synthesis of the 2'-O-TOM-protected (R)-NPEU phosphoramidite building block is described in the Supporting Information; the  $R$  configuration of the NPE group was chosen arbitrarily. The synthesis and incorporation of the corresponding (S)-NPEU phosphoramidite was recently reported by C. Höbartner, S. K. Silverman, *Angew. Chem.* **2005**, *117*, 7471–7475; *Angew. Chem. Int. Ed.* **2005**, *44*, 7305–7309. All RNA sequences were assembled from 2'-O-TOM-protected phosphoramidite building blocks according to S. Pitsch, P. A. Weiss, L. Jenny, A. Stutz, X. Wu, *Helv. Chim. Acta* **2001**, *84*, 3773–3795. Deprotection, purification and characterization of the sequences are described in the Supporting Information.
- [10] All NMR experiments were carried out on a Bruker AV 800 MHz spectrometer, equipped with 5 mm TXI-HCN cryogenic probes and z-axis gradients. Detailed NMR methods are described in the Supporting Information. All experiments were carried out at a RNA concentration of 0.3 mM in  $\text{H}_2\text{O}/\text{D}_2\text{O}$  (9:1) with potassium arsenate buffer (25 mM) pH 7.0.
- [11] T. Kühn, H. Schwalbe, *J. Am. Chem. Soc.* **2000**, *122*, 6169–6174.
- [12] In order to account for systematic differences in signal attenuation within different base pairs, the uncertainty of the  $K$  values was set to  $\pm 10\%$ ; the indicated range of the  $k$  values is given at a 95% confidence interval and includes the  $K$  uncertainty and the standard deviation of the curve fit; the indicated range of  $\Delta H$ ,  $\Delta S$ ,  $E_a$  and  $A$  is given at a 90% confidence interval and includes the standard deviation of the curve fit.
- [13] The Arrhenius activation energy  $E_a$  was translated into an activation enthalpy  $\Delta H^\ddagger$  according to  $\Delta H^\ddagger = E_a - RT$  ( $RT \approx 0.6 \text{ kcal mol}^{-1}$  at 300 K). The enthalpy of hairpin formation was determined for the two truncated 15-mer RNA hairpins representing fold A (Figure 1A;  $\Delta H = -48.2 \text{ kcal mol}^{-1}$ ) and fold B (Figure S3 in the Supporting Information;  $\Delta H =$

- 47.6 kcal mol<sup>–1</sup>) in potassium arsenate buffer (25 mM, pH 7.0) by temperature-dependent UV spectroscopy according to L. A. Marky, K. J. Breslauer, *Biopolymers* **1987**, 26, 1601–1620.
- [14] The reported  $E_a$  values are: a) 33 kcal mol<sup>–1</sup> for Ile-tRNA: D. C. Lynch, P. R. Schimmel, *Biochemistry* **1974**, 13, 1841–1852; b) 31 and 25 kcal mol<sup>–1</sup> for the 5' half of the *Leptomonas collosoma* SL RNA: K. A. LeCuyer, D. M. Crothers, *Proc. Natl. Acad. Sci. USA* **1994**, 91, 3373–3377; c) 20 kcal mol<sup>–1</sup> for the Group I intron: P. P. Zarrinkar, J. R. Williamson, *Science* **1994**, 265, 918–924; d) 34 kcal mol<sup>–1</sup> for the pseudoknot structure surrounding the first ribosome initiation site in *E. coli*  $\alpha$ -mRNA: T. C. Gluick, R. B. Gerstner, D. E. Draper, *J. Mol. Biol.* **1997**, 270, 451–463; e) 37.6 and 30.1 kcal mol<sup>–1</sup> for *hok*<sup>74</sup> and *pnd*<sup>58</sup> 5'-end RNA fragments, respectively: J. H. A. Nagel, A. P. Gulyaev, K. J. Öistämö, K. Gerdes, C. W. A. Pleij, *Nucleic Acids Res.* **2002**, 30, e63.

Received: November 11, 2005

Published online on February 2, 2006

# KINETICS AND THERMODYNAMICS OF RNA CONFORMATIONAL SWITCHING

Boris Fürtig<sup>a</sup>, Philipp Wenter<sup>b</sup>, Luc Reymond<sup>b</sup>, Christian Richter<sup>a</sup>, Stefan Pitsch<sup>b,\*</sup> and Harald Schwalbe<sup>a,\*</sup>

a: Johann Wolfgang Goethe University

Institute for Organic Chemistry and Chemical Biology

Center for Biomolecular Magnetic Resonance

Marie Curie Str. 11, 60439 Frankfurt

Fax: (+49)69-798-29515

E-mail: schwalbe@nmr.uni-frankfurt.de

b: École Polytechnique Fédérale de Lausanne

Laboratory of Nucleic Acid Chemistry, EPFL-BCH

1015 Lausanne (Switzerland)

E-mail: stefan.pitsch@epfl.ch

## Acknowledgements

We are grateful to Julia Wirmer, Kai Schlepckow, Janina Buck and Prof. Jens Wöhnert for discussions. Our work was supported by the DFG (SFB 579 “RNA-Ligand-Wechselwirkungen”) and the “Studienstiftung des Deutschen Volkes” (B.F.). The Center for Biomolecular Magnetic Resonance (BMRZ) is supported by the state of Hesse and H.S. acknowledges support from the Fonds der Chemischen Industrie.

## Abstract

The conformational switching between two alternative conformations is frequently observed in RNAs. This manuscript describes the characterization of such conformational switching by real-time NMR, the elucidation of the structural mechanism, the kinetics and the thermodynamics underlying the global conformational switching of a 34mer bistable RNA. The conformational switching of the examined bistable RNA occurs on a time scale of minutes and shows a marked temperature dependence. Comparison with other bistable system leads to the formulation of general rules that govern conformational switching in RNA.

## Introduction

The investigation of secondary structure formation in oligonucleotides has been made possible by the development of rapid temperature-jump methods. Ever since, the results of these studies have influenced our understanding of RNA folding. From the results of T-jump experiments it was concluded that RNA folding is a process starting from a completely unfolded state and proceeding towards a final, stable state with a single conformation<sup>[1] [2]</sup>. Folding is considered to follow a hierarchical pathway where rapid formation of secondary structure precedes the slower distinct formation of tertiary elements<sup>[3]</sup>. However, folding studies of the *Tetrahymena* ribozyme revealed that at least for RNAs of larger size the folding is not represented by a purely hierarchical mechanism but that secondary and tertiary folding can be coupled<sup>[4]</sup>. In the cellular context, RNA folding involves the formation of different functional states; conformational switching between those states has been reported to be functional relevant<sup>[5]</sup>. For a nascent RNA chain leaving the polymerase the newly synthesized RNA can often adopt all possible conformations that represent energetic favorable minima<sup>[6]</sup> on a rugged energy hypersurface<sup>[7]</sup>. The initial sampled conformation may have nothing in common with the conformation of the full length functional form of the RNA but may in fact act as kinetic traps<sup>[8-11]</sup>.

Furthermore, *in vivo* RNA conformational switching that involves the breakage of existing and the formation of alternative secondary and tertiary structures is intimately related to many cellular functions of RNA<sup>[12]</sup>. One important example are RNA riboswitches, structural elements located in the 5'-untranslated region of bacterial mRNAs consisting of an aptamer domain that changes conformation upon ligand binding. This conformational switching is coupled to structural rearrangements that regulate gene-expression<sup>[13, 14]</sup>. Other examples are RNA thermometers changing their conformation in response to temperature<sup>[15]</sup> and thereby acting as transcriptional regulators. Or in the case of HIV-1, the leader RNA exists in two alternative conformations<sup>[16]</sup>. In one conformation the molecule adopts an extended thermodynamically stable structure which is proposed to be the conformation necessary for translation and regulation. In the other conformation, the RNA adopts a multi-branched structure and the RNA is able to form a kissing dimer complex allowing the packaging of the viral genome. In this last example, the conformational switching is exogenously induced by binding to a viral NC protein, which acts as an RNA chaperone<sup>[17, 18]</sup>.

Here, we report about the biophysical characterization of conformational switching in RNA and derive general rules of the kinetics intrinsic to RNA structure. We have investigated a bistable 34-nucleotide RNA<sup>[19, 20]</sup> that adopts an equilibrium of two distinct folds in solution. By the introduction of a single photo-removable protecting group<sup>[21, 22]</sup> at different positions in the nucleotide sequence, either one of the conformations can exclusively be stabilized. *In situ* deprotection of the photo-protecting group by laser irradiation in

the NMR spectrometer<sup>[23, 24]</sup> enabled the determination of the kinetics of conformational switching. The results of these experiments now allow a comparative discussion of the folding behavior of secondary structural elements and the elucidation of general rules for RNA conformational switching events.

## RESULTS

In solution, the investigated 34mer RNA adopts a dynamic equilibrium between two conformations, conformation A' and conformation B'' that are almost equally populated at room temperature (Figure 1). Conformation A' consists of two hairpin stem-loop secondary structure elements connected by four single-stranded nucleotides. The hairpin stem-loop structure at the 5'-end of the molecule consists of 5 basepairs and is closed by an YNMG-tetraloop, whereas at the 3'-end of the molecule in this conformation a stem of 6 basepairs is closed by a GNRA-tetraloop. Conformation B'' consists after a single unpaired nucleotide at the 5'-end of a nine basepair canonical stem that is closed by a sequence of 7 nucleotides, on the 3'-side of the stem structure a single stranded overhang of eight nucleotides is present. Since the interconversion of the two conformations is slow on the NMR timescale, the presence of these structural elements could be confirmed by NMR spectroscopic methods by unambiguous resonance assignment of all imino-protons included in basepairs in the involved structural elements and also by comparison of spectra of reference molecules indicative of structural elements of the full length 34mer RNA equilibrium (Figure 1).

### FIGURE 1

The predicted free enthalpies  $\Delta G_{\text{theoretical}}(310\text{K}) = -0.5\text{kcal/mol}$  (mfold prediction<sup>[25]</sup>) resulting in a ratio of 1:2 of conformation A' to conformation B'' could be experimentally verified by integration of well-resolved NMR signals indicative of either of the two conformations. Integration at different temperatures yielded values ranging from  $\Delta G_{\text{experimental}}(283\text{K}) = -0.675\text{kcal/mol}$  to  $\Delta G_{\text{experimental}}(313\text{K}) = -0.1478\text{kcal/mol}$ , the experimentally determined  $\Delta G_{\text{experimental}}$  are in very good agreement with the prediction.

### Topology of structural elements as revealed by NMR structural analysis

To evaluate the results of the kinetic experiments, knowledge of the topology and the three-dimensional conformation of both conformations is necessary. The three dimensional structures of all structural elements of conformation A' are well known and can be found in the literature e.g. the YNMG tetra-loop<sup>[26]</sup>. In fact, NMR analysis including chemical shift analysis support that the structural elements of the 34mer can be represented by the structures reported in the literature<sup>[27]</sup>. Based on these data and in-line with all our NMR observations we could model the conformation as a long helix capped by the respective loops (Figure 3). In contrast to this, mfold<sup>[25]</sup> predicted conformation B'' to contain a completely unknown structural element, namely a heptaloop (5'-C<sup>11</sup>U<sup>12</sup>G<sup>13</sup>U<sup>14</sup>G<sup>15</sup>U<sup>16</sup>U<sup>17</sup>-3'). To gain insight in this structural element we established a model of the heptaloop using NMR parameters as restraints. Therefore, we synthesized a selectively labeled construct that contained <sup>13</sup>C labels in the ribose moieties of the loop nucleotides as well as in the two closing nucleotides.

### FIGURE 2

Based on a combination of through-bond correlations and NOESY experiments (Figure 2), a complete assignment of the loop resonances could be obtained (see experimental section) allowing to calculate the structure of the heptaloop based on NOE and  $^3J(\text{H,H})$  coupling constant restraints. Interestingly, the structure adopted by nucleotides 11-17 is not a heptaloop but can best be described as the combination of a biloop and triloop (Figure 3). As evident from the NOE connectivity, the base of nucleotide G15 stacks upon G18. Nucleotides C11 and G15 adopt C3'-endo sugar conformation and extend the stem whereas the ribose moieties of the nucleotides G13, U14 as well as those of U16 and U17 adopt C2'-endo conformation. Based on our data, nucleotide U12 is conformationally averaged. Although the loop is structurally restrained,  $^{13}\text{C}$  hetNOE measurements reveal that some nucleotides within this element are highly flexible. The high values for hetNOE ranging from 1.4 to 1.6 for the nucleotides U12, G13, U14 and U16, U17 illustrate that those residues exert motions. In contrast, the nucleotides C11 and G15 that are originally also assigned to be loop nucleotides exhibit significant lower hetNOE values of 1.28 and 1.33 that are in the range of the values determined for the nucleotides G9, C10 and G18, C19 being part of the helical stem and forming basepairs (as evident from independent NOESY measurements). Additionally, low temperature NMR measurement reveal that at temperature at or lower than 273K, additional exchange protected imino-signals emerge that most probably stem from exchange protected residues in the heptaloop; one signal is most likely that of G15 involved in a weak basepairing interaction with C11.

### FIGURE 3

#### Real-time NMR analysis

The investigated bistable RNA adopts a dynamic equilibrium. Due to the slow rate of interconversion, the kinetics of this process of RNA refolding cannot be measured by standard NMR methods such as line shape analysis or exchange spectroscopy. Therefore, we prepared photo-caged full-length RNA sequences that selectively destabilize one of the conformations due to modified basepairing properties<sup>[21, 22]</sup>. To stabilize conformation A', a photo-protecting group is introduced at sequence position U7, this nucleotide is selected because it does not interact with any other nucleotide in conformation A' but facilitates a interaction in conformation B'' by forming a Watson-Crick pair with nucleotide A21. Similarly, conformation B'' is exclusively stabilized by introducing the same photo-protecting group at position G30 which is a “non Watson-Crick interacting” residue in conformation B'' but would lead to a canonical G-C basepairs in conformation A' formed between residues G30 and C23.

### FIGURE 4

The predicted stabilization of the conformations can be monitored by NMR spectroscopy. The imino-region of the 1D spectrum reveals that in either case the resonances indicative of the secondary structural elements are exclusively present (Figure 4). The prepared caged starting states can be “uncaged” by a single short laser pulse (1-1.5s of 4.5W at 350nm) leading to the deprotection of the photo-protected bases. After uncaging, the sequences retain their normal basepairing potential and the molecules are able to adopt either one of the two possible conformations according to the thermodynamics of the system. The system therefore relaxes into its equilibrium form with the rate of interconversion between the two conformations. This process is monitored by subsequent recording of NMR spectra (as described in the experimental section) and the build-up and decay curves of signals originating from the different

conformations can be integrated over time in order to obtain the folding profile and to extract the kinetic rates by fitting these traces to the kinetic equations of a two-state folding model. These experiments have been repeated at different temperatures to extract information on the temperature dependence of the refolding behavior. The rates for the interconversion of the two conformations are in the range of  $k_{A' \rightarrow B''}=0.00003s^{-1}$  and  $k_{B'' \rightarrow A'}=0.00008s^{-1}$  at 288K up to  $k_{A' \rightarrow B''}=0.013s^{-1}$  and  $k_{B'' \rightarrow A'}=0.017s^{-1}$  at 313K. Comparison of the kinetic data with the previously reported refolding rates of bistable 20mer RNA systems (Figure 5 a & b) reveals that the rates for the 34mer bistable RNA are approximately 40 to 60 times smaller. Arrhenius analysis of the refolding rates leads to the determination of activation energies that are  $E_{A(A' \rightarrow B'')}=45 \pm 2kcal/mol$  and  $E_{A(B'' \rightarrow A')}=40 \pm 1.2kcal/mol$  and to frequency factors of  $A_{(A' \rightarrow B'')}=10^{29.6 \pm 2}s^{-1}$  and  $A_{(B'' \rightarrow A')}=10^{25.7 \pm 1}s^{-1}$ . From the kinetic rates, we can derive that the activation energies determined here are approximately 10kcal/mol higher compared to those of the smaller RNA systems (Figure 5c).

### FIGURE 5

Thermal denaturation experiments carried out at the same conditions as realtime NMR experiments lead to determination of basepairing enthalpies. For the truncated hairpins (Figure 1), representing secondary structure elements of the full length conformation, values of  $\Delta H_{truncatedA'(1)}=49kcal/mol$ ,  $\Delta H_{truncatedA'(1)}=56kcal/mol$ ,  $\Delta H_{truncatedA'(1)}=78kcal/mol$  were determined from thermal melting experiments. Apparently, these values are nearly equal to the double of the activation energies determined by kinetic analysis and leads to the proposal that towards the transition state the RNA has to break half of its stabilizing interactions in order to refold (Figure 6).

### FIGURE 6

#### Folding of the structural elements itself is fast

The refolding of secondary structural elements are four orders of magnitudes smaller than the folding of RNA structural elements starting from unfolded RNA molecule as reported earlier by several groups. To also show this for the investigated system, we synthesized a 34mer sequence carrying a photo-protecting group at position G9. As known from extensive structural studies carried out on UUCG tetraloops the introduction of the photo-protecting group should lead to a disruption of the structurally essential basepairing between residue G9 and U6 and thereby destabilize the UUCG loop. As a consequence, the entire stem-loop structure should open but also should not force the molecule into the second conformation where the necessary interaction between G9 and C19 is also blocked. This rationale is supported by NOESY spectra of the  $G9^{hv}$  photo protected molecule revealing that the introduction of a protecting group at G9 leads to the (at least partial) opening of the 5 basepairs containing stem  $5'-A^1C^2A^3G^4G^5-3'/5'-C^{10}C^{11}U^{12}G^{13}U^{14}-3'$  that should be present in Conformation A'. This becomes obvious comparing the NOESY spectra of the reference hairpin 1 (black colored spectrum Figure 7a) and of the reference hairpin 2 (blue colored spectrum Figure 7a) with the NOESY spectrum of the  $G9^{hv}$  molecule. The connectivity of resonances are still present for the part of the molecule that is represented by reference hairpin 2 but it is lost for the resonances of the first stem loop represented by reference hairpin 1. Although resonances of the 5'-side stem loop structure are partially present in 1D- $^1H$  NMR spectra, a NOESY walk fails indicating that



basepairs distant from the distorted loop maybe present but their stability is drastically reduced and therefore magnetization of the imino resonance is sequestered due to increased chemical exchange with the solvent water. In addition, the resonances of the tetraloop are completely absent as well as the resonances indicative of the apical basepairs e.g. G5 (blue circle in Figure 7b). Apparently, a distorted, partially unfolded conformation is present in this G9<sup>h</sup> state. The kinetics of folding RNA from unfolded or partially unfolded states is indeed fast, actually too fast to sample it with our method. As shown in Figure 7c. the peak of G5' is completely absent before deprotection with a laser pulse and becomes fully present in the first spectrum recorded after laser application. The rates for folding from an unfolded state of the RNA are higher than for conformational switching between different folded states. The lower limit for such folding from an unfolded state is  $k_{\text{fold-from-unfolded-state}} > 4\text{s}^{-1}$  with a single exponential behavior describing the process.

#### FIGURE 7

#### Folding rates as functions of topologies and basepairs involved in Sec.Str.

We have elucidated the folding rates of three different bistable RNAs (Figure 8). A decrease in folding rate by increasing number of basepairing interactions per A-helical stem can be observed. This decrease apparently follows an exponential decay and can be related via Eyring equation (equation 1)

$$(Equation\ 1) \quad k = \kappa \frac{k_B T}{h} \exp\left(-\frac{\Delta G^\ddagger}{RT}\right)$$

to an increase in  $\Delta G^\ddagger$ .

#### FIGURE 8

Comparison of activation energies  $E_{a_s}$  derived from Arrhenius analysis of the kinetic data shows a linear increase with the number of basepairs present in the initial structure. Comparison of the kinetic data with the results of prediction programs such as m-fold yields a nearly constant dependence of the ratio  $\Delta G^\ddagger / \text{initial} \Delta G$  of approx.  $1/2$ . Taking into account that  $\text{initial} \Delta G$  is a measure of interactions of the basepairs involved in the initial conformation we conclude that during these refolding processes a transition state has to be passed that retains half of the basepairs formed in the initial state. This suggests a refolding process in which an opening of basepairs towards the transition state has to occur as the initial and rate-determining step, followed by a propagation in the opening of initial basepairs that is concerted with the formation of new basepairs present in the final conformational state in the refolding process. When extrapolating this conclusion to refolding processes of RNAs in general, the refolding kinetics slow down exponentially for a linear increase in basepairs. A helical element of 20bp length would unfold with a half life of  $T = 3.7$  years.

#### FIGURE 9

In this context, it is interesting to compare the length of helices under investigations with the length of helices found in ribosomal RNAs. The analysis was conducted on the secondary structure models for the 23S rRNA of *E. coli* (J01695) and *D. radiodurans* (AE002087)<sup>[28]</sup> and cross-checked on the three-dimensional model of the large ribosomal subunit of *D. radiodurans*<sup>[29]</sup>. Such analysis

reveals that helices with canonical Watson-Crick basepairing pattern in the ribosome hardly ever exceed a length of 10 basepairs (Figure 9) and that most of the helices are consisting of 7 or fewer basepairs, although the overall dimensions of the ribosome would also allow continuous helices with up to 80 basepairs [30]. In addition, the number of non-canonical nucleotides increases with increasing helix length in the ribosome, potentially by destabilization of the potential long helices. We conclude that canonical basepaired stems longer than 10 helices are avoided in order not to generate kinetic traps that would have to refold to the wanted structures on a timescale of minutes or even hours.

## Conclusions

We could show that folding of secondary structure elements starting at partially unfolded RNA chains is a fast process. In contrast, the switching of the pre-existing secondary structure towards an energetically similar secondary structure is accompanied with significantly slower rates of interconversion. The kinetics of these processes is mainly determined by the energetic contributions of the structural building blocks. The comparison of 5 different model systems with the secondary structure elements found in the ribosomal RNA leads us to propose that natural RNAs avoid canonical helices longer than 10 basepairs per A-form helical element in order to avoid being kinetically trapped in non-native states with very significant barriers for a conformational switch into the thermodynamically most stable conformation.

## Material and Methods

### a. Synthesis

The photolabile protected nucleotides were prepared as suitable building blocks for solid phase RNA synthesis as described earlier<sup>[21, 22]</sup>. The incorporation of the modified building blocks into RNA sequences occurred efficiently with coupling yields > 98% and without side-reactions under the standard coupling and deprotection conditions, developed for the assembly of RNA sequences from 2'-O-tom protected building blocks<sup>[31, 32]</sup>.

Synthesis of the selectively labeled sequence is conducted as already described by using canonical 2'-O-TOM protected ribonucleoside phosphoramidites and solid supports containing [<sup>13</sup>C5]-labeled ribose moieties [33].

### b. Real-time NMR experiments

Kinetic NMR experiments were conducted on a Bruker AV 800MHz spectrometer equipped with a cryogenic z-grad TXI HCN probe. For triggering of the reaction, the laser setup was synchronized with the spectrometer via TTL connection<sup>[24]</sup>. The kinetic traces were recorded in a pseudo 3D dataset (described in Fürtig et al. 2007<sup>[34]</sup>). Per kinetic trace, one or two transients were averaged. Processing and analysis of these experimental data was performed using the software FELIX (Accelrys). Non-linear fitting of the kinetic traces was achieved using the program Sigmaplot (Systat Software Inc.).

### c. NMR spectroscopy

1D proton spectra: All spectra were measured at temperatures between 283K and 313 K at a Bruker AV800MHz spectrometer equipped with a cryogenic z-grad HCN probe. The water suppression was achieved either by application of jump-return<sup>[35]</sup> or Watergate <sup>[36]</sup> pulse sequences. Proton pulses were applied on resonance of the water frequency with field strength of 29.41kHz. Sufficient S/N was achieved by recoding of a single transient at 0.1mM RNA samples, nevertheless for static measurements 128 transients were averaged, the relaxation delay was set to 1.2s and the SW was 25ppm.

d. Assignment and Secondary Structure Determination

For the assignment of imino-proton resonances, jump-return-NOESY spectra were recorded at 278K in 9:1 H<sub>2</sub>O/D<sub>2</sub>O 25mM K-phosphate buffer containing 50mM NaCl pH6.5 as well for the full-length RNA construct as for the reference hairpin RNA molecules. All jump-return-NOESY spectra were recoded on a Bruker AV800MHz spectrometer equipped with a cryogenic TXI-HCN z-grad probe. Hard-pulses were applied at the midpoint of all RNA proton resonances prior to the mixing time and at the water frequency subsequent of the mixing time with field strength of 26kHz. The spectra were recoded with a spectral width of 13ppm and 25ppm for the indirect and direct dimensions; 1024pts and 4096pts were recorded for  $t_1$  and  $t_2$ , respectively. Gradients had a SINE shape and were applied with strength of 44G/cm. The relaxation delay was 2s; the mixing time ranged between 150ms and 250ms. Other proton resonances were assigned by recording soft-watergate-NOESY spectra that were applied with similar parameters. In order to facilitate water-suppression selective pulses with Gaussian shapes (as found in the Bruker shape library) of 1ms duration were applied. Unambiguous resonance assignment of the selectively <sup>13</sup>C labeled 27mer RNA was conducted using a combination of HCP<sup>[37]</sup>, HCP-TOCSY<sup>[38]</sup>, long-range HMBC and forward directed HCC-TOCSY-CCH E.COSY <sup>[39, 40]</sup> experiments. The 3D HCP and HCP-TOCSY experiments were recorded on a Bruker AVIII 600MHz spectrometer equipped with a cryogenic TCI HCP z-grad probe at 298K. Proton pulses were applied with field strength of 24kHz on-resonance at the water frequency, the carriers for <sup>13</sup>C and <sup>31</sup>P were set to 77ppm and 0ppm, respectively. The spectrum was recorded with 4096pts, 128pts, 64pts for  $t_3$ ,  $t_2$  and  $t_3$ , respectively. The relaxation delay was set to 1.5s and for each increment 16 transients were recorded. In the similarly recorded HCP-TOCSY experiment the CC-TOCSY was facilitated using a DIPSI-3 sequence with field strength of 9.3kHz for 13ms. The HMBC spectrum was recorded on a Bruker AV 700MHz spectrometer at 298K and in a fashion as optimized on long-range couplings and with low-pass J-filter to suppress unwanted one-bond correlations. No decoupling was applied during acquisition. For the selection of coherences SINE shaped gradient pulses of 27.5G/cm and 22.055G/cm were used. Proton pulses and <sup>13</sup>C pulses were applied with 23kHz, 18,5kHz on resonance at the water frequency and at 88.5ppm, respectively. The spectrum was recorded with a spectral width of 8ppm, 10ppm and 2048pts, 64pts for  $t_2$  and  $t_1$ , respectively. Off-resonant carbon Q3 pulses (512  $\mu$ s) were applied during carbon evolution with an offset 5000 Hz in order to suppress the <sup>1</sup>J(C1', C2') coupling. The 3D forward directed HCC-TOCSY-CCH E.COSY experiment was recorded on a Bruker DRX600 spectrometer at 298K. The times for the transfer were set to  $\tau_1$ =14ms for the CC-TOCSY and to  $\tau_1$ =8ms for the CT-delay to optimize the transfer amplitude as discussed in detail in Glaser et al. <sup>[40]</sup>. The <sup>13</sup>C-carrier was set to 75ppm. All <sup>1</sup>H pulses were applied at the water frequency. Hard pulses were applied with field strength of 29kHz and 19kHz for <sup>1</sup>H and <sup>13</sup>C, respectively. GARP decoupling <sup>[41]</sup> during acquisition was applied with field strength of 3.85kHz; the DIPSI-3 sequence <sup>[42]</sup> during the CC-TOCSY transfer was applied with field strength of 9.2kHz. The spectrum was recorded with 108pts, 88pt and 2048pts for  $t_1$ ,  $t_2$ ,

and  $t_3$ , respectively, resulting in an experimental duration of 36h (relaxation delay of 1.4s). Processing was conducted with XWINNMR 3.5; the Fourier transformed spectrum was converted to MSI Felix2000 for further analyzed. Scalar  $^3J(\text{H,H})$  coupling constants were also extracted from the latter experiment, the scalar coupling constants were extracted using the method described [39]. The resolution in the direct dimension where the  $^3J(\text{H,H})$  splitting was measured was set to 1.75Hz per complex point and was zero-filled during processing to a resolution of 0.4Hz per complex point. NOE-derived restraints for modeling the secondary structure were extracted from 2D double half-filtered NOESY spectra[43].

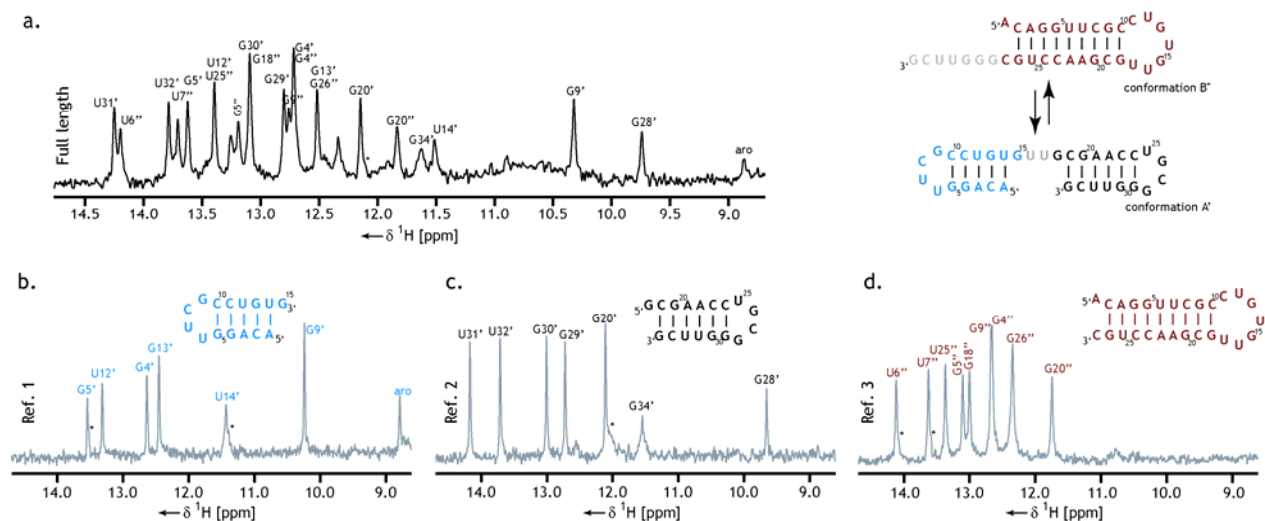
All spectra were processed using Bruker XWINNMR 3.5; the Fourier transformed spectra were converted to MSI Felix2000 for further analysis.

Modeling of the structures (as depicted in Figures 3 and 9) was achieved by structure calculation runs using CNX2000 from Accelrys with inputs of distance restraints from the NOESY-experiments, dihedral restraints derived from scalar couplings and basepair restraints derived from the presence of imino- and amino resonances.

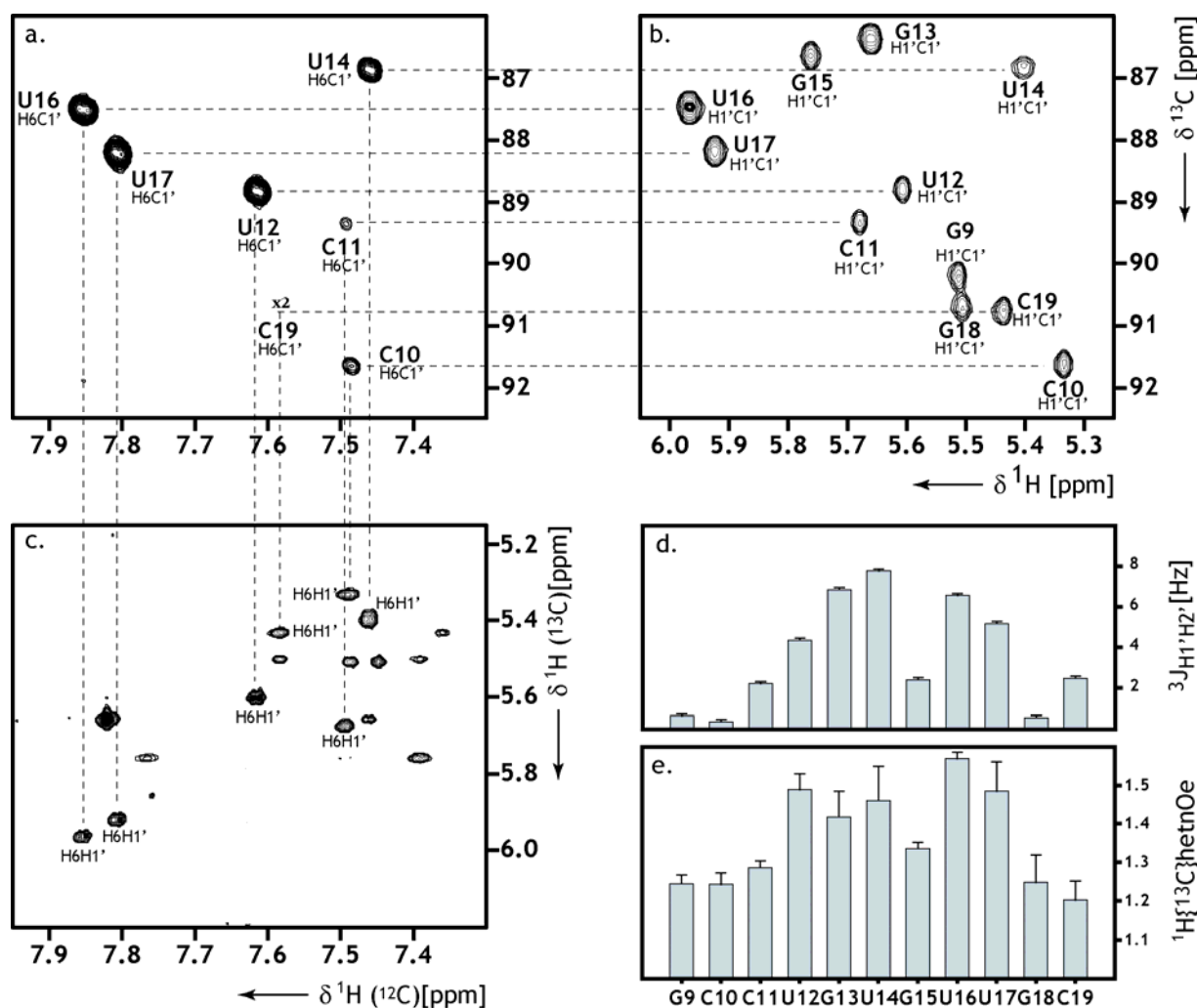
#### e. $^{13}\text{C}$ Relaxation Experiments

NMR experiments were carried out on a 600 MHz Bruker spectrometer equipped with a 5 mm  $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$  z-Grad TXI probe.  $^{13}\text{C}\{-^1\text{H}\}$  steady-state NOE data were obtained using  $^{13}\text{C}$  modifications of the pulse sequence `invnoef3gpsi` comprised in the Bruker pulse sequence library. The carrier frequency was set to 89ppm for C1'. The spectral width was 6 ppm for the ribose moiety. About 128 complex points were acquired in the indirect dimension. Off-resonant carbon Q3 pulses (512  $\mu\text{s}$ ) were applied during carbon evolution with an offset 5000 Hz in order to suppress the  $^1J(\text{C1}', \text{C2}')$  coupling. Relaxation delays of 5 s were applied between the scans the NOE measurements. NOE data were obtained with with 16 scans for each  $t_1$ -increment. The NOE experiment was recorded interleaved, with alternating proton-presaturated and non-presaturated spectra. While in the latter case, a relaxation delay of 5 s was used, proton-presaturation was applied for 3 s subsequent to a 2 s relaxation delay in the presaturated spectra. The interleaved spectra were separated by a Bruker standard macro.

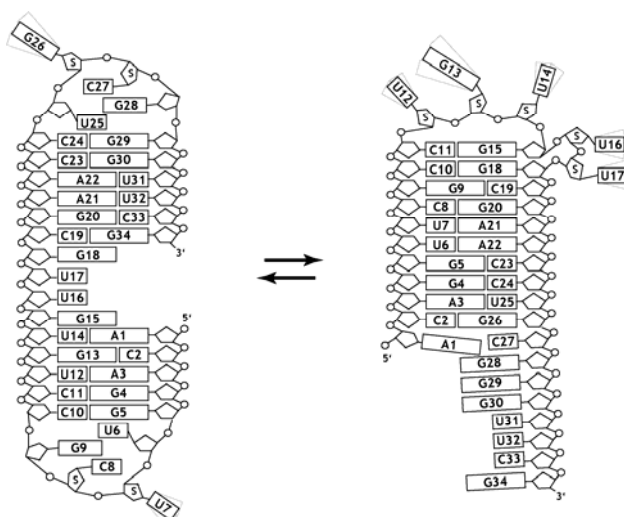
## Figures and Figure Legends



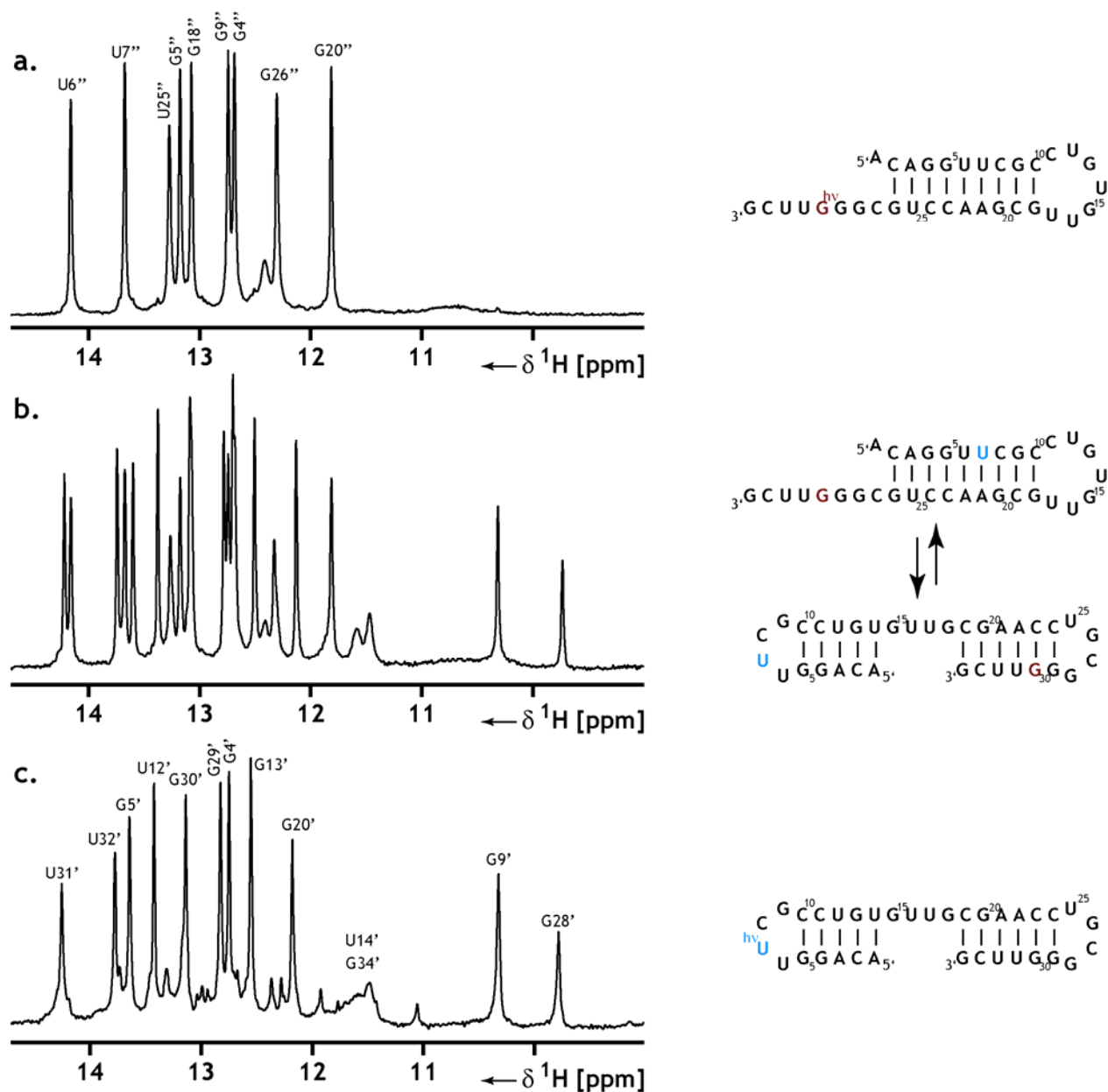
**Figure 1:** Comparison of the 1D  $^1\text{H}$  NMR spectra of the reference molecules (b-d) with the full-length RNA construct (a). Secondary structures of the reference molecules (b-d) are depicted color coded as well as their counterparts in the full length equilibrium form (a). The full-length RNA adopts two conformations that slowly interconvert on the NMR-timescale and can be described as the direct sum of the spectra of the reference hairpins at the given temperature of 310K. The spectra were recorded at 800 MHz.



**Figure 2:** Representation of collected NMR data on the 27mer reference molecule (secondary structure is depicted in figure 1d), where the two closing basepairs and the nucleotides spanning the hepta-loop are selectively  $^{13}\text{C}$  labeled in the ribose moieties. Panels a, b, c depict the  $\text{H1'}/\text{C1'}$  regions of the long-range  $^1\text{H},^{13}\text{C}$ -HMBC, the  $^1\text{H},^{13}\text{C}$ -HSQC and the  $^1\text{H}(^{13}\text{C}),^1\text{H}$ -half filtered NOESY spectra, respectively. Resonance assignment is annotated. Panel d displays the measured  $^3J(\text{H1}',\text{H2'})$  scalar coupling constants indicative for the conformation of the ribose moieties. In panel e, the  $^1\text{H}\{^{13}\text{C}\}$  het-NOE data is displayed.

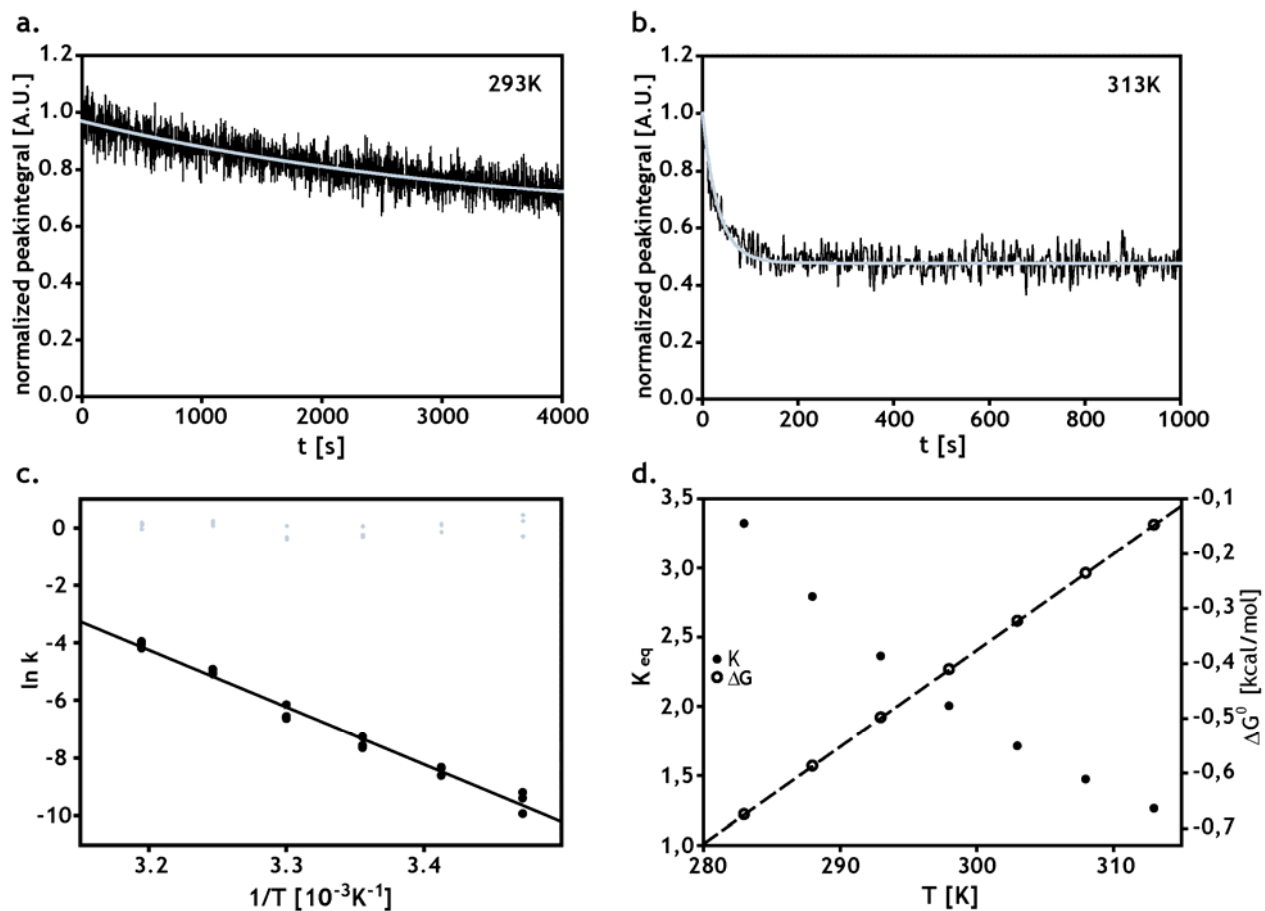


**Figure 3:** Secondary structure models as derived from NMR data. Left panel depicts the secondary structure model of conformation A', the tetra-loops U6-U7-C8-G9 and U25-G26-C27-G28 adopt the typical YNMG topology, whereas the stems adopt a coaxially stacked conformation. Right panel displays the secondary structure model of conformation B'', the proposed hepta-loop folds onto the helical stem where nucleotides C11 and G15 elongate the helix and nucleotides U12-G13-U14 and U16-U17 form a triloop and a biloop, respectively.

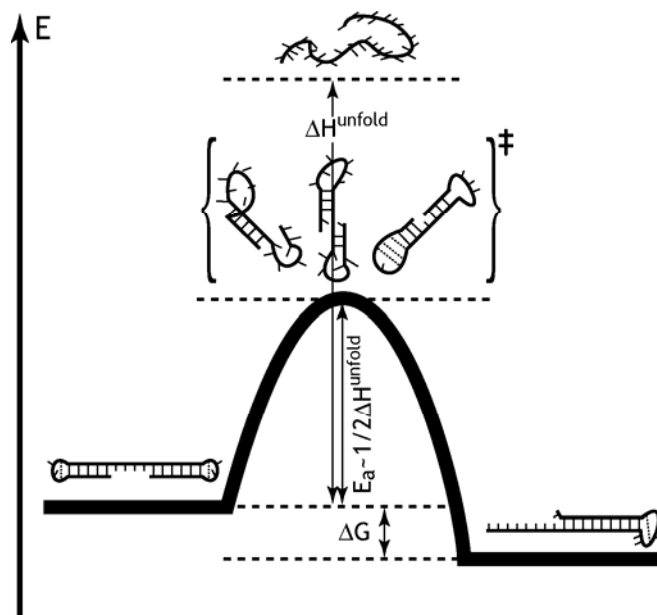


**Figure 4:** Comparison of the 1D  $^1\text{H}$  NMR spectra of the caged RNA sequences  $\text{G30}^{\text{hv}}$  (a) and  $\text{U7}^{\text{hv}}$  (c) with the full length equilibrium form (b); Color coding is indicating the positions that were chosen to stabilize either one of the conformations. NMR spectra of the caged RNAs show that the introduction of the photolabile protecting group is able to cage one desired conformation from the equilibrium for the unmodified RNA. The spectra were measured at 800MHz spectrometer.

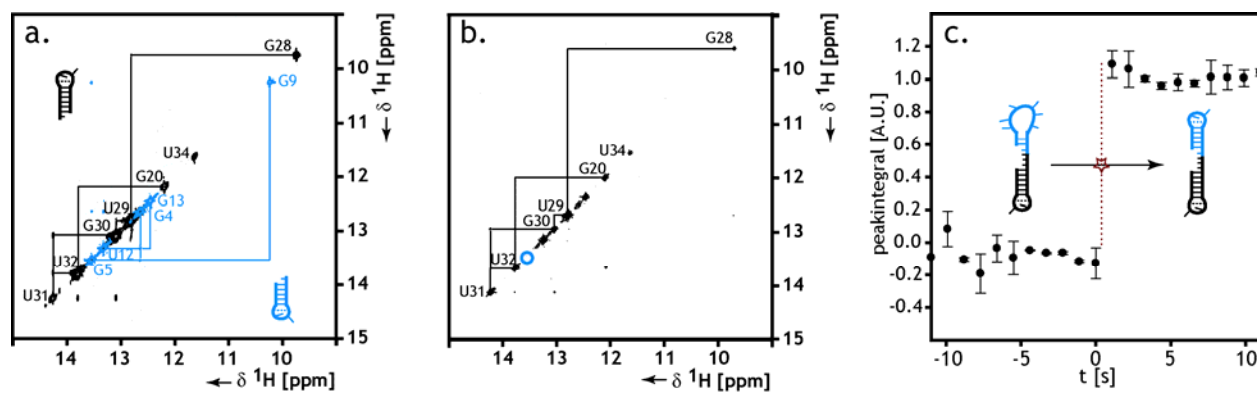




**Figure 5:** Panels a and b represent kinetic traces measured by real-time NMR of the relaxation of the photo-protected conformation G30<sup>h<sub>v</sub></sup> towards the equilibrium measured at 293K and 313K, respectively. Panel c depicts the Arrhenius analysis of the kinetics of this conformational switching process (black points), grey points indicate the residuals of the linear fit of  $\ln k = -(E_a/RT) + \ln A$ . Panel d shows the extracted equilibrium constants that can be back calculated to the free enthalpy of the equilibrium ( $\Delta G = -RT \ln K_{eq}$ ).

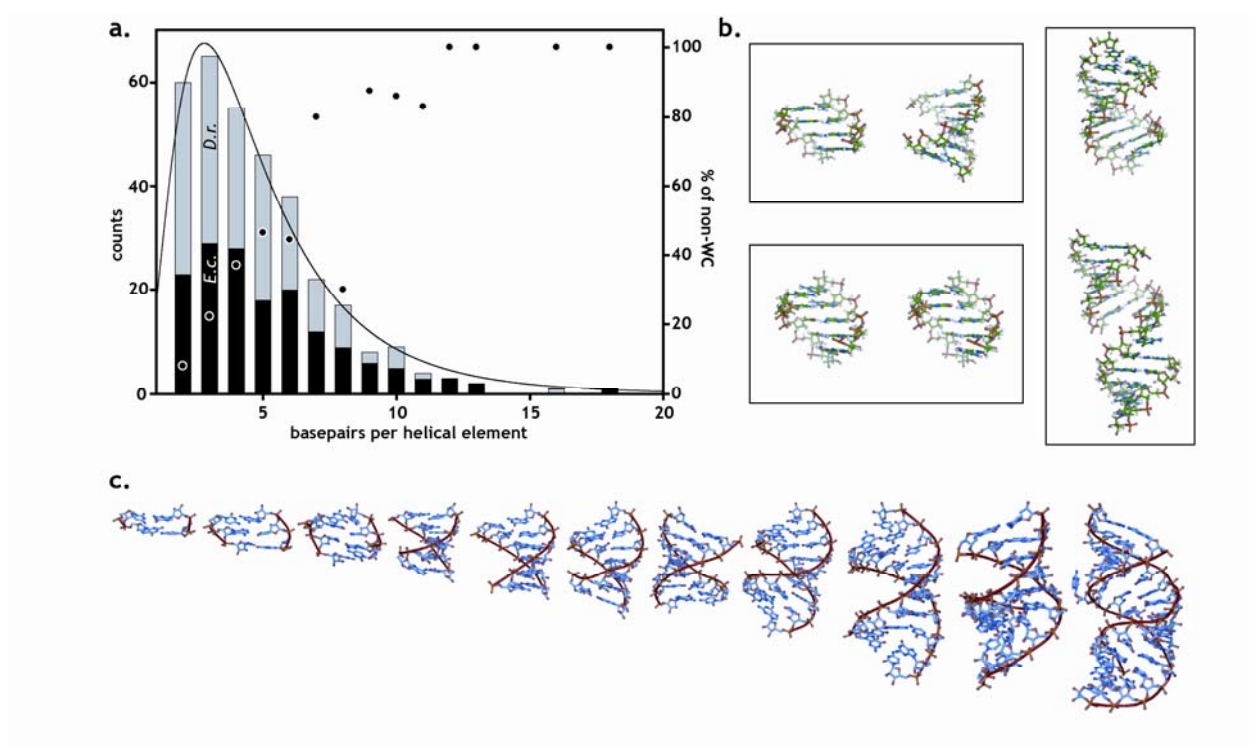


**Figure 6:** Schematic representation of the folding. In order to conformationally switch between two stable ground states the helical RNA conformation has to pass a transition state that is represented by an ensemble of conformations characterized by the presence of half of interactions present in the ground states. Therefore the rate limiting step in RNA refolding reaction is the unfolding towards the transition state, folding itself occurs very rapidly.



**Figure 7:** Panel a) corresponds to an overlay of the imino-region of the  $^1\text{H}$ - $^1\text{H}$  NOESY spectra of the two reference hairpins representing the structural elements present in conformation A' (5'-ACAGGUUCGCCUGUG-3' in blue and 5'-GCGAACCUGCGGGUUCG-3' in black). Panel b) shows the  $^1\text{H}$ - $^1\text{H}$  NOESY spectrum of the molecule G9<sup>hV</sup> (5'-ACAGGUUCG<sup>hV</sup>CCUGUGUUGCGAACCUGCGGGUUCG-3') that reveals the complete sequential NOESY walk just for one of the two stems. The second stem is partially unfolded due to the incorporation of the photolabile protecting group at a sequence position critical for the formation of the UUCG tetraloop. Panel c) shows the kinetic trace that reflects the fast kinetics of the formation of the proper sequence element (loop & stem), the peak of G5 (position indicated by blue circle in b) is present in the first spectrum recorded directly after laser deprotection (as indicated by the red star).





**Figure 9:** Evaluation of helix-length in ribosomal RNAs. Panel a depicts a histogram of the length of helical elements in 23S RNA of *E. coli* and *D. radiodurans* as found in their secondary structure model J01695 and AE002087. Counted are helical segments that consist solely of Watson-Crick basepairs and GU-basepairs, helices interrupted by non-canonical basepairs are counted as separate helical segments, the percentage of helical elements that contain GU-basepairs (non-WC) is highlighted in the histogram; the length of helical elements in the rRNA follows a log-normal distribution with a maximum at 3 Watson-Crick basepairs. Panel b displays the helical elements as modeled from the NMR data and reveals that those are ideal models to gain information on the refolding processes in natural RNAs as they depict the length of helical element found in the ribosome. For comparison reasons some helical elements as found in the three-dimensional structure of the large ribosomal subunit of *D. radiodurans* (1NKK) are depicted in panel c with increase in helical length from 2 basepairs (left) to 11 basepairs (second from right) and 16 basepairs (right most). As evident from the displayed structures the helical segment containing the high number of 16 basepairs deviates from the conformation of canonical helices and not all basepairs are formed properly but get distorted upon tertiary interactions.

## References

- [1] D. Porschke, *Mol Biol Biochem Biophys* 1977, **24**, 191-218.
- [2] S. D. Gilbert, R. T. Batey, *Chem Biol* 2006, **13**, 805-807.
- [3] D. M. Crothers, P. E. Cole, C. W. Hilbers, R. G. Shulman, *J Mol Biol* 1974, **87**, 63-88.
- [4] M. Wu, I. Tinoco, Jr., *Proc Natl Acad Sci U S A* 1998, **95**, 11555-11560.
- [5] J. H. Nagel, C. W. Pleij, *Biochimie* 2002, **84**, 913-923.
- [6] T. Pan, T. Sosnick, *Annu Rev Biophys Biomol Struct* 2006, **35**, 161-175.
- [7] N. Leulliot, G. Varani, *Biochemistry* 2001, **40**, 7947-7956.

- [8] D. K. Treiber, M. S. Rook, P. P. Zarrinkar, J. R. Williamson, *Science* 1998, 279, 1943-1946.
- [9] D. K. Treiber, J. R. Williamson, *Curr Opin Struct Biol* 1999, 9, 339-345.
- [10] D. K. Treiber, J. R. Williamson, *Curr Opin Struct Biol* 2001, 11, 309-314.
- [11] T. Pan, T. R. Sosnick, *Nat Struct Biol* 1997, 4, 931-938.
- [12] H. M. Al-Hashimi, *Chembiotech* 2005, 6, 1506-1519.
- [13] B. J. Tucker, R. R. Breaker, *Curr Opin Struct Biol* 2005, 15, 342-348.
- [14] E. Nudler, A. S. Mironov, *Trends Biochem Sci* 2004, 29, 11-17.
- [15] F. Narberhaus, T. Waldminghaus, S. Chowdhury, *FEMS Microbiol Rev* 2006, 30, 3-16.
- [16] B. Berkhout, J. L. van Wamel, *Rna* 2000, 6, 282-295.
- [17] H. Huthoff, B. Berkhout, *Rna* 2001, 7, 143-157.
- [18] M. Ooms, K. Verhoef, E. Southern, H. Huthoff, B. Berkhout, *Nucleic Acids Res* 2004, 32, 819-827.
- [19] C. Hobartner, R. Micura, *J Mol Biol* 2003, 325, 421-431.
- [20] C. Flamm, I. L. Hofacker, S. Maurer-Stroh, P. F. Stadler, M. Zehl, *Rna* 2001, 7, 254-265.
- [21] P. Wenter, B. Fürtig, A. Hainard, H. Schwalbe, S. Pitsch, *Angewandte Chemie-International Edition* 2005, 44, 2600-2603.
- [22] P. Wenter, B. Fürtig, A. Hainard, H. Schwalbe, S. Pitsch, *Chembiotech* 2006, 7, 417-420.
- [23] J. Wirmer, J. Kuhn, H. Schwalbe, *Angewandte Chemie-International Edition* 2001, 40, 4248-4251.
- [24] T. Kuhn, H. Schwalbe, *Journal of the American Chemical Society* 2000, 122, 6169-6174.
- [25] M. Zuker, *Nucleic Acids Res* 2003, 31, 3406-3415.
- [26] E. Ennifar, A. Nikulin, S. Tishchenko, A. Serganov, N. Nevskaya, M. Garber, B. Ehresmann, C. Ehresmann, S. Nikonov, P. Dumas, *J Mol Biol* 2000, 304, 35-42.
- [27] F. H. T. Allain, G. Varani, *Journal of Molecular Biology* 1995, 250, 333-353.
- [28] J. J. Cannone, S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Muller, N. Pande, Z. Shang, N. Yu, R. R. Gutell, *BMC Bioinformatics* 2002, 3, 2.
- [29] J. Harms, F. Schlutzenzen, R. Zarivach, A. Bashan, S. Gat, I. Agmon, H. Bartels, F. Franceschi, A. Yonath, *Cell* 2001, 107, 679-688.
- [30] H. F. Noller, *Science* 2005, 309, 1508-1514.
- [31] S. Pitsch, D. Ackermann, C. Denarie, F. Meylan, M. Meyyappan, E. Muller, A. Peer, S. Porcher, L. Reymond, A. Stutz, P. Wenter, X. L. Wu, *Chimia* 2005, 59, 808-811.
- [32] S. Pitsch, P. A. Weiss, L. Jenny, A. Stutz, X. L. Wu, *Helvetica Chimica Acta* 2001, 84, 3773-3795.
- [33] P. Wenter, L. Reymond, S. D. Auweter, F. H. Allain, S. Pitsch, *Nucleic Acids Res* 2006, 34, e79.
- [34] B. Fürtig, J. Buck, V. Manoharan, W. Bermel, A. Jäschke, P. Wenter, S. Pitsch, H. Schwalbe, *Biopolymers* 2007, *in press*.
- [35] V. Sklenar, A. Bax, *Journal of Magnetic Resonance* 1987, 74, 469-479.
- [36] M. L. Liu, X. A. Mao, C. H. Ye, H. Huang, J. K. Nicholson, J. C. Lindon, *Journal of Magnetic Resonance* 1998, 132, 125-129.
- [37] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *Journal of the American Chemical Society* 1994, 116, 6472-6473.
- [38] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *Journal of Biomolecular Nmr* 1995, 5, 87-92.
- [39] H. Schwalbe, J. P. Marino, S. J. Glaser, C. Griesinger, *Journal of the American Chemical Society* 1995, 117, 7251-7252.
- [40] S. J. Glaser, H. Schwalbe, J. P. Marino, C. Griesinger, *Journal of Magnetic Resonance Series B* 1996, 112, 160-180.
- [41] A. J. Shaka, P. B. Barker, R. Freeman, *Journal of Magnetic Resonance* 1985, 64, 547-552.
- [42] A. J. Shaka, C. J. Lee, A. Pines, *Journal of Magnetic Resonance* 1988, 77, 274-293.
- [43] G. Otting, K. Wuthrich, *Q Rev Biophys* 1990, 23, 39-96.

# TIME-RESOLVED NMR STUDIES OF RNA FOLDING

Boris Fürtig<sup>1</sup>, Janina Buck<sup>1</sup>, Vijayalaxmi Manoharan<sup>1</sup>, Wolfgang Bermel<sup>2</sup>, Andres Jäschke<sup>3</sup>, Philipp

Wenter<sup>4</sup>, Stefan Pitsch<sup>4</sup>, Harald Schwalbe<sup>1\*</sup>

<sup>1</sup>Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe-University Frankfurt, Max-von-Laue-Str. 7, D-60438 Frankfurt, Germany;

email: schwalbe@nmr.uni-frankfurt.de; fax: ++49-69-798-29515.

<sup>2</sup>Bruker Biospin GmbH, Silberstreifen 4, D-76287 Rheinstetten, Germany

<sup>3</sup>Institute of Pharmacy and Molecular Biotechnology, Dept. of Chemistry, University of Heidelberg, Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany

<sup>4</sup>Laboratory of Nucleic Acid Chemistry, École Polytechnique Fédérale de Lausanne, EPFL-BCH, CH-1015 Lausanne, Switzerland

**Keywords:** NMR spectroscopy, RNA folding, time-resolved NMR, photolabile caged compounds

**Acknowledgements:** Our work was supported by the DFG (SFB 579: “RNA-Ligand-Interaction”), the Studienstiftung des Deutschen Volkes with a predoctoral fellowship for B.F., the EU (Joint Research Activity within Integrated Infrastructure Initiative: EU-NMR) and the state Hesse (Center for Biomolecular Magnetic Resonance (BMRZ)). We thank Prof. Jens Wöhnert, Dr. Christian Richter, Jonas Noeske, Kai Schlepckow and Prof. Hashim Al-Hashimi for stimulating discussions.

**Abstract:** the application of real-time NMR experiments to study RNA folding is relatively new. For many RNA folding events, current investigations suggest that the timescales are in the second to minute regime. In addition, the initial investigations indicate that different folding rates are observed for one structural transition, maybe due to the hierarchical folding units of RNA. Many of the experiments developed in the field of NMR of protein folding cannot directly be transferred to RNA: hydrogen exchange experiments outside the spectrometer cannot be applied, since the intrinsic exchange rates in RNA are too fast and relaxation dispersion experiments require faster structural transitions than those observed in RNA. On the other hand, information derived from time-resolved NMR experiments, namely the acquisition of native chemical shifts, can be readily interpreted in light of formation of a single long-range hydrogen-bonding interaction for RNA systems. Together with mutational data that can readily be obtained for RNA and new ligation technologies that enhance site resolution even further, time-resolved NMR may become a powerful tool to decipher RNA folding. Such information will be of importance to understand the functions of coding and non-coding RNAs in cells.

---

## RNA-FOLDING II: CONFORMATIONAL SWITCHING OF RNA TERTIARY STRUCTURE ELEMENTS

**REVIEW ARTICLE:** STRUCTURES OF RNA SWITCHES: INSIGHT INTO MOLECULAR RECOGNITION AND TERTIARY STRUCTURE

Harald Schwalbe, Janina Buck, Boris Fürtig, Jonas Noeske, Jens Wöhnert, *Angew. Chem. Int. Ed.*, 2007, **46**, 1212-1219

[PREPARATION OF THIS MANUSCRIPT WAS SHARED BETWEEN ALL AUTHORS OF THE ARTICLE. ESPECIALLY THE PART ABOUT TERTIARY ARCHITECTURE OF RIBOSWITCH RNA MOLECULES WAS PREPARED BY THE AUTHOR OF THIS THESIS]

**RESEARCH ARTICLE:** INTERPLAY OF 'INDUCED FIT' AND PREORGANIZATION IN THE LIGAND INDUCED FOLDING OF THE APTAMER DOMAIN OF THE GUANINE BINDING RIBOSWITCH

Jonas Noeske, Janina Buck, Boris Fürtig, Hamid Nasiri, Harald Schwalbe, Jens Wöhnert, *Nucl. Acids Res.*, 2007, **35**, 572-583

[IN THE CONTEXT OF THIS THESIS THE ARTICLE IS NOT PRESENTED, THE CONTRIBUTION OF THE AUTHOR OF THIS THESIS WAS THE INITIAL ASSIGNMENT OF THE GSW-HYP COMPLEX IN ORDER TO ANALYSE THE KINETIC DATA PRESENTED IN THE FOLLOWING ARTICLE]

**RESEARCH ARTICLE:** RESOLVING LIGAND-INDUCED RNA FOLDING OF A RIBOSWITCH APTAMER DOMAIN AT ATOMIC RESOLUTION

Janina Buck, Boris Fürtig, Jonas Noeske, Jens Wöhnert, Harald Schwalbe, *PNAS*, *accepted with minor revisions*

**RESEARCH ARTICLE:** MONITORING THE HAMMERHEAD RIBOZYME REACTIVITY BY RT-NMR

Boris Fürtig, Christian Richter, Peter Schell, Philipp Wenter, Stefan Pitsch and Harald Schwalbe, *submitted*

The papers presented in this chapter of the thesis deal with the application of the earlier described RT-NMR methodology to conformational switching processes of RNA tertiary structure elements. An introduction into such tertiary structure elements is provided by the review article that compares the different structures of riboswitch molecules and their inherent functional features. The research article of Buck & Fürtig et al. discusses the ligand induced conformational switching event and deciphers the underlying mechanism of this transition. As a prerequisite for this study the resonance assignment of the imino-protons in the hypoxanthine-RNA complex had to be conducted. The results of these assignment efforts were published earlier within the paper of Noeske et al.

In the riboswitch studies all lab-work was conducted by Janina Buck, the NMR analysis and data evaluation, as well as structure modelling was done by the author of the thesis and Janina Buck equally together.

The second sizeable RNA that was examined by RT-NMR was the hammerhead ribozyme. The application of RT-NMR experiments lead to a detailed description of the catalytic reaction. Synthesis of the monomeric units was performed by Peter Schell, Stefan Pitsch performed the RNA synthesis of the substrate and ribozyme strand, NMR analysis and interpretation was conducted by the author of thesis.



# Structures of RNA Switches: Insight into Molecular Recognition and Tertiary Structure

Harald Schwalbe,\* Janina Buck, Boris Fürtig, Jonas Noeske, and Jens Wöhnert\*

## Keywords:

gene regulation · NMR spectroscopy · RNA · RNA switches · X-ray crystallography

**R**NA switches (riboswitches) have important functions in gene regulation. They comprise an aptamer domain, which is responsible for ligand binding, and an expression platform that transmits the ligand-binding state of the aptamer domain through a conformational change. Riboswitches can regulate gene expression either at the level of transcription or translation, and it has been proposed that riboswitch mechanisms are even used to regulate the processing of mRNA. This Minireview summarizes the current understanding of the structures and mode of action of RNA switches, with particular focus on secondary and tertiary interactions, which stabilize the global RNA structure and thus determine the function of the aptamer domain.

## 1. Natural and Non-Natural RNA Aptamers

It has been known for some time that aptamers, that is, RNAs derived from selection procedures (SELEX: selection of ligands by exponential amplification), can bind to a large variety of small-molecule ligands with high selectivity and affinity.<sup>[1]</sup> Aptamers bind to ligands<sup>[2]</sup> as diverse as nucleotides, amino acids, aromatic dye molecules, coenzymes such as flavinmononucleotide (FMN), aminoglycosides, and S-adenosylmethionine (SAM).<sup>[3]</sup> They stem from a randomized starting pool of RNA molecules that typically comprises  $10^{14}$ – $10^{16}$  mutants. In general, SELEX-derived aptamers bind their small-molecule ligands with affinities in the micromolar

range. Structural analysis of many of the aptamer–ligand complexes show that aptamers form intricate binding pockets for their ligands involving numerous noncanonical RNA structural elements. The interactions include electrostatic and hydrophobic

as well as hydrogen-bonding interactions. In the free form of the aptamer, however, the ligand-binding region is often completely unstructured and ligand binding is accompanied by an induced fit mechanism.

The recent discovery of naturally occurring RNAs, known as RNA switches (riboswitches), that are capable of binding to small-molecule ligands came as a surprise.<sup>[4]</sup> The binding of small-molecule ligands to RNA switches is coupled to a novel type of gene expression control in bacteria as well as in some plants and fungi.<sup>[5]</sup> So far, RNA switches have been reported that bind specifically to essential coenzymes and vitamins, amino acids, glucosamine-6-phosphate, and the purine bases guanine and adenine. RNA switches are mostly found in the 5'-untranslated regions (5'-UTR) of messenger RNA (mRNA) and therefore belong to the noncoding part of the mRNA. Many RNA switches have a modular architecture: they consist of an aptamer domain or sensor region, which is responsible for ligand binding, and the so-called expression platform that transmits the ligand-binding state of the aptamer domain and thereby modulates gene expression. Ligand binding in the aptamer domain or sensor region leads to transcription termination, which abolishes ribosome binding or leads to different mRNA stabilities as a result of altered mRNA processing (Figure 1 and Table 1).

Sequence and secondary-structure analysis indicates that 2% of all genes in *B. subtilis* are regulated at least in part by a riboswitch mechanism. Binding of a small-molecule metabolite to the aptamer domain constitutes a direct feedback loop

[\*] Prof. Dr. H. Schwalbe, J. Buck, B. Fürtig, J. Noeske  
Institute for Organic Chemistry and Chemical Biology  
Center for Biomolecular Magnetic Resonance  
Johann Wolfgang Goethe-University Frankfurt  
Max von Laue-Strasse 7, 60438 Frankfurt am Main (Germany)  
Fax: (+49) 69-798-29515  
E-mail: schwalbe@nmr.uni-frankfurt.de

Prof. Dr. J. Wöhnert  
University of Texas Health Science Center SA  
Department of Biochemistry  
7703 Floyd Curl Drive, San Antonio, TX 78229 (USA)  
Fax: (+001) 210-567-6595  
E-mail: jewoe@biochem.uthscsa.edu



Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.

to either enhance or suppress the de novo synthesis of those proteins that are involved in the metabolic or catabolic pathway of the metabolite itself. The aptamer domains of naturally occurring riboswitches are in general significantly larger than their non-natural counterparts, even if they bind to similar ligands with an affinity in the low nanomolar range.

## 2. Riboswitches as Regulation Elements for Transcription and Translation

Riboswitches have been reported that regulate gene expression either at the level of transcription or translation. In addition, it has been proposed that riboswitch mechanisms are even used to regulate the processing of mRNA and thereby presumably alter cellular mRNA stability.

The adenine riboswitch from *V. vulnificus* has been proposed to regulate the *add* operon during translation as a so-called “on switch” (Figure 1a). As proposed by Patel and

co-workers, binding of adenine induces the formation of a helical element that releases the Shine–Dalgarno sequence (SD sequence) and the start codon.<sup>[8b]</sup> The translation of the *add* operon is therefore up-regulated at high concentrations of adenine. We note, however, that the prediction of the secondary structure of this riboswitch also allows a conformation in which adenine is bound to the fully formed aptamer domain and the SD sequence is still masked by a helical rearrangement (see Supporting Information).

The thiamine pyrophosphate (TPP) riboswitch (Figure 1d) operates at the level of translation as well. In the absence of ligand, the SD sequence is single-stranded as a complementary anti-SD sequence is involved in long-range interactions with an anti-anti-SD sequence. Binding of the metabolite TPP releases the anti-SD that then pairs with the SD sequence. Therefore, in contrast to the adenine riboswitch from *V. vulnificus*, the TPP riboswitch acts as an “off switch”: high concentrations of metabolite down-regulate the translation of the *thiM* operon in *E. coli*. In addition, a TPP riboswitch was found in the 3'-untranslated region (3'-UTR) of the *thiC* operon from *A. thaliana* and is therefore proposed to operate on the level of mRNA stability instead of being involved in translational control as described for the *thiM* operon in *E. coli*.<sup>[6]</sup>

The fascinating mechanisms of regulation can, however, be even more striking: The same adenine-sensing aptamer domain that regulates translation in *V. vulnificus* acts as a transcriptional “on switch” for the *ydhL* (now named *pbuE*) operon in *B. subtilis*. Aptamer domains are apparently designed as modules and used for transcriptional or for translational control depending on the nature of the expression platform. A single nucleotide mutation (U74C) dramat-



Harald Schwalbe, born in 1966, obtained his PhD with C. Griesinger from the University in Frankfurt in 1993. After a postdoctoral stay with C. M. Dobson (Oxford, UK), he completed his habilitation at the University of Frankfurt (1999) before joining the MIT (CA, USA) as an assistant professor. He was appointed Professor of Organic Chemistry at the University of Frankfurt in 2001. His research is focused on the determination of the structure, dynamics, and functions of proteins and RNA using high-resolution NMR spectroscopy.



Jens Wöhnert, born in 1970, studied biochemistry at the Martin Luther University, Halle/Saale, and completed his PhD at the Institute of Molecular Biotechnology, Jena, where he applied NMR spectroscopy to study the structure of RNA and RNA-protein complexes. In 2000, he joined the group of Prof. Schwalbe at the Francis Bitter Magnet Laboratory at MIT. In 2002, he was appointed director of the SFB 579 “RNA–Ligand Interactions” research group at the University of Frankfurt. Since 2005, he is an assistant professor at the University of Texas, San Antonio.



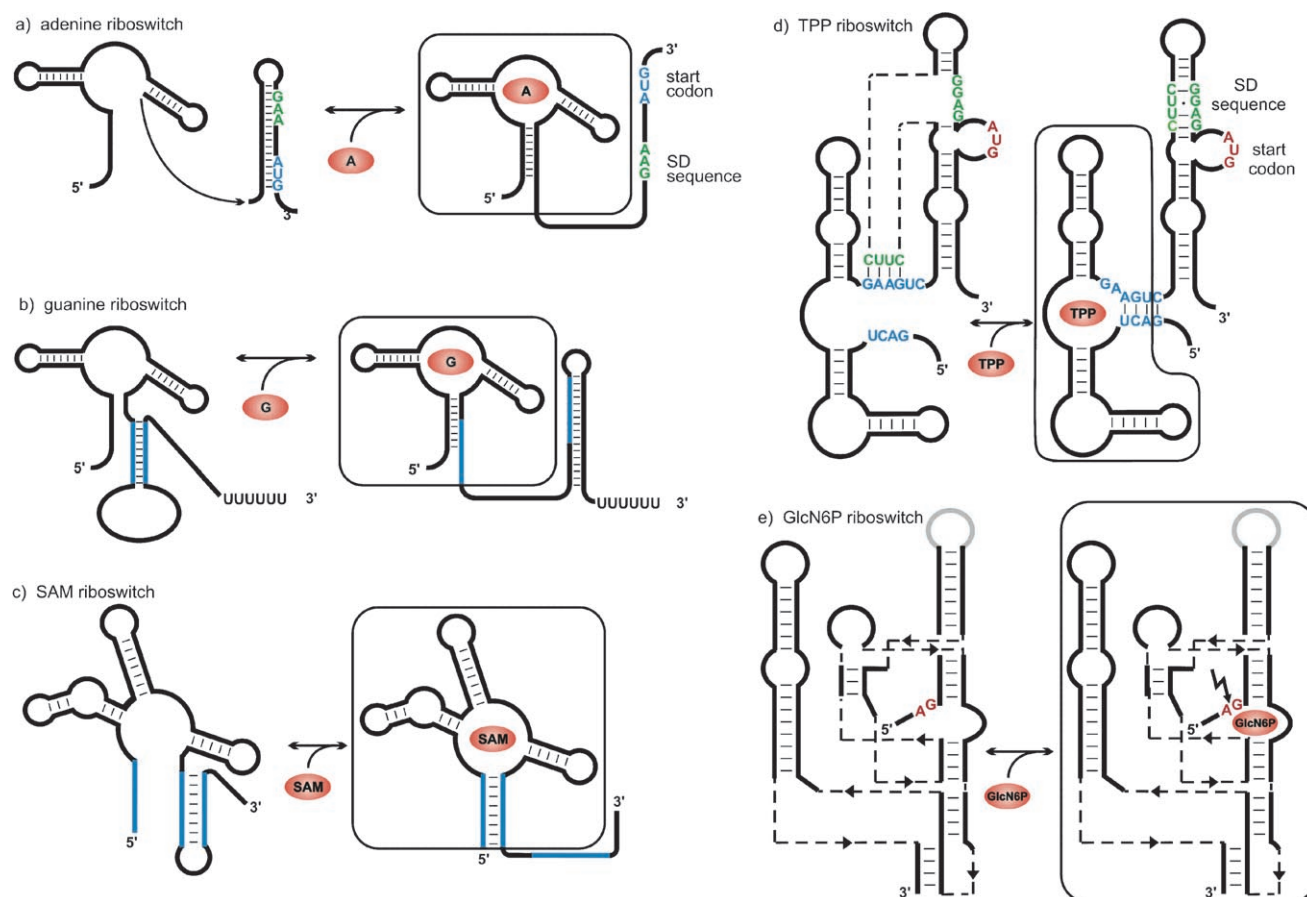
Boris Fürtig, born in 1978, studied biochemistry at the University of Frankfurt and completed his diploma research project in the group of Prof. Schwalbe in spring 2003. He is currently completing his PhD in the group of Prof. Schwalbe on the investigation of RNA by NMR spectroscopic methods.



Jonas Noeske, born in 1978, studied biochemistry at the University of Frankfurt and completed his diploma research project in the group of Prof. Schwalbe in spring 2004. He is currently working on his PhD in the group of Prof. Schwalbe, investigating riboswitches by NMR spectroscopy.



Janina Buck, born in 1980, studied chemistry at the University of Frankfurt and completed her diploma research project in the group of Prof. Schwalbe in spring 2005. She is currently working on her PhD in the group of Prof. Schwalbe, undertaking kinetic investigations of ligand-induced RNA folding by NMR spectroscopy.



**Figure 1.** RNA switches, whose aptamer domains have been solved: a) adenine riboswitch from *V. vulnificus*, b) guanine riboswitch from *B. subtilis*, c) SAM riboswitch from *T. tengcongensis*, d) TPP riboswitch from *E. coli*, and e) *glmS* ribozyme from *T. tengcongensis*.

ically changes the affinity of the adenine riboswitch to bind to guanine/hypoxanthine (Figure 1b), but no longer bind to adenine as discussed below (Figure 2a,b).<sup>[14]</sup> The guanine/hypoxanthine riboswitch regulates the *xpt-pbuX* operon in *B. subtilis* that codes for genes involved in the recycling of purine and the de novo biosynthesis of purine and acts as a transcriptional off switch.

A more indirect way of gene regulation may be involved for the *glmS* riboswitch which regulates the *glmS* gene in *B. subtilis*. The *glmS* riboswitch is a ribozyme in which gene regulation is not linked to conformational switching between alternating secondary structures. Rather, bond cleavage in the 5'-UTR of mRNA is observed when the ligand glucosamine-6-phosphate binds to the aptamer part. It is proposed that the in vitro enzymatic activity of the *glmS* riboswitch is the cellular regulation principle, as it would be in agreement with the in vivo observed down-regulation of gene expression.<sup>[7]</sup>

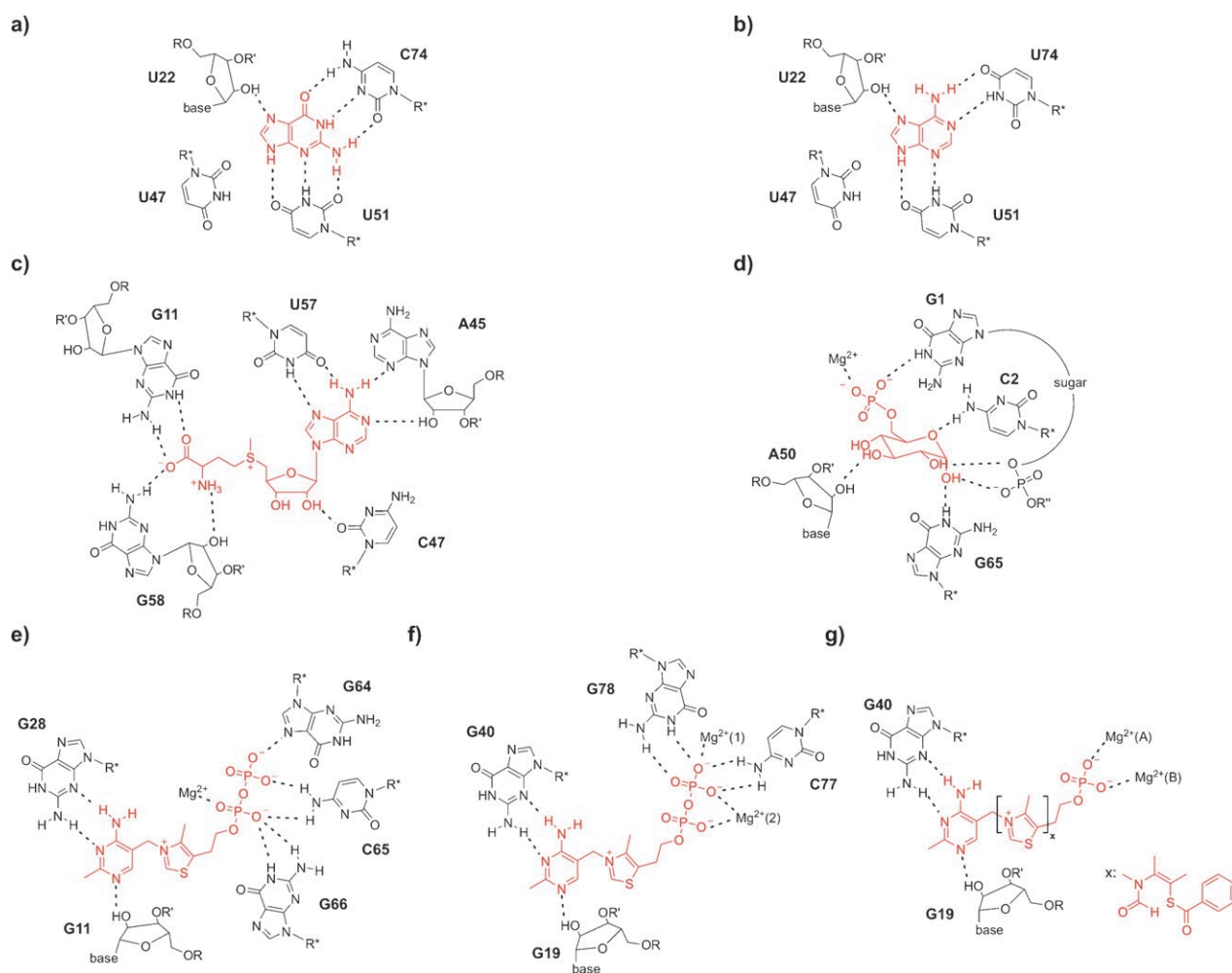
### 3. Molecular Recognition of Ligands by RNA Switches

High-resolution structures are available for the complexes of the aptamer domains of the guanine/adenine-binding riboswitches,<sup>[8]</sup> the TPP-binding riboswitches,<sup>[9]</sup> the SAM-

binding riboswitch,<sup>[10]</sup> and the *glmS* ribozyme<sup>[11,12]</sup> (see Figure 1 and Table 1). Some of these structures have already been described in review articles.<sup>[13]</sup>

The exact nature of the RNA–ligand interactions in the different complexes of RNA switches varies widely owing to the chemical diversity of the different ligands. Noncanonical structural elements such as base triples and base quadruples and non-Watson–Crick base pairs are utilized to form the ligand-binding pockets. The structures reveal a multitude of novel interaction possibilities that show how RNA can target small molecules. Therefore, beside the interest in revealing the regulation mechanisms of riboswitches, these structures are of great importance in designing novel small molecules that bind RNA with high affinities ( $K_D < 100$  nM) and specificities.

Biophysical studies have revealed that the differences in the chemical structures of purine analogues have a great influence on the binding affinities of these ligands to purine riboswitches.<sup>[5a,14]</sup> The X-ray crystal structures<sup>[8a,b,d]</sup> and NMR studies<sup>[8c]</sup> of the ligand-bound state of the purine riboswitches, notably the guanine and the adenine riboswitch, show the formation of a complex binding pocket, where almost all positions of the heterocyclic purine metabolite are recognized by the RNA (Figure 2a,b). For both riboswitches, the N3/N9 edge of the ligand is bound to uridine residue U51, the N7 of



**Figure 2.** Molecular recognition of ligands (red) by RNA switches: a) guanine by the guanine riboswitch from *B. subtilis*, b) adenine by the adenine riboswitch from *B. subtilis*/*V. vulnificus*, c) SAM by the SAM riboswitch from *T. tengcongensis*, d) glucose-6-phosphate by the *glmS* ribozyme from *T. tengcongensis*, e) TPP by the TPP riboswitch from *A. thaliana*, f) TPP by the TPP riboswitch from *E. coli*, and g) TMP by the TPP riboswitch from *E. coli* (x: benfotiamine).

**Table 1:** Summary of X-ray crystal structures of solved aptamer domains of RNA switches.

RNA switch	Regulated gene	Ligand	Resolution [Å]	PDB accession code	Ref.
guanine	<i>xpt-pbuX</i> in <i>B. subtilis</i>	hypoxanthine	1.95	1U8D	[8a]
guanine	<i>xpt-pbuX</i> in <i>B. subtilis</i>	guanine	2.4	1Y27	[8b]
guanine C74U mutant	n.a.	2,6-diaminopurine	2.15	2B57	[8d]
adenine	<i>add</i> in <i>V. vulnificus</i>	adenine	2.1	1Y26	[8b]
thiamine pyrophosphate	<i>thiC</i> in <i>A. thaliana</i>	TPP	2.9	2CKY	[9a]
thiamine pyrophosphate	<i>thiM</i> in <i>E. coli</i>	TPP	2.05	2GDI	[9b]
S-adenosylmethionine	<i>metF-metH2</i> in <i>T. tengcongensis</i>	SAM	2.9	2GIS	[10]
thiamine pyrophosphate	<i>thiM</i> in <i>E. coli</i>	TPP	2.5	2HOJ <sup>[a]</sup>	[9c]
thiamine pyrophosphate	<i>thiM</i> in <i>E. coli</i>	TMP	2.89	2HOM	[9c]
thiamine pyrophosphate	<i>thiM</i> in <i>E. coli</i>	benfotiamine	3.0	2HOO	[9c]
thiamine pyrophosphate	<i>thiM</i> in <i>E. coli</i>	pyrithiamine	3.3	2HOP	[9c]
<i>glmS</i> ribozyme	<i>glmS</i> in <i>T. tengcongensis</i>	glucose-6-phosphate	2.10	2HO7 <sup>[a]</sup>	[12]

[a] See also: 2HOK and 2HOL. [b] See also: 2GCV, 2H0S, 2H0W, 2H0X, 2H0Z, 2H06, and 2GCS. n.a.: not applicable.

the Hoogsteen edge of the purines shows a hydrogen bond to the 2'-OH group of U22, and the X-ray crystal structures report to some extent<sup>[8a,b]</sup> a hydrogen-bonding interaction of N9 to residue U47. The main difference in the two purine riboswitches is the nucleotide in the 74-position which allows

for a specific Watson–Crick base-pairing interaction of the ligand with the complementary nucleotide. In the case of the ligand guanine the nucleotide of the RNA is therefore a cytidine, while the ligand adenine is recognized by a uridine. Direct binding of the RNA to the N3/N9 edge of the purine



ligands discriminates the nucleobase against N9-glycosylated nucleosides and nucleotides which are also present in the cell.

Similarly, in the SAM riboswitch almost all functional groups of the ligand are involved in ligand recognition by the RNA (Figure 2c).<sup>[15,10]</sup> The ligand SAM can structurally be divided into two subunits: an adenosyl moiety and a chain derived from methionine which carries a methylated, positively charged sulfur atom and the charged  $\alpha$ -amino acid group attached to the 5'-end of the nucleoside. The compact conformation of the bound ligand can be pictured with the two subunits stacking on each other and therefore stabilizing this conformation through  $\pi$ -cation interactions. The adenine heterocycle is involved in the formation of a base triple with the flanking nucleobases A45 and U57. The 2'-OH sugar hydroxy group of the ligand is hydrogen-bonded to position O4 in residue C47. The oxygen atoms of the terminal carboxylate group show hydrogen-bonding interactions to the Watson-Crick site of residue G11 and to the exocyclic amino group of nucleotide G58; both nucleotides are part of the base triple G11·C44-G58. The positively charged, methylated sulfur atom is within about 4 Å distance of the O2 atoms of U7 and U88 and can therefore participate in electrostatic interactions. This interaction explains the approximately 100-fold higher  $K_D$  value of the ligand SAM compared to the neutral ligand S-adenosylhomocysteine (SAH), in which the sulfur atom is neither methylated nor positively charged. However, the methyl group bound to the sulfur atom is not directly recognized by the RNA, as seen in the crystal structure of the SAM riboswitch.<sup>[11d]</sup> Interestingly, ligand binding to the RNA switch neither involves magnesium ions in the binding pocket to compensate the negatively charged carboxylate group of the ligand nor involves the phosphate backbone in direct interactions with the ligand to compensate the net charges of the positive functional groups of the ligand.

The crystal structure of the *glmS* ribozyme bound to glucose-6-phosphate (Glc6P)<sup>[12]</sup> reveals the ligand-bound state of this catalytic riboswitch and gives insight into the catalytic activity. Despite its isosteric structure compared to the activating ligand glucosamine-6-phosphate (GlcN6P), Glc6P is a competitive inhibitor and therefore the RNA is still in its precleaved state in the presence of Glc6P (Figure 2d). The crystal structures of the ligand-free RNA and the RNA following cleavage yield information about the correlation of ligand binding and ribozyme function. The catalytic activation takes place in the binding pocket, which acts as the active site of the ribozyme, therefore ligand binding, ligand recognition, and ligand-induced activation are correlated.

Compared to other riboswitch structures, the ligand-binding pocket is rigid and preorganized<sup>[16,12]</sup> and is not characterized through a complete embedding of the bound ligand but instead through a more open, solvent-accessible form. However, ligand recognition is specific and activation of the catalytic function is significant. The structure of the riboswitch-ligand complex provides an explanation for this observation because the entire ligand with its sugar moiety and phosphate group is involved in the interaction. Cleavage requires the presence of the amino group as in GlcN6P. Activation seems to require the correct orientation of the

functional group and its capacity to take part in a proposed acid-base reaction. This observation also explains the necessity of water molecules to be present in the active site during the cleavage process. Complexes with trapped ligands that lead to cleavage have not yet been reported.

Hydrogen-bonding interactions that stabilize the RNA-ligand complex are formed between Glc6P, which is bound as the  $\alpha$ -anomer, and the unpaired nucleotide G1. The nucleotide G1 surrounds the ligand and binds to the phosphate and the sugar hydroxy groups. Furthermore, the ligand forms hydrogen bonds to C2, G65, and the 2'-OH group of A50. The negative charge of the phosphate group is compensated by the coordination of a magnesium ion, which lies approximately 10 Å from the scissile phosphate and is therefore not involved in catalysis. In addition, the 2'-OH group of the ligand forms a hydrogen bond to one water molecule found in the active site.

In contrast to many other riboswitches, not all functional groups of the biologically active metabolite TPP are recognized by the TPP riboswitch (Figure 2e-g). TPP itself can structurally be divided into a pyrimidine heterocycle, a thiazole ring, and a pyrophosphate moiety. As seen in all published crystal structures<sup>[9a,b,c]</sup> of this riboswitch (see Table 1), the ligand moiety with the two heterocycles adopts an extended form whereas the pyrophosphate adopts a bent conformation. In the TPP-bound state, two helices are oriented in parallel fashion. The pyrimidine ring of the ligand binds to one helix, and the pyrophosphate end binds to the other helix. The pyrimidine ring and the phosphate moiety of the ligand are directly recognized, in contrast to the bridging thiazole ring, which is not directly bound and therefore can also be replaced by other elements (Figure 2g).<sup>[9c,17]</sup> All reported crystal structures reveal the presence of magnesium ions in the pyrophosphate binding pocket, compensating the negatively charged phosphate groups, a finding that is in agreement with the biochemically and NMR spectroscopically observed strong magnesium dependence of ligand binding to the RNA.<sup>[9d,17]</sup> The Watson-Crick site of the pyrimidine ring is found in all published structures to form hydrogen bonds with the sugar edge of a guanosine residue. In addition, one hydrogen bond is formed between position N1 of the ligand and a 2'-OH group of another guanosine nucleotide. The pyrophosphate binding pocket and the recognition of the pyrophosphate moiety differ in the various riboswitch elements derived from different organisms, and both depend on the bound ligand.

The structure of the TPP riboswitch from *A. thaliana* reported by Ban and co-workers<sup>[9a]</sup> (Figure 2e) shows direct hydrogen bonds of G64, C65, and G66 to the pyrophosphate end of the ligand and further includes the participation of one magnesium ion in this binding pocket. The structure of the TPP riboswitch from *E. coli* published by Patel and co-workers<sup>[9b]</sup> (Figure 2f) shows a different bonding scheme in the pyrophosphate binding pocket. Here, the phosphate groups are coordinated by two magnesium ions and furthermore form hydrogen bonds to the residues C77 and G78. Edwards et al.<sup>[9c]</sup> reported that besides the magnesium-mediated contacts a hydrogen bond between G78 and a nonbridging oxygen atom of the terminal phosphate is the only direct contact from the ligand to the pyrophosphate

binding pocket. The crystal structure of the riboswitch complex with thiamine monophosphate (TMP)<sup>[9c]</sup> shows in general the same hydrogen-bonding pattern as that with TPP (Figure 2g). In contrast to the TPP-bound RNA, there is no hydrogen bond between the ligand and nucleotide G78 in TMP-bound RNA. In the TPP-bound structure, one magnesium ion binds to both phosphate groups and, in addition, a second magnesium ion binds to the terminal phosphate group. In the TMP-bound structure, two magnesium ions bind to the only (terminal) phosphate group. Replacement of the ligand TMP with benfotiamine (BTP)<sup>[9c]</sup> in which the central heterocycle is not a thiazole (Figure 2g with inset x), does not change the ligand-recognition mode compared to the TMP–RNA complex. Moreover, the crystal structure of the riboswitch complex with the ligand pyrithiamine (PT)<sup>[9c]</sup> was reported (structure not shown). Although the pyrimidine part of the ligand does not change, no hydrogen bonds could be observed between the guanosine nucleotide (G40) and the ligand. Instead, the only direct contact between the RNA and the ligand is the hydrogen bond between N1 of PT and the 2'-OH group of a guanosine residue (G19), an interaction already described for all other ligands.

#### 4. Secondary and Tertiary Interactions That Stabilize the Aptamer Domains of RNA Switches

In contrast to the smaller SELEX-derived aptamers, the structures of RNA switches also reveal numerous long-range RNA–RNA tertiary interactions, such as pseudoknots, that are not involved in the formation of the ligand-binding pocket but instead stabilize the global fold of the RNA structure. The folds of the aptamer domains of the riboswitches determined thus far are very compact owing to an extensive packing of the helical elements. This packing is facilitated by known motifs, for example, type I/II A-minor triple motifs, as well as newly observed structural elements, such as the base quadruples in the purine-binding riboswitches.<sup>[18]</sup>

On the basis of the aptamer domain architecture, we can classify the riboswitch structures into two general types:

- Type I riboswitches reveal a three-way junction as the primary global architecture of their aptamer domains; members of type I include the purine-binding riboswitches, TPP-binding riboswitches, and the *glmS* ribozyme.
- Type II riboswitches include motifs where the aptamer domain is arranged in a four-way junction architecture; the structure of the SAM-I-binding riboswitch is a representative example.

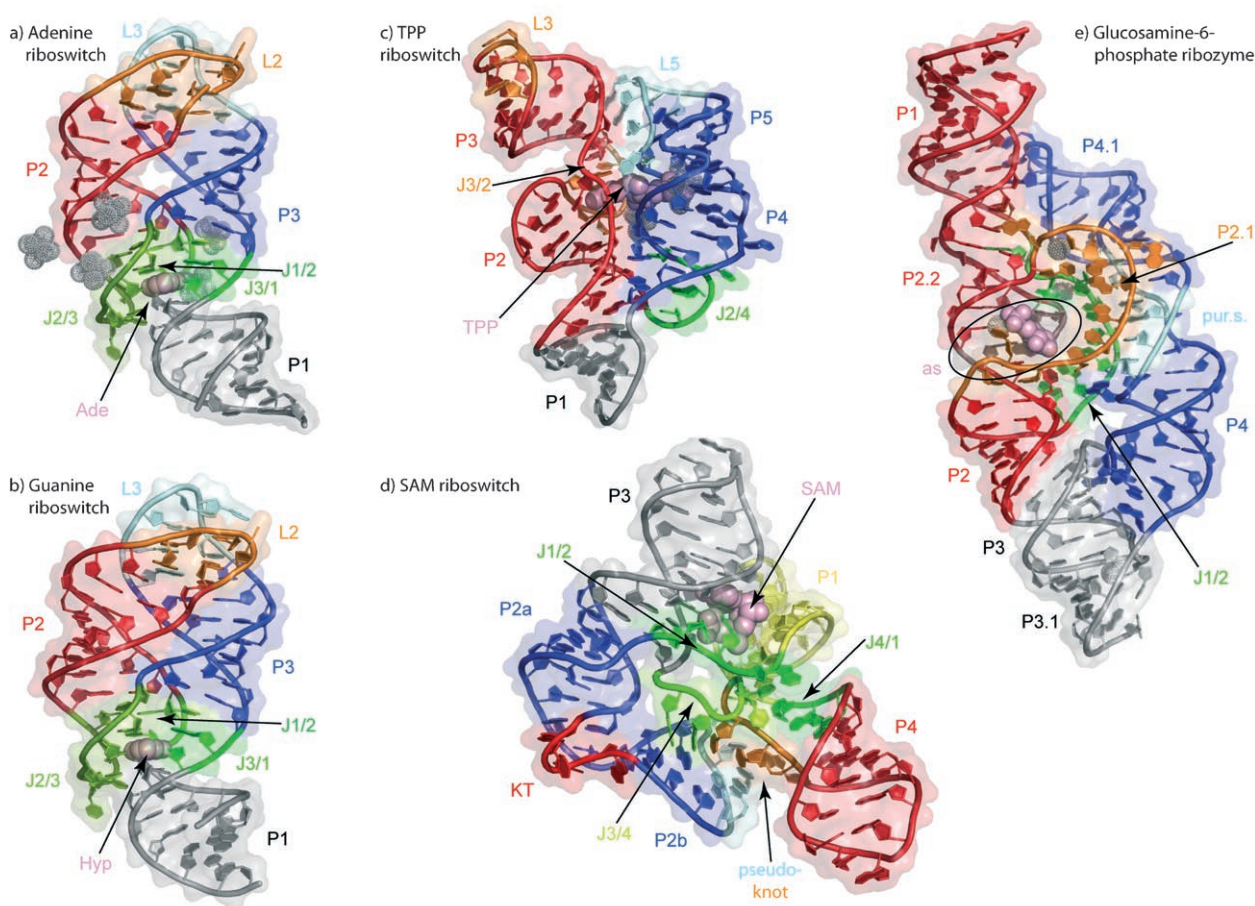
##### 4.1. Type I RNA Switches

The aptamer domains of the purine riboswitches bound to the respective ligand show a fundamental three-way junction architecture in which two parallel helices P2 and P3 are tightly packed together by a series of interconnecting hydrogen bonds between the apical loops L2 and L3 (Figure 3a,b). These hydrogen bonds span two base quadruples composed in

each case of a G–C Watson–Crick base pair and a non-canonical A·A or A·U base pair that interacts from the minor-groove side. The juxtaposition of the negatively charged RNA backbone regions which results in unfavorable electrostatic interactions is compensated by the binding of positively charged cations. The purine-binding pocket is located in the three-way junction element where base triples above (water-mediated base triple U22–A52·A73 and base triple A23·G46–C53) and below (base triple U20–A76·U49 and base triple A21–U75–C50) flank the ligand and form a nearly closed envelope. Binding of the ligand is critically dependent on the formation of the tertiary architecture, as revealed by the mutation introducing a tetraloop (four-nucleotide loop) as capping motif for the parallel-packed helices. This mutation renders the riboswitch incapable of ligand binding. NMR studies show that the loop–loop interaction is preorganized even in the absence of ligand.<sup>[18]</sup>

A similar three-way junction architecture is observed for the TPP riboswitches (Figure 3c). Here, two helical elements are arranged in a parallel manner. In contrast to the purine riboswitches, the interacting helical elements are not made up of continuous helices but in each case of two helical arms that are connected by junction elements (P2/P3 with J2/3, and P4/P5 with J4/5). While ligand binding in the purine riboswitches is facilitated by the junction element of the central three-way junction, binding of TPP and its analogues involves the junction element J4/5 as a sensing element for the pyrophosphate and J2/3 as a sensing element for the pyrimidine moiety. Binding of TPP is one of the interactions to bridge the helical arms and shapes the whole molecule into its signaling form. Furthermore, this architecture is facilitated by a helix–loop interaction between helix P3 and loop L5 that is stabilized by monovalent cations.

The structure of the *glmS* ribozyme belongs to type I riboswitches with a three-way junction element. In this case, however, the three-way junction is not the major determinant of the tertiary structure of this riboswitch. The overall structure of this RNA molecule is dominated by three coaxial stacks of helices that are arranged side by side. Two helical elements, one consisting of helices P1, P2.2, P2, P3, and P3.1 ( $\approx 100$  Å) and the other consisting of P4 and P4.1, surround the shorter central helical stack P2.1. The ribozyme is therefore a highly anisotropic molecule. The central feature of the structure is the double-pseudoknot core of the ribozyme that allows the positioning of the scissile phosphate group located at the 5'-end of helix P2.2 into the major groove of helix P2.1. To accomplish such a double-pseudoknot arrangement of the helices P2.1 and P2.2, four nonhelical crossovers are needed and are facilitated by a number of noncanonical interactions. The upper two crossovers have been classified as the roof of the active site by Klein et al.<sup>[12]</sup> They are part of two conserved base triples (G34·G7–C30 and A35·C36–G53) that are the branch points for the three-way junction between P1, P2.2, and P2.1. Interestingly, on the floor side of the active site, two nucleotides are threaded through the loop between P2 and P2.1 and then stack onto the nucleotides of either one of the lower crossovers. Whereas in the lower part of the molecule a canonical pseudoknot comprising helices P3 and P3.1 can be found that stacks onto



**Figure 3.** Three-dimensional structures of the aptamer domains of RNA switches (all RNA structures are shown as ribbon representations, ligands are shown in sphere representations and colored pale violet, and the surface is lightly shaded): a) complex of adenine riboswitch from *V. vulnificus* with adenine (pdb accession code: 1Y26), b) complex of guanine riboswitch from *B. subtilis* with hypoxanthine (1U8D), c) complex of TPP riboswitch from *E. coli* with TPP (2GDI), d) complex of SAM riboswitch from *T. tengcongensis* with SAM (2GIS), and e) complex of *glmS* ribozyme (before cleavage) from *T. tengcongensis* with Glc6P (2HO7). Structural elements are color coded (see text); the numbering follows the nomenclature given in the original publications; P: helices; J: junctional elements; L: loops; KT: kink turns; as: active side; pur.s.: purine stretch.

P2, the helices P4 and P4.1 form a coaxial stack that is a scaffolding element and acts as an abutment for P2.1. The positioning of this scaffold is maintained by several tertiary interactions: an A-minor interaction between A117 of the tetraloop L4.1 and C10–G31 of P1, and a six-purine stack consisting of A129/G128/A127/A104/A105/A106 that is formed by the internal loop between P4 and P4.1, enclosed in the minor groove of P2.1. As indicated by the low all-atom RMSD difference in the pre- and postcleavage state of the RNA molecule, the *glmS* ribozyme does not undergo a major conformational switching event upon ligand binding but just the cleavage of the scissile phosphodiester bond, and is therefore different in the mode of action to the other riboswitches described here.

#### 4.2. Type II RNA Switches

In contrast to the riboswitch structures described so far, the SAM riboswitch has a four-way junction as its central structural element (Figure 3d). Around this four-way junction, stems and stem-loop structures are found. Two sets of

coaxially stacked helices are found: P2a and P3 as well as P1 and P4 with the intervening junction element J4/1. These two stacks are held together by the pairing of residue A85 in J4/1 with U64 in J3/4 and A24 in L2. This architecture holds the two stacks at an angle of 70° which results in a binding pocket for SAM that sits in the minor grooves of P1 and P3.

Two ligand-independent tertiary structure motifs should be mentioned that are necessary for the global architecture. On such motif is a kink-turn motif, which connects the helices P2a and P2b. This kink enables the formation of a pseudoknot structure by the pairing of L2 and J3/4. The structure of the pseudoknot is further stabilized by base triples between residues A85–U64–A24 which reside in P1 and P4, and by interaction of residues from J3/4 and P2b that form adenine-mediated base triples.

#### 5. Open Questions Regarding the Regulation Mechanisms of RNA Switches

Biophysical studies pose challenges to a simple allosteric model of conformational switching to explain translational or



transcriptional regulation. According to this model, ligand binding to the aptamer domain would induce structural transitions in the secondary structure elements of the mRNA and release or mask regulation elements involved in transcription as well as translation (see Figure 1). However, as revealed by stopped-flow fluorescence studies<sup>[19,8d]</sup> and by time-resolved NMR spectroscopy<sup>[20]</sup> for the purine riboswitches and the FMN riboswitch, the rate constants at physiological ligand concentrations are too slow for a model in which ligand binding precedes the conformational switch and thereby regulates gene expression. Rather, the mechanism has to consider that the gene regulation is exerted during transcription and general protein transcription termination factors may also be important. In light of these comments, we would like to conclude this review article with a quote from Nudler:<sup>[13b]</sup> “Therefore, to fully understand riboswitch functioning, the atomic resolution picture of ligand–RNA contacts needs to be put in the context of the whole expression platform, its folding process, and kinetic coupling dictated by the RNA polymerase.”

*The work was supported by the DFG (SFB 579: “RNA–Ligand Interaction”), the Fonds der Chemischen Industrie (predoctoral fellowship to J.N.), the Studienstiftung des Deutschen Volkes (predoctoral fellowship to B.F.), and the state of Hesse (Center for Biomolecular Magnetic Resonance (BMRZ)).*

Received: October 10, 2006

Revised: November 13, 2006

Published online: January 17, 2007

- [1] a) A. D. Ellington, J. W. Szostak, *Nature* **1990**, *346*, 818–822; b) S. J. Klug, M. Famulok, *Mol. Biol. Rep.* **1994**, *20*, 97–107; c) J. Joyce, *Curr. Opin. Struct. Biol.* **1994**, *4*, 331–336.
- [2] a) M. Famulok, *Curr. Opin. Struct. Biol.* **1999**, *9*, 324–329, and references therein; b) T. Hermann, D. J. Patel, *Science* **2000**, *287*, 820–825.
- [3] D. H. Burke, L. Gold, *Nucleic Acids Res.* **1997**, *25*, 2020–2024.
- [4] a) J. Miranda-Rios, M. Navarro, M. Soberon, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9736–9741; see also: G. D. Stormo, Y. Ji, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 9465–9467; b) W. Winkler, A. Nahvi, R. R. Breaker, *Nature* **2002**, *419*, 952–956; c) A. Mironov, I. Gusarov, R. Rafikov, L. Lopez, K. Shatalin, R. Kreneva, D. Perumov, E. Nudler, *Cell* **2002**, *111*, 747–756; d) S. D. Gilbert, R. T. Batey, *Cell. Mol. Life Sci.* **2005**, *62*, 2401–2404.
- [5] a) M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, *Cell* **2003**, *113*, 577–586; b) N. Sudarsan, J. E. Barrick, R. R. Breaker, *RNA* **2003**, *9*, 644–647; c) T. Kubodera, M. Watanabe, K. Yoshiuchi, N. Yamashita, A. Nishimura, S. Nakai, K. Gomi, H. Hanamoto, *FEBS Lett.* **2003**, *555*, 516–520.
- [6] N. Sudarsan, J. E. Barrick, R. R. Breaker, *RNA* **2003**, *9*, 644–647.
- [7] W. C. Winkler, A. Nahvi, A. Roth, J. A. Collins, R. R. Breaker, *Nature* **2004**, *428*, 281–286.
- [8] a) R. T. Batey, S. D. Gilbert, R. K. Montange, *Nature* **2004**, *432*, 411–415; b) A. Serganov, Y. R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, D. J. Patel, *Chem. Biol.* **2004**, *11*, 1729–1741; c) J. Noeske, C. Richter, M. A. Grundl, H. R. Nasiri, H. Schwalbe, J. Wöhnert, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1372–1377; d) S. D. Gilbert, C. D. Stoddard, S. J. Wise, R. T. Batey, *J. Mol. Biol.* **2006**, *359*, 754–768.
- [9] a) S. Thore, M. Leibundgut, N. Ban, *Science* **2006**, *312*, 1208–1211; b) A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, *Nature* **2006**, *441*, 1167–1171; c) T. E. Edwards, A. R. Ferre-D’Amare, *Structure* **2006**, *14*, 1459–1468; d) J. Noeske, C. Richter, E. Stiral, H. Schwalbe, J. Wöhnert, *ChemBioChem* **2006**, *7*, 1451–1456.
- [10] R. K. Montange, R. T. Batey, *Nature* **2006**, *441*, 1172–1175.
- [11] a) J. E. Barrick, K. A. Corbino, W. C. Winkler, A. Nahvi, M. Mandal, J. Collins, M. Lee, A. Roth, N. Sudarsan, I. Jona, J. K. Wickiser, R. R. Breaker, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 6421–6426; b) G. Mayer, M. Famulok, *ChemBioChem* **2006**, *7*, 602–604; c) J. A. Jansen, T. J. McCarthy, G. A. Soukup, J. K. Soukup, *Nat. Struct. Mol. Biol.* **2006**, *13*, 517–523; d) J. Lim, W. C. Winkler, S. Nakamura, V. Scott, R. R. Breaker, *Angew. Chem.* **2006**, *118*, 978–982; *Angew. Chem. Int. Ed.* **2006**, *45*, 964–968.
- [12] D. J. Klein, A. R. Ferre-D’Amare, *Science* **2006**, *313*, 1752–1756.
- [13] a) S. Reichow, G. Varani, *Nature* **2006**, *441*, 1054–1055; b) E. Nudler, *Cell* **2006**, *126*, 19–22; c) S. D. Gilbert, R. T. Batey, *Chem. Biol.* **2006**, *13*, 805–807.
- [14] M. Mandal, R. R. Breaker, *Nat. Struct. Mol. Biol.* **2004**, *11*, 29–35.
- [15] W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, R. R. Breaker, *Nat. Struct. Biol.* **2003**, *10*, 701–707.
- [16] K. J. Hampel, M. M. Tinsley, *Biochemistry* **2006**, *45*, 7861–7871.
- [17] T. Yamauchi, D. Miyoshi, T. Kubodera, A. Nishimura, S. Nakai, N. Sugimoto, *FEBS Lett.* **2005**, *579*, 2583–2589.
- [18] J. Noeske, J. Buck, B. Fürtig, H. Nasiri, H. Schwalbe, J. Wöhnert, *Nucleic Acids Res.* **2007**, in press.
- [19] a) J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, *Mol. Cell* **2005**, *18*, 49–60; b) J. K. Wickiser, M. T. Cheah, R. R. Breaker, D. M. Crothers, *Biochemistry* **2005**, *44*, 13404–13414.
- [20] a) J. Buck, B. Fürtig, J. Noeske, J. Wöhnert, H. Schwalbe, unpublished results; b) J. Buck, diploma thesis, Frankfurt, **2005**.



# Time-resolved NMR methods resolving ligand-induced RNA folding at atomic resolution

Janina Buck\*, Boris Fürtig\*, Jonas Noeske\*†, Jens Wöhnert\*†, and Harald Schwalbe\*\*

\*Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Johann Wolfgang Goethe-University, Max von Laue-Strasse 7, 60438 Frankfurt am Main, Germany; and †Department of Biochemistry, University of Texas Health Science Center, 7703 Floyd Curl Drive, San Antonio, TX 78229

Edited by Alfred G. Redfield, Brandeis University, Waltham, MA, and approved August 9, 2007 (received for review April 5, 2007)

Structural transitions of RNA between alternate conformations with similar stabilities are associated with important aspects of cellular function. Few techniques presently exist that are capable of monitoring such transitions and thereby provide insight into RNA dynamics and function at atomic resolution. Riboswitches are found in the 5'-UTR of mRNA and control gene expression through structural transitions after ligand recognition. A time-resolved NMR strategy was established in conjunction with laser-triggered release of the ligand from a photocaged derivative *in situ* to monitor the hypoxanthine-induced folding of the guanine-sensing riboswitch aptamer domain of the *Bacillus subtilis* *xpt-pbuX* operon at atomic resolution. Combining selective isotope labeling of the RNA with NMR filter techniques resulted in significant spectral resolution and allowed kinetic analysis of the buildup rates for individual nucleotides in real time. Three distinct kinetic steps associated with the ligand-induced folding were delineated. After initial complex encounter the ligand-binding pocket is formed and results in subsequent stabilization of a remote long-range loop-loop interaction. Incorporation of NMR data into experimentally restrained molecular dynamics simulations provided insight into the RNA structural ensembles involved during the conformational transition.

spectroscopy | riboswitches | dynamics | purine | caged compound

Structural transitions of proteins and RNA constitute an important aspect of cellular function and information transfer. In proteins, the kinetics of these structural transitions, mainly from an unfolded ensemble to a single unique folded state, can be investigated by hydrogen exchange experiments monitored by NMR spectroscopy (1). Such studies have profoundly influenced our understanding of protein-folding pathways. In hydrogen exchange experiments, labile hydrogen atoms become protected against exchange with the bulk solvent during folding. This acquired exchange protection indicates formation of a persistent structure at atomic resolution. Incorporation of structural information derived from exchange experiments into molecular dynamics (MD) simulations (2), including data from methods exhibiting time and atomic resolution together with  $\phi$ -value analysis (3) derived from mutational work, has provided detailed structural information of intermediates populated during folding.

RNA molecules can adopt a variety of secondary and tertiary conformations (4, 5). In general, the energetic difference between alternate RNA conformations is very small, and the equilibrium distribution is strongly affected through the binding of proteins (6), ions (7), or small metabolites (8–10), or by structural modifications (11). Alternate RNA structures are stabilized by different Watson–Crick base-pairing interactions (12). To date, RNA folding has been investigated by using x-ray footprinting (13), oligonucleotide hybridization, and classical biochemical methods in conjunction with mutational data (14). Although hydrogen exchange rates can be used as reporters of RNA ground-state dynamics and stability (15–17), RNA hydrogen exchange experiments conducted in a pulse–chase manner

fail in most cases because of the high intrinsic exchange rates observed even in folded RNA structures.

Here, we report on ligand-induced conformational changes of an RNA at atomic resolution by using real-time NMR methods. We investigated the hypoxanthine-induced folding of the guanine-sensing riboswitch aptamer domain (GSR<sup>apt</sup>) of the *Bacillus subtilis* *xpt-pbuX* operon (18). Riboswitch RNAs have emerged as an important example of macromolecular structural transitions that lead to transcriptional or translational regulation of protein expression induced through the specific binding of a metabolite (reviewed in refs. 19 and 20). Riboswitches are located in the 5'-untranslated region (5'-UTR) of bacterial mRNA and consist of a highly specific metabolite receptor region (aptamer domain) coupled to a 3'-downstream sequence (expression platform). Gene expression is thought to be modulated in response to conformational differences between the ligand-bound and ligand-free conformations of the aptamer domain (21). Crystal structures of the ligand-bound aptamer domains belonging to the class of purine riboswitches (guanine- and adenine-sensing riboswitches) have revealed the presence of complex tertiary structures (22, 23). The heterocyclic ligand in these RNA–ligand complexes is almost completely embedded within the RNA through interactions between the RNA and almost all of the ligand functional groups (24). The guanine-sensing riboswitch binds the ligands guanine ( $K_d \approx 5$  nM) and hypoxanthine ( $K_d \approx 50$  nM) (18) with high affinity and specificity. The mode of molecular recognition for both ligands is the same (22, 23). To date, kinetic and thermodynamic studies concerning the ligand-induced structural rearrangement within the aptamer domain of the adenine-sensing riboswitch have revealed a folding event that operates on a slow time course over a period of seconds in the presence of  $Mg^{2+}$  ions (25–27).

We successfully trapped the ligand-free conformation of the riboswitch RNA through the use of a caged ligand. The application of photolabile “caged” compounds, which allows for the fast release of initially blocked functional groups and subsequent detection of the reaction time course, has been used in real-time NMR studies to investigate protein (28, 29) and RNA folding (30). Laser-induced deprotection within the NMR tube by direct coupling of laser optics (28, 30) allows for *in situ* photochemical release of ligand, thereby resulting in fast initiation and subsequent NMR spectroscopic detection of binding events with

Author contributions: J.B. and B.F. contributed equally to this work; J.B., B.F., J.N., J.W., and H.S. designed research; J.B., B.F., J.N., J.W., and H.S. performed research; J.B., B.F., J.N., J.W., and H.S. analyzed data; and J.B., B.F., J.N., J.W., and H.S. wrote the paper.

The authors declare no conflict of interest.

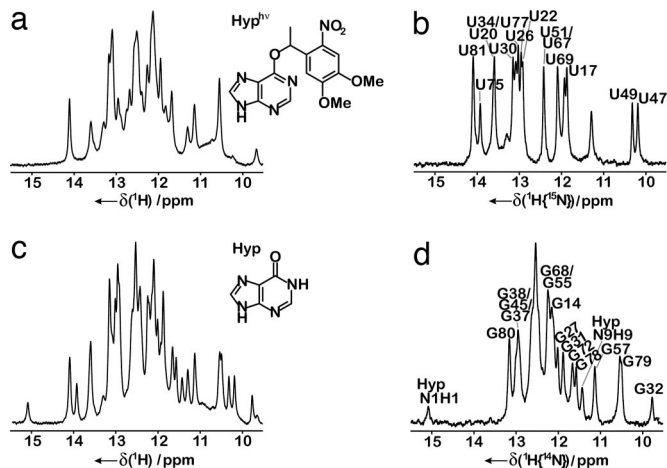
This article is a PNAS Direct Submission.

Abbreviations: DMNPE-hypoxanthine, O<sup>6</sup>-[4,5-dimethoxy-(2-nitrophenyl)ethyl]-hypoxanthine; MD, molecular dynamics; ASR<sup>apt</sup>, adenine-sensing riboswitch aptamer domain; GSR<sup>apt</sup>, guanine-sensing riboswitch aptamer domain.

\*To whom correspondence should be addressed. E-mail: schwalbe@nmr.uni-frankfurt.de.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0703182104/DC1](http://www.pnas.org/cgi/content/full/0703182104/DC1).

© 2007 by The National Academy of Sciences of the USA



**Fig. 1.** Imino proton spectra of GSR<sup>apt</sup>. (a) Unlabeled RNA before laser pulse. (*Inset*) Chemical structure of DMNPE-hypoxanthine (Hyp<sup>hu</sup>). (b) [<sup>15</sup>N]uridine-labeled RNA, uridine residues as a result of <sup>1</sup>H/<sup>15</sup>N-detection after laser pulse with annotated NMR resonance assignment (31) of resolved residues. (c) Unlabeled RNA after laser pulse. (*Inset*) Chemical chemical structure of hypoxanthine (Hyp). (d) [<sup>15</sup>N]uridine-labeled RNA, guanosine residues as a result of <sup>1</sup>H/<sup>14</sup>N-detection after laser pulse with annotated NMR resonance assignment (31) of resolved residues.

residue-specific resolution. After ligand release, three distinct kinetic steps associated with ligand-induced folding of GSR<sup>apt</sup> could be delineated. Incorporation of time-resolved NMR data into experimentally restrained MD simulations provided insight into the conformational behavior of the RNA structural ensembles involved in each step.

## Results

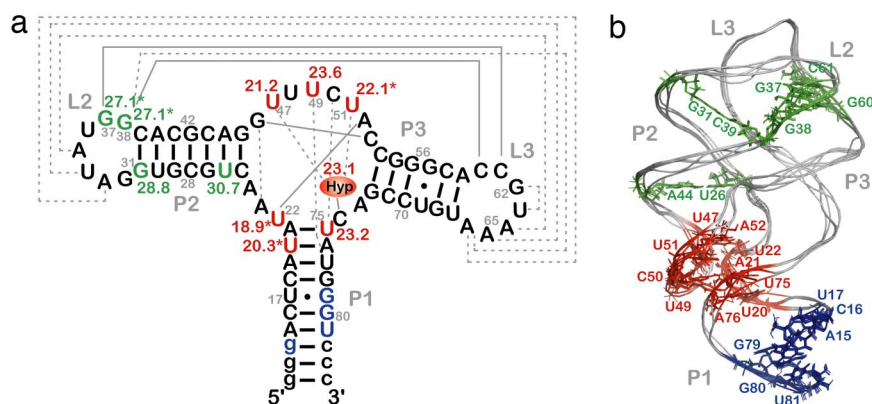
**Two-Step Specific Binding of Ligand to GSR<sup>apt</sup>.** To prevent binding of the small molecule metabolite to GSR<sup>apt</sup>-RNA, the ligand hypoxanthine was caged [*O*<sup>6</sup>-[4,5-dimethoxy-(2-nitrophenyl) ethyl]-hypoxanthine (DMNPE-hypoxanthine)]. (Here, we used the ligand hypoxanthine and its photocaged derivative instead of guanine for solubility reasons.) The photolabile protecting group blocks the Watson–Crick site of the purine ligand known to be essential for ligand recognition and binding to the RNA. Based on the imino proton assignment of GSR<sup>apt</sup> (31), the <sup>1</sup>H-NMR spectrum of the RNA in the presence of DMNPE-hypoxanthine before photolysis indicated that formation of the RNA–ligand

complex is completely suppressed. In contrast, in the time series of NMR spectra recorded after laser irradiation, imino proton signals appeared with increasing intensity representing the ligand-bound state after folding of the RNA, whereas signals corresponding to the ligand-free state decreased (Fig. 1 *a* and *c*; SI Fig. 6).

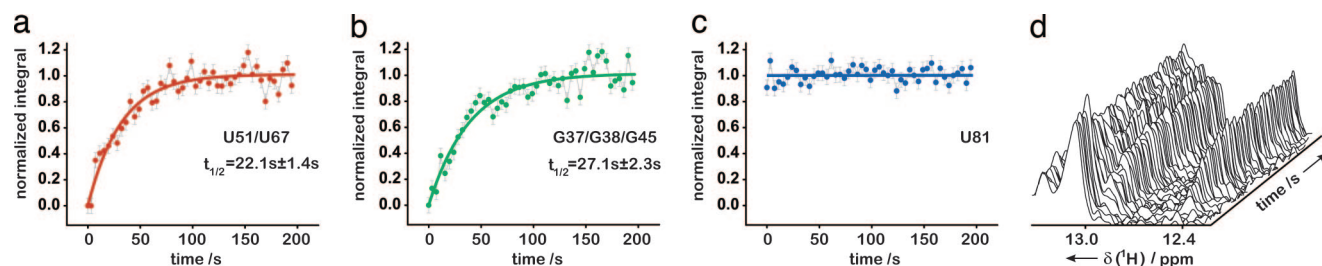
The GSR<sup>apt</sup>-RNA comprises 73 nucleotides and is at the upper size limit for NMR analysis (32). To obtain site-resolved information by one-dimensional real-time NMR methods, selective isotope labeling of the RNA in combination with NMR filter experiments (33) was used. In addition, the signal-to-noise ratio is optimized by selective excitation of imino spin magnetization under Ernst-angle conditions (34). Alternate <sup>14</sup>N/<sup>15</sup>N-labeling of the imino sites of guanosine or uridine nucleotides of the selectively ([<sup>15</sup>N]uridine) labeled RNA allowed editing of NMR resonances according to nucleotide type (Fig. 1) (see [SI Fig. 7](#) and [SI Text](#)). Thus, the time constants reflecting the kinetics of the ligand-induced conformational changes for a large fraction of individual nucleotides could be analyzed.

In general, NMR detects imino proton signals only if they are protected from exchange with solvent because of the formation of hydrogen bonds. Inaccessibility due to hydrophobic or steric exclusion is rare for RNA. Analysis of the water exchange rates for the imino proton signals of the free RNA and the RNA–ligand complex showed identical values for residues located within persistent secondary structure. Therefore, the kinetics induced by ligand binding exclusively reflect structural changes between ligand-free and ligand-bound conformations (see [SI Fig. 8](#) and [SI Text](#) for experimental details).

The site-specific half-lives [ $t_{1/2}$  (s)] for individual signal intensities fell into two different time regimes. Analysis of the kinetic data associated with the ligand-bound RNA structure (22) showed that the observed rates can be related to residues that cluster around two distinct structural elements in the riboswitch RNA (Fig. 2). A faster time course with  $t_{1/2}$  in the range 18.9–23.6 s is observed for residues that are directly involved in formation of the ligand-binding pocket (Figs. 2 and 3; **SI Fig. 9**). These include nucleotides U51, U47, and U22 with  $t_{1/2}$  values of  $22.1 \pm 1.4$  s,  $21.2 \pm 1.9$  s, and  $18.9 \pm 1.3$  s, respectively, that form direct hydrogen bonds with the ligand. In addition, the binding pocket also comprises several base triples flanking the base quadruple formed by hypoxanthine with residues involved in direct hydrogen bonds. The imino proton signals U75, U20, and U49 of the two base triples situated below (base triple A21—U75-C50 and base triple U20—A76-U49) showed  $t_{1/2}$  values of  $23.1 \pm 2.7$  s,  $20.3 \pm 2.1$  s, and  $23.5 \pm 1.9$  s, respectively.



**Fig. 2.** Secondary (a) and tertiary (b) structure of GSR<sup>apt</sup> with kinetic results. Red, half-life values [ $t_{1/2}$  (s)] in the time range 18.9–23.6 s; green, half-life values in the time range 27.1–30.7 s; blue, signals that remain unaffected during the structural transition; asterisk, overlaid signal (for further information, see text); Hyp, hypoxanthine; labeling of helices P1, P2, and P3 and loop regions L2 and L3, according to Breaker *et al.* (18); gray solid lines, Watson–Crick base-pairing interactions; gray dashed lines, noncanonical base-pairing interactions [for construct details, see [supporting information \(SI\) Fig. 5](#)].



**Fig. 3.** Representation of individual signals during time course of reaction. (a–c) Normalized integrals of imino proton signals: (a) red, core region signal U51/U67; (b) green, loop region signal G37/G38/G45; (c) blue, signal U81 that is part of helix P1 as a function of time with monoexponential fit (for signals U51/U67 and G37/G38/G45) and linear fit (for signal U81) (solid line); (d) stack plot of a series of  $^1\text{H}(^{15}\text{N})$ -NMR spectra as a function of time (imino proton subsection, 12.2–13.4 ppm).

Above the ligand-including base quadruple, a water-mediated base triple is formed (U22—A52·A73) and another base triple with residues A23·G46—C53 completes the binding pocket. The NMR resonances of residues U20, U22, and U51 were not resolved from those of residues U34, U77, and U67, respectively. However, faster kinetics of nucleotides in the core region dominated the kinetic behavior of signals U34, U77, or U67 that are all part of the loops or helices of the RNA. As for the ligand hypoxanthine, the half-life measured for the N9-bound proton signal H9 was  $23.1 \pm 3.9$  s, whereas that of the N1-bound proton signal H1 could not be determined because of an insufficient signal-to-noise ratio. Hence, the ligand revealed kinetics similar to the RNA signals of the core segment involved in its recognition.

In contrast, a slower process in the time range of 27.1–30.7 s was observed for nucleotides in helices P2 and P3 and in the loop–loop region (L2 and L3) of the RNA (see SI Fig. 9). The loop–loop interaction is characterized by two base quadruples formed by residues in L2 and L3, respectively (G38—C60 and A33·A66, and G37—C61 and U34·A65) (22, 23). Analysis of time-resolved NMR data for G38 and G37 (overlapped signal of G37/G38/G45, but all residues being located in the loop or helical region) revealed a  $t_{1/2}$  value of  $27.1 \pm 2.3$  s, and U34 could not be analyzed because of signal overlap. In addition, residue G31, which forms part of the loop L2 closing base pair, and nucleotide U26, which forms part of helix P2, showed slow half-lives with  $t_{1/2}$  values of  $28.8 \pm 4.8$  s and  $30.6 \pm 3.0$  s, respectively.

For a number of RNA imino proton resonances, no chemical shift changes in the ligand-free and ligand-bound forms were observed. This related to residues G79, G80, U81, and G14 of the lower part of helix P1, suggesting that the structure surrounding this helix remains constant during the ligand-induced structural transitions (Fig. 3c).

**A Low-Affinity Ligand-Binding Event Precedes the Folding Toward the Ligand-Bound Structure.** Given the slow time scale associated with specific ligand binding to the riboswitch RNA, we considered whether low-affinity binding events preceded specific binding and subsequent folding toward the ligand-bound structure. To this end, useful information can be obtained from the analysis of NMR line widths. The line widths of the nonexchangeable  $^{13}\text{C}$ -bound proton signals H2 and H8 of the (isotope-labeled) ligand were recorded for hypoxanthine free in solution and hypoxanthine tightly bound to the GSR<sup>apt</sup> in the RNA–ligand complex. In addition, the line widths of the same hydrogens were determined for hypoxanthine in the presence of the adenine-sensing riboswitch aptamer domain (ASR<sup>apt</sup>) of *B. subtilis* *pbuE*-mRNA. The ASR<sup>apt</sup> is similar to GSR<sup>apt</sup>, both in sequence and tertiary structure (19, 31). However, ASR<sup>apt</sup> contains a U residue at position 74, but the same position is occupied by a C residue in GSR<sup>apt</sup>. This single mutation is thought to account for the recognition of adenine by ASR<sup>apt</sup> through formation of an

intermolecular Watson–Crick base pair. ASR<sup>apt</sup>, however, showed significantly reduced affinity for the purine base hypoxanthine ( $K_d > 300 \mu\text{M}$ ) (35).

The line width of free hypoxanthine was found to be  $7.5 \pm 1.0$  Hz and increased to  $31.5 \pm 0.6$  Hz in the presence of GSR<sup>apt</sup>, as expected for a tight RNA–ligand complex in slow exchange. Remarkably, in the presence of ASR<sup>apt</sup>, no chemical shift changes were found, but line broadening was observed for the hypoxanthine proton signals. The observed line width was between the values monitored for the free ligand and for hypoxanthine complexed with GSR<sup>apt</sup> (Table 1). The line width of  $11.5 \pm 0.5$  Hz for hypoxanthine in the presence of ASR<sup>apt</sup> indicates low-affinity binding, consistent with millisecond off-rates due to exchange between the ligand-free and ligand-bound forms that is fast on the NMR time scale (Fig. 4).








The same experiments with the purine base adenine showed an identical line-broadening effect for the ligand signals in the presence of the GSR<sup>apt</sup>. In contrast, the line widths of both ligands, hypoxanthine and adenine, in the presence of an RNA construct that lacks the binding site but exhibits the structural elements helices P2 and P3, and loop regions L2 and L3 were only  $9.6 \pm 1.3$  Hz and  $10.3 \pm 0.6$  Hz, respectively (Table 1). Thus, the assumption of a low-affinity interaction with the ligand binding site in the purine riboswitch RNAs is confirmed and a general transient binding to random RNA can be neglected.

**Translation of Kinetic Rates into Structural Information.** We previously showed that in the free form of the riboswitch RNA helices P1, P2, and P3 are preformed and adopt a canonical A-form helix conformation. The core region of GSR<sup>apt</sup> is, in essence, unstructured in the absence of ligand. In contrast, the loop–loop interactions between L2 and L3 (Fig. 2) are present in the ligand-free state (31). However, small chemical shift changes of the nucleotides that form the loops indicate variations in the vicinity of this structural element on ligand binding. Our kinetic results here indicate that on ligand binding, the GSR<sup>apt</sup> folds in a three-state process, in which fast low-affinity binding precedes productive binding of the ligand to the core region of the aptamer domain, followed by a slower process that involves tightening of the tertiary structure involving nucleotides in helices P2 and P3 and the loop–loop regions L2 and L3.

To translate the kinetic rates into structural information, we used our experimental NMR data as restraints in MD simulations akin to strategies used in the protein folding field. We calculated three different structural ensembles in restrained MD simulations based on the clustering of kinetic rates around distinct regions of the aptamer domain (Fig. 2). If an NMR imino proton signal identified by its native chemical shift could be detected, it is reasonable to assume that persistent native hydrogen bonding has been formed. Therefore, the restrained MD simulations starting from a completely randomized RNA chain incorporated native hydrogen-bonding patterns (Fig. 4) when detectable. For the ensemble of



**Table 1. Summary of line width values obtained for  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled ligands in the presence of various RNA constructs**

RNA	Secondary structure	Ligand ( $^{13}\text{C}$ , $^{15}\text{N}$ -labeled)	$\text{Mg}^{2+}$ , mM	Characteristics	$\Delta\sigma$ , Hz	Line width ligand, Hz
No RNA		Hypoxanthine			0.11	$7.5 \pm 1.0$
GSR <sup>apt</sup>		Hypoxanthine		Specific hypoxanthine binding	0.13	$31.5 \pm 0.6$
GSR <sup>apt</sup>		Hypoxanthine	5	Specific hypoxanthine binding	0.12	$30.5 \pm 0.4$
ASR <sup>apt</sup>		Hypoxanthine		No binding of hypoxanthine	0.06	$11.5 \pm 0.5$
Helix P2&P3 RNA		Hypoxanthine		No binding of hypoxanthine	0.20	$9.6 \pm 1.3$
No RNA		Adenine			0.16	$7.8 \pm 0.1$
ASR <sup>apt</sup>		Adenine	5	Specific adenine binding	0.04	$27.8 \pm 0.4$
GSR <sup>apt</sup>		Adenine		No binding of adenine	0.03	$12.1 \pm 0.1$
Helix P2&P3 RNA		Adenine		No binding of adenine	0.07	$10.3 \pm 0.6$

The RNA constructs include the guanine riboswitch aptamer domain (GSR<sup>apt</sup>), the adenine riboswitch aptamer domain (ASR<sup>apt</sup>), and an RNA construct (helix P2&P3 RNA), which consists of the structural elements helices P2 and P3 and loops L2 and L3 but lacks the ligand binding site of the purine riboswitches (see [SI Fig. 5](#) for details) ( $\Delta\sigma$ , deviation of individual DSS reference signal from mean DSS line width value; line width ligand, mean line width value of two signals H2 and H8 of the respective ligand).

structures representing the free form of the RNA, residues identified by NMR to be involved in tertiary interactions were restrained (see [SI Fig. 9](#) and [SI Table 2](#)) and regarded as being similar to the interactions determined from the x-ray structure of the ligand-bound complex (22). For the second ensemble, additional restraints were included that represent the native distances for residues that showed faster folding kinetics (nucleotides U22, U47, U49, U51, and all residues that are in close proximity to hypoxanthine in the folded form). In generating the third ensemble, all residues were restrained to adopt their native conformation consistent with the NMR data and the results of the kinetic experiments. An ensemble exhibiting the native structure of the RNA–ligand complex was thus obtained.

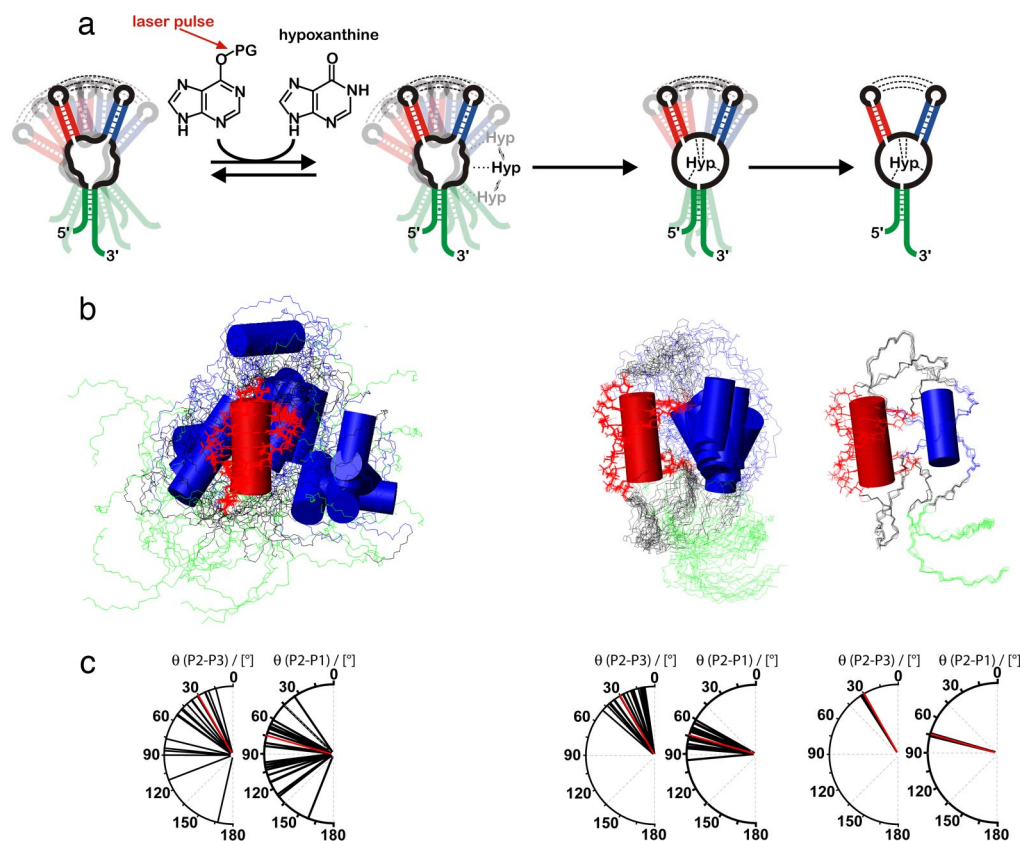
The experimentally restrained molecular dynamics revealed that, during ligand-induced folding, the relative orientation of the helices becomes defined in a two-stage process. The loop–loop interaction in the free form of the RNA is not sufficient to stabilize the entire fold. Therefore, the free form of GSR<sup>apt</sup> needs to be described as a broad structural ensemble with undetermined relative orientation of the helices resulting in a wide conformational distribution. Although the base-pairing interaction in the loop–loop region is preformed in the ligand-free state, these interactions are not sufficient to completely lock the position of the helices in a defined orientation, because nucleotides in L2 and L3 that do not contribute to direct contacts between the loops allow for sampling of a wide conformational space. This observation is consistent with increased line widths of imino proton signals in the ligand-free form of the RNA compared with the ligand-bound form (36). The MD simulations suggest that the mean interhelical angle between helix P2 and helix P3 is close to  $30.1^\circ$ , derived from the x-ray structure of the RNA–ligand complex, but fluctuates

between  $15^\circ$  and  $166^\circ$  (Fig. 4c). Binding of the ligand to the core region of the RNA tightens the complex and decreases fluctuation of the interhelical angle. However, fluctuations of the relative orientation of the structural elements were observed (helix–helix angles between  $10^\circ$  and  $50^\circ$ ). The final compact conformation, similar to the one displayed by the x-ray structure, is formed in the slowest folding step when all residues, including those exhibiting a slow folding rate, are restrained to their native conformation.

## Discussion

Here, we show that time-resolved NMR spectroscopy can be used to characterize the folding of a sizable RNA molecule. Use of time-resolved NMR data in combination with MD simulations can provide an atomic model of RNA conformational transitions as demonstrated here for the ligand-induced RNA folding of the aptamer domain of a guanine-sensing riboswitch.

In our approach, the reaction was initiated by exposure of a photolabile caged ligand to light. We used a similar approach to induce refolding in RNA model systems of smaller size in refs. 30 and 37. Furthermore, other cofactors such as metal ions could also be used in a caged form to trigger structural transitions in biomolecular systems (38). Light-induced initiation of structural transition events has the advantage of providing experimentally reproducible and precisely controlled conditions, such that multiple experiments can be coadded as done in refs. 28 and 29 and here. However, the optical density places an upper limit on the concentration of sample that can be used. Currently, we use 0.25 mM caged hypoxanthine with a laser irradiation of 1.5 s, gaining an optimized deprotection yield of 80%. The experimental conditions described here result in the fast release of ligand in adequate concentration, providing a suitable signal-



**Fig. 4.** Structural interpretation of the conformational transition. (a) Schematic illustration of the proposed folding model of GSR<sup>apt</sup>-RNA on ligand binding based on the experimentally restrained torsion angle MD simulations. The first step is low-affinity binding, the second step is the ligand-binding process, and the third step is helical tightening. (b) Overlaid structures of the three states simulated according to our NMR data, aligned on helix P2 (red). Helices P3 and P1 are blue and green, respectively. (Left) The free form of GSR<sup>apt</sup>, where only the loop-loop interaction and the canonical form of the three helices are restrained. (Center) The transition state-like form where the core is folded toward the native conformation. (Right) The native structure. (c) Distribution of helix-helix projection angles [°] between helices P2/P3 and P2/P1 as seen in the crystal structure (22) (red) and in the structural ensembles depicted in Fig. 4b (black).

to-noise ratio to observe RNA-ligand complex formation in a time-resolved manner. All NMR data obtained postrelease indicate that the final ligand-bound state is consistent with structural data of the RNA-ligand complex (22).

To resolve signals associated with individual nucleotides, application of an RNA isotope-labeling scheme in combination with use of an appropriate NMR methodology was essential in allowing us to resolve 30 of 35 imino proton signals of this sizable RNA-ligand complex. Because of limitations in the signal-to-noise ratio and because some signals are unaffected by ligand binding, we were able to analyze the kinetic rates for 11 of these signals. The observed time constants, in general, are in agreement with the time constants derived from fluorescence techniques (25, 26). The site- and time-resolved NMR methodology presented here can delineate the global folding behavior of the RNA molecule as a sequential model, detailing the different aspects of structure formation involved (Fig. 4). Generating such a model using other NMR techniques, such as those employing relaxation dispersion measurements (39), is unfeasible given the slow time scale of the conformational transition.

We also investigated the initial encounter complex of the ligand and the riboswitch RNA by using the aptamer domain of the adenine-sensing riboswitch (ASR<sup>apt</sup>) as a model, to which hypoxanthine binds with low affinity. The chemical shift values for hypoxanthine in the presence of ASR<sup>apt</sup> are identical to the ones found for the free ligand. In addition, the imino proton spectrum of ASR<sup>apt</sup> shows no changes compared with the one of the ligand-free conformation of the RNA. Hence, the NMR data support the reported lack of specificity of the adenine-sensing

riboswitch for the ligand hypoxanthine. However, the NMR resonances of hypoxanthine in this complex are broadened consistent with a model in which hypoxanthine weakly interacts with the RNA target. Additional experiments with an RNA construct that lacks the part of the ligand binding region known to be crucial for specific recognition of the ligand (denoted as helix P2&P3 RNA, secondary structure depicted in SI Fig. 5) confirm the assumption of a weak and transient interaction of the ligand with the nucleotides present in the core region of the ligand-free state of the purine riboswitch RNAs. However, ligand binding and the subsequent structural rearrangement is productive exclusively in the case of the specific ligand.

With respect to RNA folding, we describe here a general method to translate kinetic information into a structural description of the folding process. The appearance of an imino proton signal in the NMR spectrum shows that (i) the specific nucleotide acquires exchange protection, and (ii) it is locked into its native hydrogen-bonding interaction. This hydrogen-bonding interaction is often long-range in nature, and has been exploited here to use the kinetic information as a structural restraint in a restrained MD simulation.

After an initial low-affinity complex encounter, a two-step specific binding of the ligand to GSR<sup>apt</sup> could be resolved. In a faster process, the ligand binding pocket is formed that results in local stabilization of the three-way junction that anchors the two helical stems. Based on the MD simulations, this local stabilization in turn facilitates subsequent long-range stabilization of the loop-loop interactions. The emerging picture is consistent with a kinetic folding mechanism in which formation of the

ligand binding core is presumably enthalpically driven, enabling the final folding step, which relies on tightening of the helix-helix orientation. Our NMR data provide a hypothesis that can be tested with additional experimental data such as the data concerning the folding characteristics of mutant riboswitch RNAs.

## Materials and Methods

**RNA Preparation.** RNA constructs of GSR<sup>apt</sup> of the *B. subtilis* *xpt-pbuX* operon and the ASR<sup>apt</sup> of *B. subtilis* *pbuE*-mRNA (for details, see SI Fig. 5) were synthesized by *in vitro* transcription and purified as described by Noeske *et al.* (36). The unlabeled helix P2&P3 RNA construct was purchased from Dharmacon (Boulder, CO). Unlabeled rNTPs and <sup>15</sup>N-labeled rNTPs were purchased from Sigma-Aldrich (Munich, Germany) and Silantes (Munich, Germany), respectively. To use RNA for NMR studies, the RNA was exchanged into NMR buffer (25 mM potassium phosphate, pH ≈ 6.2/50 mM potassium chloride).

**Time-Resolved NMR Experiments.** The NMR experiments were performed on an AV800MHz spectrometer (Bruker, Rheinstetten, Germany) with a 5-mm z-axis gradient TXI-HCN cryogenic probe at 283 K. The NMR data were analyzed by using the software TOPSPIN 1.3, felix2000 (Accelrys, San Diego, CA) and SigmaPlot 9.0. All NMR spectra were recorded in H<sub>2</sub>O/D<sub>2</sub>O (9:1) by using standard pulse sequences with WATERGATE water suppression (40) or jump-return-echo pulse sequences (41). Light-induced reaction initiation was achieved by using a laser installation (28, 30) with direct coupling of a quartz fiber from a CW argon ion laser (Spectra-Physics, Darmstadt, Germany) into the NMR tube equipped with a quartz tip insert. Kinetic experiments were arranged as follows. The first 128 data points (eight scans, each point with an interval of 2.1 s) were recorded in the dark. The release of hypoxanthine was then induced by a laser irradiation of 1.5 s (350 nm, 4 W) followed by another 128 data points. After processing by Fourier transformation in F3, including exponential multiplication with a line-broadening factor of 20 Hz, phase correction, and baseline correction in F3, three to five spectra, which were recorded under identical conditions with 256 1D spectra, respectively,

were summed to improve the signal-to-noise ratio of the kinetic studies (for further details, see SI Table 3 and SI Text).

**Determination of Line Widths.** NMR experiments were performed on a Bruker DRX600MHz spectrometer with a 5-mm z-axis gradient TXI-HCN probe. <sup>1</sup>H, <sup>13</sup>C-HSQC spectra with <sup>13</sup>C- and <sup>15</sup>N-decoupling during acquisition were recorded in H<sub>2</sub>O/D<sub>2</sub>O (9:1) at 283 K with 2,2-dimethyl-2-silapentane-5-sulfonate as an internal standard. In all experiments, a ligand-to-RNA ratio of 1:5 was used. The line widths of signals H2 and H8 of the ligand were extracted from the appropriately zero-filled 2D spectra and analyzed by deconvolution using TOPSPIN 1.3.

**MD Simulation.** To simulate the free state of GSR<sup>apt</sup>, we considered that helices P1, P2, and P3 are present and adopt a regular A-form helix based on the NMR data. Pseudodistance restraints for C–C distances up to 8 Å were used as input. Additionally, the same kind of restraints were used for the two base quadruples of the loop-loop interaction obtained from the crystal structure. The backbone torsions were loosely constrained (±60°) to values found for the deposited structural data. An ensemble of 100 structures was calculated with a simulated annealing protocol and the 20 lowest energy structures were analyzed. The protocol consisted of the following simulation steps: TAD-MD (torsion angle dynamics-molecular dynamics) with (i) a high-temperature phase starting at  $T_{\max} = 20,000$ , with 4,000 steps and time steps of 0.01 ps; (ii) a first slow cool annealing stage starting at  $T_{\max} = 20,000$  with 20,000 steps of 0.01 ps; (iii) a second slow cool annealing stage starting from  $T_{\max} = 2,000$  with 3,000 steps of 0.005 ps; and (iv) a subsequent minimization of eight cycles each with 1,000 minimization steps. A completely randomized RNA chain was used as the starting structure. Calculations were performed with CNX2005 from Accelrys (San Diego, CA). (Restraints are summarized in SI Table 2.)

We thank Elke Stirnal and Dr. Christian Richter for excellent technical assistance. This work was supported by the Deutsche Forschungsgemeinschaft (SFB 579: “RNA–ligand-Interaction”), a Studienstiftung des Deutschen Volkes predoctoral fellowship (to B.F.), the Fonds der Chemischen Industrie predoctoral fellowship (to J.N.), and general support to H.S., and the state of Hesse [Center for Biomolecular Magnetic Resonance (BMRZ)].

- Radford SE, Dobson CM, Evans PA (1992) *Nature* 358:302–307.
- Vendruscolo M, Paci E, Dobson CM, Karplus M (2001) *Nature* 409:641–645.
- Salvatella X, Dobson CM, Fersht AR, Vendruscolo M (2005) *Proc Natl Acad Sci USA* 102:12389–12394.
- Wyatt JR, Puglisi JD, Tinoco I, Jr (1989) *BioEssays* 11:100–106.
- Schroeder R, Barta A, Semrad K (2004) *Nat Rev Mol Cell Biol* 5:908–919.
- Williamson JR (2000) *Nat Struct Biol* 7:834–837.
- Draper DE, Grilley D, Soto AM (2005) *Annu Rev Biophys Biomol Struct* 34:221–243.
- Miranda-Rios J, Navarro M, Soberon M (2001) *Proc Natl Acad Sci USA* 98:9736–9741.
- Winkler W, Nahvi A, Breaker RR (2002) *Nature* 419:952–956.
- Epshtein V, Mironov AS, Nudler E (2003) *Proc Natl Acad Sci USA* 100:5052–5056.
- Helm M (2006) *Nucleic Acids Res* 34:721–733.
- Lodmell JS, Dahlberg AE (1997) *Science* 277:1262–1267.
- Sclavi B, Woodson S, Sullivan M, Chance MR, Brenowitz M (1997) *J Mol Biol* 266:144–159.
- Treiber DK, Rook MS, Zarrinkar PP, Williamson JR (1998) *Science* 279:1943–1946.
- Gueron M, Kochoyan M, Leroy JL (1987) *Nature* 328:89–92.
- Hud NV, Schultze P, Sklenar V, Feigon J (1999) *J Mol Biol* 285:233–243.
- Pardi A, Tinoco I, Jr. (1982) *Biochemistry* 21:4686–4693.
- Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR (2003) *Cell* 113:577–586.
- Mandal M, Breaker RR (2004) *Nat Rev Mol Cell Biol* 5:451–463.
- Winkler WC, Breaker RR (2005) *Annu Rev Microbiol* 59:487–517.
- Nudler E, Mironov AS (2004) *Trends Biochem Sci* 29:11–17.
- Batey RT, Gilbert SD, Montange RK (2004) *Nature* 432:411–415.
- Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ (2004) *Chem Biol* 11:1729–1741.
- Schwalbe H, Buck J, Fürtig B, Noeske J, Wöhnert J (2007) *Angew Chem Int Ed Engl* 46:1212–1219.
- Gilbert SD, Stoddard CD, Wise SJ, Batey RT (2006) *J Mol Biol* 359:754–768.
- Wickiser JK, Cheah MT, Breaker RR, Crothers DM (2005) *Biochemistry* 44:13404–13414.
- Lemay JF, Penedo JC, Tremblay R, Lilley DM, Lafontaine DA (2006) *Chem Biol* 13:857–868.
- Kühn T, Schwalbe H (2000) *J Am Chem Soc* 122:6169–6174.
- Wirmser J, Kühn T, Schwalbe H (2001) *Angew Chem* 113:4378–4381.
- Wenter P, Fürtig B, Hainard A, Schwalbe H, Pitsch S (2005) *Angew Chem Int Ed Engl* 44:2600–2603.
- Noeske J, Buck J, Fürtig B, Nasiri HR, Schwalbe H, Wöhnert J (2007) *Nucleic Acids Res* 35:572–583.
- Fürtig B, Richter C, Wöhnert J, Schwalbe H (2003) *ChemBioChem* 4:936–962.
- Otting G, Wüthrich K (1989) *J Magn Reson* 85:586–594.
- Ernst RR, Bodenhausen G, Wokaun A (1994) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions* (Oxford Univ Press, New York).
- Mandal M, Breaker RR (2004) *Nat Struct Mol Biol* 11:29–35.
- Noeske J, Richter C, Grundl MA, Nasiri HR, Schwalbe H, Wöhnert J (2005) *Proc Natl Acad Sci USA* 102:1372–1377.
- Wenter P, Fürtig B, Hainard A, Schwalbe H, Pitsch S (2006) *ChemBioChem* 7:417–420.
- Fürtig B, Buck J, Manoharan V, Bermel W, Jäschke A, Wenter P, Pitsch S, Schwalbe H (2007) *Biopolymers* 86:360–383.
- Korzhev DM, Salvatella X, Vendruscolo M, Di Nardo AA, Davidson AR, Dobson CM, Kay LE (2004) *Nature* 430:586–590.
- Liu M, Mao X, Ye C, Huang H, Nicholson JK, Lindon JC (1998) *J Magn Reson* 132:125–129.
- Sklenar V, Bax A (1987) *J Magn Reson* 75:378–383.

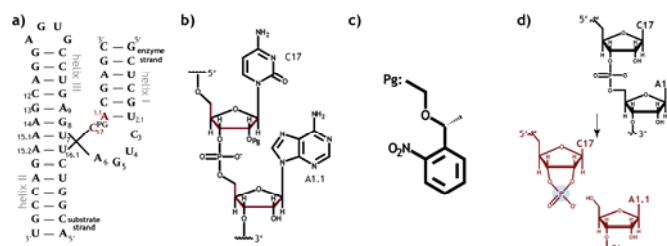


# NMR Characterization of Phosphodiester Bond Cleavage catalysed by the minimal Hammerhead Ribozyme

Boris Fürtig, Christian Richter, Peter Schell, Philipp Wenter, Stefan Pitsch and Harald Schwalbe\*

The minimal hammerhead ribozyme (mHHR) represents the smallest natural occurring ribozyme<sup>[1]</sup>. Its biological role is to cleave the satellite RNA of tobacco ringspot virus during a rolling circle replication mechanism<sup>[2]</sup>. Sequence specific hydrolysis of the 5'-3'-phosphodiester bond leads to the formation of a 2',3'-cyclic phosphate on the 5'-side of the RNA and a free 5'-OH group on the 3'-side of the cleaved strand (Figure 1d). This catalytic activity is observed both in *cis* and in *trans*. The secondary structure of the mHHR consists of a three way junction that connects three helices by three conserved single-stranded regions (Figure 1a). 14 residues comprise the conserved catalytic core of the hammerhead ribozyme<sup>[3]</sup>, including the 5'-NU<sup>16,1</sup>H<sup>17</sup>-3' triplet at the 5'-side of the cleavage side (where N represents any nucleotide, while H<sup>17</sup> is restricted to any nucleotide except G).

Over many years, the minimal hammerhead ribozyme (mHHR) served as a model to study conformational rearrangements and dynamics during RNA catalysis<sup>[4]</sup>. The kinetics of the hammerhead cleavage reaction has been characterized by a variety of biochemical and biophysical techniques<sup>[5]</sup>. The catalysis is dependent on divalent metal ions such as Mg<sup>2+</sup> that can, however, be substituted by high concentrations of monovalent ions. Depending on the exact sequence context, maximum catalytic efficiency is observed



**Figure 1.** a) Secondary structure of the investigated minimal hammerhead construct consisting of an 35nt unlabelled enzyme strand and a 14nt selectively labelled substrate strand. b) Structural representation of the catalytic core (5'-C17-A1.1-3'). The red colour indicates the introduced <sup>13</sup>C labels. c) Constitution of the npeom protecting group (Pg) that inhibits the nucleophilic attack of the C17 2'-OH onto the phosphodiester bond. d) Schematic representation of the cleavage reaction.

at pH values between 6 and 8. The mHHR accelerates RNA cleavage above the non-catalysed background RNA hydrolysis by a factor of 10<sup>6</sup><sup>[6]</sup>. From transient electric birefringence<sup>[7]</sup>, FRET<sup>[8]</sup> and nuclear magnetic resonance (NMR) spectroscopic<sup>[9]</sup> studies it is known that the three helices undergo a substantial angular movement during catalysis. In the Mg<sup>2+</sup>-free state, the ribozyme adopts a different global conformation than in the Mg<sup>2+</sup>-bound form. The angles between the helices in the Mg<sup>2+</sup>-free state adopt values of  $\theta(I/II)=38^\circ$ ,  $\theta(II/III)=169^\circ$ ,  $\theta(III/I)=152^\circ$  and give rise to an overall Y-shape conformation. In the Mg<sup>2+</sup>-bound form (using a deoxyA nucleotide at position 17 to prevent the reaction), they adopt values of  $\theta(I/II)=153^\circ$ ,  $\theta(II/III)=100^\circ$ ,  $\theta(III/I)=77^\circ$  and describe an elongated helical arrangement of the helices I and II with the third helix III oriented almost perpendicular<sup>[10]</sup>.

More recently, it was found that in the biological context the sequence does not only consist of the minimal core nucleotides but that the biological active hammerhead ribozyme is elongated (eHHR) and exhibits a loop/bulge interaction<sup>[11]</sup>. The elongated ribozyme has an enhanced cleavage rate of approx.  $5 \cdot 10^2 - 10^3$  times higher than mHHRs<sup>[12]</sup>. In 2006, the crystal structure of the elongated hammerhead ribozyme was solved at 2.2Å resolution. This crystal structure is different to structures solved for the minimal sequences. The new structure reveals the catalytic core to be in-line for a cleavage reaction<sup>[13]</sup>, leading to the proposal that the minimal hammerhead crystallises in an inactive form that can undergo conformational changes to become active. Such conformational transition is also feasible within the crystal lattice where bond cleavage has been observed<sup>[14]</sup>. As revealed by earlier NMR studies<sup>[15]</sup> and now interpreted in the light of both X-ray structures, the mHHR exerts dynamics that likely reflect the structural transitions

[\*] B. Fürtig, Dr. C. Richter, Dr. P. Schell, Prof. Dr. H. Schwalbe  
Institute for Organic Chemistry and Chemical Biology  
Center for Biomolecular Magnetic Resonance  
Johann Wolfgang Goethe-University  
Max von Laue-Str. 7, 60438 Frankfurt/Main (Germany)  
Fax: (+49)69-798-29515  
E-mail: schwalbe@nmr.uni-frankfurt.de  
Homepage: <http://schwalbe.org.chemie.uni-frankfurt.de>

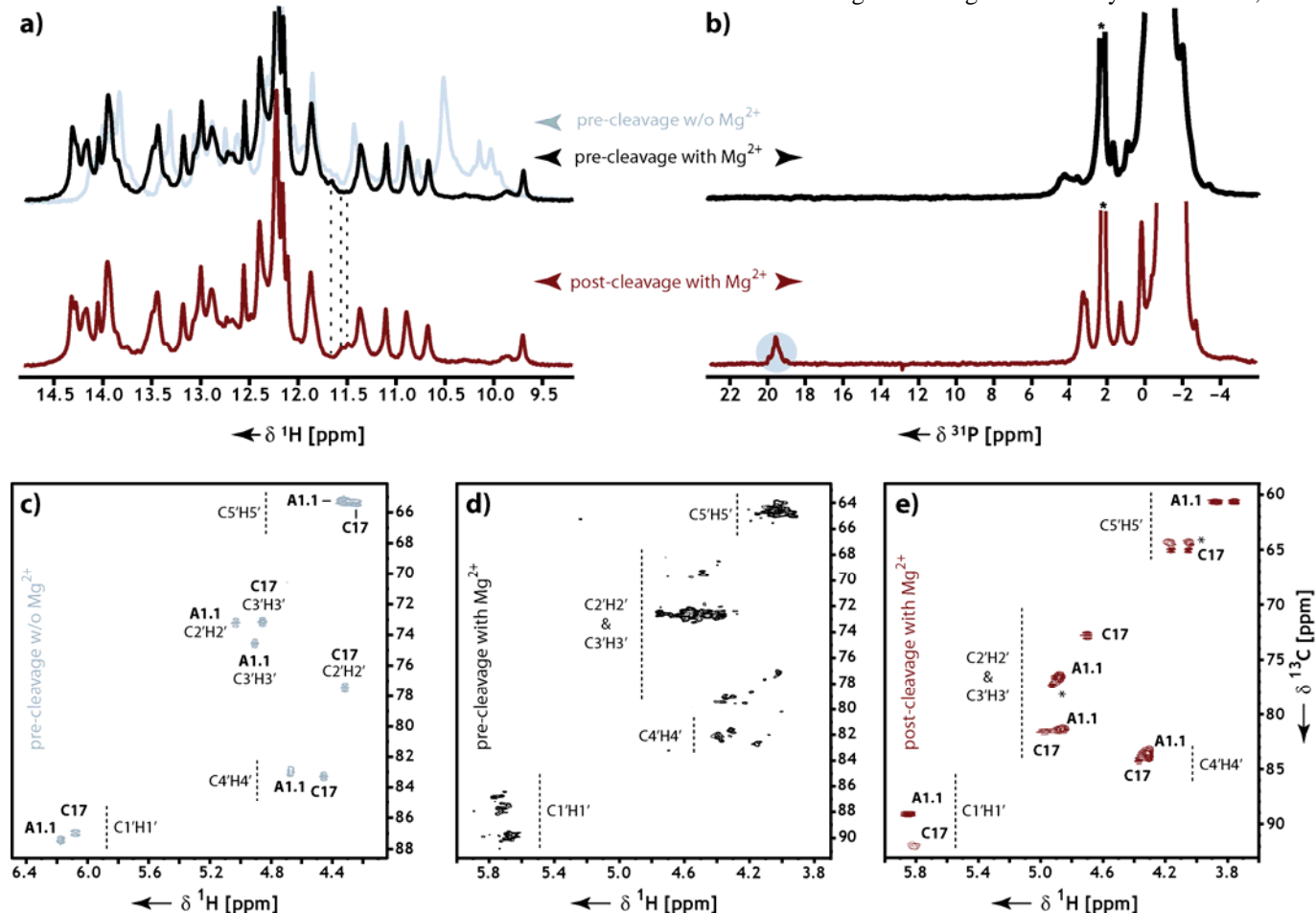
Dr. P. Wenter, Prof. Dr. S. Pitsch  
École Polytechnique Fédérale Lausanne  
Laboratory of Nucleic Acid Chemistry, EPFL-BCH  
1015 Lausanne (Switzerland)

[\*\*] We are grateful to Julia Wirmer, Kai Schlepckow, and Janina Buck for discussions. The work was funded by the DFG (SFB 579 "RNA-Ligand-Wechselwirkungen"), the "Studienstiftung des Deutschen Volkes" (B.F.), startup funds from the Massachusetts Institute of Technology, and the Fonds der Chemischen Industrie (H.S.). The solid-phase synthesis of the hammerhead ribozyme was carried out in Lausanne.

Supporting information for this article is available on the WWW under <http://www.angewandte.org> or from the author.



between the inactive and the active form.



**Figure 2.** a) Imino region of the 1D  $^1\text{H}$  spectra recorded at 293K at 800MHz, in grey photo-protected pre-cleavage state without  $\text{Mg}^{2+}$ , in black photo-protected pre-cleavage state in 10mM  $\text{Mg}^{2+}$ , in red post-cleavage state in 10mM  $\text{Mg}^{2+}$ , dashed lines indicate the minute changes in the spectrum occurring upon the cleavage reaction. b) 1D  $^{31}\text{P}$  spectra recorded at 293K at 600MHz, asterisks indicate the buffer signal, grey circle indicates the resonance of the cyclic phosphate characteristic for the cleaved ribozyme. The  $^{31}\text{P}$  signals of the pre-cleavage state are broadened. The lower panels c-e display the HSQC spectra of the selectively  $^{13}\text{C}$  ribose labelled mHHR recorded in three different states: c) Pre-cleavage photo-protected  $\text{Mg}^{2+}$  free state; d) Pre-cleavage photo-protected state containing 10mM  $\text{Mg}^{2+}$ ; e) Post-cleavage state in 10mM  $\text{Mg}^{2+}$  after photo-initiated catalysis. All  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC spectra were recorded at 800MHz at 293K.

In this communication, we investigated such structural transitions of the mHHR during catalysis with static and time-resolved NMR experiments. Often, RNA structural transitions are observed by characterising spectral changes in the imino-proton region of the 1D  $^1\text{H}$  spectrum<sup>[16, 17]</sup>. Imino-proton signals are reporters of base-pairing interactions between different nucleotides and can be incorporated into structure calculations<sup>[18]</sup>. In the case of the hammerhead ribozyme, however, few changes are observed in the imino-proton spectra during all monitored steps occurring on reaction pathway of catalysis (Figure 2a). Especially, no line broadening is detected upon  $\text{Mg}^{2+}$  addition. The cleavage reaction (Figure 1d), nevertheless, can be monitored by the characteristic chemical shift ( $^{31}\text{P}$  signal at 19.7ppm) of the reaction product, namely the cyclic 2',3'-phosphate<sup>[19]</sup> (Figure 2b). To further characterise the

structural changes during the catalytic reaction, we

synthesised a construct of the hammerhead ribozyme in which the ribose moieties of the two residues surrounding the scissile phosphodiester group are selectively  $^{13}\text{C}$ -labelled (C17 and A1.1 (Figure 1b)). We used an adenosine in position 1.1 of the hammerhead for synthetic reasons since the lower chemical yields for guanosine in many synthetic steps prohibited preparation of a  $^{13}\text{C}$ -labelled 5'-C<sup>17</sup>G<sup>1.1</sup>-3' hammerhead construct<sup>[20]</sup>. The investigated construct consisted of an unlabelled 35nt enzyme strand and a 14nt substrate strand (Figure 1a). In addition, the 2'-OH of residue C17, responsible for the intramolecular nucleophilic attack, is protected with a photolabile protecting group (npeom (Figure 1c)). Compared to a uniformly labelled or unlabelled 49nt long RNA molecule, this labelling strategy results in resolved and assignable  $^{13}\text{C}$  resonances of the two nucleotides in the core of the catalytic reaction and the photolabile protecting group prevents cleavage of the ribozyme. The npeom group with the spacer methylene group has been modelled into the ground-state prior to the experiments. Such modelling studies suggested that steric reasons should not significantly change the ground state conformation (shown in Figure 3 in the suppl. mat.). Furthermore, biochemical investigations of a photo-caged mHHR using the O-(2-nitrobenzyl)-group (npe) as photolabile group reveal identical cleavage rates and the same extend of cleavage as in the identical unmodified sequence<sup>[21]</sup>. Our synthetic strategy allows examination of the conformation of the starting ground-state and the possibility to initiate the cleavage reaction by laser irradiation in situ.



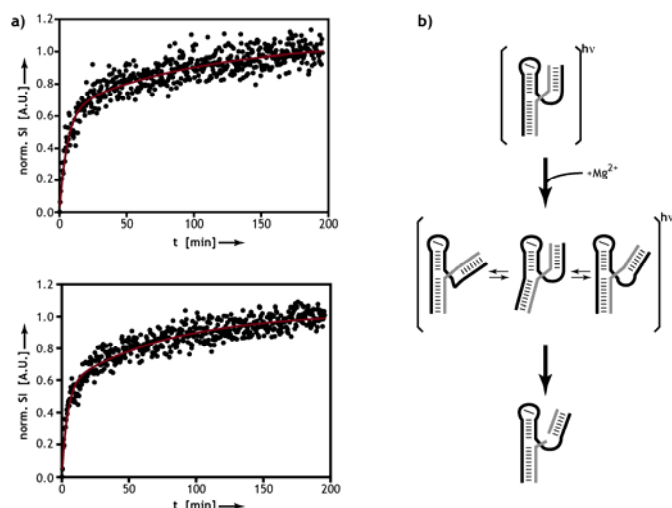
The synthesis of the labelled and photo-protected building blocks is described in the supplementary material.

In the  $\text{Mg}^{2+}$ -free state, the two nucleotides surrounding the scissile phosphodiester bond show sharp  $^{13}\text{C}$  signals. These signals can be readily assigned. Scalar coupling constants  $^3J(\text{H}1',\text{H}2')$  measured in an HCCH-E.COSY experiment<sup>[22]</sup> are 5Hz and 11Hz for the residues A1.1 and C17, respectively. The canonical coordinates calculated from the  $^{13}\text{C}$  chemical shifts<sup>[23, 24]</sup> are  $\text{can1}=-8.58$ ,  $\text{can2}=-16.94$  and  $\text{can1}=-8.77$ ,  $\text{can2}=-17.14$  for residues A1.1 and C17, respectively. Both residues adopt a predominantly South, near C2'-endo conformation in the ribose moiety and the exocyclic backbone angle  $\gamma$  of C17 is in a gt conformation ( $\gamma=60^\circ$ ). Based on the scalar coupling constants, A1.1 shows some conformational averaging. Interestingly, the three-dimensional structures as determined by X-ray crystallography for the minimal hammerhead ribozyme (see Figure 4 in the supp. material) all assign the sugar-moiety conformation to be in North, near C3'-endo conformation. In contrast, the structures of the elongated full-length ribozyme reveal a South, C2'-endo conformation for the nucleotide 5' adjacent to the scissile phosphate and a North, C3'-endo conformation for the nucleotide at the 3'-side of the cleavage side.  $^{31}\text{P}$  chemical shifts are sensitive reporters for non-canonical backbone conformations<sup>[25]</sup>. In an HCP-experiment<sup>[26]</sup>, a chemical shift of  $\delta^{31}\text{P}=-1.41\pm0.3\text{ppm}$  is observed for the scissile phosphodiester in the  $\text{Mg}^{2+}$ -free pre-cleavage state, which is consistent with a canonical backbone conformation. Compared to the structure determined for the full-length eHHR, this characteristic chemical shift indicates that the  $\text{Mg}^{2+}$ -free pre-cleavage state of mHHR is not completely preorganised for an in-line attack of the 2'-OH group towards the phosphodiester bond. The conformation in the  $\text{Mg}^{2+}$ -free state therefore represents a pre-cleavage state that has to be activated by the divalent ions to adopt the cleavable conformation.

Addition of  $\text{Mg}^{2+}$  in milli-molar concentrations, required for cleavage activity, results in an overall stabilisation of the mHHR as monitored by an increase in the melting temperature of  $\sim 18^\circ\text{C}$  ( $T_m=48.0^\circ\text{C}\pm0.1^\circ\text{C}$  in the absence of  $\text{Mg}^{2+}$ ;  $T_m=65.8^\circ\text{C}\pm0.2^\circ\text{C}$  in 10mM  $\text{Mg}^{2+}$ ; for melting profiles see supp. material figure 1) measured by CD-spectroscopy. Moreover, an increase in the measured molar ellipticity at the three maxima of 268nm, 240nm and 210nm indicates (for CD spectra also see supp. material figure 1) that the tertiary structure of the molecules is globally stabilised upon addition of  $\text{Mg}^{2+}$  (at 293K). As also evident from the 1D spectrum of the imino resonance observable only in helical regions of RNA (Fig. 2a), substantial helix rearrangements and new helical segments are formed in the presence of  $\text{Mg}^{2+}$  ions. The conformation of the catalytic core, however, is differently affected. The addition of  $\text{Mg}^{2+}$  (in 2mM steps from 0mM to 10mM) to mHHR leads to severe line broadening of all NMR signals within the two  $^{13}\text{C}$ -labelled ribose moieties (see Figure 2d). In addition, new weak signals appear. These spectral changes are indicative for an equilibrium between more than two states in slow

conformational exchange and, in addition, conformational exchange within those states in the intermediate time regime. The NMR observations are consistent with previous results and indicate that the activated complex (mHHR in 10mM  $\text{Mg}^{2+}$ ) has to be described by a conformational heterogeneous ensemble that may likely be necessary to facilitate the cleavage reaction. By EPR spectroscopy, two  $\text{Mg}^{2+}$ -induced transitions at 10mM and at 25mM  $\text{Mg}^{2+}$  concentration are monitored and the dynamics in the RNA molecule were also determined to be similar for 1mM and 10mM  $\text{Mg}^{2+}$ <sup>[27]</sup>. Additionally, results of earlier T-jump experiments<sup>[28]</sup> lead to the assumption that under comparable conditions the ribozyme molecule exists in an ensemble of at least three interconverting states that are characterised by relaxation time constants between several ms and 200ms. NMR relaxation experiments<sup>[9]</sup> are also indicating exchange lifetimes of sub-states of 10-50 ms. Based on RDC-data these dynamic transitions that reflect internal motions involving significant conformational rearrangements are solely present in the  $\text{Mg}^{2+}$ -bound state, but not in the  $\text{Mg}^{2+}$ -free state<sup>[10]</sup>. Our findings here show that chemical exchange in the millisecond time regime (intermediate exchange regime) can be detected consistent with the reported earlier findings and slow conformational exchange leading to multiple sets of signals (see Figure 2d). The latter slow conformational exchange can be detected due to the selective labelling strategy applied here. Starting from this dynamic ensemble of slowly interconverting conformers in the presence of  $\text{Mg}^{2+}$ , the hammerhead catalyzed cleavage reaction was initiated in situ by irradiation of the sample with a continuous wave laser light pulse of 1s duration (4.5W at 334nm to 380nm) in the NMR spectrometer. The laser light is routed directly into the NMR tube via spectrometer-controlled fibre optics<sup>[29, 30]</sup> (see supporting material Figure 5 for laser setup). The reaction was followed in real-time by recording a series of 1D  $^1\text{H}$ - and of  $^{13}\text{C}$ -filtered 1D  $^1\text{H}$ -spectra.

After cleavage, the NMR spectra of the mHHR reveal sharp signals arising from the two  $^{13}\text{C}$ -labelled nucleotides similar to the pre-cleavage state. The kinetic traces of the reaction can be extracted by integration of the signals arising from the product state during the real-time experiment. As expected, the stability of the RNA complex in the cleaved state is lowered as evident by the melting point of  $T_m=31.3^\circ\text{C}\pm1.3^\circ\text{C}$  compared to that of the uncleaved complex. This can also be followed by a temperature series of NMR 1D  $^1\text{H}$  measurements where a decrease of signals in the imino region and considerable shifts in the aromatic region are observable around this transition temperature (supporting material Figure 2). This melting also leads to additional peaks for C17 in the cleaved state (see peaks marked with asterisks in figure 2e). All analysed peaks reveal similar kinetics including identical build-up curves (Figure 3a). The kinetic traces can all be fitted to bi-exponential behaviour (as revealed by an F-test analysis, for examples see Table 1 in the Supp. Material). Two phases with equal amplitudes are observed that exhibit a first fast rate of  $k_1=0.2185\text{min}^{-1}\pm0.0194\text{min}^{-1}$  and a subsequent slow rate of  $k_2=0.0110\text{min}^{-1}\pm0.0011\text{min}^{-1}$ .



**Figure 3.** a) Kinetic traces of two signals extracted from real-time NMR experiment. Pseudo 3D-spectra are consisting of subsequently recorded  $^{13}\text{C}$ -filtered 1D  $^1\text{H}$  spectra. The traces of the reaction are best fitted by bi-exponential build-up behaviour (red line). b) Schematic representation of the reaction sequence monitored.

In the kinetic experiments, the two detectable phases are consistent with the following model: The dynamic pre-cleavage state in the presence of  $\text{Mg}^{2+}$  consists of an equilibrium between different states, which subsequently have to undergo different conformational transitions to facilitate catalysis (Figure 3b). These distinct rearrangements from cleavage-incompetent to cleavage-competent conformations are characterised by the two rate revealed in the real-time experiment. As evident by our NMR data, these conformational transitions do not include changes in the base-pairing pattern, since otherwise transient changes in the imino-region of the spectra would have to be observed which is not the case. Thus, the dynamic reorientation movement of the helices have to induce changes in the catalytic core. Such changes could facilitate different binding modes of the metal ions required for catalysis which may be only transiently bound in certain backbone conformations. Alternatively, the conformational changes could modulate the  $\text{pK}_\text{A}$  value of the hydroxyl group in a conformational dependent manner and thereby distinguish between the catalytic activity of the subsets of molecules<sup>[14]</sup>. Data of Persson et al.<sup>[31]</sup> support the idea that increasing flexibility by introduction of a UUUU tetraloop in comparison to the more stable GNRA tetraloop may allow increased sampling of those conformations of the ribozyme that are catalytically more active<sup>[32]</sup>. For all signals of the catalytic core, the same kinetic behaviour can be monitored consistent with a model in which the nucleotides adjacent to the scissile phosphate group are exerting the dynamic transition needed to facilitate catalysis, in a concerted manner.

In summary, the mHHR in the  $\text{Mg}^{2+}$ -free form adopts a pre-cleavage form in which the sugar conformation is prone for the catalytic reaction but where the scissile phosphate still remains an A-form like conformation. Addition of  $\text{Mg}^{2+}$  then induces dynamic transitions not only regarding the orientation of the helices but also in the active side, resulting in an interconverting ensemble of molecules that exhibit either a conformation of the scissile phosphate capable for the

chemical step of catalysis or still reside in the ground state conformation that is populated in the absence of divalent metal ions. The dynamics of mHHR are more pronounced for the non-canonical region of the mHHR ribozyme and involve the ribose moieties of the catalytic core. The biphasic kinetic behaviour of the mHHR is consistent with the requirement of two distinct kinetic events: the faster rate represent the kinetic trace of those molecules that adopt conformations competent for catalysis in the dynamic equilibrium and the slower rate represents those that first have to undergo a conformational transition at the active side phosphate to be in-line for the nucleophilic attack (Figure 3b).

Detection of RNA dynamics by NMR spectroscopy is often restricted to the analysis of base-pairing changes revealed from imino-proton resonances. Since no changes in the base-pairing pattern occur during the reaction catalysed by the hammerhead ribozyme, the use of selective  $^{13}\text{C}$ -labelling schemes reporting on changes in the ribose moieties of the involved nucleotides was found to be mandatory. An intrinsic trigger of the reactivity was necessary in order to facilitate analysis of the molecular ensemble in the different states under identical conditions. We could design such a system using organic synthesis to introduce a photolabile protecting group caging the reactive 2'-OH group. Both synthetic strategies were prerequisite to the application of real time NMR experiments that for the first time combine high spatial resolution with good temporal resolution in order to examine the reactivity of any ribozyme. The NMR experiments reveal a concerted motion of both nucleotides of the catalytic centre during the catalysed cleavage reaction.

## Experimental Section

**Static NMR spectroscopy:**  $^1\text{H}$ ,  $^{13}\text{C}$ -HSQC; HCC-TOCSY and NOESY spectra were recorded at mostly 293K on a Bruker AV 800MHz spectrometer equipped with a cryogenic TXI HCN z-grad probe head. The pulse sequences provided in the Bruker library were applied. Processing and analysis were performed using XWin-NMR 3.5 and TopSpin 2.0 (both Bruker) software suites. HCP and 1D  $^{31}\text{P}$  spectra were recorded on a Bruker AVIII 600MHz spectrometer equipped with a cryogenic TCI HCP z-grad probe.

**Real-time NMR spectroscopy:** Kinetic NMR experiments were also conducted on a Bruker AV 800MHz spectrometer equipped with a cryogenic TXI HCN probe head. For triggering of the reaction the laser setup (see supporting information) was synchronised with the spectrometer via TTL connection. The kinetic traces were recoded in a pseudo-3D dataset (described in Fürtig et al. 2007<sup>[33]</sup>). Per kinetic trace point 16 transients were averaged. Processing and analysis of these experimental data was performed using the software FELIX (Accelrys). Non-linear fitting of the kinetic traces was achieved using the program Sigmaplot (Systat Software Inc.).

Received: ((will be filled in by the editorial staff))

Published online on ((will be filled in by the editorial staff))

**Keywords:** hammerhead ribozyme • real-time NMR • RNA dynamics • NMR spectroscopy • selective NMR isotope labelling

[1] A. C. Forster, R. H. Symons, *Cell* **1987**, 50, 9-16.

- [2] G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider, G. Bruening, *Science* **1986**, *231*, 1577-1580.
- [3] G. Ferbeyre, J. M. Smith, R. Cedergren, *Mol. Cell. Biol.* **1998**, *18*, 3880-3888.
- [4] R. Przybilski, C. Hammann, *Chembiochem* **2006**, *7*, 1641-1644.
- [5] K. F. Blount, O. C. Uhlenbeck, *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 415-440.
- [6] M. J. Fedor, O. C. Uhlenbeck, *Biochemistry* **1992**, *31*, 12042-12054.
- [7] K. M. Amiri, P. J. Hagerman, *Biochemistry* **1994**, *33*, 13172-13177.
- [8] T. Tuschl, C. Gohlke, T. M. Jovin, E. Westhof, F. Eckstein, *Science* **1994**, *266*, 785-789.
- [9] M. P. Latham, D. J. Brown, S. A. McCallum, A. Pardi, *Chembiochem* **2005**, *6*, 1492-1505.
- [10] K. Bondensgaard, E. T. Mollova, A. Pardi, *Biochemistry* **2002**, *41*, 11532-11542.
- [11] A. Khvorova, A. Lescoute, E. Westhof, S. D. Jayasena, *Nat. Struct. Biol.* **2003**, *10*, 708-712.
- [12] M. De la Pena, S. Gago, R. Flores, *EMBO J.* **2003**, *22*, 5561-5570.
- [13] M. Martick, W. G. Scott, *Cell* **2006**, *126*, 309-320.
- [14] J. B. Murray, C. M. Dunham, W. G. Scott, *J. Mol. Biol.* **2002**, *315*, 121-130.
- [15] C. Hammann, D. G. Norman, D. M. Lilley, *Proc. Natl Acad. Sci. USA* **2001**, *98*, 5503-5508.
- [16] P. Wenter, B. Fürtig, A. Hainard, H. Schwalbe, S. Pitsch, *Angew. Chem. Int. Ed.* **2005**, *44*, 2600-2603.
- [17] P. Wenter, B. Fürtig, A. Hainard, H. Schwalbe, S. Pitsch, *Chembiochem* **2006**, *7*, 417-420.
- [18] J. Buck, B. Fürtig, J. Noeske, J. Wöhnert, H. Schwalbe, *submitted* **2007**.
- [19] D. G. Gorenstein, *Annu Rev Biophys Bioeng* **1981**, *10*, 355-386.
- [20] S. S. P. Quant, R. W. Wechselberger, M. A. Wolter, K. H. Wörner, P. Schell, J. W. Engels, C. Griesinger, H. J. Schwalbe, *Tetrahedron Lett.* **1994**, *35*, 6649-6652.
- [21] S. G. Chaulk, A. M. MacMillan, *Nucl. Acids Res.* **1998**, *26*, 3173-3178.
- [22] H. Schwalbe, J. P. Marino, G. C. King, R. Wechselberger, W. Bermel, C. Griesinger, *J. Biomol. NMR* **1994**, *4*, 631-644.
- [23] M. Ebrahimi, P. Rossi, C. Rogers, G. S. Harbison, *J. Magn. Reson.* **2001**, *150*, 1-9.
- [24] B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *Chembiochem* **2003**, *4*, 936-962.
- [25] S. S. Wijmenga, B. N. M. van Buuren, *Prog. Nucl. Magn. Reson. Spect.* **1998**, *32*, 287-387.
- [26] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, *J. Am. Chem. Soc.* **1994**, *116*, 6472-6473.
- [27] T. E. Edwards, S. T. Sigurdsson, *Biochemistry* **2005**, *44*, 12870-12878.
- [28] M. Menger, F. Eckstein, D. Porschke, *Nucl. Acids Res.* **2000**, *28*, 4428-4434.
- [29] T. Kühn, Schwalbe, H., *J. Am. Chem. Soc.* **2000**, *122*, 6169-6174.
- [30] J. Wirmer, J. Kuhn, H. Schwalbe, *Ang. Chem. Int. Ed.* **2001**, *40*, 4248-4251.
- [31] T. Persson, R. K. Hartmann, F. Eckstein, *Chembiochem* **2002**, *3*, 1066-1071.
- [32] J. Koplin, Y. Mu, C. Richter, H. Schwalbe, G. Stock, *Structure* **2005**, *13*, 1255-1267.
- [33] B. Fürtig, J. Buck, V. Manoharan, W. Bermel, A. Jäschke, P. Wenter, S. Pitsch, H. Schwalbe, *Biopolymers* **2007**, *in press*.

## Entry for the Table of Contents

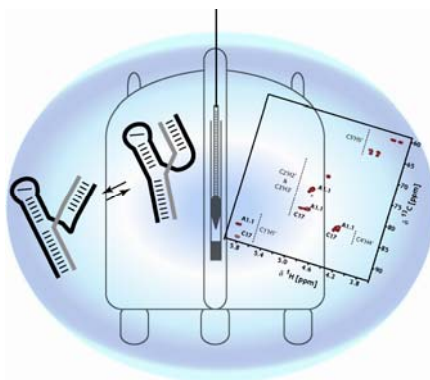
Layout 1:

### RNA dynamics

Boris Fürtig, Christian Richter, Peter Schell, Philipp Wenter, Stefan Pitsch and Harald Schwalbe\*

Page – Page

NMR Characterization of  
Phosphodiester Bond Cleavage  
catalysed by the minimal Hammerhead  
Ribozyme



By the combination of selective labelling schemes and the application of real-time NMR spectroscopy, the RNA cleavage reaction catalysed by the hammerhead ribozyme could be analysed. Substantial changes in the dynamics are observed during the reaction. These dynamics are shown to be important for understanding the kinetics of the reaction when followed by time-resolved NMR with atomic resolution.

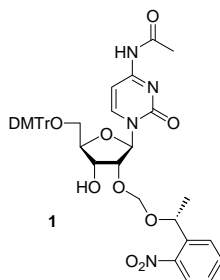
## Supplementary material

### Monitoring phosphodiester bond cleavage catalysed by the minimal hammerhead ribozyme by NMR

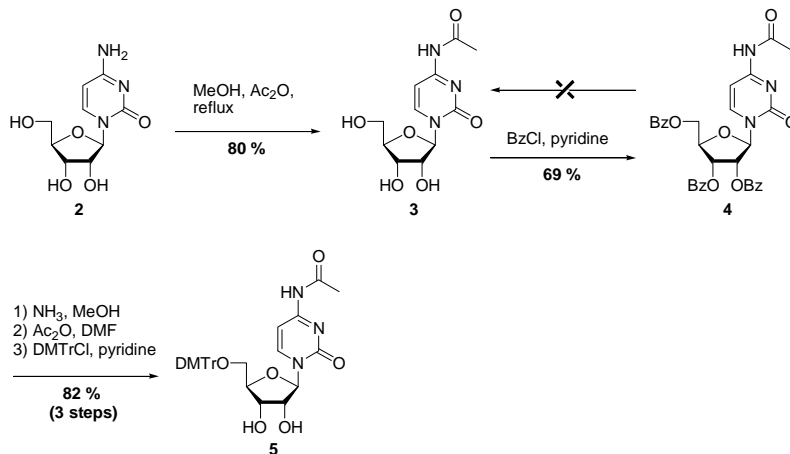
*Boris Fürtig<sup>1</sup>, Christian Richter<sup>1</sup>, Peter Schell<sup>1</sup>, Philipp Wenter<sup>2</sup>, Stefan Pitsch<sup>2</sup> and Harald Schwalbe<sup>1\*</sup>*

#### A) SYNTHESIS OF BUILDING BLOCKS FOR RNA SOLID PHASE SYNTHESIS BASED ON PHOSPHOAMIDITE CHEMISTRY

##### 1 Synthesis of the C-building block 1

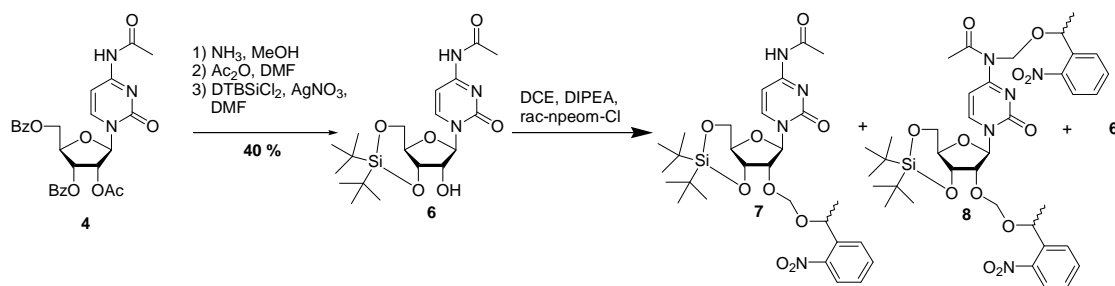


Our syntheses started from the nucleosides prepared as described in the literature<sup>1</sup>. As photoprotected nucleotide, C-nucleotide **1** was prepared. Using a literature procedure<sup>2</sup> cytidine **2** was reacted with acetic anhydride in refluxing methanol to give the *N*-acetyl derivative **3** in 80 % yield. **3** was further benzoylated to provide **4** as a model compound to study the selective cleavage of the ester groups without affecting the *N*-acetyl functionality. These studies proved to be unsuccessful. **4** was then used to explore a different approach to synthesize **1**. Fully protected **4** was completely deprotected using NH<sub>3</sub> in methanol<sup>3</sup> and then selectively acetylated by a procedure communicated by Pitsch. Crude **3** was further dimethoxytritylated to give **5** in 82 % over all three steps without purification of the intermediates. This procedure proved to be suitable for the application with labelled material (Scheme S1).



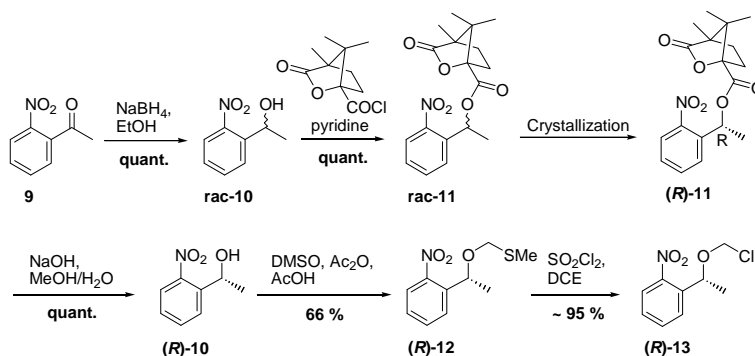
**Scheme S1** Studies with unlabeled material towards **1**

Since npeom is not cleaved by fluoride ions, the use of a silyl group to simultaneously protect the 3' and 5' position was explored. Stimulated by a recent publication from Agrofolio<sup>4</sup>, the di-*tert*-butylsilyl-group was studied. Initial studies gave unsatisfactory yields for the sequence **4** → **6**, and further reaction of **6** with racemic npeom-Cl gave a mixture of **7**, **8** and unreacted **6**. Since alkylation of **5** with npeom-Cl mediated by the tin-acetal<sup>4</sup> strongly favors the 2'-*O*-isomer, these studies were abandoned (Scheme S2). Despite these results the silyl pathway has a good potential for the application for other nucleosides and further studies (e.g. using the Markiewicz-group) are worthwhile.



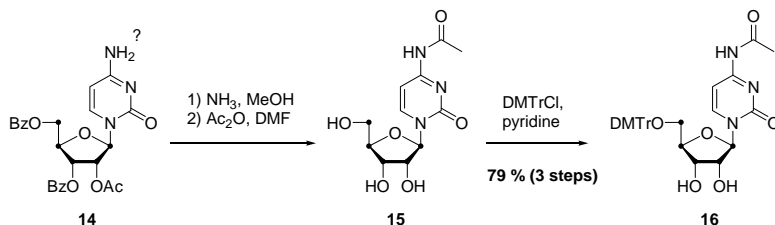
**Scheme S2** Studies using the di-*tert*-butylsilyl-group

For NMR-studies, the npeom-group should be used enantiomerically pure to avoid formation of diastereomers in the building blocks and the RNA and thereby complicate the NMR spectra. The synthesis was undertaken using literature procedures<sup>6-8</sup> and is summarized in Scheme S3. (*R*)-**13** was generated from (*R*)-**12** directly before use and was used without further purification in the alkylation reactions.



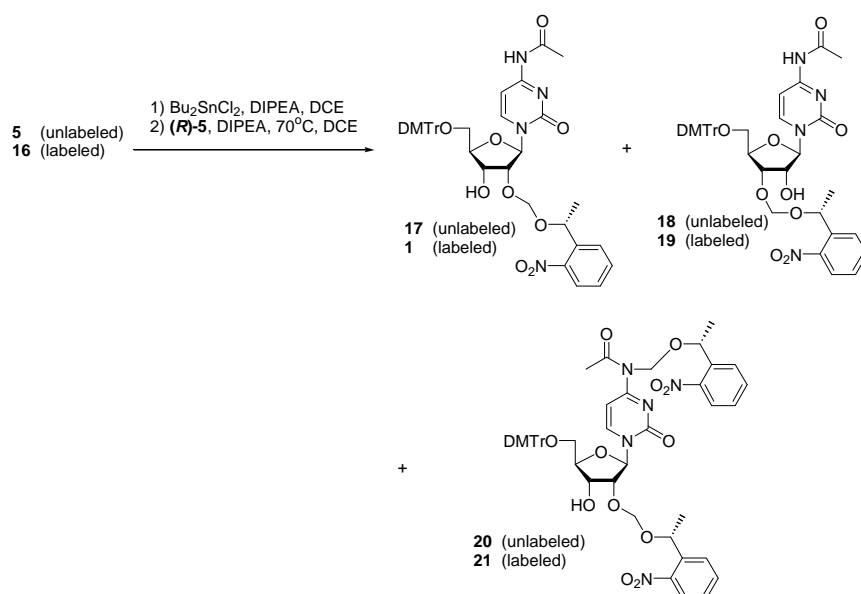
**Scheme S3** Synthesis of enantiomerically pure npeom-Cl (*R*)-**13**

The abovementioned (Scheme S1) three step sequence could be carried out in identical manner also for the labelled material and **16** was obtained in 79 % overall yield (Scheme S4).



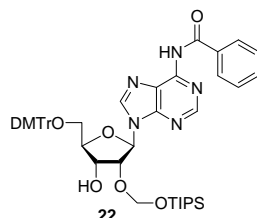
**Scheme S4** Synthesis of labelled cytidine derivative **16**

Studies with unlabelled material showed that a slight excess of npeom-Cl (*R*)-**13** gave the best results for the synthesis of **1**. Nonetheless, the undesired alkylation of the base to form **21** could not be avoided (Scheme S3). Starting from **16**, the yields were as follows: **1** (44 %), **19** (~16 %, contains **16**) and **21** (18 %).

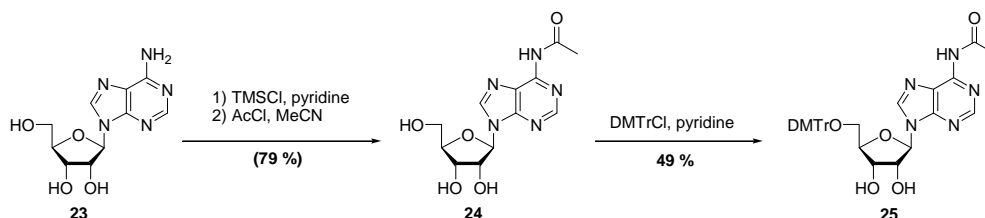


**Scheme S5** Introduction of npeom to give labeled **1**

## 2 Synthesis of the A-building block **22**

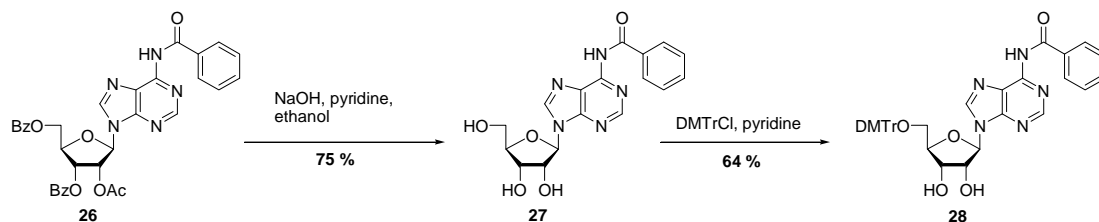


The labelled starting material had a benzoate group in the base. Since acetyl would be the preferred group, methods for their introduction were examined. Following a procedure by Pitsch, transient protection of adenosine **23** gave **24**, but in an impure form in about 79 % yield. Subsequent dimethoxytritylation of crude **24** was straight forward, albeit with a relative low yield of 49 % of **25** (Scheme S6).



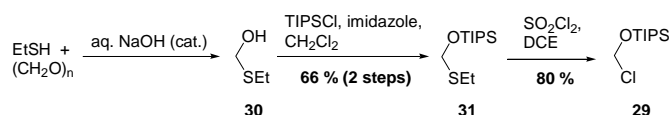
**Scheme S6** Studies for the synthesis of **25**

Since literature procedures<sup>10</sup> showed excellent yields for the selective deprotection of peracylated *N*-benzoyl-protected adenosines, this pathway was pursued. After a model study with 10 mg of unlabelled material showed that deprotection and dimethoxytritylation was successful, the synthesis was undertaken with the labelled material. Surprisingly, selective deprotection of **26** to give **27** had a yield of only 75 % and subsequent dimethoxytritylation to give **28** yielded only 64 % (Scheme 7). The starting material was about 10 % impure, and throughout the synthesis incomplete dissolution of reactants could be observed. Nonetheless, enough **28** could be obtained for the introduction of the Tom-group.



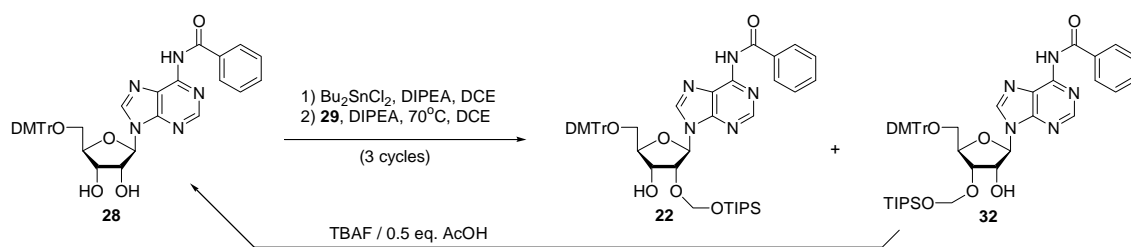
**Scheme S7** Synthesis of labelled **28**

Tom-Cl **29** was synthesized according to a procedure by Pitsch and generated directly before use (Scheme S8).



**Scheme S8** Synthesis of Tom-Cl **29**

The reaction to introduce the Tom protecting group was repeated three times, following a procedure communicated by Pitsch. Between the reactions, undesired **32** was cleaved with TBAF/0.5 eq. AcOH and resulting **28** was resubmitted into the reaction. The overall yield of desired **22** was 29 % (Scheme S9).



**Scheme S9** Synthesis of labelled **22**

The prepared nucleotide building blocks were used in RNA solid-phase synthesis. Phosphoramidites were prepared in the laboratory of S. Pitsch using standard procedures.

#### Experimental section:

**N<sup>d</sup>-Acetyl-cytidine 3.** Cytidine (2.6 g, 10.7 mmol) was suspended in methanol (200 mL) and heated under reflux. Acetic anhydride (10 mL, 106 mmol) was added in 5 portions, with at least 45 min between every addition. After addition of the final portion heating was continued for 1 h then the mixture was cooled to room temperature. A colorless precipitate formed. The volume was reduced to ~100 mL and the colorless solid was collected. Yield: 2.48 g (8.7 mmol, 80 %).

**N<sup>d</sup>-Acetyl-2',3',5'-tri-*O*-benzoyl-cytidine 4.** **3** (1.0 g, 3.5 mmol) was suspended in pyridine (40 mL). Then benzoyl chloride (2 mL, 17.3 mmol) was added dropwise. After 5 min the mixture is clear. After stirring overnight the yellow mixture was poured into water and extracted with dichloromethane. The organic layer was subsequently extracted with 1N HCl, water, sat. NaHCO<sub>3</sub>, dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. After column chromatography (DCM/MeOH 98/2) the product containing fractions were combined and crystallized from DCM/hexanes. Yield: 1.43 g (2.4 mmol, 69 %) colorless solid **4**. The mother liquor was discarded, since there was enough material for the studies.

**N<sup>d</sup>-Acetyl-5'-*O*-DMTr-cytidine 5 (and labeled 16).** **4** (400 mg, 0.67 mmol) was suspended in 7N NH<sub>3</sub> in methanol (10 mL) at room temperature and stirred overnight. The mixture was evaporated and dried at an oil pump. Then, DMF (4 mL) and acetic anhydride (70 µL, 0.74 mmol) were added and the mixture was stirred overnight. If the reaction is incomplete, methanol is added and the mixture is evaporated and resubmitted using DMF (4 mL) and acetic anhydride (18 µL, 0.17 mmol) under the same conditions. When the reaction is complete, methanol is added and the mixture is evaporated and dried at an oil pump overnight. The residue is suspended in pyridine (5 mL) and DMTr-Cl (340 mg, 1.5 eq.) is added. After stirring overnight, the mixture is poured into dichloromethane and extracted with sat. NaHCO<sub>3</sub>. The organic layer is dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. Coevaporation with toluene removes residual pyridine. Column chromatography (DCM/MeOH gradient 0-4 %) gives a purple foam of **5**. Yield: 322 mg (0.55 mmol, 82 %).

In the same manner **14** (300 mg, ~0.5 mmol) was reacted to give **16** (234 mg, 0.4 mmol, 79 %) as a purple foam.

**(RS)-1-(2-Nitrophenyl)ethanol rac-10.** 2-Nitroacetophenone **9** (16.68 g, 101 mmol) was dissolved in ethanol (400 mL) and cooled in an ice bath. NaBH<sub>4</sub> (4.00 g, 105 mmol) was added in one portion. This mixture was stirred for 4 h in the ice bath. Then acetic acid (10 mL) was added slowly (!) (pH = 5-6) and the mixture was evaporated. The residue was partitioned between water and ether. After phase separation the organic layer was washed with sat. NaHCO<sub>3</sub> and dried over Na<sub>2</sub>SO<sub>4</sub>. Evaporation gave a yellow oil (17 g, 101 mmol, quant.), which was used without further purification.

**(R)- and (S)-1-(2-Nitrophenyl)ethyl (1S)-camphanates rac-11.** Crude (RS)-1-(2-nitrophenyl)ethanol **rac-10** (6 g, 36 mmol) was dissolved in pyridine (50 mL) and (-)-camphanic acid chloride (8 g, 37 mmol) was added in one portion. The mixture warmed and a colorless precipitate formed. After stirring overnight water was added and the mixture was evaporated. The residue was dissolved in EtOAc and washed with water, 1N HCl, water sat. NaHCO<sub>3</sub> and dried over Na<sub>2</sub>SO<sub>4</sub>. Evaporation gave an oil, which crystallized upon standing as an off-white solid (12.5 g, 36 mmol, quant.).

**(R)-1-(2-Nitrophenyl)ethyl (1S)-camphanate (R)-11.** The above compound **rac-11** (12.5 g) was dissolved in boiling methanol (100 mL) and the mixture was allowed to cool. Long needles formed, after some time also prisms crystallized. Reheated again, until everything was dissolved and again cooled. When the prisms started to form the mother liquor was decanted. The needles were crystallized from MeOH (2 x) to give the pure (NMR) (R,S)-ester **(R)-11** (1.32 g, 3.8 mmol). The mother liquors were combined and saved.



**((R)-1-(2-Nitrophenyl)ethoxy)methyl ethyl sulfide (R)-12.** (R)-1-(2-Nitrophenyl)ethyl (1S)-camphanate **(R)-11** (1.32 g, 3.8 mmol) was dissolved in methanol (20 mL) and 2N KOH (2.8 mL) and refluxed for 30 min. The mixture was cooled and neutralized with 1N HCl. After reduction to the aqueous phase, it was partitioned between ether and sat. NaHCO<sub>3</sub>. The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. The residue was dissolved in DMSO (4 mL) and acetic acid (2 mL) and acetic anhydride (2.9 mL) was added. This mixture was stirred at room temperature under exclusion of light for 6 days. This mixture was added slowly to sat. NaHCO<sub>3</sub> and extracted with hexanes (2 x). The organic layer was washed with water (2 x), dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. Flash chromatography on silica gel (toluene - toluene/EtOAc 98/2) gave a yellow oil of **(R)-12** (570 mg, 2.5 mmol, 66 %).

**Npeom-Cl (R)-13.** Directly before use, **(R)-12** (1 eq.) was dissolved in dichloroethane and cooled in an ice bath. SO<sub>2</sub>Cl<sub>2</sub> (1 eq.) was added and the mixture stirred for 30 min in an ice bath and 10 min at room temperature. The solvent was evaporated and the residue dried at an oil pump and directly used for the next step.

**Building Block 1. 16** (230 mg, 0.39 mmol) was dissolved in dichloroethane (5 mL) and Huenig's Base (270 µL, 1.56 mmol). Bu<sub>2</sub>SnCl<sub>2</sub> (150 mg, 0.5 mmol) was added and the mixture was stirred for 4 h at room temperature. Then, the mixture was heated to 70°C and a solution of **(R)-13** (prepared from **(R)-12** (106 mg, 0.47 mmol)) in dichloroethane (3 mL) was added. After 25 min at 70°C sat. NaHCO<sub>3</sub> was added and the resulting slurry was filtered through Celite. The aqueous phase was extracted with dichloromethane (2 x). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. Column chromatography (hexanes/ethyl acetate (2 % TEA) gradient: 3/1, 1/1, 1/3, 100 % Ee. Then the column is washed to get the 3'-isomer **19** and remaining **16** off) gave **21** (68 mg, 0.072 mmol, 18 %), **1** (133 mg, 0.172 mmol, 44 %) and a mixture of **16** and **19** (50 mg, ~16 %) as slightly yellow foams.

**N<sup>6</sup>-Acetyl-adenosine 24.** Adenosine **23** (5 g, 18.71 mmol) was suspended in pyridine (40 mL) and cooled in an ice bath. TMSCl (23.65 mL, 187 mmol) was added dropwise over 3 h and then stirred overnight at room temperature. The mixture was cooled in an ice bath and acetonitrile (60 mL, precooled in an ice bath) was added, followed by acetyl chloride (2 mL, 28.2 mmol) dropwise. This mixture was stirred for 2 h in an ice bath, poured into cold dichloromethane (110 mL) and ice water (75 mL) was added with vigorous stirring. The organic layer was separated, washed with water, dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. The residue was coevaporated with toluene (2 x), dissolved in methanol (40 mL) and acetic acid (11.5 mL) was added. This mixture was stirred for 90 min and stored in a freezer (-20°C) for 2 days. No crystals. Then more toluene and acetic acid were added. Over the course of two weeks a gel formed, which slowly turned into a colorless solid. This solid was collected and washed with ether. Yield: 4.43 g (14.3 mmol, 79 %), but impure (TLC 4 more spots, but NMR shows only acetic acid).

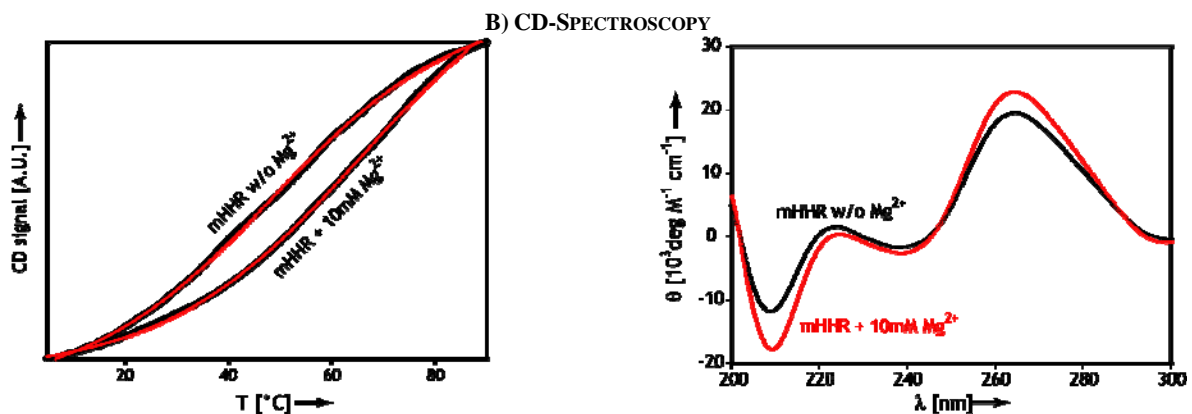
**N<sup>6</sup>-Acetyl-5'-O-DMTr-adenosine 25.** **24** (crude) (1.55 g, 5 mmol) was dissolved in pyridine (15 mL) and DMTrCl (2.54 g, 7.5 mmol) was added. This mixture was stirred overnight and evaporated after the addition of methanol. The residue was dissolved in dichloromethane, extracted with sat. NaHCO<sub>3</sub> and the organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>. Coevaporation with toluene (2 x) removes remaining pyridine. Column chromatography (DCM/MeOH gradient 0-5% (packed with 2 % TEA)) gave **25** (1.5 g, 2.45 mmol, 49 %) as a slightly purple foam.

**N<sup>6</sup>-Benzoyl-5'-O-DMTr-adenosine (labeled) 28.** **26** (439 mg, 0.7 mmol) was dissolved (something remains undissolved) in ethanol (3 mL) and pyridine (3 mL) and cooled in an ice bath. To this mixture 2N NaOH was added and stirred for 15 min. The mixture is neutralized with DOWEX 50 and filtered. The resin is washed thoroughly with water. The aqueous layer is washed with dichloromethane (2 x) and evaporated. Bath temperature below 35°C. The residue (195 mg, max. 0.52 mmol) was dissolved in pyridine (5 mL) and DMTrCl (250 mg, 0.74 mmol) was added. This mixture was stirred overnight and evaporated after the addition of water. The residue was dissolved in dichloromethane, extracted with sat. NaHCO<sub>3</sub> and the organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>. Coevaporation with toluene (2 x) removes remaining pyridine. Column chromatography (DCM/MeOH gradient 0-5% (packed with 2 % TEA)) gave **28** (225 mg, 0.33 mmol, 49 %) as a slightly purple foam.

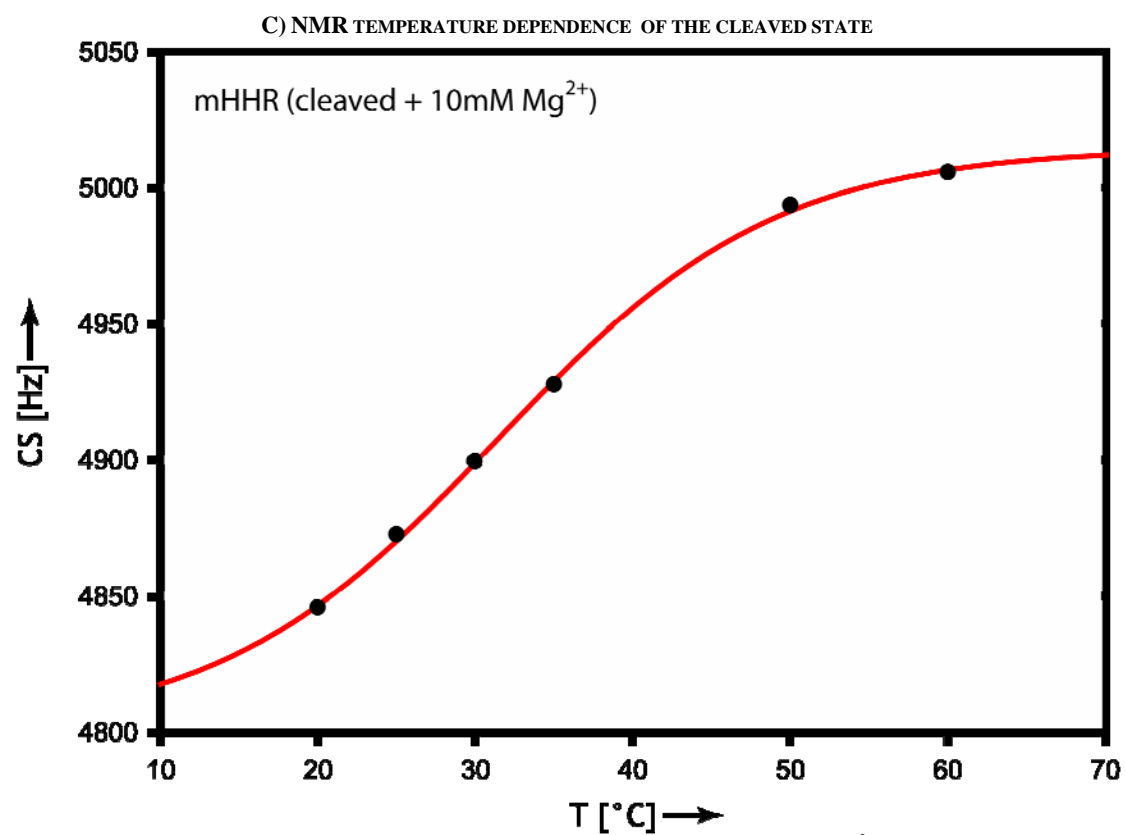
**Tom-Cl 29.** Ethanethiol (7.4 mL, 100 mmol) and formaldehyde (3.0 g, 100 mmol) were mixed and 1 drop of 10N NaOH was added. After the initial reaction ceased the mixture was warmed to 40°C for 1 h. Dichloromethane (100 mL) and imidazole (13.6 g, 200 mmol) were added, followed by TIPSCl (21.4 mL, 100 mmol). The mixture was stirred overnight at room temperature and then poured into hexanes (200 mL). This mixture was stirred with 10% NaH<sub>2</sub>PO<sub>4</sub> (100 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. Distillation in an oil pump vacuum gave **31** (16.4 g, 66 mmol, 66 %) as a colorless liquid.

Directly before use, **31** (1 eq.) was dissolved in dichloroethane and cooled in an ice bath. SO<sub>2</sub>Cl<sub>2</sub> (1 eq.) was added and the mixture stirred for 30 min in an ice bath and 10 min at room temperature. The solvent was evaporated and the residue dried at an oil pump and directly used for the next step.

**Building Block 22. 28** (220 mg, 0.324 mmol) was dissolved in dichloroethane (5 mL) and Huenig's Base (230 µL, 1.4 mmol). Bu<sub>2</sub>SnCl<sub>2</sub> (130 mg, 0.44 mmol) was added and the mixture was stirred for 90 min at room temperature. Then, the mixture was heated to 70°C and a solution of **29** (prepared from **31** (105 mg, 0.423 mmol)) in dichloroethane (3 mL) was added. After 20 min at 70°C sat. NaHCO<sub>3</sub> was added and the resulting slurry was filtered through Celite. The aqueous phase was extracted with dichloromethane (2 x). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. Column chromatography (hexanes/ethyl acetate (2 % TEA) gradient: 9/1, 8/2, 7/3, 6/4, 5/5. Then the column is washed to get the 3'-isomer **32** and remaining **28** off) yields the 2'-O-isomer **22**. The fractions with the wrong isomer and remaining **28** are combined and treated with TBAF/0.5 eq. AcOH in THF. The mixture is poured into ether and extracted with brine (3 x) and sat. NaHCO<sub>3</sub>. The organic layer is dried over Na<sub>2</sub>SO<sub>4</sub> and evaporated. Column chromatography (DCM/MeOH gradient: 0-5 % (packed in TEA)) yields **28**. After 3 cycles **22** (80 mg, 0.092 mmol, 29 %) was obtained as a colorless foam. **32** together with **28** gave also 80 mg. Some loss of **22** was accounted to acidic chloroform. Great care has to be taken with chloroform as a solvent. Always filter through basic alumina before use.



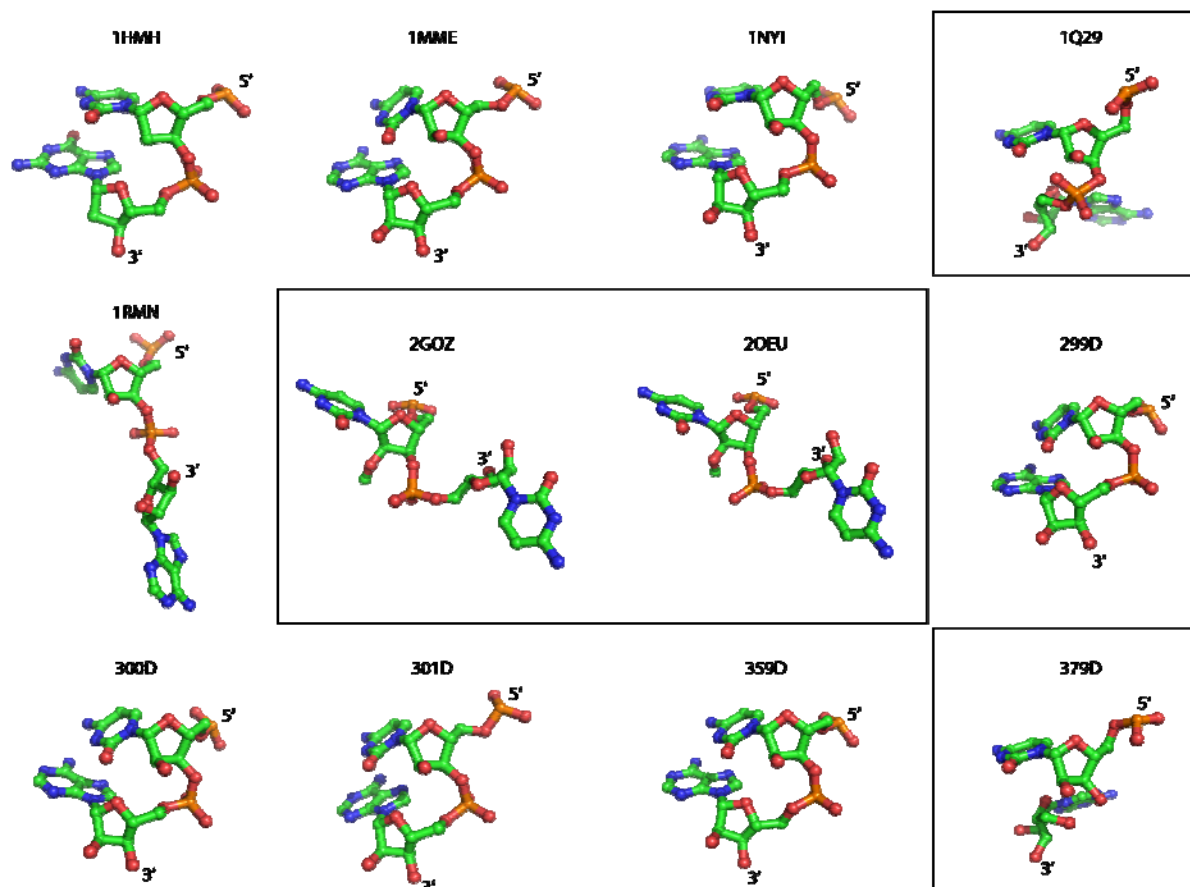
**Figure 1:** Left panel: CD-Temperature row for the mHHR in the uncleaved form with and without Mg<sup>2+</sup>; right panel CD-spectra of the molecule under the same conditions



**Figure 2:** NMR-Temperature row for the mHHR in the cleaved form in the presence of 10mM  $\text{Mg}^{2+}$ ; The figure depicts the chemical shift changes dependent on the sample temperature. Transition is observed at  $T_m=31.3^\circ\text{C}$ .

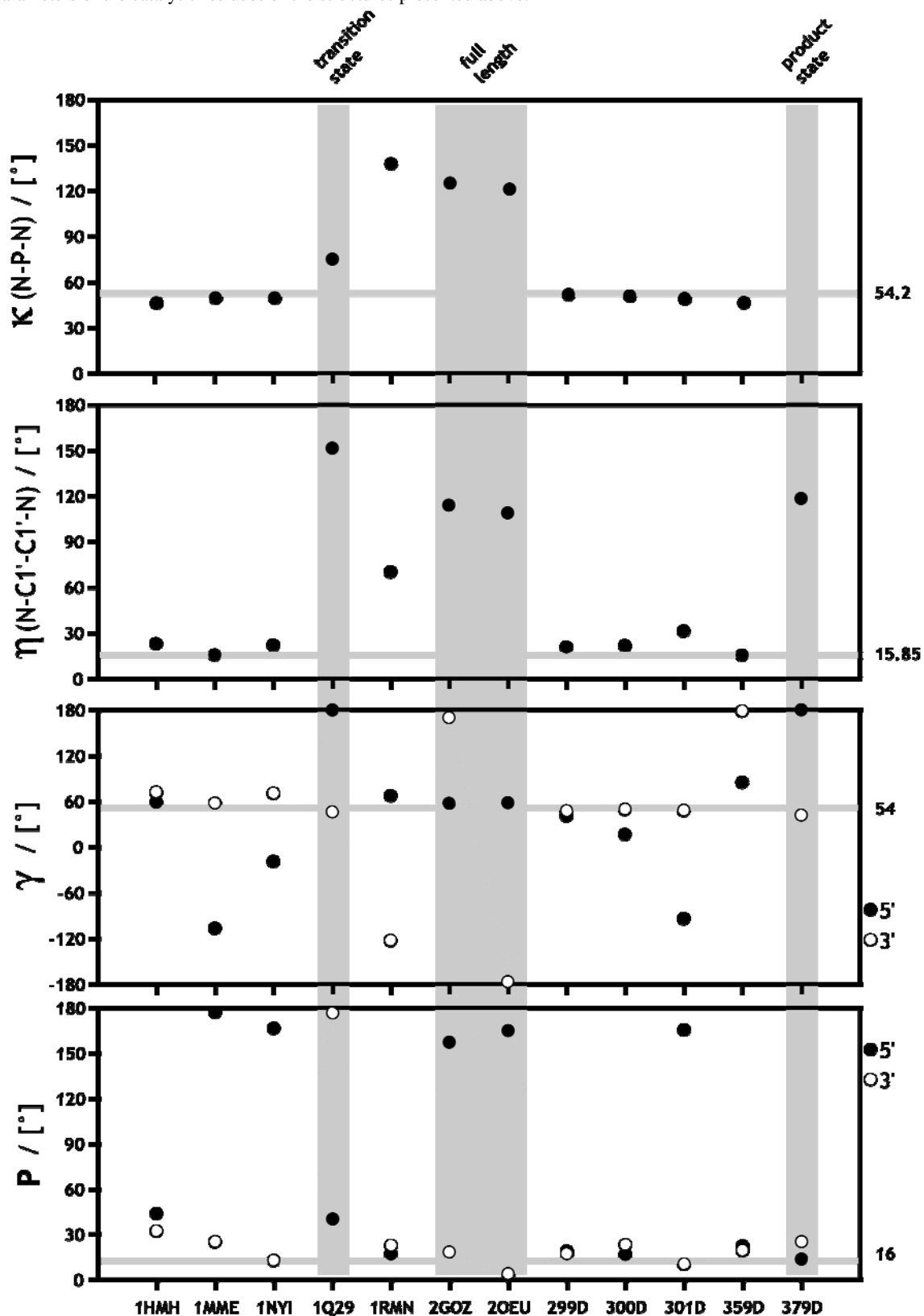
#### D) COMPARATIVE STRUCTURAL DATA

Core nucleotides as found in the different hammerhead X-ray structures deposited in the pdb:



**Figure 3:** Residues of the catalytic core 5'- and 3'- adjacent to the scissile phosphate as found in the different X-ray crystallographic structures (1RMN is based on fluorescence measurements). Except for 2GOZ and 2OEU, that represent structures of the elongated HHR, all structures are reporting on the minimal HHR. The structure 379D is determined on a cleaved ribozyme.

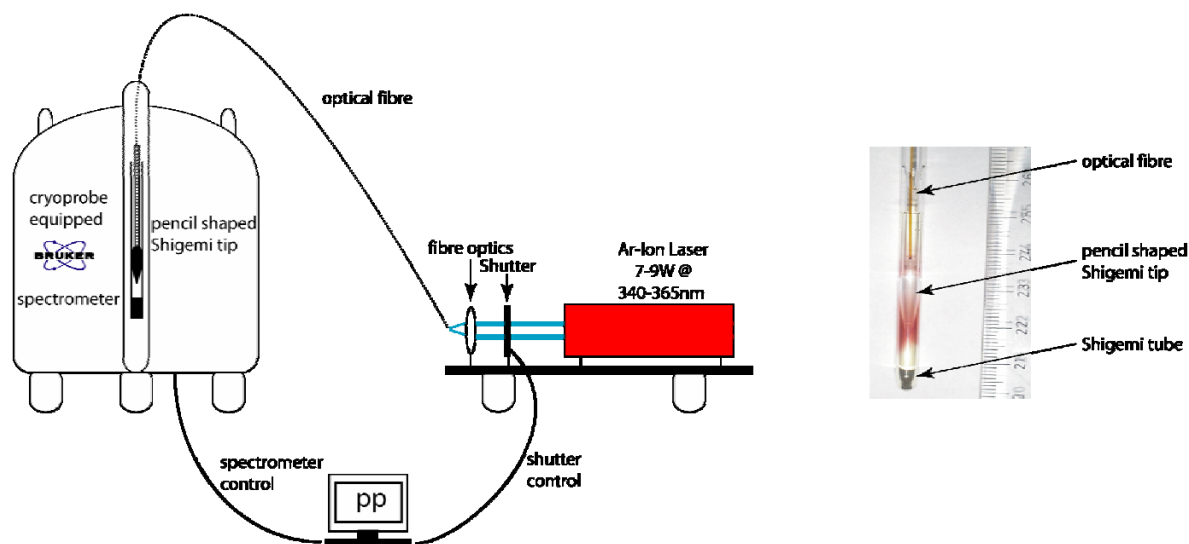
Structural parameters of the catalytic residues of the structures presented above:



**Figure 4:** P: Pseudorotation-Phase;  $\gamma$ : exocyclic backbone angle (filled circles: 5'-nucleotide adjacent to scissile phosphate, open circles: 3'-nucleotide adjacent to scissile phosphate);  $\eta$ : pseudo-dihedral angle of glycosidic bond upon the C1'-C1' vector;  $\kappa$ : angle between glycosidic N's and scissile phosphate group; special featured structures are highlighted vertically in grey; canonical A-Form values are indicated by grey horizontal lines, values annotated.

### E) LASER-NMR SETUP

Photolytic deprotection is achieved in the NMR spectrometer using a setup that was initially developed for protein folding studies (see Figure 44). The slightly modified system consists of a CW-Ar-ion laser {BeamLok 2060 (Spectra Physics) 9.7W power operating at 340-365nm} whose beam is conducted to the sample tube via an optical fibre {LWL-Kabel UV1000/1100N (Ceram Optec)}. A high power resistant lens {lens: U13X (Newport) mounted in a combined lens and fibre holder: F-91-C1-T, FPR1-C1A (Newport)} obtains coupling of the laser beam into the fibre {other optical components used -beamsplitter 200-0 (CVI), -mirror 200-10 (CVI)}. Irradiation of the complete sample is maintained using pencil shaped Shigemi tips. A shutter {shutter: uniblitz electronic VS14S2ZM1 (Vincent Associates); shutter-driver: uniblitz VCM-D1 (Vincent Associates)}, connected to the spectrometer blocks the laser beam in front of the fibre and can trigger the laser irradiation.



**Figure 5:** Left panel: Schematic representation of laser coupling to the NMR-sample: The laser beam is transmitted via an optical fibre into the sample residing in the active site of the NMR spectrometer; right panel: a pencil shaped shigemi tip connected to the optical fibre and sitting in a sample of reduced volume achieves uniform sample illumination.

## F) RESULTS OF RT-NMR EXPERIMENTS

Table 1: Results of the real-time NMR experiments: kinetic for the 5 resolved sugar resonances:

Peak #	a	$\sigma a$	b	$\sigma b$	c	$\sigma c$	d	$\sigma d$	$R^2$	F
1	0,4872	0,0427	0,0959	0,0141	0,7586	0,1157	0,0056	0,0021	0,9876	97,95381813
2	0,6278	0,0229	0,1734	0,0149	0,4321	0,0177	0,0102	0,0018	0,9944	204,4247515
3	0,5641	0,017	0,2392	0,02	0,464	0,0122	0,0127	0,0013	0,9961	344,6112364
4	0,5008	0,0206	0,4172	0,068	0,5156	0,0166	0,0138	0,0015	0,9907	192,1366827
5	0,5052	0,0287	0,3209	0,0647	0,549	0,0261	0,0112	0,002	0,9825	111,2254408
MW	0,5242	0,017	0,2185	0,0194	0,5276	0,0133	0,011	0,0011	0,9958	

Kinetic traces are fitted to the following biexponential equation:  $f=a*(1-\exp(-b*x))+c*(1-\exp(-d*x))$

a & c: amplitudes ;  $\sigma a$  &  $\sigma c$ : fitting error

b & d: rate constants;  $\sigma b$  &  $\sigma d$ : fitting error

F. F-value of the statistical F test

#### G) REFERENCES FOR THE SUPP. MAT.

- 1 S. Quant, R. W. Wechselberger, M. A. Wolter, K. H. Worner, P. Schell, J. W. Engels, C. Griesinger, H. Schwalbe, *Tetrahedron Letters* **1994**, 35, 6649-6652.
- 2 Foeldesi, A.; Maltseva, T.; Dinya, Z.; Chattopadhyaya, J.; *Tetrahedron* **1998**, 54, 14487-14514
- 3 Agrofoglio, L. A.; Jacquinet, J.-C.; Lancelot, G. *Tetrahedron Letters* **1997**, 38, 1411-1412
- 4 Saito Y.; Nyilas, A.; Agrofoglio, L. A. *Carbohydrate Research* **2001**, 331, 83-90
- 5 Pitsch, S.; Weiss, P. A.; Wu, X.; Ackermann, D.; Honegger, T. *Helvetica Chimica Acta* **1999**, 82, 1753-1761
- 6 Corrie, J. E. T.; Reid, G. P.; Trentham, D. R.; Hursthouse, M. B.; Mazid, M. A. *J. Chem. Soc. Perkin I* **1992**, 1015-1019
- 7 Schwartz, M. E.; Breaker, R. R.; Asteriadis, G. T.; deBear, J. S.; Gough, G. R.; *Bioorganic & Medicinal Chemistry Letters* **1992**, 2, 1019-1024
- 8 Benneche, T.; Gundersen, L.-L.; Undheim, K. *Acta Chemica Scandinavica* **1988**, B 42, 384-389
- 9 Tiecco, M.; Testaferri, L.; Bagnoli, L.; Marini, F.; Temperini, A.; Tomassini, C.; Santi, C. *Tetrahedron Letters* **2000**, 41, 3241-3245
- 10 Milecki, J.; Zamaratski, E.; Maltseva, T. V.; Foeldesi, A.; Adamiak, R. W.; Chattopadhyaya, J.; *Tetrahedron* **1999**, 55, 6603-6629



## APPENDIX

**AP 1:** Pulse-sequence and parameters for pseudo 3D kinetic experiment (sequential array of 1D-<sup>1</sup>H- experiments before and after laser trigger, with <sup>15</sup>N decoupling)

```

#i ncl ude <Avance. i ncl >
#i ncl ude <Grad. i ncl >
#i ncl ude <Delay. i ncl >

"l 6=td2"
"l 7=td1"

"d11=30m"

1 ze
2 20u
3 d12 pl 16: f3
4 50u BLKGRAD
  d1
  10u setnmr3^14
  d27 ;duration laser
  10u setnmr3^14
  d28
  goto 9
5 d12
6 10u do: f3
  10u setnmr3^14
  d29 ;laser gobo
  10u setnmr3^14
  d28
7 ze
  d11 pl 16: f3
8 30m
  d1 do: f3

9 10u pl 1: f1
  p1 ph1
  d19
  p1 ph2
  50u UNBLKGRAD
  p16: gp1
  d16
  p1 ph3
  d19*2
  p1 ph4
  50u
  p16: gp1
  d16
  go=8 ph31 cpd3: f3
  30m do: f3 wr #0 if #0 zd
  50u BLKGRAD
  lo to 5 times l 6
  lo to 3 times l 7
  50u BLKGRAD
exit

ph1=0 2
ph2=2 0
ph3=0
ph4=2
ph31=0 2
; l 6 number of fid after first laser pulse
; l 7 number of laser irradiations

```

**AP 2:** Pulse-sequence and parameters for pseudo-3D kinetic experiment (sequential array of 1D- $^1\text{H}$ - $\{^{15}\text{N}/^{14}\text{N}\}$ -experiments before and after laser trigger, with  $^{15}\text{N}$  decoupling and flip-angle variation for Ernst-angle excitation)

```
#i ncl ude <Avance. i ncl >
#i ncl ude <Grad. i ncl >
#i ncl ude <Del ay. i ncl >

"d2=1s/2*cnst11"      ; ~1/(2*JNH)
"d16=100u"
"d22=d2-p16-d16-p1*2-d19*2"
"d23=d2-p16-d16-p21*2-3u"

"p22=p21*2"
"p2=p1*2"
"cnst10=78/90" ; EA calculation
"p0=p1*cnst10"

"l0=1"
"l6=td2"
"l7=td1"

1 ze
2 20u
3 d12 do: f3
4 10u do: f3
  10u setnmr3|14
  d27
  10u setnmr3^14
  d28
  goto 9
5 d12 do: f3
6 10u do: f3
  10u setnmr3^14
  d29
  10u setnmr3^14
7 d28
  d1
  goto 9
8 4u do: f3
  d1
9 4u do: f3
  10u pl1: f1
  10u pl3: f3
  p0 ph1
  d19

p0 ph2
d22 UNBLKGRAD
p16: gp1
d16
p1 ph3
d19*2
p1 ph4

if "l0 %2 ==1"
{
  (p21 ph5): f3
}
else
{
  (p21 ph8): f3
}
3u
(p21 ph6): f3
p16: gp1
d16 pl16: f3
d23 BLKGRAD
go=8 ph31 cpd3: f3
30m do: f3 wr #0 if #0 zd
0.1u iu0
50u BLKGRAD
l0 to 5 times l6
l0 to 3 times l7
50u ; BLKGRAD
exit

ph1=0
ph2=2
ph3=0 0 1 1 2 2 3 3
ph4=2 2 3 3 0 0 1 1
ph5=0
ph6=0
ph7=0
ph8=2
ph31=0 0 2 2
; l6 number of fid after first laser pulse
; l7 number of laser irradiations
```

**AP 3:** Pulse-sequence for 2D Hadamard HSQC

```
; Hadamard 15N-HSQC
```

```
prosol relations=<triple>
```

```
#include <Avance.inc>
```

```
#include <Grad.inc>
```

```
#include <Delay.inc>
```

```
define list<shape> SPL=<sp_hadam>
```

```
"p2=p1*2"
```

```
"p22=p21*2"
```

```
"d0=3u"
```

```
"d11=30m"
```

```
"d12=20u"
```

```
"d13=4u"
```

```
"d26=1s/(cnst4*4)"
```

```
"DELTA=d19-p22/2"
```

```
"DELTA1=d26-p16-d16-p27*3-d19*5"
```

```
"DELTA2=d26-p16-d16-p27*2-p0-d19*5-8u"
```

```
"DELTA3=d26*2-p16-d16-4u"
```

```
"DELTA4=p21*2/3.1416"
```

```
"TAU=d26-p16-4u"
```

```
1 ze
```

```
    d11 pl 16: f3
```

```
2 d11 do: f3
```

```
3 d1 pl 1: f1
```

```
    50u UNBLKGRAD
```

```
    (p1 ph1)
```

```
    4u
```

```
    p16: gp1
```

```
    TAU pl 3: f3
```

```
    (center (p2 ph1) (p22 ph6): f3 )
```

```
    4u
```

```
    p16: gp1
```

```
    TAU
```

```
    (p1 ph2)
```

```
    4u
```

```
    p16: gp2
```

```
    d16 pl 19: f1
```

```
    (p21 ph3): f3
```

```
    p16: gp4
```

```
    d16
```

```
    DELTA3
```

```
    4u cpd1: f1
```

```
    (p14: SPL ph1: r): f3
```

```
    4u do: f1
```

```
    DELTA3
```

```
    p16: gp4
```

```
    d16 pl 1: f1
```

```
    (p21 ph4): f3
```

```
    4u
```

```
p16: gp2
```

```
d16
```

```
(p1 ph7)
```

```
DELTA1
```

```
p16: gp3
```

```
d16 pl 18: f1
```

```
p27*0.231 ph8
```

```
d19*2
```

```
p27*0.692 ph8
```

```
d19*2
```

```
p27*1.462 ph8
```

```
DELTA
```

```
(p22 ph1): f3
```

```
DELTA
```

```
p27*1.462 ph9
```

```
d19*2
```

```
p27*0.692 ph9
```

```
d19*2
```

```
p0*0.231 ph9
```

```
4u
```

```
p16: gp3
```

```
d16
```

```
4u BLKGRAD
```

```
DELTA2 pl 16: f3
```

```
go=2 ph31 cpd3: f3
```

```
d11 do: f3 mc #0 to 2
```

```
    F1QF(SPL.inc)
```

```
exit
```

```
ph1=0
```

```
ph2=1
```

```
ph3=0 2
```

```
ph4=0 0 0 0 2 2 2 2
```

```
ph5=0 0 2 2
```

```
ph6=0
```

```
ph7=2
```

```
ph8=1
```

```
ph9=3
```

```
ph31=0 2 0 2 2 0 2 0
```

```
; d1 : relaxation delay; 1-5 * T1
```

```
; d11: delay for disk I/O [30 msec]
```

```
; d12: delay for power switching [20 usec]
```

```
; d13: short delay [4 usec]
```

```
; d16: delay for homospoil/gradient recovery
```

```
; d19: delay for binomial water suppression
```

```
; d19 = (1/(2*d)), d = distance of next null
```

```
; d26 : 1/(4J(YH))
```

```
; cnst4: = J(YH)
```

```
; in0: 1/(2 * SW(X)) = DW(X)
```

```
; FnMODE: QF
```

**AP 4:** Pulse-sequence for 2D long range HMBC with C2'-decoupling

```

;HMBC
;with low-pass J-filter to suppress one-
bond correlations
;and decoupling of C2'
;using gradient pulses for selection

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

"p2=p1*2"
"d0=3u"
"d2=1s/(cnst2*2)"
"d6=1s/(cnst13*2)"

1 ze
2 30m
  d12 pl 9: f1
  d1 cw: f1 ph29
  4u do: f1
  d12 pl 1: f1
3 p1 ph1
  d2
  p3: f2 ph3
  d6
  p3: f2 ph4
  d0
  50u UNBLKGRAD
  p16: gp1
  d16
  center (p2 ph2): f1 (p8: sp13 ph6): f2
  50u
  p16: gp2
  d16
  d0
  p3: f2 ph5
  4u
  p16: gp3
  d16
  4u BLKGRAD
  go=2 ph31
  d1 mc #0 to 2 F10F(i d0)
exit

ph1=0
ph2=0 0 0 0 2 2 2 2
ph3=0 0 2 2
ph4=0 2
ph5=0 0 0 0 0 0 0 2 2 2 2 2 2 2 2
ph29=0
ph31=0 2 0 2 0 2 0 2 2 0 2 0 2 0 2 0 0

;pl 1 : f1 channel - power level for pulse
;          (default)
;pl 2 : f2 channel - power level for pulse
;          (default)
;pl 9 : f1 channel - power level for
;          presaturation
;p1 : f1 channel - 90 degree high power
;          pulse
;p2 : f1 channel - 180 degree high power
;          pulse
;p3 : f2 channel - 90 degree high power
;          pulse
;p8 : f2 channel - off resonance 180 shaped
;          pulse
;p16: homospoil/gradient pulse
;d0 : incremented delay (2D) [3
;          usec]
;d1 : relaxation delay; 1-5 * T1
;d2 : 1/(2J)XH
;d6 : delay for evolution of long range
;          couplings
;d12: delay for power switching [20
;          usec]
;d16: delay for homospoil/gradient recovery
;cnst2: = J(XH)
;cnst13: = J(XH) long range
;in0: 1/(2 * SW(X)) = DW(X)
;nd0: 2
;NS: 2 * n
;DS: 16
;td1: number of experiments
;FnMODE: QF

```

**AP 5:** Pulse-sequence for 2D  $^1\text{H}$ - $^{31}\text{P}$  HEHAHA

```
;rna_hpdi . t2. be
;G.W. Kellogg, J. Magn. Reson. 98, 176-182
;(1992)
```

```
#i ncl ude <Avance. i ncl >
#i ncl ude <Grad. i ncl >
#i ncl ude <Del ay. i ncl >
```

```
"p2=p1*2"
"d11=30m"
"d12=20u"
"p0=p1*4/3"
"d0=i n0/2-p2/2-p3*2/3. 1416"
```

```
"DELTA1=d21-p16-d16-d0"
"FACTOR1=(d21/(p6*115. 112))/2+0. 5"
"l 1=FACTOR1*2"
"l 4=d1/(p0+5m)"
```

```
1 ze
  d11 pl 12: f2
2 d11 do: f2
3 d12 pl 1: f1 pl 2: f2
4 (p0 ph1)
  5m
  l o to 4 times l 4
  4u
  50u UNBLKGRAD
```

```
p16: gp1
d16
(p3 ph3): f2
d0
(p2 ph5)
d0 pl 10: f1 pl 23: f2
(p20 ph4): f2
;begi n DI PSI 2: f1, DI PSI 2: f3
5 (p6*3. 556 ph11): f1 (p6*3. 556 ph21): f2
  (p6*4. 556 ph12): f1 (p6*4. 556 ph22): f2
  (p6*3. 222 ph11): f1 (p6*3. 222 ph21): f2
  (p6*3. 167 ph12): f1 (p6*3. 167 ph22): f2
  (p6*0. 333 ph11): f1 (p6*0. 333 ph21): f2
  (p6*2. 722 ph12): f1 (p6*2. 722 ph22): f2
  (p6*4. 167 ph11): f1 (p6*4. 167 ph21): f2
  (p6*2. 944 ph12): f1 (p6*2. 944 ph22): f2
  (p6*4. 111 ph11): f1 (p6*4. 111 ph21): f2
```

```
(p6*3. 556 ph12): f1 (p6*3. 556 ph22): f2
(p6*4. 556 ph11): f1 (p6*4. 556 ph21): f2
(p6*3. 222 ph12): f1 (p6*3. 222 ph22): f2
(p6*3. 167 ph11): f1 (p6*3. 167 ph21): f2
(p6*0. 333 ph12): f1 (p6*0. 333 ph22): f2
(p6*2. 722 ph11): f1 (p6*2. 722 ph21): f2
(p6*4. 167 ph12): f1 (p6*4. 167 ph22): f2
(p6*2. 944 ph11): f1 (p6*2. 944 ph21): f2
(p6*4. 111 ph12): f1 (p6*4. 111 ph22): f2
```

```
(p6*3. 556 ph12): f1 (p6*3. 556 ph22): f2
(p6*4. 556 ph11): f1 (p6*4. 556 ph21): f2
```

```
(p6*3. 222 ph12): f1 (p6*3. 222 ph22): f2
(p6*3. 167 ph11): f1 (p6*3. 167 ph21): f2
(p6*0. 333 ph12): f1 (p6*0. 333 ph22): f2
(p6*2. 722 ph11): f1 (p6*2. 722 ph21): f2
(p6*4. 167 ph12): f1 (p6*4. 167 ph22): f2
(p6*2. 944 ph11): f1 (p6*2. 944 ph21): f2
(p6*4. 111 ph12): f1 (p6*4. 111 ph22): f2
```

```
(p6*3. 556 ph11): f1 (p6*3. 556 ph21): f2
(p6*4. 556 ph12): f1 (p6*4. 556 ph22): f2
(p6*3. 222 ph11): f1 (p6*3. 222 ph21): f2
(p6*3. 167 ph12): f1 (p6*3. 167 ph22): f2
(p6*0. 333 ph11): f1 (p6*0. 333 ph21): f2
(p6*2. 722 ph12): f1 (p6*2. 722 ph22): f2
(p6*4. 167 ph11): f1 (p6*4. 167 ph21): f2
(p6*2. 944 ph12): f1 (p6*2. 944 ph22): f2
(p6*4. 111 ph11): f1 (p6*4. 111 ph21): f2
```

```
(p6*3. 556 ph11): f1 (p6*3. 556 ph21): f2
(p6*4. 556 ph12): f1 (p6*4. 556 ph22): f2
(p6*3. 222 ph11): f1 (p6*3. 222 ph21): f2
(p6*3. 167 ph12): f1 (p6*3. 167 ph22): f2
(p6*0. 333 ph11): f1 (p6*0. 333 ph21): f2
(p6*2. 722 ph12): f1 (p6*2. 722 ph22): f2
(p6*4. 167 ph11): f1 (p6*4. 167 ph21): f2
(p6*2. 944 ph12): f1 (p6*2. 944 ph22): f2
(p6*4. 111 ph11): f1 (p6*4. 111 ph21): f2
```

```
(p6*3. 556 ph12): f1 (p6*3. 556 ph22): f2
(p6*4. 556 ph11): f1 (p6*4. 556 ph21): f2
(p6*3. 222 ph12): f1 (p6*3. 222 ph22): f2
(p6*3. 167 ph11): f1 (p6*3. 167 ph21): f2
(p6*0. 333 ph12): f1 (p6*0. 333 ph22): f2
(p6*2. 722 ph11): f1 (p6*2. 722 ph21): f2
(p6*4. 167 ph12): f1 (p6*4. 167 ph22): f2
(p6*2. 944 ph11): f1 (p6*2. 944 ph21): f2
(p6*4. 111 ph12): f1 (p6*4. 111 ph22): f2
```

```
l o to 5 times l 1
;end DI PSI 2: f1, DI PSI 2: f3
```

```
(p17 ph11)
4u pl 12: f2
4u BLKGRAD
go=2 ph31 cpd2: f2
d11 do: f2 mc #0 to 2
F1PH(i p3, i d0)
```

```
exi t
```

```
ph1=0
ph2=1
ph3=1 3
ph4=0 0 0 0 2 2 2 2
ph5=0 0 2 2
ph11=0
ph12=2
ph21=0 0 0 0 2 2 2 2
ph22=2 2 2 2 0 0 0 0
ph31=0 2 0 2 2 0 2 0
```

**AP 6:** complex Lorentzian function

$$L(\Omega, \Omega^0, \lambda) = \frac{1}{\lambda + i(\Omega - \Omega^0)}$$

## PHOTO-CIDNP OF RNA

Generally, the sensitivity of a NMR experiment is determined by the population difference of the spin states. This difference is described by a Boltzman distribution according to

$$\frac{N_{\beta}}{N_{\alpha}} = \exp\left(\frac{\gamma B_0 h}{kT}\right)$$

and is inherently low in NMR. Methods that dramatically increase the sensitivity of NMR experiments therefore deal with the excitation of non-Boltzmann states of the nuclear spins. Among the different NMR methods that exploit non-Boltzmann excitation such as CIDNP<sup>226</sup> or DNP<sup>227</sup>. In the following, CIDNP experiments are commented, since they have been successfully used in the field of protein folding and might have also impact in the structural characterisation of RNA molecules.

### PHOTO-CIDNP

Chemical induced dynamic nuclear polarisation (CIDNP) induces a non-Boltzmann nuclear spin state distribution in thermal or photochemical reactions, usually from collision and diffusion or disproportion of radical pairs, and is detected by NMR by enhanced absorption or emission signals. It allows large improvements of sensitivity and reduction of spectral crowding. In the field of biomolecular NMR, photo-CIDNP experiments provide insight in the solvent accessibility of aromatic residues. The polarisation of aromatic residues in proteins is achieved by photo-sensitizers such as flavins and aza-aromatics. Early studies<sup>228</sup> demonstrated that also for RNA the solvent accessibility of adenine and guanine residues can be monitored.

### PHOTO-CIDNP of RNA

The first photo-CIDNP studies in the nucleic acids-field were carried out on single nucleobases, nucleosides and nucleotides. Thereby signal enhancement was found for A and G in all three constitutions and for T, 5MeCyt and 3MeCyt in the base form<sup>229</sup>. Interestingly also the H2 and H8 signal revealed CIDNP effects in the dinucleotides d(ApA) and d(pApA)<sup>230</sup>. The following studies focussed merely on the CIDNP of purines since they show the greatest signal enhancement effects. The strongest amplification is observed on the resonances of G(H8) and if both G and A are present in the sample the CIDNP effect of A is suppressed by G. Although the magnitude of the effect is lower it is worthwhile noting that in pyrimidines U(H5) displays a negative polarisation<sup>228a</sup>.

---

<sup>226</sup>R. Kaptein, K. Dijkstra, K. Nicolay, *Nature*, 1978, **274**, 293-294

<sup>227</sup>C.T. Farrar, D.A. Hall, G. J. Gerfen, M. Rosay, J.H. Ardenkjaer-Larsen, R.G. Griffin, *J Magn Reson.*, 2000, **144**, 134-141; J.H. Ardenkjaer-Larsen, B. Fridlund, A. Gram, G. Hansson, L. Hansson, M.H. Lerche, R. Servin, M. Thaning, K. Golman, *Proc. Natl. Acad. Sci.*, 2003, **100**, 10158-10163

<sup>228</sup>E.F. Mccord, K.M. Morden, A. Pardi, I. Tinoco, S.G. Boxer, *Biochemistry*, 1984, **23**, 1926-1934; E.F. Mccord, K.M. Morden, I. Tinoco, S.G. Boxer, *Biochemistry*, 1984, **23**, 1935-1939

<sup>229</sup>R. Kaptein, K. Nicolay, K. Dijkstra, *J. Chem. Soc. Chem. Commun.*, 1979, 1092-1094

<sup>230</sup>R. Kaptein, S. Stob, R.M. Scheek, K. Dijkstra, T.W. Schleich, *Bull. Magn. Reson.*, 1981, **3**, 28-30

Studies on double stranded RNA molecules of the sequence  $[5'-AGCU-3']_2$  showed no enhancement at low temperatures as 5°C-10°C, but significant signals at elevated temperatures as 20°C to 30°C. This results in the conclusion that melting ( $T_m=17.5^\circ\text{C}\pm 0.5^\circ\text{C}$ ) has to occur first and CIDNP signals are just observable for single stranded, unpaired residues of the RNA. Or in other words being in the context of a double helix has shields the nucleotides from being polarised by the photo-sensitizer. Additionally external sample parameters are determined to have a considerable effect on the CIDNP intensity. So the CIDNP-signals of I(H8) and G(H8) decreases if the concentration of Phosphate increases (approx. signal is attenuated to half if going from 0mM to 750mM phosphate). As optimal a pH range between pH7-8 is verified and generally the rule applies that the higher the temperature the lower the CIDNP effect is. In several cases also polarisation of the G(H1') proton is observed, arising from a  $\chi$  angle dependent cross-relaxation between this and the aromatic proton of the base. The first study on a biologically relevant RNA the yeast tRNAPhe reveals again that only solvent accessible G and modified G nucleotides are polarised<sup>228b</sup>.

## MATERIALS AND METHODS

### Laser-Setup

Similar to the setup described in chapter IV an argon ion-laser (with a maximum power of 7.5W at 488/515nm) is connected to the sample via the described fibre optics. However, illumination of the sample tube for the photo-CIDNP experiments is realized by routing the fibre directly into the tube as described early by Berliner and co-workers<sup>231</sup>. The usage of pencil shaped Shigemitsu tubes can be omitted because compared to the photochemical reactions described earlier the quantum yield is much lower. An advantage of the direct routing is that the mandatory magnetic field homogeneity is achieved more easily.

### Sample and Conditions

Typical samples had a volume of 500 $\mu\text{l}$  and contained 500 $\mu\text{M}$  to 100 $\mu\text{M}$  analyte (nucleobase, nucleotide or RNA) and 200 $\mu\text{M}$  to 50 $\mu\text{M}$  photosensitizer (normally FMN) in 25mM or 200mM phosphate buffer. All samples were prepared with 10% D<sub>2</sub>O as lock substance.

As reported for photo-CIDNP of proteins there are different photosensitizers than FMN, some of those were also tested which are thionine and 2-2'-bipyridyl.

---

<sup>231</sup>J.E. Scheffler, C.E. Cottrell, L.J. Berliner, 1985, *J. Magn. Reson.*, **63**, 199-201



## NMR-Experiments

All photo-CIDNP experiments were conducted as described in the following scheme:

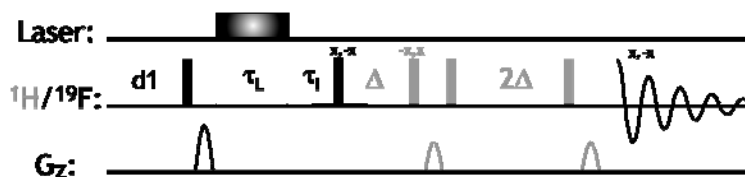


Figure 63: Schematic representation of pulse-sequence used for photo-CIDNP measurements; small rectangulars represent hard  $\pi/2$  pulses, the black bar indicates the laser irradiation during the time period  $\tau_L$ ;  $d1$  is the relaxation delay; in the case of proton detected photo-CIDNP experiments water suppression is achieved by an additional jump return pulse element indicated in grey (therefore  $\Delta$  is the  $j$ -r-delay);  $\tau_1$  is the time between the end of the laser irradiation and the start of the NMR read-out sequence.

The initial pulse-gradient element was used to destroy the equilibrium magnetisation. Subsequently laser irradiation induces the CIDNP-effect and lead to generation of polarisation which can be detected by the subsequent NMR pulse-sequence read-out.

## PHOTO-CIDNP AND RNA-LIGAND INTERACTIONS

As described before, X-ray structures showed that ligand-binding by the guanine riboswitch RNA is associated with the formation of a tight binding pocket (see section chapter III & IV). Comparison of  $^1\text{H}$ -NMR and photo-CIDNP spectra of the free ligand hypoxanthine (in the presence of the photosensitizer flavin mononucleotide (FMN) before and after laser irradiation of 0.5s) revealed enhanced signals for its protons H2/H8 (Figure 64). In contrast, no enhancement effect on the signals of the free RNA could be detected (Figure 64). Hence, the accessibility of the ligand complexed with the RNA in solution was studied by NMR using photo-CIDNP (Figure 64). While the signals at 8ppm, corresponding to free hypoxanthine still showed a CIDNP effect that increased as the ligand-RNA ratio increased, no effect could be obtained for signals of bound hypoxanthine, confirming the lack of accessibility of the ligand in complex with the RNA also in solution. Therefore, it should be possible that a RNA folding reaction can be monitored by this methodology. This could be further improved by the utilisation of fluorinated nucleotides<sup>232</sup> that in close analogy to the protein case might show higher CIDNP intensities<sup>233</sup>.

<sup>232</sup>L.G. Scott, B.H. Geierstanger, J.R. Williamson, M. Hennig, *J. Am. Chem. Soc.*, 2004, **126**, 11776-11777

<sup>233</sup>I. Kuprov, P.J. Hore, *J. Magn. Reson.*, 2004, **168**, 1-7

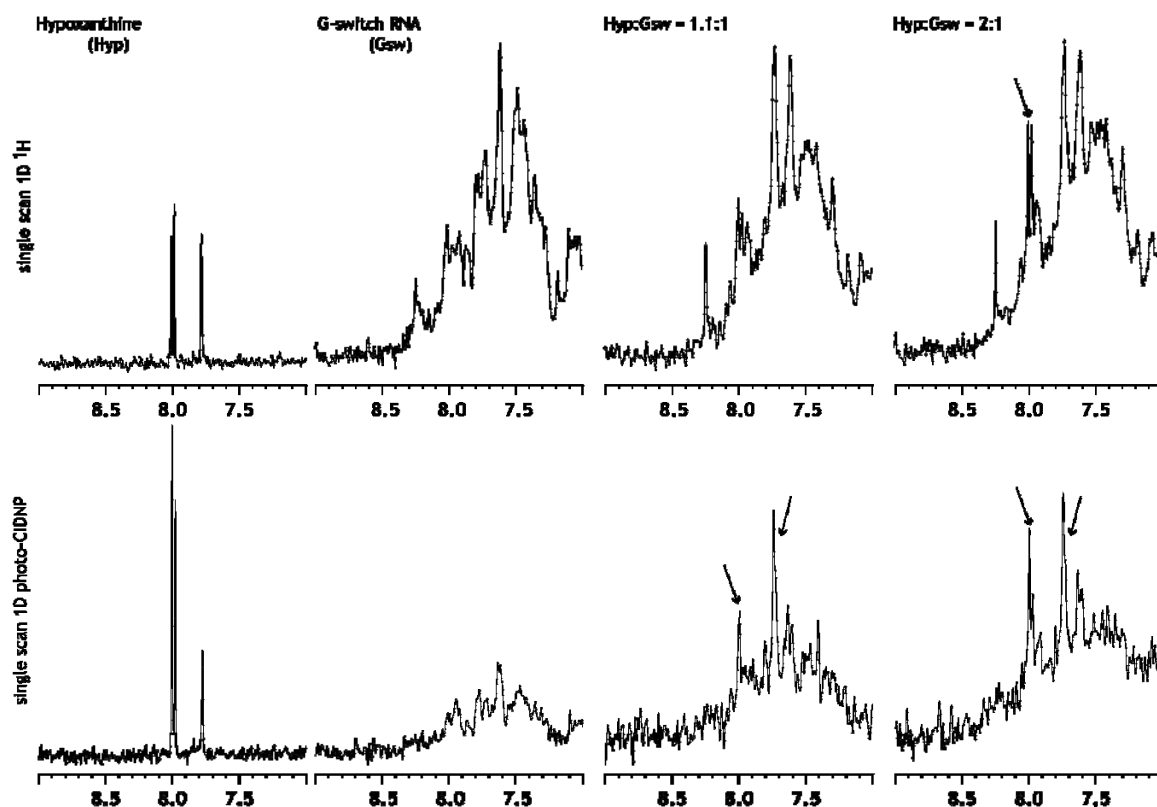
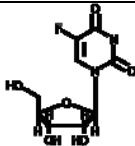
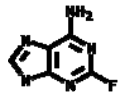
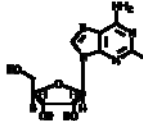


Figure 64: Upper panel from left to right: <sup>1</sup>H-NMR spectrum of hypoxanthine (0.25mM) with FMN (0.2mM) before laser pulse; <sup>1</sup>H-NMR spectrum of guanine-sensing riboswitch RNA (0.1mM) with FMN (0.05mM) before laser pulse; <sup>1</sup>H-NMR spectrum of hypoxanthine (0.11mM) and guanine-sensing riboswitch RNA (0.1mM) with FMN (0.2mM) before laser pulse; <sup>1</sup>H-NMR spectrum of hypoxanthine (0.2mM) and guanine-sensing riboswitch RNA (0.1mM) with FMN (0.2mM) before laser pulse; all spectra were recorded at 600MHz with 4096points and 1scan (temperature 288K). Lower panel from left to right: 1D photo-CIDNP spectrum of hypoxanthine (0.25mM) with FMN (0.2mM) directly after laser irradiation (500ms); the signals at 8ppm correspond to H2 and H8 of hypoxanthine, the signal at 7.75ppm corresponds to FMN; 1D photo-CIDNP spectrum of guanine-sensing riboswitch RNA (0.1mM) with FMN (0.05mM) directly after laser irradiation (500ms); 1D photo-CIDNP spectrum of hypoxanthine (0.11mM) and guanine-sensing riboswitch RNA (0.1mM) with FMN (0.2mM) directly after laser irradiation (500ms); 1D photo-CIDNP spectrum of hypoxanthine (0.2mM) and guanine-sensing riboswitch RNA (0.1mM) with FMN (0.2mM) directly after laser irradiation (500ms); all spectra were recorded at 600MHz with 4096points and 1scan (H<sub>2</sub>O/D<sub>2</sub>O (9:1); temperature 288K).

#### PHOTO-CIDNP OF FLUORINATED COMPOUNDS

In order to test if photo-CIDNP on fluorinated RNA building blocks is detectable we use primarily four different commercially available compounds: 5-Fluorocytidine, 5-Fluorouridine, 2-Fluoroadenine and 2-Fluoroadenosine. As photosensitizer FMN is used. Irradiation is applied for 1s at a laser power of 1W at 488/515nm. Whereas 5-Fluorocytidine does not show any detectable CIDNP signal, the other three compounds do exhibit a signal enhancement in photo-CIDNP measurements (for results see table 10). The strongest effect is monitored for the fluorinated nucleobase of adenine, compared to this the other two compounds show an enhancement decreased by a factor of ~5.

Table 10: Positive results of  $^{19}\text{F}$ -CIDNP measurements

Substance	S/N $^{19}\text{F}$ -1D	S/N $^{19}\text{F}$ -CIDNP	Enhancement	
5-Fluorouridine	10.1	22.22	2.2	
2-Fluoroadenine	6.54	71.04	10.86	
2-Fluoroadenosine	37.03	92.16	2.48	

As reported for the proton detected photo-CIDNP experiments the enhancement effect for purine bases seems to be higher than those for the pyrimidines, being in line with the explanation that because of the more extended aromatic ring system a better association of the photosensitizer towards the analyte is achieved. Interestingly, the intensity of the effect drops if going from the nucleobase constitution to the nucleosides.

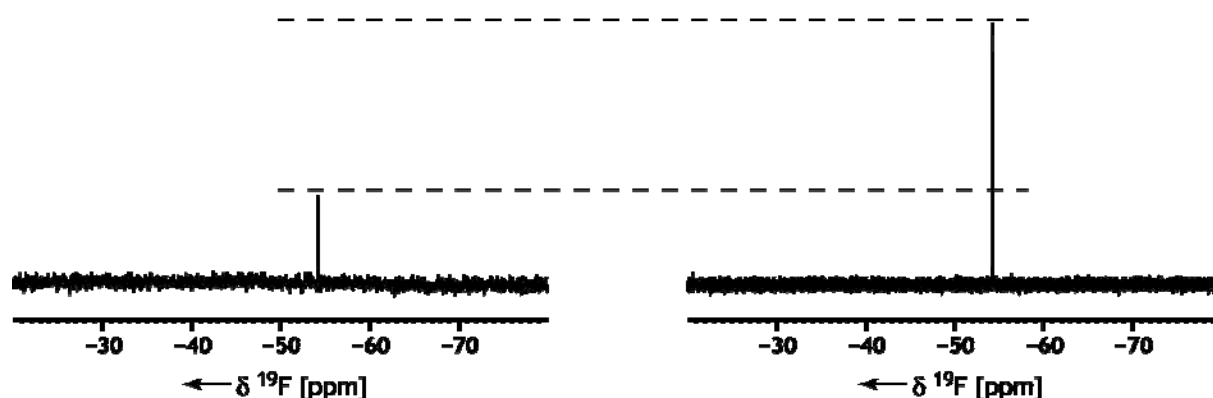


Figure 65: 1D  $^{19}\text{F}$  spectrum (left panel) and 1D  $^{19}\text{F}$  photo-CIDNP spectrum of 2-fluoroadenine (right panel), both measured at 600 MHz at 288 K. Hard  $^{19}\text{F}$   $\pi/2$  pulse was applied with field strength of 23.8 kHz. Transmitter frequency was set to -60 ppm, spectra were recorded with a SW of 177 ppm and digitized with 128 k points. For the 1D spectrum 8 transients were acquired, for the photo-CIDNP spectrum just a single transient was acquired, laser irradiation was set to 1 s at 1 W laser output power. Spectra are plotted at the same noise level; dashed lines indicate the difference in signal intensity.

Also in line with the earlier findings is that for C nucleosides no effect is detectable.

Based on these findings photo-CIDNP experiments were conducted on 2-F-A labelled RNA molecules. The HIV-2 transactivation response element (TAR) was used as example molecule, where all A-nucleotides were substituted by 2-FA. The RNA was biochemically synthesised by Prof. Dr. Mirko Hennig as described in Scott et al.. The  $^{19}\text{F}$ -1D spectrum showed the expected

pattern of four peaks as depicted in Figure 66. Addition of the specific ligand arginine-amide leads to changes in the chemical shift. The aim of the experiments was to possibly detect changes in the photo-CIDNP effect between these two states.

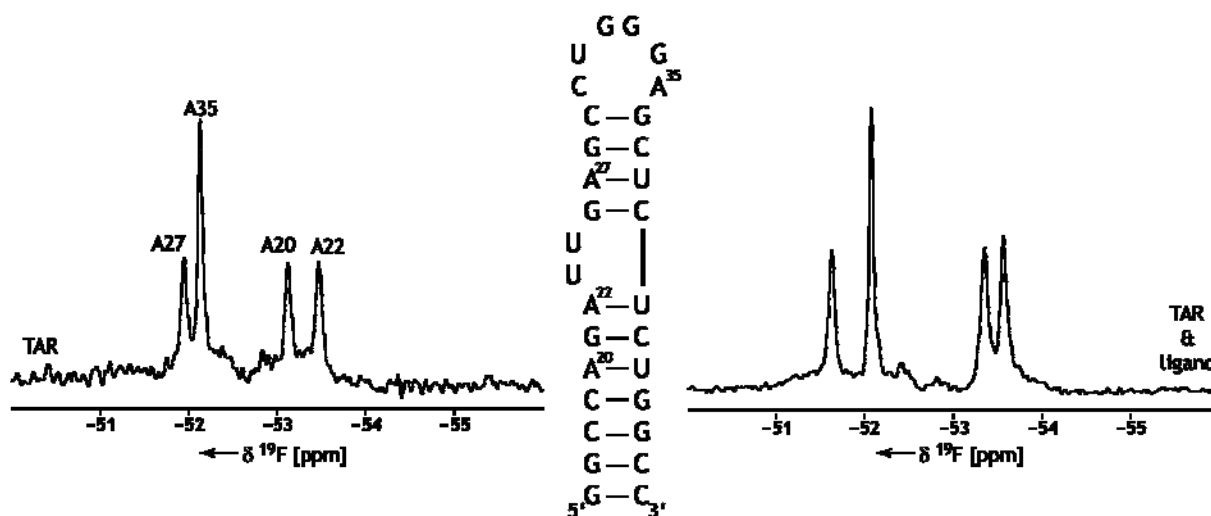


Figure 66: 1D  $^{19}\text{F}$  spectra of 2FA-labelled TAR RNA (secondary structure as depicted in the middle panel), TAR RNA with FMN as photosensitizer (left panel) and TAR RNA plus arginine-amide as ligand and with FMN as photosensitizer (right panel).

Unfortunately, under several tested conditions (FMN, thionine, and bipyridyl as photosensitizer; different phosphate buffer concentrations between 25mM and 200mM) no photo-CIDNP effect could be detected. This could be due to the fact that a) incorporated into the context of a structured RNA 2-FA does not show any photo-CINP effect at all or b) that none of the labelled bases in TAR is as solvent exposed so that association of the photo-sensitizer can occur.

As revealed on non-fluorinated nucleotides the photo-CIDNP intensity is higher by a factor of  $\sim 4$  to  $\sim 40$  depending on the phosphate buffer concentration for the H8 proton towards the H2 proton. If this also would hold true for the fluorinated case this could circumvent the negative results of  $^{19}\text{F}$  photo-CINP experiments on RNA. Unfortunately again, the synthesis of such labelled RNA molecules is in the moment not available.

#### PHOTO-CIDNP AND RT-NMR

Recapitulating the application of photo-CIDNP experiments on RNA it could be possible to exploit the effect in RT-NMR studies of RNA. The advantage of such an approach would be that due to the fact that non-Boltzman magnetisation is utilized the need of a long recycling delay is omitted and therefore fast pulsing is possible<sup>234</sup>.

- 1) Monitoring the time course of ligand binding: If the resonances of the ligand arise in photo-CIDNP experiments they can be followed in a time resolved manner during

<sup>234</sup>J. Wirmer, T. Kühn, H. Schwalbe, *Angew. Chem. Int. Ed.*, 2001, **40**, 4248-4251

the binding process. Preconditions would be that either the chemical shift or the polarisation is modulated during the reaction.

- 2) Monitoring the time course of fluorinated-ligand binding: If fluorinated ligands happen to bind the RNA molecule and show photo-CIDNP effects (as 2F-Adenosine to the adenine-dependent riboswitch), than similar to (1) the binding event could be monitored. The inherent advantage would be that exclusively resonance of the ligand would arise and possibly report on the direct binding event, whereas the RNA resonances allow obtaining a model of how the RNA behaves during ligand binding. Dissected structural models could be obtained by such an approach.

A general pitfall might occur that in the case of the concrete application might be easily overcome by changing the ratio of solutes: If the reaction initiation is photochemical radical reactions are involved that might get quenched by the aromatic system of the photosensitizer. But this was also overcome in the case of the protein folding studies.



## ZUSAMMENFASSUNG

Die vorliegende Doktorarbeit beschäftigt sich mit den strukturellen Änderungen in RNA Molekülen während dynamischer konformationeller Änderungen, die gemeinhin als RNA-Faltung bezeichnet werden. Im Gegensatz zur Proteinfaltung sind RNA-Faltungsprozesse nicht exklusiv als die Faltung einer definierten Konformation aus einem Ensemble an ungefalteten, d.h. ausgehend von unstrukturierten Molekülen, zu verstehen. RNA-Faltung beinhaltet vielmehr die strukturelle Umwandlung verschiedener stabiler Konformationen (die als RNA-Umfaltung benannt wird) und den Aufbau von molekularen Komplexen aus mehreren Molekülen (siehe Kapitel III). Die experimentelle Technik, die hier zur Untersuchung dieser Prozesse genutzt wurde, ist die hochauflösende Flüssig-NMR-Spektroskopie. Das Verständnis der strukturellen und biophysikalischen Grundlagen solcher Umfaltungsreaktionen von RNA ist essentiell, da solche konformationellen Änderungen die biologische Funktion der Moleküle modulieren. Dabei ist zu bemerken, dass eine Umfaltungsreaktion eine intrinsische Eigenschaft einer gegebenen RNA-Sequenz sein kann oder die Antwort auf ein externes zelluläres Signal, wie die Bindung eines niedermolekularen Liganden (z.B. in Aptameren und in Riboswitch RNAs), eines Proteins oder eines Metall-Ions.

Der erste Teil dieser Doktorarbeit (Kapitel I & II) hält einen Überblick über die Themengebiete RNA-Struktur und RNA-Faltung bereit. Beide Kapitel führen in den derzeitigen Stand der Forschung ein. Kapitel II führt dabei entlang der hierarchischen Ordnung von RNA Molekülen und diskutiert die Eigenschaften von Primär-, Sekundär- und Tertiär-Strukturelementen. Ein besonderes Augenmerk wird dabei auf bistabile RNA Systeme gelegt; ihre wichtige biologische Funktionalität wird dargestellt, ebenso wird das Potential ausgeleuchtet, diese funktionale Klasse von RNA Molekülen als Modellsysteme zu nutzen, um fundamentale Fragen zu konformationellen Übergängen in RNA zu beantworten. In Kapitel III folgt sodann die Diskussion über RNA-Faltung in *in vitro* Experimenten als auch im zellulären Kontext (*in vivo*). Die Kapitel IV und V besprechen die NMR-spektroskopischen Techniken, die genutzt werden, um die Art und die dynamischen Eigenschaften von konformationellen/strukturellen Umwandlungen in RNA zu untersuchen. Hierbei wird der Schwerpunkt auf die verwendeten Techniken des Wasseraustauschs an labilen Protonen und der zeitaufgelösten NMR-Spektroskopie gelegt.

Der zweite Teil der Doktorarbeit fasst kumulativ die durchgeführten Studien zusammen. Kapitel VI bespricht hierbei die grundlegenden NMR Techniken, die zur Strukturaufklärung von RNA Molekülen angewendet werden und zeigt deren Anwendungsmöglichkeiten an unterschiedlichen Beispielen von strukturellen und biochemischen Studien.

Der Übersichts-Artikel beschreibt, dass die NMR Spektroskopie eine wertvolle Technik zur Charakterisierung der molekularen Struktur, der Dynamik von RNA Molekülen und der Interaktion zwischen RNA und Proteinen, RNA und Metallionen, RNA und niedermolekularen

Liganden und RNA und Solvenz-Molekülen ist. Eine besondere Betonung fanden die für biochemische Fragestellungen am meisten relevanten Aspekte:

- Beschreibung des Musters der Basenpaarungen: Hierbei wird erläutert, wie kanonische und nicht-kanonische Basenpaarungen in RNA Molekülen mittels NMR-Spektroskopie beschrieben werden können und wie daraus Modelle für die Sekundärstruktur von RNA-Molekülen abgeleitet werden. Des Weiteren wird bedacht, wie die Dynamik einer Basenpaarung mittels NMR-Methoden aufgeklärt und in dynamische Modelle übersetzt werden kann.
- Beschreibung von konformationellen Gleichgewichten: Hierbei wird der Beschreibung von Gleichgewichten zwischen verschiedenen Konformationen eines RNA-Moleküls gedacht. Die unterschiedlichen Zeitskalen auf denen solche Prozesse ablaufen und die nötigen NMR-Experimente zur Beschreibung dieser werden diskutiert.
- Beschreibung von orts aufgelösten Bindungsstudien.
- Beschreibung der Versatzbauweise von RNA Molekülen: Hier wird aufgezeigt, dass der modulare Aufbau der RNA eine strukturelle Beschreibung eines großen Moleküls mittels der Verwendung kleiner Subfragmente ermöglicht.
- Beschreibung der Dynamik, der lokalen und der globalen Struktur eines RNA Moleküls mittels neuartiger Parameter z.B. dipolare Restkopplungen (RDCs).

Weiterhin werden die Synthese-Methoden zur Herstellung von isotope markierter RNA und die nötigen Experimente für vollständige Resonanzzuordnung und strukturelle Charakterisierung beschrieben und an eigenen Beispielen verdeutlicht.

Haarnadel-Schleifen (Hairpin-loops) werden aus einer antiparallelen Helix –dem so genannten Stem–, die von einer Schleife (loop) geschlossen wird, gebildet. Sie stellen unter allen bekannten Sekundärstrukturelementen von RNA-Molekülen ein dominantes Modell dar. Der Loop kann prinzipiell eine unterschiedliche Anzahl von ungepaarten Nukleotiden tragen. Die am weitesten verbreiteten Haarnadelschleifen jedoch sind vier ungepaarte Nukleotide enthaltende Tetraloops. Tetraloops findet man in vielfältigen Funktionen, so sind sie bei der Faltung von RNA als Nukleationspunkte beschrieben. Wichtig sind sie auch bei der Protein-RNA-Erkennung, so sind Tetraloops in die Wechselwirkung zwischen HIV-TAR-RNA und TAT-Protein eingebunden. Interessanterweise verfügen Tetraloops der Sequenz cYNMGg über die höchste thermodynamische Stabilität. verfügen. Das folgende Kapitel VII beschreibt die komplette Resonanzzuordnung eines RNA Modell-Moleküls (14mer cUUCGg tetra-loop RNA) und stellt eine neue Pulstechnik vor, die zur Zuordnung der Resonanzen von quaternären Kohlenstoffen in Purinbasen benützt werden kann. Da eine strukturelle Charakterisierung vorliegt (sowohl aus eigenen NMR-Parametern als auch aus vergleichenden Literaturdaten), ermöglichte die bis dato vollständigste Resonanzzuordnung die Validierung der chemischen Verschiebung als strukturabhängigen Parameter. Die möglich gemachte Zuordnung der quaternären Kohlenstoffe der Nukleobasen mittels der neuen Experimente wird ihre Wirkungskraft im Laufe der zurzeit



stattfindenden Renaissance  $^{13}\text{C}$ -detektierter Experimente, auch an großen RNA-Molekülen, entfalten. Weiterhin schließt sich ein Report an, wie die Konformation des Zuckerrückgrates in RNA-Molekülen bestimmt wird und schlägt mittels einer an oben genanntem Modellsystem durchgeführte Parametrisierung  $^1\text{J}$  skalare Kopplungen als neue Strukturparameter vor.

### Ergebnisse:

- Erste vollständige NMR Resonanzzuordnung eines RNA-Moleküls (14mer cUUCGg tetra-loop RNA)
- Etablierung einer neuen Pulssequenz, welche die Zuordnung von quaternären Kohlenstoffen ermöglicht
- Beschreibung der Rückgratkonformation eines RNA-Moleküls anhand eines vollständigen Satzes an NMR Parametern
- Beschreibung der Abhängigkeit neuer NMR Parameter ( $^1\text{J}$  skalare Kopplungen) zur Bestimmung der Zuckerkonformation in RNA

-----

Die katalytischen und regulatorischen Eigenschaften von Ribonucleinsäuren hängen häufig von konformationellen Umlagerungen sowohl der Tertiär- als auch der Sekundärkonformation ab. Ihre vielfältige Dynamik wird eindrucksvoll durch das Auftreten von parallelen Faltungswegen sowie von metastabilen und koexistierenden („bistabilen“) Konformationen belegt. Die Faltung von Ribonucleinsäuren aus einem denaturierten, nicht-nativen Zustand wurde eingehend mit unterschiedlichsten Methoden untersucht. Dagegen sind nur wenige zeitaufgelöste Studien zur RNA-Faltung aus einem ungestörten, nativen Zustand bekannt. Kapitel VII & VIII fassen die hierzu durchgeführten RT-NMR Studien zusammen. Kapitel VIII gibt hierbei einen Überblick über die Untersuchungen an drei bistabilen RNA-Systemen. Diese Moleküle wurden ausgewählt, da sie als Modelle für RNA-Umfaltungsreaktionen dienen. Eine neue Methode für die Untersuchung struktur- und zeitaufgelöster RNA-Faltung wird eingeführt, die auf der photolytischen Freisetzung voreingestellter Faltungskonformationen in einem thermodynamischen Ungleichgewichtszustand und anschließendem Verfolgen des Umfaltungsprozesses mittels Echtzeit-NMR-Spektroskopie beruht. Diese Methode wurde für eine detaillierte kinetische Charakterisierung der Umfaltung jener drei bistabiler Ribonucleinsäuren angewendet, die jeweils in zwei koexistierenden Haarnadelstrukturen vorliegen. Um eine der beiden Konformationen selektiv zu destabilisieren, wurden ausgewählte Guanosine oder Uracile, die nur in einer Konformation Basenpaare bilden, durch (S)-[(Nitrophenyl)ethyl]-guanosin oder (R)-[(Nitrophenyl)ethyl]-uracil ersetzt. Wie zu erwarten, bricht die sterisch anspruchsvolle NPE-Gruppe Watson-Crick-Basenpaare auf und destabilisiert dadurch die assoziierte Sekundärkonformation; durch nachfolgende Photolyse werden jedoch die ursprünglichen Paarungseigenschaften zurückerhalten und der Umfaltungsprozess kann stattfinden.

**Ergebnisse:**

- Erste Studien zu RNA Umfaltungsreaktionen mit atomarer Auflösung
- Anwendung neuer RT-NMR Techniken
  - sowohl bei der photolytischen Initiierung der Reaktion
  - als auch bei der Aufnahme der zeitaufgelösten NMR Spektren
- Beschreibung unterschiedlicher Umfaltungsmechanismen für unterschiedliche RNA-Moleküle
- Formulierung genereller RNA Faltungs-Regeln

Das finale Kapitel IX behandelt die Anwendung der oben ausgeführten neuen Methodologie auf konformationelle Umwandlungen von RNA Tertiär-Strukturelementen: a) Guanin-abhängige Riboswitch RNA (GSW) und b) Minimales „hammerhead“ Ribozym (mHHR).

Bei Riboswitch-RNAs handelt es sich um RNA Moleküle, welche Liganden niederen Molekulargewichts binden. Sowohl in Bakterien als auch in Pflanzen und Pilzen ist diese Ligandbindung an eine neue Form der Regulation der Genexpression gekoppelt. Bisher wurden Riboswitche gefunden, die spezifisch an essentielle Koenzyme und Vitamine, an Aminosäuren, an Glukosamin-6-phosphat und die Purinbasen Guanin und Adenin binden. Riboswitch-RNAs sind vor allem in den nicht-translatierten Bereichen am 5'-Ende (5'-UTR) der messenger RNAs (mRNA) lokalisiert und gehören damit zum nicht-kodierenden Teil der mRNA. Viele Riboswitch-RNAs weisen eine modulare Architektur auf: sie bestehen aus einer Aptamerdomäne auch als Sensorregion bezeichnet, die den Ligand bindet, und einer sogenannten Expressionplattform, die durch eine allosterische Umoorientierung auf den Ligandbindungszustand der Aptamerdomäne reagiert und in Folge dessen die Genexpression moduliert. Die Bindung des Liganden an die Aptamerdomäne führt zum Abbruch der Transkription, hebt die Ribosomenbindung auf oder induziert unterschiedliche Stabilitäten der mRNA aufgrund unterschiedlicher mRNA-Prozessierung.

Die Analysen von Sequenz und Sekundärstruktur weisen darauf hin, dass 2% aller Gene in *B.subtilis* zumindest zum Teil durch einen Riboswitchmechanismus reguliert werden. Die Bindung des Liganden niederen Molekulargewichts an die Aptamerdomäne erzeugt einen direkten Regelkreislauf, um die Häufigkeit der *de-novo* Synthese von denjenigen Proteinen entweder zu erhöhen oder zu unterdrücken, die am metabolischen oder katabolischen Kreislauf des Metaboliten selbst beteiligt sind. Die Aptamerdomänen natürlich vorkommender Riboswitche sind für gewöhnlich viel größer als diejenigen nicht-natürlicher Aptamere, wobei die Riboswitch-RNAs ähnliche Liganden mit Affinitäten im nanomolaren Bereich binden. Die Ratenkonstanten für die Ligandbindung bei physiologischen Ligandkonzentrationen sind zu langsam für ein Modell demzufolge Ligandbindung der konformationellen Umwandlung vorangehe und dadurch die Genexpression reguliere. Stattdessen wird im Mechanismus berücksichtigt werden müssen, dass die Regulation während der Transkription stattfindet, und dass allgemeine Protein-Transkriptionsterminationsfaktoren ebenfalls eine Rolle spielen könnten. Der Beitrag dieser Doktorarbeit kann als sich an folgendem Zitat orientierend beschrieben werden: "Therefore, to fully understand riboswitch functioning, the atomic resolution picture of ligand-RNA contacts

needs to be put in the context of the whole expression platform, its folding process, and kinetic coupling dictated by the RNA polymerase.”

Das minimale Hammerhead-Ribozym wird seit Jahren als Model bei der Beschreibung von dynamischen Übergängen in RNA Molekülen genutzt. Mit dieser Arbeit wird das molekulare Bild der sequenzspezifischen Katalyse der Spaltung des Phosphatrückgrates um einen NMR-spektroskopischen Beitrag erweitert.

**Ergebnisse:**

- Resonanzzuordnung der Imino-Signale des Hypoxanthin-GSW RNA-ligand Komplexes
- Anwendung von RT-NMR Techniken zur Untersuchung der Ligand induzierten Umfaltung des GSW mit atomarer Auflösung
- Umwandlung von kinetischer Information in strukturelle Parameter zur Beschreibung der beobachteten Prozesse
- Formulierung eines Faltungsmechanismus für das GSW-RNA Molekül
- Anwendung von RT-NMR Techniken zur Untersuchung der katalytischen Reaktion des mHHR mit atomarer Auflösung



**BORIS FÜRTIG**

Schulstrasse 4  
61206 Nieder-Wöllstadt

**PERSONAL DATA**

Date of birth 23<sup>rd</sup> May 1978  
Place of birth Bad Nauheim (Germany)  
Marital status married to Julia Christina Fürtig nee Kahl  
Citizenship German

**UNIVERSITY EDUCATION**

Graduation  
*06/2003 – 04/2007* Graduation at the BMRZ (Centre for biomolecular magnetic resonance – at the Institute for Organic Chemistry and Chemical Biology) under supervision of Prof. Dr. Harald J. Schwalbe

Course of studies  
*10/1998 – 04/2003* Study of Biochemistry at the Johann Wolfgang Goethe-University in Frankfurt am Main, Certificate Diploma

*09/2002* Prediploma in Biochemistry

**ALTERNATIVE CIVILIAN SERVICE**

*09/1997 – 09/1998* Johann Peter Schäfer-Schule Friedberg, Residential School and School for blind and visually handicapped people

**SCHOOLING**

*08/1988 – 07/1998* Augustiner Schule Friedberg, Gymnasium (secondary school)  
Certificate Abitur

*07/1984 – 07/1988* Joseph von Eichendorff-Schule Ilbenstadt, (primary school)

**ACADEMIC TEACHERS**

Prof. Dr. E. Bamberg, Prof. Dr. J. Bereiter-Hahn, Prof. Dr. W. Brüggemann, Prof. Dr. E. Egert, Prof. Dr. J.W. Engels, Prof. Dr. K.D. Entian, Prof. Dr. Dr. H. Fasold, Prof. Dr. J. Feierabend, Prof. Dr. W. Gnatzy, Prof. Dr. M. Göbel, PD Dr. U. Günther, Prof. Dr. C. Griesinger, Prof. Dr. S.K. Hashmi, Prof. Dr. Kolbesen, Prof. Dr. B. Ludwig, Prof. Dr. Dr. R. Protsch von Zieten, Prof. Dr. H. Roskos, Prof. Dr. H. Rüterjans, Prof. Dr. M. Sattler, Prof. Dr. H. Schwalbe, Prof. Dr. R. Tampé, Prof. Dr. M. Wagner, Prof. Dr. S. Wiedenhofer, Prof. Dr. J. Wöhnert, Prof. Dr. H. Zimmermann



**ACKNOWLEDGEMENT**

„Mache es bitte horizontal.“ (Harald)

First of all I would like to thank my supervisor Harald Schwalbe. I am deeply indebted to him for giving me the opportunity to join his group with this most interesting project imaginable. His continuous support and interest in our projects, although he is strongly limited in time, is greatly acknowledged. He is providing an atmosphere in his group best suitable to conduct research in an independent manner without compromising in support. Dear Harald thanks a lot for this way of supported freedom and for all the opportunities you made available to me.

„Das ist ja mit Schweinen auf Marder geschossen.“ (Christian R.)

I can hardly express how gratified I am to Christian Richter. He introduced me best into the operation of spectrometers as no one else could have done it. His patience with me and his high enthusiasm for our shared projects is greatly acknowledged. It is really a huge pleasure to work with him, to attend conferences together with him and to plant grass in his garden. Dear Christian all the best to you and to your growing family.

Paul: Die Anopheles Mücke produziert den Malaria-Erreger

Christian S.: Nein, der ist in einer anderen Phase als im Menschen.

Kai: Anaphase!

Anna: Was?

The most of the time in the last four years I had the pleasure to be located in the boyz-room. It is impossible for me to describe what made this combination of individuals so special, but it was really a time to remember, therefore thanks a lot boyz: Jürgen Graf, Sergey Ilin, Jan Ferner, Martin Hähnke, Tomislav Argirevic and especially to Christian Schlörb, Kai Schlepckow, S. Kaspar Grimm, who started together with me the endeavour of NMR and with whose I also had a lot of fun during our studies. Also special thanks to Jürgen for the balcony-conversations. To Jan, Christian and Martin also a great thank for the maintenance of the computers.

“This is a kinetic analysis of going from something in between  
to something in between somewhere else.” (Viji)

For the fruitful collaboration, for all discussions and the marvellous RNA samples I am grateful to Prof. Dr. Stefan Pitsch and Philipp Wenter. Although one of them is vanished in Switzerland and the other in Alabama I hope to hear from you for to say: thank you.

Wenn das nicht meine Diplomarbeit wäre, würde ich mich über  
die komischen Ergebnisse prächtig amüsieren. (Senada)

This thesis would have to be completely different if I had not the chance to conduct research on several shared projects within the Schwalbe group. Therefore I am very glad to get the opportunity to share projects with Janina Buck, Vijayalaxmi Manoharan and Senada Nozinovic. It was always a pleasure to work with all of you. Nevertheless special thanks to Janina for having

great fun at spectrometers, lasers and elsewhere, thanks for all the discussions on all imaginable topics.

Ich bin für meinen Pessimismus eingekauft worden. (Jens)  
Gibt's noch Weißwurst? (Jens)

Jens Wöhnert was the one who really impressed me with his deep knowledge about RNA and who infected me with the enthusiasm for the topic. Thanks to him for all shared ideas on what one could do and for all the things we then finally did together. Hopefully there will be the opportunity to have some more beers at the "Lahmer Esel" or elsewhere in future...

„Und dann vergleichen wir, wer den Längsten hat.“ (Jan)

I also want to thank all the former and actual inhabitants of the several other rooms at the AK Schwalbe, who made up a very comfortable working atmosphere not only in the social-room during extensive and discussion-rich coffee-breaks: Karin Abarca, Katrin Ackermann, Aphrodite Anastasiadis-Pool, Neda Bakhtiari, Katja Barthelmes, Holger Berk, Emily Collins, Elke Duchardt, Anna Fröhlich, Christian Gerum, Friederike Heinike, Fabian Hiller, Henry Jonker, Ajit Kaur, Melanie Koschinat, Jitendra Kumar, Anna Lena Lieblein, Daniel Mathieu, Sarah Mensch, Hillary Moberly, Hamid Nasiri, Gerd Nielsen, Jonas Noeske, Stephan Rehm, Anke Reining, Jörg Rinnenthal, Nicole Schmut, Sridhar Sreeramulu, Max Stadler, Karla Werner.

Special thanks go to Julia Wirmer for the fabulous introduction into the laser system (mostly done late night) and to Elke Stirnal who shared the morning coffee with me, although I constantly asked for buffers and more and tried to expand my chaotic lab space towards her very proper one.

"This is impossible – even to my standards" (Viji)

A huge, huge thanks to Anna Paulus and Elena Hartmann for the organisation of our life in the AK Schwalbe. Nothing would have worked out without you- thanks!

„Luft hat kein Deuterium“ (Jan)

The heart of the NMR service is Johannes Gottfried Zimmermann aka "Zim". He is in the best sense of the word special. Dear Zim thanks for all help and also for the kind accommodation in your room – I love it.

„Ich bin eher ein softpulser.“ (Christian Schlörb)

A special thanks to Clemens Glaubitz for taking over the "Zweitgutachten" and also for the collaborative work on the Phosphor-CSA project. Also thanks to Jakob Lopez for all discussions and for the jogging in the morning.

I also want to acknowledge Wolfgang Bermel and Mirko Hennig for fruitful collaborations.



„Der ist so was wie Schrödingers Harald.“ (Martin)

I also want to thank Jacqueline Burré not only for being a great companion during out studies but also for being a real friend. Jacqueline all the best to you and it was always a great pleasure for me to work and to share time with you.

Am Ende möchte ich noch jenen danken, die mir am nächsten sind und auf deren Vertrauen und Unterstützung ich immer und in jeder Situation zurückgreifen kann: meiner Familie. Ich denke nun dankend an Euch alle, da Ihr mich zu dem Menschen habt werden lassen, der ich nun bin: Ihr meine Großeltern Leni und Georg, Rudi und Maria, Du Karl und in ganz besonders Ihr meine lieben Eltern, ohne die weder das Studium noch die Promotion möglich gewesen wären. Liebe Mama, lieber Papa für Alles meinen tiefen Dank, den ich nicht in Worte zu fassen vermag. Meiner Frau gebührt der letzte Satz dieser Danksagung: Julia ich danke Dir für Deine Liebe und Deine Geduld und dafür, dass Du mich immer aufzubauen wusstest, wenn es nötig war.

„Gruß, Kuss, Julius.“ (Jürgen)

