

# **Spectroscopical Investigations to Determine RNA-Ligand Interactions and RNA Dynamics**

Dissertation  
Zur Erlangung des Doktorgrades  
der Naturwissenschaften

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

von  
**Jan Ferner**  
aus New Rochelle, USA

Frankfurt 2009

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

Dekan: Prof. Dr. D. Steinhilber

Gutachter: Prof. Dr. Harald J. Schwalbe  
Prof. Dr. Michael Göbel

Datum der Disputation: 2009

*Für meine Eltern und meine Frau Elke*

This thesis was prepared under the supervision of Prof. Dr. Harald J. Schwalbe between November 2002 and July 2009 at the Institute for Organic Chemistry and Chemical Biology of the Goethe-University Frankfurt am Main.

## TABLE OF CONTENTS

### CHAPTER I

<b>SUMMARY &amp; OVERVIEW</b> .....	1
List of Publications .....	4

### CHAPTER II

<b>RNA BIOMACROMOLECULES</b> .....	7
<b>RNA - History and Functions</b> .....	7
<b>RNA – Structure and Dynamics</b> .....	12
Primary Structure .....	12
Secondary Structure .....	13
Double Helices .....	13
Hairpin Loops .....	16
Internal Loops .....	18
Bulges .....	19
Tertiary Structure .....	20
Dynamics .....	21

**CHAPTER III**

<b>RNA TARGETS &amp; LIGANDS</b> .....	27
<b>The Human Immunodeficiency Virus (HIV)</b> .....	28
Packaging Signal Element $\Psi$ .....	29
Tryptophane-Rich Peptide Ligands of HIV $\Psi$ -RNA .....	31
Transactivation Response Element TAR .....	32
Arginine-Flanked Synthetic Peptide Ligands of HIV TAR-RNA .....	34
<b>The Coxsackievirus B3</b> .....	36
Stemloop D Subdomain of the Cloverleaf RNA .....	36

**CHAPTER IV**

<b>NMR THEORY &amp; METHODOLOGIES</b> .....	39
<b>Ligand Binding Interactions</b> .....	39
<b>Resonance Assignment</b> .....	42
RNA .....	42
Peptide Ligand .....	53
<b>Structure Determination of RNA-Ligand Complexes</b> .....	55
Determination of Structural Restraints .....	55
Hydrogen-Bond Distance Restraints .....	55
NOE Distance Restraints .....	57
Dihedral Restraints .....	58
Orientational Restraints .....	60

## TABLE OF CONTENTS

---

Structure Calculation.....	62
ARIA.....	63
HADDOCK.....	64
<b>Determination of global and local dynamics .....</b>	<b>66</b>
Autocorrelated Relaxation.....	66
Longitudinal Relaxation .....	66
Transverse Relaxation .....	67
Longitudinal Relaxation in a Rotating Frame.....	69
Sources of Spin Relaxation.....	69
Relaxation of a Spin S in an I-S Spin System.....	71
Derivation of Dynamical Information from Autocorrelated Relaxation Rates.....	73
Global Dynamics.....	73
Local Dynamics .....	74

## CHAPTER V

### THE $\Psi$ -RNA HEXAPEPTIDE LIGAND

<b>Research Article:</b> Identification of Peptide Ligands for Target RNA Structures Derived from the HIV-1 Packaging Signal $\Psi$ by Screening Phage-Displayed Peptide Libraries .....	81
<b>Research Article:</b> A Tryptophan-Rich Hexapeptide Inhibits Nucleic Acid Destabilization Chaperoned by the HIV-1 Nucleocapsid Protein.....	89

**Research Article:** Inhibition of HIV-1 by a Peptide Ligand of the  
Genomic RNA Packaging Signal  $\Psi$  ..... 105

## CHAPTER VI

### STRUCTURE OF A TAR RNA-LIGAND COMPLEX

**Research Article:** Tripeptides from Synthetic Amino Acids Block the  
Tat-TAR Association and Slow Down HIV Spread in Cell Cultures ..... 119

**Research Article:** Structures of HIV TAR RNA-Ligand Complexes Reveal  
Higher Binding Stoichiometries ..... 129

## CHAPTER VII

### FAST DYNAMICS IN RNA TETRALOOPS

**Research Article:** Quantitative  $\Gamma$ -HCNCH: Determination of the  
Glycosidic Torsion Angle  $\chi$  in RNA Oligonucleotides from the Analysis of  
CH Dipolar Cross-Correlated Relaxation by Solution NMR Spectroscopy  
..... 147

**Research Article:** NMR and MD Studies of the Temperature-Dependent  
Dynamics of RNA YNMG-Tetraloops ..... 173



**GERMAN SUMMARY**

**SPEKTROSKOPISCHE UNTERSUCHUNGEN ZUR BESTIMMUNG VON RNA-  
LIGAND WECHSELWIRKUNGEN UND RNA DYNAMIKEN..... 203**

**REFERENCES**

..... 209

**CURRICULUM VITAE**

..... 227

**ACKNOWLEDGEMENT**

..... 230

## TABLE OF CONTENTS

---

**CHAPTER I****SUMMARY & OVERVIEW**

This thesis describes the structural characterization of interactions between biological relevant ribonucleic acid biomacromolecules (RNAs) and selected ligands to optimize the methodologies for the design of pharmacological lead compounds. To achieve this aim, not only the structures of the RNA, the ligand and their complexes need to be known, but also information about the inherent dynamics, especially of the target RNA, are necessary. To determine the structure and dynamics of these molecules and their complexes, liquid state nuclear magnetic resonance spectroscopy (NMR) is a suitable and powerful method. The necessity for these investigations arises from the lack of knowledge in RNA-ligand interactions, e.g. for the development of new medicinal drugs targeting crucial RNA sequences.

In the first chapters of this thesis (Chapters II to IV), an introduction into RNA research is given with a focus on RNA structural features (Chapter II), into the interacting molecules, the biology of the specific RNA targets and the further development of their ligands (Chapter III) and into the NMR theory and methodologies used within this thesis (Chapter IV).

Chapter II begins with a description of RNA characteristics and functions, placing the focus on the increasing attention that these biomacromolecules have attracted in recent years due to their diverse biological functionalities. This is followed by a detailed description of general structural features of RNA molecules.

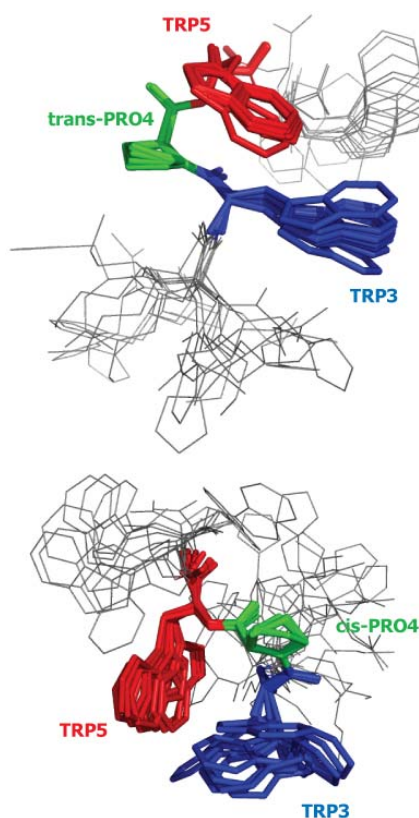
The biological functions of the RNAs investigated in this thesis (Human immunodeficiency virus PSI- and TAR-RNA and Coxsackievirus B3 Stemloop D in the 5'-cloverleaf element), together with their known structural characteristics are introduced in Chapter III. Furthermore, a description of the investigated ligands is given, focusing on the methods how their affinity and specificity were determined.

The introduction is completed in Chapter IV, where the relevant NMR theory and methodologies are explained. First, kinetics and thermodynamics of ligand binding are summarized from an NMR point of view. Subsequently, a detailed description of the resonance assignment procedures for RNAs and peptidic ligands is given. This procedure mainly concentrates on the assignment of the proton resonances, which are essential for the later structure calculation from NMR restraints. The procedure for NMR structure calculation of RNA and its complexes follows with a short introduction into the programs ARIA and HADDOCK. The final part of this chapter explains the relaxation theory and the methodology to extract dynamic information from autocorrelated relaxation rates via the model-free formalism.

In the Chapters V to VII of this thesis, the original publications are included and grouped into three topics. Chapter V comprehends the publications on the investigations of HIV  $\Psi$ -RNA and its hexapeptidic ligand. These three publications<sup>[1-3]</sup> focus on the characterization<sup>[1-3]</sup> of the ligand and its binding properties, its structure and the optimization of its composition aiming to improve its usage for further spectroscopic investigations.

#### Achievements:

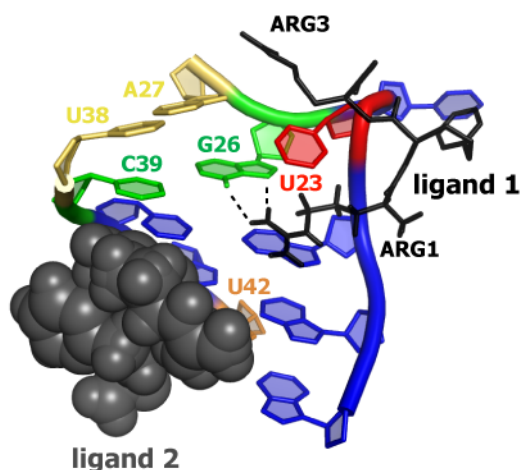
- determination of a consensus peptide sequence (HWWPWW) as ligand for HIV PSI-RNA with a phage-display assay and confirmation of binding by circular dichroism (CD) spectroscopy
- ligand optimization with peptide arrays on membrane supports (SPOT-assay), which resulted in a peptide (HKWPWW) with improved properties for structure determination
- characterization of ligand binding to further nucleic acid targets (TAR-RNA,



PBS-DNA) of the natural  $\Psi$ -ligand, NCp7, by NMR and fluorescence spectroscopy

- structure determination of two peptide conformations present in solution, which both interact with the target RNAs and DNAs
- elucidation of an equilibrium between the *cis*- and *trans*-conformations of the central proline peptidic bond revealing almost equal population of the two conformations

The publications in Chapter VI describe the selection of tripeptide ligands synthesized in the group of Prof. Göbel and the spectroscopical investigation of their complex with HIV TAR-RNA. The combination of chemical synthesis, biochemical screening and determination of the complex structure provided insights into RNA-ligand interactions and ligand optimization.<sup>[4,5]</sup>



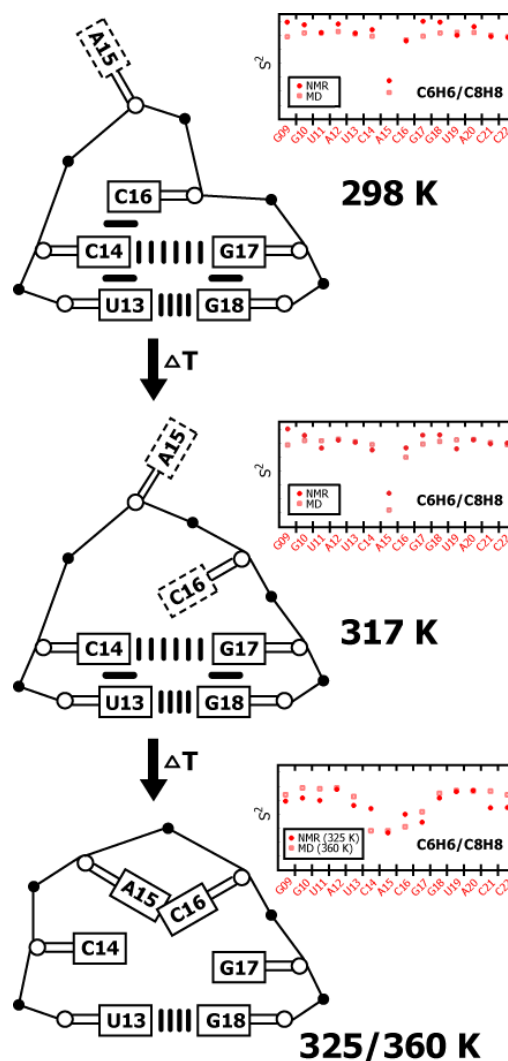
Achievements:

- synthesis and selection of tripeptide ligands for HIV TAR-RNA containing a central non-natural aromatic amino acid (by the Göbel group)
- NMR structure determination of a complex with a 2:1 ligand binding stoichiometry
- characterization of a second weaker binding site on TAR-RNA for arginine-bearing ligands

In the final chapter, Chapter VII, the publications on RNA dynamics are placed. By optimization of the determination and analysis of NMR relaxation data and the combination of the experimental results with computational MD simulations the pico- to nanosecond dynamics could be illustrated.<sup>[6,7]</sup>

## Achievements:

- incorporation of newly derived zero-point motion averaged bond lengths and nucleotide- and conformation-specific  $^{13}\text{C}$  chemical shift anisotropies (CSAs) in the model-free analysis of auto-correlated relaxation rates
- more precise determination of the glycosidic bond angle  $\chi$  with a parameterization of the angular dependency of the cross-correlated relaxation rates between the dipoles  $\text{C1}'\text{H1}'$ - $\text{C8H8}/\text{C6H6}$
- correlation of autocorrelated relaxation rates extracted from NMR experiments and from an MD trajectory of two RNA hairpins
- comparison of the dynamical changes of two structurally similar tetraloops with increasing temperature
- proposal of an atomic picture of RNA melting



## List of Publications

- 1) Identification of Peptide Ligands for Target RNA Structures Derived from the HIV-1 Packaging Signal  $\Psi$  by Screening Phage-Displayed Peptide Libraries. Anette Pustowka, Julia Dietz, Jan Ferner, Michael Baumann, Margot Landersz, Christoph Königs, Harald Schwalbe, and Ursula Dietrich, *ChemBioChem* **2003**, 4, 1093-1097.<sup>[1]</sup>

- 2) A Tryptophan-Rich Hexapeptide Inhibits Nucleic Acid Destabilization Chaperoned by the HIV-1 Nucleocapsid Protein. Chinappan Raja, Jan Ferner, Ursula Dietrich, Sergey Avilov, Damien Ficheux, Jean-Luc Darlix, Hugues de Rocquigny, Harald Schwalbe, and Yves Mély, *Biochemistry* **2006**, 45, 9254-9265.<sup>[2]</sup>
- 3) Inhibition of HIV-1 by a Peptide Ligand of the Genomic RNA Packaging Signal  $\Psi$ . Julia Dietz, Joachim Koch, Ajit Kaur, Chinappan Raja, Stefan Stein, Manuel Grez, Anette Pustowka, Sarah Mensch, Jan Ferner, Lars Möller, Norbert Bannert, Robert Tampé, Gille Divita, Yves Mély, Harald Schwalbe, and Ursula Dietrich, *ChemMedChem* **2008**, 3, 749-755.<sup>[3]</sup>
- 4) Tripeptides from Synthetic Amino Acids Block the Tat-TAR Association and Slow Down HIV Spread in Cell Cultures. Verena Ludwig, Andreas Krebs, Michaela Stoll, Ursula Dietrich, Jan Ferner, Harald Schwalbe, Ute Scheffer, Gerd Dürner, and Michael W. Göbel, *ChemBioChem* **2007**, 8, 1850-1856.<sup>[4]</sup>
- 5) Structures of HIV TAR RNA-Ligand Complexes Reveal Higher Binding Stoichiometries. Jan Ferner, Marcel Suhartono, Sven Breitung, Henry R. Jonker, Mirko Hennig, Jens Wöhnert, Michael W. Göbel, and Harald Schwalbe, *ChemBioChem* **2009**, 10, 1490-1494.<sup>[5]</sup>
- 6) Quantitative Gamma-HCNCH: Determination of the Glycosidic Torsion Angle  $\chi$  in RNA Oligonucleotides from the Analysis of CH Dipolar Cross-Correlated Relaxation by Solution NMR Spectroscopy. Jörg Rinnenthal, Christian Richter, Jan Ferner, Elke Duchardt, and Harald Schwalbe, *J. Biomol. NMR* **2007**, 39, 17-29.<sup>[6]</sup>
- 7) NMR and MD Studies of the Temperature-Dependent Dynamics of RNA YNMG-Tetraloops. Jan Ferner, Alessandra Villa, Elke Duchardt, Elisabeth Widjajakusuma, Jens Wöhnert, Gerhard Stock, and Harald Schwalbe, *Nucleic Acids Res.* **2008**, 36, 1928-1940.<sup>[7]</sup>





## CHAPTER II

## RNA BIOMACROMOLECULES

## RNA - History and Functions

For a long time, the main function of ribonucleic acids (RNA) was assigned to be a translator of the coding genomic deoxyribonucleic acids (DNA) into functional proteins. Over the past decade, however, RNA has been subject to increasing attention due to its remarkable functional diversity in biological systems<sup>[8]</sup> as well as its usage in synthetical approaches, e.g. for catalysis<sup>[9, 10]</sup>. Following the discovery of the structure<sup>[11-13]</sup> and mechanism of the ribosome<sup>[14]</sup> and of the central role of RNA as its catalytic moiety, more and more investigations have been focused on further biological roles of RNA.

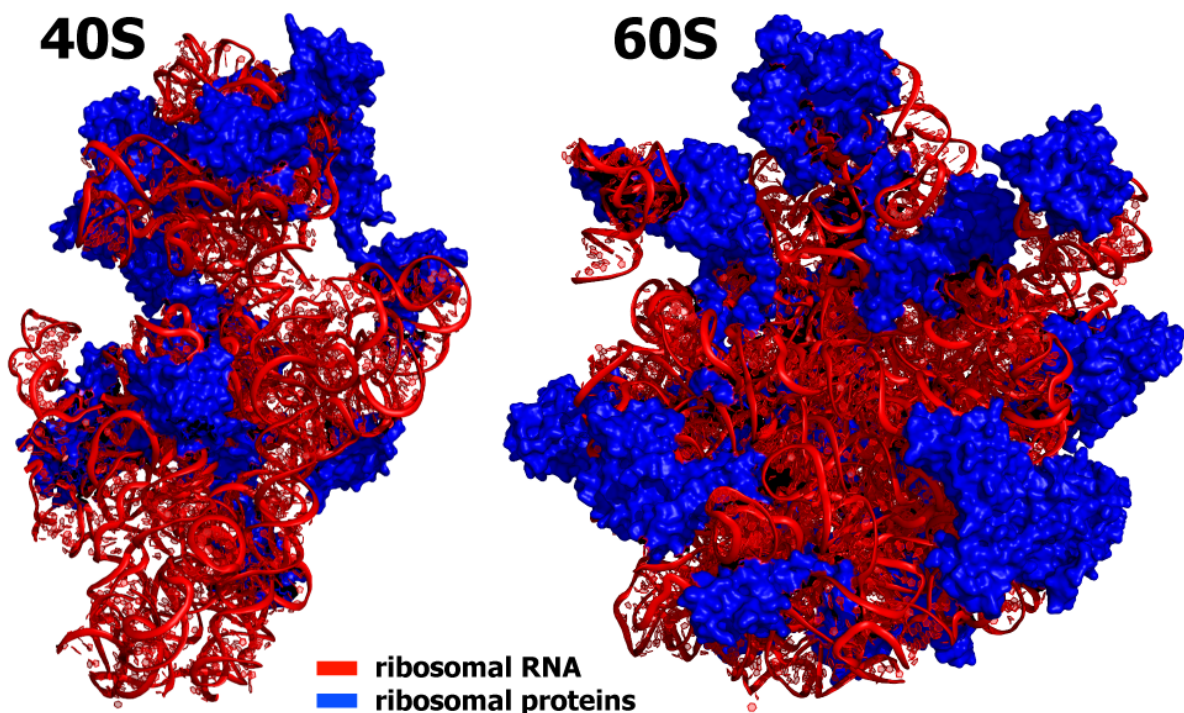


Figure 1: Structures of the ribosomal subunits 40S (left) and 60S (right) from yeast.<sup>[15]</sup> The ribosomal RNAs are colored red, the ribosomal proteins blue.

The ribosome is a multimolecular system (Figure 1) in all living cells which constitutes the center of the translational apparatus. Herein, ribonucleic acids

(rRNA) play the main catalytic role for the translation of messenger RNA (mRNA) into an amino acid sequence using transfer RNAs (tRNA) which transfer a specific active amino acid to the growing polypeptide chain. A second significant multimolecular system with a catalytic RNA is the spliceosome which prepares pre-mRNA for its translation.

These findings support the idea of an „RNA-world“<sup>[16]</sup>, which could have formed the precursor of today's biological setup of DNA, RNA and proteins. Retroviruses and the ribosome are possible relicts of this world. The retroviruses, e.g. the human immunodeficiency virus (HIV), store their genetic information as RNA rather than DNA and nowadays use the more efficient expression system of infected cells for their proliferation. In the foretime, the retroviruses possibly were self-replicating organisms.

Several projects were started at the end of the last century to map the genome of many different species, e.g. the Human Genome Project.<sup>[17]</sup> A surprising outcome of these projects was the small percentage of protein encoding genes in eukaryotic genomes. Furthermore, the amount of non-coding parts of the genome is strongly correlated with the complexity of the organism. While prokaryotic cells contain between 6 and 24 % of non-coding DNA, the percentage in Homo sapiens is at about 97 % (Figure 2).<sup>[18, 19]</sup> This correlation raised the question of the function of these non-coding parts and led to the finding that nucleic acids possess major diverse regulatory functions at several steps of the transcriptional and translational process. Most of the genes contain, for example, introns, which are removed by the spliceosome. By regulating the spliceosome apparatus, different mRNAs are produced from the same pre-mRNA, which are then translated into different protein isoforms increasing protein diversity. The excised introns can further act as regulatory factors, e.g. as microRNA<sup>[20]</sup>, as described in the following paragraph.

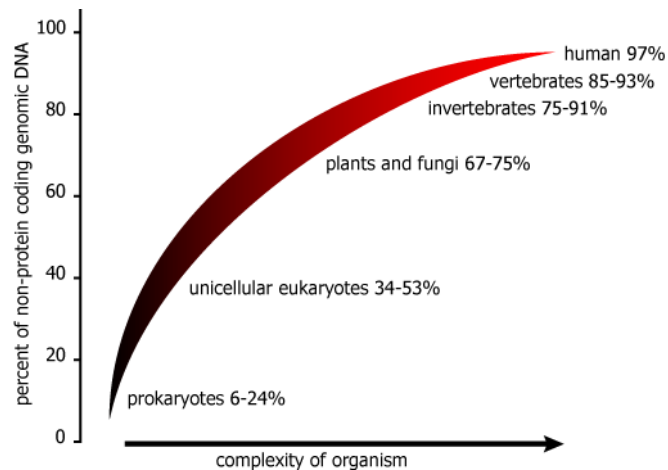


Figure 2: Percentage of non-coding genomic DNA versus the complexity of organism.<sup>[18]</sup>

This important regulatory function is summarized under the term RNA interference (RNAi).<sup>[21]</sup> While microRNAs (miRNA) are the biological tools for this mechanism<sup>[22]</sup>, biotechnology and science use small interfering RNA strands (siRNA)<sup>[23]</sup> to interfere with selected RNA targets. Both have a length of about 20-25 nucleotides and interact with the genomic DNA or mRNAs to suppress transcription or translation, respectively. This suppression is achieved by binding to a totally (siRNA) or partially (miRNA) complementary strand of DNA or RNA, followed by DNA/histone methylation, mRNA degradation or blockade of translation. The selectivity and reliability of this interference makes it attractive to investigate their potential for scientific and medical application in cell cultures and living organisms.<sup>[24]</sup>

Further attractive regulatory functions lie within the 5' untranslated region (5'-UTR) of mRNAs. Since 2002, riboswitches<sup>[25-28]</sup> are in the focus of the RNA community. They are found in eubacteria, archaea, fungi and plants and exert their function by binding of a small, selected molecule which initiates a structural rearrangement of the RNA, the „switching“. Switching allows the ligand-dependent control of mRNA transcription or translation. An example of these small molecules is the nucleobase guanine which binds to the guanine sensing riboswitch in *Bacillus subtilis* with affinities comparable to those of enzymes binding their ligand.<sup>[29, 30]</sup> This binding of a small ligand to an RNA

sequence with such high affinity had been unprecedented in biological systems so far. Other nucleobases<sup>[31]</sup>, nucleotides<sup>[32]</sup>, amino acids<sup>[33, 34]</sup> and vitamins<sup>[28]</sup> are further examples of riboswitch ligands. They all have in common to bind to riboswitch sequences located in the 5'-UTRs of mRNAs coding for proteins involved in their own biosynthesis.

Other sequences in the 5'-UTR of mRNAs bind biomacromolecules (e.g. proteins or other RNA sequences) to fulfill their function. The retrovirus HIV (human immunodeficiency virus) is a representative organism with well known RNA sequences which possesses different functionalities. The  $\Psi$ -RNA (packaging signal element) is essential for the dimerization of two viral genomes prior to their encapsidation, as well as for the encapsidation itself.<sup>[35]</sup> A second interesting non-coding sequence in HIV is the TAR RNA (trans-activation response element) which binds the protein Tat (transcriptional transactivator) to regulate the activity of the viral promoter controlling the viral replication.<sup>[36]</sup> In another viral organism, the enterovirus Coxsackievirus B3, the cloverleaf-like RNA element contains several protein binding sites to form a ribonucleoprotein complex which plays a crucial role in the assembly of the replication initiation complexes.<sup>[37, 38]</sup> All three elements are introduced more precisely in CHAPTER III.

These diverse functionalities of RNA were unexpected given the small chemical diversity of its nucleotide building blocks. While the phosphodiester backbone and the sugar moieties do not show a variation at all, the four nucleobases contain the main differences in the nucleotide configuration. However, even these moieties exhibit very similar properties with regard to their ability to act as hydrogen bond donors and acceptors as well as their tendency to stack due to their aromatic properties. Therefore, nucleic acid structures are constituted to a large extent of helical parts mainly stabilized by Watson-Crick base pairing. Non-Watson-Crick base pairs and the sequences connecting the double helices, e.g. loops and bulges, are the interesting regions for interactions with ligands and receptors. They are often

dynamic and partly unstructured which enables them to adopt different binding structures in varying complexes.

To understand these biological relevant functions of RNA and to influence RNA-ligand interactions, detailed knowledge of the ligand induced RNA structural changes, the specific interactions of RNA with its ligands and last but not least of the intrinsic RNA dynamics is necessary. The aim of this doctoral thesis is to contribute to this field of RNA-ligand interactions through investigations of selected RNA targets and ligands with spectroscopic methods, especially NMR spectroscopy in solution.

## RNA – Structure and Dynamics

The structure of RNA is ordered hierarchically into primary, secondary and tertiary structure elements, analogous to proteins. In the following paragraphs the basic definitions and selected elements are explained. The chapter ends with an introduction of motional dynamics occurring in RNA molecules since the RNA dynamics are essential for the diverse functionality.

### Primary Structure

The four nucleotides adenosine, guanosine, uridine and cytidine (Figure 3) constitute the RNA building blocks. The primary structure of RNA is defined as the sequence of the nucleotides starting from the 5'-end to the 3'-end.

The nucleotides are connected through their ribose sugar moieties via a phosphodiester group, which links the 5'- and the 3'-ends.<sup>[39]</sup> Attached to the glycosidic carbon C1' are the specific nucleobases, the two purines adenine and guanine and the two pyrimidines, uracil and cytosine. Energetically, two distinct orientations of the nucleobases are favorable along the glycosidic angle  $\chi$ , either a *syn* orientation (50°-80°) or an *anti* orientation (180°-280°) with respect to the ribose moiety. The five-membered ribose moiety is constrained to a non-planar conformation, envelope or twist. Most favored are the C2'-endo and C3'-endo conformations. While deoxyribonucleotides prefer the C2'-endo conformation, also known as the *south*-conformation, the 2'-hydroxyl group in ribonucleotides causes a stereoelectronic effect which results in a preference of the C3'-endo or *north*-conformation. Since the energetic difference between these two conformations is rather small in ribonucleotides, the ribose can adopt a C2'-endo conformation in some sequential arrangements. The further degrees of freedom are along the ribose-phosphodiester backbone. Their torsion angles are strongly restricted to specific regions depending on the conformation of the sugar (Table 1).

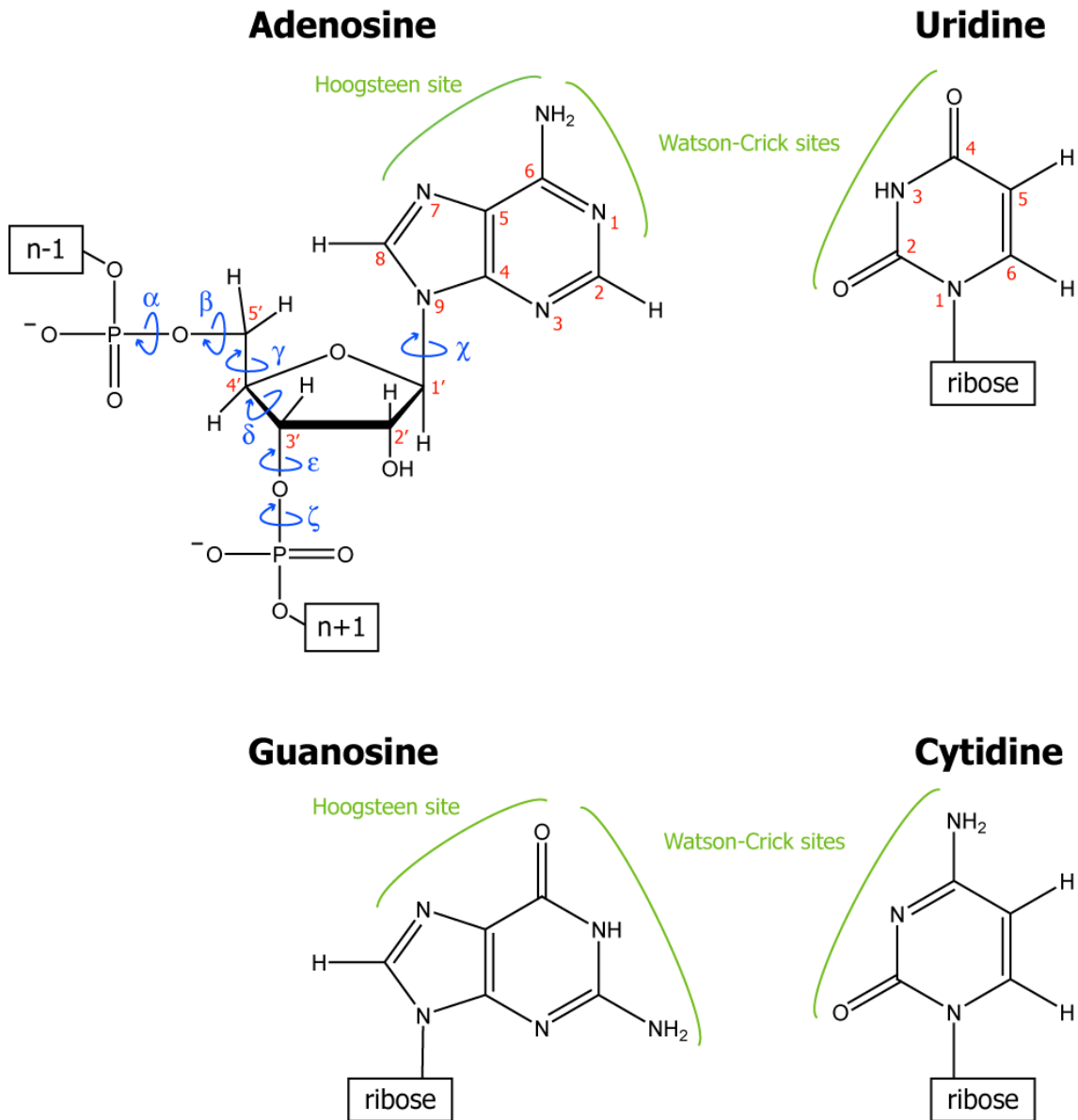


Figure 3: Configuration, numbering (red) and angles (blue) of a ribonucleotide and the four nucleobases in RNA: adenosine, uridine, guanosine and cytidine. The nucleobase sites for hydrogen bonding patterns, Watson-Crick or Hoogsteen, are drawn in (green).

## Secondary Structure

### Double Helices

The most abundant structure composition of nucleic acids is the double helix where hydrogen bonds between opposite nucleobases connect the two strands, forming a base pair (Figure 4).

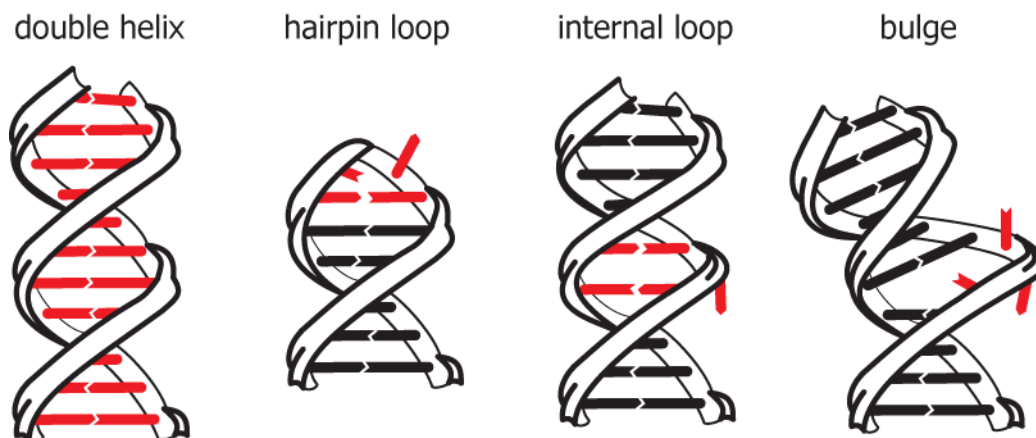


Figure 4: Schemes of RNA secondary structure elements: double helix, hairpin loop, internal loop and bulge. The nucleotides which are part of the corresponding secondary structure element are colored in red.

An RNA stem which is characterized by the following canonical features, is described as A-form RNA: (i) the opposite nucleobases of two antiparallel strands form Watson-Crick A-U and G-C base pairs (Figure 5 A); (ii) the ribose moieties are in a C3'-endo sugar pucker conformation, avoiding sterical hindrances of the 2'-hydroxyl-group and enabling it to hydrogen-bond to the O4'-acceptor of the 3'-sequential nucleotide, as well as to form a water-bridged hydrogen-bond to the phosphodiester backbone<sup>[40]</sup>; (iii) the nucleobase moieties are in an *anti*-conformation ( $\chi \approx 200^\circ$ ), with the bulk of the base pointing away from the ribose<sup>[39]</sup>; (iv) in Table 1 the backbone torsion angles in A-form RNA are defined. These local settings result in a phosphorous-phosphorous distance of about 6 Å, a rise of 11 nucleotides per turn and a pitch between the stacking base pairs of 3 to 4 Å. The base pairs are slightly tilted from a perpendicular arrangement with respect to the helix axis and the antiparallel orientation leaves gaps between the two backbones, the major and the minor groove. The major groove exposes the edges of the nucleobases and is deep and narrow. In contrast, the minor groove is wide and shallow disclosing the ribose moieties.



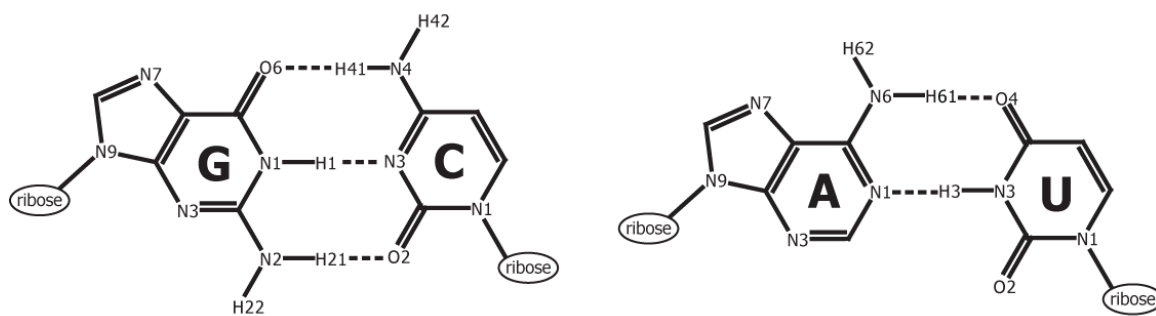
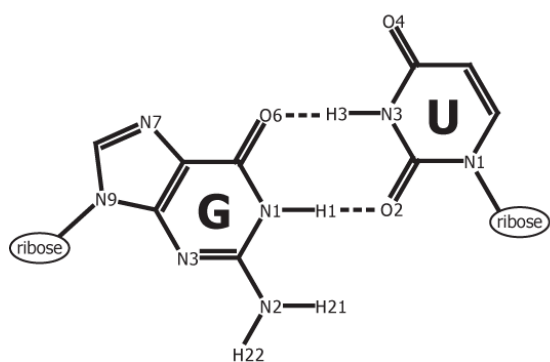
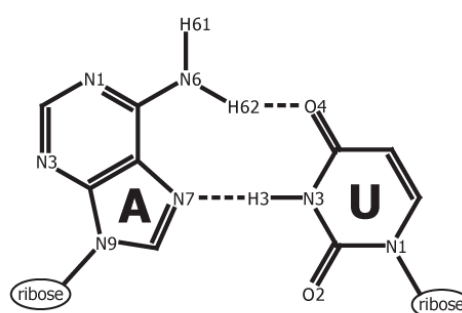
**(A) Watson-Crick base pairs****(B) G-U wobble base pair****(C) A-U Hoogsteen base pair**

Figure 5: Selected base pairs found in RNA biomacromolecules: (A) the canonical G-C and A-U Watson-Crick base pairs, (B) a G-U wobble base pair and (C) an A-U Hoogsteen base pair. The hydrogen bonds are indicated by dashed lines.

Table 1: Definition of the backbone angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\varepsilon$  and  $\zeta$  and the glycosidic bond angle  $\chi$ . The denoted angles account for an A-form RNA-conformation.<sup>[39]</sup>

angle	angle definition	A-RNA
$\alpha$	$O3'_{n-1} - P_n - O5'_n - C5'_n$	-68
$\beta$	$P_n - O5'_n - C5'_n - C4'_n$	178
$\gamma$	$O5'_n - C5'_n - C4'_n - C3'_n$	54
$\delta$	$C5'_n - C4'_n - C3'_n - O3'_n$	82
$\varepsilon$	$C4'_n - C3'_n - O3'_n - P_{n+1}$	-153
$\zeta$	$C3'_n - O3'_n - P_{n+1} - O5'_{n+1}$	-71
$\chi$	$C2'_n - C1'_n - N1_n - C2_n$ (pyrimidines) $C2'_n - C1'_n - N9_n - C4_n$ (purines)	-158

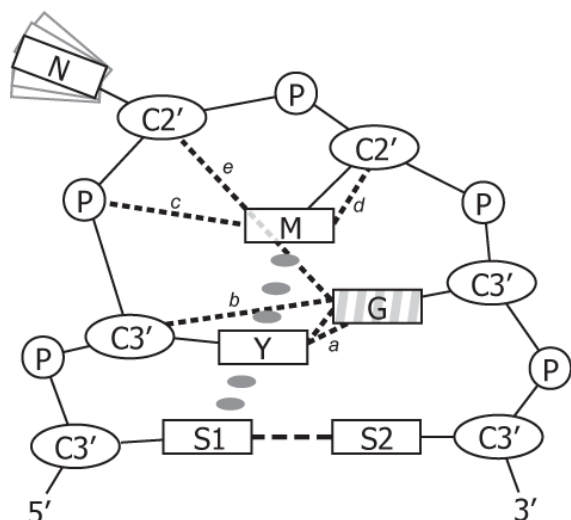
### Hairpin Loops

Loops allow a single chain of RNA to form a hairpin where the subsequent nucleotides can form a double helical stem with the preceding nucleotides. These loop stretches need to consist of at least two nucleotides to accomplish the reversion of the strand orientation. Structure and stability of hairpin loops can differ significantly. For some loops, no definite structure can be determined while others are well structured due to a high number of stabilizing interactions.

Tetraloops are a good example for the different types of structure and degrees of stability. Structurally well defined tetraloops are grouped into families.<sup>[41]</sup> The most abundant RNA tetraloop families detected in all biological organisms are the YNMG- and the GNRA-tetraloops.

In the YNMG-motif, the Y stands for a pyrimidine nucleotide, uridine or cytidine. At the second position of the tetraloop (N) any of the four nucleotides can be incorporated while the third position (M) is occupied by either an adenosine or a cytidine. There are several stabilizing interactions within the loop region (Figure 6).<sup>[41]</sup> Only the nucleobase of the second loop residue is not involved in any stabilizing interactions neither intra- nor interresidual, which explains the lack of selectivity at this position. The sugar moieties of the central two residues possess major populations of C2'-endo conformation which enables the loop to reverse the direction of the backbone. The UUCG tetraloop, closed by a Watson-Crick C-G base pair constitutes the most stable known tetraloop.<sup>[42]</sup> Due to its high stability it often appears as folding nucleation point in functional biological RNAs but has no known ligands.

## YNMG-motif



## GNRA-motif

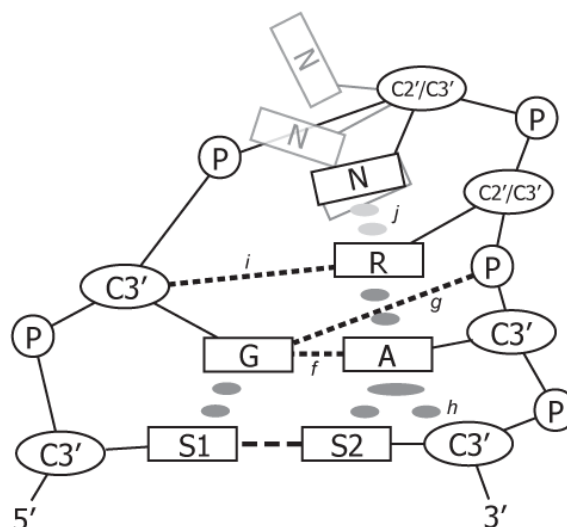


Figure 6: Schematic representation of the tetraloop motifs YNMG (left) and GNRA (right). The sugar moieties are displayed as ellipses with their sugar pucker (C2' or C3') quoted, the nucleobases as rectangles, phosphodiester groups as circles labeled P, stacking interactions indicated by gray ellipsoids, and hydrogen bonds with dashed lines. In the YNMG-motif, the nucleotides Y and G form a base pair which stacks onto the pyrimidine nucleobase on the adjacent stem base pair S1-S2. The guanosine imino H1 and amino H21 protons clasp the carbonyl oxygen O2 of the pyrimidine Y (a). The guanosine nucleobase is arranged in a syn-conformation (striped rectangle). A strong hydrogen bond between the hydroxyl-group (2'OH) of Y and the guanosine carbonyl oxygen O6 stabilizes this base pair additionally (b). On top of Y, the nucleobase of the third loop residue (M) is stacked. This moiety is further stabilized by 2 hydrogen bonds: One between the amino group (cytidine N4 or adenosine N6) in the nucleobase and a phosphodiester backbone oxygen between the first and second nucleotide (c) and the second is an intraresidual hydrogen bond of the ribose hydroxyl to the carbonyl oxygen O2 in case of a cytidine in position 3 or the nitrogen N3 of an adenosine (d). The nucleotide at position 2 only possesses a hydrogen bond of its ribose hydroxyl proton to the guanosine oxygen O6 (e).<sup>[41]</sup> The GNRA-motif also forms a base pair between its first and last residue. This unusual G-A base pair is stabilized by hydrogen bonds between adenine amino N6 and guanine N3, between guanine amino N2 and adenine N7 (f) as well as the oxygen  $O_{\text{pro-R}}$  of the phosphate group between loop residues 3 and 4 (g). Both nucleobases stack on the adjacent stem base pair and the adenosine nucleobase in addition causes a large ring current effect on the H1' proton of S2 (h). The third nucleobase stacks on the terminal pyrimidine and its N7 accepts a hydrogen bond from the hydroxyl group of the first loop nucleotide (i). The type of the nucleobase in the second position defines the further loop conformation. Purines favor to stack on the succeeding purines (j), while pyrimidines are more flexible.<sup>[43-45]</sup>

GNRA loops constitute the second large RNA tetraloop family (Figure 6). The nucleotide in the second position is unspecific (N) but a purine nucleotide is located in the third position (R).<sup>[44]</sup> Among the members of this family the backbone orientation does not vary strongly (0.6 Å to 0.9 Å) while equivalent nucleobases can dislocate by up to 48° from co-planarity.<sup>[43]</sup> The last three nucleobases stack on each other and are nearly parallel to the closing Watson-Crick base pair.<sup>[45]</sup> For most tetraloops in the GNRA-family the sugar moieties of all loop residues are in a C3'-endo conformation. However, about 20 % of the investigated loops possess a C2'-endo sugar pucker for the central two nucleotides. This structural arrangement with different possible conformations allows a high specificity for possible ligands since all three stacked loop nucleobases are accessible with their Watson-Crick sites and possess large motional freedom for a selective recognition.<sup>[43]</sup>

### Internal Loops

A double helical stem can be disrupted in its canonical structure by an internal loop when the nucleotides do not form Watson-Crick or G-U/U-G wobble base pairs (Figure 5). This secondary element can either be totally unstructured leading to unrelated motions of the adjacent double helices or the nucleotides can form non-canonical base pairs sustaining the double helix<sup>[46, 47]</sup> but changing the local structure and its dynamics.<sup>[48, 49]</sup> All types of non-canonical base pairs can appear in internal loops providing different surfaces in the major and minor groove, as well as differences in stability and dynamics. Therefore these RNA elements are targets for interactions with other biomacromolecules.<sup>[50, 51]</sup>

A distinction is drawn between a symmetric and an asymmetric internal loop depending on the number of nucleotides in both strands. If the stretches are equal in number (symmetric internal loop), each nucleotide has a possible binding partner to form a base pair. In the case of an asymmetric internal loop at least one nucleotide has no binding partner and can either stack unpaired within the helix or extrude pointing towards the solvent.

This conformational difference may depend on further binding interactions as in the case of the internal loop J6a/6b in the group I intron ribozyme. This internal loop is a GAAA tetraloop receptor (Figure 7 C) where loop nucleotides change their stacking conformation between a state free in solution<sup>[52]</sup> and a state in the presence of the tetraloop.<sup>[53]</sup>

### **Bulges**

A bulge is defined as a non-base-paired nucleotide sequence in one of the strands which connects two helical segments of ribonucleotides. Therefore it is closely related to an asymmetric internal loop. The bulge permits flexibility and a curvature of the adjacent domains. A grouping into sequence families is not possible due to the high variability in length and nucleotide composition. Single nucleotide bulges can adopt two structural arrangements: while pyrimidines tend to orient extrahelical, the purine nucleotides exhibit a high bias to stack within the double helical parts. This finding is in line with the known stacking properties of the nucleobases. The stacking propensity decreases in the following order: G > A > C >> U.<sup>[54, 55]</sup> Adding more nucleotides to the bulge leads to a bending of the stems. The bending is not only dependent on the sequence but also divalent ions such as Mg<sup>2+</sup> influence the structure by compensating the repulsive forces along the negatively charged phosphodiester backbone. Concomitant with this bending is also an increased flexibility of the free RNA. The angle between two helical parts connected by a trinucleotide bulge can exceed 90° as determined for the TAR RNA hairpin.<sup>[56]</sup> The freedom of the bulge nucleotides enables them to interact with the adjacent stems to form specific structural features. In this case, the TAR RNA hairpin is also a good example for this: In complex with the small ligand argininamide the nucleobase of the first bulge residue (U23) was found to form a base triple with an A-U Watson-Crick base pair in the subsequent double helix.<sup>[57, 58]</sup>

## Tertiary Structure

RNA secondary structure elements can further orient each other to form tertiary structure elements. The function for these arrangements can be to stabilize the global structure, protect specific vulnerable parts of the RNA or to form specific binding sites for small ligands or other biomacromolecules. Tertiary structure elements can be constituted either of arrangements of two or more double helical stems, of two unpaired sequences or of a single stranded stretch interacting with a double helix.<sup>[59]</sup>

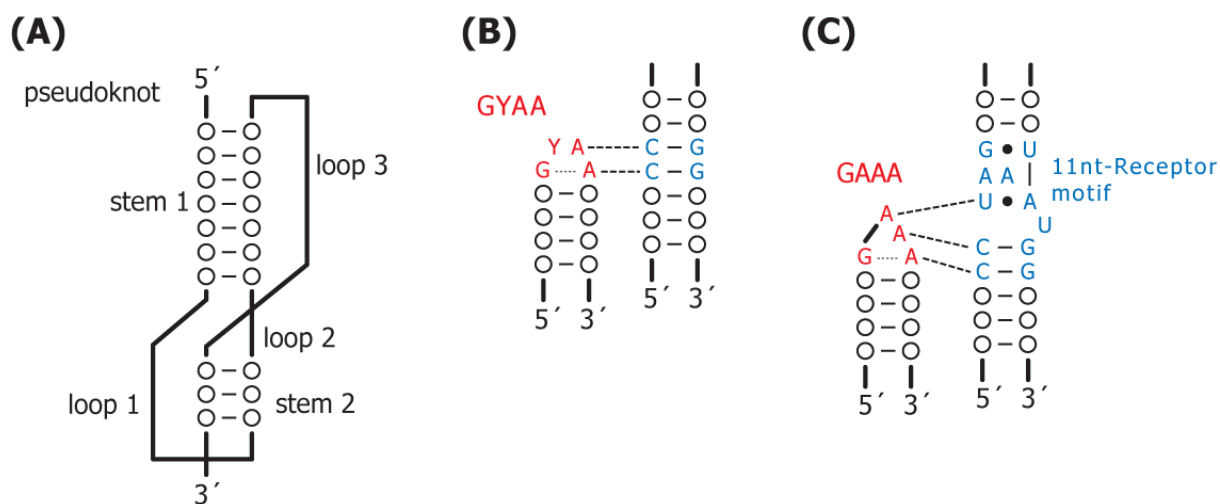


Figure 7: Schematic representations of (A) a pseudoknot with co-axial stacking and the loops L1 and L3 crossing the deep and shallow grooves; (B and C) two tertiary motifs involving GNRA-tetraloops (residues colored in red). In the first, the two adenines form base triples in the shallow groove of a helix (B). In the second, a purine stretch of three adenine binds to a 11-nt-receptor motif (colored blue) (C).<sup>[53]</sup>

The orientation of two stems can lead to an elongation of the double helix mediated by stacking interactions (coaxial stacking). The connection between these stems can either be a bulge or a third stem as in a three-way junction. Helical stems can also bind each other by fitting their grooves, major or minor, into each other, e.g. in the A-minor motif.<sup>[60]</sup> An interesting tertiary structure element with a special setup of double helical stems is the pseudoknot (Figure 7 A). In the simplest pseudoknot loop nucleotides of a hairpin form a second double helix with nucleotides downstream of the stem.<sup>[61]</sup> The

complexity can be increased by incorporation of further stems within the loops as in the Diels-Alder ribozyme.<sup>[62]</sup>

Two hairpins can bind each other by formation of base pairs between its loop nucleotides (kissing loops). These can, for example, be standard Watson-Crick base pairs which are incorporated in more complex base quadruples with several hydrogen bonds between nucleotides of the same loop and to nucleobases in the second loop.<sup>[29]</sup> In the guanine-sensing riboswitch, e.g., these loop-loop kissing interactions help to preorganize the ligand binding site<sup>[63]</sup>, which is located in the bulge connecting the two hairpins and is 6-7 base pairs apart from the loops. Binding of the ligand then further stabilizes the entire structure displaying a long range effect on the tertiary loop-loop interactions.

Single as well as stretches of unpaired nucleotides can bind to double helical stems forming triple-stranded helices as in the binding of GNRA-tetraloops to receptor motifs (Figure 7 B and C). A further example is the base triple in HIV TAR, described previously, which is the smallest type of this tertiary structure element.<sup>[57, 58]</sup> The third nucleotide in this base triple interacts with the Hoogsteen site of the purine nucleobase, where the purine N7 and guanine O6 act as hydrogen bond acceptor and the adenine N6 as hydrogen bond donor.

Higher orders of structural elements are not specified for RNA biomacromolecules but they exist similarly to the quaternary structure in protein complexes. The most prominent example of an RNA quaternary structure is the ribosome (Figure 1) which consists of three RNA strands and several ribosomal proteins.

## Dynamics

Dynamics are a fundamental characteristic of RNA and therefore are essential for the functional diversity of RNA. All bonds, angles and torsion angles are

subject to vibrations, librations and fluctuations on the picosecond time scale or faster but to a different extent (Figure 8). In the case of nucleotides, four partly unrelated motional sites can be separated: the glycosidic angle, the ribose sugar pucker, and the 5'- and 3'-connections to the phosphodiester backbone.

Due to the inherent planarity of the nucleobase the glycosidic angle  $\chi$  (O4' – C1' – N1/9 – C2/4) is the only degree of freedom for conformational motions of the base moiety. An interconversion between the two energetically favored conformations, *syn* and *anti*, takes place on a time scale of nano- to microseconds (Figure 8). But large sterical clashes mostly prevent an interconversion in structured RNA segments, especially of the purine nucleobases. In addition, the different charge distributions of the two heteroaromatic sides afford a large structural change of the environment to allow base flipping. For pyrimidines a base flipping exchanges the hydrophobic C5/C6 side with the strongly charged C2/N3 side which possesses the ability to form hydrogen bonds as donor (uridine N3H3) or acceptor (cytidine N3 and cytidine/uridine O2). Therefore, most dynamics are angular fluctuations around the two favored conformations.

The repuckering between C2'-endo and C3'-endo display the largest conformational rearrangements in the ribose moieties in RNA, since the interconversion between these two conformations requires only small activation energies. The process of interconversion takes place on a timescale of pico- to nanoseconds (Figure 8) and mechanistically leads over an O4'-endo intermediate stage which is energetically more favorable than the O4'-exo conformation.

The most degrees of freedom within the nucleotides lie along the ribose-phosphodiester backbone. Especially the torsion angles  $\alpha$  (O3'(n-1) – P – O5' – C5') and  $\zeta$  (C3' – O3' – P – O5'(n+1)) display the highest distribution in RNA biomacromolecules. But the strong electrostatic charges restrict the angles to distinct regions in structured RNA segments. The other angles are additional restricted to specific regions depending on the conformation of the ribose.



Overall the hydrogen bonding network with cations and solvent molecules prevent large motions of the backbone.

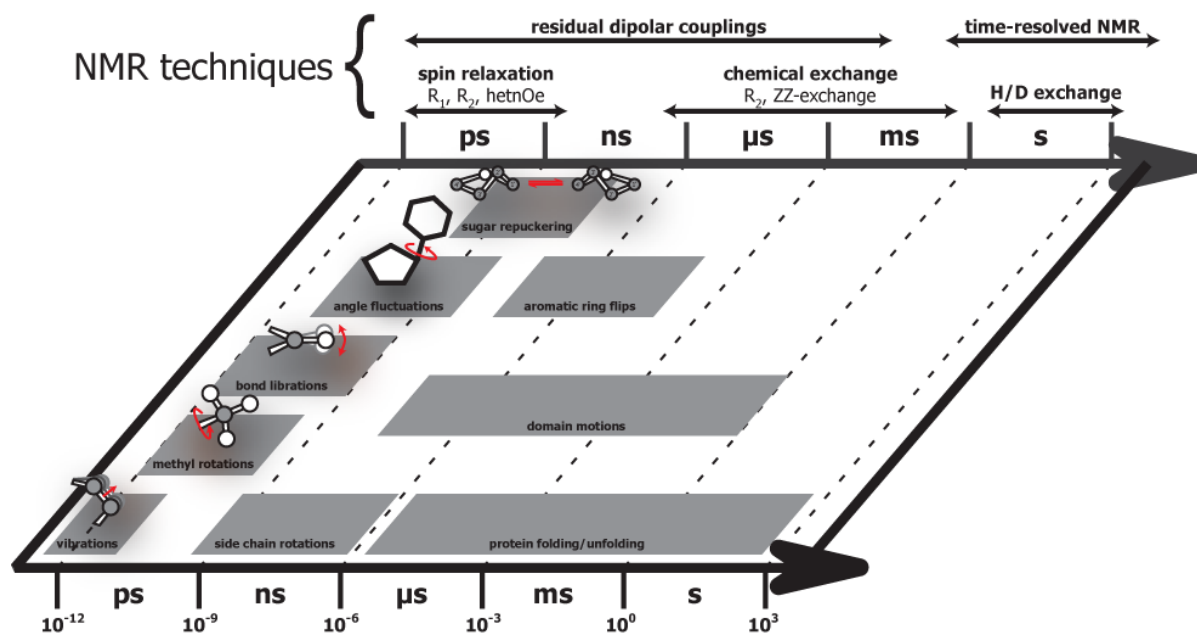


Figure 8: Time scales of molecular motions from fast local dynamics to global rearrangements are displayed in the plane. Above the plane, NMR parameters, which are sensitive for internal dynamics, are given together with the window of sensitivity.

These local flexibility sites can be translated into dynamics within secondary structure elements. Among these the canonical double helices are expected to display only minor dynamics since the base pairing and stacking forces rigidify the nucleobases and the water-mediated hydrogen bonding network and coordinated metal cations stabilize the 3'-endo ribose conformation and the phosphodiester backbone. However, recent detailed investigations of canonical DNA double helices revealed that cytidine riboses can display increased mobility with an interchange of the ribose conformation between a C2'-endo and a C3'-endo sugar pucker. This also increases the mobility of the cytidine nucleobase which then can, for example, flip out and bind to an enzymatic protein to become methylated.<sup>[64, 65]</sup>

The other secondary structure elements with unpaired nucleotides, hairpin loops, internal loops and bulges display much more dynamics and

conformational flexibility which enables these sequences to adapt to ligands and receptors to fulfill a function. As described above, the tetraloop YNMG is highly structured and stabilized (Figure 6), only allowing the motion of the second loop nucleobase moiety.<sup>[66]</sup> The dynamics measured in this moiety possible arise from angular fluctuations around the glycosidic angle. This motional restriction may explain, why the most stable tetraloop with the sequence UUCG closed by a C-G canonical base pair functions as a structural stabilisator but has none known binding partner in nature.

In contrast, the loops in the GNRA-family display a much wider region of flexibility, although they are also highly structured and stabilized by several hydrogen bonds and stacking interactions (Figure 6).<sup>[43, 44]</sup> The distinct influence of the nucleotide type at the second loop position is a significant observation.<sup>[45]</sup> While purines extend the stacking on the 3'-end of the loop, pyrimidines possess a higher degree in flexibility following the decreased stacking properties of cytidine and uridine.<sup>[55]</sup> In addition, the central two ribose moieties are variable between the two most favorable sugar puckers, C2'- and C3'-endo, and a correlation with the unstacking of the second loop nucleobase can be observed.<sup>[67]</sup> This conformational flexibility enables the GNRA tetraloops to adopt to specific receptors, as e.g. in the case of the GAAA tetraloop receptor in the group I intron ribozyme (Figure 7 C).<sup>[53]</sup>

Since bulges always constitute of unpaired nucleotides, these secondary structure elements are often flexible hinges between double helices. A highly investigated bulge is the tri- or dinucleotide bulge in the HIV transactivation response element TAR (Figure 10). Free in solution no defined structure of the bulge could be determined and a flexible bending of the stems up to an angle of 90° was extracted.<sup>[56]</sup>

All dynamical RNA segments have in common that they rigidify significantly when complexed to a ligand or receptor. In the case of TAR, Al-Hashimi and co-workers could prove that the stable TAR-ligand complexes are also possible conformations in free TAR concluding that the ligands stabilize existing conformations rather than inducing new ones.<sup>[68]</sup> This „conformational capture“ adds a further RNA-ligand binding mechanism to the

widely accepted feature claimed “induced fit”.<sup>[69]</sup> In the “induced fit” mechanism a stable RNA motif reorganizes its structure to a different conformation to bind the ligand or receptor. To distinguish between these two binding mechanism a precise knowledge of the structure and the dynamics in the free and the bound state is necessary.



**CHAPTER III****RNA TARGETS & LIGANDS**

To understand RNA-ligand interactions and RNA dynamics, specific RNA targets need to be selected which possess defined functions and natural ligands. With knowledge of these characteristics, ligands can be designed or selected from libraries by different assays.

For long, it was assumed that RNA ligands need to be biomacromolecules with several interaction sites to bind to a specific RNA sequence since the differences between the four nucleotides are not significant. Therefore only protein domains and complementary RNA sequences were expected as binding partners. Different protein elements could be identified as RNA binding motifs, e.g. the zinc-finger motif. These binding motifs often appear more than once in the RNA binding protein to increase affinity and selectivity. In the case of the NCp7-domain of the HIV Gag-polyprotein, two zinc-fingers bind to the HIV packaging signal  $\Psi$ -RNA. This necessity for multiple binding motifs prevented scientists from expecting small molecules as RNA ligands. The RNA field was surprised in the past decade, as RNA aptamer domains were located in biological systems that bind small ligands like single nucleobases or amino acids to fulfill their function. These RNA domains were entitled "riboswitches"<sup>[25-28]</sup>, since they changed their structure globally upon binding of the ligand, switching "on" or "off" a subsequent process.

The first attempts to a synthetic design of small molecules binding RNA were adapted from the knowledge of DNA-ligand interactions. Thus, intercalators and groove binders were the first compounds under investigation.<sup>[70]</sup> But since the regular undistorted A-form helices of RNA do not represent the functionally interesting target regions<sup>[71]</sup>, these compounds are not attractive as starting points for therapeutic drugs. Therefore, the development of RNA-targeting therapeutics is conceptionally closer to protein-ligand drug

development, because the binding sites and enzymatical regions of RNA biomolecules are structurally diverse. An additional aggravating circumstance is the high flexibility of many of these functional regions. It follows that a structural investigation of such a flexible target in its free state often provides no guideline for a rational ligand design. Also the high negative charge, inherent in each RNA sequence, hinders the ligand screening, since positively charged ligands bind strongly but maybe unspecific.

Therefore, screening approaches of combinatorial chemistry libraries, biochemical *in vitro* assays or of other biotechnological methods constitute a sensible starting point for RNA drug design. After the introduction of the RNA targets, the selected methods for ligand candidate finding and lead optimization of each RNA target are described. In case of the first target, the HIV packaging signal  $\Psi$ , a phage-display assay of short peptides provided a hexapeptide as lead sequence.<sup>[1]</sup> The second approach targeted the HIV transactivation response element TAR with a combinatorial screening of synthetic tripeptides using the FRET-technology (Fluorescence Resonance Energy Transfer) for detection.<sup>[72]</sup>

## The Human Immunodeficiency Virus (HIV)

The human immunodeficiency virus (HIV) has attracted attention throughout the past decades as inductor of the incurable pandemic disease AIDS (*Acquired Immune Deficiency Syndrome*). This virus belongs to the family of retroviruses (*reverse transcriptase oncoviruses*), which are characterized by an RNA genome and replicate via a DNA intermediate. The enzyme *reverse transcriptase* is essential for transformation of the genetic information into a DNA sequence which is inserted into the host DNA genome. Later, when the virus becomes virulent the hosts' transcriptional and translational systems are used for the viral proliferation. Since no cure or immunization has been found so far, therapy has focused on suppression of disease outbreak and viral proliferation. Most of these medical efforts aim at protein targets, e.g.

through inhibitory drugs against the *reverse transcriptase*.<sup>[73]</sup> In the recent past, more and more viral strains have built up resistances against the drugs requiring the development of new drugs and the application of orthogonal therapies.<sup>[74]</sup> This issue and the latest results on the functional diversity of RNA led to new targets in the retroviral system. Since the viral RNA possesses several regulatory functions, interfering with these can help to optimize therapies circumventing the resistances.<sup>[75]</sup> Two RNA sequences have been widely investigated to understand their biological function as well as model sequences for the methodological development in general RNA research: the packaging signal element  $\Psi$  and the transactivation response element TAR.

### Packaging Signal Element $\Psi$

The first RNA sequence widely investigated and generally assumed as drug target in HIV is the packaging signal element  $\Psi$  in the 5' untranslated region of the viral genome. The sequence contains around 120 nucleotides and folds into four stemloops, denoted SL1-4 (Figure 9 A). Different functions are attributed to the different loops.<sup>[35]</sup> The first stemloop, SL1, contains a palindromic sequence in the loop which is the main dimerization site, DIS (dimerization initiation site). This sequence is essential to insure the encapsidation of two viral genomes in a new virus particle prior to the budding from the host cell. In addition, an influence on the encapsidation could be determined.<sup>[76-78]</sup> SL2 has also been abbreviated as SD stemloop for its function as the major splice donor.<sup>[79]</sup> Additional interactions with the nucleocapsid protein (NC) of the Gag polyprotein indicate its participation in genome recognition and packaging.<sup>[80, 81]</sup> For the main function of the  $\Psi$ -RNA, the viral genome packaging, the third stemloop, SL3 (Figure 9 B), is essential since it can package RNA into virus-like particles by itself.<sup>[82]</sup> Its main binding partner is the NCp7 protein domain in the Gag polyprotein. Following the AUG start-codon of the Gag protein is the last stemloop, SL4, of the  $\Psi$ -RNA. The close proximity to the start codon implies a regulatory function which is not

properly understood yet.<sup>[83]</sup> So far, the combined effects of the stemloops are not well known since the individual functions overlap and are diverse.

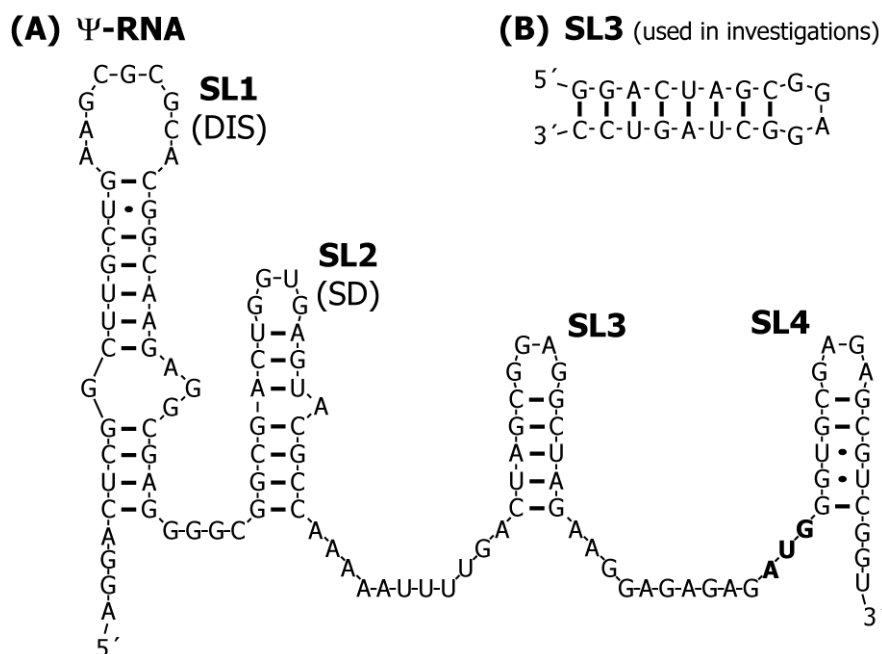


Figure 9: RNA secondary structures of the HIV packaging signal element  $\Psi$  (A) and its stemloop SL3 used in further investigations (B). The stemloops SL1 and SL2 are also coined as the dimerization initiation site (DIS) and the major splice donor (SD), respectively. The AUG start codon right in front of SL4 is highlighted with bold letters. Watson-Crick base pairs are sketched as thick lines and G-U wobble base pairs as ellipsoids.

The structures of the individual stemloops have been solved in solution by NMR (SL1<sup>[84]</sup>, SL2<sup>[85]</sup>, SL3<sup>[86]</sup> and SL4<sup>[87]</sup>). The considerable size and high flexibility of the whole  $\Psi$ -RNA prevented a determination of the overall structural arrangement until just recently. By combination of biochemical cross-linking and mass spectroscopic analysis, a structural model could be derived by Fabris and co-workers.<sup>[88]</sup> In this model, the four helical stems are oriented parallel to each other, SL2 and SL4 enclose SL1 and the GNRA-tetraloop of SL4 binds to the bulge of SL1. There is evidence that the isolated stemloop structures are similar to the structures in their complete RNA, supporting the assumption that studies with the individual hairpins are biologically relevant.



### **Tryptophan-Rich Peptide Ligands of HIV $\Psi$ -RNA**

The phage-display assay is a widely used methodology to screen for ligands of a chosen target. These targets can range from large biomacromolecules, like antibodies, to short peptides.<sup>[89]</sup> RNA sequences have also been successfully used as targets, like e.g. the Rev responsive element stem loop IIB (RRE-IIB)<sup>[90]</sup>, the Tat transactivating region (TAR)<sup>[91]</sup> or the packaging signal ( $\Psi$ ).<sup>[1]</sup> These RNA targets are biotinylated directly or via a linker and immobilized on streptavidine coated plates. These plates are then incubated with a library of bacteriophages which present peptide sequences of selected lengths or fragments of selected proteins on their surface. Phages displaying a tightly binding peptidic ligand remain bound to the plates during the subsequent washing steps. To sort out unspecific binders, rounds of negative selection are performed with mutated or other distinct targets. After several rounds of positive and negative selection the genome inserts of the positive phages are analyzed and the displayed sequences are determined.

With the full length RNA packaging structure  $\Psi$  and with its individual stemloops SL1, SL2 and SL3 as targets, peptide motifs were extracted from phage libraries displaying 7 or 12 amino acids linearly or 7 amino acids in a cyclic form. Among most motifs detected as RNA ligands were clusters of aromatic amino acids, especially tryptophan. From these, the consensus sequence HWWPWW was selected as most prominent lead peptide.<sup>[1]</sup> The presence of multiple tryptophan residues in the lead peptide is not surprising since the natural ligand of  $\Psi$ , the nucleocapsid protein NCp7, shows specific interactions of its residue W37 with unpaired guanosine residues within SL3.<sup>[86, 92]</sup> NCp7 also shows interactions with further guanosine residues in the other  $\Psi$  stemloops.<sup>[93, 94]</sup> This consensus sequence is also in agreement with a characterization of RNA ligands by Dinshaw Patel stating that tryptophan-rich ligands interact by aromatic stacking with hairpin loops.<sup>[95]</sup> The further investigations and results are presented in CHAPTER V and were published.<sup>[1-3]</sup>

## Transactivation Response Element TAR

The transactivation response element TAR is a 59 nucleotide RNA stem-loop (Figure 10 A) structure located at the 5'-end of all nascently transcribed messenger RNAs.<sup>[36]</sup> It contains a six nucleotide loop, two single-nucleotide bulges and a two- or three-nucleotide bulge. The size of the larger bulge is the only difference between the main HIV strain, classified as Type 1 (HIV-1), and a variant occurring in Africa, HIV Type 2 (HIV-2). HIV-1 TAR has a trinucleotide bulge with the sequence U23-C24-U25 while the HIV-2 TAR lacks the central cytidine. This bulge is essential for the binding of the transactivator of transcription protein (Tat) which thereby increases the level of transcription of the downstream mRNA. Tat mainly binds through its basic domain of sequence RKKRRQRRR. Further interactions with other proteins, in particular cyclin T1, then activate RNA polymerase II by phosphorylation and lead to the explosive increase of transcription of viral proteins.<sup>[36, 96]</sup>

Studies on TAR RNA structure, dynamics and its interactions have mainly been focused on the upper hairpin including the larger bulge (Figure 10 B, C and D). The structure has been solved free in solution<sup>[97]</sup> as well as complexed to a Tat protein<sup>[98]</sup> and several small ligands.<sup>[57, 58, 99-110]</sup> Free TAR consists of two canonical stems connected by the flexible bulge linker. The angular fluctuation of the two stems can range up to 94°<sup>[56]</sup> and the bulge nucleobases of U23 and C24 can stack on the lower stem, while U25 is mostly unrestricted in its motions.<sup>[58, 97]</sup> A convincing evidence for the disruption in stacking at the bulge region is that the imino proton of U40 right beneath the bulge cannot be detected in NMR experiments. Increased dynamics and the fast exchange with solvent broaden the resonance extremely. A possible base triple of U23 with the canonical base pair A27-U38 has been assumed due to the close spatial proximity of the nucleobases, but the high flexibility of the bulge nucleotides prevents a direct experimental proof by NMR spectroscopy or X-ray crystallography.

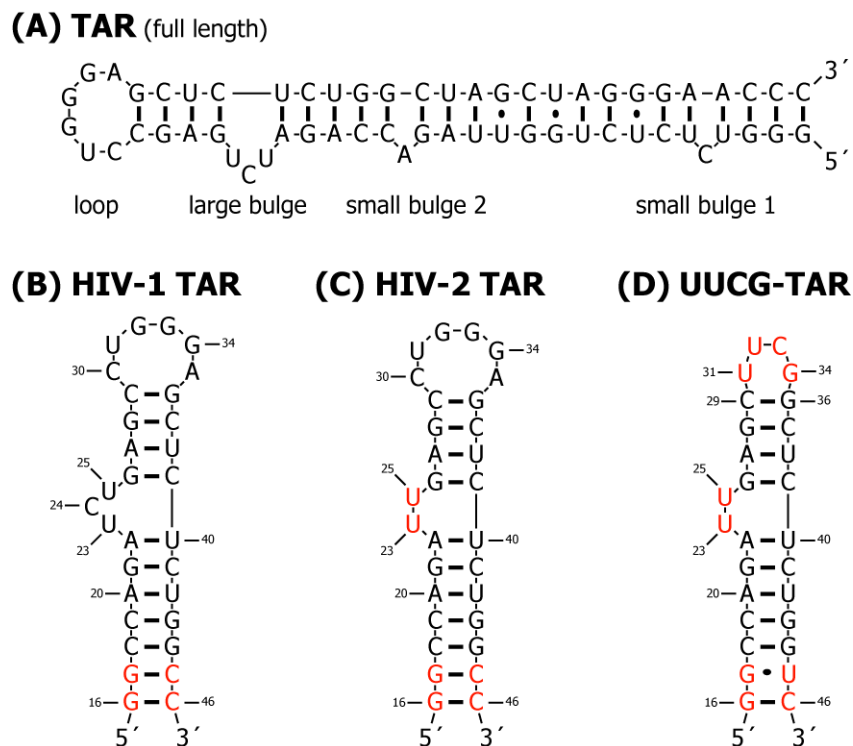


Figure 10: RNA secondary structures of the full length HIV transactivation response element TAR **(A)** and variants used in further investigations: the HIV-1 TAR **(B)**, the HIV-2 TAR **(C)** and a UUCG-tetraloop mutant of HIV-2 TAR **(D)**. Watson-Crick base pairs are sketched as thick lines and G-U wobble base pairs as ellipsoids. Differences in the sequences **B**, **C** and **D** from the wt HIV-1 TAR **(A)** are colored red.

The hexaloop 5'-CUGGGA-3' also displays a highly dynamic structure, which is assumed to be stabilized by stacking interactions and a cross-loop base pair between C30-G34. Colvin et al. could detect some sequential NOEs between the loop nucleotides suggesting that the nucleobases are stacked, but no base pairs are formed in the loop.<sup>[111]</sup> In contrast Kulinski et al., who derive their conclusions from biochemical studies and molecular dynamics simulations, claim C30 and G34 could form a base pair stacking on the last stem base pair C29-G36.<sup>[112]</sup> The intermediate residue A35 displays abnormal NMR chemical shifts suggesting an unusual environment, which is consistent with an extruding residue.<sup>[113]</sup>

In order to investigate and understand its binding characteristics, the complex of TAR with argininamide has widely been used.<sup>[56-58, 68, 114-117]</sup> This ligand

binds with a low millimolar dissociation constant in the bulge region.<sup>[116]</sup> It induces, both in HIV-1<sup>[58]</sup> and HIV-2 TAR<sup>[114]</sup>, the formation of a base triple between the bulge residue U23 and the canonical A27-U38 base pair. U23 interacts in a reverse Hoogsteen manner with A27 (Figure 5 C). This region also displays the highest number of contacts to the ligand.<sup>[58]</sup> Ligand binding is accomplished by a significant change in internal flexibility of these residues, as observed experimentally by a NMR relaxation analysis.<sup>[115]</sup> In contrast, the other bulge nucleotides, C24 and U25, preserve their flexibility and the NMR chemical shifts indicate that their nucleobase moieties are pointing out of the helical stems. This complexation also rigidifies the two adjacent stems to an almost linear and continuous double helix<sup>[117]</sup> with stacking interactions between the base pairs A22-U40 and G26-C39. In addition, the imino proton of U40 becomes visible confirming the closure of the bulge region and the reduction of its dynamics.

Several further ligands have been investigated in complexation with TAR RNA, beginning from Tat derived peptides<sup>[99, 101, 118]</sup>, through ligands with hetero-aromatic moieties<sup>[102-104]</sup> and guanidinium group-mimicking molecules<sup>[106]</sup> to aminoglycosides<sup>[107]</sup>. Almost all of them bind in the bulge region and therefore compete with the Tat protein. Although they all cause different structural arrangements, they rigidify the interhelical angle. The angular distribution of the TAR-complexes ranges from 5° to 47°.

### **Arginine-Flanked Synthetic Peptide Ligands of HIV TAR-RNA**

Since the selectivity of arginine moieties for the bulge of TAR RNA has been thoroughly investigated by different groups, two arginine residues frame the ligands from a combinatorial library of synthetic tripeptides in a screening assay by Michael Göbel and coworkers. For detection, an established fluorimetric competition assay (FRET) was used in which the ligands compete with a Tat-derived peptide, labeled with fluorescein and rhodamine, for the TAR bulge binding site.<sup>[4]</sup> Complexed to the RNA the two fluorimetric labels of

the peptide are separated from each other and therefore a high level of fluorescence is detectable. In contrast, when the labeled peptide is free in solution the fluorophores quench each other significantly. Therefore, a reduction of the fluorescence during titration with a ligand indicates the competition between the peptide and the ligand for a binding site. This results in an  $IC_{50}$ -value of the ligand, which can then be translated into a binding constant with knowledge of the binding constant of the peptide.

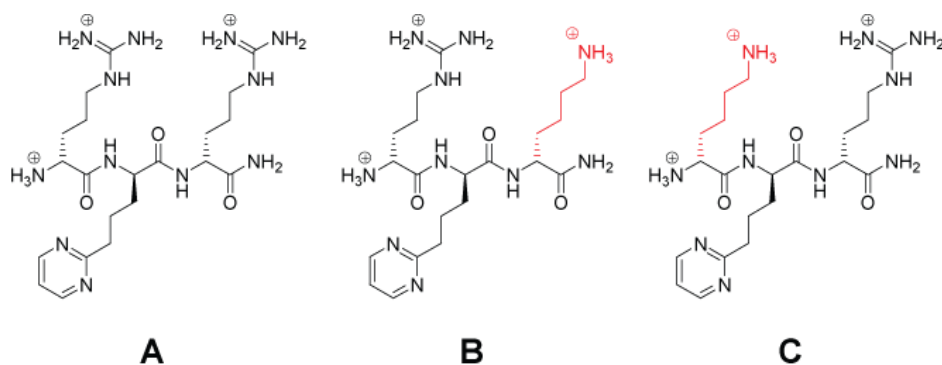


Figure 11: Constitution of the TAR-ligands investigated in the publication Ferner et al.<sup>[5]</sup> Ligand **A** was used in the intensive structural investigation of the RNA-ligand complex. Ligands **B** and **C** are lysine mutants of **A**. The differences are colored in red.

To increase the binding affinity and selectivity, a central amino acid was introduced with an aromatic or heteroaromatic side chain (Figure 11).<sup>[119]</sup> This moiety should enforce hydrophobic stacking interactions of the ligand with nucleobases of the bulge and adjacent residues. Heteroaromatic side chains may also contribute by their possibility to engage in hydrogen bonds. To increase the stability of the peptide against proteolytic degradation, the ligands were mostly synthesized with the D-amino acid configuration. The stereochemistry was shown to have only marginal influence on the affinity in the case of short Tat-derived peptides.<sup>[120]</sup> The spectroscopic investigations and the structure determination of an RNA-ligand 1:2 complex is presented in CHAPTER VI and published.<sup>[4, 5]</sup>

## The Coxsackievirus B3

### Stemloop D Subdomain of the Cloverleaf RNA

Coxsackieviruses are the cause of several human diseases and belong to the group of enteroviruses. Their RNA genome encodes for a polyprotein of about 2200 amino acids which is processed by viral proteases. All enteroviruses possess a highly structured and conserved 5'-untranslated region including the internal ribosomal entry site (IRES) and a cloverleaf-like RNA element required for translation and replication, respectively.<sup>[37, 38]</sup> The cloverleaf-like element mediates the assembly of several viral and host proteins to build a ribonucleoprotein complex which initiates the replication process. One component crucial for the assembly, the proteinase 3C<sup>pro</sup>, specifically recognizes the element alone *in vitro* by binding to the stemloop D subdomain (Figure 12). The main binding site of the protein can be mapped adjacent to the symmetrical internal loop and to the apical tetraloop of the RNA.<sup>[50]</sup> While the specific identity of the four loop nucleotides seems dispensable, the structural features and the U/C-G closing base pair appear to be more relevant for protein binding. The apical loop has to be a member of the well-structured YNMG-family (CHAPTER II, page 17).<sup>[50]</sup>

#### Stemloop D of the Cloverleaf Element in Coxsackievirus B3

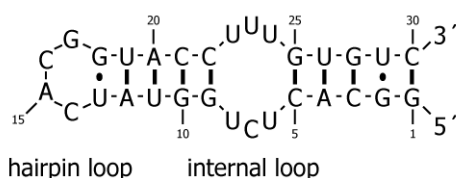


Figure 12: RNA secondary structure of the Stemloop D (SLD) of the 5'-Cloverleaf Element in Coxsackievirus B3. Watson-Crick base pairs are sketched as thick lines and G-U wobble base pairs as ellipsoids.

The symmetrical internal loop contains three pyrimidine base pairs: two asymmetric U-U base pairs flanking a central C-U pair. This non-canonical

region displays a narrowed minor groove and a high electronegativity in the major and minor groove, since 5 and 6 carbonyl groups are positioned here, respectively. These may also act as hydrogen bond acceptors for a ligand.<sup>[50]</sup>





## CHAPTER IV

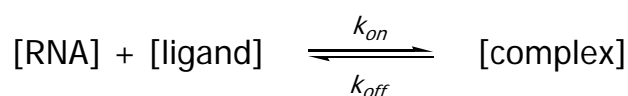
## NMR THEORY &amp; METHODOLOGIES

Within this chapter the NMR theory and methodologies are introduced which were used in the investigations of RNA-ligand binding interactions, RNA-ligand structure elucidation and RNA dynamics.

## Ligand Binding Interactions

The study of interactions between small ligands and biomacromolecules by NMR spectroscopy requires the consideration of some peculiar factors. The low inherent sensitivity of NMR leads to the necessity of rather high concentrations of the components ( $>10 \mu\text{Mol}$ ), prohibiting a precise determination of the thermodynamic dissociation constant  $K_D$  for high affinity ligands, where the  $K_D$  is in the nanomolar region to compete with the natural binding partner. However, 1D  $^1\text{H}$  NMR can easily provide information on the binding epitope and the binding kinetics. With the assignment of the RNA imino resonances (described in the following paragraph), changes in this spectral region can be followed during a stepwise titration of the ligand to an RNA sample. Imino resonances in or close to the binding site normally display the largest changes in chemical shifts and/or line widths, while signals further away from the binding site retain their resonance positions and line shape. The changes of resonance position and line shape are indicative for the binding epitope and kinetics (Figure 13).

The binding kinetics have a large influence on the NMR signals. While the on-rate  $k_{on}$  of the ligand binding to RNA is often dominated by the diffusion limit, which lies between  $10^8$  to  $10^9 \text{ M}^{-1}\text{s}^{-1}$  for small ligands in water, the off-rate  $k_{off}$  is dependent on the binding characteristics. By NMR, these characteristics are separated into three cases (**1-3**), dependent on the lifetime  $\tau$  ( $=1/k_{binding}$ ) (with  $k_{binding} = k_{on} + k_{off}$ ) of the RNA-ligand complex.



$$K_D = \frac{[\text{RNA}] \cdot [\text{ligand}]}{[\text{complex}]} = \frac{k_{\text{off}}}{k_{\text{on}}} \quad (\text{Equation 1})$$

### 1. fast exchange

Here, the sum of  $k_{\text{on}}$  and  $k_{\text{off}}$  is significantly larger than the absolute difference of the resonance frequencies in the bound and the free state:

$$k_{\text{binding}} \gg |\delta_{\text{bound}} - \delta_{\text{free}}| \quad (\text{Equation 2}),$$

where  $\delta_{\text{free}}$  and  $\delta_{\text{bound}}$  are the resonances of the same nucleus in the free and the bound state, respectively. In this case, one NMR signal per nucleus is observed ( $\delta_{\text{obs}}$ ) with the resonance position being the weighted average of the chemical shifts arising from the nucleus in the free and the bound state:

$$\delta_{\text{obs}} = f_{\text{bound}} \delta_{\text{bound}} + f_{\text{free}} \delta_{\text{free}} \quad (\text{Equation 3}),$$

where  $f_{\text{free}}$  and  $f_{\text{bound}}$  are the populations in the free and the bound state and sum up to 1.

### 2. intermediate exchange

$$k_{\text{binding}} \approx |\delta_{\text{bound}} - \delta_{\text{free}}| \quad (\text{Equation 4})$$

If the sum of the rates,  $k_{\text{binding}}$ , is similar to the absolute difference in chemical shifts of the free and the bound state, the signals broaden extremely and can even become undetectable. Therefore, this timescale of binding is the most undesired to investigate by NMR spectroscopy. A change of temperature (change of  $k_{\text{off}}/k_{\text{on}}$ ) and of the magnetic field (change of  $\Delta\delta$ ) can help in these cases.

### 3. slow exchange

$$k_{\text{binding}} \ll |\delta_{\text{bound}} - \delta_{\text{free}}| \quad (\text{Equation 5})$$

In this case, which is typical for high affinity complexes (dissociation constant in the nanomolar range), two signals are observed for each proton, one for the free and one for the bound form. During the

titration the signals of the free form decrease, while the signals from the bound form increase according to their populations.

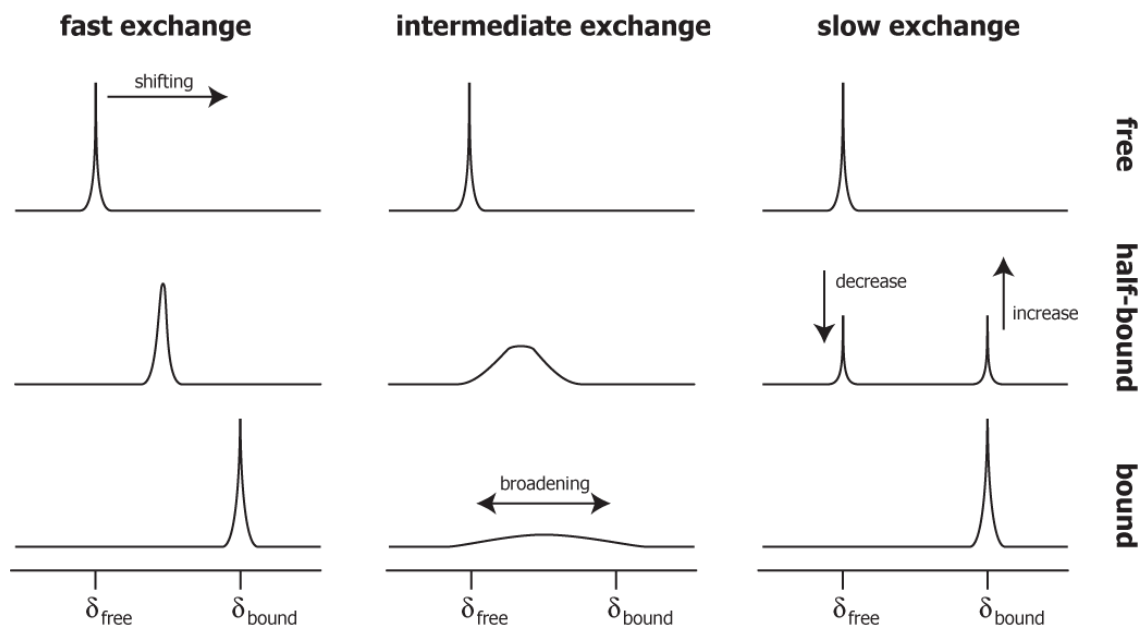


Figure 13: Schematic representation of NMR signal changes during titration with a ligand in the case of a binding process in fast, intermediate and slow exchange. In fast exchange the signal shifts from the resonance of the free-state to the bound-state; in intermediate exchange the signal broadens and in slow exchange the signal representing the free-state decreases while the bound-state signal increases.

## Resonance Assignment

To gain NOE distance restraints for the structural calculation, an almost complete assignment of all proton resonances of all investigated molecules is necessary. Two main approaches can be selected to achieve the assignment with a  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled RNA sample. The first approach arises from the NOE-based assignment process used for unlabeled molecules and is complemented with through-bond homo- and heteronuclear correlations (scalar couplings  $^nJ$ , Figure 17).<sup>[121, 122]</sup> In the second approach, as many as possible of the NOE-based correlations are replaced by through-bond correlations.<sup>[122]</sup> In Figure 18 a flowchart is displayed combining elements of both approaches. This approach may achieve a complete assignment for structural calculation and verify assignments determined by NOE- and/or through-bond correlations.

## RNA

For the assignment of proton and other nuclear resonances in an RNA macromolecule of around 30 nucleotides, a uniformly  $^{13}\text{C}$ -/ $^{15}\text{N}$ -labeled sample can be sufficient.<sup>[123-125]</sup> For larger RNAs, it may become necessary to synthesize nucleotide-specific labeled samples to especially overcome the overlap of the sugar proton and carbon resonances but also of the aromatic carbons and protons at positions 5 and 6 of the pyrimidines and 8 of the purines.<sup>[50]</sup> To further increase the size limit for NMR investigations, ligation of labeled and unlabeled segments is promising but so far the yields are not satisfying for a routine application.<sup>[126, 127]</sup>

A proper choice of a consistent temperature for all experiments is important. While a temperature of 298 K (~25 °C) is appropriate for the measurement of carbon-bound protons, the nitrogen- and oxygen-bound protons might exchange too fast with the solvent to be detectable or to determine connectivities among each other. Therefore, a decreased temperature of

around 283 K (10 °C) may be advantageous which decreases the exchange rates. To gain a complete assignment, the different resonances, shifted by the temperature, of each nucleus have to be connected.

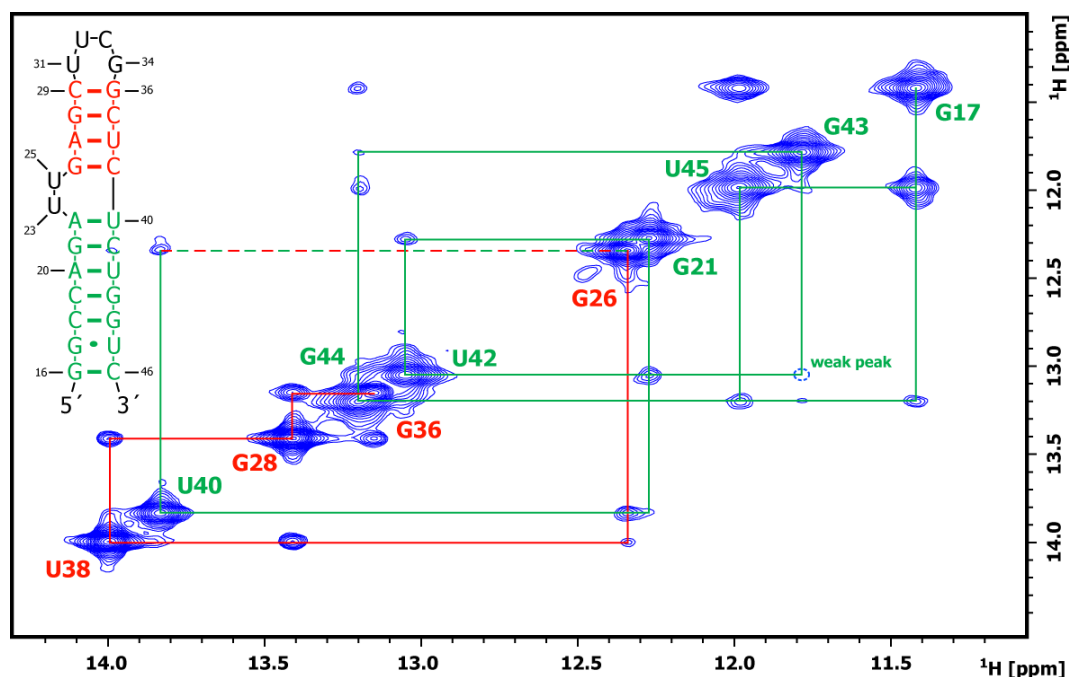


Figure 14: 2D  $^1\text{H},^1\text{H}$ -NOESY spectrum of the UUCG-TAR imino protons (secondary structure is embedded in the figure). The vertical and horizontal lines display the NOESY-walk connecting the sequential iminos of the upper (red) and lower (green) stem, as well as a direct connection of the two stems (red and green dashed line). The cross peak between U42 and G43 is very weak but visual at lower contour levels. The strong cross peak between G17 and U45 confirms the G-U wobble base pair.

As initial step to confirm a stable uniform structure of the RNA sample and to determine the optimal temperature, the 1D  $^1\text{H}$  imino region serves as a good indicator. Most detectable imino signals stem from protons that are involved in hydrogen bonds. In particular, the base-paired imino protons of double-helical stems are visible and give significant evidence of a properly folded RNA. In rare cases, signals of imino protons that are not involved in hydrogen-bonds can be observed. They need to be protected otherwise from solvent exchange. An optimal temperature for the determination of these exchanging protons is identified, when as many as possible signals can be detected (preferentially at low temperatures), but the linewidths of the

signals do not lead to significant overlap (preferentially at a higher temperature).

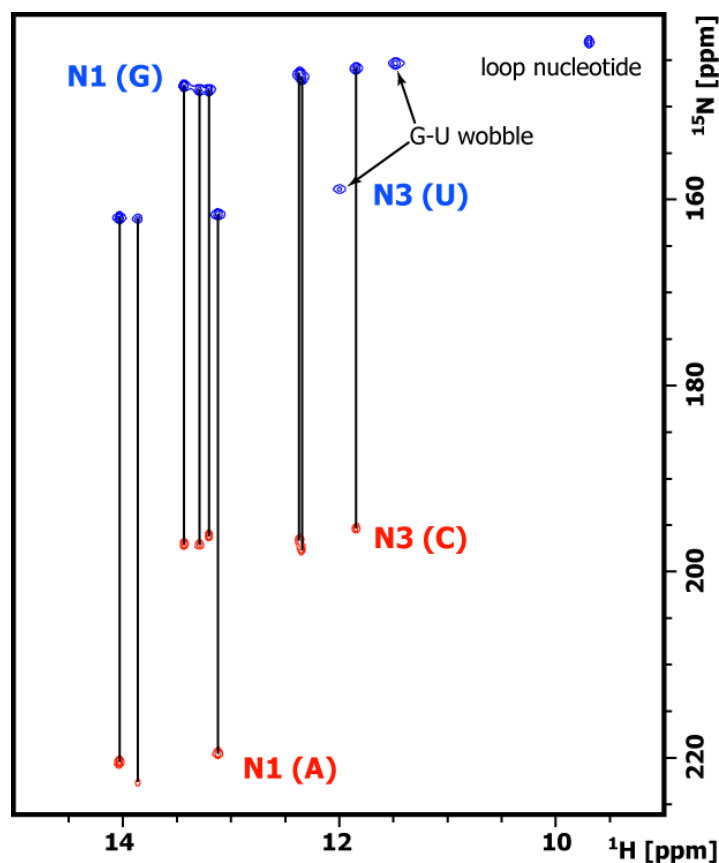


Figure 15: 2D HNN-COSY spectrum of UUCG-TAR (Figure 10 D). The diagonal peaks of the imino hydrogen bond donor nitrogens (N1 (G) and N3 (U)) are colored in blue, of the acceptor nitrogens (N1 (A) and N3 (C) in red). Two imino resonances without cross peaks are depicted as the G-U wobble base pair where carbonyl oxygens are the hydrogen bond acceptors. The most upfield shifted imino resonance (~9.5 ppm) is assigned to a loop nucleotide which also forms a hydrogen bond to a carbonyl oxygen.

Via a  $^{15}\text{N}$ -HSQC, centered between the N1-resonances of guanosine and N3 of uridine (Table 2), the proton resonances can be assigned base type-specifically to a nitrogen nucleus. The first sequential and therefore structural information can be obtained with a 2D  $^1\text{H}$ ,  $^1\text{H}$ -NOESY optimized for the imino protons.<sup>[128, 129]</sup> Strong cross peaks herein between imino protons of a guanosine and a uridine give a strong indication for a G-U wobble base pair (iminos G17 and U45 in Figure 14). With the knowledge of the imino base

identity, sequential walks can be carried out in structured segments, e.g. helical stems (Figure 14). A comparison of these sequences with secondary structure predictions (e.g. from the program MFOLD<sup>[130, 131]</sup>) serves as an indicator of a proper RNA fold. If overlap in the 2D NOESY prevents an unambiguous sequential assignment, a 3D <sup>15</sup>N-edited HSQC-NOESY experiment can help. To determine the identity of the hydrogen bond acceptor, a HNN-COSY<sup>[132]</sup> experiment is necessary which correlates the imino protons also to the acceptor nitrogen (Figure 15). In Watson-Crick base pairs the H1 protons of guanine are correlated to the cytidine nitrogen N3 and the uridine H3 to the adenine N1. The HNN-COSY experiment also reports on unusual hydrogen bonds, e.g. Hoogsteen base pairs where the uridine imino is base-paired to the adenine N7 (Figure 5 C). Also G-U wobble base pairs (Figure 5 B) can be validated since the iminos of guanine and uridine do not coordinate with a nitrogen and therefore should not display a cross peak.

The amino protons display a more significant degree of exchange than the imino protons even when they are base-paired, hindering the assignment and the identification of important connectivities. In addition to the increased solvent exchange, the two amino protons are subject to conformational exchange between each other. Therefore, a special 2D-NOESY experiment is needed which first correlates the exchanging protons with their adjacent nitrogen via a CPMG (Carr-Purcell-Meiboom-Gill) pulse train.<sup>[133]</sup> The nitrogen resonance is then representative for the amino group. A subsequent NOE mixing period establishes correlations with spatially close protons.<sup>[134]</sup> With help of a <sup>15</sup>N-HSQC, centered in the region of the amino nitrogen resonances, the assignment of all base-paired protons and nitrogen nuclei can be completed and first starting points for the further assignment are set.

Table 2: Chemical shifts ranges of  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  and  $^{31}\text{P}$  derived from the BMRB database for all included RNA biomacromolecules. The regions represent the standard deviations around the average value. The values in italic are based on less than 20 inputs.

Atom Type		Atom Name	Chemical Shift Regions [ppm]				
			Adenosine	Guanosine	Uridine	Cytidine	
$^1\text{H}$	exchangeable	H1		11.4 – 13.4			
		H21 <sup>a</sup>		6.3 – 8.5			
		H22 <sup>a</sup>		5.8 – 7.5			
		H3			11.7 – 14.5		
		H41 <sup>a</sup>				7.4 – 8.7	
		H42 <sup>a</sup>				6.6 – 7.9	
		H61 <sup>a</sup>	6.7 – 8.3				
		H62 <sup>a</sup>	6.3 – 7.5				
		HO2'	<i>6.7 – 7.0</i>	<i>6.7 – 7.0</i>	<i>6.7 – 6.9</i>	<i>6.3 – 8.6</i>	
	aromatic	H2	7.3 – 8.1				
		H5			5.2 – 5.8	5.2 – 5.8	
		H6			7.6 – 8.0	7.5 – 7.9	
		H8	7.7 – 8.3	7.2 – 7.9			
		aliphatic	H1'	5.6 – 6.1	5.3 – 6.0	5.4 – 5.8	5.3 – 5.8
			H2'	4.4 – 4.8	4.3 – 4.8	4.2 – 4.6	4.1 – 4.6
H3'	4.4 – 4.8		4.3 – 4.8	4.3 – 4.7	4.2 – 4.6		
H4'	4.2 – 4.7		4.2 – 4.6	4.3 – 4.5	4.0 – 4.8		
H5'	4.1 – 4.5		4.0 – 4.5	4.0 – 4.5	4.0 – 4.5		
	H5''	4.0 – 4.5	4.0 – 4.4	4.0 – 4.4	3.9 – 4.5		
$^{13}\text{C}$	aromatic	C2	150.1 – 155.9	150.0 – 158.6	144.9 – 164.8	123.2 – 192.0	
		C4	<i>146.3 – 147.9</i>	<i>149.1 – 151.1</i>	165.3 – 168.7	160.6 – 170.1	
		C5	<i>94.9 – 136.9</i>	104.6 – 143.4	101.4 – 104.9	93.4 – 101.2	
		C6	<i>150.3 – 159.6</i>	151.4 – 165.1	138.5 – 143.7	138.8 – 142.6	
		C8	135.4 – 142.7	133.3 – 139.1			
		aliphatic	C1'	89.2 – 93.2	89.1 – 93.2	90.0 – 93.9	90.7 – 94.3
	C2'		69.7 – 79.5	69.9 – 78.4	72.5 – 76.6	71.7 – 77.8	
	C3'		70.2 – 76.0	70.1 – 75.7	70.2 – 75.7	68.7 – 75.1	
	C4'		79.6 – 84.5	79.4 – 84.6	79.8 – 84.7	79.2 – 83.8	
		C5'	60.4 – 70.5	59.7 – 71.9	58.9 – 72.6	59.7 – 69.1	
$^{15}\text{N}$	aromatic	N1	211.1 – 230.4	140.4 – 152.3	142.2 – 152.8	135.0 – 162.3	
		N3	207.8 – 220.7	<i>98.3 – 194.2</i>	157.9 – 163.2	178.8 – 202.8	
		N7	227.8 – 233.0	226.9 – 237.3			
		N9	167.2 – 171.9	167.4 – 170.5			
	amino	N2		72.9 – 76.7			
		N4				95.9 – 99.5	
	N6	78.6 – 84.3					
$^{31}\text{P}$	backbone	P	-4.7 – -1.4	-4.5 – -1.4	-4.9 – -1.9	-5.0 – -2.0	

<sup>a</sup> The nomenclature of the two amino protons in cytidine (H41/H42), guanosine (H21/H22) and adenosine (H61/H62) need to be handled in consideration of differing definitions. While IUPAC and BMRB distinguish by the chemical shift, most force fields refer to the conformation, where the protons H42, H22 and H62 point to the Watson-Crick site. Within this thesis, the IUPAC nomenclature is used.



In the next step, the carbon-bound protons are investigated. To reduce the problems caused by the strong solvent signal, e.g. overlap and radiation damping, an exchange of the solvent to 100% D<sub>2</sub>O is necessary. This exchange can be achieved by freeze drying of the previously used sample and refilling with the same volume of D<sub>2</sub>O which preserves the concentrations and buffer conditions. Freeze drying of the sample should be repeated two to three times to achieve a nearly complete D<sub>2</sub>O to H<sub>2</sub>O exchange and also the NMR-tube should be washed with D<sub>2</sub>O and dried. After solvent exchange, the integrity of the sample should be checked by native gel electrophoresis.

<sup>13</sup>C-HSQC spectra of the sugar (~56 ppm to ~93 ppm) and C5 (~95 ppm to ~105 ppm) region and of the aromatic carbons C2, C6 and C8 (130-155 ppm) give a complete view of the RNA. Each nucleotide should result in a H1'/C1' cross peak and a H8/C8 or H6/C6 cross peak for purines and pyrimidines, respectively, also enabling to check the correct number of residues. Scalar couplings with neighboring carbon spins in the sugar moieties (C1' to C5') and in the pyrimidine bases (C4, C5, C6) give rise to undesired splittings in the indirect <sup>13</sup>C-dimension which result in increased spectral overlap and can impair resonance determination and peak counting. This problem can be circumvented by incrementing the indirect dimension within a constant time delay, which refocuses the corresponding coupling ( $\tau = 2/(^1J_{CC})$ ). For the measurement of the aromatic carbons ( $^1J_{CC} \approx 66$  Hz) this delay is usually set to 30 ms and in the sugar region ( $^1J_{CC} \approx 40$  Hz) to 50 ms (Figure 16). During each increment the transverse relaxation reduces the signal intensity, resulting in weakened or even undetectable resonances of fast relaxing carbon nuclei. Therefore, an additional experiment with a delay of  $1/(^1J_{CC}) = 25$  ms is measured for the ribose moieties but the maximum possible resolution is halved. As side effect, this shorter experiment helps differentiating the C1' and C5' nuclei from C2', C3' and C4' since their cross peaks possess a negative sign due to only one coupling carbon neighbor.

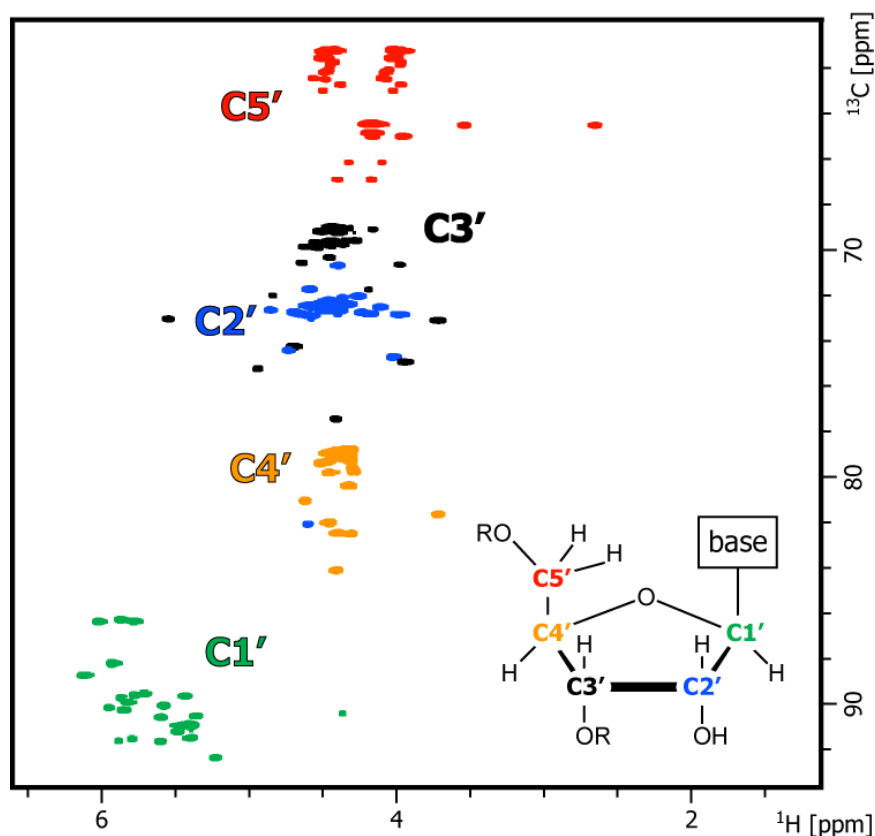


Figure 16: 2D Constant time  $^1\text{H},^{13}\text{C}$ -HSQC spectrum of the RNA ribose resonances ( $^{13}\text{C},^{15}\text{N}$ -labeled UUCG-TAR (Figure 10 D) with an eightfold excess of the unlabeled pyrimidinyl-ligand (Figure 11 A) in  $\text{D}_2\text{O}$ ). The signals of the individual ribose carbons are colored uniformly, enunciating the typical spectral regions. Outliers arise from nucleotides in the loop, bulge or at the terminus.

Due to the good resolution of the anomeric carbon resonances in respect to the other ribose carbon resonances, the C1'-nuclei represent a good starting point for the assignment of the individual ribose spin systems. With a 3D HCCH-COSY and a 3D HCCH-TOCSY<sup>[135]</sup> the carbon and proton resonances of each spin system can be determined. After incrementation of  $^1\text{H}$  in the first dimension and of the directly bonded  $^{13}\text{C}$  in the second dimension, a COSY or TOCSY mixing period follows to correlate the carbons with the directly neighbored carbon (COSY) or through the whole carbon spin system (TOCSY). To further reduce signal overlap a variant of the HCCH-TOCSY is useful, the forward-directed HCC-TOCSY-CCH-E.COSY.<sup>[136, 137]</sup> By skillful choice of delays and phase-cycling, cross peaks between  $\text{C}_i'$  and  $\text{H}_{(i+1)'}$  (with  $i = 1-4$ ) are obtained in the resonance plane of each anomeric proton. In addition, due to the E.COSY transfer also  $^3\text{J}(^1\text{H},^1\text{H})$ -coupling constants can be extracted,

providing information on the furanose sugar conformation. The measurement of  ${}^3J(\text{H1}',\text{H2}')$  and  ${}^3J(\text{H3}',\text{H4}')$  helps to determine the sugar conformation in an RNA molecule. The canonical A-form RNA has a C3'-endo sugar conformation resulting in a small  ${}^3J(\text{H1}',\text{H2}')$  and a large  ${}^3J(\text{H3}',\text{H4}')$ . In the inverse case of a large  ${}^3J(\text{H1}',\text{H2}')$  and small  ${}^3J(\text{H3}',\text{H4}')$  the sugar is in the second most populated conformation, C2'-endo.

Connection via scalar couplings of the intraresidual spin systems of ribose and nucleobase are achieved by the 2D H(C)N-experiment.<sup>[138]</sup> Via INEPT steps, magnetization is transferred from the protons (H1' and H8/H6) across the adjacent carbons to the anomeric nitrogen (N1 in pyrimidines and N9 in purines). The nitrogen resonances are then incremented in the first dimension. In the following, the magnetization is transferred back to the proton it originates from (out-and-back transfer) and further to the proton of the corresponding ribose or nucleobase (all-the-way-through transfer). Within this transfer, an incrementation of the carbon resonances is possible to gain further resonance dispersion (3D HCN). If the resolution of nitrogen and proton resonances, however, is sufficient, this should be avoided to reduce magnetization loss due to carbon relaxation. To optimize this experiment, selective delays and pulses can be chosen for the  ${}^1\text{H}$ - ${}^{13}\text{C}$  INEPT transfers, since the couplings differ in the nucleobase and in the ribose (Figure 17). This results in an all-the-way-through transfer and can be applied in both directions, benefitting from resolution and relaxation differences of the aromatic and aliphatic nuclei.<sup>[139, 140]</sup>



For a sequential assignment of the ribose spin systems the couplings between carbons C3', C4' and C5' to the phosphorous nuclei of the phosphodiester backbone can be used in a HCP-TOCSY-experiment.<sup>[141]</sup> Similar to the HCN-experiment, magnetization is transferred from the ribose protons across the carbons to the <sup>31</sup>P nuclei via INEPT steps. After incrementation of the phosphorous resonances, magnetization is transferred back and further to the C3', C4' and C5' carbons the magnetization originated from and to the carbons of the adjacent residue, respectively. Following this, a CC-TOCSY step forwards the magnetization along the ribose spin system to the anomeric C1'. Incrementation of the C1' resonances in the second dimension is advantageous due to the better resolution of these carbon nuclei. Detection is then accomplished on the ribose protons. With this pulse sequence, the backbone <sup>31</sup>P is correlated to two ribose spin systems enabling a sequential assignment.

To gain further assignments and to circumvent problems due to specific characteristics of the investigated RNA molecule, the following experiments may help: The H5(C5)NN-COSY<sup>[142]</sup> experiment is a variant of the HNN-COSY experiment explained previously, which correlates the non-exchangeable H5 proton of uracil with its N3 nitrogen and with a base-paired nitrogen if existent. This experiment is helpful if the proton of the base pair hydrogen bond is not detectable or overlap of imino protons hinders the assignment of a correlation. The sufficiently strong <sup>2</sup>J coupling between C5 and N3 of about 5.7 Hz (Figure 17 C) in the uridine nucleobase constitutes the essential basis of this experiment which is not applicable for cytosines.

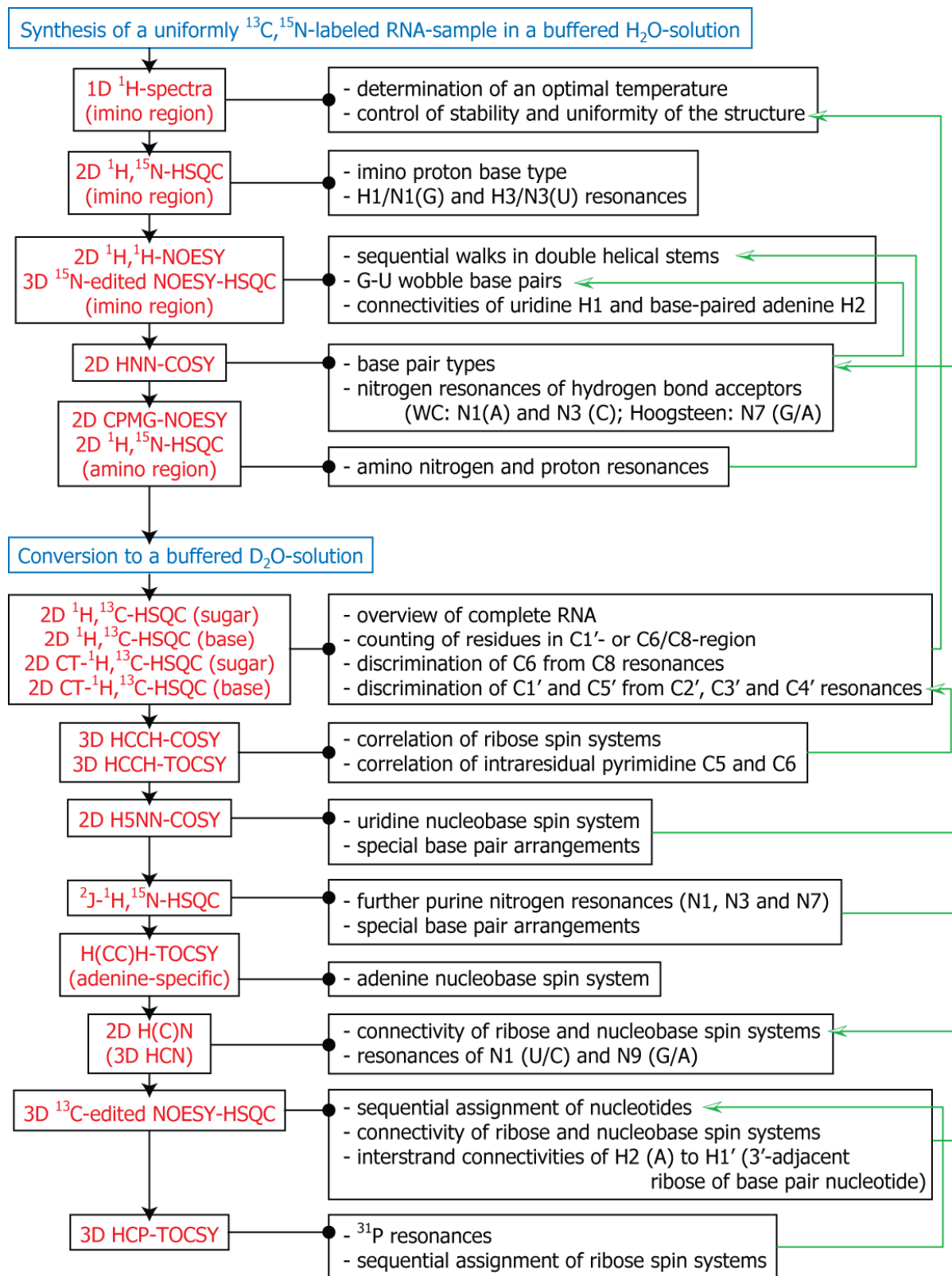


Figure 18: Flowchart of NMR experiments (red) to obtain a complete assignment of an RNA molecule. Corresponding sample conditions are colored in blue. The conclusions of the experiments are listed aside. Green arrows point out conclusions, verifications and additional information on the outcome of previous experiments.

For detection of weakly bound base pairs also the knowledge of the chemical shift of the hydrogen bond acceptor can be useful. With a  ${}^2J\text{-}^{15}\text{N}\text{-HSQC}$ <sup>[143]</sup> the non-exchangeable protons H2 and H8 in adenosine and H8 in guanosine can be correlated to the nitrogens N1, N3, N7 and N9, respectively. Especially for Hoogsteen base pairs where the purine N7 act as hydrogen bond acceptors this experiment is helpful to determine their chemical shifts, which can be indicative for a hydrogen bond by a downfield shift of over 2 ppm in comparison to nucleobases in canonical Watson-Crick base-pairs.<sup>[144]</sup> The correlations of H2 to N1 and N3 in adenines can be used to verify the N1-assignments in canonical stems or for determination of unusual hydrogen bonds to N3.<sup>[145]</sup>

To correlate the two non-exchanging hydrogens in the adenosine nucleobase, H2 and H8, with each other a variant of the HCCH-TOCSY experiment explained previously can be used. This variant results in a 2D  ${}^1\text{H},{}^1\text{H}$ -correlation using a  ${}^{13}\text{C},{}^{13}\text{C}$ -TOCSY with optimized delays for the stronger  ${}^1J$ -couplings between the aromatic protons and carbons, an optimized  ${}^{13}\text{C}$ -offset resonance and a longer TOCSY mixing time due to the weaker homonuclear carbon couplings.<sup>[146]</sup> This experiment may be necessary to correlate the Watson-Crick site with the sugar moiety in purines.

## Peptide Ligand

Standard NMR experiments ( ${}^1\text{H},{}^{13}\text{C}\text{-HSQC}$ ,  ${}^1\text{H},{}^{15}\text{N}\text{-HSQC}$ , 2D  ${}^1\text{H},{}^1\text{H}\text{-COSY}$ , 2D  ${}^1\text{H},{}^1\text{H}\text{-TOCSY}$  and 2D  ${}^1\text{H},{}^1\text{H}\text{-NOESY}$ ) as partly described previously for RNA are used for the assignment of small ligand molecules. Due to the slower relaxation in small molecules and the higher achievable concentrations in comparison to the RNA biomacromolecules a selective labeling of the ligands is not necessary and the heteronuclear experiments rely on natural abundance of  ${}^{13}\text{C}$  and  ${}^{15}\text{N}$ -isotopes.

For ligands binding to RNA in fast exchange the fast correlation time of the ligand may lead to problems in the NOESY spectra. The nuclear Overhauser effect between two nuclei is dependent on their correlation time and

possesses a zero crossing. At this critical point, no NOE between the two nuclei can be observed. For a  $^1\text{H}, ^1\text{H}$ -interaction it lies at  $\omega_0\tau_c \approx 1.12$  which corresponds to a correlation time of  $\sim 300$  ps at 600 MHz. To avoid this pitfall, the ROESY (= Rotating Frame NOESY) experiment should be chosen.<sup>[147]</sup> Although the ROE provides overall weaker signal intensities, it is rather similar for a wide range of correlation times and does not possess a zero crossing. By applying a spin-lock field during the mixing time, transverse cross relaxation can take place. Due to the faster transverse autocorrelated relaxation compared to longitudinal relaxation (see paragraph on Autocorrelated Relaxation, page 66), ROE signals decay more quickly compared to NOE signals for larger correlation times and therefore for bigger biomacromolecules. A further characteristic of the ROESY experiment is the possibility to distinguish signals arising from cross relaxation from those originating from chemical exchange due to their different sign.<sup>[148]</sup> Due to the zero crossing of the NOE this is complicated in NOESY spectra. For fast correlation times, cross peaks in a NOESY spectra can be separated in cross relaxation and chemical exchange but at correlation times above the zero crossing the sign is the same for both.



## Structure Determination of RNA-Ligand Complexes

The assignment of the RNA and ligand resonances in the complex is the requisite to obtain restraints for the structure calculation of the complex. These restraints include distance restraints such as hydrogen bonds and NOE contacts, torsion angle restraints derived from  $^3J$ -couplings and orientational restraints from residual dipolar couplings (RDC). In combination with molecule specific constraints from force fields (e.g. AMBER, CHARMM, GROMACS) structure calculations of molecules and complexes are performed.

### Determination of Structural Restraints

#### Hydrogen-Bond Distance Restraints

Detectability of RNA imino proton signals in 1D  $^1\text{H}$ -NMR-spectra is evidence for their involvement in hydrogen bonds, in most cases forming base pairs. The assignment gives complete information on the base pair type, e.g. Watson-Crick or G-U wobble (Figure 5). Watson-Crick base pairs are characterized by cross peaks in the HNN-COSY experiment from the imino protons guanine H1 and uridine H3 to the acceptor nitrogens cytidine N3 and adenine N1.<sup>[132]</sup> Further, G-C Watson-Crick base pairs display strong NOE cross peaks between the imino proton H1 and the amino protons H21 (guanine) and H41 (cytidine) which are hydrogen bonded to base-paired carbonyl oxygens. In A-U Watson-Crick base pairs NOE cross peaks of the imino H3 to the adenine protons H2 and H61 are detectable. In particular, the cross peak to H2 is significant since the adenine H2 does not exchange with the solvent. In the incremented dimension of the H2 the line width of the cross peak is therefore narrow and the peak stands out from the broad imino-amino proton cross peaks. G-U wobble base pairs are characterized by a strong NOE between the imino protons guanine H1 and uridine H3 and the clear upfield shift of both. Also, no cross peaks are detected for both in the HNN-COSY since the hydrogen bond acceptors are carbonyl oxygens. Unusual base pairs can be determined by further experiments: Hydrogen bonds of the

imino protons to other nitrogen acceptors can be detected in the HNN-COSY, since the resonances of the purine N1, N3 and N7 nuclei are in distinct regions (see Table 2). With a long range HNCO-experiment, where the delay for the magnetization transfer between the protonated nitrogen and the acceptor carbonyl-carbon is optimized, further hydrogen bonds are determined.<sup>[149, 150]</sup> In addition, the resonances of carbonyl-carbons indicate an involvement as hydrogen bond acceptor, since these resonances are downfield-shifted when the carbonyl-oxygen is an acceptor of a hydrogen bond.<sup>[151]</sup>

After determination of the base pairs the distances between the proton and the acceptor nucleus and between the donor and acceptor nuclei are defined as hydrogen bond restraints. Standard values for canonical base pairs are listed in Table 3.

*Table 3: Average distances between nuclei in hydrogen bonds of standard canonical Watson-Crick base pairs for structural calculations.<sup>[39]</sup> In general, the margin is set to 0.5 Å for the shorter limit and 0.3 Å for the longer limit.*

WC-base pair	nucleus 1	nucleus 2	average distance [Å]
<b>G-C</b>	guanosine N1	cytidine N3	2.91
	guanosine H1	cytidine N3	1.89
	guanosine O6	cytidine N4	2.72
	guanosine O6	cytidine H41 <sup>a</sup>	1.71
	guanosine N2	cytidine O2	3.08
	guanosine H21 <sup>a</sup>	cytidine O2	2.08
<b>A-U</b>	adenosine N1	uridine N3	2.95
	adenosine N1	uridine H3	1.93
	adenosine N6	uridine O4	2.83
	adenosine H61 <sup>a</sup>	uridine O4	1.82

<sup>a</sup> The nomenclature of the amino protons are defined in Table 2.

To retain base pair planarity, a constraint is implemented in the calculation for each base pair. Thereby, three heavy atoms from both nucleobases are

selected and these 6 nuclei are defined to lie in a plane. This avoids twisting, buckling and staggering of the nucleobases in respect to each other.

### NOE Distance Restraints

The most important input for the structure calculation are the distance restraints derived from 2D  $^1\text{H},^1\text{H}$ -NOESY (and ROESY) spectra<sup>[128, 129]</sup>, because of the large number of restraints that can be derived. Protons within  $\sim 6 \text{ \AA}$  distance can give rise to cross peaks, which possess volumes/intensities proportional to the inverse sixth power of the distance of the two protons. A considerable advantage is also the occurrence of intermolecular information combining the interacting molecules. The volumes/intensities are decreased by chemical or conformational exchange processes and are further influenced by spin diffusion. Spin diffusion is a relaxation process which correlates spin resonances of protons with larger distances via a proton correlated to both by a NOE. Therefore, comparison of NOESY spectra with different mixing times enable to avoid misinterpretations, since influences by spin diffusion occur considerably when using longer mixing times.

For a comparable conversion of volume/intensity into a distance throughout the whole spectral range a homogeneous excitation is essential. Therefore, to gain information from peaks close to the strong water signal, which counts for almost all ribose and  $\text{H}\alpha$  protons in RNA and peptides, respectively, a transfer into  $\sim 100\%$   $\text{D}_2\text{O}$  as solvent is necessary.

To translate the volumes/intensities into distances, an assignment and calibration of each spectra is necessary. This can either be done manually or automatically. Manually, the peaks are assigned with knowledge of all proton resonances and knowledge of the structural setup, e.g. secondary structure elements. For calibration the volumes/intensities of cross peaks between protons with known distances are chosen. In RNA, e.g., these can be the cross peaks from  $\text{H}5'$  to  $\text{H}5''$  ( $1.78 \text{ \AA}$ ) or  $\text{H}5$  to  $\text{H}6$  ( $2.46 \text{ \AA}$ ), since there is no conformational degree of freedom between them. With these starting points the other cross peaks are then grouped into distance ranges. Normally three

different categories are discriminated: strong ( $<2.5 \text{ \AA}$ ), medium ( $<4 \text{ \AA}$ ) and weak ( $<6 \text{ \AA}$ ). A minimum distance is not necessarily set, since signal reduction due to exchange may falsify the classification. A minimum of distances is defined by the force field constraints (bond length and van der Waals radii). With this method of restraints determination only a rough evaluation is possible. More strikingly, cross peaks that can not be assigned unambiguously may have to be excluded as restraints. This is particularly unfortunate for RNA, because of its small spectral distribution, especially of proton resonances within the ribose moieties. To avoid these problems of an estimated evaluation of the NOE-distance calibration and refusal of ambiguous NOE cross peaks an assignment and calibration in an iterative manner is helpful. This is the main benefit of the program ARIA (Ambiguous Restraints for Iterative Assignment).<sup>[152]</sup> Provided only with a resonance list, it can assign the cross peaks of a NOESY spectra, even to multiple spin-spin correlations with a weighting function. With a primary structure the volumes/intensities are calibrated and translated into distances. The assignment, weighting and calibration are then optimized in the process of the structure calculation. ARIA also can take into account influences on the NOE peak volumes/intensities arising from anisotropic diffusion and spin diffusion. With a first structural model and some unambiguously assigned cross peaks this automatic method enables to determine significantly more distance restraints and account for misleading side effects.

### Dihedral Restraints

RNA molecules possess a high number of flexible torsion angles along the backbone (angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\varepsilon$  and  $\xi$ ), within the sugar moiety (correlated by the sugar pucker) and the glycosidic bond angle  $\chi$ . This accentuates the necessity to determine dihedral restraints. The glycosidic bond angle  $\chi$  can be derived by different methods: either by measurement of heteronuclear  $^3\text{J}$ -couplings, e.g. from H1' to the carbons C2/C6 of pyrimidines or C4/C8 of purines<sup>[153-155]</sup> or by measurement of cross-correlated relaxation rates

between the dipole-dipole C1'H1' and the dipole-dipole C6H6/C8H8 ( $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  (pyrimidine) and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  (purine))<sup>[6]</sup> or the nitrogen CSA ( $\Gamma_{N1,C1'H1'}^{CSA,DD}$  (pyrimidine) and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  (purine)).<sup>[156]</sup> Since the nucleobase normally adopts only two regions, also a close look at the intraresidual NOESY cross peaks can help to differentiate between the *syn*- and *anti*-conformation. In the *syn*-conformation, strong cross peaks between the anomeric H1' and the aromatic H6/H8 protons can be detected, while the intraresidual cross peaks between the ribose H2' and H3' and the aromatic H6/H8 are smaller than in a canonical *anti*-conformation. Information on the sugar pucker can be derived from the forward-directed HCC-TOCSY-CCH-E.COSY experiment<sup>[136, 137]</sup> which is explained in the previous section on RNA assignment. Extraction of H1',H2'- and H3',H4'-coupling constants serve to assign the ribose pucker to a C2'-endo or C3'-endo conformation. This determination of the ribose pucker also restrains the backbone angles  $\gamma$  and  $\delta$  in the ribose-phosphodiester backbone. Additionally, a quantitative analysis of the HCP-experiments used in the RNA assignment can further classify the backbone angles, especially  $\beta$  and  $\varepsilon$ .<sup>[141]</sup> The scalar couplings between the phosphorous nucleus and the H3' and H5'/H5'' protons ( ${}^3J_{HP}$ ), as well as between the phosphorous nucleus and the two adjacent C4' carbon nuclei ( ${}^3J_{CP}$ ) can be converted into the angles  $\beta$  and  $\varepsilon$  with a Karplus parameterization.<sup>[153]</sup> The further backbone angles  $\alpha$  and  $\xi$  are determinable by measurement of the cross-correlated relaxation rates between the CH-dipolar coupling and the phosphorous CSA with knowledge of the angles  $\beta$  and  $\varepsilon$ .<sup>[157]</sup>

In double helical regions, an assumption of a canonical A-RNA conformation can be made if a specific NOE peak pattern is given (Table 4)<sup>[158]</sup> and the sugar pucker is C3'-endo. This assumption permits to define the backbone angles  $\alpha$ ,  $\beta$ ,  $\varepsilon$  and  $\xi$  to canonical regions without hesitation (Table 1).

Table 4: Intensities of NOE cross peaks between sequential and intraresidual nuclei in canonical A-RNA. The intensities are distinguished in weak (*w*), medium (*m*) and strong (*s*).

nucleus 1	nucleus 2	sequential (n+1)	intraresidual
H1'	H6/H8	w	w
H2'	H6/H8	s	w
H3'	H6/H8	m	m
H2'	H1'	w	s
H6	H5	w	s

### Orientational Restraints

Since distance restraints derived from NOE data only report on short- to medium-range distances, restraints lack for long stretched RNA sequences as in double helices. Without these long-range restraints the structure calculations of RNA biomacromolecules can hardly converge. Stems may therefore bend during the structural calculation, although they are linear in solution. In addition, regions with intrinsic conformational flexibility may prevent a restriction of the local conformation, due to the lack of distance or dihedral restraints which cannot be determined experimentally.

In this case, residual dipolar couplings (RDC) can improve the structure, since they provide angular information of individual I-S vectors in respect to the main magnetic field across the whole molecule.<sup>[159-162]</sup> In solution the dipolar interactions between the dipoles I and S normally average to zero due to the rapid molecular tumbling. By introduction of an aligning medium, e.g. phages, this averaging is hindered and a residual alignment (~0.1%) under steric and/or electrostatic influences is maintained which leads to residual dipolar couplings up to a few tens of Hertz. In RNA, RDCs between carbons or nitrogens and their attached protons are usually measured. Hereby the protonated carbons are more advantageous, due to their higher abundance and their wider angular distribution in respect to the global alignment tensor. In contrast, N-H vectors only exist in the nucleobase and mostly lie parallel to each other in a plane perpendicular to the main component of an axial symmetric alignment tensor which is often found for RNA hairpin structures.

These conditions decrease the orientational information that can be derived from N-H residual dipolar couplings.

RDCs can be determined in a HSQC where the decoupling between the two spins is switched off. During the chemical shift evolution period (direct or indirect dimension) scalar and dipolar couplings then evolve leading to a signal splitting. The difference between the splittings in a solution containing an alignment medium and an isotropic solution represents the RDC. The RDC can be converted into an angular correlation of the measured vector to the molecular alignment tensor, orienting the vectors in respect to each other and irrespective of their distances:

$$\text{RDC}_{\text{IS}} = -\left(\frac{\mu_0}{4\pi}\right) \frac{\gamma_I \gamma_S \hbar}{2\pi^2 r_{\text{IS}}^3} S_{zz} \left\{ \frac{1}{2} (3 \cos^2 \theta - 1) + \frac{1}{2} \eta \sin^2 \theta \cos 2\phi \right\} \quad (\text{Equation 6}),$$

where  $S_{zz}$  (main component) and  $\eta$  (asymmetry) are parameters of the alignment tensor and  $\theta$  and  $\phi$  are the polar angles relating the vector to the molecule-fixed principal axis system (PAS) of the tensor. The PAS can be determined primarily with a starting structure and under the assumption of mainly steric interactions between aligning medium and the molecule, e.g. with the program PALES.<sup>[163]</sup> For RNA, negatively charged filamentous bacteriophages (Pf1) are often used, which align mainly by steric influences but also electrostatic repulsions between the negative phages and the negative phosphodiester backbone play a role.<sup>[164-166]</sup> Due to the aromatic groups stacked in helical stems, RNA biomacromolecules possess a high diamagnetic susceptibility, which induces a field alignment of the molecule.<sup>[159, 160]</sup> The degree of alignment scales with the squared value of the magnetic field, so that measurements of the splittings at two distinct fields enables to predict the splitting at a field strength of zero, which corresponds to the scalar coupling. The two magnetic field strengths should be reasonably separated to assure a precise analysis. However, the reduced resolution at lower fields needs to be accounted for. And since the overall alignment is low, the resulting RDCs are only about one tenth of RDCs determined with an alignment medium. The scalar coupling of RNA imino

groups,  ${}^1J_{NH}$ , can additionally be calculated with the empirically determined equation 7, since the value is proportional to the imino chemical shift  $\delta_H$ .<sup>[167]</sup>

$${}^1J_{NH} = (1.21\text{Hz} / \text{ppm})\delta_H - 103.5 \pm 0.6\text{Hz} \quad (\text{Equation 7}),$$

## Structure Calculation

To determine a structure of a single molecule or a complex a minimization of a target function is required, which calculates the agreement between a conformation of the investigated system and the set of con- and restraints defined for this system. This minimization is performed by either of two major techniques: (i) metric matrix distance geometry (DG) and (ii) Cartesian or torsion angle restrained molecular dynamics (rMD). The metric matrix distance geometry technique is e.g. implemented in the programs DIG-II<sup>[168]</sup> and DIANA<sup>[169]</sup> and calculates the structures from a matrix of atomic distances. The second technique, rMD, is realized in the programs like AMBER<sup>[170]</sup>, CHARMM<sup>[171]</sup> and CNX/CNS<sup>[172]</sup>. In the case of rMD, structures are calculated from selected force field constraints and the experimental restraints with the aim to minimize the potential function,  $V_{tot}$ :

$$V_{tot} = V_{ff} + V_{NMR} \quad (\text{Equation 8})$$

with the potential  $V_{ff}$  including the force field parameters:

$$V_{ff} = \omega_{bond}V_{bond} + \omega_{angle}V_{angle} + \omega_{dihedral}V_{dihedral} + \omega_{improper}V_{improper} + \omega_{vdW}V_{vdW} + \omega_{electrostatics}V_{electrostatics} \quad (\text{Equation 9})$$

and the potential  $V_{NMR}$  containing the experimental restraints:

$$V_{NMR} = \omega_{H-bond}V_{H-bond} + \omega_{NOE}V_{NOE} + \omega_{J-coupling}V_{J-coupling} + \omega_{RDC}V_{RDC} \quad (\text{Equation 10})$$

The terms  $\omega_i$  herein describe the force constants that weight the corresponding potential. These constants are predefined in the chosen force field (e.g. AMBER<sup>[173]</sup>, GROMACS<sup>[174, 175]</sup>, OPLS<sup>[176, 177]</sup>) which comprehends parameters for bond lengths, bond angles, torsion angles, improper, van der Waals- and electrostatic interactions. The weighting factors for the



experimental potentials are defined empirically, as described above for the NOE distance restraints.

Different forms of the potential functions,  $V_i$ , can be selected. Mostly based on a biharmonic function, flat-bottom (or flat-well) or soft potentials are often used (Figure 19). The chosen boundaries of the respective parameter define the extent of the flat-bottom, while the upper limit may be restrained gently with a soft flat-bottom potential.

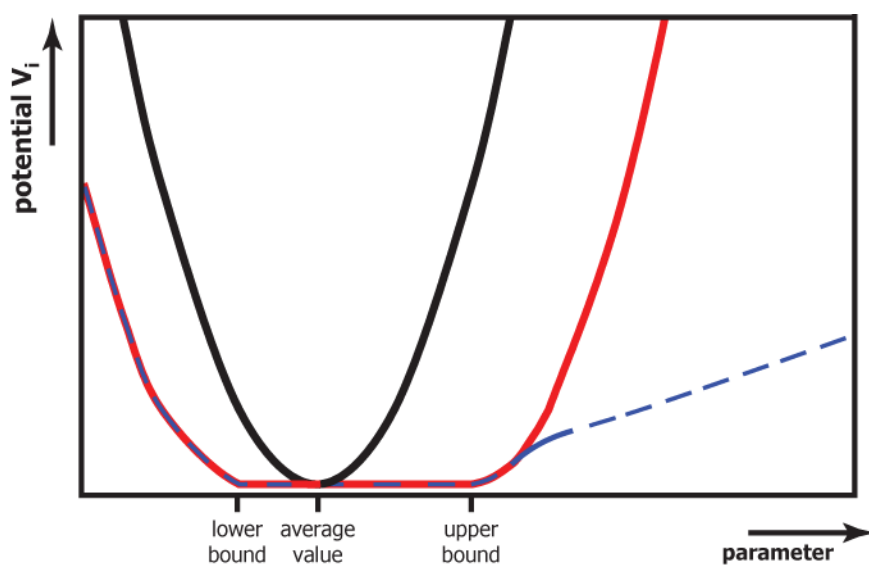


Figure 19: Graphical representation of possible potentials,  $V_i$ , used for con- and restraint parameters in structure calculations. The black line represents a biharmonic function, the red line a flat-bottom potential dependent on a lower and upper bound and the blue dashed line for soft flat-bottom potential.

In the following the programs ARIA and HADDOCK are shortly introduced which are recent developments based on the restrained molecular dynamics technique. The presented setup parameters are extracted from the TAR RNA/pyrimidinyl-ligand investigation.<sup>[5]</sup>

## ARIA

The program ARIA 1.2<sup>[152]</sup> is based on the program CNS 1.1 (Crystallography and NMR Systems)<sup>[172]</sup>. Its advantage is the use of ambiguous distance restraints which are optimized in an iterative manner. Starting from a

preliminary structure the NOESY cross peaks are calibrated and ambiguous assignments are weighted to the possible contributing distances. With these distances and the other non-distance restraints, CNS calculates a number of structures following a simulated annealing (SA) protocol with torsion angle dynamics (TAD). Starting at a high initial temperature (e.g. 10000 K) SA is performed for a defined number of steps. The simulation temperature is then reduced stepwise in three cooling stages (e.g. to 2000 K and 1000 K) to almost 0 K. In the distinct stages and iterations the contributions of the different restraints are weighted individually. After each iteration the lowest energy structures are analyzed to optimize the calibration and weighting of the ambiguous cross peaks for the following structure calibration.

### **HADDOCK**

If a structure of a biomacromolecule is known and the interactions with ligands do not change the structure globally, the program HADDOCK 2.1 (High Ambiguity Driven Docking)<sup>[178, 179]</sup> can be used to model this complex. It is a further development of ARIA and predicts possible interactions between the macromolecule and the ligand(s) based on the structures and ambiguous interaction restraints (AIRs). These AIRs are derived from biochemical and/or biophysical interaction data. Biochemical data can, for example, be extracted from mutagenesis experiments while NMR spectroscopy provides biophysical interaction data derived from chemical shift perturbations and/or from intermolecular NOEs. This interaction data is then introduced as ambiguous distance between defined sites on the interacting molecules to drive the docking process. For NOEs a manually classification as described above for the distance restraints is suitable. At first, arbitrary starting structures of the complex are generated from reliable structures of the macromolecule and the ligand(s) determined previously. A rigid docking is then performed at a high temperature (e.g. 2000 K) in iteration 1 followed by a first cooling stage to a temperature around the experimental setup (e.g. 300 K). Starting from this temperature a semi-flexible docking is allowed

during a cooling step to a lower temperature (e.g. 50 K) where the flexible regions on each molecule are user defined. In the final cooling stage down to a temperature of 0 K, more flexible residues are selected to allow local structural adjustments of the macromolecule and the ligand(s). The restraints introduced for the ARIA calculations before can also be implemented to maintain the overall structural integrity. This is especially important in the case of RNA complexes, since HADDOCK strongly uses electrostatic interactions for the complex assembly and repulsions along the phosphodiester backbone may lead to distortions within the RNA structure.

## Determination of global and local RNA dynamics

### Autocorrelated Relaxation

Knowledge of the structure cannot explain the high diversity in functions of RNA. Therefore, investigations of dynamics are necessary. Autocorrelated relaxation measurements provide information about local and global dynamics.

After a perturbation, e.g. by a radio frequency pulse, the magnetization of nuclear spins relax back to a Boltzmann-equilibrium. This relaxation can be divided into two types:

(i) Relaxation of the magnetization parallel to the main field along the z-axis (longitudinal relaxation) and (ii) dephasing of the coherence in the x-y plane (transverse relaxation).<sup>[180]</sup>

### Longitudinal Relaxation

Longitudinal relaxation implies a change in the energy of the spin system. This energy is released to the environment, in this special case called lattice which leads to the term of spin-lattice relaxation. Generally, it is an exponential decay with the rate and time constants  $R_1$  and  $T_1$ , respectively. A mono-exponential behavior is observed for isolated or weakly coupled spins, while in the case of strongly anisotropic molecular motions, cross relaxation or cross-correlated relaxation a non-exponential spin-lattice relaxation can occur. To measure the longitudinal relaxation mostly a series of inversion-recovery experiments is recorded (Figure 20). After inversion into  $-z$ -magnetization by a  $180^\circ$ -pulse the recovery delay  $\tau_M$  is varied to achieve relaxation time-dependent signal intensities  $I$  (or volumes). By fitting of the mono-exponential function in Equation 11 to the data points,  $T_1$  or its inverse  $R_1$  can be determined.

$$I = I_0 \cdot e^{-R_1 \tau_M} \quad (\text{Equation 11})$$

To suppress undesired influences by dipole-CSA cross-correlated and cross relaxation to other spins the observed spins need to be decoupled during the relaxation delay. The effects of global molecular anisotropy and cross-correlated relaxation need to be considered especially for larger macromolecules due to their dependence on the global correlation time  $\tau_c$ .

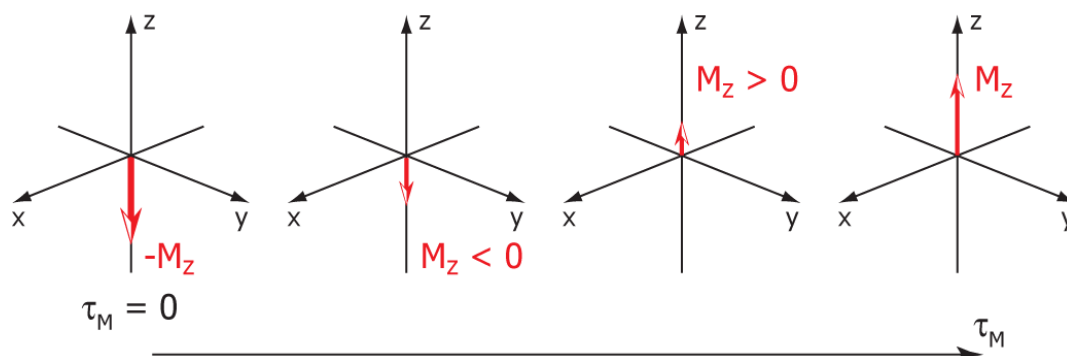


Figure 20: Scheme of the change in longitudinal magnetization from  $-M_z$  to the equilibrium magnetization  $M_z$  due to  $T_1$  relaxation.

### Transverse Relaxation

In contrast to longitudinal relaxation, transverse relaxation does not involve exchange of energy between the spin system and the lattice and therefore is adiabatic. It is a destructive interference between the individual spin vectors and leads to a dephasing of these magnetic vectors in the x-y plane resulting in a total loss of transverse magnetization (Figure 21). This mechanism is also called spin-spin relaxation and its rate and time constant are abbreviated with  $R_2$  and  $T_2$ , respectively. Parallel to the transverse dephasing also longitudinal relaxation leads to a return of magnetization back to the thermal equilibrium. Because  $R_2$  defines the lifetime of the observable transverse magnetization, it can be measured by determining the line width at half the maximum height with Equation 12. However, this rate is inaccurate due to the influence of field inhomogeneities.

$$\Delta\nu_{1/2} = \frac{R_2^*}{\pi} \quad (\text{Equation 12})$$

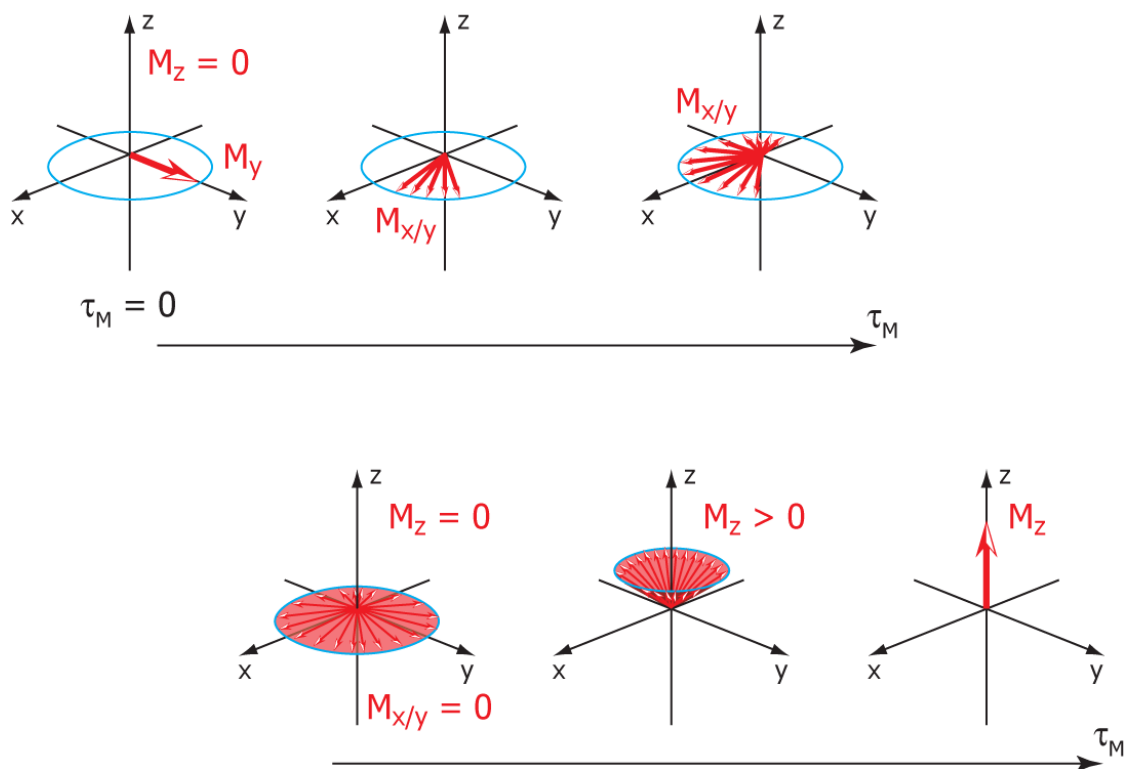


Figure 21: Scheme of the change in transversal magnetization from  $M_y$  to the equilibrium magnetization  $M_z$ . The fast transversal relaxation leads to a dephasing of the  $x/y$ -magnetization until no magnetization is left ( $M_{x/y} = M_z = 0$ ). The slower longitudinal relaxation recovers magnetization to the equilibrium  $z$ -magnetization.

Therefore, a Hahn echo experiment<sup>[181]</sup> should be used to eliminate these inhomogeneities. To suppress the influences of spin diffusion on the relaxation rate determination, the echo sequence can be replaced by a CPMG (Carr-Purcell-Meiboom-Gill) pulse train.<sup>[133]</sup> By variation of the relaxation delay relaxation time-dependent intensities and volumes are measured and can be fitted to a mono-exponential decay function as in Equation 11. The extracted relaxation rate  $R_2$  also contains a contribution from chemical exchange processes sensed by the spin. This rate is abbreviated with  $R_{ex}$  and will be discussed in more detail later.

### Longitudinal Relaxation in a Rotating Frame

Transverse relaxation may alternatively be determined by measurement of  $T_1$  with a spin-lock sequence in a rotating frame. After the magnetization is flipped into the transverse plane a strong radio frequency field  $B_1$  locks the magnetization along the axis this field is applied from. This spin-lock creates a new effective field noticed by the spin, oriented in an angle  $\theta$  with respect to  $B_0$ :

$$\tan \theta = \frac{B_1}{\frac{\Delta\omega}{\gamma}} \quad (\text{Equation 13}),$$

where  $\Delta\omega$  is the offset of the applied spin-lock. The resulting relaxation rate  $R_{1\rho}$  contains contributions from both relaxation types, the longitudinal  $R_1$  and the transverse  $R_2$ :

$$R_{1\rho} = R_1 \cos^2 \theta + R_2 \sin^2 \theta \quad (\text{Equation 14})$$

$R_2$  determined in this approach contains much less unwanted influences, such as contributions from exchange ( $R_{ex}$ ) and therefore represents a relaxation rate closer to the spin specific transverse relaxation rate. Unwanted influences originate from the experimental disturbances. However, two assumptions need to be considered: the spin-lock field has to be much weaker than the external magnetic field and the following inequation needs to be satisfied:

$$\omega_1^2 \tau^2 \ll 1 \quad (\text{Equation 15}),$$

where  $\omega_1$  is the amplitude of the spin-lock field and  $\tau$  is the correlation time. This condition is always the case in liquid state NMR due to the rapid tumbling motions on the subnano- and nanosecond timescale.

### Sources of Spin Relaxation

Although spin relaxation is always observed in NMR, the probability of an isolated magnetic dipole to undergo a spontaneous transition from the upper to the lower energy state is extremely low.

$$W = \frac{\mu_0}{4\pi} \frac{2\hbar\gamma^2\omega^3}{3c^3} \quad (\text{Equation 16})$$

Solving Equation 16 with  $\mu_0$  being the permeability of vacuum,  $\hbar$  the Planck constant,  $\gamma$  the gyromagnetic ratio and  $\omega$  the Larmor frequency for a proton in a 600 MHz magnet results in a transition probability of  $W \approx 10^{-21} \text{ s}^{-1}$ . Therefore, spontaneous transitions are ineffective for a relaxation process. An interaction with the spin environment (lattice and other spins) is necessary to enhance relaxation. This interaction takes place via fluctuations of the external magnetic field sensed by the nuclear spin. Two spin properties mainly contribute to the relaxation process of spins in biomacromolecules in liquid state NMR: the chemical shift anisotropy (CSA) and dipolar interactions. While the CSA originates from the inhomogeneous distribution of the electrons surrounding the nucleus, the dipolar interactions evolve between adjacent nuclear spins. Both properties do not affect the spin resonances in liquid-state, since their angular dependence with respect to the main field cancel in an isotropic environment. However, the magnetic field fluctuations arising from molecular tumbling in solution and local dynamics influence these anisotropic spin properties and therefore induce spin relaxation. Further interactions, e.g. arising from a quadrupolar nature of nuclei or from unpaired electrons, can be neglected in relaxation studies of biomacromolecules. Hence, measurement of autocorrelated relaxation rates can in first approximation be assigned to dipolar and CSA relaxation processes. This assumption simplifies the extraction of global and local dynamic information.

Magnetic field fluctuations can be converted into a correlation function with a specific parameter, the correlation time  $\tau$ . The link between the relaxation rates and the dynamics-dependent correlation time are the Lorentian-shaped spectral density functions  $J(\omega)$ . In case of an isolated spin the following equations are valid for slow molecular tumbling:



- Spin-lattice (longitudinal) relaxation

$$R_1 = 2\gamma^2 J(\omega_0) = 4\gamma^2 b^2 \frac{\tau}{1 + \omega_0^2 \tau^2} \quad (\text{Equation 17})$$

- Spin-spin (transverse) relaxation

$$R_2 = \gamma^2 \left( J(0) + \frac{1}{2} J(\omega_0) \right) = 2\gamma^2 b^2 \left( \tau + \frac{\tau}{1 + \omega_0^2 \tau^2} \right) \quad (\text{Equation 18})$$

- Spin-lattice relaxation in the rotating frame

$$R_{1\rho} = \gamma^2 \left( J(\omega_1) + \frac{1}{2} J(\omega_0) \right) = 2\gamma^2 b^2 \left( \frac{\tau}{1 + \omega_1^2 \tau^2} + \frac{\tau}{1 + \omega_0^2 \tau^2} \right) \quad (\text{Equation 19}),$$

where  $b^2$  contains parameters relevant for relaxation (e.g. CSA and dipolar interactions) and  $\omega_1$  is the amplitude of the spin-lock field.

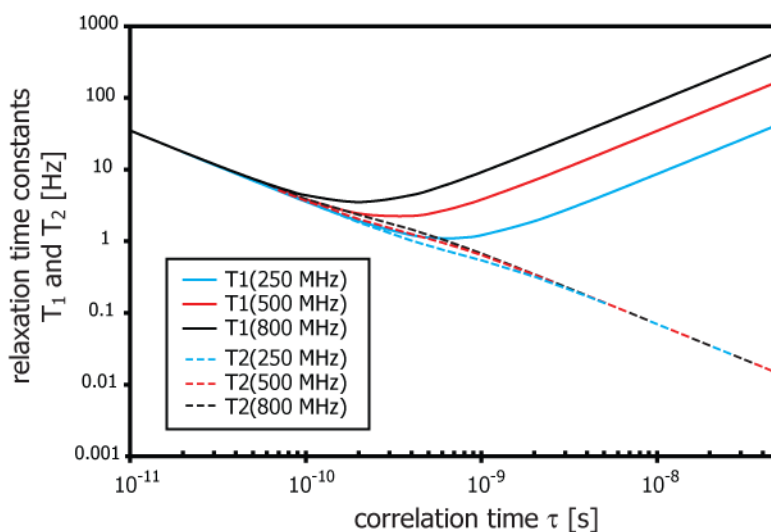


Figure 22: Dependence of the proton relaxation times  $T_1$  and  $T_2$  on the correlation time  $\tau$  at the Larmor frequencies of 250 MHz, 500 MHz and 800 MHz assuming a fluctuating magnetic field with an amplitude  $b^2 = 10^{-8} T^2$ .

## Relaxation of a Spin S in an I-S Spin System

As introduced above, the relaxation of a spin S in an isolated I-S spin system is described in a first approximation by the I-S dipolar interaction and the CSA of the nucleus S. In biomacromolecules,  $^{15}\text{N}$ - $^1\text{H}$  or  $^{13}\text{C}$ - $^1\text{H}$  groups represent

these isolated spin systems. Measurement of the autocorrelated relaxation rates of the heavy atoms provide numerous reporters for the determination of local and global flexibility of the macromolecule. Equations 17 and 18 can be expanded for I-S spin systems and correlate the relaxation rates to the spin density functions at discrete frequencies,  $J(\omega)$ :

- Spin-lattice (longitudinal) relaxation

$$R_1 = \frac{d^2}{4} [J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I + \omega_S)] + c^2 J(\omega_S) \quad (\text{Equation 20})$$

- Spin-spin (transverse) relaxation

$$R_2 = \frac{d^2}{8} [4J(0) + J(\omega_I - \omega_S) + 3J(\omega_S) + 6J(\omega_I) + 6J(\omega_I + \omega_S)] \\ + \frac{c^2}{6} [4J(0) + 3J(\omega_S)] + R_{ex} \quad (\text{Equation 21})$$

with the parameters for the dipolar interaction  $d = \frac{\mu_0 \hbar \gamma_S \gamma_I}{8r_{IS}^3 \pi^2}$  and for the CSA interaction  $c = \frac{\omega_0 \cdot \Delta\sigma_S}{\sqrt{3}}$ , where  $r_{IS}$  is the bond distance between the nuclei I and S and  $\Delta\sigma_S$  is the chemical shift anisotropy of nucleus S.

A third independent relaxation parameter, which can be measured in an I-S spin system, is the steady-state heteronuclear Overhauser effect (hetnOe). The hetnOe is a cross relaxation phenomenon depending on the dipolar relaxation mechanism. By saturation of the spin I for a sufficient period of time the starting magnetization of the bound spin S is influenced through the hetnOe. This leads to an increase or decrease of detected magnetization:

$$\text{hetnOe} = \left( \frac{d^2}{4R_1} \right) \left( \frac{\gamma_S}{\gamma_I} \right) [6J(\omega_I + \omega_S) - J(\omega_I - \omega_S)] \quad (\text{Equation 22})$$

The ratio of volumes (or intensities) of the signals of spin S measured with and without presaturation results in the NOE enhancement factor  $\varepsilon_{NOE} = 1 + \text{hetnOe}$  (Equation 23).

## Derivation of Dynamical Information from Autocorrelated Relaxation Rates

Autocorrelated relaxation rates are indicative for dynamics on the pico- to nanosecond timescale. Within this time regime, motions like vibrations, methyl group rotations, bond librations, side chain rotations and small conformational rearrangements take place (Figure 8).<sup>[180, 182-187]</sup> As seen in the equations 17, 18 and 19 the correlation time of the corresponding spin has an essential influence on the relaxation rates. The correlation time  $\tau$  of a spin can be separated into the global correlation time of the whole molecule ( $\tau_c$ ) and into an internal correlation time ( $\tau_e$ ), provided that these motions are independent from each other:

$$\frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e} \quad (\text{Equation 24})$$

### Global Dynamics

This global correlation time can be converted into the rotational diffusion coefficient  $D_R$  which for a spherical molecule is as following:

$$\frac{1}{\tau_c} = 6D_R \quad (\text{Equation 25})$$

From the Stokes-Einstein-Debye hydrodynamic approach,  $\tau_c$  can be converted into further global parameters, such as the molecular volume  $V$ , or a temperature dependence can be extracted.

$$\tau_c = \frac{V\eta}{kT} \quad (\text{Equation 26}),$$

where  $\eta$  is the solvent viscosity. In the case of axial symmetric or anisotropic tumbling, these equations become extended and are therefore used from programs, like HYDRONMR<sup>[188]</sup>, to calculate relaxation rates from a given 3-dimensional structure.

For a spheric biological macromolecule, the correlation time is defined as the time taken for the molecule to rotate one radian ( $\sim 60^\circ$ ) or move a distance of the order of its own dimension. The trimmed mean value of the ratio of the relaxation rates can serve as rough estimate for the global correlation time of the whole molecule by Equation 27, if the individual spins do not undergo increased local dynamics and the molecule tumbles isotropically in solution.

$$\frac{R_2}{R_1} \approx 1 + \frac{1}{2} \omega_S^2 \tau_c^2 \quad (\text{Equation 27})$$

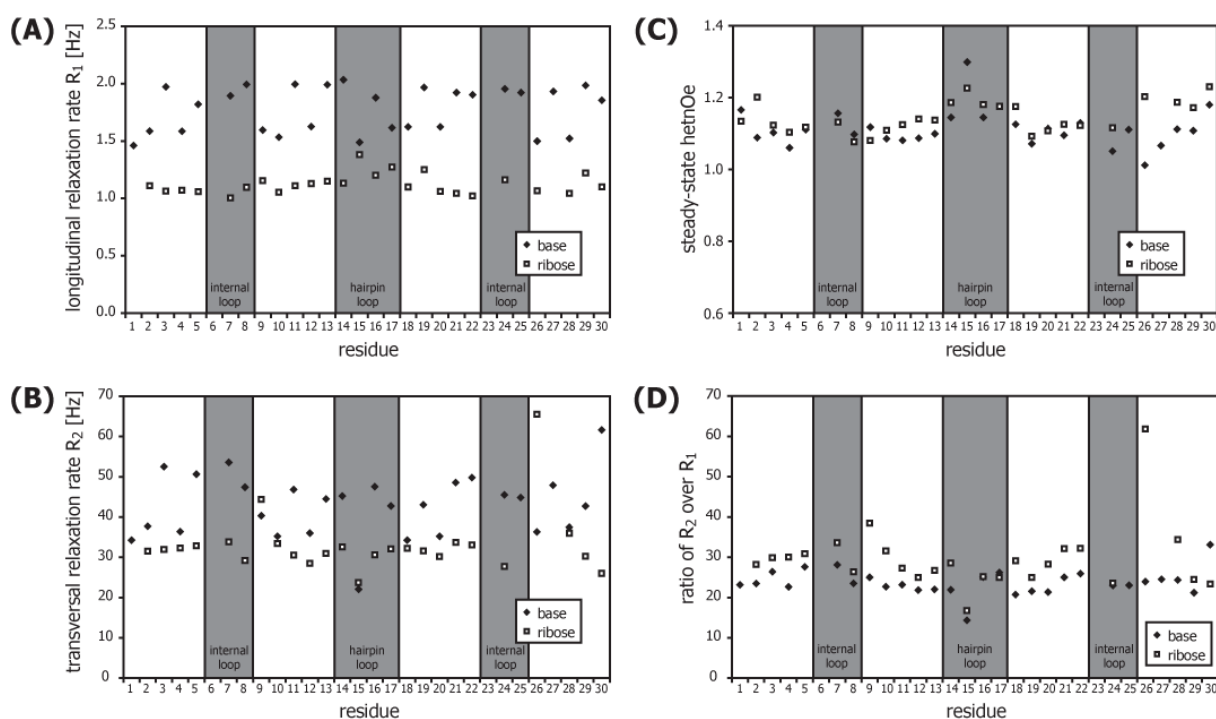


Figure 23:  $^{13}\text{C}$  autocorrelated relaxation rates  $R_1$  (A),  $R_2$  (B), the  $^{13}\text{C}$   $\{^1\text{H}\}$ -steady-state heteronuclear NOE (C) and the ratio  $R_2/R_1$  (D) of the aromatic C6/C8 (filled diamonds) and aliphatic C1' (open squares) nuclei in the SLD hairpin (Figure 12). The regions of the internal and the hairpin loop are shaded gray.

## Local Dynamics

Spins which are subject to large amplitude motions on time scales faster than the global correlation time  $\tau_c$  yield smaller  $R_2/R_1$ -ratios (e.g. residue 15 in Figure 23 D), especially because of a reduced  $R_2$ . Significantly larger ratios

are observed for spins undergoing chemical or conformational exchange processes (e.g. residue 26 in Figure 23 D), since the transverse relaxation rate contains the exchange parameter  $R_{ex}$ . An increased ratio also can indicate rotational diffusion anisotropy for bond vectors aligned almost collinear with the unique axis of the diffusion tensor. Therefore the plot of the individual ratios along the sequence gives a first indication for rigid and dynamical sites in the molecule.

Determination of the hetnOe from a spin S directly gives further indications of the dynamics of the considered I-S vector. While spins with correlation times similar to the global correlation time in the nanosecond regime sense only small nuclear Overhauser effects from the spin I, the effect increases with shorter correlation times and consequentially with faster motions of the dipole-dipole vector (terminal residues and residue 15 in Figure 23 C). While  $R_1$  and  $R_2$  do not display an easy to interpret graph, especially for the nucleobases where the large and varying CSAs scramble the data, the hetnOe and the ratio  $R_2/R_1$  give first indications of dynamics: the second hairpin loop residue (A15) has a big hetnOe and a small ratio, indicating increased local dynamics of this structurally unrestricted residue. The significantly increased ratios for the ribose G26 and the 3'-terminal nucleobase C30 are indications for a chemical exchange contribution to the transverse relaxation rate (Figure 23).

To extract a more pictorial presentation of the local dynamics from the autocorrelated relaxation rates, Lipari and Szabo presented the model-free formalism in 1982.<sup>[182, 183]</sup> This formalism assumes a form of the correlation function without solving an equation of motion. For a molecule tumbling isotropically with  $\tau_c$  the corresponding spectral density function  $J_{LS}(\omega)$  for an isolated I-S spin system is as follows:

$$J_{LS}(\omega) = \frac{2}{5} \left( \frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S^2) \tau}{1 + \omega^2 \tau^2} \right) \quad (\text{Equation 28})$$

Without knowledge of an equation of motion, one cannot extract the real local movements of the investigated spin from this analysis but the parameters  $S^2$  and  $\tau_e$ .  $S^2$  is called the order parameter, which can adopt a value between 0 and 1 correlated to the amplitude of the local motions. For  $S^2 = 1$  the spectral density function reduces to a simple Lorentzian function with the global correlation time  $\tau_c$ , implying no local dynamics of the I-S spin vector. An order parameter  $S^2 = 0$  signifies a totally flexible vector which is moving independently from the global molecule. For example methyl groups come close to this value due to their mostly unrestricted rotation. The internal correlation time  $\tau_e$  is connected to the effective correlation time  $\tau$  by Equation 24. It represents the time scale for the underlying local dynamics.  $\tau_e$  and  $\tau_c$  need to be independent for the model-free formalism to be applicable. Therefore it is necessary to pay attention that the internal correlation time is at least a magnitude smaller, e.g. the motions are faster, than the global  $\tau_c$ . With this model-free formalism it is also possible to extract  $R_{ex}$ , a relaxation rate caused by chemical and/or conformational exchange, since it is incorporated in the transverse relaxation. But also field inhomogeneities or a wrong definition of the global anisotropy contribute to this rate.

In 1990, Clore and coworkers<sup>[189]</sup> extended the formalism to further differentiate the time scales of motions by separating the general order parameter  $S^2$  in a slow and a fast contribution:

$$S^2 = S_s^2 * S_f^2 \quad (\text{Equation 29})$$

These order parameters,  $S_s^2$  and  $S_f^2$ , need to account for independent dynamics and therefore should be separated in their timescale by at least one order of magnitude. The following spectral density function is derived:

$$J_{LS}(\omega) = \frac{2}{5} \left( \frac{S^2 \tau_c}{1 + \omega^2 \tau_c^2} + \frac{(1 - S_f^2) \tau_f}{1 + \omega^2 \tau_f^2} + \frac{(S_f^2 - S^2) \tau}{1 + \omega^2 \tau^2} \right) \quad (\text{Equation 30})$$

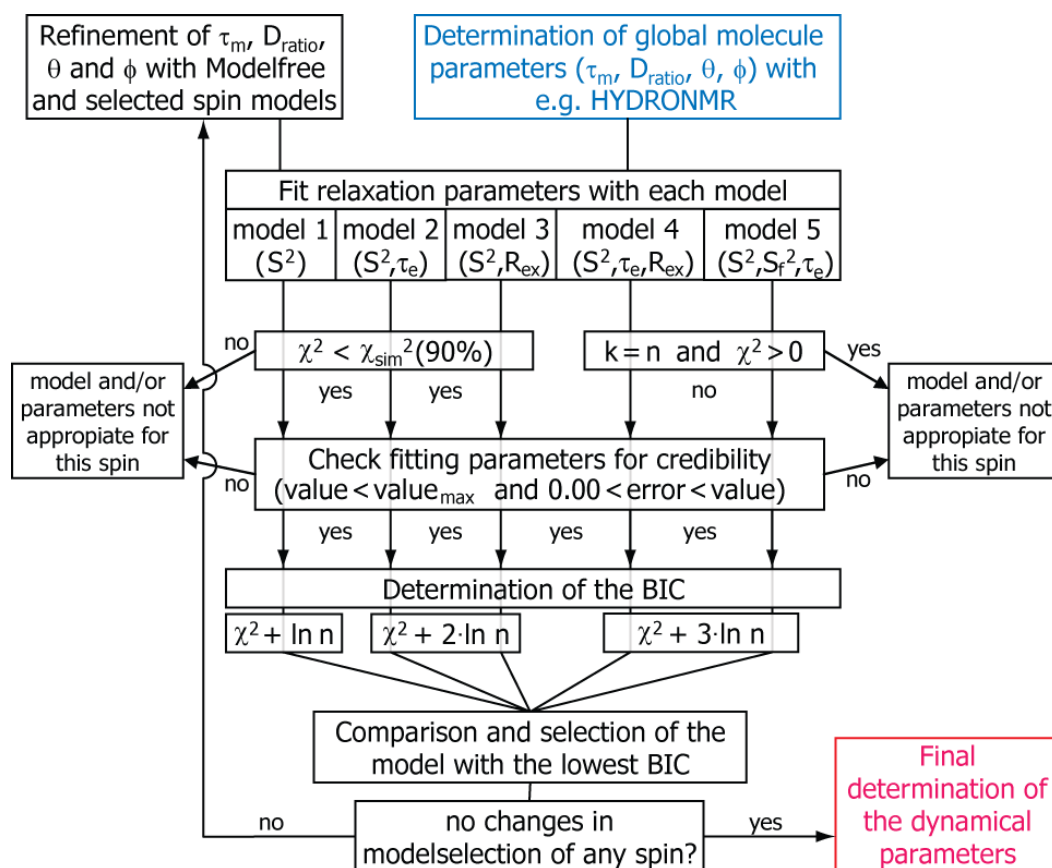


Figure 24: Flowchart of the model selection process to extract global and internal dynamical parameters from autocorrelated relaxation rates following the model-free formalism<sup>[182, 183]</sup> with the program Modelfree 4.x.<sup>[185, 191]</sup> Arrows, labeled with **yes** and **no**, display branching points in the flowchart. At first (blue box) a determination is necessary of the global molecule parameters: the global correlation time  $\tau_m$ , the axial symmetry  $D_{ratio}=D_{||}/D_{\perp}$  and the polar angles  $\theta$  and  $\phi$ , which convert the diffusion tensor into the coordinate frame of the pdb-file. These can be determined e.g. computationally with HYDRONMR<sup>[188]</sup>. These parameters are set fixed and each spin is fitted with all 5 models. The results need to be checked for each spin individually. First the square deviation  $\chi^2$  (Equation 31) needs to be lower than  $\chi_{sim}^2$  from 90% of an adequate number of simulated data points. The fitted internal parameters, the order parameter  $S^2$ , the internal correlation time  $\tau_e$  and the exchange rate  $R_{ex}$ , need to be credible. The value should not hit the defined maximum value and its error should lie above zero and clearly below the actual value. If the number of fitted parameters ( $k$ ) is equal to the number of input data points ( $n$ ), which is often given for models 4 and 5,  $\chi^2$  must be zero. If these conditions are not fulfilled, the model and/or the fixed parameters are not appropriate for this spin. With credible results the  $\chi^2$  is converted to the Bayesian Information Criteria value (BIC), which takes into account the number of data points ( $n$ ).<sup>[192]</sup> For each spin the model is selected with the lowest BIC. With a new set of defined models, the global parameters can be refined with Modelfree and a next model selection iteration begins. If no changes in the model selection occur in respect to the previous iteration, a final determination of the dynamical parameters follows (red box).

To convert the experimental data, the relaxation rates  $R_{1i}$ ,  $R_{2i}/R_{1\rho}$  and the  $hetnOe$  can be fitted by five different models to extract the order parameter and up to two further parameters.<sup>[190]</sup> The easiest model, *model 1*, only results in an order parameter, with the corresponding internal correlation time assumed to be very fast ( $< 10$  ps). With a slower internal motion *model 2* should be applicable which assigns a  $S^2$ - and a  $\tau_e$ -value to the I-S vector. In the *model 3* I-S vectors with fast internal motions and contributions of  $R_{ex}$  are best fitted. *Model 4* is a complex extension of the previous two models, resulting in a  $S^2$ -, a  $\tau_e$ - and a  $R_{ex}$ -value. If the local motions can be further separated into a fast and a slow motion, then with equation 30 the *model 5* should fit the data best, resulting in a  $S^2$ -, a  $\tau_e$ - and a  $S_f^2$ -value. The  $\tau_e$ -value represents the internal correlation time of the slow motion.

With increase in complexity of the models (*model 1* < *model 2/3* < *model 4/5*) more fittable parameters are used. To distinguish that the better fitting with a more complex model does not arise from the increased number of parameters and is significant, statistical (a mathematical F-Test) or probabilistical (the Akaike or the Bayesian information criteria, abbreviated AIC or BIC, respectively) analyses are necessary. D'Auvergne and Gooley showed in 2003 that the probabilistic approaches are better for the model-free analysis of NMR relaxation data, since the statistical approach<sup>[191]</sup> tends to underestimate the complexity in the model selection step.<sup>[192]</sup>

Figure 24 displays the flowchart to determine the fitting model for each spin with the program *Modelfree* from the lab of Art Palmer<sup>[185, 191]</sup>. Model selection is accomplished with the Bayesian Information Criteria (BIC) which is derived from the square deviation  $\chi^2$ :

$$\chi^2 = \sum_{i=1}^M \left\{ \frac{(R_{1i} - R_{1i}^{fit})^2}{\sigma_{R_{1i}}^2} + \frac{(R_{2i} - R_{2i}^{fit})^2}{\sigma_{R_{2i}}^2} + \frac{(NOE_i - NOE_i^{fit})^2}{\sigma_{NOE_i}^2} \right\} \quad (\text{Equation 31})$$

$R_{1i}$ ,  $R_{2i}$  and  $NOE_i$  are the relaxation parameters from the  $i$ /th magnetic field  $M$ ;  $R_{1i}^{fit}$ ,  $R_{2i}^{fit}$  and  $NOE_i^{fit}$  are the corresponding fitted values; and  $\sigma_{R_{1i}}$ ,  $\sigma_{R_{2i}}$  and  $\sigma_{NOE_i}$  are the experimental uncertainties in the relaxation parameters.



In addition, a precise knowledge of the bond length of the I-S vector and the chemical shielding anisotropy CSA is necessary to extract authentic dynamical parameters<sup>[66]</sup>, since the dipolar and the CSA relaxation are the main contributors to spin relaxation (Equations 20, 21 and 22). The normally-assumed bond lengths,  $R$ , need to be slightly increased to a zero-point motion averaged value,  $r_{eff}$ , to extract the right contribution to the rates by the dipolar relaxation mechanism (Table 5).<sup>[193, 194]</sup>

Table 5: Selected C-H bond lengths in nucleotide moieties used in earlier studies for relaxation analysis ( $R$ ) in RNA<sup>[66, 195]</sup> and zero-point motion averaged bond lengths which give an effective bond length  $r_{eff}$ .<sup>[193, 194]</sup>

moiety	bond	R [Å]	$r_{eff}$ [Å]
ribose	C1'-H1'	1.095	1.118
pyrimidine	C6-H6	1.082	1.102
purine	C8-H8	1.079	1.102
	C2-H2	1.079	1.102

In former model-free analyses of relaxation rates, the CSA was accepted as constant for nuclei in configurationally similar environments, such as the amide nitrogens in amino acid chains. But recently, it could be demonstrated that the amide nitrogen CSA is site-specific and can vary by over 10 %.<sup>[196]</sup> Also for nucleotides, <sup>13</sup>C CSAs determined by liquid-crystal NMR on helical A-form RNA<sup>[197]</sup> differed significantly from the previously used values which were either calculated on model compounds<sup>[198]</sup> or derived by solid-state NMR on isolated nucleotides<sup>[199]</sup>. This discrepancy (Table 6) displayed the necessity to use nucleobase-specific values to extract reasonable dynamical parameters. In addition, it was published, that the conformation of the nucleobase in respect to the ribose moiety, *syn* or *anti*, has a significant influence on the CSA of the surrounding carbon and nitrogen nuclei.<sup>[154, 200]</sup>

Table 6: Chemical shielding anisotropies (CSA) of aromatic nuclei in nucleobases. The old values were determined by DFT calculations<sup>[198]</sup> and solid-state NMR<sup>[199]</sup>, while the new values were extracted from liquid-state NMR experiments.<sup>[197]</sup>

nucleobase	carbon nucleus	CSA [ppm]	
		old	new
adenine	C8	-134	-144
	C2		-168
guanine ( <i>anti</i> -conformation)	C8	-134	-133
guanine ( <i>syn</i> -conformation)	C8		-122
cytosine	C6	-179	-208
uracil	C6	-179	-208

The experimental determination of the CSA depends on the assumed bond length, for which reason the newly derived values need to be used in combination. For the C1' nucleus so far no new derivation of the CSA has been performed taking the zero-point motion averaged bond lengths into account. Therefore, the old bond lengths should be used with a CSA of 45 ppm which is the middle of the published CSA values. These CSAs ranged from 29 to 60 ppm and were determined by DFT calculations on ribose and deoxyribose model compounds<sup>[198, 201]</sup>, solid-state NMR measurements on isolated nucleotides<sup>[199]</sup> and liquid-crystal NMR experiments on helical A-RNA.<sup>[202]</sup> All investigations showed no measureable differences between the four different nucleotides.

## CHAPTER V

**THE  $\Psi$ -RNA HEXAPEPTIDE LIGAND****Research Article:** Identification of Peptide Ligands for Target RNA Structures Derived from the HIV-1 Packaging Signal  $\Psi$  by Screening Phage-Displayed Peptide Libraries

Anette Pustowka, Julia Dietz, [Jan Ferner](#), Michael Baumann, Margot Landersz, Christoph Königs, Harald Schwalbe, and Ursula Dietrich

*ChemBioChem* **2003**, 4, 1093-1097

This publication describes the determination of peptidic ligands with affinity for the HIV-1 packaging signal  $\Psi$  and its single stemloops by a phage-display assay. The screening resulted in the consensus sequence HWWPWW, where the high number of tryptophan residues attracted attention. The binding specificity was confirmed and the affinity measured by ELISA and CD spectroscopy. An exchange of each residue with an alanine (ALA-scan) extracted the binding contributions individually.

The phage-display screening and ELISA studies were accomplished in the group of Dr. Dietrich. The author of this thesis performed all CD-spectroscopical measurements and their analysis. The corresponding paragraphs within the discussion and the methodological explanations were composed by the author.



## Identification of Peptide Ligands for Target RNA Structures Derived from the HIV-1 Packaging Signal $\psi$ by Screening Phage-Displayed Peptide Libraries

Anette Pustowka,<sup>[a]</sup> Julia Dietz,<sup>[a]</sup> Jan Ferner,<sup>[b]</sup> Michael Baumann,<sup>[b]</sup> Margot Landersz,<sup>[a]</sup> Christoph Königs,<sup>[a]</sup> Harald Schwalbe,<sup>[b]</sup> and Ursula Dietrich\*<sup>[a]</sup>

RNA molecules have the capacity to adopt complex three-dimensional structures, which allow specific recognition of proteins or RNA ligands. This property enables RNA molecules to perform essential regulatory and catalytic functions in the cell; these include modulation of gene activities, RNA transport, protein biosynthesis, splicing, viral genome packaging, and many more.<sup>[1]</sup> Consequently, RNA molecules have also become an attractive target for therapeutic interventions, especially since sequence information of the human genome combined with modern chip technologies allows the association of defined RNA molecules with specific pathological alterations.

RNA molecules mostly exert their regulatory functions by specific interactions with proteins. Clearly, posttranscriptional base modifications play an important role in protein recognition, as shown for the specific interaction of tRNA molecules with aminoacyl-tRNA synthetases.<sup>[2]</sup> In the last few years, however, X-ray crystallography and NMR spectroscopy have provided additional insights into some rules governing specific RNA–protein recognition.<sup>[3]</sup> According to these analyses, structural features of RNA molecules like A-form helices, single-stranded loops of a hairpin, or bulged regions are also important elements for protein recognition.<sup>[4]</sup> Arginine-rich motifs bind to RNA structures by adopting different conformations within different complexes.<sup>[5]</sup> Other specific contacts between single-stranded RNA loops and proteins are often mediated by aromatic amino acids. In any case, induced fit of either one or both binding partners is very frequently observed in RNA–protein recognition and is mechanistically important for biological regulation.<sup>[6–9]</sup>

In terms of therapeutic interventions, one approach to identify molecules that interfere with specific RNA–protein interactions is to select those molecules from complex chemical or molecular

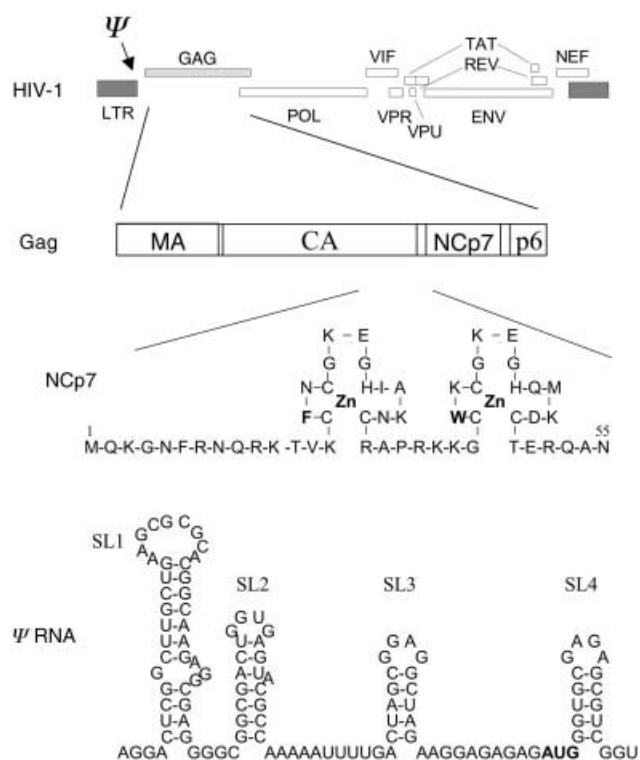
libraries. The huge structural space of RNA phenotypes is exploited in screening procedures in which RNA aptamers are selected and amplified over multiple rounds as the best-fitting ligands for a target molecule. The name SELEX (systemic evolution of ligands by exponential enrichment) has been coined for this procedure.<sup>[10]</sup> Other approaches use a given RNA structure as a target to select for specific ligands from compound libraries.<sup>[11, 12]</sup>

Phage-displayed peptide libraries have been widely used to select peptide ligands for antibodies, protein domains, or peptides.<sup>[13]</sup> Since the selection of peptide ligands is based on the structural recognition of the peptide motifs presented at the phage surface, phage-display technology should also be suited for the selection of peptide ligands for RNA structures. In fact, some groups have already used phage-display technology for this purpose: the U1 small nuclear RNA has been a target to select proteins or antibody binding fragments (Fab) from phage libraries.<sup>[14–17]</sup> Tailored phage libraries expressing zinc finger motifs or arginine-rich motifs were generated to select protein domains binding to the Rev responsive element stem-loop IIB (RRE-IIB) from HIV-1 and the 5S rRNA<sup>[18, 19]</sup> or the Tat transactivating region (TAR).<sup>[20, 21]</sup>

The aim of this study was to select specific peptide ligands for the RNA packaging structure  $\psi$  of HIV-1. The  $\psi$  RNA is a highly structured region located at the 5'-untranslated end of unspliced HIV-1 RNA molecules. The  $\psi$  RNA is responsible for specific encapsidation of two viral RNA genomes into the virions during their assembly process at the plasma membrane.<sup>[22]</sup> Specific recognition of viral genomic RNAs is mediated by the interaction of the  $\psi$  RNA with the viral Gag polyprotein, in particular, with the nucleocapsid protein NCp7 containing two zinc fingers.<sup>[23, 24]</sup> The entire  $\psi$  region consists of about 120 nucleotides predicted to form four individual stem-loop structures, SL1–SL4 (Figure 1). Several additional functions are also located within the  $\psi$  structure like the dimerization sequence (DIS) in SL1, the splice donor site (SD) in SL2, and the Gag initiation codon preceding SL4. Although all four stem-loops are involved in the RNA encapsidation process, SL3 is the major packaging signal, as it is capable of direct packaging of heterologous RNAs into virus-like particles.<sup>[25]</sup> A number of NMR studies have proven the specific interaction of SL3, and also of the other SL structures, with the NCp7 protein.<sup>[26–28]</sup> The key structural determinants are interactions between aromatic residues of the zinc fingers and the SL3 RNA sequence.<sup>[29–31]</sup> A detailed study combining biochemical analysis and NMR spectroscopy of SL3 has shown the specific interaction of full-length Gag with the GGAG tetraloop as well as with a purine-rich internal loop of SL3.<sup>[32]</sup> Furthermore, Gag binding and polymerization leads to progressive unwinding of the SL3 secondary structure.<sup>[33]</sup> As packaging of viral RNA genomes into virions is an essential step in the viral replication cycle, this process should also be suited for antiviral interventions. In fact, antisense RNA directed against the  $\psi$  region and inhibitors for the zinc finger protein NCp7 have shown antiviral activity.<sup>[34, 35]</sup> Our approach focuses on the selection of peptide ligands from phage-displayed peptide libraries, which specifically bind to the complete  $\psi$  RNA or to single stem-loops thereof.

[a] Dr. U. Dietrich, A. Pustowka, J. Dietz, M. Landersz, C. Königs  
Georg-Speyer-Haus  
Institute for Biomedical Research  
Paul-Ehrlich-Strasse 42–44  
60596 Frankfurt (Germany)  
Fax: (+49) 69-63395-297  
E-mail: ursula.dietrich@em.uni-frankfurt.de

[b] J. Ferner, Dr. M. Baumann, Prof. Dr. H. Schwalbe  
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)



**Figure 1.** Location of the packaging signal  $\psi$  and the gag gene within the HIV-1 genome. The nucleocapsid protein NCp7 of the Gag polyprotein interacts through its zinc fingers (zinc in bold) with the  $\psi$ -RNA structure and in particular with the third stem-loop SL3. The Gag initiation codon is shown in bold.

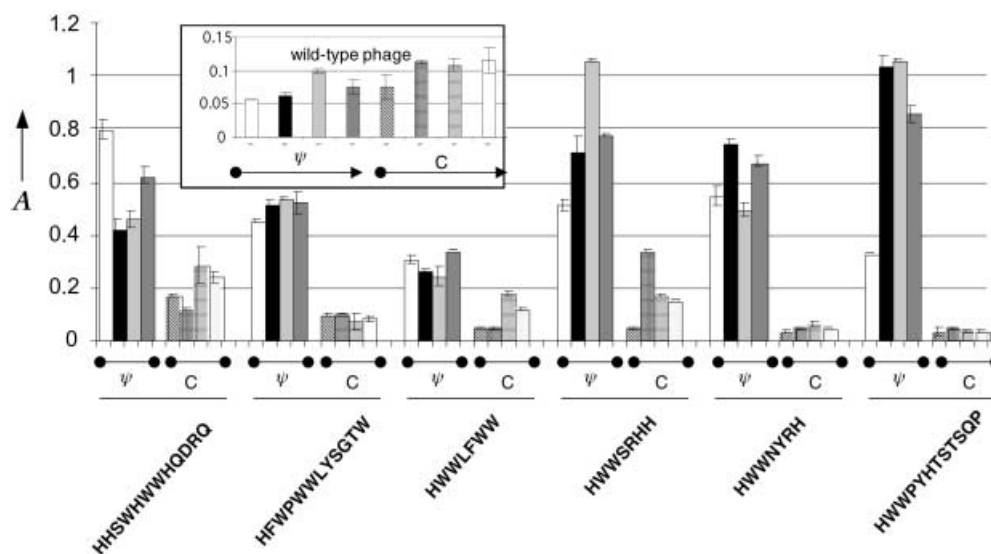
Immobilized full-length in vitro transcribed  $\psi$  RNA was stable under screening conditions and was in a conformation that would allow binding of the natural ligand, HIV-1 Gag protein p55. Full-length  $\psi$  RNA and short synthetic RNA molecules representing stem-loops SL1, SL2, and SL3 were used as the

targets in the biopanning rounds. For negative selections, we used an unrelated RNA of similar size (CCR5) or mutated stem-loop RNAs. After several rounds of positive and negative selections, phages were analyzed for specific binding to the target RNAs derived from the  $\psi$  region by ELISA. Between 8.3 and 19.8% of the phage clones were positive depending on the library used. Peptide motifs of specific phages were identified by sequencing the genome inserts encoding them. Among the most prominent motifs selected was a cluster of aromatic amino acids in conjunction with positively charged amino acids. Interestingly, similar aromatic motifs could be selected with full-length  $\psi$  RNA and with the single stem-loop RNAs (all or SL1, SL2, and SL3; Table 1).

The specificity of the corresponding phage for the  $\psi$  RNA and the stem-loops SL1, SL2, and SL3 was confirmed by ELISA (Figure 2). Specific binding of the phage to full-length  $\psi$  RNA was

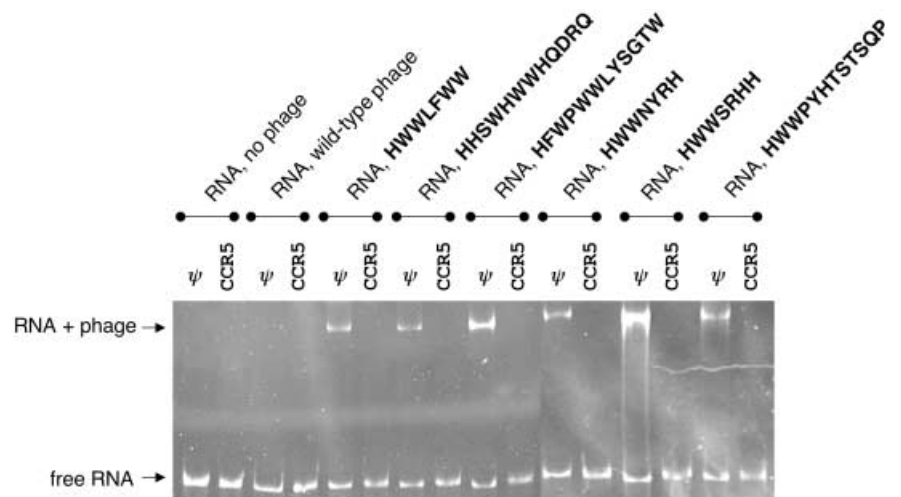
**Table 1.** Aromatic peptide motifs selected with RNA structures derived from the HIV-1 packaging signal  $\psi$ .

Phage clones	Target RNA	Peptide motifs
12.3/12.9/12.18	$\psi$	HHSWHWWHQDRQ
12.21	$\psi$	RWWSWPSYQSS
12.25	$\psi$	WPMTNWFHYHSW
12.55	$\psi$	HFWPWWLYSGTW
7.5	$\psi$	HWWLFWW
7c.47	$\psi$	HWPFLLHS
12.64	SL1	HWWPYHTSTSQP
D4E12.7	SL1	FPWHFHRAPSIH
A557C.64/82/95	SL1	HWWSRHH
7C.8/7C.56	SL2	HWWSRHH
C.7x	SL1	SPWHPHR
A557C.41	SL2	HPHWWHR
A557C.122	SL2	HWWNYRH
A557C.156	SL2	HWWSWRH
A557C.166	SL2	HQHWWRK
C5E7.12.64	SL3	IPWTQHMAMSPM

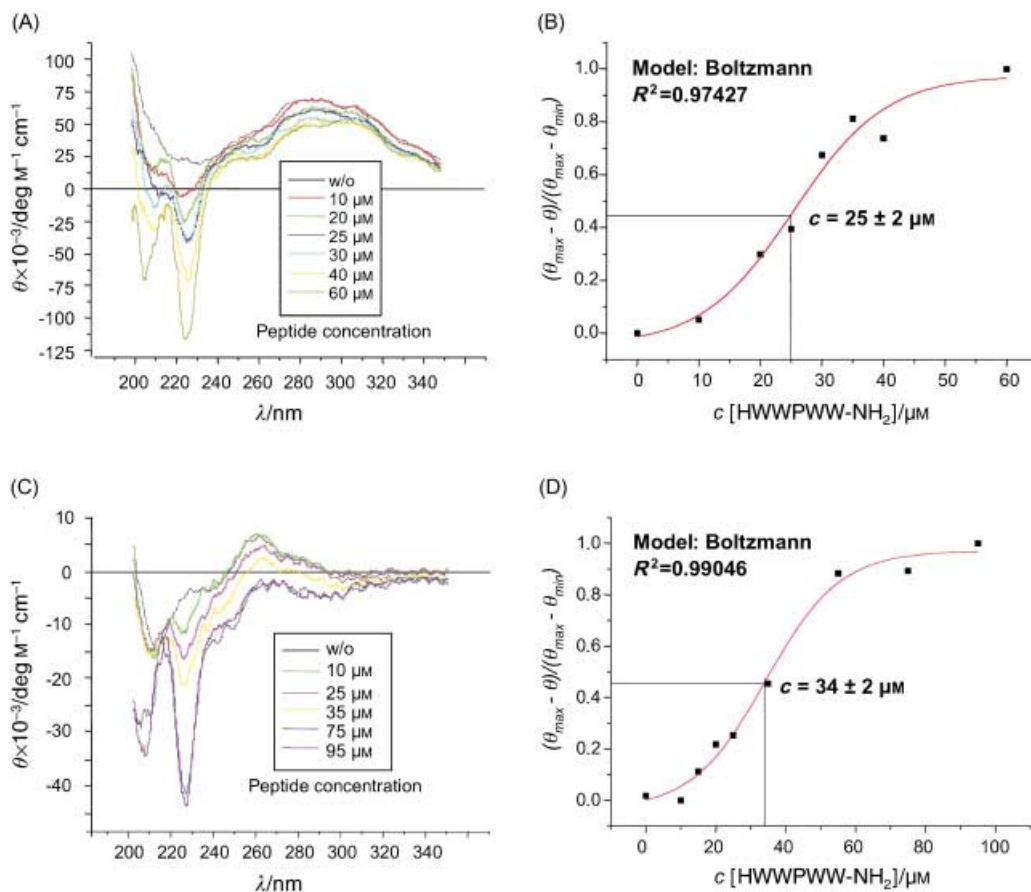


**Figure 2.** Analysis by ELISA of binding of phages selected with RNAs derived from the  $\psi$  region after several rounds of positive and negative selection to  $\psi$  RNAs ( $\psi$ : full-length  $\psi$  RNA, white bars; SL1, black bars; SL2, light grey bars; SL3, dark grey bars) and control RNAs (C: CCR5-RNA, diagonal stripes; mutated SL1, grids; mutated SL2, horizontal stripes; mutated SL3, dots). The insert graph shows binding of wild-type phages to the same RNAs. The optical density values (A) are the mean of triplicate measurements at 492 nm.

also visualized by band-shift analysis by separation on polyacrylamide gels (Figure 3). Specific binding of the hydrophobic peptide motifs to  $\psi$  RNA was confirmed for synthetic peptides corresponding to the peptide sequences on the selected phages by ELISA (data not shown) and CD spectroscopy (Figure 4). After Boltzmann transformation ( $y = (A_1 - A_2)/(1 + e^{(x-x_0)/dx}) + A_2$ ; fitting the data asymptotically to the initial value with no peptide ( $A_1$ ), to the final value with a large excess of peptide ( $A_2$ ), and to the value and the slope of the inflection point  $x_0$  which corresponds to the binding constant,  $K_D$ ), the deduced affinity of a consensus peptide HWWPWW for  $\psi$  RNA was  $25 \pm 2 \mu\text{M}$  and for SL3 was  $34 \pm 2 \mu\text{M}$ . Exchange of single amino acids within the HWWPWW peptide for alanine



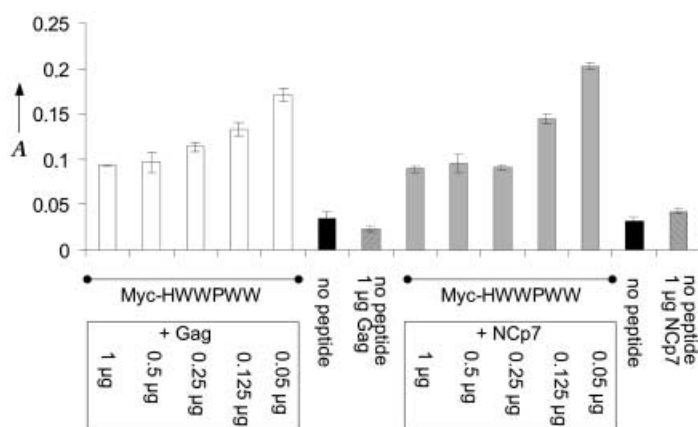
**Figure 3.** Analysis of binding of phages selected with RNAs derived from the  $\psi$  region after several rounds of positive and negative selection to full-length  $\psi$  RNA and CCR5 control RNA by band-shift analysis. Specific band shifts are observed for the selected phages in the presence of  $\psi$  RNA.



**Figure 4.** Detection of binding between peptide HWWPWW-NH<sub>2</sub> and  $\psi$  RNA (A and B) and stem-loop SL3 (C and D) by CD spectroscopy. A) CD spectra (at 20 °C) show the decrease of the positive molar ellipticity at around 290 nm which stems from the secondary structure of the  $\psi$  RNA (0.2  $\mu\text{M}$ ) induced by titration with the peptide HWWPWW-NH<sub>2</sub> in 100 mM potassium phosphate buffer (pH 6.5). B) Plot of the molar ellipticity decrease at 290 nm ( $\theta_{\text{max}} - \theta$ ) in relation to the maximum decrease ( $\theta_{\text{max}} - \theta_{\text{min}}$ ) against the peptide concentration.  $\theta_{\text{max}}$  is the maximum molar ellipticity at 290 nm during the experiment and  $\theta_{\text{min}}$  is the minimum value. Curve fitting assumed the sigmoidal Boltzmann function ( $y = (A_1 - A_2)/(1 + e^{(x-x_0)/dx}) + A_2$ ) and resulted in a binding constant  $K_D$  of  $25 \pm 2 \mu\text{M}$ . C) CD spectra (at 20 °C) show the decrease of the positive molar ellipticity around 260 nm which stems from the secondary structure of the SL3 RNA (0.7  $\mu\text{M}$ ) induced by titration with the peptide HWWPWW-NH<sub>2</sub> in 100 mM potassium phosphate buffer (pH 6.5). D) Plot of the molar ellipticity decrease at 260 nm ( $\theta_{\text{max}} - \theta$ ) in relation to the maximum decrease ( $\theta_{\text{max}} - \theta_{\text{min}}$ ) against the peptide concentration.  $\theta_{\text{max}}$  is the maximum molar ellipticity at 260 nm during the experiment and  $\theta_{\text{min}}$  is the minimum value. Curve fitting assumed the sigmoidal Boltzmann function (as above) and resulted in a binding constant  $K_D$  of  $34 \pm 2 \mu\text{M}$ .

was followed by CD spectroscopy and showed residues H1, W3, and W5 to be important for RNA binding; in contrast, mutation of W2 and P4 resulted in improved RNA binding.

Tryptophan motifs are known to be important for RNA binding. NMR analysis of SL3 RNA with the nucleocapsid protein NCp7 showed the specific stacking interaction of W37 of the second zinc finger of NCp7 with unpaired guanosine residues (especially G318) within the SL3 tetraloop.<sup>[36, 37]</sup> NCp7 also binds to the single-stranded loops of SL1, SL2, and SL4 with affinities of 20–140 nM.<sup>[38]</sup> High-resolution structures revealed similar interactions between NCp7 and the different stem-loops involving the hydrophobic cleft of the zinc fingers and unpaired guanosine residues. In fact, this stacking interaction was used in binding assays to study the affinities of wild-type and mutated stem-loops SL1–SL4 for NCp7, based on the quenching of the fluorescence of W37 by guanosine residues in the single-stranded loops of SL1–SL4.<sup>[32, 38]</sup> Interestingly, we also selected similar tryptophan-containing peptide motifs independently of which RNA was used as the target. Furthermore, we could confirm specific binding of the hydrophobic peptides to  $\psi$  RNA as well as to SL1, SL2, and SL3 by ELISA. This binding could be competed with by the natural ligands Gagp55 and NCp7 (Figure 5), but not by an unrelated RNA binding protein, the HIV-1 transactivator protein Tat (data not shown).



**Figure 5.** Binding of the myc-labeled HWWPWW peptide to full-length  $\psi$  RNA is competed with by the natural ligands Gagp55 (white bars) and NCp7 (grey bars). Black bars indicate control wells with RNA and no peptide, diagonal stripes indicate control wells without peptide but with a competitor protein. Bound peptide is detected by a horse raddish peroxidase labeled anti-myc antibody.

Thus, by the phage-display approach we selected peptide ligands for RNA structures derived from the HIV-1  $\psi$  region, and these peptides resemble the natural NCp7 ligand in terms of the binding features, including the tryptophan residues involved and the cross-reactivity with different RNA stem-loop structures derived from the  $\psi$  region. The selected peptides can now serve to derive modified peptides or small molecules with optimized binding properties to the target RNA, which could then potentially interfere with the packaging process of HIV-1 RNA genomes.

## Experimental Section

**Target RNAs:** Full-length  $\psi$  RNA and CCR5 RNA of similar length were in vitro transcribed after PCR amplification and cloning (nucleotides 20–202 from HIV-1 pNL4-3 and nucleotides 737–890 from pcDNA3-CCR5) into pTRikan18 (Ambion). Shorter stem-loops SL1–SL3, as well as mutated stem-loops SL1–SL3, were chemically synthesized and biotinylated. All RNAs were checked on polyacrylamide gels for purity. The sequences were: SL1, biotin-TTTTTCUCGGCUGAAGCGCGCA CGGCAA GAGGCGAG; SL1 mut, biotin-TTTTTCUCGGCUGAAGCGCGCA GUCGUU-CAGGCGAG; SL2, biotin-TTTTGGCGACUGGUGA GUA CGCC; SL2 mut, biotin-TTTTGGCGAC UAA AAA GUA CGCC; SL3, biotin-TTTTGGACUA GCG GAG GCU AGUCC; SL3 mut, biotin-TTTTGGACUA GCA AAA GCU AGUCC.

**Screening of phage-displayed peptide libraries:** RNAs (50 ng per well) in psi buffer (5 mM 2-[4-(2-hydroxyethyl)-1-piperazinyl]ethanesulfonic acid (HEPES), 100 mM KCl, (pH 7.4)) were denatured for 3 min at 95 °C and renatured for 20 min at room temperature before being immobilized on streptavidin-coated plates, either directly (SL1–SL3) or through a biotinylated primer (5'-oligo: GCTTAATACTGACGCTCTCGC, 100 pmol) complementary to the 3'-end of the  $\psi$  RNA or (3'-oligo: CCTTTA GTG AGG GTA ATT CTC G) complementary to the 5'-end of CCR5 RNA by incubating for 3 h at 4 °C. After washing, the plates were incubated with 3 different phage libraries (Biolabs) displaying peptides of 7 or 12 amino acids in a linear (7-mer, 12-mer) or a cyclic form (7-mer) for 4 h at 4 °C. After extensive washing, binding phages were eluted with glycine (100  $\mu$ L, 0.2 M, (pH 2.2)) and neutralized with tris(hydroxymethyl)aminomethane (Tris)/HCl (15  $\mu$ L, (pH 9.1)). After four rounds of positive selection, CCR5 RNA or mutated SL1, SL2, or SL3 RNAs were used for negative selection. Finally, phages were amplified and titered, and phage DNA was prepared for sequencing of the peptide inserts.

**Peptide synthesis:** All chemicals and solvents used were of analytical grade. Peptides were synthesized on an ABI 433A peptide synthesizer (Applied Biosystems, USA) by solid-phase chemistry with in situ neutralization (2 M diisopropylethylamine (DIEA) in *N*-methylpyrrolidone (NMP)) and activation with 19% *O*-(benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU) and 7.66% 1-hydroxy-1*H*-benzotriazole (HOBT) in *N,N*-dimethylformamide (DMF) on 9-fluorenylmethoxycarbonyl-amide (Fmoc-amide) resins (Applied Biosystems, USA). After chain assembly was complete, peptides were deprotected and simultaneously cleaved from the resin by treatment with trifluoroacetic acid containing phenol (1.25%), chlorotriisopropylsilane (1.25%), 1,2-ethanedithiol (1.25%), and water (1.25%). Peptides were then lyophilized and purified by preparative HPLC. Fmoc-protected amino acids were obtained from NovaBiochem, USA. Side-chain protecting groups were as follows: Arg(Pmc), Asn(Trt), Asp(tBu), Gln(Trt), Glu(tBu), His(Trt), Lys(Boc), Ser(tBu), Thr(tBu), Trp(Boc), and Tyr(tBu), where Pmc = 2,2,5,7,8-pentamethylchromen-6-sulfonyl, Trt = trityl = triphenylmethyl, and Boc = *tert*-butoxycarbonyl.

**ELISA:** Streptavidin-coated plates were blocked with milk powder (3% in phosphate-buffered saline (PBS)) for 2 h at room temperature. After washing, RNAs (50 ng) from the  $\psi$  region or control RNAs were immobilized for 90 min at room temperature as described above. After washing, the amplified phage stocks (10  $\mu$ L) were added for 2 h at room temperature. Binding phage were detected with mouse anti-phage antibodies conjugated with horseradish peroxidase (HRP). After addition of the HRP substrate *ortho*-phenylenediamine dihydrochloride (OPD) plates were read at 492 nm.



For peptide competition ELISA measurements, the myc-labeled HWWPWW peptide (500 ng) was added in milk powder (5% in PBS with 5 µg of yeast tRNA) to immobilized full-length  $\psi$  RNA. After washing, Gag or NCp7 protein (50 ng – 1 µg) was added for 30 min at room temperature. The bound myc-HWWPWW peptide was detected with an HRP-labeled anti-myc antibody.

**Band-shift analysis:** RNA (500 ng) was denatured for 5 min at 95 °C and renatured for 25 min at room temperature. The phage stock (10 µL) was added and incubated for 20 min at room temperature. Samples were run on a native polyacrylamide gel (5%) for 150 min at 75 V and 4 °C. Bands were visualized by ethidium bromide staining.

**Circular dichroism spectroscopy:** Spectra were obtained with a Jasco J810 spectropolarimeter (Jasco, USA) at 20 °C. The peptides had no elliptic maxima in the range of 260–290 nm, thereby allowing for direct observation of changes in the RNA spectrum with the addition of peptides. The concentration of  $\psi$  RNA was 0.2 µM and of SL3-RNA was 0.7 µM in a 100 mM potassium phosphate buffer (pH 6.5). The peptide concentrations ranged from 5–100 µM.

## Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (SFB579 on "RNA–ligand interactions"). The following reagents were obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, National Institutes of Health: HIV-1SF2 p55 Gag (Chiron Corporation) and HIV-1MN p7 (Dr. L. Henderson). Tat protein was provided by Dr. E. Loret (France).

**Keywords:** HIV-1 packaging · peptides · phage display · RNA · RNA–peptide interactions

- [1] M. G. Caprara, T. W. Nilsen, *Nat. Struct. Biol.* **2000**, *7*, 831–833.
- [2] P. Mucha, A. Szyk, P. Rekowski, P. A. Weiss, P. A. Agris, *Biochemistry* **2001**, *40*, 14 191–14 199.
- [3] R. N. De Guzman, R. B. Turner, M. F. Summers, *Biopolymers* **1998**, *48*, 181–195.
- [4] J. M. Pérez-Cañadillas, G. Varani, *Curr. Opin. Struct. Biol.* **2001**, *11*, 53–58.
- [5] M. A. Weiss, N. Narayana, *Biopolymers* **1998**, *48*, 167–180.
- [6] J. R. Williamson, *Nat. Struct. Biol.* **2000**, *7*, 834–837.
- [7] N. Leulliot, G. Varani, *Biochemistry* **2001**, *40*, 7947–7956.
- [8] D. J. Patel, *Curr. Opin. Struct. Biol.* **1999**, *9*, 75–87.
- [9] A. D. Frankel, *Curr. Opin. Struct. Biol.* **2000**, *10*, 332–340.
- [10] S. D. Jayasena, *Clin. Chem.* **1999**, *45*, 1628–1650.
- [11] E. S. DeJong, B. Luy, J. P. Marino, *Curr. Top. Med. Chem.* **2002**, *2*, 289–302.
- [12] G. J. Zaman, P. J. Michiels, C. A. van Boeckel, *Drug Discov. Today* **2003**, *8*, 297–306.
- [13] H. M. E. Azzazy, W. E. Highsmith, Jr., *Clin. Biochem.* **2002**, *35*, 425–445.
- [14] I. A. Laird-Offringa, J. G. Balesco, *Methods Enzymol.* **1996**, *267*, 149–168.
- [15] J. E. Powers, M. T. Marchbank, S. L. Deutscher, *Nucleic Acids Symp. Ser.* **1995**, *33*, 240–243.
- [16] R. M. Hoet, M. Pieffers, M. H. Stassen, J. Raats, R. de Wildt, G. J. Pruijn, F. van den Hoogen, W. J. van Venrooij, *J. Immunol.* **1999**, *163*, 3304–3312.
- [17] P. F. Agris, M. T. Marchbank, W. Newman, R. Guenther, P. Ingram, J. Swallow, P. Mucha, A. Szyk, P. Rekowski, E. Peletskaya, S. L. Deutscher, *J. Protein Chem.* **1999**, *18*, 425–435.
- [18] W. J. Friesen, M. K. Darby, *Nat. Struct. Biol.* **1998**, *5*, 543–546.
- [19] W. J. Friesen, M. K. Darby, *J. Biol. Chem.* **2001**, *276*, 1968–1973.
- [20] S. Hoffmann, D. Willbold, *Biochem. Biophys. Res. Commun.* **1997**, *235*, 806–811.
- [21] G. Jonas, S. Hoffmann, D. Willbold, *J. Biomed. Sci.* **2001**, *8*, 430–436.
- [22] R. Berkowitz, J. Fisher, S. P. Goff, *Curr. Top. Microbiol. Immunol.* **1996**, *214*, 177–218.
- [23] J. Dannull, A. Surovoy, G. Jung, K. Moelling, *EMBO J.* **1994**, *13*, 1525–1533.
- [24] J. Clever, C. Sasseti, T. G. Parslow, *J. Virol.* **1995**, *69*, 2101–2109.
- [25] T. Hayashi, T. Shioda, Y. Iwakura, H. Shibuta, *Virology* **1992**, *188*, 590–599.
- [26] G. K. Amarasinghe, R. N. De Guzman, R. B. Turner, M. F. Summers, *J. Mol. Biol.* **2000**, *299*, 145–156.
- [27] G. K. Amarasinghe, J. Zhou, M. Miskimon, K. J. Chancellor, J. A. McDonald, A. G. Matthews, R. R. Miller, M. D. Rouse, M. F. Summers, *J. Mol. Biol.* **2001**, *314*, 961–970.
- [28] A. Mujeeb, T. G. Parslow, A. Zarrinpar, C. Das, T. L. James, *FEBS Lett.* **1999**, *458*, 387–392.
- [29] A. H. Maki, A. Ozarowski, A. Misra, M. A. Urbaneja, J. R. Casas-Finet, *Biochemistry* **2001**, *40*, 1403–1412.
- [30] C. Vuilleumier, E. Bombarda, N. Morellet, D. Gérard, B. P. Roques, Y. Mély, *Biochemistry* **1999**, *38*, 16 816–16 825.
- [31] N. Morellet, H. Déméné, V. Teilleux, T. Huynh-Dinh, H. de Rocquigny, M. C. Fournié-Zaluski, B. P. Roques, *J. Mol. Biol.* **1998**, *283*, 419–434.
- [32] A. C. Paoletti, M. F. Shubsda, B. S. Hudson, P. N. Borer, *Biochemistry* **2002**, *41*, 15 423–15 428.
- [33] A. Zeffman, S. Hassard, G. Varani, A. Lever, *J. Mol. Biol.* **2000**, *297*, 877–893.
- [34] J. A. Turpin, S. J. Terpening, C. A. Schaeffer, G. Yu, C. J. Glover, R. L. Felsted, E. A. Sausville, W. G. Rice, *J. Virol.* **1996**, *70*, 6180–6189.
- [35] D. R. Chadwick, A. M. L. Lever, *Gene Ther.* **2000**, *7*, 1362–1368.
- [36] R. N. De Guzman, Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, M. F. Summers, *Science* **1998**, *279*, 384–388.
- [37] L. Pappalardo, D. J. Kerwood, I. Pelczar, P. N. Borer, *J. Mol. Biol.* **1998**, *282*, 801–818.
- [38] M. F. Shubsda, A. C. Paoletti, B. S. Hudson, P. N. Borer, *Biochemistry* **2002**, *41*, 5276–5282.

Received: June 2, 2003 [Z 681]



## **Research Article:** A Tryptophan-Rich Hexapeptide Inhibits Nucleic Acid Destabilization Chaperoned by the HIV-1 Nucleocapsid Protein

Chinappan Raja, Jan Ferner, Ursula Dietrich, Sergey Avilov, Damien Ficheux, Jean-Luc Darlix, Hugues de Rocquigny, Harald Schwalbe, and Yves Mély  
*Biochemistry* **2006**, 45, 9254-9265

In this publication, the binding characteristics of the optimized ligand HKWPWW (derivation in the publication Dietz et al., 2008) to further viral targets in HIV (TAR-RNA and PBS-DNA) were investigated by NMR, steady-state and time-resolved fluorescence spectroscopy. The peptide exists in two conformations which were determined by NMR and ARIA calculations to support the analysis of the fluorescence data.

The fluorescence measurements and analysis were carried out by our collaborators in France, the workgroup of Yves Mély. All NMR measurements to determine the RNA-ligand interactions and the ligand structure were performed by the author of this thesis. He was in charge of writing the corresponding methodological parts and contributed strongly to the results and discussion sections of the manuscript.



# A Tryptophan-Rich Hexapeptide Inhibits Nucleic Acid Destabilization Chaperoned by the HIV-1 Nucleocapsid Protein<sup>†</sup>

Chinappan Raja,<sup>‡</sup> Jan Ferner,<sup>§</sup> Ursula Dietrich,<sup>||</sup> Sergiy Avilov,<sup>‡,⊥</sup> Damien Ficheux,<sup>#</sup> Jean-Luc Darlix,<sup>△</sup> Hugues de Rocquigny,<sup>‡</sup> Harald Schwalbe,<sup>§</sup> and Yves Mély\*<sup>‡</sup>

Département Pharmacologie et Physicochimie, Faculté de Pharmacie, Institut Gilbert-Laustriat, UMR 7175 CNRS/Université Louis Pasteur (Strasbourg I), 74 route du Rhin, 67401 Illkirch, France, 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, Georg Speyer Haus, Institute for Biomedical Research, Paul Ehrlich Strasse 42-44, 60596 Frankfurt, Germany, Palladin Institute of Biochemistry, 9, Leontovich Street, 01030 Kiev, Ukraine, IBCP, 7, passage du Vercors, 69367 Lyon Cedex 07, France, and LaboRétro, Unité de Virologie Humaine INSERM 412, Ecole Normale Supérieure de Lyon, 46 allée d'Italie, 69364 Lyon, France

Received December 16, 2005; Revised Manuscript Received April 13, 2006

**ABSTRACT:** The nucleocapsid protein (NC) of HIV-1 exerts critical functions in viral genome replication and virus assembly. Since the recognition of target nucleic acids is required in the initial step of most NC-mediated processes, attempts were made to find small molecules capable of competing with this recognition. In particular, several Trp-rich hexapeptides were recently found to strongly bind RNA sequences targeted by NC. To further validate these peptides as potential anti-NC agents, we studied the ability of Ac-HKWPWW-NH<sub>2</sub>, taken as a representative, to interfere with the NC chaperone properties required during reverse transcription. Using NMR and steady-state and time-resolved fluorescence spectroscopy, we characterized the structure of Ac-HKWPWW-NH<sub>2</sub> as well as its binding to viral sequences such as TAR and PBS involved in the two obligatory strand transfers of reverse transcription. Results show that Ac-HKWPWW-NH<sub>2</sub> exhibits an almost symmetric *cis*–*trans* equilibrium at the level of the Pro residue where it is structured. The peptide binds both TAR and PBS sequences with low micromolar affinities. The *cis*-Pro and *trans*-Pro conformations of the peptide bind with comparable affinities to (–)PBS, mainly through stacking interactions between the Trp residues and the (–)PBS bases. Though all three Trp residues may contribute to the (–)PBS/Ac-HKWPWW-NH<sub>2</sub> complex formation, Trp3 and Trp5 residues are the key residues in the complexes with the *cis*-Pro and *trans*-Pro conformations, respectively. Moreover, Ac-HKWPWW-NH<sub>2</sub> stabilizes cTAR secondary structure and largely inhibits the NC-directed melting of cTAR. This further strengthens the interest of this peptide for deriving modified peptides capable of inhibiting NC and HIV-1 replication.

The nucleocapsid protein (NC)<sup>1</sup> of the human immunodeficiency virus type 1 (HIV-1) plays several key roles in the viral life cycle, either as a domain of the Gag polyprotein precursor or as the mature protein. NC notably selects and directs dimerization of the genomic RNA, which is subsequently efficiently packaged into assembling viral particles (1, 2). Moreover, NC acts as a chaperone protein during reverse transcription by promoting the annealing of the

primer tRNA to the primer binding site (PBS) and directing the two obligatory strand transfers necessary for the synthesis of a complete proviral DNA by reverse transcriptase (3–10). In the first transfer, the cTAR from the strong stop cDNA [(–)sscDNA] is hybridized to the TAR located at the 3' end of the genomic RNA while the second transfer occurs via the annealing of the (+)PBS and (–)PBS sequences located at the 3' end of the (+)ssDNA and cDNA(–), respectively.

The mature nucleocapsid protein is a small highly basic protein with two CX<sub>2</sub>CX<sub>4</sub>HX<sub>4</sub>C zinc fingers connected by a short flexible basic linker (11). Zinc fingers are essential for

<sup>†</sup> The Agence Nationale de Recherches sur le SIDA (ANRS), the European TRIOH Consortium, the Deutsche Forschungsgemeinschaft (SFB579 on RNA-Ligand Interactions), the Center for Biomolecular Magnetic Resonance, and the State of Hesse are gratefully acknowledged for financial support to this work. C.R. was a fellow from Sidaction. S.A. was a fellow from the French Ministère de la Recherche and Fondation pour la Recherche Médicale.

\* To whom correspondence should be addressed. Tel: +33 (0)3 90 24 42 63. Fax: +33 (0)3 90 24 43 12. E-mail: yves.mely@pharma.u-strasbg.fr.

<sup>‡</sup> Institut Gilbert-Laustriat.

<sup>§</sup> Johann Wolfgang Goethe University.

<sup>||</sup> Institute for Biomedical Research.

<sup>⊥</sup> Palladin Institute of Biochemistry.

<sup>#</sup> IBCP.

<sup>△</sup> Ecole Normale Supérieure de Lyon.

<sup>1</sup> Abbreviations: DQF-COSY, double-quantum-filtered correlation spectroscopy; FI, fluorescein; HIV-1, human immunodeficiency virus type 1; HSQC, heteronuclear single-quantum correlation; MES, 2-(*N*-morpholino)ethanesulfonic acid; NC, nucleocapsid protein; NMR, nuclear magnetic resonance; PAGE, polyacrylamide gel electrophoresis; PBS, primer binding site; ROE, rotating-frame Overhauser enhancement; ROESY, rotating-frame Overhauser enhancement spectroscopy; SL3, stem-loop 3 from the Psi encapsidation sequence of the HIV-1 genome; TAR, transactivation response element; cTAR, complementary DNA sequence to TAR; TMR, 5- and 6-carboxytetramethylrhodamine; TOCSY, total correlated spectroscopy.

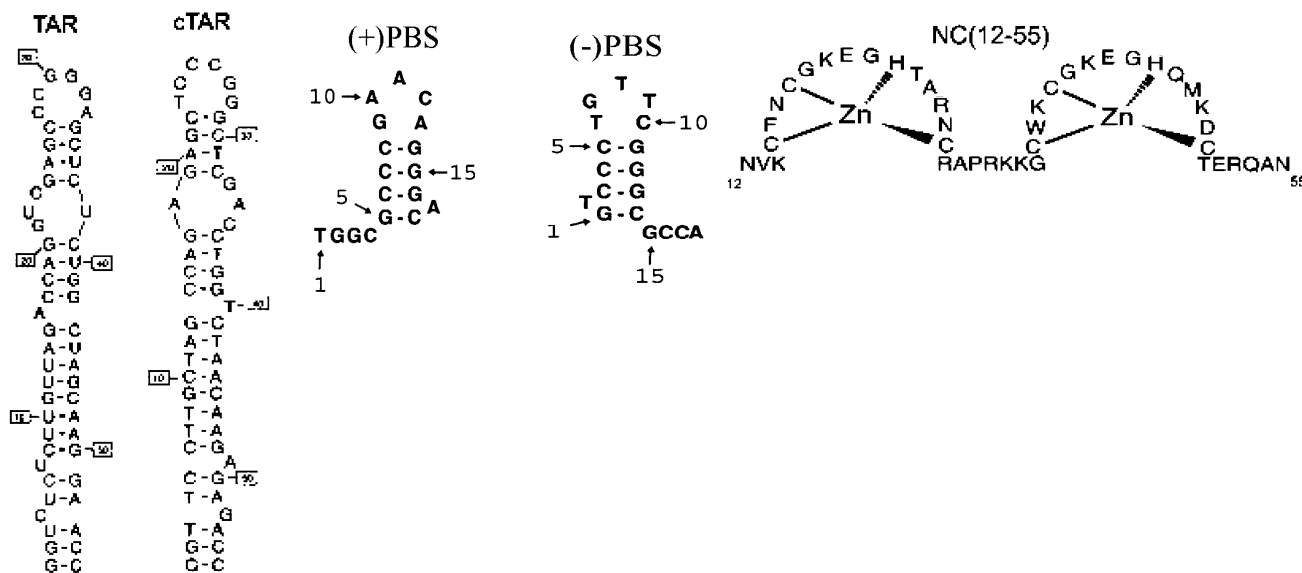


FIGURE 1: Structure of TAR RNA, cTAR DNA, (+)PBS DNA, (-)PBS DNA, and NC(12-55). The selected TAR and PBS sequences are from the HIV-1 Mal strain. The secondary structures of cTAR and PBS were predicted from that of TAR (72) and (-)PBS (73), respectively, and the mfold program (74).

NC activities. Indeed, mutations of amino acids involved in zinc binding completely abolish virus infectivity and impair virion structure (12–15). The NC mutant virions have low amounts of genomic RNA and exhibit defects in reverse transcription (10, 12, 13, 16–22). These results were corroborated by *in vitro* studies describing the NC determinants in nucleic acid binding and chaperone activities (13, 19, 20, 23–25). Nucleic acid chaperoning by NC is thought to involve at least three steps: (i) binding of NC to its target nucleic acid sequences, (ii) destabilization of the nucleic acid secondary and tertiary structures, and (iii) promotion of the annealing of the destabilized complementary sequences (for a review, see ref 5 and references cited therein). NC destabilization properties mainly rely on the zinc fingers while NC-mediated annealing depends also on the basic amino acids located in both N- and C-terminal regions (13, 19, 20, 23–25).

Due to its numerous functions in the viral life cycle and its highly conserved sequence, NC constitutes an attractive target for the development of antiviral agents (26–30). In this respect, many attempts were performed to target directly NC molecules with either (i) zinc-ejecting agents that react with Cys residues within the zinc fingers and induce subsequent loss of functionally active conformation (31) or (ii) RNA aptamers and xanthenyl ring-containing compounds that bind NC with high affinity (32–35). Since the initial step of most NC-promoted activities corresponds to a nucleic acid recognition process, an alternative strategy is to synthesize small molecules able to compete with NC to its target nucleic acids. In this respect, phage display libraries were recently used to identify small peptides able to interact with the encapsidation  $\psi$  RNA sequence of HIV-1 (36). This study identified Trp-rich hexapeptides that bind with micromolar affinities to the encapsidation sequence of the genomic RNA and may thus potentially interfere with the selection, dimerization, and packaging of the HIV-1 genome.

To further validate the potential of these Trp-rich peptides to interfere with NC functions during reverse transcription, we performed a detailed structural and functional analysis

of Ac-HKWPWW-NH<sub>2</sub>, taken as a representative of these peptides. The structure of this peptide and its interaction with (-)PBS DNA were studied by NMR. Moreover, steady-state and lifetime fluorescence spectroscopy was used to further characterize its binding to the TAR (TAR RNA and cTAR DNA) and PBS [(–)PBS DNA and (+)PBS DNA] sequences (Figure 1) involved in the first and second strand transfers during reverse transcription. In addition, using doubly labeled cTAR DNA sequences, we characterized the ability of Ac-HKWPWW-NH<sub>2</sub> to inhibit the NC-directed destabilization of the cTAR secondary structure. Taken together, our results show that this peptide exhibits an almost symmetric *cis-trans* equilibrium at the level of its Pro residue, binds to both TAR and PBS sequences with low micromolar affinities mainly through stacking interactions with Trp residues, stabilizes the secondary structure of cTAR, and inhibits its NC-mediated melting *in vitro*. This further strengthens the importance of this peptide to derive modified peptides or small molecules able to inhibit NC functions during HIV-1 replication.

## MATERIALS AND METHODS

NC(12–55) (Figure 1) and Ac-HKWPWW-NH<sub>2</sub> were synthesized by solid-phase chemistry as previously described (36, 37). The peptides were purified on reversed-phase HPLC columns and lyophilized. NC(12–55) was stored in its zinc-bound form, and its concentration was determined using an extinction coefficient of 5700 M<sup>-1</sup> cm<sup>-1</sup> at 280 nm.

Doubly and singly labeled DNA oligonucleotides were synthesized at a 0.2  $\mu$ mol scale by IBA GmbH Nucleic Acids Product Supply (Göttingen, Germany). For this study, we used a pair of dyes, 5- and 6-carboxytetramethylrhodamine (TMR) and 5- and 6-carboxyfluorescein (Fl) (38). The 5' terminus was labeled with TMR via an amino linker with a six carbon spacer arm. The 3' terminus was labeled with Fl using a special solid support with the dye already attached. Oligonucleotides were purified by PAGE, and the purity was verified by electrospray ionization mass spectrometry. Fluorescence experiments were performed in 25 mM Tris-HCl,

pH 7.5, 30 mM NaCl, and 0.2 mM MgCl<sub>2</sub> (25, 38, 39). Concentrations of oligonucleotides and peptides were measured by UV absorption on a Cary 400 UV–visible spectrophotometer. Extinction coefficients of 521910, 396600, 153900, and 175680 M<sup>-1</sup> cm<sup>-1</sup> at 260 nm were used for cTAR, TAR, (-)PBS, and (+)PBS sequences, respectively (10, 38). All experiments were performed at 20 °C. Ac-HKWPWW-NH<sub>2</sub>-cTAR binding, the effect of Ac-HKWPWW-NH<sub>2</sub> on NC destabilizing activity, and NMR spectra of Ac-HKWPWW-NH<sub>2</sub> were also studied at 37 °C.

**Homo- and Heteronuclear NMR Measurements.** For the structure determination of the peptide, NMR experiments were recorded at 298 and 310 K on a 2.0 mM Ac-HKWPWW-NH<sub>2</sub> peptide sample which was synthesized as described above with an N-terminal acetylation and an amide group at the C-terminus. The NMR buffer contained 25 mM potassium phosphate and 50 mM potassium chloride, and the pH was adjusted to 6.2. NMR resonance assignment was accomplished by analyzing the following experiments: a <sup>1</sup>H, <sup>15</sup>N HSQC and a <sup>1</sup>H, <sup>13</sup>C HSQC (40–42) a <sup>1</sup>H, <sup>1</sup>H DQF-COSY, a <sup>1</sup>H, <sup>1</sup>H TOCSY (43) (mixing time of 80 ms), and a <sup>1</sup>H, <sup>1</sup>H ROESY (44) (mixing time of 400 ms). Full experimental details are provided in the Supporting Information.

The interactions between the peptide and (-)PBS DNA (from MWG-Biotech AG) were monitored by 1D homonuclear <sup>1</sup>H NMR during the titration of a 0.4 mM solution of Ac-HKWPWW-NH<sub>2</sub> with (-)PBS in the above-mentioned NMR buffer. The 1D experiments were recorded with a jump–return–echo sequence where the excitation maximum was set to the chemical shift range of the Trp indole signals. The NMR data were recorded on Bruker 400 and 600 MHz spectrometers at 298 and 310 K.

All structure calculations were performed using the program CNS 1.1 (45) and the ARIA 1.2 (46) setup and protocols. The protein allhdg 5.3 force field (47) was used and was slightly adapted to allow structural calculations with the N-terminal acetyl group, the C-terminal amide, and the *cis*-proline. After each of the first eight iterations, 50 structures were calculated. Then, the ROE distance restraints were recalibrated by ARIA based on the 10 lowest energy structures. The violation tolerance was progressively reduced to 0.1 Å in the last iteration in which 200 structures were calculated. The final 20 lowest energy structures were further analyzed and visualized.

**Steady-State and Time-Resolved Fluorescence Measurements.** Fluorescence emission spectra were recorded on either a Fluorolog or a FluoroMax spectrofluorometer (Jobin Yvon) equipped with a thermostated cell compartment. The fluorescence quantum yield of Ac-HKWPWW-NH<sub>2</sub> was calculated by using Trp in water as a reference ( $\phi_{\text{Trp}} = 0.14$ ) (48).

Time-resolved fluorescence spectroscopy measurements were performed with a time-correlated single photon counting technique as previously described (10). The excitation wavelengths for Trp and Fl were 295 and 480 nm, respectively, and their emission were monitored at 350 and 520 nm, respectively. The fluorescence lifetime measurements were performed with a 1 μM peptide concentration. Time-resolved data analysis was performed by the maximum entropy method using the Pulse5 software (49). The mean lifetime  $\langle\tau\rangle$  was calculated from the individual fluorescence lifetimes,  $\tau_i$ , and their relative amplitudes,  $\alpha_i$ , by  $\langle\tau\rangle = \sum_i \alpha_i \tau_i$ . The population,  $\alpha_0$ , of dark species in peptide–oligonucleo-

tide complexes and Fl-5'-cTAR-3'-TMR was calculated by

$$\alpha_0 = 1 - \frac{\langle\tau\rangle_R}{\langle\tau\rangle_S R_m} \quad (1)$$

where  $\langle\tau\rangle_R$  is the measured mean lifetime of the free Ac-HKWPWW-NH<sub>2</sub> peptide or the singly labeled Fl-5'-cTAR derivative, while  $\langle\tau\rangle_S$  is the measured lifetime of the peptide–oligonucleotide complex or the doubly (Fl-5'-cTAR-3'-TMR) labeled derivative, respectively.  $R_m$  corresponds to the ratio of their corresponding steady-state fluorescence intensities. The remaining amplitudes,  $\alpha_{ic}$ , were recalculated from the measured amplitudes,  $\alpha_i$ , according to  $\alpha_{ic} = \alpha_i(1 - \alpha_0)$ .

The effect of His1 deprotonation on Ac-HKWPWW-NH<sub>2</sub> fluorescence was monitored in 25 mM 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer. Direct addition of NaOH to adjust the pH value leads to turbidity, probably due to local peptide precipitation at the point of addition. To avoid this, separated MES buffer solutions with different pH values ranging from 4.5 to 7.5 were prepared. To each solution was added Ac-HKWPWW-NH<sub>2</sub> at a final 1 μM concentration, and its fluorescence intensity at 350 nm was recorded.

For binding studies, fluorescence titrations were performed by adding increasing oligonucleotide concentrations to a fixed amount of peptide in 25 mM Tris-HCl buffer, 30 mM NaCl, and 0.2 mM MgCl<sub>2</sub>, pH 7.5. The binding stoichiometry was determined at a peptide concentration of 1.5–2.0 μM while the binding constants were determined at peptide concentrations of 0.2–1.5 μM. Fluorescence intensities were corrected for dilution, buffer fluorescence, and screening effects due to the oligonucleotide absorbance. The fluorescence intensity,  $I$ , at 350 nm was fitted according to

$$I = I_0 - [(I_0 - I_t)/L_t] \{ [1 + (L_t + nN_t)K_{\text{app}}] - \sqrt{[1 + (L_t + nN_t)K_{\text{app}}]^2 - 4L_t n N_t K_{\text{app}}^2} \} / 2K_{\text{app}} \quad (2)$$

where  $L_t$  and  $N_t$  designate the total concentration of peptide and oligonucleotide, respectively.  $I_t$  represents the fluorescence at the plateau when all of the peptide is bound, whereas  $I_0$  and  $I$  correspond to the fluorescence intensities of the peptide in the absence and in the presence of a given concentration of oligonucleotide.  $K_{\text{app}}$  and  $n$  correspond to the apparent binding constant and the number of peptide binding sites, respectively.

## RESULTS

**Characterization of Ac-HKWPWW-NH<sub>2</sub> by NMR.** Figure 2 shows the <sup>1</sup>H, <sup>15</sup>N HSQC spectrum of the peptide. For a single peptide conformation, 10 signals (5 backbone amide resonances, 3 side chain indole resonances, and 2 side chain amide resonances) are expected. The observation of a total of 20 cross-peaks suggests the existence of two conformations in slow chemical exchange; i.e., the rate of interconversion is slow on the NMR time scale (millisecond to second range). The findings of two full and independent data sets described in Figure 2 for the <sup>1</sup>H, <sup>15</sup>N HSQC are in agreement with the proton 1D spectra and evident in all recorded NMR spectra. By analysis of the <sup>1</sup>H, <sup>1</sup>H TOCSY and the <sup>1</sup>H, <sup>1</sup>H DQF-COSY spectra, the intraresidual spin systems for both conformers could be assigned. From the

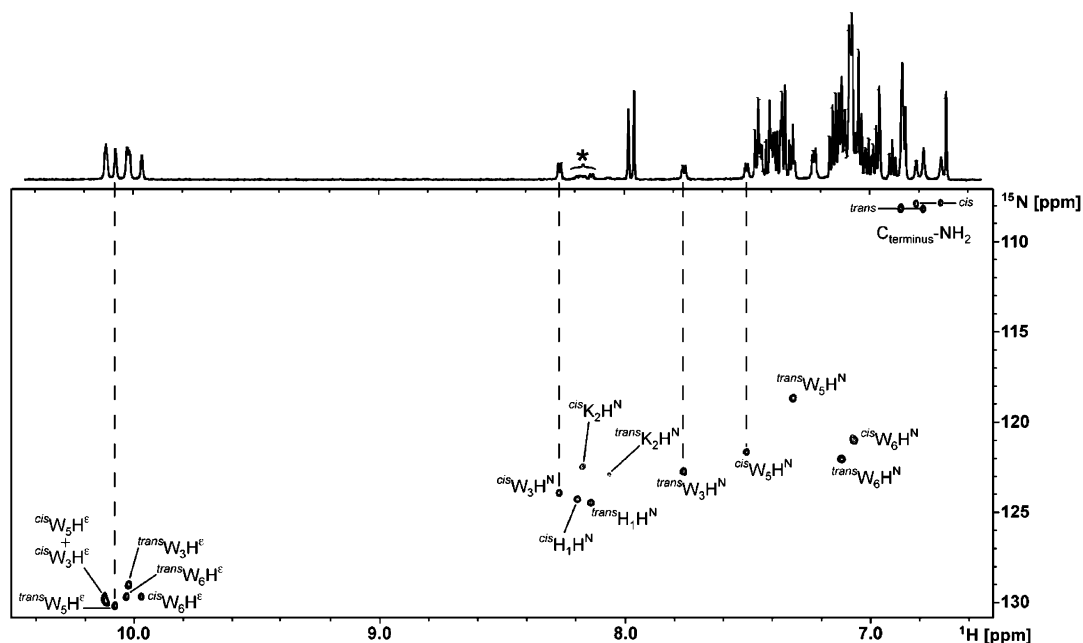


FIGURE 2: NMR spectra of the Ac-HKWPWW-NH<sub>2</sub> peptide. (Top) <sup>1</sup>H 1D spectrum of the amide and aromatic region of the peptide recorded at 600 MHz with water presaturation during the relaxation delay. (Bottom) <sup>1</sup>H,<sup>15</sup>N HSQC recorded at 600 MHz showing the assignment of the Trp side chain, the backbone amide, and the C-terminal amide resonances. The labeled signals (\*) in the 1D spectrum belong to the backbone amide resonances of the N-terminal residues His1 and Lys2 in both proline conformations. Their weak intensities are due to the water presaturation and indicate a fast exchange with water and no defined structure of the N-terminus.

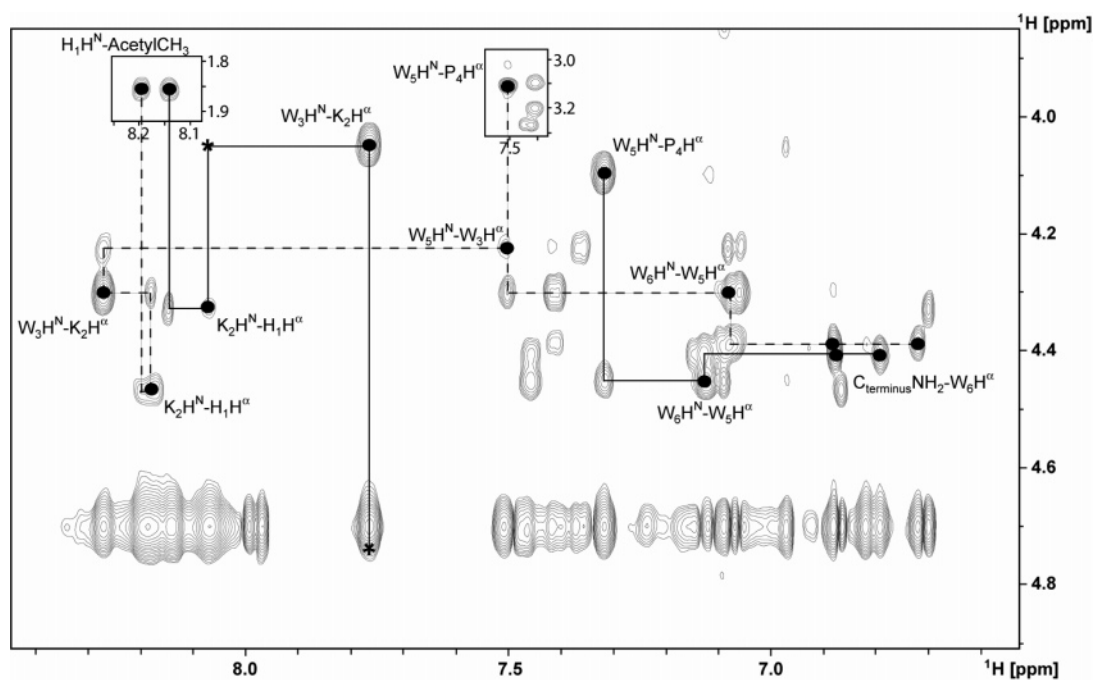


FIGURE 3: <sup>1</sup>H<sup>N</sup>,<sup>1</sup>H<sup>α</sup> region in the <sup>1</sup>H,<sup>1</sup>H ROESY spectrum of the Ac-HKWPWW-NH<sub>2</sub> peptide. The ROESY spectrum was recorded at 600 MHz with a mixing time of 400 ms. The solid and the dashed lines are the connectivity walks assigned to the *trans*-proline and the *cis*-proline conformations, respectively. The interresidual cross-peaks along the connections are labeled with filled circles (●) and assigned. The intrasidual peaks labeled with an asterisk (\*) are observable at lower contour levels and in the TOCSY experiment.

<sup>1</sup>H<sup>N</sup>,<sup>1</sup>H<sup>α</sup> region of the <sup>1</sup>H,<sup>1</sup>H ROESY spectrum, the spin systems identified in the TOCSY and COSY experiments could be assigned sequentially for each of the two conformations as shown in Figure 3.

The assignment of the two sets of cross-peaks in the 2D NMR spectra to either *cis*- or *trans*-proline conformation was based on the methods of Lee et al. (50) and Schubert et al. (51). The method of Lee et al. (50) derives the Pro conformation from the NOESY patterns around the Pro

residue. In the case of the Ac-HKWPWW-NH<sub>2</sub>, the ROESY cross-peaks with medium intensity between the H<sup>δ</sup>s of Pro4 and the H<sup>α</sup> proton of the preceding amino acid (Trp3) in one resonance set indicate that the Pro residue likely adopts a *trans* conformation. In contrast, strong cross-peaks from the preceding H<sup>α</sup> to the H<sup>α</sup> of the proline and medium intensity signals to the amide proton of the following residue (Trp5) indicate a *cis* conformation for the second set of resonances. This assignment is further confirmed by the



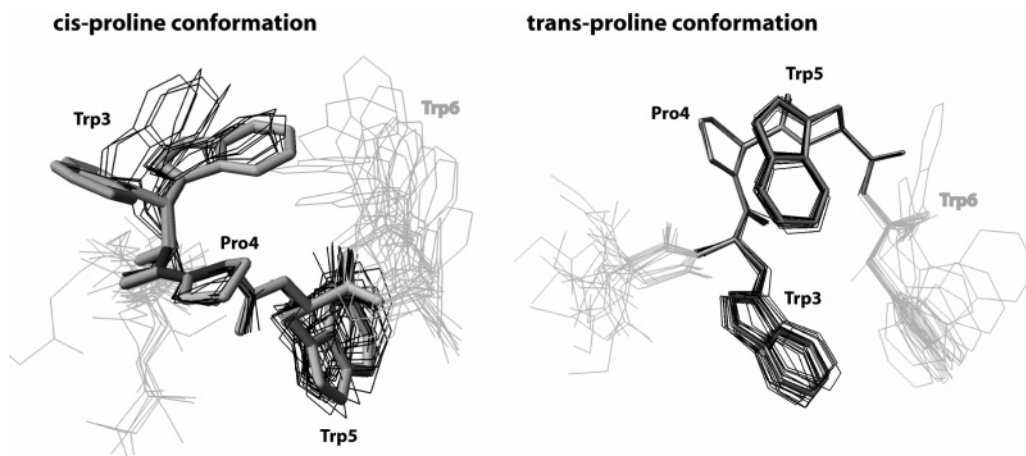


FIGURE 4: Structural ensemble of the two Ac-HKWPWW-NH<sub>2</sub> conformations. The 20 lowest energy structures are shown for both *cis*-proline (left) and *trans*-proline (right) conformations. In both ensembles, the side chains of His1 and Lys2 are not represented since they do not exhibit a preferred conformation. Trp6 shows some preferred conformations and is therefore shaded. The residues in the core region (Trp3–Trp5) are thickened. While in the *trans*-proline conformation, all members of the family of structures superimpose well, the Trp3 side chain in the *cis*-proline conformation wobbles between two conformations.

method of Schubert et al. (51), which is based on the chemical shift difference between the carbons at the  $\beta$ -position and at the  $\gamma$ -position of the Pro residue. This chemical shift difference constitutes a reference-independent indicator of the Pro peptide bond conformation based on a statistical analysis of 1033 prolines in 304 proteins deposited in the BioMagRes database. A difference  $\Delta_{\beta\gamma} = \delta[^{13}\text{C}^\beta] - \delta[^{13}\text{C}^\gamma]$  of around  $4.5 \pm 1.2$  ppm indicates a high probability of a *trans*-proline amide bond conformation, while a difference of  $9.6 \pm 1.3$  ppm is characteristic for the *cis* conformation. The Ac-HKWPWW-NH<sub>2</sub> peptide gives differences of 9.0 and 4.0 ppm for the two observed conformations, in agreement with the ROESY analysis. Peak integration and comparison in 1D and 2D spectra resulted in a *cis:trans* ratio of 46:54 with a variation of 4%. Interestingly, no significant change in this ratio was detected when the temperature was raised to 37 °C, indicating that the *cis-trans* equilibrium is preserved at physiological temperature.

Analysis of the ROESY spectrum showed no cross-peaks between protons of the two conformations, indicating a slow exchange rate between the two conformations. In addition to the intraresidual ROE cross-peaks, 15 sequential and 4 medium-range (all between Trp3 and Trp5) cross-peaks were observed between the resonances assigned to the *cis* conformation. For the *trans* conformation, 16 sequential resonances and 3 medium-range (from the H <sup>$\delta$</sup> 1 of Trp3 to H <sup>$\alpha$</sup>  of Trp5 and to one H <sup>$\beta$</sup>  of Trp6 as well as from the H <sup>$\alpha$</sup>  of Pro4 to the backbone H<sup>N</sup> of Trp6) cross-peaks were observed. The cross-peaks identified in the ROESY spectrum were converted into distance restraints for separate structure calculations using the CNS 1.1 and the ARIA 1.2 programs.

The structure calculations confirmed the results described above. Examination of the 20 lowest energy structures showed a defined structure around the Pro residue in both conformations. The peptide with the *cis*-Pro conformation (Figure 4, left) adopts a folded structure, which is supported by the 4 observed ROE cross-peaks between Trp3 and Trp5. In contrast, the *trans* conformation (Figure 4, right) exhibits an extended backbone. While in the *trans*-Pro conformation both the backbone and the side chains of the residues Trp3, Pro4, and Trp5 can be superimposed, the NMR data for the *cis*-proline conformation suggest that the Trp3 side chain

adopts two different conformations. The  $\chi^1$  angles of these two conformations are around 60° and slightly above 180°, respectively. The terminal residues, His1, Lys2, and Trp6, exhibit no preferred structure in both conformers.

The average distances and the relative orientations between the aromatic side chains of the Trp residues were determined over the 20 lowest energy structures of both conformations. In the *cis*-Pro conformation, the side chains of Trp5 and Trp6 are about 7–8 Å apart, with a planar orientation. Since the indole ring of Trp3 can adopt two conformations, there is no conserved orientation with respect to the other indole rings and the inter-ring distances are between 7 and 11 Å. Significantly shorter distances between the Trp side chains are observed in the *trans*-Pro conformation. Indeed, in the well-resolved structure of the core of the *trans*-Pro conformation, the indole side chains of Trp3 and Trp5 are only 6.2 ( $\pm 0.2$ ) Å apart, with a fixed planar orientation. Moreover, Trp6 comes even closer to Trp3 (4.6–6.5 Å) but with no conserved orientation.

*Characterization of Ac-HKWPWW-NH<sub>2</sub> by Fluorescence.* Trp residues constitute intrinsic fluorescent probes of peptides and proteins. Due to their exquisite sensitivity to the environment, Trp fluorescence properties can be used to obtain information on the overall structural properties and conformational changes of the protein or peptide in which these Trp residues are included (52). In this respect, we characterized the fluorescence properties of the three Trp residues in Ac-HKWPWW-NH<sub>2</sub>. The fluorescence intensity maximum of Ac-HKWPWW-NH<sub>2</sub> is 351 ( $\pm 1$ ) nm (Figure 5a), which is typical for fully solvent exposed Trp residues. This, in line with the NMR data, confirms that the three Trp residues do not form a hydrophobic cluster. The fluorescence quantum yield ( $\phi$ ) of Ac-HKWPWW-NH<sub>2</sub> in 20 mM Hepes, pH 7.5, buffer was found to be 0.058 ( $\pm 0.002$ ), which is about 3-fold less than the quantum yield of free Trp ( $\phi = 0.14$ ).

Moreover, the time-resolved fluorescence decay of Ac-HKWPWW-NH<sub>2</sub> is characterized by three lifetimes, 0.42, 1.35, and 3.0 ns, with relative amplitudes of 26%, 52%, and 22%, respectively (Table 1). Interestingly, the radiative rate constant,  $k_r = \phi/\langle\tau\rangle$ , was found to be  $3.9 \times 10^7 \text{ s}^{-1}$  and thus significantly smaller than that of *N*-acetyl-L-tryptophanamide

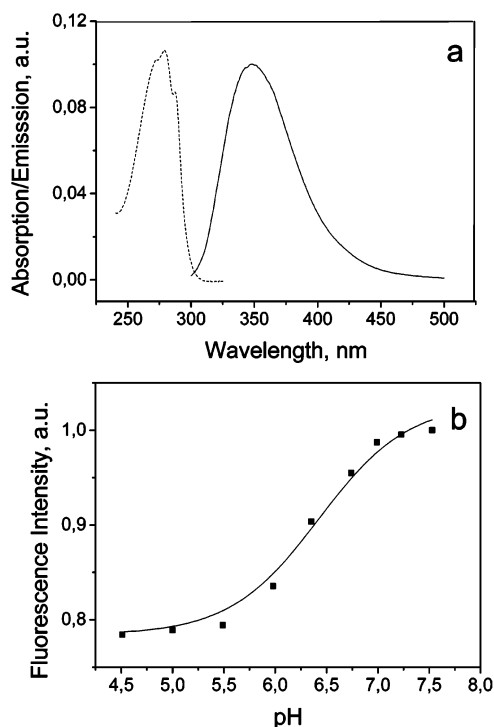


FIGURE 5: Fluorescence characterization of Ac-HKWPWW-NH<sub>2</sub>. (a) Absorption and emission spectra of Ac-HKWPWW-NH<sub>2</sub> in 50 mM Hepes, pH 7.5. The dotted and solid lines represent the absorption and emission spectrum, respectively. The excitation wavelength was at 280 nm. (b) Effect of His1 deprotonation on Ac-HKWPWW-NH<sub>2</sub> intrinsic fluorescence. The deprotonation of His1 was monitored through the intrinsic fluorescence of Ac-HKWPWW-NH<sub>2</sub>. The experiments were performed in 25 mM MES at different pH values. Excitation and emission wavelengths were 280 and 350 nm, respectively.

in water (53), unambiguously indicating static quenching. This quenching may be attributed to ground state stacking interactions that are thought to occur between Trp pairs, notably in the *trans*-proline conformation where the Trp side chains are close.

In a next step, the pK<sub>a</sub> of His1 was determined through the pH dependence of the peptide quantum yield. The rationale of this approach is that only the protonated form of His markedly quenches the fluorescence of spatially close Trp residues (54, 55). The quantum yield of Ac-HKWPWW-NH<sub>2</sub> was monitored in 25 mM Mes buffer in the pH range 4.5–7.5 (Figure 5b). We observed a steep decrease in  $\phi$  when the pH decreases. This pH dependence of  $\phi$  was fitted according to (56)

$$\phi = \phi_{\text{MH}} + (\phi_{\text{M}} - \phi_{\text{MH}})/(1 + 10^{\text{pK}_a} \times 10^{-\text{pH}}) \quad (3)$$

where  $\phi_{\text{MH}}$  and  $\phi_{\text{M}}$  are the fluorescence quantum yields of the peptide with protonated and deprotonated His, respectively. A pK<sub>a</sub> value of 6.3 ( $\pm 0.1$ ) close to that of free His was obtained, in line with its random orientation observed by NMR spectroscopy. This indicates that, at pH 7.5, about 95% of the His residues are nonprotonated.

**Binding of Ac-HKWPWW-NH<sub>2</sub> to the Viral cTAR and PBS Sequences.** To characterize the binding of Ac-HKWPWW-NH<sub>2</sub> to TAR and PBS, reverse titrations were performed by adding increasing oligonucleotide concentrations to a fixed peptide concentration (Figure 6). Binding was monitored through the decrease of Ac-HKWPWW-NH<sub>2</sub> intrinsic fluo-

rescence resulting from the binding to the oligonucleotides. Binding experiments were performed in a 25 mM Tris-HCl, pH 7.5, 30 mM NaCl, 0.2 mM MgCl<sub>2</sub> buffer which is currently used for evaluating NC destabilizing properties (10, 38, 39). For both PBS and TAR derivatives, the titrations indicate a sharp fluorescence decrease that reduces the peptide fluorescence by about 40% and 25% for TAR and PBS derivatives, respectively.

In contrast to the large decrease in the steady-state fluorescence intensity, the three fluorescence lifetimes (0.42, 1.35, and 3.0 ns) and their relative amplitudes remained essentially unchanged upon the formation of the hexapeptide–nucleic acid complexes (Table 1), suggesting the formation of dark species with a very short or null lifetime. These dark species are indicative of stacking interactions between the Trp residues of Ac-HKWPWW-NH<sub>2</sub> and both (+)PBS and cTAR (57). Similarly, very low fluorescent species have also been observed with the single Trp residue of NC. However, while these low fluorescent species represent about 90% of the total species in the complexes of NC with its viral nucleic acid targets (9, 23, 58), the dark species in Ac-HKWPWW-NH<sub>2</sub> complexed with PBS and cTAR represent only 25% and 34%, respectively. This indicates that, in the complexes, only a fraction of the Trp residues of Ac-HKWPWW-NH<sub>2</sub> is stacked with the oligonucleotide bases.

In a first step, the binding stoichiometry of Ac-HKWPWW-NH<sub>2</sub> to the TAR and PBS oligonucleotides was estimated from the intersection of the initial slope of the titration with the fluorescence plateau. The number of binding sites was found to be  $7 \pm 1$  and  $12 \pm 2$  for PBS and TAR derivatives, respectively (Table 2), giving occluded binding sizes of about 2.5 and 4.5 nucleotides per peptide, respectively. These differences in the occluded binding sizes suggest that the Ac-HKWPWW-NH<sub>2</sub> molecules probably do not saturate the TAR species and, thus, that the true occluded binding size for Ac-HKWPWW-NH<sub>2</sub> on oligonucleotides may be equal to or smaller than 2.5. In a second step, the apparent binding constant,  $K_{\text{app}}$ , was determined by fitting the binding data to eq 2. This approach is only semiquantitative, since it is based on the assumption that the binding sites for the peptide are identical and independent. The binding constant for all TAR derivatives (cTAR DNA, TAR RNA, and its DNA equivalent, TAR DNA) was found to be about  $10^6 \text{ M}^{-1}$  (Table 2), indicating that there is no strong dependence on the oligonucleotide sequence or nature (DNA or RNA). A similar conclusion also applies for PBS derivatives since Ac-HKWPWW-NH<sub>2</sub> binds to both (+)PBS and (–)PBS sequences with an affinity of  $(3\text{--}4) \times 10^6 \text{ M}^{-1}$ . Noticeably, the binding constants of Ac-HKWPWW-NH<sub>2</sub> to both TAR and PBS sequences are about 1 order of magnitude higher than those described for related peptides to the SL3 sequence of the encapsidation sequence (36). These differences may be ascribed to differences in both the experimental conditions and the peptide sequence since binding of Ac-HKWPWW-NH<sub>2</sub> to SL3 in our conditions led to  $K_{\text{app}}$  values comparable to those obtained with PBS and TAR (data not shown). Interestingly, though the Ac-HKWPWW-NH<sub>2</sub> peptide contains only six amino acids, its binding constants for both TAR and PBS derivatives are of the same order of magnitude as those reported for NC(12–55) (9, 23). This confirms that this peptide may efficiently compete with NC for its nucleic acid targets.

Table 1: Time-Resolved Fluorescence Parameters of Ac-HKWPWW-NH<sub>2</sub><sup>a</sup>

	$\alpha_0$	$\tau_1$ (ns)	$\alpha_1$	$\tau_2$ (ns)	$\alpha_2$	$\tau_3$ (ns)	$\alpha_3$	$\langle\tau\rangle$ (ns)
(+)PBS	0.25	0.30	0.27 (0.20)	1.28	0.50 (0.38)	2.90	0.23 (0.17)	1.39
cTAR	0.34	0.45	0.30 (0.20)	1.47	0.50 (0.33)	3.08	0.20 (0.13)	1.48

<sup>a</sup> The experiments were performed with 1  $\mu$ M Ac-HKWPWW-NH<sub>2</sub> either in the absence or in the presence of 1  $\mu$ M (+)PBS or cTAR. Excitation and emission wavelengths are 295 and 350 nm, respectively. The fluorescence lifetimes,  $\tau_1$  to  $\tau_3$ , the measured relative amplitudes,  $\alpha_1$  to  $\alpha_3$ , and the mean lifetime,  $\langle\tau\rangle$ , are expressed as means for three experiments. The standard errors of the mean are usually below 10% for the lifetimes and 15% for the amplitudes. The relative amplitude,  $\alpha_0$ , of the dark species is calculated by eq 1. The remaining amplitudes,  $\alpha_{ic}$  (in parentheses), were recalculated according to  $\alpha_{ic} = \alpha_i(1 - \alpha_0)$ .

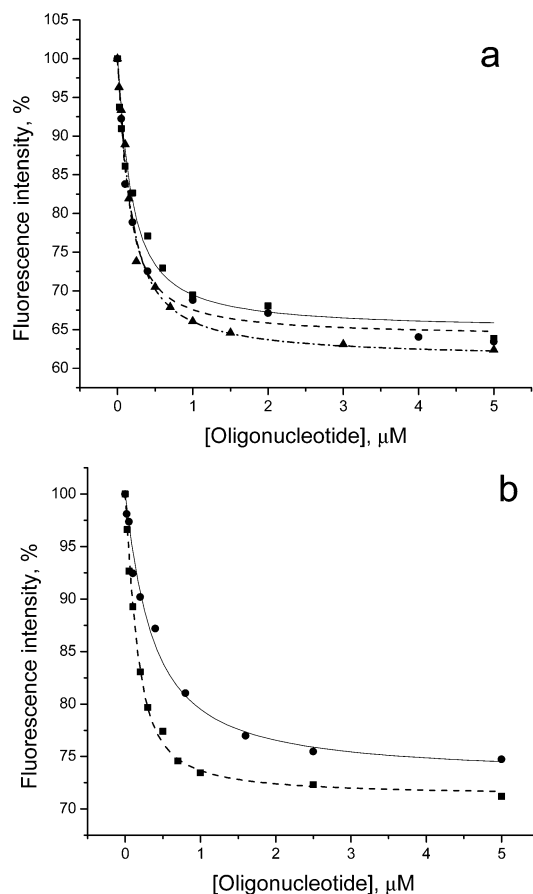


FIGURE 6: Binding curves of Ac-HKWPWW-NH<sub>2</sub> with (a) TAR and (b) PBS derivatives. The peptide concentration was 1  $\mu$ M in 25 mM Tris-HCl, 30 mM NaCl, and 0.2 mM MgCl<sub>2</sub>. In (a), the peptide was titrated with TAR RNA ( $\blacktriangle$ , dash-dotted line), TAR DNA ( $\bullet$ , dashed line), and cTAR DNA ( $\blacksquare$ , solid line). In (b), the peptide was titrated with (+)PBS ( $\bullet$ , solid line) and (-)PBS ( $\blacksquare$ , dashed line). Excitation and emission wavelengths were 295 and 350 nm, respectively. The lines represent the fits of the experimental points with eq 2.

To further characterize the binding process, the salt and temperature dependence of the binding constant of Ac-HKWPWW-NH<sub>2</sub> to cTAR was investigated (59). The dependence of the binding constant on NaCl concentration (in the range 30–500 mM) appears very small (data not shown), indicating that the binding strength relies essentially on nonelectrostatic contributions. Moreover, no change in the binding stoichiometry and only moderate affinity decrease (3-fold) of Ac-HKWPWW-NH<sub>2</sub> to cTAR were observed when the temperature was raised to 37 °C, indicating that the hexapeptide can compete with NCp7 for the binding to cTAR also at physiological temperature.

To identify by NMR the peptide conformation which is mostly responsible for the binding to the oligonucleotides,

Table 2: Binding Parameters of Ac-HKWPWW-NH<sub>2</sub> for TAR and PBS Derivatives<sup>a</sup>

sequence	$K_{app}$ ( $\times 10^{-6}$ ) (M <sup>-1</sup> )	$n$
TAR RNA	1.8	12
cTAR DNA	0.8	12
TAR DNA	1.1	12
(+)PBS	3.0	7
(-)PBS	4.0	7

<sup>a</sup> Experiments were performed in 25 mM Tris-HCl, 30 mM NaCl, and 0.2 mM MgCl<sub>2</sub>, pH 7.5. The number of binding sites was determined as described in the text. The  $K_{app}$  values were calculated by fitting the data in Figure 6 to eq 2. Indicated values correspond to the mean of two to three experiments. The standard error of the mean was usually of 2 for  $n$  and about 30% for  $K_{app}$ .

Table 3: Chemical Shift Changes of Trp Indole Resonances upon Titration of Ac-HKWPWW-NH<sub>2</sub> with (-)PBS<sup>a</sup>

peptide: DNA ratio	chemical shift changes of Trp indole resonances in the two peptide conformations (ppm)					
	cis W3	cis W5	cis W6	trans W3	trans W5	trans W6
4:1	0.07	0.03	0.05	0.04	0.06	0.05
2:1	0.16	0.08	0.11	0.11	0.14	0.12
1:1	0.24	0.12	0.17	0.16	0.21	0.17
1:4	0.37	0.22	0.28	0.25	0.30	0.26

<sup>a</sup> The experiments were performed with 0.4 mM Ac-HKWPWW-NH<sub>2</sub> and increasing concentrations of (-)PBS at 310 K. The 1D spectra were calibrated on the water signal. The signals overlapping in the 1D experiments were resolved in 2D experiments.

a titration was performed by adding increasing amounts of (-)PBS DNA to a fixed concentration of Ac-HKWPWW-NH<sub>2</sub>. Already at the first step of the titration at 298 K where (-)PBS was added to the peptide at a 1:4 molar ratio, a large broadening of all peptide proton signals together with a slight turbidity was observed. This turbidity disappeared when (-)PBS was added to a 1:1 molar ratio, but the signals remained broad and unresolved. The signal intensities and resolution were recovered partly by increasing the temperature to 310 K and by switching to a lower frequency spectrometer (400 MHz). These observations are characteristic of an intermediate exchange of the peptide with the DNA, in line with the binding constant of the peptide to (-)PBS. Moreover, the significant changes of the Trp indole resonances in both *cis*-Pro and *trans*-Pro conformations during titration with (-)PBS (Table 3) indicate that both peptide conformations bind to (-)PBS with comparable affinities. All Trp indole signals shifted upfield during the titration, in line with a stacking of the indole rings with the bases of (-)PBS. In addition, the amplitudes of these shifts can be used to identify which Trp residues are the most important for interaction. In the *cis*-Pro conformation, the large shift displayed by Trp3 suggests that this residue plays a central role in the interaction with (-)PBS. In contrast,

Table 4: Effects of Ac-HKWPWW-NH<sub>2</sub> and NC on cTAR Secondary Structure As Measured by Time-Resolved Fluorescence Spectroscopy<sup>a</sup>

TMR-5'-cTAR-3'-Fl	$\alpha_0$	$\tau_1$ (ns)	$\alpha_1$	$\tau_2$ (ns)	$\alpha_2$	$\tau_3$ (ns)	$\alpha_3$	$\langle\tau\rangle_S$ (ns)	$\langle\tau\rangle_R$ (ns)	$R_m$
	0.732	0.16 ± 0.01	0.130	1.23 ± 0.10	0.043	3.92 ± 0.01	0.096	1.77 ± 0.04	3.86	8.1
+NC	0.443	0.15 ± 0.01	0.089	1.30 ± 0.03	0.236	3.7 ± 0.1	0.227	2.27 ± 0.02	3.91	3.1
+Ac-HKWPWW-NH <sub>2</sub>	0.852	0.14 ± 0.03	0.052	1.23 ± 0.09	0.020	4.05 ± 0.01	0.075	2.3 ± 0.1	3.95	11.8
+Ac-HKWPWW-NH <sub>2</sub> + NC	0.601	0.30 ± 0.10	0.067	1.42 ± 0.09	0.153	3.7 ± 0.1	0.178	2.3 ± 0.1	3.85	4.2

<sup>a</sup> Experiments were performed in the same buffer as in Table 2. Excitation and emission wavelengths were 480 and 520 nm, respectively. Concentrations used: TMR-5'-cTAR-3'-Fl, 0.1  $\mu$ M; Ac-HKWPWW-NH<sub>2</sub>, 7  $\mu$ M; NC(12-55), 1.1  $\mu$ M. The fluorescence lifetimes,  $\tau_1$  to  $\tau_3$ , the relative amplitudes,  $\alpha_1$  to  $\alpha_3$ , and the mean lifetime,  $\langle\tau\rangle$ , are expressed as means for three experiments. The relative amplitude,  $\alpha_0$ , of the dark species and the remaining amplitudes,  $\alpha_{ic}$ , were recalculated as in Table 1.

the modest shift observed with Trp5 suggests that this residue is only moderately involved in the complexes formed with this peptide conformation. A more symmetric situation was observed with the *trans*-Pro conformation, since the three Trp residues displayed comparable chemical shifts during their interaction with (-)PBS. Nevertheless, since at all peptide:oligonucleotide ratios, the changes in the chemical shifts of Trp5 were systematically higher than those of Trp3 and Trp6, this suggests that Trp5 may contribute slightly more than the two other Trp residues in the complexes of the *trans*-Pro conformation with (-)PBS. Moreover, since time-resolved fluorescence data show that only a fraction of the Trp residues is involved in stacking interactions, the binding of both peptide conformations is probably heterogeneous with the number and nature of stacked Trp residues differing from one bound peptide to another.

*Ac-HKWPWW-NH<sub>2</sub> Stabilizes the Secondary Structure of PBS and cTAR.* Since Ac-HKWPWW-NH<sub>2</sub> binds to TAR and PBS with affinities similar to that of NC, this prompted us to determine if, by analogy to NC, the peptide exhibits nucleic acid chaperone activities. To this end, we analyzed the interaction of Ac-HKWPWW-NH<sub>2</sub> with cTAR doubly labeled at its 5' and 3' ends by TMR and Fl, respectively. As previously shown, the TMR and Fl dyes form a nonfluorescent heterodimer when the stem of cTAR is closed (10, 38). However, due to fraying of the terminal bases of cTAR stem, the nonfluorescent closed species was shown to be in equilibrium with partly melted fluorescent species (where Fl was removed away from TMR) (10, 38, 39). Since, in melted species, Fl and TMR undergo FRET, the time-resolved fluorescence technique has been shown to accurately identify the melting degree of the various species in solution, as well as their population (10). In this respect, the three lifetime components,  $\tau_1$  to  $\tau_3$ , of TMR-5'-cTAR-3'-Fl (Table 4) have been associated with the melting of the terminal 3 bp segment, the lower half of cTAR, and the whole cTAR molecule, respectively (10, 38). These partly melted species are in equilibrium with a population,  $\alpha_0$ , of closed dark species (associated with a null lifetime) representing 73% in the present batch of cTAR molecules. NC shifted the equilibrium toward the partly melted species (as could be seen from the increase in  $\alpha_1$  to  $\alpha_3$  values), increasing the fluorescence intensity of TMR-5'-cTAR-3'-Fl by a factor of about 3 (Table 4). In sharp contrast, addition of increasing concentrations of Ac-HKWPWW-NH<sub>2</sub> induced a fluorescence decrease (amounting to about 20% in the presence of an excess of peptide) with no significant change in the ratio of the Fl to TMR peaks (Figure 7a). This fluorescence decrease is due to a significant increase of the  $\alpha_0$  value at the expense of all other species, suggesting that Ac-HKWPWW-NH<sub>2</sub> decreases the population of melted species

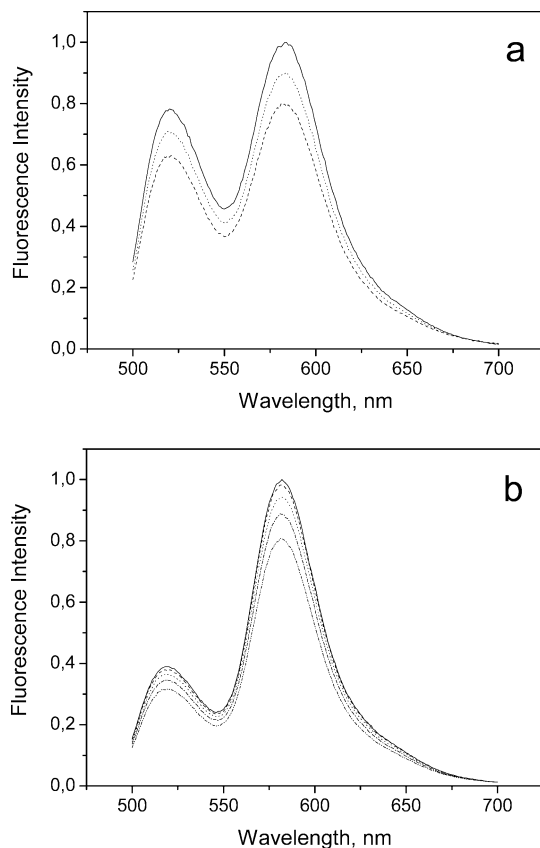


FIGURE 7: Stabilization by Ac-HKWPWW-NH<sub>2</sub> of the secondary structures of cTAR and (-)PBS. (a) The fluorescence spectrum of TMR-5'-cTAR-3'-Fl (100 nM) was recorded in the absence (solid line) or in the presence of Ac-HKWPWW-NH<sub>2</sub> at a concentration of 5  $\mu$ M (dotted line) or 10  $\mu$ M (dashed line), respectively. (b) Fluorescence spectra of 100 nM TMR-5'-(-)PBS-3'-Fl in the absence (solid line) and in the presence (dotted lines) of 1, 2, 5, and 10  $\mu$ M Ac-HKWPWW-NH<sub>2</sub>.

to the benefit of the dark species and thus that the peptide stabilizes the cTAR secondary structure. Interestingly, Ac-HKWPWW-NH<sub>2</sub> induced also a decrease of the fluorescence intensity (by about 8%) at 37 °C, although to a lesser extent than at 20 °C.

In a next step, we analyzed the interaction of Ac-HKWPWW-NH<sub>2</sub> with TMR-5'-(-)PBS-3'-Fl. This oligonucleotide was also shown to be in equilibrium between closed and open species (9, 60). In contrast to cTAR, NC has been shown to induce only a limited 1.4-fold increase in TMR-5'-(-)PBS-3'-Fl fluorescence, due to a limited redistribution between the already partly opened species. As for cTAR, addition of an excess of Ac-HKWPWW-NH<sub>2</sub> decreases the fluorescence of TMR-5'-(-)PBS-3'-Fl by about 20% (Figure 7b), suggesting that Ac-HKWPWW-NH<sub>2</sub> also stabilizes the secondary structure of the PBS. The absence

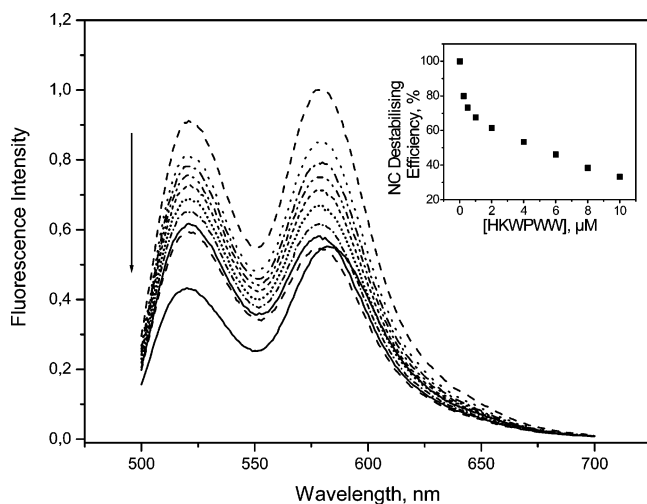


FIGURE 8: Inhibition by Ac-HKWPWW-NH<sub>2</sub> of NC(12–55) destabilizing activity. The bottom spectrum (solid line) and top spectrum (dashed line) correspond to 100 nM TMR-5'-cTAR-3'-Fl in the absence and the presence of 1.1 μM NC(12–55), respectively. The dotted lines describe the spectra of the TMR-5'-cTAR-3'-Fl/NC(12–55) complexes in the presence of increasing concentrations (as indicated by the arrow) of Ac-HKWPWW-NH<sub>2</sub>. Inset: Inhibition by Ac-HKWPWW-NH<sub>2</sub> of the NC(12–55)-directed destabilization of cTAR. The inhibition was calculated on basis of the Fl peak intensity of TMR-5'-cTAR-3'-Fl.

of Ac-HKWPWW-NH<sub>2</sub>-promoted destabilization of (–)PBS stem is further supported by the absence of changes in the chemical shift, line width, and relative intensity of the imino proton signals of the four (–)PBS base pairs in the NMR titration.

*Ac-HKWPWW-NH<sub>2</sub> Inhibits NC-Directed Destabilization of cTAR.* Since Ac-HKWPWW-NH<sub>2</sub> and NC(12–55) show similar binding affinities but opposite effects on cTAR stability, we examined the competing effect of Ac-HKWPWW-NH<sub>2</sub> on NC-directed destabilization of cTAR secondary structure. For these experiments, 100 nM TMR-5'-cTAR-3'-Fl was first incubated with NC(12–55) at a nucleotide to protein molar ratio of 5. As shown in Figure 8, this results in a sharp fluorescence intensity increase due to the partial melting of cTAR secondary structure. Addition of increasing concentrations of Ac-HKWPWW-NH<sub>2</sub> gradually reduces the fluorescence intensity of both Fl and TMR peaks, suggesting that Ac-HKWPWW-NH<sub>2</sub> competitively inhibits NC destabilizing activity. At the highest concentration (10 μM) of Ac-HKWPWW-NH<sub>2</sub>, the intensity of the Fl peak was still 40% higher than the Fl intensity in the free oligonucleotide, while the intensity of the TMR peak was reduced to the level of the TMR peak in the free oligonucleotide. This indicates that the FRET between Fl and TMR was reduced with respect to that in the free oligonucleotide and, thus, that cTAR equilibrium is still partly shifted toward the melted forms. In line with this partial inhibition of NC destabilizing activity, the time-resolved measurements showed that in the presence of both peptides the amplitude of the closed species ( $\alpha_0$ ) and the amplitudes of the partly melted species ( $\alpha_1$  to  $\alpha_3$ ) were intermediate to those of the free cTAR and the cTAR/NC complexes (Table 4). Taken together, our data indicate that Ac-HKWPWW-NH<sub>2</sub> inhibits NC destabilizing activity, probably by competing with NC for binding cTAR. This competition between NC and Ac-HKWPWW-NH<sub>2</sub> for binding cTAR was further confirmed by adding increasing

NC concentrations to a cTAR–Ac-HKWPWW-NH<sub>2</sub> complex. We observed a clear fluorescence increase of TMR-5'-cTAR-3'-Fl fluorescence, indicating that NC(12–55) displaces Ac-HKWPWW-NH<sub>2</sub> and exerts its destabilizing activity (data not shown). On the basis of these conclusions, the EC<sub>50</sub> concentration of Ac-HKWPWW-NH<sub>2</sub> that inhibits 50% of the NC destabilizing activity (under the conditions described in the legend of Figure 8) was found to be about 4 μM as determined from the Fl intensity peak (Figure 8, inset). Interestingly, the effect of Ac-HKWPWW-NH<sub>2</sub> on NC destabilizing activity at 37 °C was similar to that described at 20 °C, with an EC<sub>50</sub> of 3.5 μM, indicating that the competing effect of the peptide is preserved at physiological temperature.

## DISCUSSION

In the present work, we characterized the structure and the fluorescence properties of Ac-HKWPWW-NH<sub>2</sub> and its potency to inhibit NC destabilizing activity. After structural calculations and free energy minimization, a well-defined structure was found around the central Pro residue. Moreover, the peptide was found to exhibit two slowly interconverting conformations that were attributed to *cis*- and *trans*-Pro conformations. In contrast to His1, Lys2, and Trp6 residues which showed no preferred orientation, Trp3 and Trp5 adopted defined orientations in both conformations. In the *trans*-Pro conformation, the proximity of Trp3 and Trp5 indole rings as well as their planar orientation suggests a stacking interaction between the two indole rings. Such an interaction may also exist between Trp3 and Trp6 side chains in the *trans* conformation as well as between Trp5 and Trp6 in the *cis* conformation. Since fluorescence quenching due to ground state stacking interactions has been reported for various fluorescence pairs (61–63), stacking interactions between Trp side chains may explain the low quantum yield of the peptide as well as the static fluorescence quenching (Table 1). As described by the exciton theory, these ground state interactions between parallel Trp side chains in an H-type geometry [in which the transition dipoles of the individual chromophores are parallel to each other and normal to the radius vector connecting them (64)] probably allow strong coupling between the transition dipoles of the Trp side chains and delocalization of excitation over them, preventing fluorescence emission (64–66). In contrast, since Lys, Pro, and deprotonated His residues are poor quenchers of Trp fluorescence (ref 55 and references cited therein), these residues probably do not significantly contribute to the low quantum yield of Ac-HKWPWW-NH<sub>2</sub>.

Ac-HKWPWW-NH<sub>2</sub> was found to bind with good affinity to both the PBS and TAR viral sequences, confirming the potency of this type of peptide to compete with NC for its target nucleic acid elements (36). Since the binding appears to be mainly nonelectrostatic, it is likely that the stacking of the Trp residues with the bases of PBS and TAR derivatives (Tables 2 and 3) governs the binding of Ac-HKWPWW-NH<sub>2</sub> to both PBS and TAR target sequences. In the case of (–)PBS, the two peptide conformations were found to bind with comparable affinities. Though all three Trp residues of Ac-HKWPWW-NH<sub>2</sub> contribute (but not necessarily simultaneously) to the binding of both peptide conformations to (–)PBS, the key residues in the complexes with *cis*-Pro and *trans*-Pro conformations are Trp3 and Trp5, respectively.

This conclusion is in line with the critical role of Trp3 and Trp5 evidenced for the binding to SL3 (36). The central contribution of the Trp residues of Ac-HKWPWW-NH<sub>2</sub> in the binding to oligonucleotides is also fully consistent with the critical role of the Trp37 residue of NC, which strongly contributes through its stacking (mainly with guanine residues) (67, 68) to the overall stability of the NC-oligonucleotide complexes (23, 58, 69). In addition to Trp residues, His1 has also been reported to play an important role in the binding process (36). Since His is largely unprotonated at pH 7.5 (Figure 5b), it should thus mainly interact with oligonucleotides through nonelectrostatic interactions.

The good affinity of Ac-HKWPWW-NH<sub>2</sub> for NC nucleic acid targets suggests that it partly mimics the hydrophobic platform formed by the Val13, Phe16, Thr 24, Ala25, Trp37, and Met46 residues in the two folded zinc finger motifs (11, 16). This platform has been shown to be critical for strong binding (16, 67, 68, 70, 71) as well as for nucleic acid destabilization (23). This latter property is clearly not realized with Ac-HKWPWW-NH<sub>2</sub> which is unable to destabilize both PBS and TAR derivatives. This may be attributed to the fact that the proper orientation of the residues in the NC hydrophobic platform is much more critical for destabilization than for binding (23). Moreover, the lack of ionic interactions involved in the destabilization properties of NC may also contribute to the inability of the hexapeptide to destabilize the nucleic acids. Due to its flexible structure, Ac-HKWPWW-NH<sub>2</sub> can probably adopt appropriate conformations for binding; however, the subsequent molecular steps related to NC unwinding activity which rely on the constrained structure of the two zinc finger motifs cannot be mimicked by Ac-HKWPWW-NH<sub>2</sub>.

In fact, Ac-HKWPWW-NH<sub>2</sub> rather stabilizes both cTAR and PBS secondary structures. Since the stacking of the Trp residues of Ac-HKWPWW-NH<sub>2</sub> with the oligonucleotide bases strongly contributes to the binding affinity, this stacking may also be involved in the stabilization of the oligonucleotides. An exactly opposite conclusion was drawn for NC where the stacking of Trp37 is required to unwind cTAR secondary structure (23). This indicates that the destabilization of cTAR by NC probably results from a concerted activity of Trp37 with the other residues of the NC hydrophobic platform that cannot be mimicked by only the Trp residues of the hexapeptide.

Thus, in full line with its good affinity for cTAR, Ac-HKWPWW-NH<sub>2</sub> efficiently inhibits NC destabilizing activity on the cTAR secondary structure (at both 20 and 37 °C) through competition for the same binding sites on cTAR. This suggests that, in addition to its potential interference with the packaging process of HIV-1 RNA genome (36), Ac-HKWPWW-NH<sub>2</sub> may potentially interfere with NC during the two-strand transfer reactions in reverse transcription in vivo. This peptide can now serve to derive modified peptides or small molecules with improved binding properties in order to inhibit NC functions during HIV-1 replication.

#### SUPPORTING INFORMATION AVAILABLE

Expanded Materials and Methods section. This material is available free of charge via the Internet at <http://pubs.acs.org>.

#### REFERENCES

- Clever, J., Sasseti, C., and Parslow, T. G. (1995) RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1, *J. Virol.* 69, 2101–2109.
- Mihailescu, M. R., and Marino, J. P. (2004) A proton-coupled dynamic conformational switch in the HIV-1 dimerization initiation site kissing complex, *Proc. Natl. Acad. Sci. U.S.A.* 101, 1189–1194.
- Tisne, C., Roques, B. P., and Dardel, F. (2004) The annealing mechanism of HIV-1 reverse transcription primer onto the viral genome, *J. Biol. Chem.* 279, 3588–3595.
- Lapadat-Tapolsky, M., Gabus, C., Rau, M., and Darlix, J. L. (1997) Possible roles of HIV-1 nucleocapsid protein in the specificity of proviral DNA synthesis and in its variability, *J. Mol. Biol.* 268, 250–260.
- Levin, J. G., Guo, J., Rouzina, I., and Musier-Forsyth, K. (2005) Nucleic acid chaperone activity of HIV-1 nucleocapsid protein: critical role in reverse transcription and molecular mechanism, *Prog. Nucleic Acid Res. Mol. Biol.* 80, 217–286.
- Bampi, C., Jacquenet, S., Lener, D., Decimo, D., and Darlix, J. L. (2004) The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance, *Int. J. Biochem. Cell Biol.* 36, 1668–1686.
- Rein, A., Henderson, L. E., and Levin, J. G. (1998) Nucleic-acid-chaperone activity of retroviral nucleocapsid proteins: significance for viral replication, *Trends Biochem. Sci.* 23, 297–301.
- Hargittai, M. R., Gorelick, R. J., Rouzina, I., and Musier-Forsyth, K. (2004) Mechanistic insights into the kinetics of HIV-1 nucleocapsid protein-facilitated tRNA annealing to the primer binding site, *J. Mol. Biol.* 337, 951–968.
- Egele, C., Schaub, E., Ramalanjaona, N., Piemont, E., Ficheux, D., Roques, B., Darlix, J. L., and Mely, Y. (2004) HIV-1 nucleocapsid protein binds to the viral DNA initiation sequences and chaperones their kissing interactions, *J. Mol. Biol.* 342, 453–466.
- Bernacchi, S., Stoylov, S., Piemont, E., Ficheux, D., Roques, B. P., Darlix, J. L., and Mely, Y. (2002) HIV-1 nucleocapsid protein activates transient melting of least stable parts of the secondary structure of TAR and its complementary sequence, *J. Mol. Biol.* 317, 385–399.
- Summers, M. F., Henderson, L. E., Chance, M. R., Bess, J. W., Jr., South, T. L., Blake, P. R., Sagi, I., Perez-Alvarado, G., Sowder, R. C., III, Hare, D. R., et al. (1992) Nucleocapsid zinc fingers detected in retroviruses: EXAFS studies of intact viruses and the solution-state structure of the nucleocapsid protein from HIV-1, *Protein Sci.* 1, 563–574.
- Gorelick, R. J., Gagliardi, T. D., Bosche, W. J., Wiltrout, T. A., Coren, L. V., Chabot, D. J., Lifson, J. D., Henderson, L. E., and Arthur, L. O. (1999) Strict conservation of the retroviral nucleocapsid protein zinc finger is strongly influenced by its role in viral infection processes: characterization of HIV-1 particles containing mutant nucleocapsid zinc-coordinating sequences, *Virology* 256, 92–104.
- Guo, J., Wu, T., Anderson, J., Kane, B. F., Johnson, D. G., Gorelick, R. J., Henderson, L. E., and Levin, J. G. (2000) Zinc finger structures in the human immunodeficiency virus type 1 nucleocapsid protein facilitate efficient minus- and plus-strand transfer, *J. Virol.* 74, 8980–8988.
- Berthou, L., Pechoux, C., and Darlix, J. L. (1999) Multiple effects of an anti-human immunodeficiency virus nucleocapsid inhibitor on virus morphology and replication, *J. Virol.* 73, 10000–10009.
- Ottmann, M., Gabus, C., and Darlix, J. L. (1995) The central globular domain of the nucleocapsid protein of human immunodeficiency virus type 1 is critical for virion structure and infectivity, *J. Virol.* 69, 1778–1784.
- Morellet, N., de Rocquigny, H., Mely, Y., Jullian, N., Demene, H., Ottmann, M., Gerard, D., Darlix, J. L., Fournie-Zaluski, M. C., and Roques, B. P. (1994) Conformational behaviour of the active and inactive forms of the nucleocapsid NcP7 of HIV-1 studied by <sup>1</sup>H NMR, *J. Mol. Biol.* 235, 287–301.
- Remy, E., de Rocquigny, H., Petitjean, P., Muriaux, D., Theilleux, V., Paoletti, J., and Roques, B. P. (1998) The annealing of tRNA<sup>Lys</sup> to human immunodeficiency virus type 1 primer binding site is critically dependent on the NcP7 zinc fingers structure, *J. Biol. Chem.* 273, 4819–4822.
- Rong, L., Liang, C., Hsu, M., Kleiman, L., Petitjean, P., de Rocquigny, H., Roques, B. P., and Wainberg, M. A. (1998) Roles

- of the human immunodeficiency virus type 1 nucleocapsid protein in annealing and initiation versus elongation in reverse transcription of viral negative-strand strong-stop DNA, *J. Virol.* **72**, 9353–9358.
19. Williams, M. C., Rouzina, I., Wenner, J. R., Gorelick, R. J., Musier-Forsyth, K., and Bloomfield, V. A. (2001) Mechanism for nucleic acid chaperone activity of HIV-1 nucleocapsid protein revealed by single molecule stretching, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6121–6126.
  20. Guo, J., Wu, T., Kane, B. F., Johnson, D. G., Henderson, L. E., Gorelick, R. J., and Levin, J. G. (2002) Subtle alterations of the native zinc finger structures have dramatic effects on the nucleic acid chaperone activity of human immunodeficiency virus type 1 nucleocapsid protein, *J. Virol.* **76**, 4370–4378.
  21. Heath, M. J., Derebail, S. S., Gorelick, R. J., and DeStefano, J. J. (2003) Differing roles of the N- and C-terminal zinc fingers in human immunodeficiency virus nucleocapsid protein-enhanced nucleic acid annealing, *J. Biol. Chem.* **278**, 30755–30763.
  22. Lee, N., Gorelick, R. J., and Musier-Forsyth, K. (2003) Zinc finger-dependent HIV-1 nucleocapsid protein-TAR RNA interactions, *Nucleic Acids Res.* **31**, 4847–4855.
  23. Beltz, H., Clauss, C., Piemont, E., Ficheux, D., Gorelick, R. J., Roques, B., Gabus, C., Darlix, J. L., de Rocquigny, H., and Mely, Y. (2005) Structural determinants of HIV-1 nucleocapsid protein for cTAR DNA binding and destabilization, and correlation with inhibition of self-primed DNA synthesis, *J. Mol. Biol.* **348**, 1113–1126.
  24. Williams, M. C., Rouzina, I., and Bloomfield, V. A. (2002) Thermodynamics of DNA interactions from single molecule stretching experiments, *Acc. Chem. Res.* **35**, 159–166.
  25. De Rocquigny, H., Gabus, C., Vincent, A., Fournie-Zaluski, M. C., Roques, B., and Darlix, J. L. (1992) Viral RNA annealing activities of human immunodeficiency virus type 1 nucleocapsid protein require only peptide domains outside the zinc fingers, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 6472–6476.
  26. Mayasundari, A., Rice, W. G., Diminnie, J. B., and Baker, D. C. (2003) Synthesis, resolution, and determination of the absolute configuration of the enantiomers of *cis*-4,5-dihydroxy-1,2-dithiane 1,1-dioxide, an HIV-1NCp7 inhibitor, *Bioorg. Med. Chem.* **11**, 3215–3219.
  27. Basurur, V., Song, Y., Mazur, S. J., Higashimoto, Y., Turpin, J. A., Rice, W. G., Inman, J. K., and Appella, E. (2000) Inactivation of HIV-1 nucleocapsid protein P7 by pyridinioalkanoyl thioesters. Characterization of reaction products and proposed mechanism of action, *J. Biol. Chem.* **275**, 14890–14897.
  28. Gorelick, R. J., Benveniste, R. E., Lifson, J. D., Yovandich, J. L., Morton, W. R., Kuller, L., Flynn, B. M., Fisher, B. A., Rossio, J. L., Piatak, M., Jr., Bess, J. W., Jr., Henderson, L. E., and Arthur, L. O. (2000) Protection of Macaca nemestrina from disease following pathogenic simian immunodeficiency virus (SIV) challenge: utilization of SIV nucleocapsid mutant DNA vaccines with and without an SIV protein boost, *J. Virol.* **74**, 11935–11949.
  29. Druillennec, S., Meudal, H., Roques, B. P., and Fournie-Zaluski, M. C. (1999) Nucleomimetic strategy for the inhibition of HIV-1 nucleocapsid protein NCp7 activities, *Bioorg. Med. Chem. Lett.* **9**, 627–632.
  30. Darlix, J. L., Cristofari, G., Rau, M., Pechoux, C., Berthoux, L., and Roques, B. (2000) Nucleocapsid protein of human immunodeficiency virus as a model protein with chaperoning functions and as a target for antiviral drugs, *Adv. Pharmacol.* **48**, 345–372.
  31. Rice, W. G., Supko, J. G., Malspeis, L., Buckheit, R. W., Jr., Clanton, D., Bu, M., Graham, L., Schaeffer, C. A., Turpin, J. A., Domagala, J., Gogliotti, R., Bader, J. P., Halliday, S. M., Coren, L., Sowder, R. C., II, Arthur, L. O., and Henderson, L. E. (1995) Inhibitors of HIV nucleocapsid protein zinc fingers as candidates for the treatment of AIDS, *Science* **270**, 1194–1197.
  32. Allen, P., Collins, B., Brown, D., Hostomsky, Z., and Gold, L. (1996) A specific RNA structural motif mediates high affinity binding by the HIV-1 nucleocapsid protein (NCp7), *Virology* **225**, 306–315.
  33. Stephen, A. G., Worthy, K. M., Towler, E., Mikovits, J. A., Sei, S., Roberts, P., Yang, Q. E., Akee, R. K., Klausmeyer, P., McCloud, T. G., Henderson, L., Rein, A., Covell, D. G., Currens, M., Shoemaker, R. H., and Fisher, R. J. (2002) Identification of HIV-1 nucleocapsid protein: nucleic acid antagonists with cellular anti-HIV activity, *Biochem. Biophys. Res. Commun.* **296**, 1228–1237.
  34. Berglund, J. A., Charpentier, B., and Rosbash, M. (1997) A high affinity binding site for the HIV-1 nucleocapsid protein, *Nucleic Acids Res.* **25**, 1042–1049.
  35. Kim, M. Y., and Jeong, S. (2003) RNA aptamers that bind the nucleocapsid protein contain pseudoknots, *Mol. Cells* **16**, 413–417.
  36. Pustowka, A., Dietz, J., Ferner, J., Baumann, M., Landersz, M., Konigs, C., Schwalbe, H., and Dietrich, U. (2003) Identification of peptide ligands for target RNA structures derived from the HIV-1 packaging signal psi by screening phage-displayed peptide libraries, *ChemBioChem* **4**, 1093–1097.
  37. de Rocquigny, H., Ficheux, D., Gabus, C., Fournie-Zaluski, M. C., Darlix, J. L., and Roques, B. P. (1991) First large scale chemical synthesis of the 72 amino acid HIV-1 nucleocapsid protein NCp7 in an active form, *Biochem. Biophys. Res. Commun.* **180**, 1010–1018.
  38. Beltz, H., Azoulay, J., Bernacchi, S., Clamme, J. P., Ficheux, D., Roques, B., Darlix, J. L., and Mely, Y. (2003) Impact of the terminal bulges of HIV-1 cTAR DNA on its stability and the destabilizing activity of the nucleocapsid protein NCp7, *J. Mol. Biol.* **328**, 95–108.
  39. Azoulay, J., Clamme, J. P., Darlix, J. L., Roques, B. P., and Mely, Y. (2003) Destabilization of the HIV-1 complementary sequence of TAR by the nucleocapsid protein through activation of conformational fluctuations, *J. Mol. Biol.* **326**, 691–700.
  40. Bodenhausen, G., and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy, *Chem. Phys. Lett.* **69**, 185–189.
  41. Cavanagh, J., Palmer, A. G., III, Wright, P. E., and Rance, M. (1991) Sensitivity improvement in proton-detected 2-dimensional heteronuclear relay spectroscopy, *J. Magn. Reson.* **91**, 429–436.
  42. Kay, L. E., Keifer, P., and Saarinen, T. (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity, *J. Am. Chem. Soc.* **114**, 10663–10665.
  43. Braunschweiler, L., and Ernst, R. R. (1983) Coherence transfer by isotropic mixing: application to proton correlation spectroscopy, *J. Magn. Reson.* **53**, 521–528.
  44. Bothner-By, A. A., and Noggle, J. H. (1979) Time development of nuclear Overhauser effects in multispin systems, *J. Am. Chem. Soc.* **101**, 5152–5155.
  45. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **54** (Part 5), 905–921.
  46. Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA, *Methods Enzymol.* **339**, 71–90.
  47. Linge, J. P., O'Donoghue, S. I., and Nilges, M. (1999) Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation, *J. Biomol. NMR* **13**, 51–59.
  48. Eisinger, J., and Navon, G. (1969) Fluorescence quenching and isotope effect of tryptophan, *J. Chem. Phys.* **50**, 2069–2077.
  49. Livesey, A. K., and Brochon, J. C. (1987) Analyzing the distribution of decay constants in pulse-fluorimetry using the maximum entropy method, *Biophys. J.* **67**, 2013–2023.
  50. Lee, Y. C., Jackson, P. L., Jablonsky, M. J., Muccio, D. D., Pfister, R. R., Haddox, J. L., Sommers, C. I., Anantharamaiah, G. M., and Chaddha, M. (2001) NMR conformational analysis of *cis* and *trans* proline isomers in the neutrophil chemoattractant, *N*-acetylproline-glycine-proline, *Biopolymers* **58**, 548–561.
  51. Schubert, M., Labudde, D., Oschkinat, H., and Schmieder, P. (2002) A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on <sup>13</sup>C chemical shift statistics, *J. Biomol. NMR* **24**, 149–154.
  52. Lakowicz, J. R. (1999) *Principles of Fluorescence Spectroscopy*, 2nd ed., Plenum Publishing, New York.
  53. Werner, T. C., and Forster, L. S. (1979) The fluorescence of tryptophanyl peptides, *Photochem. Photobiol.* **29**, 905–914.
  54. Shinitzky, M., and Goldman, R. (1967) Fluorometric detection of histiine-tryptophan complexes in peptides and proteins, *Eur. J. Biochem.* **3**, 139–144.
  55. Chen, Y., and Barkley, M. D. (1998) Toward understanding tryptophan fluorescence in proteins, *Biochemistry* **37**, 9976–9982.
  56. Mely, Y., Cornille, F., Fournie-Zaluski, M. C., Darlix, J. L., Roques, B. P., and Gerard, D. (1991) Investigation of zinc-binding

- affinities of Moloney murine leukemia virus nucleocapsid protein and its related zinc finger and modified peptides, *Biopolymers* 31, 899–906.
57. Montenay-Garestier, T., Toulme, F., Fidy, J., Toulme, J.-J., and Le Doan, T. (1983) in *Structure, Dynamics and Interactions and Evolution of Biological Macromolecules* (Helene, C., Ed.) pp 113–128, D. Reidel Publishing Co., Dordrecht, The Netherlands.
58. Bombarda, E., Ababou, A., Vuilleumier, C., Gerard, D., Roques, B. P., Piemont, E., and Mely, Y. (1999) Time-resolved fluorescence investigation of the human immunodeficiency virus type 1 nucleocapsid protein: influence of the binding of nucleic acids, *Biophys. J.* 76, 1561–1570.
59. Record, M. T., Lohman, M. L., and De Haseth, P. (1976) Ion effects on ligand-nucleic acid interactions, *J. Mol. Biol.* 107, 145–158.
60. Egele, C., Schaub, E., Piemont, E., de Rocquigny, H., and Mely, Y. (2005) Investigation by fluorescence correlation spectroscopy of the chaperoning interactions of HIV-1 nucleocapsid protein with the viral DNA initiation sequences, *C. R. Biol.* 328, 1041–1051.
61. Geoghegan, K. F., Rosner, P. J., and Hoth, L. R. (2000) Dye-pair reporter systems for protein-peptide molecular interactions, *Bioconjugate Chem.* 11, 71–77.
62. Krishnamoorthy, G., Roques, B., Darlix, J. L., and Mely, Y. (2003) DNA condensation by the nucleocapsid protein of HIV-1: a mechanism ensuring DNA protection, *Nucleic Acids Res.* 31, 5425–5432.
63. Packard, B. Z., Topygin, D. D., Komoriya, A., and Brand, L. (1997) Design of profluorescent protease substrates guided by exciton theory, *Methods Enzymol.* 278, 15–23.
64. Kasha, M. (1963) Energy transfer mechanisms and the molecular exciton model for molecular composite systems, *Radiat. Res.* 20, 55–71.
65. Kasha, M. (1991) Energy transfer, charge transfer, and proton transfer in molecular composite systems, *Basic Life Sci.* 58, 231–251.
66. Scholes, D. S., and Ghiggino, K. P. (1994) Electronic interactions and interchromophore excitation transfer, *J. Phys. Chem.* 98, 4580–4590.
67. De Guzman, R. N., Wu, Z. R., Stalling, C. C., Pappalardo, L., Borer, P. N., and Summers, M. F. (1998) Structure of the HIV-1 nucleocapsid protein bound to the SL3 psi-RNA recognition element, *Science* 279, 384–388.
68. Morellet, N., Demene, H., Teilleux, V., Huynh-Dinh, T., de Rocquigny, H., Fournie-Zaluski, M. C., and Roques, B. P. (1998) Structure of the complex between the HIV-1 nucleocapsid protein NCp7 and the single-stranded pentanucleotide d(ACGCC), *J. Mol. Biol.* 283, 419–434.
69. Vuilleumier, C., Bombarda, E., Morellet, N., Gerard, D., Roques, B. P., and Mely, Y. (1999) Nucleic acid sequence discrimination by the HIV-1 nucleocapsid protein NCp7: a fluorescence study, *Biochemistry* 38, 16816–16825.
70. Amarasinghe, G. K., De Guzman, R. N., Turner, R. B., and Summers, M. F. (2000) NMR structure of stem-loop SL2 of the HIV-1 psi RNA packaging signal reveals a novel A-U-A base-triple platform, *J. Mol. Biol.* 299, 145–156.
71. Stote, R. H., Kellenberger, E., Muller, H., Bombarda, E., Roques, B. P., Kieffer, B., and Mely, Y. (2004) Structure of the His44 → Ala single point mutant of the distal finger motif of HIV-1 nucleocapsid protein: a combined NMR, molecular dynamics simulation, and fluorescence study, *Biochemistry* 43, 7687–7697.
72. Baudin, F., Marquet, R., Isel, C., Darlix, J. L., Ehresmann, B., and Ehresmann, C. (1993) Functional sites in the 5' region of human immunodeficiency virus type 1 RNA form defined structural domains, *J. Mol. Biol.* 229, 382–397.
73. Johnson, P. E., Turner, R. B., Wu, Z. R., Hairston, L., Guo, J., Levin, J. G., and Summers, M. F. (2000) A mechanism for plus-strand transfer enhancement by the HIV-1 nucleocapsid protein during reverse transcription, *Biochemistry* 39, 9084–9091.
74. SantaLucia, J., Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics, *Proc. Natl. Acad. Sci. U.S.A.* 95, 1460–1465.

BI052560M



## Supplementary material

### Homo- and heteronuclear NMR measurements

All NMR experiments were carried out on a 600 MHz Bruker spectrometer equipped with a 5mm  $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$  Z-Grad TXI probe. Spectra were processed with TOPSPIN 1.3 (Bruker) and analyzed with SPARKY 3 (University of California, San Francisco, USA). The following experiments were used for assignment and structural investigations:  $^1\text{H}$  one-dimensional spectrum with the pulse sequence zgpr,  $^1\text{H},^{15}\text{N}$  HSQC with fhsqcf3gpqh,  $^1\text{H},^{13}\text{C}$  HSQC (36, 37, 38) with hsqcetgpsisp2.2,  $^1\text{H},^1\text{H}$  TOCSY (39) with mlevsgpqh,  $^1\text{H},^1\text{H}$  DQF-COSY with cosydfesgpqh and  $^1\text{H},^1\text{H}$  ROESY (40) with roesyegpqh. All pulse sequences were used as given in the Bruker pulse sequence library. In the  $^1\text{H},^{15}\text{N}$  HSQC experiment, 96 increments were acquired with a nitrogen carrier frequency of 117 ppm and a spectral width of 34 ppm. A relaxation delay of 1.5 s was applied between the 512 scans for each  $t_1$ -increment. In the carbon  $^1\text{H},^{13}\text{C}$  HSQC experiment 256 increments were acquired with a carbon carrier frequency of 75 ppm, a spectral width of 120 ppm and a 1.3 s relaxation delay between the 64 scans. During the INEPT transfers, adiabatic shaped pulses (500  $\mu\text{s}$  CHIRP pulse with a bandwidth of 60 kHz) were used on the carbon nuclei for inversion. The TOCSY experiment with 80 ms MLEV17 mixing sequence was setup with 800 increments, a relaxation delay of 2 s between the 32 scans and excitation sculpting with gradients for water suppression. Equivalent settings were used for the ROESY experiment with a 350 ms 2.5 kHz spinlock pulse and for the DQF-COSY which was measured with 1024 increments.



## **Research Article:** Inhibition of HIV-1 by a Peptide Ligand of the Genomic RNA Packaging Signal $\Psi$

Julia Dietz, Joachim Koch, Ajit Kaur, Chinappan Raja, Stefan Stein, Manuel Grez, Anette Pustowka, Sarah Mensch, Jan Ferner, Lars Möller, Norbert Bannert, Robert Tampé, Gille Divita, Yves Mély, Harald Schwalbe, and Ursula Dietrich

*ChemMedChem* **2008**, 3, 749-755

Since the consensus sequence of the peptidic ligand to HIV  $\Psi$  RNA derived from the phage-display assay, HWWPWW,<sup>[1]</sup> possessed unfortunate spectroscopical applicability due to precipitation when in a mixture with RNA, a SPOT-assay was performed to vary the individual amino acids randomly, but retain the affinity and selectivity. Substitution of one tryptophan with a lysine resulted in the desired outcome. The most favored ligand, HKWPWW, was further investigated in its binding characteristics by NMR and fluorescence spectroscopy and in its in-vivo effects.

The group of Prof. Tampé and PD Joachim Koch mainly provided and organized the SPOT-technology with the aid from the group of Prof. Schwalbe (Ajit Kaur), while Mély and co-workers performed the fluorescence spectroscopy. The in-vivo experiments were accomplished by the groups of Dr. Dietrich and Prof. Divita. The thesis author conducted all NMR spectroscopic work and identified that both conformers of the peptide bind the RNA similarly with their central tryptophans W3 and W5 mainly incorporated. He also was responsible for the chemical synthesis of the fluorescence labeled peptides used in the in-vivo experiments, as well as for the writing of all corresponding paragraphs.



# Inhibition of HIV-1 by a Peptide Ligand of the Genomic RNA Packaging Signal $\Psi$

Julia Dietz,<sup>[a]</sup> Joachim Koch,<sup>[b]</sup> Ajit Kaur,<sup>[c]</sup> Chinnappan Raja,<sup>[d]</sup> Stefan Stein,<sup>[a]</sup> Manuel Grez,<sup>[a]</sup> Anette Pustowka,<sup>[a]</sup> Sarah Mensch,<sup>[c]</sup> Jan Ferner,<sup>[c]</sup> Lars Möller,<sup>[e]</sup> Norbert Bannert,<sup>[e]</sup> Robert Tampé,<sup>[b]</sup> Gilles Divita,<sup>[f]</sup> Yves Mély,<sup>[d]</sup> Harald Schwalbe,<sup>\*,[c]</sup> and Ursula Dietrich<sup>\*,[a]</sup>

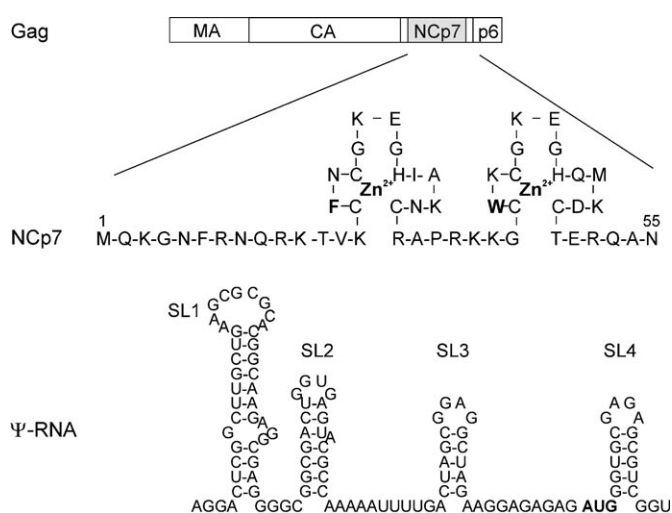
The interaction of the nucleocapsid NCp7 of the human immunodeficiency virus type 1 (HIV-1) Gag polyprotein with the RNA packaging signal  $\Psi$  ensures specific encapsidation of the dimeric full length viral genome into nascent virus particles. Being an essential step in the HIV-1 replication cycle, specific genome encapsidation represents a promising target for therapeutic intervention. We previously selected peptides binding to HIV-1  $\Psi$ -RNA or stem loops (SL) thereof by phage display. Herein, we describe synthesis of peptide variants of the consensus HWWPWW motif on

membrane supports to optimize  $\Psi$ -RNA binding. The optimized peptide, psi-pepB, was characterized in detail with respect to its conformation and binding properties for the SL3 of the  $\Psi$  packaging signal by NMR and tryptophan fluorescence quenching. Functional analysis revealed that psi-pepB caused a strong reduction of virus release by infected cells as monitored by reduced transduction efficiencies, capsid p24 antigen levels, and electron microscopy. Thus, this peptide shows antiviral activity and could serve as a lead compound to develop new drugs targeting HIV-1.

## Introduction

During HIV-1 assembly the viral genome is efficiently encapsidated over cytoplasmic mRNA because of the specific recognition of the highly structured RNA packaging signal  $\Psi$  by the two zinc fingers of the viral NCp7 domain of the Gag polyprotein precursor (Figure 1). The  $\Psi$ -region consists of four stem loops, SL1 to SL4, within the 5' UTR of the unspliced HIV-1 RNAs.<sup>[2-5]</sup> High affinity binding of aromatic and hydrophobic amino acids of the NC zinc fingers to SL2 and the major packaging signal SL3, in particular tryptophan W37 to the GGAG tetraloop, is followed by less specific interactions of basic

amino acids in NCp7 leading to the co-assembly of Gag with the genomic RNA in the form of a dimeric molecule.<sup>[2-6]</sup> As genomic RNA encapsidation is essential for HIV-1 replication, disruption of the specific interaction between  $\Psi$ -RNA and NCp7 zinc fingers should interfere with virus production. Indeed, Zn<sup>2+</sup> ejection, application of antisense RNA against  $\Psi$ -RNA elements or  $\Psi$ -RNA decoys resulted in antiviral activity in vitro.<sup>[7-13]</sup>



**Figure 1.** The nucleocapsid domain NCp7 of HIV-1 Gag interacts with the  $\Psi$ -RNA structure. Aromatic amino acids within the zinc fingers of NCp7 (bold) interact with  $\Psi$ -RNA, in particular with the GGAG tetraloop of SL3.

- [a] J. Dietz,<sup>+</sup> Dr. S. Stein, Dr. M. Grez, Dr. A. Pustowka, Dr. U. Dietrich  
Georg-Speyer-Haus, Institute for Biomedical Research  
Paul-Ehrlich-Str. 42-44, 60596 Frankfurt (Germany)  
Fax: (+49) 69 6339 5297  
E-mail: ursula.dietrich@em.uni-frankfurt.de
- [b] Dr. J. Koch,<sup>+</sup> Prof. Dr. R. Tampé  
Institute of Biochemistry  
Johann Wolfgang Goethe-University Frankfurt (Germany)
- [c] A. Kaur, S. Mensch, J. Ferner, Prof. Dr. H. Schwalbe  
Institute for Organic Chemistry and Chemical Biology  
Center for Biomolecular Magnetic Resonance  
Johann Wolfgang Goethe-University Frankfurt  
Max-von-Laue-Str. 9, 60438 Frankfurt (Germany)  
Fax: (+49) 69 7982 9515  
E-mail: schwalbe@nmr.uni-frankfurt.de

- [d] Dr. C. Raja, Prof. Dr. Y. Mély  
Département Pharmacologie et Physicochimie  
Faculté de Pharmacie, Institut Gilbert-Laustriat  
UMR 7175 CNRS/Université Louis Pasteur
- [e] L. Möller, Dr. N. Bannert  
Center for Biological Safety 4, Robert Koch Institute
- [f] Dr. G. Divita  
Centre de Recherches de Biochimie Macromoléculaires  
CRBM-CNRS

[<sup>+</sup>] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://www.chemmedchem.org> or from the author.

We and others selected peptide ligands for  $\Psi$ -RNA from phage displayed peptide libraries to interfere with NCp7 binding.<sup>[1,14]</sup> Interestingly, the selected peptides were rich in tryptophans and these amino acids were important for binding to  $\Psi$ -RNA, such as W37 in the natural ligand NCp7 (Figure 1). Although the selected peptides interacted with  $\Psi$ -RNA, their binding affinity was in the high micromolar range and thus required optimization in view of therapeutic applications.<sup>[1]</sup> Herein, we describe the selection and functional analysis of an optimized peptide variant, *psi-pepB*, that shows antiviral activity in cell culture.

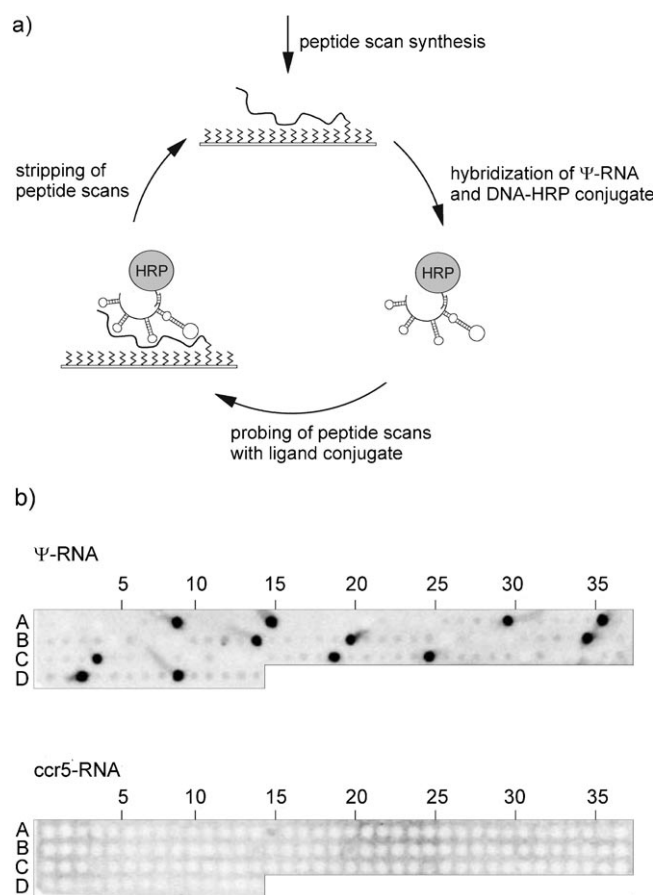
## Results and Discussion

### Peptide optimization on solid-support peptide arrays

To optimize the peptide ligands based on the previously identified consensus peptide HWWPWW (*psi-pepA*), we used peptide arrays on membrane supports for screening.<sup>[1,15,16]</sup> This method has been widely applied for epitope mapping and the identification of protein–protein interaction sites and to investigate peptide–DNA interactions.<sup>[17]</sup> To our knowledge, this is the first study adapting this method for peptide–RNA interactions. The experimental setup is illustrated in Figure 2a. Peptide arrays of permutations of the peptide GGSGSHWWPWWGGSG (parental sequence underlined) with one position each exchanged against all other 19 natural amino acids were probed with either the complete  $\Psi$ -RNA or a control RNA of the same length. Strikingly, only peptides containing either an arginine or lysine at any of the positions within the peptide HWWPWW specifically interacted with  $\Psi$ -RNA whereas interaction with *ccr5* control RNA was below the detection limit (Figure 2b and Table 1). The peptide HKWPWW (*psi-pepB*) was selected for further analysis based on its solubility and binding characteristics for  $\Psi$ -RNA.

### Binding characteristics of *psi-pepB* to SL3-RNA

The binding of *psi-pepB* to SL3 was characterized by tryptophan fluorescence quenching (Figure 3a). At saturating concentrations, the interaction between SL3 and *psi-pepB* resulted in 60% reduction of the tryptophan fluorescence. This quenching was significantly higher than that previously observed with PBS and TAR sequences.<sup>[18]</sup> As fluorescence quenching results mainly from the  $\pi$ - $\pi$  stacking of the tryptophan residues with the oligonucleotide bases, this suggests that a large fraction of the tryptophan residues of *psi-pepB* was stacked with the bases of SL3. Extensive stacking would favor inhibition of NCp7 binding to SL3, as the stacking of W37 of NCp7 with guanines is a major driving force in the binding of NCp7 to its RNA substrates.<sup>[5,19,20]</sup> The number of *psi-pepB* binding sites on SL3 was found to be  $3 \pm 1$  under the conditions used, giving an occluded binding site of about six nucleotides per peptide, close to that found for TAR.<sup>[18]</sup> Assuming that the binding sites are identical and noninteracting, we obtained an apparent dissociation constant for *psi-pepB* of  $1.1 \pm 0.4 \mu\text{M}$  derived from three experiments. This affinity is about one order of magni-



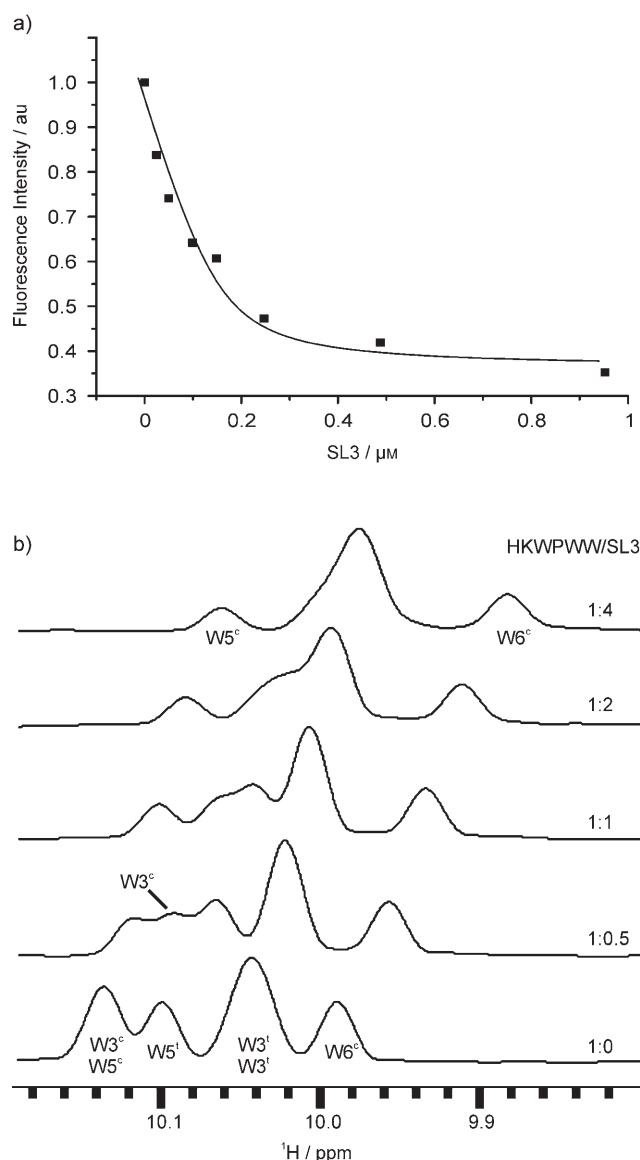
**Figure 2.** HIV-1  $\Psi$ -RNA binding to peptide arrays. a) Scheme of the peptide optimization on membrane supports. b) Arrays of permutations of the peptide GGSGSHWWPWWGGSG (target residues underlined) were probed with  $\Psi$ -RNA and *ccr5*-RNA. Bound RNA was visualized by chemiluminescence imaging. The parental peptide GGSGSHWWPWWGGSG was included in each set of permutations (line 1: spot 7; line 2: spots 3, 24; line 3: spots 2, 29; line 4: spot 13).<sup>[1]</sup>

**Table 1.** Novel peptide ligands of HIV-1  $\Psi$ -RNA.<sup>[a]</sup>

Line	Spot	Sequence
1	9	GGSGS <b>K</b> WWPWWGGSG
	15	GGSGS <b>R</b> WWPWWGGSG
	30	GGSGS <b>H</b> KWPWWGGSG
	36	GGSGS <b>H</b> RWPWWGGSG
2	14	GGSGS <b>H</b> WKPWWGGSG
	20	GGSGS <b>H</b> WRPWWGGSG
	35	GGSGS <b>H</b> WWKWWGGSG
3	4	GGSGS <b>H</b> WWRWWGGSG
	19	GGSGS <b>H</b> WWPKWGGSG
4	25	GGSGS <b>H</b> WWPRWGGSG
	3	GGSGS <b>H</b> WWPWKGGSG
	9	GGSGS <b>H</b> WWPW <b>R</b> GGSG

[a] Peptide sequences specifically recognized by  $\Psi$ -RNA are listed. The positions of the peptide spots depicted from Figure 2b are annotated, amino acids exchanged are given in bold. Sequences derived from the consensus peptide HWWPWW identified previously are underlined.<sup>[1]</sup>

tude higher than that of *psi-pepA*, suggesting that the substitution of W2 by a lysine residue significantly stabilizes the complex.<sup>[1]</sup>



**Figure 3.** Binding of *psi-pepB* to SL3-RNA. a) Binding curve of HKWPWW-NH<sub>2</sub> with SL3-RNA. Excitation and emission wavelengths were 295 and 350 nm, respectively. The solid line represents the mathematical fit of the experimental points as described previously.<sup>[21]</sup> b) Titration of SL3-RNA to *psi-pepB*. One-dimensional <sup>1</sup>H NMR spectra of the peptide focusing on the tryptophan indole imino signals. The spectra correspond to the free *psi-pepB* and to *psi-pepB* after addition of SL3-RNA up to a fourfold excess. Each tryptophan residue results in two resonances, one resulting from the *cis* (c) and the other for the *trans* (t) conformation of proline.

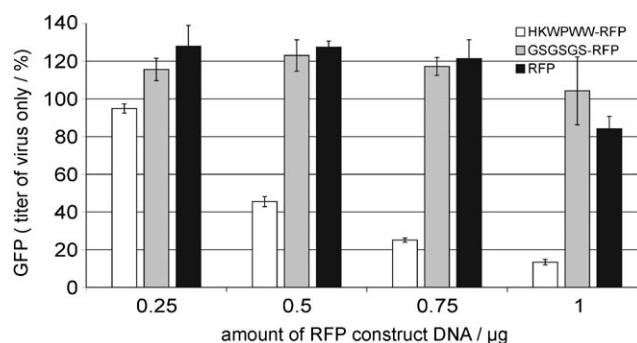
As previously shown by NMR spectroscopy, *psi-pepB* adopted two global conformations with nearly equal populations depending on the *cis/trans* conformation of the peptide bond of the central proline.<sup>[18]</sup> By titrating SL3 up to a fourfold excess to a *psi-pepB* peptide solution followed by <sup>1</sup>H NMR analysis, distinctive shifts of the well-resolved tryptophan side chain imino proton resonances were observed for both conformations (Figure 3b). All resonances shifted upfield and increased in line width (~30%) during the titration indicating an interaction of both conformations with the SL3 in the fast to intermediate NMR timescale. The largest shifts of the indole imino

protons were observed for W5 in the *trans*-proline conformation and W3 in the *cis*-proline conformation. The conformational *cis/trans* equilibrium remained constant upon interaction with RNA. During the reverse titration of *psi-pepB* to SL3, no shifts of the imino resonances were detectable in the oligonucleotide base paired stem (data not shown). This assigns the specific binding region for the *psi-pepB* peptide to the GGAG tetraloop where no imino resonances are detectable. As *psi-pepB* also bound to the tetraloop of SL3 similar to the natural NCp7 ligand and due to its improved affinity, we expected that the peptide could inhibit HIV-1 replication.<sup>[5]</sup>

#### Antiviral activity of *psi-pepB*

We previously showed by ELISA that binding of *psi-pepA* to Ψ-RNA can be competed by Gagp55 and NCp7 underlining the specificity of binding for this RNA.<sup>[11]</sup> Herein, we assessed the capacity of the optimized peptide *psi-pepB* to inhibit HIV-1. The antiviral effect of *psi-pepB* was first analyzed by determining the transduction efficiencies of lentiviral vectors (LV) encoding the green fluorescence marker gene (*gfp*) in the presence of *psi-pepB* or control peptides by FACS analysis.<sup>[21]</sup> Expression of *psi-pepB* in cells was achieved upon transfection of a plasmid encoding the peptide sequence as a fusion with the marker gene for the red fluorescence protein (RFP). Transduction efficiencies of viral supernatants were then determined for different amounts of transfected RFP constructs in comparison with viruses generated in the presence of a control peptide (GSGSGS-RFP) or RFP alone. A representative experiment is shown in Figure 4. A clear reduction in LV transduction efficiencies was observed with increasing concentrations of transfected *psi-pepB* DNA. Whereas no reduction in LV titers was observed for the control peptide or RFP alone, *psi-pepB* expression reduced titers up to 85% in cells transfected with 1 μg of the corresponding DNA construct. However, the total amount of Gag proteins in transfected cells was comparable in the cell lysates as analyzed by Western blot analysis (data not shown).

Reduced LV titers were not due to toxicity of the peptides in the producer cells as determined by counting viable cells



**Figure 4.** Antiviral activity of the optimized *psi-pepB* peptide on replication incompetent lentiviral particles. Transduction efficiencies of lentiviral particles produced in the presence of *psi-pepB* (white), a control peptide (gray), or RFP alone (black) were evaluated by titrating lentiviral supernatants on 293T cells based on determining the percentage of GFP-positive cells by FACS analysis.

based on digital resistance determinations (data not shown). Furthermore, *psi-pepB* specifically inhibited the production of LV particles based on HIV-1  $\Psi$  and Gag, whereas no specific inhibition was observed for  $\gamma$ -retroviral vectors containing the  $\Psi$ -signal and Gag of murine leukaemia virus (data not shown).

We next infected PM1 cells stably expressing the peptide-RFP-construct with HIV-1<sub>NL4-3</sub> to confirm the antiviral activity of the *psi-pepB* peptide on replication-competent HIV-1. The amount of CAp24 antigen in culture supernatants was reduced by 70% at day five in the presence of *psi-pepB* (Supporting Information, SI Figure 1). Further, we analyzed the virus particles generated in the presence of *psi-pepB* or the control peptide by electron microscopy. Clearly, less budding viruses were observed in cells transduced with the *psi-pepB* peptide as compared to the control peptide and these viruses had a less condensed core (Figure 5). A detailed evaluation indeed revealed reduced virus production in *psi-pepB* expressing cells as compared to cells expressing the control peptide. This was observed at several levels (Supporting Information, SI Table 1): the mean number of virus producing cells (4.6 versus 3.5), the number of free virions per cell (21.0 versus 45.6), and the number of buddings (1.5 versus 2.1). The phenotype observed here is reminiscent of viruses with mutated Zn fingers in NCp7 that led to altered localization of Gag in the cells as well as impaired assembly and budding.<sup>[22]</sup>

To further prove that the antiviral activity is due to the *psi-pepB* peptide, we synthesized a rhodamine labeled *psi-pepB* peptide in conjunction with a protein transduction domain.<sup>[23]</sup> After delivery into cells infected with replication competent HIV-1<sub>LAI</sub>, the antiviral effect was analyzed by titrating HIV-1 supernatants generated in the presence of *psi-pepB* or a control peptide on TZM-bl cells. *Psi-pepB* inhibited HIV-1<sub>LAI</sub> by more than 90% after transduction of  $2 \times 40 \mu\text{M}$  of peptide, whereas the control peptide had only minimal effects (Figure 6). In these experiments, the first peptide addition was performed about 24 h after infection to circumvent a potential inhibitory effect of the peptides on reverse transcription, a process, which also depends on the chaperone activity of NCp7 during strand transfer reactions.<sup>[3,24]</sup> No toxic effects of the peptides were observed on P4.R5 or TZM-bl cells at any concentration tested (data not shown).

Interestingly, *psi-pepB* (HKWPWW) is highly homologous to short cationic antimicrobial peptides that disrupt large unilamellar vesicles.<sup>[25,26]</sup> As indolicidin (ILPWKWPWWPWR) was previously shown to be virucidal against HIV-1 virions at a concentration of  $174 \mu\text{M}$  after 60 min, we analyzed *psi-pepB* for potential additional virucidal activity after incubation with lentiviral particles.<sup>[25]</sup> At  $35 \mu\text{M}$ , a concentration close to that showing 90% antiviral activity in our assays ( $40 \mu\text{M}$ , Figure 6), no virucidal activity of *psi-pepB* was found relative to the control peptide (Supporting Information, SI Figure 2). Thus, the antiviral activity observed in our assays is not due to virucidal activity. Virucidal activity on virions was only observed at much higher *psi-pepB* concentrations, being 35% at  $174 \mu\text{M}$  relative to the control peptide. Therefore, virucidal activity is very unlikely to contribute significantly to the antiviral activity ob-

served in our assays, in particular, as the cells were already infected with HIV-1 when the peptides were added.

Taken together, *psi-pepB* represents a peptide with antiviral activity against HIV-1, causing a strong reduction in particle production. From this study and our previous analyses it seems very likely that the tryptophan-rich peptide *psi-pepB* may mimic the hydrophobic platform of NCp7 that is essential for several functions.<sup>[18]</sup> These include the two strand transfer reactions during reverse transcription by binding to PBS and TAR sequences and RNA packaging by interaction with the  $\Psi$ -RNA. In line with this, *psi-pepB*, although originally selected with  $\Psi$ -RNA structures, also binds to sequences derived from PBS and TAR.<sup>[18]</sup> Thus, the antiviral activity observed here may result from the combined inhibitory effects at multiple steps during HIV-1 replication, where the action of NCp7 is needed. Based on the intriguing findings described here, the molecular details underlying the inhibitory action of the *psi-pepB* peptide at possibly more than one target have to be elucidated in further studies.

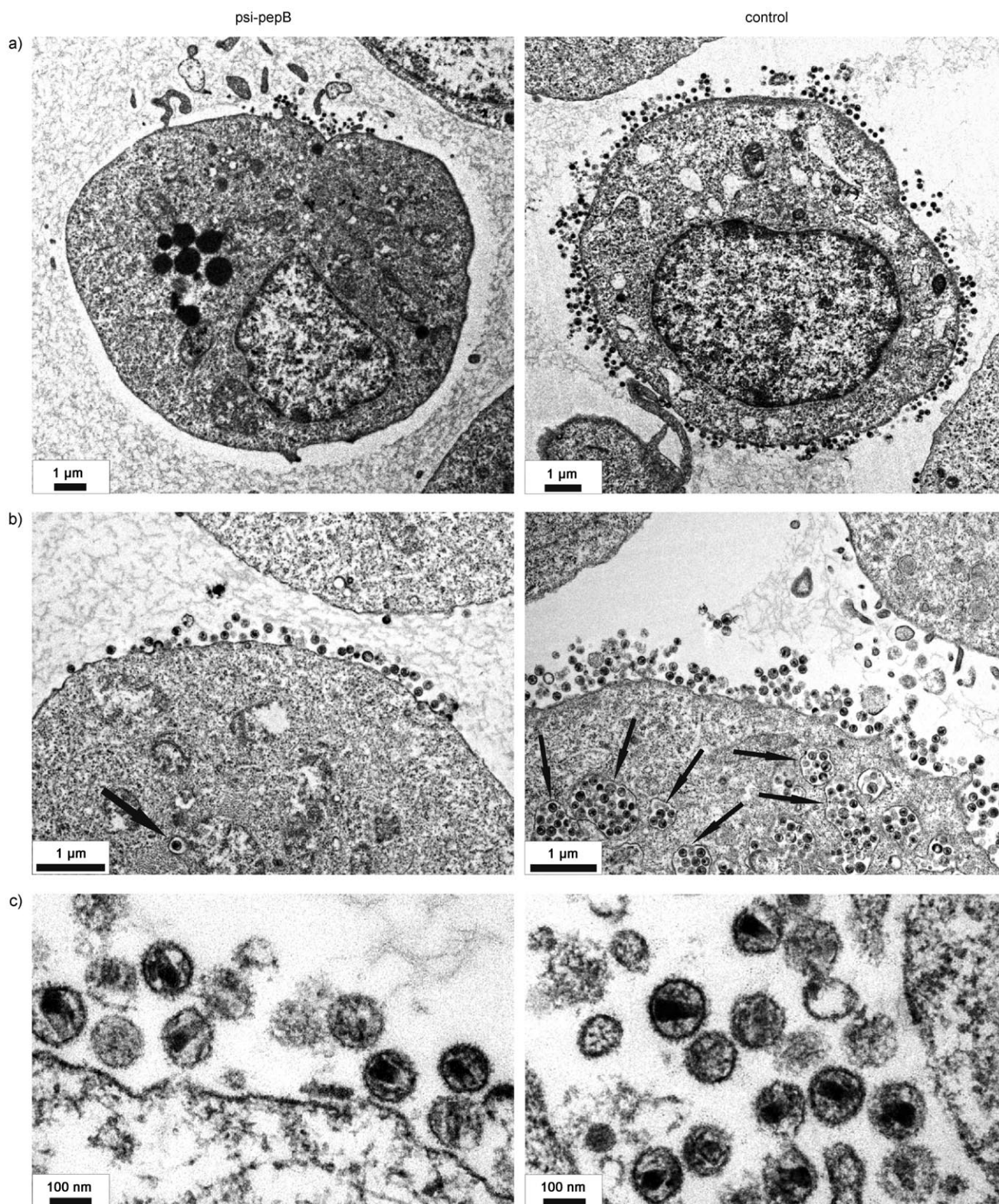
## Experimental Section

**Spot synthesis peptide membranes.** Libraries of cellulose-bound peptides were synthesized by Fmoc-chemistry semiautomatically on a spot robot (ASP222, Intavis, Germany) as described previously.<sup>[15]</sup> Fmoc protected amino acids were obtained from Bachem, Switzerland. Peptide libraries were synthesized semiautomatically using Fmoc chemistry as separate spots on cellulose membranes carrying activated polyethylene glycol spacers.<sup>[15,16]</sup> To reduce steric hindrance for later RNA binding experiments, permutations of the peptide HWWPWW were synthesized in between an N-terminal GGSGS- and a C-terminal GGSG-linker. In vitro transcribed  $\Psi$ -RNA of HIV-1 was hybridized to an oligonucleotide complementary to the 5' end (5'-CGAGAAUUACCCUCACUAAAGG-3') coupled with horseradish peroxidase (HRP). The  $\Psi$ -RNA or the control ccr5-RNA (8.5 pmol) was hybridized with the HRP/DNA conjugate (20 pmol) in  $10 \mu\text{L}$  at  $45^\circ\text{C}$  for 1 h. Peptide scans were preincubated with  $\Psi$ -buffer (0.1 mM KCl, 5 mM HEPES, pH 7.4) and then saturated for 1 h with 2% blocking reagent (Roche, Germany) in blocking buffer (100 mM maleic acid, 150 mM NaCl, pH 7.5). Membranes were probed with the  $\Psi$ -RNA/DNA/HRP-hybrid for 2 h in  $\Psi$ -buffer (10 mL). After washing, bound RNA was visualized by chemiluminescence imaging. Membranes were reused after stripping using a three-step procedure: 1) incubation for 30 min with urea (8 M), SDS (20%),  $\beta$ -mercaptoethanol (0.1%), 2) incubation for 30 min with EDTA (10 mM), RNase A (0.1 mg mL<sup>-1</sup>), Tris/HCl (50 mM, pH 8), and 3) incubation for 5 min with trifluoroacetic acid.

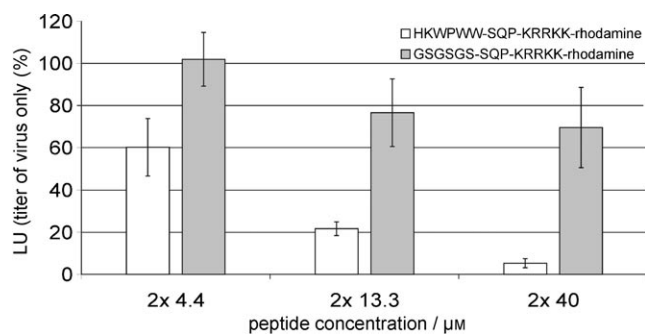
**Fluorescence titrations.** Fluorescence titrations were performed on a thermostated Fluorolog spectrofluorometer (Jobin Yvon) by adding increasing SL3-RNA concentrations to *psi-pepB* ( $0.2 \mu\text{M}$ ) in potassium phosphate buffer (100 mM, pH 6.5). The fluorescence intensity  $I$  at 350 nm was calculated as described previously and corrected for dilution, buffer fluorescence, and screening effects due to the oligonucleotide absorbance.<sup>[18]</sup>

**<sup>1</sup>H NMR Measurements.** The titration of SL3-RNA to *psi-pepB* was monitored up to fourfold excess of RNA by 1D <sup>1</sup>H NMR on a Bruker 700 MHz spectrometer equipped with a <sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N triple resonance cryoprobe at 298 K. A 2D homonuclear NOESY spectrum was recorded at the final state to confirm the assignment. The peptide concentration was 0.2 mM in a buffer of potassium phosphate (25 mM) and potassium chloride (50 mM) at pH 6.2. All ex-





**Figure 5.** Electron microscopy of viral particles produced in cells expressing *psi-pepB* or control peptides. Effect of the *psi-pepB* peptide on synthesis and morphology of HIV-1 particles. Representative electron micrographs of infected PM1 cells stably expressing the HKWPWW peptide (left panel) or a control peptide (right panel). a) The expression of the peptide targeting the packaging signal leads to a significantly lower particle production and b) to a strongly reduced budding activity into the endosomal compartment (indicated by arrows). c) Mature HIV-1 cores of particles produced in the presence of the *psi-pepB* peptide appear more electron lucent and slightly less condensed.



**Figure 6.** Antiviral activity of the optimized *psi-pepB* peptide linked to a protein transduction domain on replication competent HIV-1. Infection efficiencies of HIV-1<sub>LAI</sub> virions generated in the presence of *psi-pepB* (white) or control peptide (gray) were analyzed by titrating HIV-1<sub>LAI</sub> supernatants on TZM-bl cells. The infected cells were grown for 44 h and the extent of infection was determined by measuring luciferase activity in the cell lysates.

periments were recorded with a jump-return-echo sequence, where the excitation maximum was centered on the chemical shift range of the tryptophan indole signals.

**Cell lines.** 293T, P4.R5 MAGI, and TZM-bl cells were maintained in DMEM supplemented with FCS (10%), penicillin ( $100 \text{ U mL}^{-1}$ ),  $100 \mu\text{g mL}^{-1}$  streptomycin ( $100 \mu\text{g mL}^{-1}$ ), and L-glutamine (2 mM) at  $37^\circ\text{C}$  in a 5%  $\text{CO}_2$  atmosphere.

**Vector construction.** For expression of peptides in mammalian cells, the plasmid pF25-GFP-Vpr kindly provided by Roland Stauber was modified.<sup>[27]</sup> The *gfp* gene was substituted by the *rfp* gene (Ds Red1) after PCR amplification from pHR'SINcPPT-SRW and cloning into the NheI/NarI sites. The *vpr* gene was deleted by restriction using NarI and XbaI and fill in with the Klenow fragment. A linker encoding the *psi-pepB* peptide or a GSGSGS control peptide was cloned upstream of *rfp* into the SacII/NheI sites to generate peptide-RFP fusion proteins.

**Production of lentiviral particles.** Replication defective lentiviral particles were generated by transient co-transfection of 293T cells with the plasmids pHR'SINcPPT-SEW, pCMV $\Delta$ R8.91, and pMD2.VSVG by the calcium phosphate method.<sup>[21,28,29]</sup> For inhibition experiments, lentiviral particles were generated in the presence of the respective cotransfected pF25-peptide-RFP constructs. 293T cells were grown to 80% confluence and cotransfected with the three plasmids for lentivirus production ( $0.5 \mu\text{g}$  pHRSEWcPPT,  $0.4 \mu\text{g}$  pCMV $\Delta$ R8.91, and  $0.9 \mu\text{g}$  pMD2.VSVG) and varying amounts of the pF25-peptide-RFP constructs. After 12 h, the supernatants were discarded and the cells were fed with fresh medium (2 mL). Viral stocks were harvested 48 h post transfection and filtered through  $0.22 \mu\text{m}$  syringe filters before storage at  $-70^\circ\text{C}$ .

**Transduction and determination of lentiviral titers.** 293T cells were seeded at  $1 \times 10^5$  cells and transduced with serially diluted viral vector stocks the next day in the presence of polybrene ( $8 \mu\text{g mL}^{-1}$ ). Three days after transduction, cells were harvested, re-suspended in PBS, and EGFP positive fluorescent cells were quantified by FACS analysis. The viral vector titer was determined as the percentage of EGFP positive cells multiplied by a factor to account for the dilution of the viral stock and total cell number.

**Electron microscopy.** PM1 cells stably expressing *psi-pepB* peptide or a control peptide in fusion with RFP were infected with HIV-1<sub>NL4-3</sub> (MOI of 0.006), fixed at day five post infection with glutaraldehyde (2.5%) in HEPES (0.05 M), pH 7.2, and embedded in low melting point agarose (3%). The cells were postfixed with 1% osmium tetroxide, treated with 0.1% tannic acid (0.1%), and contrasted with uranyl acetate (2%). After dehydration the pellets were finally

embedded in epoxy resin (Epon). Ultrathin sections were post-stained with 2% uranyl acetate (2%) and lead citrate (0.1%) and were examined with a Tecnai Spirit transmission electron microscope (FEI Co., USA) at 120 keV. Images were taken with a Mega-view III (Olympus Soft Imaging Solutions) camera. For each single infection, 25 complete cross sections of HIV-1 producing cells were analyzed.

**Peptide synthesis.** For synthesis of the rhodamine-labeled peptides Ac-HKWPWWWSQP $\Delta$ KRRKK and Ac-GSGSGSSQP $\Delta$ KRRKK, an *N*- $\alpha$ -Fmoc-Lys(Mtt)-Wang resin was used. After removal of the lysine side-chain protecting group (4-Methyltrityl) with TFA (1%), TIS (5%), DCM (94%), and attachment of the dye 5(6)-Carboxytetramethyl-rhodamine (Novabiochem, Switzerland) to the free lysine side-chain amino group, this resin was used in automatic peptide synthesis on an ABI433A peptide synthesizer as previously described.<sup>[1]</sup>

**Anti-HIV assays with the transduced peptides.** P4.R5 MAGI cells were seeded at  $1 \times 10^4$  cells and infected with HIV-1<sub>LAI</sub> (MOI of 0.5). The next day (24 h after infection), the virus was washed off and rhodamine-labeled peptides were added to the cells in fresh medium for 24 h. Thereafter, the same amount of peptides was added again for a further 24 h until the supernatants were harvested. The viral titer of the preparations was determined on TZM-bl cells expressing firefly luciferase under the control of the HIV-1 LTR.  $1 \times 10^4$  TZM-bl cells were seeded and infected with the viral supernatants in the presence of polybrene ( $8 \mu\text{g mL}^{-1}$ ) for 44 h. The viral titer was quantified by luciferase assay. Cells were washed with PBS, lysed with harvest buffer (0.5 M Mes-Tris, 1 M DTT, 10% Triton-X-100 and glycerol), and light emission was measured on a Lumistar Galaxy Luminometer (BMG Labbiotechnologies, Germany).

**Toxicity and virucidal activity assays.** The viability of 293T cells transfected with the different peptide encoding plasmids was assessed by measuring resistance based on digital pulse processing. Briefly, transfected cells were harvested 48 h post transfection, re-suspended in PBS, and viable cells were counted in an automated cell counting system (CasyTT, Schärfe Systems). Viability of peptide treated P4.R5 MAGI or TZM-bl cells was determined by the ViaLight kit (Cambrex) based on the luminometric measurement of ATP.

To assess the virolytic activity of the peptides on lentiviral particles, culture supernatants ( $10 \mu\text{L}$ ) of producer cells (see production of lentiviral particles) were incubated with an equal volume of peptide solutions (0–174  $\mu\text{M}$ ) for 60 min at  $37^\circ\text{C}$  in duplicates. After diluting 1:200 in growth medium, transduction efficiencies were determined on 293T cells as described above.

## Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (SFB 579), the EU Integrated Project TRIoH (LSHB-CT-2003-503480) and the state of Hessen (Center of Biomolecular Magnetic Resonance). We thank Michaela Stoll and Hana Kunkel for expert technical assistance. We particularly thank Jean-Luc Darlix (Lyon) for helpful discussions. The following reagents were obtained through the NIH AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: TZM-bl, P4.R5 MAGI, HIV-1<sub>LAI</sub>, HIV-1<sub>NL4-3</sub>. The authors declare that they have no competing financial interests.

**Keywords:** antiviral peptides • HIV-1 • NMR spectroscopy • RNA packaging • spot synthesis

- [1] A. Pustowka, J. Dietz, J. Ferner, M. Baumann, M. Landersz, C. Konigs, H. Schwalbe, U. Dietrich, *ChemBioChem* **2003**, *4*, 1093–1097.
- [2] J. L. Darlix, J. L. Garrido, N. Morellet, Y. Mely, H. De Rocquigny, *Adv. Pharmacol.* **2007**, *55*, 299–346.
- [3] J. L. Darlix, M. Lapadat-Tapolsky, H. de Rocquigny, B. P. Roques, *J. Mol. Biol.* **1995**, *254*, 523–537.
- [4] G. K. Amarasinghe, R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu, M. F. Summers, *J. Mol. Biol.* **2000**, *301*, 491–511.
- [5] R. N. De Guzman, Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, M. F. Summers, *Science* **1998**, *279*, 384–388.
- [6] S. Campbell, V. M. Vogt, *J. Virol.* **1995**, *69*, 6487–6497.
- [7] W. G. Rice, J. G. Supko, L. Malspeis, R. W. Buckheit, Jr., D. Clanton, M. Bu, L. Graham, C. A. Schaeffer, J. A. Turpin, J. Domagala, R. Gogliotti, J. P. Bader, S. M. Halliday, L. Coren, R. C. Sowder II, L. O. Arthur, L. E. Henderson, *Science* **1995**, *270*, 1194–1197.
- [8] D. R. Chadwick, A. M. Lever, *Gene Ther.* **2000**, *7*, 1362–1368.
- [9] H. De Rocquigny, V. Shvadchak, S. Avilov, C. Z. Dong, U. Dietrich, J. L. Darlix, Y. Mely, *Minirev. Med. Chem.* **2008**, *8*, 24–35.
- [10] N. M. Dorman, A. M. Lever, *Gene Ther.* **2001**, *8*, 157–165.
- [11] L. M. Miller Jenkins, J. C. Byrd, T. Hara, P. Srivastava, S. J. Mazur, S. J. Stahl, J. K. Inman, E. Appella, J. G. Omichinski, P. Legault, *J. Med. Chem.* **2005**, *48*, 2847–2858.
- [12] J. C. Paillart, M. Shehu-Xhilaga, R. Marquet, J. Mak, *Nat. Rev. Microbiol.* **2004**, *2*, 461–472.
- [13] E. Skripkin, J. C. Paillart, R. Marquet, M. Blumenfeld, B. Ehresmann, C. Ehresmann, *J. Biol. Chem.* **1996**, *271*, 28812–28817.
- [14] M. Y. Park, J. Kwon, S. Lee, J. You, H. Myung, *Virus Res.* **2004**, *106*, 77–81.
- [15] R. Frank, *Tetrahedron* **1992**, *48*, 9217–9232.
- [16] J. Koch, M. Mahler, *Peptide Arrays on Membrane Supports*, Springer, Heidelberg, **2002**.
- [17] M. Reuter, J. Schneider-Mergener, D. Kupper, A. Meisel, P. Mackeldanz, D. H. Kruger, C. Schroeder, *J. Biol. Chem.* **1999**, *274*, 5213–5221.
- [18] C. Raja, J. Ferner, U. Dietrich, S. Avilov, D. Ficheux, J. L. Darlix, H. de Rocquigny, H. Schwalbe, Y. Mely, *Biochemistry* **2006**, *45*, 9254–9265.
- [19] G. K. Amarasinghe, R. N. De Guzman, R. B. Turner, M. F. Summers, *J. Mol. Biol.* **2000**, *299*, 145–156.
- [20] E. Bombarda, A. Ababou, C. Vuilleumier, D. Gerard, B. P. Roques, E. Piemont, Y. Mely, *Biophys. J.* **1999**, *76*, 1561–1570.
- [21] L. Naldini, U. Blomer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, D. Trono, *Science* **1996**, *272*, 263–267.
- [22] B. Grigorov, D. Decimo, F. Smagulova, C. Pechoux, M. Mougel, D. Muriaux, J. L. Darlix, *Retrovirology* **2007**, *4*, 54.
- [23] M. C. Morris, J. Depollier, J. Mery, F. Heitz, G. Divita, *Nat. Biotechnol.* **2001**, *19*, 1173–1176.
- [24] C. Bampi, S. Jacquenet, D. Lener, D. Decimo, J. L. Darlix, *Int. J. Biochem. Cell Biol.* **2004**, *36*, 1668–1686.
- [25] W. E. Robinson, Jr., B. McDougall, D. Tran, M. E. Selsted, *J. Leukocyte Biol.* **1998**, *63*, 94–100.
- [26] D. J. Schibli, R. F. Eband, H. J. Vogel, R. M. Eband, *Biochem. Cell Biol.* **2002**, *80*, 667–677.
- [27] R. H. Stauber, S. Rulong, G. Palm, N. I. Tarasova, *Biochem. Biophys. Res. Commun.* **1999**, *258*, 695–702.
- [28] C. Demaison, K. Parsley, G. Brouns, M. Scherr, K. Battmer, C. Kinnon, M. Grez, A. J. Thrasher, *Hum. Gene Ther.* **2002**, *13*, 803–813.
- [29] R. Zufferey, D. Nagy, R. J. Mandel, L. Naldini, D. Trono, *Nat. Biotechnol.* **1997**, *15*, 871–875.

---

Received: August 2, 2007

Revised: December 14, 2007

Published online on January 18, 2008

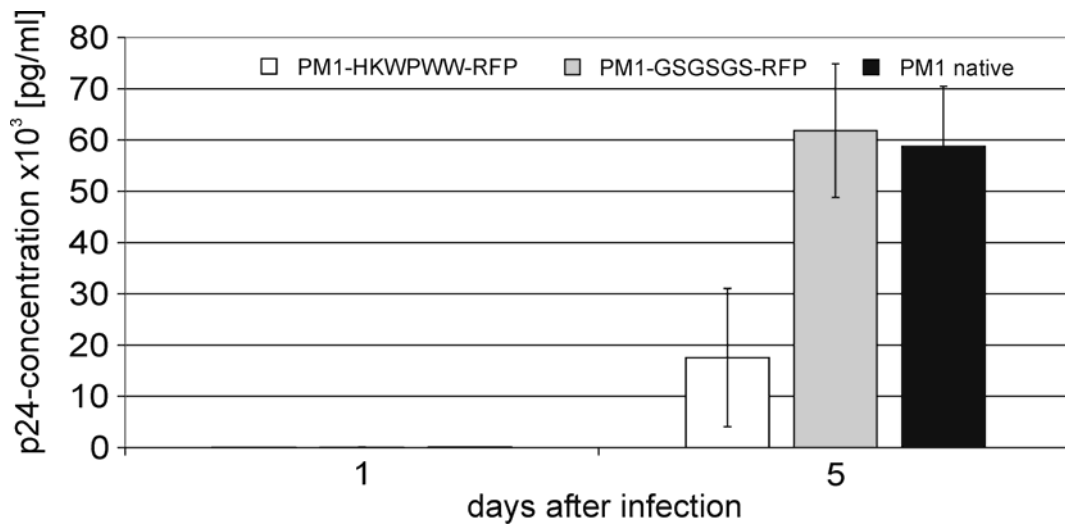
Supporting Information to Manuscript:

# Inhibition of HIV-1 by a Peptide Ligand of the Genomic RNA Packaging Signal $\Psi$

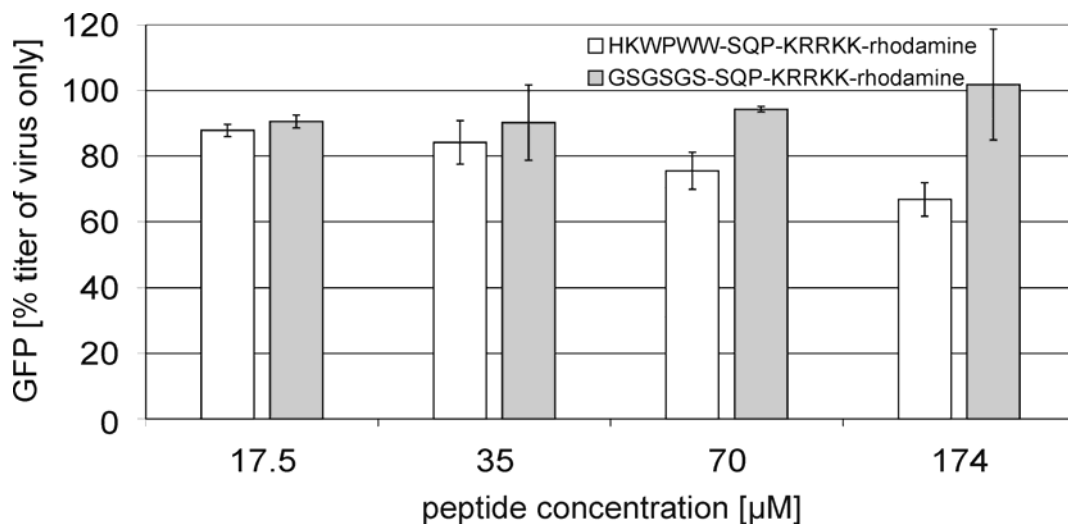
*Julia Dietz, Joachim Koch, Ajit Kaur, Chinappan Raja, Stefan Stein, Manuel Grez, Anette Pustowka, Sarah Mensch, Jan Ferner, Lars Möller, Norbert Bannert, Robert Tampé, Gille Divita, Yves Mély, Harald Schwalbe, and Ursula Dietrich*

*ChemMedChem* **2008**, 3, 749-755

## Supporting Figures:



Supporting Fig. 1: Inhibition of HIV-1 propagation after stable expression of *psi-pepB* in PM1 cells. HIV-replication in PM1 cells was measured by determining the p24-antigen release in culture supernatants from cells stably expressing *psi-pepB* (white), a control peptide (grey) or no peptide (native PM1 cells) (black) five days after infection of the cell lines with HIV-1<sub>NL4-3</sub>.



Supporting Figure 2: Virucidal activity of *psi-pepB*. Transduction efficiencies of lentiviral particles preincubated for 1 h at 37°C with *psi-pepB* (white) or control peptide (grey), respectively, were analyzed by titrating on 293T cells based on determining the percentage of GFP-positive cells by FACS analysis.

## Supporting Tables:

Supporting Table 1: Quantitative evaluation of virions produced in the presence of the peptides by electron microscopy

a) Virions from PM1 cells stably expressing *psi-pepB*

b) Virions from PM1 cells stably expressing the control peptide

(\*mean numbers; x: not evaluated)

a)

Virions outside cell	Endosomal vesicles	Buddings	nuclear pores	infected cells/mesh	cells/mesh	mesh no.	countings	grid
17	0	0	0	1	15	1	1	1
6	0	2	0	5	20	2	1	1
5	0	0	0	x	x	2	1	1
6	3	2	0	x	x	2	1	1
8	2	0	3	6	20	3	1	1
41	1	5	0	x	x	3	1	1
x	x	x	x	4	16	4	x	1
27	25	1	17	4	18	5	1	1
25	2	3	14	x	x	5	1	1
x	x	x	x	4	18	6	x	1
x	x	x	x	2	15	7	x	1
75	0	3	no nucleus	5	20	8	1	1
2	7	3	0	x	x	8	1	1
10	0	2	6	x	x	8	1	1
33	2	2	9	3	15	9	1	1
x	x	x	x	2	15	10	x	1
75	17	8	no nucleus	6	18	11	1	1
17	2	1	7	x	x	11	1	1
45	3	1	0	3	15	12	1	1
8	0	1	0	2	16	13	1	2
8	0	0	0	7	20	14	1	2
14	0	0	0	x	x	14	1	2
12	0	1	0	3	20	15	1	2
10	17	1	4	4	20	16	1	2
19	18	0	0	x	x	16	1	2
30	10	2	0	3	20	17	1	2
12	7	0	0	5	16	18	1	2
7	10	0	0	3	18	19	1	2
26	20	0	9	5	20	20	1	2
8	10	0	2	x	x	20	1	2
<b>21,0*</b>	<b>6,0*</b>	<b>1,5*</b>	<b>3,0*</b>	<b>3,9*</b>	<b>17,8*</b>	<b>20</b>	<b>26</b>	<b>2,0</b>

b)

Virions outside cell	Endosomal vesicles	Buddings	nuclear pores	infected cells/mesh	cells/mesh	mesh no.	countings	grid
16	20	3	5	6	17	1	1	1
x	x	x	x	2	12	2	x	1
8	10	0	4	4	20	3	1	1
40	4	4	12	4	20	4	1	1
x	x	x	x	5	16	5	x	1
45	10	3	9	6	20	6	1	1
x	x	x	x	5	20	7	x	1
30	5	2	no nucleus	7	20	8	1	1
70	5	10	5	2	15	9	1	1
11	7	4	2	6	20	10	1	1
17	2	1	8	x	x	10	1	1
12	6	0	8	3	10	11	1	1
35	4	0	no nucleus	5	20	12	1	1
40	2	1	no nucleus	x	x	12	1	1
30	5	0	4	4	20	13	1	1
x	x	x	x	7	20	14	x	1
x	x	x	x	2	15	15	x	1
45	5	3	7	6	20	16	1	1
x	x	x	x	4	15	17	x	1
x	x	x	x	6	18	18	x	1
110	6	1	no nucleus	6	20	19	1	1
35	3	1	10	3	15	20	1	1
50	13	1	9	5	18	21	1	1
40	4	8	18	6	18	22	1	1
200	10	1	8	6	20	23	1	1
27	6	1	6	7	20	24	1	2
120	20	1	no nucleus	6	18	25	1	2
x	x	x	x	7	18	26	x	2
35	10	2	10	7	20	27	1	2
50	10	5	5	5	20	28	1	2
25	4	2	2	6	20	29	1	2
25	6	1	5	5	20	30	1	2
40	10	0	30	4	16	31	1	2
30	20	0	12	7	18	32	1	2
<b>45,6*</b>	<b>8,0*</b>	<b>2,1*</b>	<b>8,5*</b>	<b>5,1*</b>	<b>18,1*</b>	<b>32</b>	<b>26</b>	<b>2</b>





## CHAPTER VI

**STRUCTURE OF A TAR RNA-LIGAND COMPLEX****Research Article:** Tripeptides from Synthetic Amino Acids Block the Tat-TAR Association and Slow Down HIV Spread in Cell Cultures

Verena Ludwig, Andreas Krebs, Michaela Stoll, Ursula Dietrich, Jan Ferner, Harald Schwalbe, Ute Scheffer, Gerd Dürner, and Michael W. Göbel  
*ChemBioChem* **2007**, 8, 1850-1856

Tripeptides were screened for their binding properties to HIV TAR-RNA by a biochemical FRET-assay, a microbiological reporter gene assay and NMR spectroscopy. These investigations resulted in a tripeptide containing a side chain pyrimidinyl group in its central residue as best ligand for further structural analyses.

The peptides were synthesized and the FRET-assays carried out in the group of Prof. Göbel, while the reporter gene assay was carried out in the group of Dr. Dietrich. The author of this thesis performed titrations of the most promising ligands to RNA by NMR spectroscopy to extract binding strengths and binding sites. Explanations, results and the discussion of the spectroscopic work were also supplied by the author.



# Tripeptides from Synthetic Amino Acids Block the Tat–TAR Association and Slow Down HIV Spread in Cell Cultures

Verena Ludwig,<sup>[a]</sup> Andreas Krebs,<sup>[a]</sup> Michaela Stoll,<sup>[b]</sup> Ursula Dietrich,<sup>[b]</sup> Jan Ferner,<sup>[a]</sup> Harald Schwalbe,<sup>[a]</sup> Ute Scheffer,<sup>[a]</sup> Gerd Dürner,<sup>[a]</sup> and Michael W. Göbel<sup>\*[a]</sup>

*Non-natural amino acids with aromatic or heteroaromatic side chains were incorporated into tripeptides of the general structure Arg-X-Arg and tested as ligands of the HIV RNA element TAR. Some of these compounds could compete efficiently with the association of TAR and Tat and downregulated a TAR-controlled re-*

*porter gene in HeLa cells. Peptide 7, which contains a 2-pyrimidinyl-alkyl chain, also inhibited the spread of HIV-1 in cell cultures. NMR studies of 7 bound to HIV-2-TAR gave evidence for contacts in the bulge region.*

## Introduction

HIV infection and AIDS still pose a major threat to human health. Current chemotherapies consist of drug combinations mainly targeted towards two viral proteins: the reverse transcriptase and the viral protease.<sup>[1]</sup> While these therapeutic schemes are powerful and life-saving, considerable drawbacks are related to toxicity problems, development of drug resistance, and price. The search for new small-molecule inhibitors of HIV therefore remains a challenge for chemistry and biology.<sup>[2]</sup>

Blocking viral entry into cells is one successful strategy that has recently led to a novel class of antiviral drugs.<sup>[3]</sup> The TAR–Tat system constitutes another interesting opportunity for chemotherapeutic intervention. TAR, the 59 mer *trans*-activation response element, is a stem–loop structure located at the 5' ends of all viral transcripts.<sup>[4]</sup> It forms a tight complex with the arginine-rich domain of the viral protein Tat. Upon binding of human proteins, in particular cyclin T1, to the RNA Tat complex, RNA polymerase II is activated by phosphorylation. This initiates efficient transcription of viral genes. When binding of Tat to TAR is prevented, transcription stalls shortly after the stem–loop structure of TAR. As a product of coevolution of RNA and viral and human proteins, TAR is well conserved compared to other parts of the HIV genome.<sup>[5]</sup> Great efforts to identify antiviral small molecules acting by TAR–Tat inhibition have consequently been made.<sup>[6–9]</sup>

Molecular recognition of RNA results from charge–charge attraction, hydrogen bonds, and stacking interactions. While Coulomb attraction alone is sufficient for high ligand affinities, the common drawback is promiscuous binding. Finding the proper balance between polar and hydrophobic interactions might be the key to simultaneously improving selectivity. We have started a research program targeted towards synthetic  $\alpha$ -amino acids that carry aromatic and heteroaromatic side

chains.<sup>[10]</sup> Such residues might contribute to RNA binding either through stacking with exposed bases or through hydrogen bonding. The main purpose of the peptidic backbone is to assemble several side chains with polar or hydrophobic properties in a chemically efficient way and to allow, if required, easy combinatorial variation. Here we report on tripeptides of the general structure  $D$ -Arg-X- $D$ -Arg-CONH<sub>2</sub>, where X is a non-natural aromatic or heterocyclic amino acid. Tripeptides selected from combinatorial libraries have been shown previously to act as efficient TAR–Tat inhibitors.<sup>[8c,d]</sup>

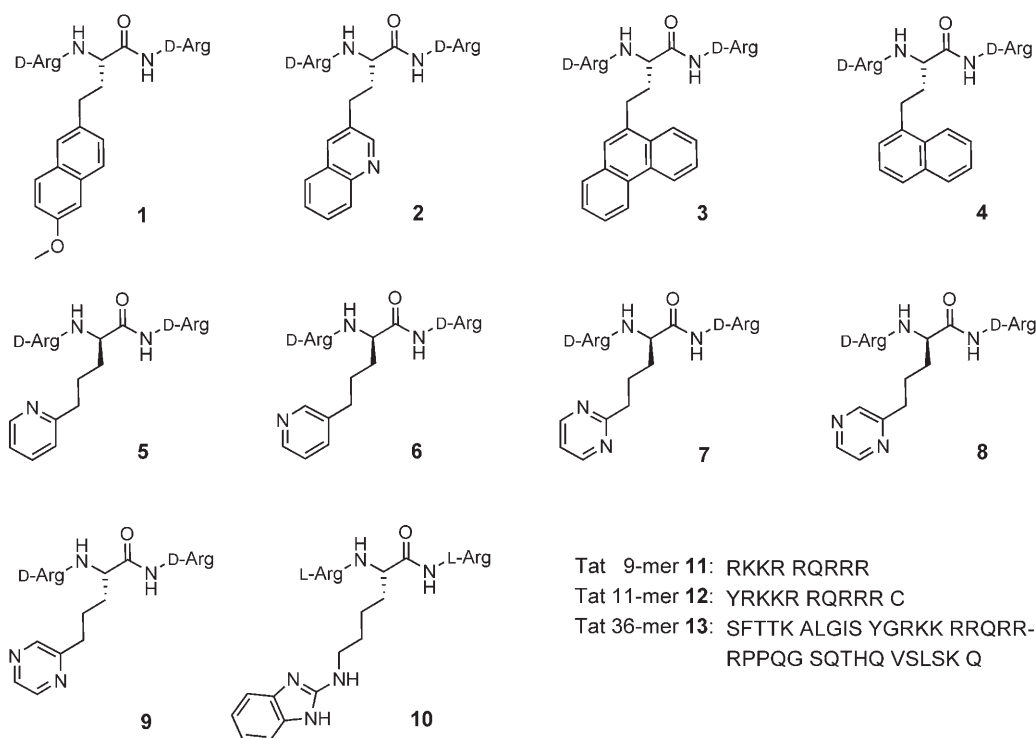
## Results and Discussion

The non-natural amino acids required to build tripeptides 1–9 were synthesized from unsaturated chiral precursors by hydroboration and Pd-catalyzed cross-coupling.<sup>[10]</sup> Standard Fmoc solid-phase synthesis then led to the peptides shown in Scheme 1.

Electrophilic species produced upon deprotection of arginine have the potential to interfere with aromatic side chains. One reason for the selection of  $D$ -Arg-X- $D$ -Arg-CONH<sub>2</sub>, therefore, was to check the stability of new amino acid building blocks and to examine their applicability for peptide libraries.

[a] Dr. V. Ludwig, Dr. A. Krebs, J. Ferner, Prof. Dr. H. Schwalbe, Dr. U. Scheffer, Dr. G. Dürner, Prof. Dr. M. W. Göbel  
Institute of Organic Chemistry and Chemical Biology  
Johann Wolfgang Goethe Universität Frankfurt  
Max-von-Laue-Strasse 7, 60438 Frankfurt am Main (Germany)  
Fax: (+49)69-798-29464  
E-mail: M.Gobel@chemie.uni-frankfurt.de

[b] M. Stoll, Dr. U. Dietrich  
Georg-Speyer-Haus, Institute for Biomedical Research  
Paul-Ehrlich-Strasse 42–44, 60596 Frankfurt am Main (Germany)

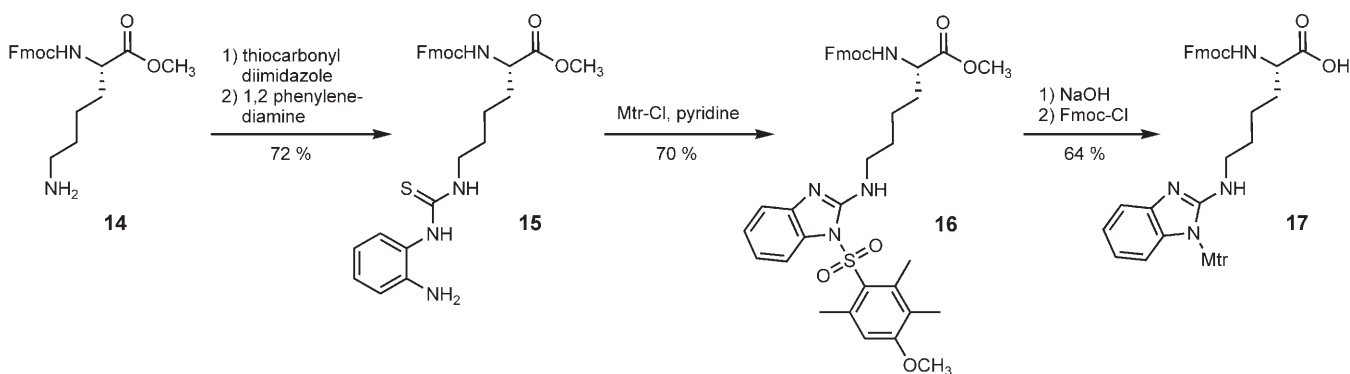


**Scheme 1.** Structures of peptides **1–13**. N terminus:  $\text{NH}_2$ , C terminus: with exception of peptide **13** (COOH), carboxamide.

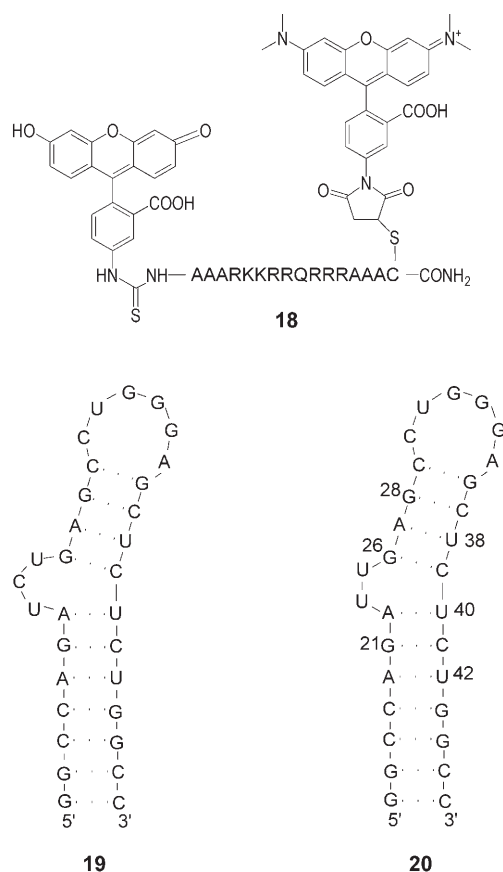
On the other hand, arginine-rich peptides can be expected to bind the bulge region of TAR RNA with high affinities. D-Arginine was chosen in order to enhance the stability against proteolytic degradation in cell cultures. Tripeptide **10** is based on the benzimidazole amino acid **17**, prepared from L-Fmoc-lysine methyl ester (**14**) as follows. Consecutive addition of thiocarbonyldiimidazole and benzene-1,2-diamine converted **14** into thiourea **15**. Upon protection of the free amino group, **15** spontaneously cyclized to **16**. Ester cleavage and reintroduction of the Fmoc group then led to the protected amino acid **17** (Scheme 2).

RNA-binding affinities of peptide ligands (Scheme 3) were determined in a fluorimetric competition assay.<sup>[69,7h,11]</sup> Peptide **18**, labeled with fluoresceine and rhodamine, adopts random coil conformations in the absence of RNA and shows efficient reciprocal quenching of the dyes. Association of **18** with the

TAR model **19** results in a 2.6-fold increase in fluorescence, presumably by driving the peptide into more extended conformations. The assay can conveniently be run in 96-well microplates at pH 7.4 and 37 °C with 100 nM concentrations of **18** and **19**, respectively. Titrating peptide **18** with the TAR models **19** and **20** established the 1:1 complex stoichiometries.  $K_d$  values of  $0.8 \pm 0.3$  nM (**18–19**) and  $0.4 \pm 0.1$  nM (**18–20**) could be calculated from these data. RNA ligands competing for the Tat binding site of **19** displace peptide **18** and reduce the fluorescence emission, thus allowing  $\text{IC}_{50}$  values to be determined as a measure of relative TAR affinities. These numbers do not necessarily reflect the formation of RNA–competitor complexes with exactly defined structures and stoichiometries. The assay will also respond to less specific and multiple interactions. We were reluctant, therefore, to calculate the  $K_d$  of a compound from its  $\text{IC}_{50}$  value.



**Scheme 2.** Synthesis of the Fmoc-protected benzimidazole amino acid **17**.



**Scheme 3.** Structures of the Tat peptide **18**, the HIV-1 TAR model **19**, and the HIV-2 TAR model **20**. Nucleotides exhibiting imino signal shifts in  $^1\text{H}$  NMR titrations with argininamide or peptide **7** are numbered.

$\text{IC}_{50}$  values for streptomycin, neomycin, argininamide, and a series of tripeptides are summarized in Table 1. Each compound was shown in a separate control not to quench the fluorescence of **18** in the absence of RNA. Unless a minimum of one arginine unit is present, most cationic tripeptides exhibit low to medium affinities, although a notable exception is Lys-Trp-Lys ( $\text{IC}_{50}=30\ \mu\text{M}$ ). For L-Lys-D-Lys-L-Asn, the major dis-

Competing ligand	$\text{IC}_{50}$ [ $\mu\text{M}$ ]	Competing ligand	$\text{IC}_{50}$ [ $\mu\text{M}$ ]
streptomycin	150	neomycin	7
argininamide	1400	His-Phe-His	3000
His-His-His	4200	His-Arg-His	250
Lys-Phe-Lys	9000	Lys-Tyr-Lys	3000
Lys-Trp-Lys	30	Lys-Lys-Lys	1000
Lys-Arg-Lys	400	L-Lys-D-Lys-L-Asn-COOH	30000
L-Lys-D-Lys-L-Asn	400	D-Arg-D-Phe-D-Arg	75
D-Arg-D-Tyr-D-Arg	80	D-Arg-D-Trp-D-Arg	2
D-Trp-D-Arg-D-Arg	90	D-Arg-D-Lys-D-Arg	30
Arg-Arg-Arg	4	D-Arg-D-Arg-D-Arg	4
Tat 11-mer <b>12</b>	0.50	Tat 36-mer <b>13</b>	0.12

Tripeptides were prepared by standard Fmoc solid-phase synthesis. The configurations of the amino acids are L if not otherwise stated. N terminus:  $\text{NH}_2$ , C terminus: with exception of peptide **13** (COOH), carboxamide. RNA: **19**. Experimental error:  $\pm 30\%$ .

crepancy between our and the published data ( $K_d$  in the nanomolar range) is not yet explained.<sup>[8c]</sup>  $\text{IC}_{50}$  values for all tripeptides of the general structure Arg-X-Arg range between 2 and 80  $\mu\text{M}$ . The tight binding of the tryptophan derivative, once more, points to the importance of stacking interactions. Not unexpectedly, Arg-Arg-Arg is among the best peptidic ligands, showing affinity comparable to that of its enantiomer D-Arg-D-Arg-D-Arg. Tat-TAR complexation has been characterized earlier as an unselective process.<sup>[12]</sup>

Table 2 shows the TAR affinities of the non-natural tripeptides **1–10**. All  $\text{IC}_{50}$  values are below 100  $\mu\text{M}$ . The best candi-

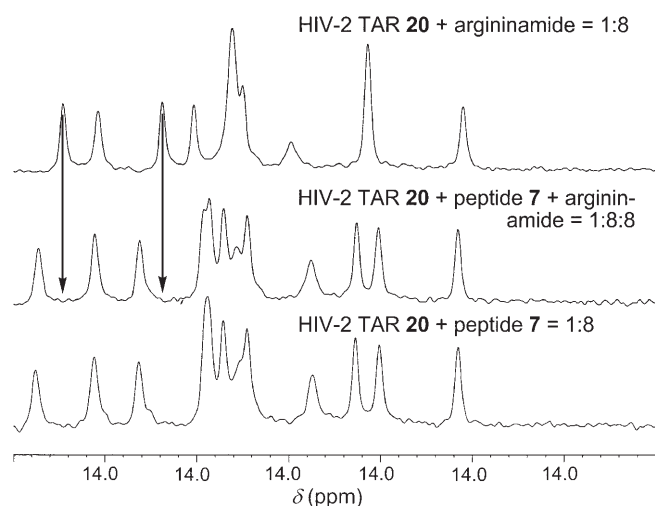
**Table 2.** TAR affinities of the non-natural tripeptides **1–10**.

Peptide	$\text{IC}_{50}$ [ $\mu\text{M}$ ]	Peptide	$\text{IC}_{50}$ [ $\mu\text{M}$ ]
<b>1</b>	15	<b>2</b>	30
<b>3</b>	3	<b>4</b>	60
<b>5</b>	80	<b>6</b>	20
<b>7</b>	2	<b>8</b>	80
<b>9</b>	80	<b>10</b>	2

N and C termini:  $\text{NH}_2$ , carboxamide. RNA: **19**. Experimental error:  $\pm 30\%$ .

dates (**3**, **7**, and **10**) are comparable to D-Arg-D-Trp-D-Arg, but not superior. The indole, aminobenzimidazole, and phenanthrene moieties of these peptides are excellent candidates for stacking interactions. Conjugation of cationic RNA ligands with such moieties is known to result in major affinity gains.<sup>[8f,13]</sup> The pyrimidine residue of peptide **7**, in contrast, is not a good stacking partner and is unable to be protonated at physiological pH. The remarkable  $\text{IC}_{50}$  value of 2  $\mu\text{M}$ , therefore, must result from other properties of the heterocycle, presumably its ability to accept hydrogen bonds.

When the fluorimetric competition assay was repeated with RNA **20**, a model of HIV-2 TAR exhibiting a two-nucleotide bulge (Scheme 3), tight binding of peptide **7** to RNA was again observed ( $\text{IC}_{50}=2\ \mu\text{M}$ ). This allowed the complex of **7** and **20** to be studied by NMR. Spectra of HIV-2 TAR generally show better resolution as a result of reduced conformational mobility [ $\text{K}_2\text{HPO}_4/\text{KH}_2\text{PO}_4$  (25 mM) and KCl (50 mM) at pH 6.2; 283 K]. Consistent with fluorimetric binding data, the phosphate buffer used for NMR experiments drastically lowered the TAR affinities of arginine-containing peptides. When TAR model **20** (200  $\mu\text{M}$ ) was titrated with argininamide, a characteristic set of imino proton signals—assigned to nucleotides located around the bulge region—was shifted (Figure 1).<sup>[14]</sup> These shift effects could be saturated with 8 equivalents of argininamide (Figure 1, top). The same set of signals was also shifted upon titration with **7** (Figure 1, bottom). The pyrimidine peptide therefore interacts with RNA **20** in the same binding mode as the classical bulge ligand argininamide. In a competition experiment with eightfold excesses of both peptide **7** and argininamide, only the signals of the **7**·**20** complex are observed (Figure 1, center). Even with a 12-fold excess of argininamide over peptide **7**, the imino spectra did not change significantly to indicate a displacement of peptide **7**, thus demon-



**Figure 1.**  $^1\text{H}$  NMR spectra of the HIV-2 TAR model **20** in the presence of argininamide (8 equiv; top), peptide **7** (8 equiv; bottom), and a mixture of both (8 equiv of each ligand; center). The imino region is shown. Note that the characteristic signals of the complex **20**-argininamide (arrows) are not detectable in the mixture of **20** with both ligands.

strating the superior TAR affinity of this compound. For the complex **7**-**20** no line-broadening or other signs of nonspecific interactions were found when the peptide concentration was further increased. In contrast, the phenanthrene **3** caused intense line-broadening in titrations with RNA **20**.

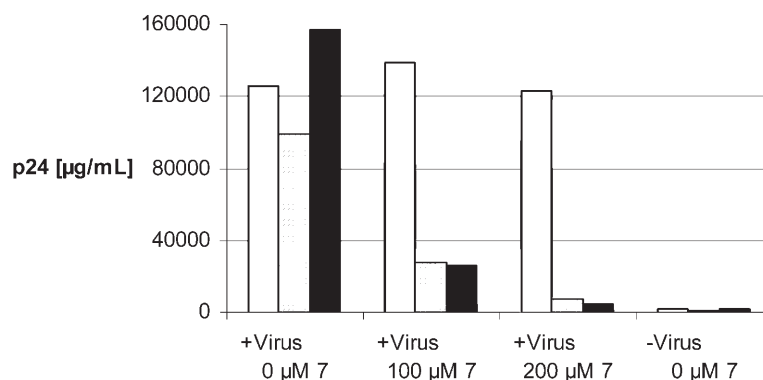
A reporter gene assay in HeLa P4 cells expressing the relevant receptors CD4, CCR5, and CXCR4, as well as the  $\beta$ -galactosidase gene under the control of the HIV-1 LTR promoter, was executed next in order to test the Tat–TAR inhibition in cell cultures. Upon infection with HIV-1, the newly translated Tat upregulates the  $\beta$ -galactosidase gene. The enzymatic activity is then quantified colorimetrically. Thus, compounds inhibiting the Tat–TAR interaction should reduce  $\beta$ -galactosidase activities. To exclude interference of this assay from toxic effects, the toxicities of the peptides were assessed by luminometric measurement of ATP in living cells. The results are summarized in Table 3. In a first set of experiments, the virus was added simultaneously with the peptides. The reference inhibitor, Tat peptide **11**,<sup>[15]</sup> is characterized by an  $\text{IC}_{50}$  value of  $20\ \mu\text{M}$ . D-Arg-D-Arg-D-Arg shows only weak effects ( $\text{IC}_{50} = 500\ \mu\text{M}$ ), probably due to promiscuous binding to all kinds of polyanions. Interesting  $\text{IC}_{50}$  values between 40 and  $70\ \mu\text{M}$ , however, were found with tripeptides **3**, **7**, and **10**. Cationic peptides such as **11**, apart from blocking the Tat–TAR association, are also known to interfere with viral entry by complexation of CXCR4.<sup>[7a,b,16]</sup> To distinguish between both possible mechanisms of reporter gene inhibition, in the

**Table 3.** Inhibition of a TAR–Tat-dependent reporter gene in HeLa cells by peptides **1**–**10**.

Peptide	$\text{IC}_{50}$ [ $\mu\text{M}$ ]		Toxicity ( $\text{TC}_{50}$ )
	Virus added simultaneously	Virus added first	
tat nonamer <b>11</b>	20	50	no effect <sup>[a]</sup>
His-His-His	> 1000	n.d.	n.d.
L-Lys-D-Lys-L-Asn	> 1000	n.d.	n.d.
D-Arg-D-Arg-D-Arg	500	no effect	n.d.
D-Arg-D-Trp-D-Arg	100	250	500
<b>1</b>	500	no effect	n.d.
<b>2</b>	250	no effect	no effect <sup>[d]</sup>
<b>3</b>	50	no effect	no effect <sup>[b]</sup>
<b>4</b>	100	no effect	no effect <sup>[c]</sup>
<b>5</b>	100	> 500	2000
<b>6</b>	150	> 500	no effect <sup>[d]</sup>
<b>7</b>	40	< 200	no effect <sup>[b]</sup>
<b>8</b>	500	no effect	n.d.
<b>9</b>	750	no effect	2000
<b>10</b>	70	500	no effect <sup>[c]</sup>

N and C termini:  $\text{NH}_2$ , carboxamide. Concentration maxima in toxicity assays: [a] up to  $100\ \mu\text{M}$ , [b] up to  $200\ \mu\text{M}$ , [c] up to  $500\ \mu\text{M}$ , [d] up to  $1000\ \mu\text{M}$ .

second set of experiments cells were infected first and then treated with peptides after 2 h. This time lag allows efficient infection of cells prior to incubation with peptides. Accordingly, higher concentrations were required to achieve 50% inhibition:  $50\ \mu\text{M}$  for the Tat peptide **11**, for example. The phenanthrene peptide **3** proved inactive under such conditions. It therefore seems probable that **3** blocks the assay mainly by inhibition of viral entry. In contrast, for two peptides, the pyrimidine compound **7** and D-Arg-D-Trp-D-Arg,  $\text{IC}_{50}$  values around  $200\ \mu\text{M}$  were observed. The antiviral activity of compound **7** was examined by measuring the amount of HIV produced in the culture supernatants after 2, 4, and 6 days in the continuous presence of the peptide. The viral p24 antigen, detectable by ELISA, served as a monitor to quantify HIV. Important reductions of p24 levels were observed in the presence of **7** at concentrations of 100 and  $200\ \mu\text{M}$  (Figure 2).



**Figure 2.** Spread of HIV in cell culture in the absence or presence of peptide **7**. The viral protein p24 serves to monitor the viral load after 2 days (white), 4 days (gray), and 6 days (black). A representative experiment is shown.

## Conclusions

The aromatic and heteroaromatic sidechains present in peptides **1–10** were not designed specifically to fit the bulge of HIV-TAR. Because of their general Arg-X-Arg structure, considerable affinities were expected, especially for compound **3**, characterized by its hydrophobic phenanthrene moiety. The good binding properties of peptide **7** to TAR, however, came as a surprise to us. This compound also showed the highest activity in the reporter-gene assay and significant inhibition of HIV-1. A closer structural investigation of the complex **7·20** is underway, directed at identifying potent non-peptidic analogues of **7**. The unexpected properties of peptide **7** also highlight the importance of structural diversity for ligand discovery. In consequence, peptide libraries of the non-natural amino acids present in compounds **1–10** are currently being investigated.

## Experimental Section

**General:** All reagents were of the highest grades commercially available. TLC: glass plates coated with silica gel F 254 (0.25 mm, Merck). Flash chromatography: silica 60 (40–63  $\mu\text{m}$ , 230–400 mesh, Merck). M.p.: Kofler hot plate microscope, uncorrected. UV: Varian Cary 1E. FT-IR: Jasco 420; in  $\tilde{\nu}$  [ $\text{cm}^{-1}$ ].  $^1\text{H}$  NMR: Bruker AM 250; chemical shifts ( $\delta$ ) in ppm relative to  $\text{Me}_4\text{Si}$  (0.00 ppm) or  $[\text{D}_6]\text{DMSO}$  (2.50 ppm) as internal standards. ESI-MS: Fisons VG Platform II. Elemental analysis: Heraeus HCN-Rapid.

**Hydrochloride of (S)-Fmoc-lysine methyl ester (14):**  $\text{SOCl}_2$  (2.25 mL, 3.8 equiv) was slowly added to a suspension of Fmoc-protected L-lysine (3.00 g, 8.14 mmol) in dry methanol (100 mL). The resulting clear solution was kept at room temperature for 2 h and then evaporated to dryness. Upon redissolving in methanol (20 mL) and dropwise addition into diethyl ether, the hydrochloride of ester **14** precipitated. A colorless solid (3.34 g, 98%) was obtained after filtration and drying. m.p. 154–155  $^\circ\text{C}$ ;  $^1\text{H}$  NMR ( $[\text{D}_6]\text{DMSO}$ ):  $\delta$  = 7.88 (m, 4H), 7.76–7.69 (m, 3H), 7.43–7.29 (m, 5H), 4.36–4.18 (m, 3H), 4.03–3.94 (m, 1H), 3.61 (s, 3H), 2.72 (brs, 2H), 1.70–1.33 ppm (m, 6H); IR (KBr):  $\tilde{\nu}$  = 3356, 2964, 2914, 1738, 1692, 1530, 1268, 1052, 736  $\text{cm}^{-1}$ ; elemental analysis calcd (%) for  $\text{C}_{22}\text{H}_{27}\text{ClN}_2\text{O}_4$ : C 63.08, H 6.50, N 6.69; found: C 62.94, H 6.35, N 6.75.

**Synthesis of thiourea 15:** Thiocarbonyl diimidazole (2.72 g, 1.5 equiv) and imidazole (0.20 g, 0.3 equiv) were dissolved in acetonitrile (50 mL) with gentle heating. Lysine ester hydrochloride **14** (4.30 g, 10.3 mmol) was then added at room temperature. After the system had been stirred for 2 h, benzene-1,2-diamine (1.66 g, 1.5 equiv) was added. Stirring at room temperature was continued overnight. Evaporation of the solvent and flash chromatography on silica gel (hexanes/AcOEt 2:1  $\rightarrow$  hexanes/AcOEt 1:1  $\rightarrow$  AcOEt) allowed thiourea **15** (3.96 g, 72%) to be isolated as a colorless solid. m.p. 69–73  $^\circ\text{C}$ ;  $^1\text{H}$  NMR ( $[\text{D}_6]\text{DMSO}$ ):  $\delta$  = 8.78 (s, 1H), 7.91 (d,  $J$  = 7.2 Hz, 2H), 7.84–7.72 (m, 3H), 7.46–7.29 (m, 4H), 7.01–6.95 (m, 2H), 6.75 (d,  $J$  = 7.2 Hz, 2H), 6.57 (t,  $J$  = 7.4 Hz, 1H), 4.81 (s, 2H), 4.34–4.21 (m, 3H), 4.06–4.00 (m, 1H), 3.64 (s, 3H), 3.41 (brs, 2H), 1.70–1.48 (m, 4H), 1.28 ppm (m, 2H); IR (KBr):  $\tilde{\nu}$  = 3356, 2947, 1718, 1534, 1450, 1217, 1078, 741  $\text{cm}^{-1}$ ; elemental analysis calcd (%) for  $\text{C}_{29}\text{H}_{32}\text{N}_4\text{O}_4\text{S}$ : C 65.39, H 6.06, N 10.52; found: C 65.49, H 6.04, N 10.46.

**N-Protection and cyclization to 16:** 4-Methoxy-2,3,6-trimethylbenzenesulfonyl chloride (Mtr chloride; 0.366 g, 4 equiv) and pyridine (0.12 mL, 4 equiv) were added at 0  $^\circ\text{C}$  to a solution of thiourea **15** (0.195 g, 0.367 mmol) in dichloromethane (20 mL). The mixture was stirred at 0  $^\circ\text{C}$  for 2 h and at room temperature overnight. Chromatographic purification on silica gel (hexanes/AcOEt 1:1  $\rightarrow$  hexanes/AcOEt 2:3  $\rightarrow$  AcOEt) and removal of solvents afforded benzimidazole **16** (0.183 g, 70%, colorless solid). m.p. 81–82  $^\circ\text{C}$ ;  $^1\text{H}$  NMR ( $[\text{D}_6]\text{DMSO}$ ):  $\delta$  = 7.88 (d,  $J$  = 7.2 Hz, 1H), 7.77 (d,  $J$  = 7.8 Hz, 1H), 7.67 (d,  $J$  = 7.4 Hz, 2H), 7.41–7.22 (m, 6H), 7.07–6.78 (m, 4H), 6.67 (d,  $J$  = 7.8 Hz, 1H), 4.28–4.15 (m, 4H), 4.06–3.98 (dd,  $J$  = 8.4, 4.9 Hz, 1H), 3.61 (s, 3H), 3.47–3.41 (m, 3H), 2.60 (s, 3H), 2.23 (s, 3H), 1.97 (s, 3H), 1.72–1.58 (m, 6H), 1.38–1.27 ppm (m, 2H); IR (KBr):  $\tilde{\nu}$  = 3342, 2942, 1720, 1579, 1450, 1309, 1177, 1143, 740  $\text{cm}^{-1}$ ; MS (ESI) calcd for  $\text{C}_{39}\text{H}_{42}\text{N}_4\text{O}_7\text{S}$ : 710.3; found:  $m/z$  (%) = 711.3 (100)  $[\text{M}+\text{H}^+]$ ; elemental analysis calcd (%) for  $\text{C}_{39}\text{H}_{42}\text{N}_4\text{O}_7\text{S}\cdot\text{H}_2\text{O}$ : C 64.27, H 6.08, N 7.69; found: C 64.19, H 6.12, N 7.65.

**N-Protected amino acid 17:** Methyl ester **16** (0.496 g, 0.698 mmol) was dissolved at room temperature in 1,4-dioxane (30 mL) and aqueous NaOH (1 N, 30 mL). After 2 h, the mixture was neutralized with aqueous hydrochloric acid (1 N, 30 mL). Solid  $\text{NaHCO}_3$  (0.5 g) was added, followed by 1,4-dioxane (30 mL) and Fmoc chloride (0.198 g, 1.1 equiv). After stirring at room temperature for 3 h, the mixture was saturated with NaCl and extracted several times with AcOEt. Evaporation of the solvent and flash chromatography on silica gel (AcOEt  $\rightarrow$  MeOH/AcOEt 1:9  $\rightarrow$  MeOH/AcOEt 1:2) allowed the protected amino acid **17** (0.312 g, 64%) to be isolated as a colorless solid. m.p. 168  $^\circ\text{C}$ ;  $^1\text{H}$  NMR ( $[\text{D}_6]\text{DMSO}$ ):  $\delta$  = 7.88 (d,  $J$  = 7.3 Hz, 2H), 7.78 (t,  $J$  = 6.2 Hz, 2H), 7.44–7.23 (m, 5H), 7.08–6.80 (m, 5H), 6.67 (d,  $J$  = 7.9 Hz, 1H), 4.31–4.27 (m, 2H), 3.88 (s, 3H), 3.47–3.44 (m, 2H), 2.62 (s, 3H), 2.24 (s, 3H), 2.02 (s, 3H), 1.87–1.36 ppm (m, 8H); IR (KBr):  $\tilde{\nu}$  = 3410, 2938, 1725, 1685, 1578, 1459, 1142, 1045, 739, 660  $\text{cm}^{-1}$ ; MS (ESI) calcd for  $\text{C}_{38}\text{H}_{40}\text{N}_4\text{O}_7\text{S}$ : 696.3; found:  $m/z$  (%) = 697.4 (100)  $[\text{M}+\text{H}^+]$ ; elemental analysis calcd (%) for  $\text{C}_{38}\text{H}_{40}\text{N}_4\text{O}_7\text{S}\cdot 2\text{H}_2\text{O}$ : C 62.28, H 6.05, N 7.65; found: C 62.30, H 5.82; N 7.66.

**Tripeptide 10 from amino acid 17:** The synthesis was performed on Rink amide MBHA resin (150 mg, theoretical loading 0.78  $\text{mmol g}^{-1}$ , 0.19 mmol) by standard procedures as published earlier.<sup>[10]</sup> After removal of the last Fmoc group, the resin was washed with DCM and dried in vacuo. The resin was then treated with TFA (1650  $\mu\text{L}$ ), thioanisole (100  $\mu\text{L}$ ), water (100  $\mu\text{L}$ ), phenol (50 mg), and EDT (50  $\mu\text{L}$ ). The resin turned red and after 5 h of gentle agitation it was filtered off. Upon trituration with  $\text{Et}_2\text{O}$ , the peptide precipitated from the filtrate. The precipitate was suspended in  $\text{Et}_2\text{O}$  several times and spun down in a centrifuge. After removal of the organic solvent, the pellet was dissolved in water, lyophilized, and purified by reversed-phase HPLC: Phenomenex AQUA C18 5  $\mu\text{m}$ , 250  $\times$  10 mm;  $\text{H}_2\text{O} + \text{MeOH} = 67 + 33 + 0.1\%$  trifluoroacetic acid. M ( $\text{C}_{25}\text{H}_{43}\text{N}_3\text{O}_3$ ) calcd = 573.4; MS (ESI)  $m/z$  (%) = 574.5 (100)  $[\text{M}+\text{H}]^+$ , 287.6 (81)  $[\text{M}+2\text{H}]^+$ .

**HPLC purification of other tripeptides:** Columns: Reprosil-pur-C18-AQ (250  $\times$  20), Prontosil-C18-H (250  $\times$  16), or Phenomenex AQUA C18 5  $\mu\text{m}$  (250  $\times$  20). Isocratic mixtures of 0.1% aqueous trifluoroacetic acid + acetonitrile ranging from 98 + 2 to 71 + 29 were used as solvents.

**In vitro transcription:** TAR model **19** was prepared by T7 transcription as published previously.<sup>[69]</sup> After desalting and lyophilization, the RNA pellet was redissolved in DEPC-treated water to a concentration of 100  $\mu\text{M}$  (stock solution) or 100 nM (final dilution). Before

quantification of RNA by UV absorption, monomers were completely removed by polyacrylamide gel electrophoresis.

**Inactivation of RNases:** All experimental steps were performed under sterile conditions. Plasticware, tubes, and most solutions were treated with diethylpyrocarbonate (DEPC). Solutions that were not compatible with DEPC treatment were prepared by mixing up molecular biology grade powdered reagents in DEPC-treated ultrapure water. All glassware was baked at 180 °C for 6 h.

**Determination of  $K_d$ :** Fluorescence-based binding assays were performed in 96-well microplates (Corning 6860, black, nonbinding surface) at 37 °C with use of a final volume of 100  $\mu$ L in TK buffer [Tris-HCl (50 mM), KCl (20 mM), Triton-X 100 (0.01%), pH 7.4].<sup>[69,11]</sup> Prior to titration, the RNAs [100 nM in Tris-HCl (5 mM), pH 7.4] were heated to 90 °C for 5 min and then immediately placed on ice for an additional 2–5 min. Tat peptide **18** (Thermo Electron Corporation, 10 nM) was titrated either with TAR RNA **19** or with TAR RNA **20**. The following concentration range for the RNA was chosen: 0, 4, 6, 8, 10, 12, 14, 16, 20, 30, 40 nM. The blank contained only TK buffer. After incubation for 5 min at 37 °C the samples were measured in a fluorescence microplate reader (Safire<sup>2</sup>; Tecan; excitation wavelength 489 nm, emission wavelength 590 nm). Fluorescence readouts are corrected by the blank value. A 1:1 binding model with  $K_d$  as a variable was then fitted to the experimental data points by a nonlinear least-squares procedure. Each  $K_d$  was calculated from four independent experiments.

**Determination of  $IC_{50}$ :** TAR RNA **19** and Tat peptide **18** were both used at final concentrations of 100 nM. The fluorescence of pure peptide **18** and of the Tat-TAR complex was determined first (Reader: Fluostar Galaxy, BMG Labtechnologies; excitation wavelength 540 nm, emission wavelength 590 nm). Titration curves were determined from eleven data points. The competitor concentration at which the fitted titration curve intersected with the mean fluorescence counts of the Tat-TAR complex and of uncomplexed Tat was taken as the  $IC_{50}$  value of the peptide.

**Reporter gene inhibition:** HeLa P4 cells expressing the relevant HIV-1 receptors CD4, CCR5, and CXCR4, as well as the  $\beta$ -galactosidase reporter gene under the control of the HIV-1 LTR promoter, were used for all cell culture experiments (obtained through the NIH AIDS Research and Reference Program). Cells ( $1.2 \times 10^4$  per well) were added to a 96-well microplate in medium (100  $\mu$ L) and incubated overnight at 37 °C in a cell incubator. The next day, cells were washed and infected with HIV-1<sub>Lai</sub> (30  $\mu$ L) in medium (total volume of 100  $\mu$ L) at 37 °C for 48 h in the presence of increasing concentrations of the compound to be tested. Two days later, cells were washed and lysed with harvest buffer [50  $\mu$ L, glycerol (2.5 mL), MES-Tris (1.25 mL), DTT (1 M, 25  $\mu$ L), Triton X-100 (250  $\mu$ L), H<sub>2</sub>O (up to 25 mL)] for 10 min on ice. Cells were centrifuged at 1200 rpm for 10 min, and cell lysate (3  $\mu$ L) was added to reaction buffer [33  $\mu$ L, MgCl<sub>2</sub> (1 M, 15  $\mu$ L), NaH<sub>2</sub>PO<sub>4</sub>/Na<sub>2</sub>HPO<sub>4</sub> (0.5 M, 3 mL), Galacton (100 $\times$ , 150  $\mu$ L), H<sub>2</sub>O (up to 15 mL)] in a microtiter plate well. The plate was shaken for 45 min in the dark and measured in the luminometer (Lumistar Galaxy, BMG Labtechnologies) after addition of amplifier [25  $\mu$ L per well, NaOH (80  $\mu$ L), Emerald (10 $\times$ , 400  $\mu$ L), H<sub>2</sub>O (up to 4 mL)].

**Cytotoxicity assays:** Cytotoxicity assays were performed with the Vialight Plus kit from Cambrex, which quantifies the amount of ATP in living cells luminometrically. HeLa P4 cells were seeded ( $1.5 \times 10^4$  per well) and incubated overnight with medium [100  $\mu$ L, DMEM (10% FCS), L-glutamine (2%), penicillin/streptomycin (1%), gentamicin (500  $\mu$ g mL<sup>-1</sup>), Pyromycin (1  $\mu$ g mL<sup>-1</sup>)] in a cell incubator at 37 °C. The next day, medium was replaced by fresh medium con-

taining different concentrations of the compounds to be tested. Measurements were performed in triplicate. After 2 days, cells were washed carefully with fresh medium and were then lysed according to the instructions of the manufacturer to extract the ATP from the cells. After 10 min, the ATP monitoring reagent was added to generate the luminescent signal, which was quantified 2 min later with the aid of a luminometer (Lumistar Galaxy, BMG Labtechnologies).

**Antiviral activity:** For long-term inhibition studies, HeLa P4 cells ( $0.4 \times 10^4$  per well) were plated (24 well plate) in medium (400  $\mu$ L) and incubated at 37 °C overnight. The next day, cells were washed and infected with HIV-1<sub>Lai</sub> (120  $\mu$ L) in a total volume of 400  $\mu$ L for 2 h at 37 °C. After several washings, peptide **7** was added to give final concentrations of 0, 100, and 200  $\mu$ M. After 2 and 4 days the supernatants were withdrawn and replaced by fresh medium containing peptide **7** at the given concentrations. After 6 days, all supernatants were analyzed for p24 antigen by a commercial ELISA based on a biotinylated monoclonal p24 antibody, the binding of which is quantified with a chromogenic substrate for peroxidase conjugated to streptavidin (Innotest HIV Antigen, Innogenetics, Heiden, Germany).

**1D proton NMR:** The interaction between peptide **7** and TAR model **20** (MWG-Biotech AG) was monitored by 1D homonuclear <sup>1</sup>H NMR during the titration of a solution of RNA (0.15 mM) with the peptide up to an eightfold excess of peptide **7**. The solutions were buffered with potassium phosphate (25 mM) and potassium chloride (50 mM) at a pH of 6.2. The 1D experiments were recorded with a Jump-Return-Echo sequence where the excitation maximum was set to the chemical shift range of the nucleobase imino signals.<sup>[17]</sup> They were recorded on a Bruker 600 MHz spectrometer at 283 K. For the competition experiments with argininamide (from Sigma-Aldrich) the amino acid was titrated to the solution of **20** and **7** (ratio of 1 to 8) up to a 12-fold excess of argininamide relative to peptide **7**.

## Acknowledgements

Financial support by the Deutsche Forschungsgemeinschaft (SFB 579 "RNA Ligand Interactions") is gratefully acknowledged. We also thank Prof. Dr. Mirko Hennig, Scripps Research Institute, La Jolla, for a sample of HIV-2 TAR and Dr. Klaus Langer, University of Frankfurt, for offering access to his microplate reader (BMG Fluostar).

**Keywords:** antiviral agents • fluorescent probes • HIV • peptidomimetics • RNA recognition

- [1] a) S. G. Deeks, *Br. Med. J.* **2006**, 332, 1489; b) F. J. Piacenti, *Pharmacotherapy* **2006**, 26, 1111.
- [2] Discovery of a maturation inhibitor: F. Li, R. Goila-Gaur, K. Salzwedel, N. R. Kilgore, M. Reddick, C. Matallana, A. Castillo, D. Zoumplis, D. E. Martin, J. M. Orenstein, G. P. Allaway, E. O. Freed, C. T. Wild, *Proc. Natl. Acad. Sci. USA* **2003**, 100, 13555.
- [3] a) J. P. Moore, R. W. Doms, *Proc. Natl. Acad. Sci. USA* **2003**, 100, 10598; b) J. Kadow, H. G. Wang, P. F. Lin, *Curr. Opin. Investig. Drugs* **2006**, 7, 721; c) V. Briz, E. Poveda, V. J. Soriano, *J. Antimicrob. Chemother.* **2006**, 57, 619.
- [4] a) M. Stevens, E. De Clercq, J. Balzarini, *Med. Res. Rev.* **2006**, 26, 595; b) S. N. Richter, G. Palu, *Curr. Med. Chem.* **2006**, 13, 1305; c) S. Bannwarth, A. Gatignol, *Curr. HIV Res.* **2005**, 3, 61; d) M. Baba, *Curr. Top. Med.*



- Chem.* **2004**, *4*, 871; e) A. Krebs, V. Ludwig, O. Boden, M. W. Göbel, *ChemBioChem* **2003**, *4*, 972; f) J. Karn, *J. Mol. Biol.* **1999**, *293*, 235.
- [5] a) B. Klaver, B. Berkhout, *EMBO J.* **1994**, *13*, 2650; b) J.-C. Paillart, M. Dettenhofer, X.-F. Yu, C. Ehresmann, B. Ehresmann, R. Marquet, *J. Biol. Chem.* **2004**, *279*, 48397; c) E. Ramirez de Arellano, V. Soriano, A. Holguin, *AIDS Res. Hum. Retroviruses* **2005**, *21*, 949.
- [6] Small-molecule ligands of TAR interfering with binding of Tat: a) C. Parolin, B. Gatto, C. Del Vecchio, T. Pecere, E. Tramontano, V. Cecchetti, A. Fravolini, S. Masiero, M. Palumbo, G. Palu, *Antimicrob. Agents Chemother.* **2003**, *47*, 889; b) S. N. Richter, B. Gatto, O. Tabarrini, A. Fravolini, M. Palumbo, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4247; c) M. He, D. Yuan, W. Lin, R. Pang, X. Yu, M. Yang, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 3978; d) H.-Y. Mei, D. P. Mack, A. A. Galan, N. S. Halim, A. Heldsinger, J. A. Loo, D. W. Moreland, K. A. Sannes-Lowery, L. Sharmeen, H. N. Truong, A. W. Czarnik, *Bioorg. Med. Chem.* **1997**, *5*, 1173; e) H.-Y. Mei, M. Cui, A. Heldsinger, S. M. Lemrow, J. A. Loo, K. A. Sannes-Lowery, L. Sharmeen, A. W. Czarnik, *Biochemistry* **1998**, *37*, 14204; f) F. Hamy, V. Brondani, A. Flörsheimer, W. Stark, M. J. J. Blommers, T. Klimkait, *Biochemistry* **1998**, *37*, 5086; g) S. Renner, V. Ludwig, O. Boden, U. Scheffer, M. Göbel, G. Schneider, *ChemBioChem* **2005**, *6*, 1119; h) A. I. H. Murchie, B. Davis, C. Isel, M. Afshar, M. J. Drysdale, J. Bower, A. J. Potter, I. D. Starkey, T. M. Swarbrick, S. Mirza, C. D. Prescott, P. Vaglio, F. Aboul-ela, J. Karn, *J. Mol. Biol.* **2004**, *336*, 625; i) B. Davis, M. Afshar, G. Varani, A. I. H. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A. J. Potter, I. D. Starkey, T. Swarbrick, F. Aboul-ela, *J. Mol. Biol.* **2004**, *336*, 343; j) K. E. Lind, Z. Du, K. Fujinaga, B. M. Peterlin, T. L. James, *Chem. Biol.* **2002**, *9*, 185; k) Z. Du, K. E. Lind, T. L. James, *Chem. Biol.* **2002**, *9*, 707; l) M. Mayer, T. L. James, *J. Am. Chem. Soc.* **2004**, *126*, 4453; m) M. Mayer, P. T. Lang, S. Gerber, P. B. Madrid, I. Gómez Pinto, R. K. Guy, T. L. James, *Chem. Biol.* **2006**, *13*, 993; n) Z. Xiao, N. Zhang, Y. Lin, G. B. Jones, I. H. Goldberg, *Chem. Commun.* **2006**, 4431; o) D. Yuan, M. He, R. Pang, S. S. Lin, Z. Li, M. Yang, *Bioorg. Med. Chem.* **2007**, *15*, 265.
- [7] Tat-TAR inhibition by aminoglycosides: a) A. Litovchick, A. Lapidot, M. Eisenstein, A. Kalinkovich, G. Borkow, *Biochemistry* **2001**, *40*, 15612; b) A. Lapidot, V. Vijayabashkar, A. Litovchick, J. Yu, T. L. James, *FEBS Lett.* **2004**, *577*, 415; c) K. F. Blount, Y. Tor, *Nucleic Acids Res.* **2003**, *31*, 5490; d) K. F. Blount, F. Zhao, T. Hermann, Y. Tor, *J. Am. Chem. Soc.* **2005**, *127*, 9818; e) K. F. Blount, Y. Tor, *ChemBioChem* **2006**, *7*, 1612; f) E. Riguet, J. Désiré, O. Boden, V. Ludwig, M. Göbel, C. Bailly, J.-L. Décout, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4651; g) D. Raghunathan, V. M. Sanchez-Pedregal, J. Junker, C. Schwiegel, M. Kalesse, A. Kirschning, T. Carlomagno, *Nucleic Acids Res.* **2006**, *34*, 3599; h) S. Yajima, H. Shionoya, T. Akagi, K. Hamasaki, *Bioorg. Med. Chem.* **2006**, *14*, 2799; i) B. Clique, A. Ironmonger, B. Whittaker, J. Colley, J. Titchmarsh, P. Stockley, A. Nelson, *Org. Biomol. Chem.* **2005**, *3*, 2776; j) A. Ironmonger, B. Whittaker, A. J. Baron, B. Clique, C. J. Adams, A. E. Ashcroft, P. G. Stockley, A. Nelson, *Org. Biomol. Chem.* **2007**, *5*, 1081.
- [8] Tat-TAR inhibition by peptides: a) F. Hamy, E. R. Felder, G. Heizmann, J. Lazdins, F. Aboul-ela, G. Varani, J. Karns, T. Klimkait, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 3548; b) T. Klimkait, E. R. Felder, G. Albrecht, F. Hamy, *Bio-technol. Bioeng.* **1998**, *61*, 155; c) S. Hwang, N. Tamilarasu, K. Ryan, I. Huq, S. Richter, W. C. Still, T. M. Rana, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 12997; d) S. Hwang, N. Tamilarasu, K. Kibler, H. Cao, A. Ali, Y.-H. Ping, K.-T. Jeang, T. M. Rana, *J. Biol. Chem.* **2003**, *278*, 39092; e) C.-W. Lee, H. Cao, K. Ichiyama, T. M. Rana, *Bioorg. Med. Chem. Lett.* **2005**, *15*, 4243; f) V. Peytoux, R. Condom, N. Patino, R. Guedj, A.-M. Aubertin, N. Gelus, C. Bailly, R. Terreux, D. Cabrol-Bass, *J. Med. Chem.* **1999**, *42*, 4042; g) Z. Athanassiou, K. Patora, R. L. A. Dias, K. Moehle, J. A. Robinson, G. Varani, *Biochemistry* **2007**, *46*, 741.
- [9] Tat-TAR inhibition by oligonucleotides: a) J. J. Turner, G. D. Ivanova, B. Verbeure, D. Williams, A. A. Arzumanov, S. Abes, B. Lebleu, M. J. Gait, *Nucleic Acids Res.* **2005**, *33*, 6837; b) C. Di Primo, I. Rudloff, S. Reigadas, A. A. Arzumanov, M. J. Gait, J. J. Toulmé, *FEBS Lett.* **2007**, *581*, 771; c) S. Tripathi, B. Chaubey, S. Ganguly, D. Harris, R. A. Casale, V. N. Pandey, *Nucleic Acids Res.* **2005**, *33*, 4345.
- [10] A. Krebs, V. Ludwig, J. Pfizer, G. Dürner, M. W. Göbel, *Chem. Eur. J.* **2004**, *10*, 544.
- [11] C. Matsumoto, K. Hamasaki, H. Mihara, A. Ueno, *Bioorg. Med. Chem. Lett.* **2000**, *10*, 1857.
- [12] A. Garbesi, F. Hamy, M. Maffini, G. Albrecht, T. Klimkait, *Nucleic Acids Res.* **1998**, *26*, 2886.
- [13] a) K. Hamasaki, A. Ueno, *Bioorg. Med. Chem. Lett.* **2001**, *11*, 591; b) N. W. Luedtke, Q. Liu, Y. Tor, *Biochemistry* **2003**, *42*, 11391.
- [14] a) A. S. Brodsky, J. R. Williamson, *J. Mol. Biol.* **1997**, *267*, 624; b) M. Hennig, J. R. Williamson, *Nucleic Acids Res.* **2000**, *28*, 1585; c) K. T. Dayie, A. S. Brodsky, J. R. Williamson, *J. Mol. Biol.* **2002**, *317*, 263.
- [15] I. Choudhury, J. Wang, A. B. Rabson, S. Stein, S. Pooyan, S. Stein, M. J. Leibowitz, *J. Acquired Immune Defic. Syndr. Hum. Retrovirol.* **1998**, *17*, 104.
- [16] a) D. Schols, *Curr. Top. Med. Chem.* **2004**, *4*, 883; b) H. Xiao, C. Neuveut, H. L. Tiffany, M. Benkirane, E. A. Rich, P. M. Murphy, K.-T. Jeang, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11466; c) S. Ghezzi, D. M. Noonan, M. G. Aluigi, G. Vallanti, M. Cota, R. Benelli, M. Morini, J. D. Reeves, E. Vicenzi, G. Poli, A. Albini, *Biochem. Biophys. Res. Commun.* **2000**, *270*, 992; d) D. Daelemans, D. Schols, M. Witvrouw, C. Pannecouque, S. Hatse, S. van Dooren, F. Hamy, T. Klimkait, E. De Clercq, A.-M. Vandamme, *Mol. Pharmacol.* **2000**, *57*, 116.
- [17] P. Plateau, M. Gueron, *J. Am. Chem. Soc.* **1982**, *104*, 7310.

Received: May 2, 2007

Published online on September 21, 2007



## **Research Article:** Structures of HIV TAR RNA-Ligand Complexes Reveal Higher Binding Stoichiometries

Jan Ferner, Marcel Suhartono, Sven Breitung, Henry R. Jonker, Mirko Hennig, Jens Wöhnert, Michael W. Göbel, and Harald Schwalbe  
*ChemBioChem* **2009**, 10, 1490-1494

The structural determination of the tripeptide ligand in complex with the TAR-RNA is summarized in this article. An intensive investigation was necessary to find out that the complex has a ligand-RNA stoichiometry of 2:1. The strong binding site is located within the RNA bulge as observed with other TAR ligands. The second, lower affinity binding site, however, could be verified as potential help for further ligand optimization.

The thesis author was in charge of almost all experiments presented in this work, as well as for the data analysis and the writing. Göbel and co-workers provided the ligand and Dr. Jonker was responsible for the set-up and maintenance of the structural calculations.

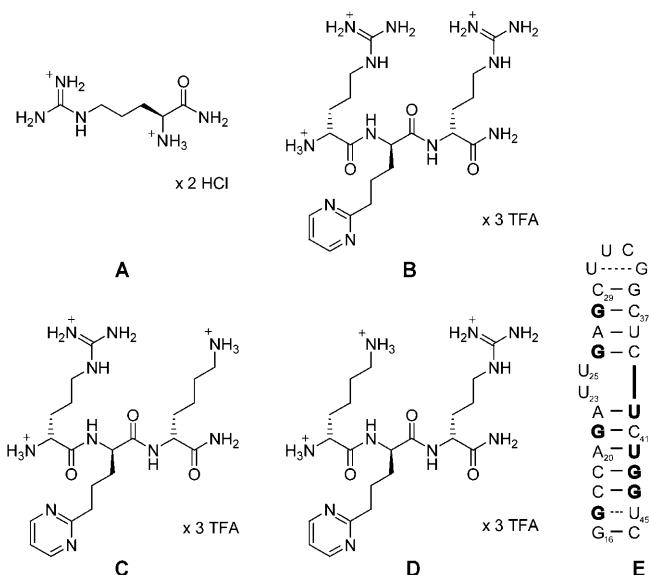


## Structures of HIV TAR RNA–Ligand Complexes Reveal Higher Binding Stoichiometries

Jan Ferner,<sup>[a]</sup> Marcel Suhartono,<sup>[b]</sup> Sven Breitung,<sup>[b]</sup> Hendrik R. A. Jonker,<sup>[a]</sup> Mirko Hennig,<sup>[c]</sup> Jens Wöhnert,<sup>[d]</sup> Michael Göbel,<sup>\*[b]</sup> and Harald Schwalbe<sup>\*[a]</sup>

The development of low molecular weight ligands that bind to RNA with high affinity and specificity remains challenging.<sup>[1]</sup> The trans-activation response element TAR is regarded as an important RNA target to control the replication cycle of the human immunodeficiency virus (HIV). In particular, the internal bulge within the apical hairpin of TAR (E' in Scheme S1 in the Supporting Information) has been investigated widely as a drug target because its interaction with the trans-activator protein (Tat) is essential for the transcription of viral proteins.<sup>[2]</sup> The hairpin contains a six nucleotide loop and an internal bulge. The size of the bulge is the only difference between HIV-1 and HIV-2 strains. HIV-1 TAR has a trinucleotide bulge with the sequence U23-C24-U25 while HIV-2 TAR lacks the central cytidine.

The bulge functions as a flexible linker that leads to substantial angular fluctuation of the two canonical stems.<sup>[3]</sup> Initially, the complex of TAR with argininamide (Scheme 1 A) served as a mimic of the complex between TAR and Tat and provided important insights into the conformation of TAR in the protein–RNA complex. Argininamide binds with a low millimolar dissociation constant in the bulge region.<sup>[4]</sup> In HIV-1<sup>[5]</sup> and HIV-2<sup>[6]</sup> TAR, argininamide induces the formation of a base triple, which is formed between the bulge residue U23 and the canonical A27–U38 base pair. In conjunction with G26, this region also displays the most contacts to argininamide.<sup>[4,7]</sup> Complexes of TAR with several further ligands have been investigated including Tat-derived peptides,<sup>[8]</sup> ligands with heteroaromatic moieties,<sup>[9]</sup> guanidinium-like ligands,<sup>[10]</sup> aminoglycosides<sup>[11]</sup> as well as divalent cations.<sup>[12]</sup> Almost all ligands bind in the bulge



**Scheme 1.** Constitution of the ligands (A–D) and secondary structure of the HIV-2 TAR RNA (E) investigated in this study. Residues of the RNA whose CSPs are presented in Figure 1 are highlighted.

region and therefore compete directly with Tat. Despite different recognition modes, all ligands have been reported to rigidify the interhelical angle.<sup>[3a,13]</sup> Recently, Zhang et al. showed that the conformation of free TAR represents a dynamic ensemble that samples all possible interhelical angles detected in the various RNA–ligand structures. Ligands therefore select and stabilize free state conformations rather than induce new ones.<sup>[14]</sup>

In our investigations, we were interested whether a ligand could occupy two binding sites present in the ensemble of free state structures. Although many TAR–ligand complexes are formed with 1:1 stoichiometry, one of the first structural studies of TAR–ligand complexes in 1995 provided evidence that the tightly binding peptide ligand ADP-1 binds to two distinct sites of HIV-1 TAR.<sup>[15]</sup> Very recently, a cyclic aminoglycoside analogue was shown to bind TAR with RNA/ligand-stoichiometry of 2:3.<sup>[11b]</sup>

Here, we report the structural characterization of complexes between TAR and novel peptidic ligands developed by Göbel and co-workers. These compounds block Tat–TAR association in vitro and attenuate HIV proliferation in cell cultures.<sup>[16]</sup> The tripeptides contain two D-arginines flanking a non-natural amino acid with a heteroaromatic side chain and two variants with D-lysine replacing D-arginine (Scheme 1 B–D). The ligands were designed to contain a heteroaromatic amino acid to provide hydrophobic stacking interactions in addition to the elec-

[a] J. Ferner, Dr. H. R. A. Jonker, Prof. Dr. H. Schwalbe  
Institut für Organische Chemie und Chemische Biologie  
Zentrum für Biomolekulare Magnetische Resonanz (BMRZ)  
Johann Wolfgang Goethe-Universität Frankfurt am Main  
Max-von-Laue-Strasse 7, 60438 Frankfurt am Main (Germany)  
Fax: (+49) 69-798-29515  
E-mail: schwalbe@nmr.uni-frankfurt.de

[b] Dr. M. Suhartono, S. Breitung, Prof. Dr. M. Göbel  
Institut für Organische Chemie und Chemische Biologie  
Johann Wolfgang Goethe-Universität Frankfurt am Main  
Max-von-Laue-Strasse 7, 60438 Frankfurt am Main (Germany)

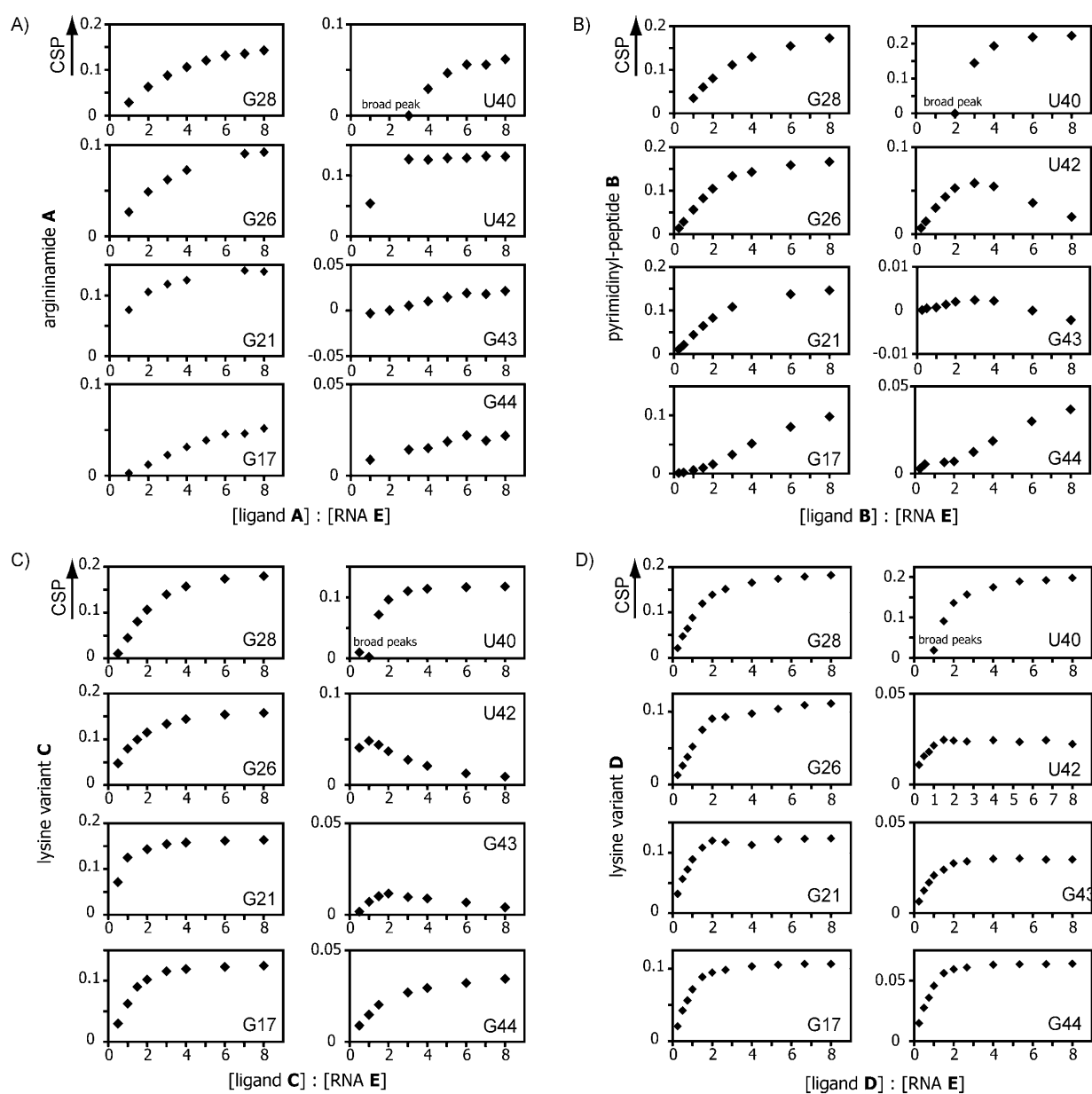
[c] Prof. Dr. M. Hennig  
Department of Biochemistry and Molecular Biology,  
Medical University of South Carolina  
173 Ashley Avenue, PO Box 250509, Charleston, SC 29425 (USA)

[d] Prof. Dr. J. Wöhnert  
Institut für Molekulare Biowissenschaften  
Zentrum für Biomolekulare Magnetische Resonanz (BMRZ)  
Johann Wolfgang Goethe-Universität Frankfurt am Main  
Max-von-Laue-Strasse 9, 60438 Frankfurt am Main (Germany)

Supporting information for this article is available on the WWW under <http://dx.doi.org/10.1002/cbic.200900220>.

trostatic attraction exerted by the two arginine side chains. Low micromolar  $IC_{50}$  values have been observed in competition experiments with a dye-labeled Tat-peptide. The affinities to TAR RNA result from specific binding to the bulge region in combination with weaker interactions in at least one alternative binding site. Our NMR data provide a structural model and reveal that ligands in excess, with similar constitution to **B** can bind with higher stoichiometries in line with previous mass spectrometry data.<sup>[17]</sup> Our findings of higher binding stoichiometries are not unprecedented; in fact, similar ligands have been reported to exhibit higher binding stoichiometries at ligand excess.<sup>[9d,e]</sup> Structural models, however, have not been reported.

We performed a titration of argininamide **A** to HIV-2 TAR (**E**) that resulted in continuous chemical shift perturbations (CSPs) of the imino  $^1H$  NMR resonances around the bulge up to a saturation level and indicated binding in the fast exchange limit (Figure 1A); residues further away show marginal CSPs upon addition of ligand (Figure S1A). The imino resonance of residue U40 is not detectable in the free RNA but can be detected upon addition of argininamide due to acquired protection against exchange with the solvent water. In agreement with Brodsky et al.,<sup>[7]</sup> argininamide binds directly below the bulge, selects a single RNA conformation from the dynamic ensemble of structures in the apo state, and stabilizes the A22-U40 base pair and the base triple of U23-A27-U38.

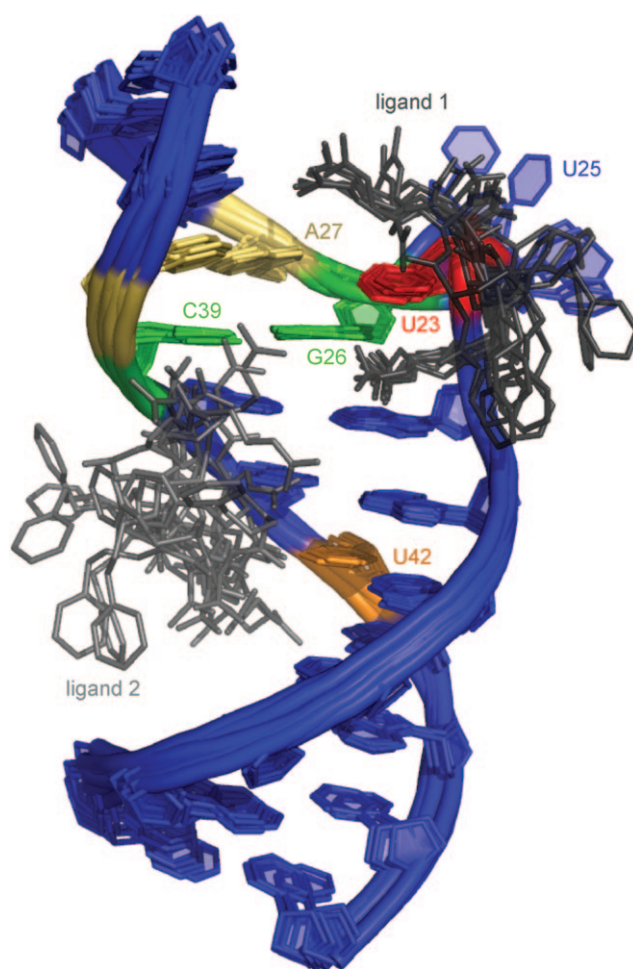


**Figure 1.** Chemical shift perturbations (CSPs) of the imino protons of RNA **E** as a function of the [ligand]/[RNA] ratio for the ligands **A**) argininamide, **B**) the pyrimidinyl-peptide and **C**) and **D**) the lysine variants.

For the newly designed ligand **B**, a more complex titration behavior is observed (Figure 1 B). The CSPs are substantially larger (by a factor of two) than those observed with argininamide. Interestingly, the NMR chemical shift response changes in the course of titration. At intermediate [ligand]/[RNA] ratio, the CSPs are the largest. Up to a ratio of 2:1, the observed shift changes resemble the argininamide titration profile. At higher [ligand]/[RNA] ratios, the so far unaffected signals of the imino protons of residues G17 and G44 start to change. Concomitantly, the imino resonances of U42 and G43 start to move into the opposite direction. A similar behavior is observed for the CSPs of the H5/H6 cross peaks detected in TOCSY spectra during the titration (Figure S2). While the residues in and around the bulge (U23, U25, C39 and U40) change their chemical shifts of H5 and H6 significantly at the beginning of the titration, resonances of residues C18, C41, U42, C45 either start to change or the CSPs redirect at a [ligand]/[RNA] ratio above 2:1. Taken together, this indicates the formation of a distinct 1:1 complex at intermediate [ligand]/[RNA] ratios and occupation of a second, weaker binding site for ligand **B** within the lower stem of TAR at higher [ligand]/[RNA] ratios.

In order to calculate the structure of the complex, we investigated a TAR construct containing a UUCG-tetraloop (Scheme 1 E) due to its favourable NMR characteristics. The ligand binding characteristics of this RNA remain unaltered. Structure calculations assuming a 1:1 complex resulted in two different structural models in which the ligand either binds in the major groove of the bulge similar to argininamide<sup>[4]</sup> and ligand rbt203 from Davis et al.<sup>[9d]</sup> or on the opposing side similar to the binding site of aminoglycosides such as neomycin B.<sup>[11a]</sup> In both structural models, the ligand is in close proximity to the nucleobase of residue U23 to which it shows strong intermolecular NOEs. The orientation of this bulge residue is different in the two models (Figure S3). For both 1:1 complexes, however, nine out of 48 intermolecular distance restraints localized in the lower stem are severely violated and the convergence rate of the structure calculations is low. We further analysed the subnanosecond dynamics of the bulge by using <sup>13</sup>C relaxation data.<sup>[18]</sup> The dynamics revealed that only nucleobase U25 remains flexible upon addition of the ligand while U23 is as rigid as the stem residues (Figure S4). Since neither doubling of the resonances nor peculiar dynamics of residue U23 could be detected, we conclude that the solution structure at saturation levels of ligand can neither be described by both structures nor by a rapid conformational equilibrium between the two structures.

Therefore, we performed NMR-based structural calculations assuming a 2:1 complex stoichiometry in agreement with our previous mass spectrometry results of related compounds.<sup>[17]</sup> In the ternary complex residue U23 is located in the major groove close to the base pairs G26-C39 and A27-U38 (Figure 2). However, no evidence for the formation of the base triple U23-A27-U38 could be obtained experimentally (data not shown). For the first ligand molecule with a larger number of stronger intermolecular NOEs, binding around the nucleobase of U23 is observed with the N-terminal Arg1 side chain below and the C-terminal Arg3 side chain above the plane of the



**Figure 2.** Ten best structures of the ternary complex (red = U23, green = G26-C39, yellow = A27-U38, orange = U42, black = ligand 1, gray = ligand 2).

base. This arrangement of the guanidinium groups is very similar to the 1:1 TAR-complex with rbt203.<sup>[9d]</sup> The N-terminal arginine is structurally well defined, and its guanidinium group is in close proximity to the Hoogsteen side of G26 to which it can form hydrogen bonds with the acceptor atoms N7 and O6. The more flexible C-terminal arginine shows transient interactions with the nucleobase of A27 and therefore prevents the contact of U23 and A27. This observation is in agreement with the absence of a base triplet in the ternary complex. The central pyrimidine residue points out to the solution and does not interact with the RNA. The second ligand molecule is localized in the major groove of the bottom stem along the pyrimidine stretch U40-C41-U42 but its structure is less defined. The intermolecular contacts are exclusively mediated by the two arginine residues. Occupation of this second binding site has been previously reported for other TAR complexes with structurally unrelated ligands.<sup>[11b,15]</sup>

Based on these results and in order to test whether specificity could be improved without compromising affinity, two additional tripeptides were synthesized and investigated. Replacement of the C-terminal Arg3 with lysine (Scheme 1 C) improved the binding affinity reducing the IC<sub>50</sub> value by a factor of four. Importantly, it showed identical patterns of CSPs during the

NMR titration when compared to **B**. In particular, U42 and G43 represent the region where the effects of the two binding sites are accentuated. Since the influence of the second ligand binding occurs already at a lower [ligand]/[RNA] ratio (Figure 1C), the first, higher-affinity binding site in the bulge is saturated at an earlier stage: in agreement with the observed lower  $IC_{50}$ , both affinity and specificity for the first binding site are increased. The replacement of the N-terminal arginine with a lysine (Scheme 1D) leads to a slight increase of the  $IC_{50}$  value and does not show the characteristics in the NMR titration as observed with ligands **B** or **C**. This observation points to the importance of an N-terminal arginine residue for higher affinity to the binding site in the bulge.

From a structural point of view, the role of the heteroaromatic side chain in the central residue of the tripeptide remains undefined. The lack of intermolecular NOEs is likely due to the absence of a persistent direct contact of the heteroaromatic side chain with the target RNA. Substitution with amino acids containing other aromatic side chains (for example, phenanthrene or pyrazinyl) results in similar binding affinities and influences on the imino resonance shifts but leads to a broadening of the imino NMR signals (Supporting Information, Figure S5).

We conclude that the high affinity of our novel TAR-ligands is based on two properties:

- 1) The terminal residues clasp around the nucleobase of residue U23 in the bulge capable to form hydrogen bonds to the nucleobases of G26 and A27. A C-terminal lysine is favored since its side chain amine group can bind to the N7 of A27 but experiences less repulsion from the adenine amine group compared to a guanidinium group of an arginine. Additionally, cation- $\pi$  interactions between the arginine guanidinium group and nucleobases are likely determinants that stabilize this binding. The central amino acid acts as spacer. Its aromatic side chain has an indirect influence on the complexation by its effects on ligand solubility, dynamics and probably entropic contributions to binding.
- 2) Ligands with arginine side chains bind to at least one additional site on TAR. This site is localized along the pyrimidine stretch U40-C41-U42 directly below the bulge adjacent to residue G43 and has previously been observed in the context of ligand-bound HIV-1 and HIV-2 TAR.<sup>[11b,15]</sup> The specificity is surprising since this region is expected to adopt a standard double helical A-RNA structure. It is interesting to note that Varani and co-workers have recently detected subnanosecond dynamics for the stem-residue G43 in free HIV-1 TAR,<sup>[19]</sup> a dynamical feature that might be linked to the lower stems capability to bind additional ligand molecules.

In conclusion, NMR spectroscopy revealed two binding sites of varying affinity for peptidic ligands containing non-natural amino acids on TAR RNA. We wish to discuss our findings here in the context of the netropsin-DNA complex by Wemmer and co-workers, a seminal finding changing further development of specific DNA binders.<sup>[20]</sup> The NMR data in this case revealed a

2:1 binding stoichiometry in the ligand-DNA complex, different to proposed models at the time. In our case, we find that the bulge region of TAR RNA represents the major determinant for selectivity; affinity, however, can be further increased by additional positive charges targeting the lower stem of TAR. The detection of multiple binding sites, although some are weak, may inspire novel medicinal chemistry and enable the exploitation of fragment-based ligand design to target HIV TAR RNA.<sup>[21]</sup>

## Experimental Section

Titration of RNA (**E** and **E'**) with the ligands were monitored by 1D  $^1H$  NMR with RNA (150  $\mu M$ ) in a buffer (at pH 6.2) containing  $K_3PO_4$  (25 mM) and KCl (50 mM) and ligand (up to eightfold excess).

For resonance assignment and structure determination, the following NMR experiments were measured in a RNA sample (**E**, 0.65 mM) with eightfold excess of ligand **B**: normal and constant time  $^1H,^{13}C$  HSQC,  $^1H,^{15}N$  HSQC,  $^2J$   $^1H,^{15}N$  HSQC,<sup>[22]</sup> HNN-COSY,<sup>[23]</sup> H5NN-COSY,<sup>[24]</sup> 2D H(C)N,<sup>[25]</sup> 3D  $^{13}C$ -edited NOESY-HSQC, 3D  $^{15}N$ -edited NOESY-HSQC, 3D HCCH-COSY, 3D HCCH-TOCSY,<sup>[26]</sup> 3D forward-directed HCC-TOCSY-CCH E.COSY,<sup>[27]</sup> 2D  $^{15}N$ -edited CPMG-NOESY,<sup>[28]</sup> 2D  $^1H^1H$ -NOESY.<sup>[29]</sup> To cross-validate the structures, residual dipolar couplings (RDC) measured with IPAP-HSQC and with pf1 phages as aligning medium were used.<sup>[30]</sup> All experimental data were processed with Topspin 1.3 (Bruker, Germany) and analyzed with Sparky 3.114 (T. D. Goddard and D. G. Keller, UCSF, USA).

Calculations of the RNA-complex structure were performed with CNS 1.1<sup>[31]</sup> by using the ARIA 1.2 setup and protocols.<sup>[32]</sup> The dna-rna-allatom force field was used with OPLS parameters<sup>[33]</sup> and restraints are summarized in Table S1. Modelling of the complexes with 1:1 and 1:2 stoichiometries (RNA/ligand) were achieved using a high ambiguity driven docking approach with the program HADDOCK 2.1.<sup>[34]</sup> The ambiguous interaction restraints (AIRs) were defined from intermolecular NOEs which were classified by their intensity into either strong (distance closer than 4 Å), medium (distance closer than 5 Å) and weak (distance closer than 6 Å).

## Acknowledgements

H.S. is member of the DFG-funded Center of Excellence: Macromolecular Complexes. The work was supported by the state of Hessen (BMRZ), the Sonderforschungsbereich 579 "RNA-Ligand-Interactions" and the European Commission (Spinell and eNMR).

**Keywords:** HIV TAR ligands • NMR spectroscopy • peptides • RNA structures • RNA

- [1] a) H. Schwalbe, M. W. Göbel, G. Wess, *ChemBioChem* **2003**, *4*, 915–917; b) J. R. Thomas, P. J. Hergenrother, *Chem. Rev.* **2008**, *108*, 1171–1224.
- [2] a) B. Berkhout, R. H. Silverman, K. T. Jeang, *Cell* **1989**, *59*, 273–282; b) J. Karn, *J. Mol. Biol.* **1999**, *293*, 235–254.
- [3] a) K. T. Dayie, A. S. Brodsky, J. R. Williamson, *J. Mol. Biol.* **2002**, *317*, 263–278; b) Q. Zhang, A. C. Stelzer, C. K. Fischer, H. M. Al-Hashimi, *Nature* **2007**, *450*, 1263–1267.
- [4] J. Tao, A. D. Frankel, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 2723–2726.
- [5] J. D. Puglisi, R. Tan, B. J. Calnan, A. D. Frankel, J. R. Williamson, *Science* **1992**, *257*, 76–80.
- [6] A. S. Brodsky, H. A. Erlacher, J. R. Williamson, *Nucleic Acids Res.* **1998**, *26*, 1991–1995.
- [7] A. S. Brodsky, J. R. Williamson, *J. Mol. Biol.* **1997**, *267*, 624–639.



- [8] a) K. M. Weeks, D. M. Crothers, *Cell* **1991**, *66*, 577–588; b) M. J. Churcher, C. Lamont, F. Hamy, C. Dingwall, S. M. Green, A. D. Lowe, P. J. G. Butler, M. J. Gait, J. Karn, *J. Mol. Biol.* **1993**, *230*, 90–110; c) Z. Athanassiou, R. L. A. Dias, K. Moehle, N. Dobson, G. Varani, J. A. Robinson, *J. Am. Chem. Soc.* **2004**, *126*, 6906–6913.
- [9] a) L. Dassonneville, F. Hamy, P. Colson, C. Houssier, C. Bailly, *Nucleic Acids Res.* **1997**, *25*, 4487–4492; b) N. Gelus, F. Hamy, C. Bailly, *Bioorg. Med. Chem.* **1999**, *7*, 1075–1079; c) K. E. Lind, Z. Du, K. Fujinaga, B. M. Peterlin, T. L. James, *Chem. Biol.* **2002**, *9*, 185–193; d) B. Davis, *J. Mol. Biol.* **2004**, *336*, 343–356; e) A. I. H. Murchie, *J. Mol. Biol.* **2004**, *336*, 625–638.
- [10] A. Lapidot, E. Ben-Asher, M. Eisenstein, *FEBS Lett.* **1995**, *367*, 33–38.
- [11] a) C. Faber, H. Sticht, K. Schweimer, P. Röscher, *J. Biol. Chem.* **2000**, *275*, 20660–20666; b) D. Raghunathan, V. M. Sánchez-Pedregal, J. Junker, C. Schwiégk, M. Kalesse, A. Kirschning, T. Carlomagno, *Nucleic Acids Res.* **2006**, *34*, 3599–3608.
- [12] a) J. A. Ippolito, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9819–9824; b) T. E. Edwards, S. T. Sigurdsson, *Biochem. Biophys. Res. Commun.* **2003**, *303*, 721–725.
- [13] M. Zacharias, P. J. Hagerman, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6052–6056.
- [14] Q. Zhang, X. Sun, E. D. Watt, H. M. Al-Hashimi, *Science* **2006**, *311*, 653–656.
- [15] F. Aboul-ela, J. Karn, G. Varani, *J. Mol. Biol.* **1995**, *253*, 313–332.
- [16] a) A. Krebs, V. Ludwig, J. Pfizer, G. Dürner, M. W. Göbel, *Chem. Eur. J.* **2004**, *10*, 544–553; b) V. Ludwig, A. Krebs, M. Stoll, U. Dietrich, J. Ferner, H. Schwalbe, U. Scheffer, G. Dürner, M. W. Göbel, *ChemBioChem* **2007**, *8*, 1850–1856.
- [17] N. Morgner, H.-D. Barth, B. Brutschy, U. Scheffer, S. Breitung, M. W. Göbel, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1600–1611.
- [18] a) E. Duchardt, H. Schwalbe, *J. Biomol. NMR* **2005**, *32*, 295–308; b) J. Ferner, A. Villa, E. Duchardt, E. Widjajakusuma, J. Wöhnert, G. Stock, H. Schwalbe, *Nucleic Acids Res.* **2008**, *36*, 1928–1940.
- [19] M. F. Bardaro Jr, Z. Shajani, K. Patora-Komisarska, J. A. Robinson, G. Varani, *Nucleic Acids Res.* **2009**, *37*, 1529–1540.
- [20] a) M. Mrksich, W. S. Wade, T. J. Dwyer, B. H. Geierstanger, D. E. Wemmer, P. B. Dervan, *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 7586–7590; b) B. H. Geierstanger, J. P. Jacobsen, M. Mrksich, P. B. Dervan, D. E. Wemmer, *Biochemistry* **1994**, *33*, 3055–3062.
- [21] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531–1534.
- [22] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1994**, *4*, 117–122.
- [23] a) A. J. Dingley, S. Grzesiek, *J. Am. Chem. Soc.* **1998**, *120*, 8293–8297; b) K. Pervushin, A. Ono, C. Fernandez, T. Zyperski, M. Kainosho, K. Wüthrich, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14147–14151.
- [24] J. Wöhnert, A. J. Dingley, M. Stoldt, M. Görlach, S. Grzesiek, L. R. Brown, *Nucleic Acids Res.* **1999**, *27*, 3104–3110.
- [25] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, *J. Biomol. NMR* **1993**, *3*, 721–727.
- [26] A. Pardi, E. P. Nikonowicz, *J. Am. Chem. Soc.* **1992**, *114*, 9202–9203.
- [27] a) H. Schwalbe, J. P. Marino, S. J. Glaser, C. Griesinger, *J. Am. Chem. Soc.* **1995**, *117*, 7251–7252; b) J. P. Marino, H. Schwalbe, S. J. Glaser, C. Griesinger, *J. Am. Chem. Soc.* **1996**, *118*, 4388–4395; c) S. J. Glaser, H. Schwalbe, J. P. Marino, C. Griesinger, *J. Magn. Reson. Ser. B* **1996**, *112*, 160–180.
- [28] L. Mueller, P. Legault, A. Pardi, *J. Am. Chem. Soc.* **1995**, *117*, 11043–11048.
- [29] a) G. Varani, I. Tinoco, Jr., *Q. Rev. Biophys.* **1991**, *24*, 479–532; b) K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
- [30] a) F. Tian, H. M. Al-Hashimi, J. L. Craighead, J. H. Prestegard, *J. Am. Chem. Soc.* **2001**, *123*, 485–492; b) M. Ottiger, F. Delaglio, A. Bax, *J. Magn. Reson.* **1998**, *131*, 373–378.
- [31] A. T. Brünger, *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54*, 905–921.
- [32] J. P. Linge, S. I. O'Donoghue, M. Nilges, *Methods Enzymol.* **2001**, *339*, 71–90.
- [33] a) W. L. Jorgensen, J. Tirado Rives, *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666; b) J. Pranata, S. G. Wierschke, W. L. Jorgensen, *J. Am. Chem. Soc.* **1991**, *113*, 2810–2819.
- [34] a) C. Dominguez, R. Boelens, A. M. J. J. Bonvin, *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737; b) S. J. de Vries, A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. M. J. J. Bonvin, *Proteins Struct. Funct. Bioinf.* **2007**, *69*, 726–733.

Received: April 8, 2009

Published online on May 14, 2009

Please note: Minor changes have been made to this manuscript since its publication in *ChemBioChem* Early View. The Editor

**CHEMBIOCHEM**

## Supporting Information

© Copyright Wiley-VCH Verlag GmbH & Co. KGaA, 69451 Weinheim, 2009

# Supporting Information

for

## Structures of HIV TAR RNA-Ligand Complexes Reveal Higher Binding Stoichiometries

Jan Ferner, Marcel Suhartono, Sven Breitung, Hendrik R. A. Jonker, Mirko Hennig, Jens Wöhnert, Michael Göbel\*, and Harald Schwalbe\*

**RNA synthesis:** TAR RNA was either purchased from MWG-Biotech or prepared by in-vitro transcription with T7 RNA polymerase and purified by HPLC.<sup>[S1]</sup> All samples were dissolved in a phosphate buffer (25 mM K<sub>3</sub>PO<sub>4</sub>, 50 mM KCl, pH 6.2) in either 90% H<sub>2</sub>O / 10% D<sub>2</sub>O or 100% D<sub>2</sub>O.

**Preparation of the pyrimidine amino acid:** Fmoc-D<sup>2</sup>-Pyrim-OH was synthesized as published previously.<sup>[S2]</sup> A more efficient alternative route, starting from D-glutamic acid is accepted for publication in *Synthesis*.

**General procedure for solid-phase peptide synthesis:** Coupling of *N*-Fmoc-amino acids to Rink amide MBHA resin: After swelling of the resin (30 mg) with CH<sub>2</sub>Cl<sub>2</sub> (2 x 2 mL) for 15 min, the solvent was replaced by NMP. The Fmoc-protected resin was treated with piperidine (25% in DMF) three times (15 min, 5 min, 2 min) to liberate the amine. Afterwards, the resin was washed 5 times with NMP. For coupling, the Fmoc-amino acid, e.g. Fmoc-D-Arg<sup>Pmc</sup>-OH (2 equiv), was dissolved in NMP (1.5 mL) together with HOBt (3 equiv) and DIC (3 equiv). This mixture was shaken for 10 min, before it was added to the resin. After 3 h of gentle agitation, the resin was washed with NMP and tested for quantitative coupling by the Kaiser test. When the chain extension was shown to be complete, the Fmoc group was removed as described above. After washing with NMP (5 times) subsequent Fmoc-protected amino acids were attached using the same procedure as before.

Cleavage from the resin and removal of protecting groups: Having disconnected the final Fmoc group, the resin was washed 5 times with NMP, 5 times with CH<sub>2</sub>Cl<sub>2</sub> and then dried in vacuo. Afterwards, the resin was treated with TFA (1630 μL), thioanisol (100 μL), H<sub>2</sub>O (100 μL), PhOH (100 mg), EDT (50 μL), and TIS (20 μL). The resin turned red and after shaking for 90 min the filtered solution was poured into ice cooled Et<sub>2</sub>O. The peptide precipitated from the filtrate. The precipitate was suspended several times in Et<sub>2</sub>O (supersonic bath) and spun down in a centrifuge (1800 G). After removal of the organic solvent, the pellet was redissolved in H<sub>2</sub>O and concentrated by use of a speedvac to eliminate volatile impurities. For the use of the peptides as TFA salts in RNA binding assays a further purification by HPLC is absolutely necessary. To avoid impurities, which may disturb the assay, it is recommended to use a fresh HPLC column for the purification step.

**Analytical data:** H<sub>2</sub>N-DArg-D2Pyrim-DArg-CONH<sub>2</sub>. Yield: 4.2 mg (30 %, 30 mg Rink Amide MBHA resin subst.: 0.54 mmol/g). HPLC: analytical: Reprosil AQ, 125 x 4.6, 0.1 % TFA/MeCN 100:23, 0.8 mL/min,  $t_R = 2.37$  min; preparative separation: Reprosil AQ, 250x20, 0.1 % TFA/MeCN 100:23, 7 mL/min,  $t_R = 10.0$  min; second preparative preparation: Bischoff-Prontosil, 250x16, 0.1% TFA/MeCN 100:6.5, 7 mL/min,  $t_R = 11.6$  min. MS (ESI) m/z (%) = 507.31 (1.11) [M+H]<sup>+</sup>, 254.11 (100) [M+2H]<sup>2+</sup>, 170.40 (2.29) [M+3H]<sup>3+</sup>; C<sub>21</sub>H<sub>38</sub>N<sub>12</sub>O<sub>3</sub> calcd 506.32.

H<sub>2</sub>N-DLys-D2Pyrim-DArg-CONH<sub>2</sub>. Yield: 5.4 mg (40 %, 30 mg Rink Amide MBHA resin subst.: 0.54 mmol/g). HPLC: analytical: Reprosil AQ, 125x4.6, 0.1 % TFA/MeCN 100:23, 0.8 mL/min,  $t_R = 2.35$  min; preparative separation: Reprosil AQ, 250 x 20, 0.1 % TFA/MeCN 100:23, 7 mL/min,  $t_R = 9.8$  min; second preparative separation: Bischoff-Prontosil, 250x16, 0.1% TFA/MeCN 100:6.5, 7 mL/min,  $t_R = 10.6$  min. MS (ESI) m/z (%) = 479.10 (11.55) [M+H]<sup>+</sup>, 240.00 (100) [M+2H]<sup>2+</sup>; C<sub>21</sub>H<sub>38</sub>N<sub>10</sub>O<sub>3</sub> calcd 478.31.

H<sub>2</sub>N-DArg-D2Pyrim-DLys-CONH<sub>2</sub>. Yield: 2.3 mg (17 %, 30 mg Rink Amide MBHA resin subst.: 0.54 mmol/g). HPLC: analytical: Reprosil AQ, 125x4.6, 0.1 % TFA/MeCN 100:23, 0.8 mL/min,  $t_R = 2.39$  min; preparative separation: Reprosil AQ, 250 x 20, 0.1 % TFA/MeCN 100:23, 7 mL/min,  $t_R = 10.2$  min; second preparative separation: Bischoff-Prontosil, 250x16, 0.1 % TFA/MeCN 100:6.5, 7 mL/min,  $t_R = 10.0$  min. MS (ESI) m/z (%) = 479.2 (14.09) [M+H]<sup>+</sup>, 240.1 (100) [M+2H]<sup>2+</sup>; C<sub>21</sub>H<sub>38</sub>N<sub>10</sub>O<sub>3</sub> calcd 478.31.

H<sub>2</sub>N-dLys-D2Pyrim-dLys-CONH<sub>2</sub>. Yield: 2.4 mg (18 %, 30 mg Rink Amide MBHA resin subst.: 0.54 mmol/g). HPLC: analytical: Reprosil AQ, 125x4.6, 0.1 % TFA/MeCN 100:23, 0.8 mL/min,  $t_R$  = 2.35 min; preparative separation: Reprosil AQ, 250 x 20, 0.1 % TFA/MeCN 100:23, 7 mL/min,  $t_R$  = 9.2 min; MS (ESI)  $m/z$  (%) = 451.18 (12.43)  $[M+H]^+$ , 226.01 (100)  $[M+2H]^{2+}$ ; C<sub>21</sub>H<sub>38</sub>N<sub>8</sub>O<sub>3</sub> calcd 450.31.

**NMR spectroscopy:** The titration experiments were recorded on Bruker 600 MHz spectrometer at 283 K with a jump-return-Echo.<sup>[S3]</sup> For resonance assignment and structure determination, <sup>13</sup>C- and <sup>15</sup>N-labeled RNA was synthesized<sup>[S4]</sup> and the experiments mentioned in figure capture 3 were recorded at 283 K and 298 K.

**Structure calculation:** Calculations of the RNA-complex structure were performed with CNS 1.1<sup>[31]</sup> by using the ARIA 1.2 setup and protocols.<sup>[32]</sup> The dna-rna-allatom force field was used with OPLS parameters<sup>[33]</sup> and the restraints (Table S1). In the first eight iterations, 50 structures were calculated and the NOE distance restraints were calibrated from the ten lowest energy structures. The violation tolerance was progressively reduced to 0.1 Å in the final iteration, in which 200 structures were cal-

**Table S1:** Experimental restraints for the calculations with ARIA1.2 and HADDOCK-2.1 and the RMSD of the heavy atoms in the ternary complex between TAR RNA **E** and two ligand molecules **B**.

Experimental Restraints	
RNA-RNA distances (unambiguous)	468
RNA-RNA distances (ambiguous)	51
ligand-ligand distances	98
RNA-ligand1 distances	39
RNA-ligand2 distances	9
hydrogen bond distances	60
base pair planarities	10
residual dipolar couplings	43
dihedrals	175
Total	951
RMSD (all)	1.399
RMSD (RNA only)	0.766
RMSD (rigid RNA only)	0.507

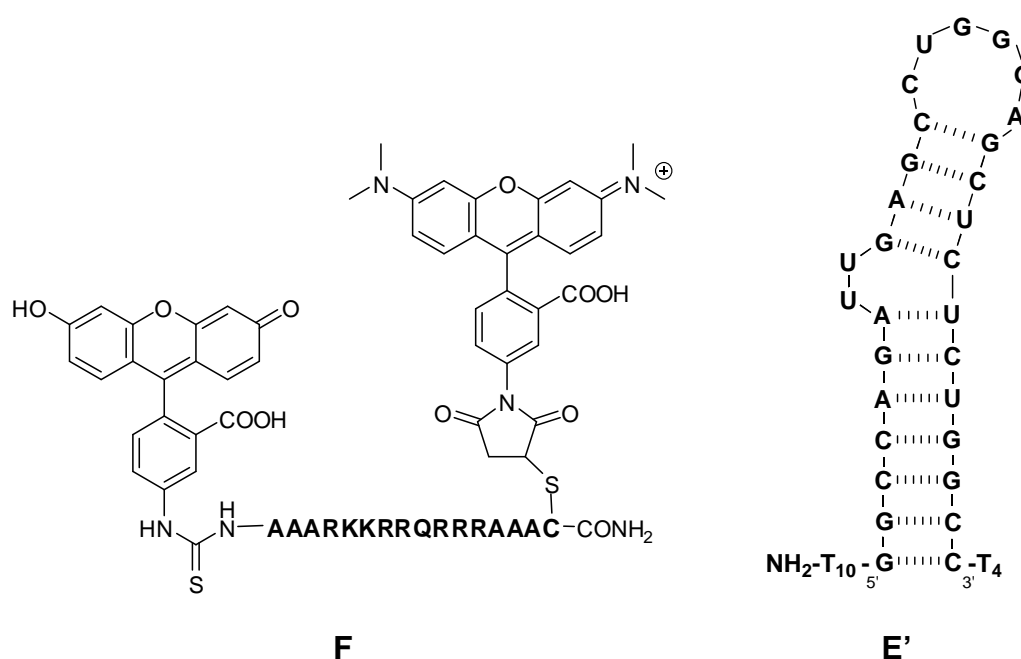
culated. For structure calculations, the standard simulated annealing (SA) protocol in ARIA was used with torsion angle dynamics (TAD). The initial high-temperature stage consisted of 10000 steps at 10000 K and a high-temperature conformational search of 8000 steps to 2000 K. The following two cooling stages reduced the temperature in 80 000 steps to 1000 K and 40000 to 0.01 K. During the SA protocol the force constants for the NOE distance and hydrogen bond restraints was set to  $10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  during the high-temperature stages and  $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$  during the cooling stages. Dihedral restraints were incorporated with force constants of  $5 \text{ kcal mol}^{-1} \text{ rad}^{-2}$  in the high-temperature stage,  $25 \text{ kcal mol}^{-1} \text{ rad}^{-2}$  in the first cooling and  $200 \text{ kcal mol}^{-1} \text{ rad}^{-2}$  in the second cooling stage. RDC restraints were not used in the high-temperature stage but in the first and second cooling stage with  $0.1 \text{ kcal mol}^{-1} \text{ Hz}^{-2}$  and  $0.5 \text{ kcal mol}^{-1} \text{ Hz}^{-2}$ , respectively.

Modelling of the complex between HIV-2 TAR RNA and one or two molecules of ligand **B** was achieved using a high ambiguity driven docking approach with the program HADDOCK 2.1.<sup>[34]</sup> The ambiguous interaction restraints (AIRs) were defined from intermolecular NOEs which were classified by their intensity into either strong (distance closer than 4 Å), medium (distance closer than 5 Å) and weak (distance closer than 6 Å). Previous to the docking, arbitrary starting structures were generated from the 20 best RNA structures and the 20 best ligand structures to first calculate the coordinates of the 1:1 complex where the ligand binds to the high affinity site. In the first iteration, a rigid docking was performed starting at 2000 K (5000 steps) and cooling down to 300 K in 1000 steps. Then, a semi-flexible docking was accomplished from 300 K to 50 K in 2000 steps with only the RNA bulge residues U23, U25 and the ligand flexible. The final docking stage cooled the complex from 300 K to 0 K in 100 steps with only a fully flexible ligand. In these calculations, the intermolecular distance restraints were assigned a force constraint weaker than the intramolecular RNA restraints ( $30 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ) during the final cooling stage. In the next step, the strongly violated intermolecular distance restraints were assigned to a second ligand and a ternary complex was calculated starting with 20 RNA structures and 5 ligand structures each. The remaining parameters were used as previously.

**Determination of peptide concentrations:** The correct concentration of each peptide was obtained by UV spectrometry. A protected derivative of the 2-pyrimidinyl amino alcohol was used as a standard. UV/Vis (EtOH/H<sub>2</sub>O 2:1):  $\lambda_{\text{max}} = 248 \text{ nm}$  ( $\epsilon_{248} =$

3400 L·mol<sup>-1</sup>·cm<sup>-1</sup>). The gravimetric method cannot be recommended since it leads to fluctuations of up to 50 %.

**RNA binding assay:** A fluorimetric competition assay was used to determine the RNA binding affinities of compounds **B-D**, **G**. Tat peptide **F**, labeled with fluorescein and rhodamine, has a flexible structure and shows efficient fluorescence quenching of the dyes when RNA is absent. Upon binding to TAR, the mean distance between the dyes is enlarged leading to increased quantum yields of fluorescence. When the Tat peptide is displaced from TAR by a competing ligand, fluorescence goes down to the initial intensity, allowing to determine IC<sub>50</sub> values. Since peptide **B** may form higher order complexes with TAR beyond the 1:1 stoichiometry, these numbers should not be interpreted as thermodynamic constants. However, they allow a precise ranking of relative ligand affinities to the TAR RNA. The binding assay is performed in 96 well microtiter plates (Corning 6860, black, non-binding surface) at 37 °C with a final volume of 100 μL in TK buffer (50 mM Tris-HCl, 20 mM KCl, 0.01 % Triton-X100, pH 7.4). Prior to titration, the RNA [100 nm in Tris-HCl (5 mM), pH 7.4] was heated to 90°C for 5 min and then immediately placed on ice for additional 2–5 min.



**Scheme 1:** Model of the Fluorescein/Rhodamine labeled Tat peptide and HIV-2 TAR-RNA.

**Determination of IC<sub>50</sub>:** TAR RNA **E'** and Tat peptide **F** were both used at final concentrations of 10 nM. The fluorescence of pure peptide **F** and of the Tat-TAR complex was determined first (Reader: Tecan safire<sup>2</sup>; excitation wavelength 489 nm, emission

wavelength 590 nm). Titration curves with the competitor were then determined from at least eleven data points. The competitor concentration at which the fitted titration curve intersected with the mean fluorescence counts of the Tat–TAR complex and of uncomplexed Tat was taken as the IC<sub>50</sub> value of the peptide. The deviation of TAR affinities published here (Table S2) and in a previous report <sup>[9b]</sup> is only in part explained by the different assay conditions. It also turned out that a scrupulous purification of the peptides is essential for obtaining reliable IC<sub>50</sub> data. Trace impurities in the old samples of **B-D,G** not visible under the published HPLC conditions must have influenced to some extent the binding data reported before. <sup>[9b]</sup>

**Table S2:** Analysis of the HIV-2 TAR-RNA affinity with tripeptides containing 2-Pyrimidinyl amino acid via FRET assay.

H <sub>2</sub> N-(D)X-(D)2Pyrim-(D)X-CONH <sub>2</sub>	IC <sub>50</sub> [μM] HIV-2
H <sub>2</sub> N-(D)Arg-(D)2Pyrim-(D)Arg-CONH <sub>2</sub> ( <b>B</b> )	~17
H <sub>2</sub> N-(D)Arg-(D)2Pyrim-(D)Lys-CONH <sub>2</sub> ( <b>C</b> )	4-5
H <sub>2</sub> N-(D)Lys-(D)2Pyrim-(D)Arg-CONH <sub>2</sub> ( <b>D</b> )	~26
H <sub>2</sub> N-(D)Lys-(D)2Pyrim-(D)Lys-CONH <sub>2</sub> ( <b>G</b> )	50-55

#### References:

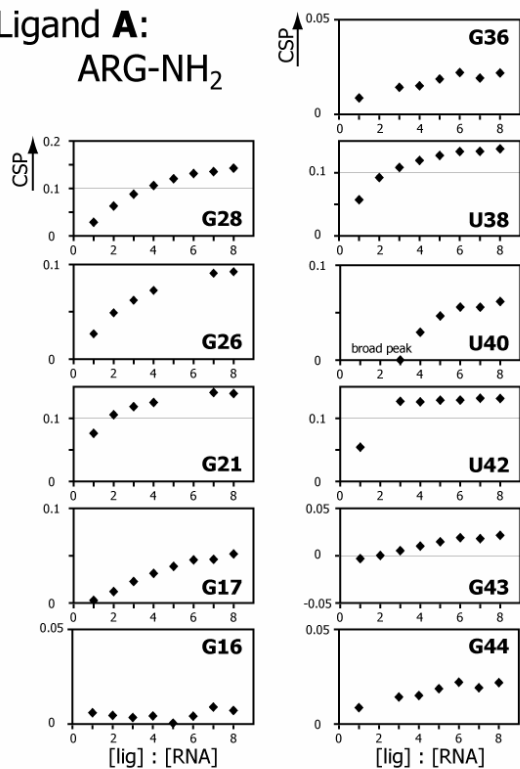
- [9d] B. Davis, M. Afshar, G. Varani, A. I. H. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A. J. Potter, I. D. Starkey, T. Swarbrick, F. Aboul-ela, *J. Mol. Biol.* **2004**, 336, 343-356.
- [9e] A. I. H. Murchie, B. Davis, C. Isel, M. Afshar, M. J. Drysdale, J. Bower, A. J. Potter, I. D. Starkey, T. M. Swarbrick, S. Mirza, C. D. Prescott, P. Vaglio, F. Aboul-ela, J. Karn, *J. Mol. Biol.* **2004**, 336, 625-638.
- [31] A. T. Brünger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr D Biol Crystallogr.* **1998**, 54, 905-921.
- [S1] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenback, *Nucleic Acids Res.* **1987**, 15, 8783-8798.
- [S2] A. Krebs, V. Ludwig, J. Pfizer, G. Dürner, M.W. Göbel, *Chem. Eur. J.* **2004**, 10, 544–553.
- [S3] P. Plateau, M. Guéron, *J. Am. Chem. Soc.* **1982**, 104, 7310-7311.
- [S4] B. Fürtig, C. Richter, J. Wöhnert, H. Schwalbe, *ChemBioChem* **2003**, 4, 936-962.



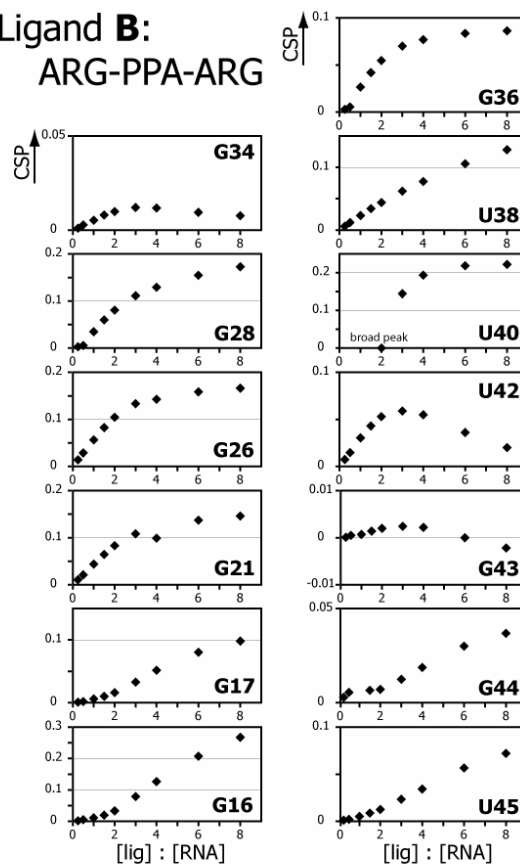
## Iminoshifts of HIV-2 TAR

Ligand A:

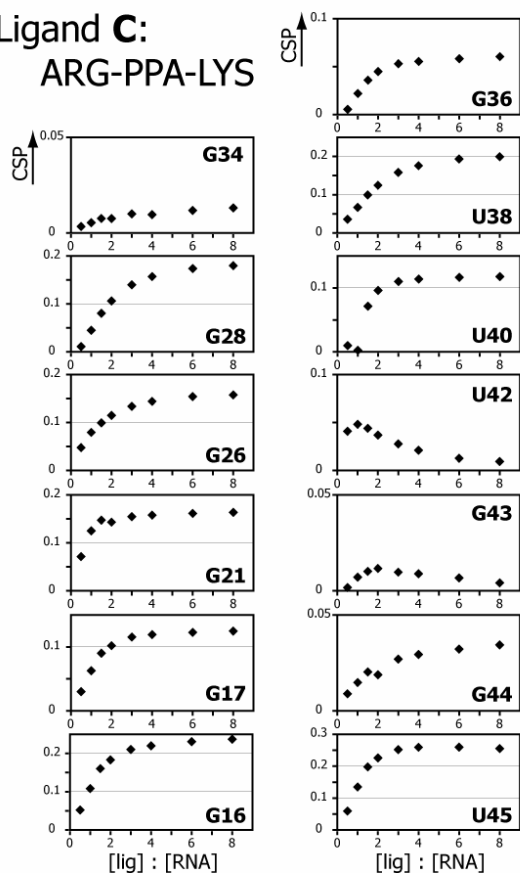
ARG-NH<sub>2</sub>



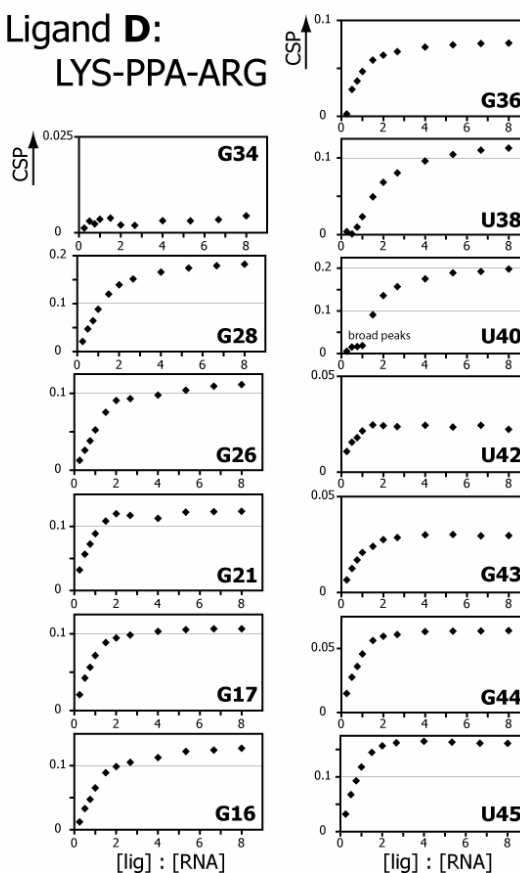
Ligand B:  
ARG-PPA-ARG



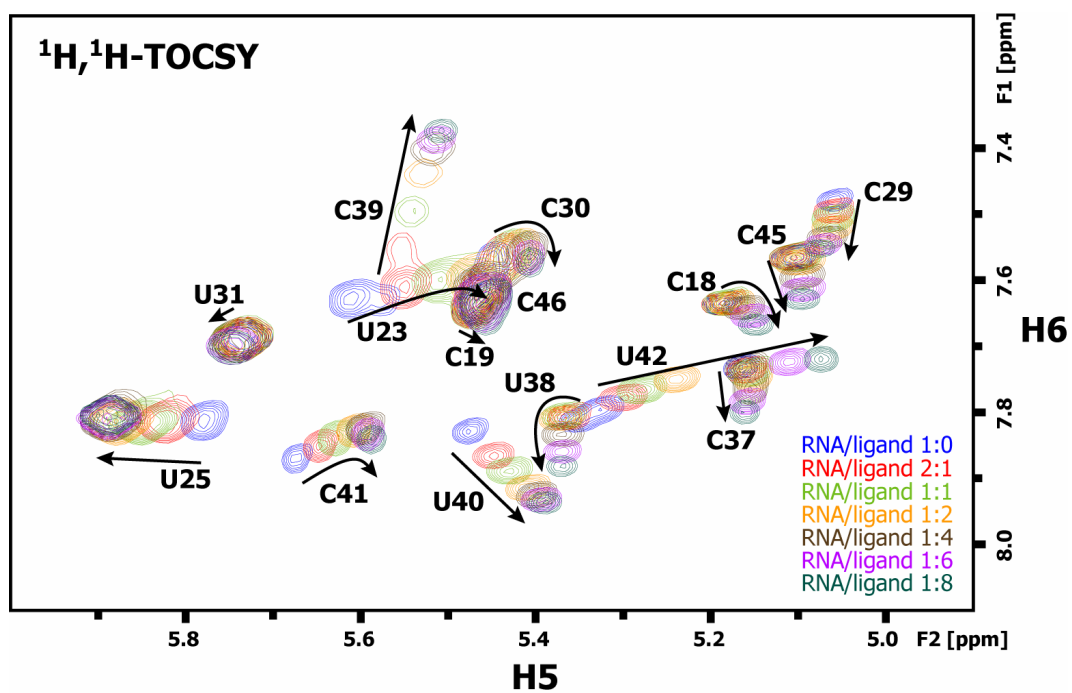
Ligand C:  
ARG-PPA-LYS



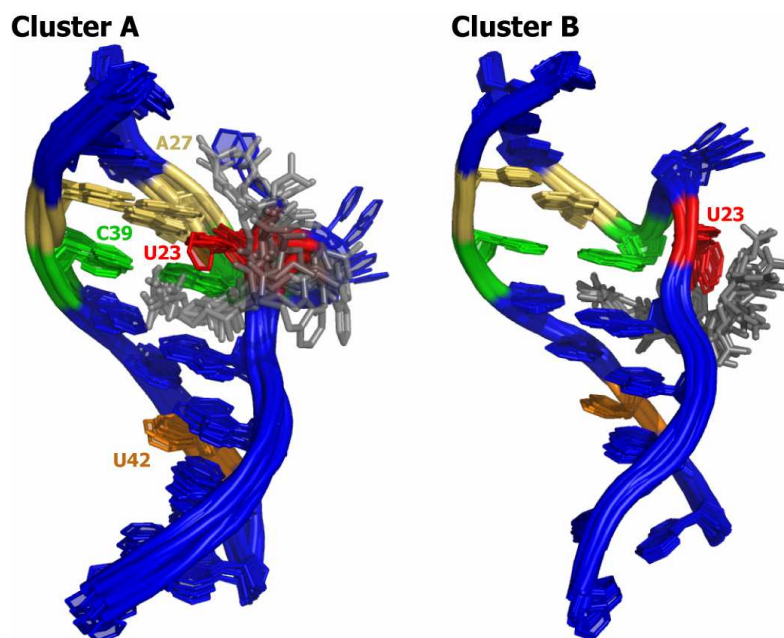
Ligand D:  
LYS-PPA-ARG



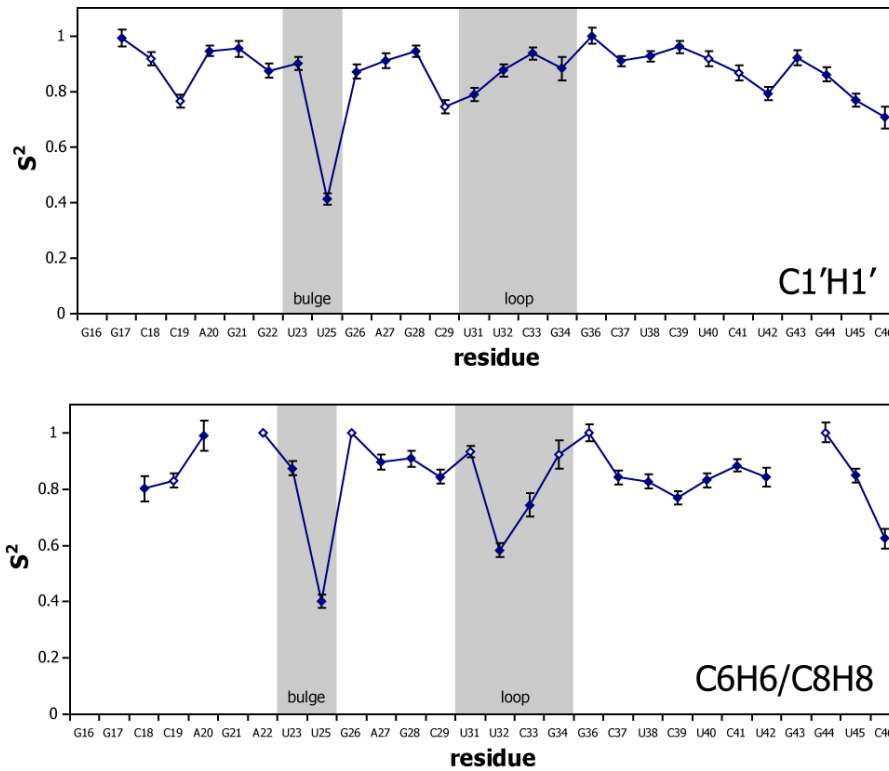
**Figure S1:** Chemical shift perturbations (CSP) as a function of the [ligand]:[RNA]-ratio for the ligands argininamide (A), the pyrimidinyl-peptide (B) and the lysine-variants (C and D).



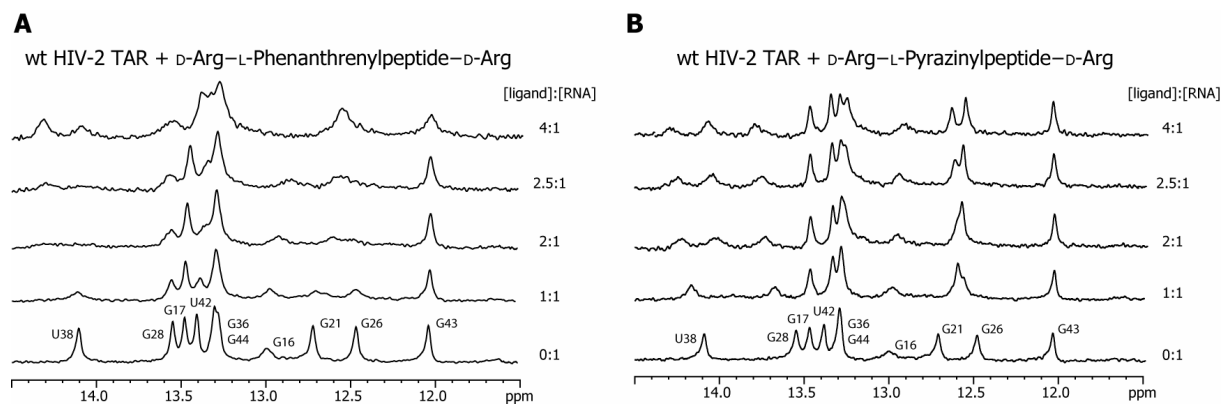
**Figure S2:**  $^1\text{H}, ^1\text{H}$ -TOCSY spectra of the H5/H6-crosspeaks from pyrimidines in HIV-2 TAR (wt) at different [ligand B]:[RNA]-ratios



**Figure S3:** Structures of the two clusters resulting from a 1:1-complex of HIV-2 TAR and ligand B.



**Figure S4:** Order parameters ( $S^2$ ) calculated from carbon relaxation rates ( $R_1$ ,  $R_2$ , *hetNOE*) for the vectors C1'H1' (top) reporting on dynamics of the ribose moiety and the vectors C6H6 and C8H8 (bottom) reporting on dynamics of the pyrimidine and purine nucleobases, respectively. The bulge and loop regions are shaded gray.



**Figure S5:** 1D NMR spectra of the wt TAR (HIV-2) imino protons at different [ligand]:[RNA]-ratios with the phenanthrenyl-peptide (A) and the pyrazinyl-peptide (B) as ligands.



## CHAPTER VII

**FAST DYNAMICS IN RNA TETRALOOPS**

**Research Article:** Quantitative  $\Gamma$ -HCNCH: Determination of the Glycosidic Torsion Angle  $\chi$  in RNA Oligonucleotides from the Analysis of CH Dipolar Cross-Correlated Relaxation by Solution NMR Spectroscopy

Jörg Rinnenthal, Christian Richter, Jan Ferner, Elke Duchardt, and Harald Schwalbe

*J. Biomol. NMR* **2007**, 39, 17-29

This article describes a novel liquid state NMR method for the determination of the glycosidic torsion angle  $\chi$  in nucleic acids by analysis of cross-correlated relaxation rates. This method facilitates to distinguish between a *syn*- and an *anti*-conformation for all four nucleotides, even in larger nucleic acid molecules. For the parameterization, the determination of order parameters derived from autocorrelated relaxation rates and a MD simulation were necessary.

This publication is based on work exclusively carried out in the group of Prof. Schwalbe. The author of this thesis provided the order parameters for the two investigated RNA molecules.



# Quantitative $\Gamma$ -HCNCH: determination of the glycosidic torsion angle $\chi$ in RNA oligonucleotides from the analysis of CH dipolar cross-correlated relaxation by solution NMR spectroscopy

Jörg Rinnenthal · Christian Richter ·  
Jan Ferner · Elke Duchardt · Harald Schwalbe

Received: 5 March 2007 / Accepted: 22 May 2007 / Published online: 20 July 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** A novel NMR pulse sequence is introduced to determine the glycosidic torsion angle  $\chi$  in  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled oligonucleotides. The quantitative  $\Gamma$ -HCNCH measures the dipolar cross-correlated relaxation rates  $\Gamma_{\text{C6H6,C1'H1'}}^{\text{DD,DD}}$  (pyrimidines) and  $\Gamma_{\text{C8H8,C1'H1'}}^{\text{DD,DD}}$  (purines). Cross-correlated relaxation rates of a  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled RNA 14mer containing a cUUCGg tetraloop were determined and yielded  $\chi$ -angles that agreed remarkably well with data derived from the X-ray structure of the tetraloop. In addition, the method was applied to the larger stemloop D (SLD) subdomain of the Coxsackievirus B3 cloverleaf 30mer RNA and the effect of anisotropic rotational motion was examined for this molecule. It could be shown that the  $\chi$ -angle determination especially for nucleotides in the anti conformation was very accurate and the method was ideally suited to distinguish between the syn and the anti-conformation of all four types of nucleotides.

**Keywords** NMR spectroscopy · Isotope labeled RNA ·  $\chi$ -angle determination · Cross-correlated relaxation

## Introduction

The availability of isotope labeled RNA oligonucleotides (Batey et al. 1992; Nikonowicz et al. 1992; Quant et al. 1994; Batey et al. 1995) has stimulated the development of a large number of different heteronuclear NMR experiments (Varani and Tinoco 1991; Varani et al. 1996; Wijmenga 1998; Cromsigt et al. 2001; Fürtig et al. 2003) and allowed de-novo structure determination of sizeable RNA molecules by solution-state NMR spectroscopy. In comparison to the structure determination of proteins, where NOEs are the most important source of information, additional angular information is needed for RNA because the proton density is much lower than in proteins. Direct angular information can be obtained either from the analysis of homo- and heteronuclear  $^3J$  coupling constants or from the interpretation of cross-correlated relaxation rates. Among the torsion angles of interest, the  $\chi$ -angle determines the relative orientation of the nucleobase with respect to the ribose moiety, linking the nucleobases to the backbone scaffold of the RNA. The nucleobase can assume either the syn ( $\chi \approx 40^\circ$ – $80^\circ$ ) or the anti ( $\chi \approx 180^\circ$ – $240^\circ$ ) conformation. The analysis of populated torsion angles in RNA deposited in the data base shows a spread in  $\chi$  of around  $50^\circ$  for purine and of around  $30^\circ$  for pyrimidine nucleotides. (see Fig. 2)

Using NMR-spectroscopy, the  $\chi$ -angle can be determined by measurement of the  $^3J(\text{C2/4,H1}')$  and the  $^3J(\text{C6/8,H1}')$  coupling constants (Schwalbe et al. 1994; Trantirek et al. 2002; Munzarova and Sklenar 2003), using a Karplus parametrization. However, the accuracy of the angle determination depends on the quality of these Karplus parametrizations. The validity of the available parametrizations for RNA still remains to be investigated (Duchardt et al. 2004). More recently, cross-correlated relaxation  $\Gamma$ -rates were exploited to obtain structural

**Electronic Supplementary Material** The online version of this article (doi:10.1007/s10858-007-9167-5) contains supplementary material, which is available to authorized users.

J. Rinnenthal · C. Richter · J. Ferner · E. Duchardt ·  
H. Schwalbe (✉)  
Institute for Organic Chemistry and Chemical Biology, Center  
for Biomolecular Magnetic Resonance, Johann Wolfgang  
Goethe-University, Max-von-Laue-Strasse 7, 60438  
Frankfurt/Main, Germany  
e-mail: schwalbe@nmr.uni-frankfurt.de

information in proteins (Reif et al. 1997; Pelupessy et al. 1999; Carlomagno et al. 2001; Schwalbe et al. 2001; Boisbouvier and Bax 2002; Kloiber et al. 2002; Zwahlen and Vincent 2002) and in oligosaccharides (Ilin et al. 2003). In oligonucleotides, the sugar pucker (Carlomagno et al. 1999; Felli et al. 1999; Richter et al. 1999), the angles of the sugar-phosphate backbone (Richter et al. 2000) and the  $\chi$ -angle (Duchardt et al. 2004; Sychrovsky et al. 2005) can be determined from cross-correlated relaxation rates. In addition, if the structural parameters are known, cross-correlated relaxation  $\Gamma$ -rates can be used to analyze the dynamics of a biomacromolecule (Banci et al. 2001; Ravindranathan et al. 2003; Tugarinov and Kay 2004; Markwick et al. 2005; Wang et al. 2005).

Previously, we have reported the determination of the  $\chi$ -angle from the measurement of nitrogen chemical shift anisotropy (N-CSA), C–H dipolar cross-correlated relaxation rates ( $\Gamma_{N1,C1'H1'}^{CSA,DD}$  for pyrimidines and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  for purines) (Duchardt et al. 2004). The method requires the knowledge of the size and the orientation of nitrogen (N1/N9) CSA tensors. These tensors have been determined for desoxymononucleotides by solid state NMR (Stueber and Grant 2002). However, the structural interpretation of cross-correlated relaxation rates that involve the chemical shift anisotropy as one relaxation mechanism is difficult, since little is known about the dependence of the CS tensors on either the sugar pucker mode or the  $\chi$ -angle itself. Recent quantum mechanical calculations have indicated that the nitrogen chemical shift anisotropy can in fact be dependent on the  $\chi$ -angle as well as on the conformation of the ribose moiety (Sychrovsky et al. 2005).

In this report, we therefore propose a new, CSA-independent method to determine the  $\chi$ -angle based on the measurement of the dipolar cross-correlated relaxation rates  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  in a novel NMR experiment named the quantitative  $\Gamma$ -HCNCH. The method is applied to a 14mer cUUCGg RNA tetraloop and compared to the  $\chi$ -angles derived from the previous analysis of the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$ ,  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  and the X-ray structure (Ennifar et al. 2000; Duchardt et al. 2004). Furthermore, we demonstrate the applicability of the novel method for a larger RNA, namely the stemloop D (SLD) subdomain of the Coxsackievirus B3 cloverleaf 30mer RNA (Ohlenschläger et al. 2004). In addition, we provide an easy to use  $\Gamma^{DD,DD}(\chi)$  parametrization in order to facilitate the  $\chi$ -angle determination.

## Materials and methods

### NMR spectroscopy

The measurements were performed using a uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled 14mer cUUCGg RNA tetraloop sample with

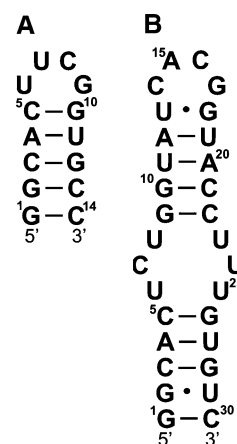
the sequence 5'-PO<sub>4</sub><sup>2-</sup>-PO<sub>3</sub><sup>-</sup>-PO<sub>3</sub><sup>-</sup>-GGCAC(UUCG)GUGC C-OH-3'. The secondary structure of the molecule is shown in Fig. 1A. The relevant chemical shifts for the resonance assignment were taken from Fürtig et al. (2004). The RNA was purchased from Silantes (Munich, Germany). The sample contained 0.7 mM RNA in 20 mM potassium phosphate, pH 6.4, 0.4 mM EDTA and 10% D<sub>2</sub>O.

Additional measurements were performed on an adenosine, uridine- $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled 30mer RNA (SLD of the CVB3 genomic RNA). The secondary structure of the molecule is shown in Fig. 1B (Ohlenschläger et al. 2004). The sample contained 1.2 mM RNA in 40 mM potassium phosphate, pH 6.2, 0.2 mM EDTA and 10% D<sub>2</sub>O.

The NMR measurements were carried out on a Bruker 700 MHz spectrometer equipped with a 5 mm  $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$   $z$ -axis-gradient cryogenic probe. Measurements were carried out at 298 K for the 14mer and 310 K for the 30mer. The data was processed and analyzed using the program TOPSPIN 1.3 (Bruker Biospin). Peak intensities were taken for the quantitative analysis of  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  according to Eq. 4. The stated error is equal to the RMSD of the multiple determined  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  rates recorded with different mixing times, different frequency labeling in  $t_1$  (N1/9 or C1') and two different pulse sequences that use signal detection either on the H1' (Fig. 4) or the H8 (Supplementary Material).

### MD simulation

A 5 ns molecular dynamics simulation of the 14mer was performed with the program CHARMM using the CHARMM27 nucleic acid force field (Foloppe and MacKerell Jr. 2000; MacKerell Jr and Banavali 2000). A recent X-ray structure (pdb ID 1F7Y) of the UUCG loop elongated by a standard A-form double strand of



**Fig. 1** (A) Secondary structure of the cUUCGg 14mer RNA. (B) Secondary structure of the uCACGg SLD 30mer RNA



appropriate sequence served as the starting structure for the simulation (Duchardt et al. 2004).

Simulations were carried out using periodic boundary conditions in a rhombic dodecahedron cut out of a cube with 50 Å side-length, filled with preequilibrated TPI3P water (Jorgensen et al. 1983). The system was neutralized with sodium counterions, which were placed manually by random replacement of water molecules. Non-bonded interactions were smoothly shifted to zero using a 12 Å cutoff. The pair-list of non-bonded interactions was generated using a 14 Å cutoff, and updated whenever any atom had moved >1 Å since the last list update. The calculation of the trajectory took place under constant pressure conditions using the Langevin piston method (Feller et al. 1995) with a check every 4 ps that the average temperature remained within ±5 K of 298 K. Conformations of the system were saved every 0.2 ps for further analysis.  $\chi$ -angle extraction was performed using CHARMM input scripts. The first 500 ps of the trajectory were not considered in the analysis.

### Theory

For a RNA macromolecule that shows isotropic rotational tumbling in solution, the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  cross-correlated relaxation rates are given by

$$\begin{aligned} \Gamma_{C6H6,C1'H1'}^{DD,DD} &\equiv \frac{1}{5} \frac{\gamma_H^2 \gamma_C^2}{r_{C6H6}^3 r_{C1'H1'}^3} \left(\frac{\mu_0}{4\pi}\right)^2 \hbar^2 \\ &\times \left(S_{C6H6,C1'H1'}^{DD,DD}\right)^2 (3 \cos^2 \theta_{C6H6,C1'H1'} - 1) \tau_c \\ \Gamma_{C8H8,C1'H1'}^{DD,DD} &\equiv \frac{1}{5} \frac{\gamma_H^2 \gamma_C^2}{r_{C8H8}^3 r_{C1'H1'}^3} \left(\frac{\mu_0}{4\pi}\right)^2 \hbar^2 \\ &\times \left(S_{C8H8,C1'H1'}^{DD,DD}\right)^2 (3 \cos^2 \theta_{C8H8,C1'H1'} - 1) \tau_c \end{aligned} \tag{1}$$

where  $\gamma_H$  and  $\gamma_C$  are the gyromagnetic ratios for the hydrogen and carbon nuclei,  $r_{C6H6(C8H8)}$  and  $r_{C1'H1'}$  are the lengths of the  $\overrightarrow{C6H6(C8H8)}$  and  $\overrightarrow{C1'H1'}$  bond vectors (Fig. 3C) which are set to 1.104 and 1.090 Å, respectively (Ying et al. 2006),  $\mu_0$  is the magnetic susceptibility of the vacuum,  $\tau_c$  is the isotropic rotational correlation time of the molecule,  $\hbar$  is the Planck constant divided by  $2\pi$ ,  $S_{C6H6,C1'H1'}^{DD,DD}$  ( $S_{C8H8,C1'H1'}^{DD,DD}$ ) are the cross-correlated order parameters for internal motion and  $\theta_{C6H6,C1'H1'}$  ( $\theta_{C8H8,C1'H1'}$ ) are the projection angles between the  $\overrightarrow{C6H6(C8H8)}$  and the  $\overrightarrow{C1'H1'}$  bond vectors in pyrimidine and purine nucleotides, respectively. The projection angle  $\theta_{C6H6,C1'H1'}$  ( $\theta_{C8H8,C1'H1'}$ ) is a function of the glycosidic torsion angle  $\chi$ . It follows that also  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  can be related to torsion angle  $\chi$  by

application of basic geometric considerations. The derivation of this relationship is given in the supplementary material (S2). From this, a parametrization similar to the Karplus equations for  $J$  couplings can be deduced also for  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$ . This parametrization is only dependent on  $\chi$ , the order parameter  $(S_{C6H6,C1'H1'}^{DD,DD})^2$  or  $(S_{C8H8,C1'H1'}^{DD,DD})^2$  and  $\tau_c$ .

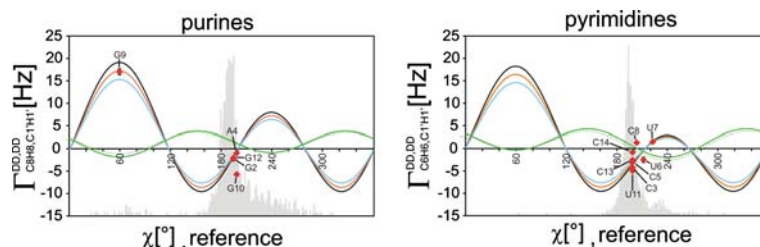
$$\begin{aligned} \Gamma_{C6H6,C1'H1'}^{DD,DD} &= 4.1370 \tau_c \left(S_{C6H6,C1'H1'}^{DD,DD}\right)^2 10^9 [A \cos(\chi - 59.01) \\ &\quad + B \cos(2\chi - 118.01) + C] \tag{2} \\ \Gamma_{C8H8,C1'H1'}^{DD,DD} &= 4.1370 \tau_c \left(S_{C8H8,C1'H1'}^{DD,DD}\right)^2 10^9 [A \cos(\chi - 59.01) \\ &\quad + B \cos(2\chi - 118.01) + C] \end{aligned}$$

The parameters  $A$ ,  $B$  and  $C$  are summarized in Table 1 for purines and pyrimidines and are shown in Fig. 2 for the 14mer RNA assuming an isotropic overall rotational correlation time  $\tau_c$  equal to 2.31 ns (Duchardt and Schwalbe 2005). Since the parametrizations for the nucleobases C and U are almost identical to each other and also the parametrizations for the nucleobases G and A, it makes sense to summarize the different parametrization curves for the four different kinds of nucleotides to two different parametrization curves distinguishing only between purines and pyrimidines. The exact parametrizations for the nucleobases C, U, A and G are given in the supplementary material (Table S1). For the cross-correlated order parameters  $(S_{C8H8,C1'H1'}^{DD,DD})^2$  and  $(S_{C6H6,C1'H1'}^{DD,DD})^2$  (Table 2) we used a value derived from the autocorrelated order parameters derived from  $^{13}C$  relaxation analysis (Duchardt and Schwalbe 2005) as approximation for the cross-correlated order parameters. As pointed out by Bodenhausen and coworkers, this approach is valid if one assumes fast uncorrelated motion of the  $\overrightarrow{C6H6(C8H8)}$  bond vectors with respect to the  $\overrightarrow{C1'H1'}$  bond vectors (Vugmeyster et al. 2004).

For the cUUCGg 14mer RNA the assumption of isotropic rotational diffusion is valid since the rotational diffusion anisotropy  $A$  is small ( $A = 1.35$ ) (Duchardt and Schwalbe 2005). For molecules that have a rather extended shape, as it is the case for the SLD 30mer RNA (Fig. 1B), the effect of anisotropic rotational tumbling cannot be neglected. Therefore, we analyzed the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  rates by taking the size and the orientation of the axially symmetric rotational diffusion tensor into

**Table 1** Parameters for the calculation of the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  according to equation (2)

	$A$	$B$	$C$
Purines	0.583	1.191	0.226
Pyrimidines	0.801	1.015	0.093



**Fig. 2**  $\Gamma_{C8H8,C1'H1'}^{DD,DD}(\chi)$  for purines and  $\Gamma_{C6H6,C1'H1'}^{DD,DD}(\chi)$  for pyrimidines, curves are shown for  $(S_{C8H8,C1'H1'}^{DD,DD})^2$  and  $(S_{C6H6,C1'H1'}^{DD,DD})^2$  equal to 0.8 (light blue), 0.9 (orange) and 1.0 (black).  $\tau_c$  was set to 2.31 ns (Duchardt and Schwalbe 2005) for the 14mer at 298 K. In addition, the experimental  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  from the 14mer RNA are plotted against the reference  $\chi$ -angles (Duchardt et al. 2004) (red

squares).  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  for pyrimidines and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  for purines are illustrated as green solid (guanine, cytosine) and dashed (adenine, uracil) lines, respectively. The gray bars represent the  $\chi$ -angle distribution in the RNA fraction of the large ribosomal subunit (PDB entry 1FFK (Ban et al. 2000))

**Table 2**  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$ ,  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  and the corresponding  $\chi$ -angles in comparison to the reference (crystal structure), the  $\chi$ -angles extracted from the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$ ,  $\Gamma_{N9,C1'H1'}^{CSA,DD}$ , and the  $^3J(C,H)$ -coupling constants (Duchardt et al. 2004)

Nucleotide	(Hz)		$\chi$ [°]	$\chi$ [°]	$\chi$ [°]	$\chi$ [°]	$\chi$ [°]
	$\Gamma_{C6H6,C1'H1'}^{DD,DD}$	$(S_{C6H6,C1'H1'}^{DD,DD})^2$	$\Gamma_{C6H6,C1'H1'}^{DD,DD}$	(MD simulation)	$^3J(C,H)$	$\Gamma_{N1,C1'H1'}^{CSA,DD}$	(X-ray)
	$\Gamma_{C8H8,C1'H1'}^{DD,DD}$	$(S_{C8H8,C1'H1'}^{DD,DD})^2$	$\Gamma_{C8H8,C1'H1'}^{DD,DD}$			$\Gamma_{N9,C1'H1'}^{CSA,DD}$	
G1							194.8 ± 4.4
G2	-2.3 ± 0.4	0.959	192.4 ± 1.4	197.5 ± 7.8	192 ± 2	205 ± 1	194.8 ± 4.4
C3	-4.3 ± 0.1	0.926	193.7 ± 0.4	202.4 ± 7.4		203 ± 1	198.4 ± 3.1
A4	-0.9 ± 0.1	0.968	196.3 ± 0.4	198.9 ± 7.7	198.5 ± 1.5	207 ± 1	198.2 ± 3.5
C5	-3.1 ± 0.3	0.920	198.8 ± 1.3	196.7 ± 7.3	191 ± 8	201 ± 1	197.85
U6	-2.5 ± 0.6	0.935	201.2 ± 2.5	204.4 ± 7.5	196.5 ± 4.5	221	211.03
U7	1.5 ± 0.2	0.777	223.6 ± 2.1	237.2 ± 17.6	217.5 ± 9.5	209 ± 4	222.23
C8	1.3 ± 0.2	0.880	221.6 ± 1.5	204.6 ± 10.2	210.5 ± 6.5	218 ± 2	202.64
G9	17.0 ± 0.6	0.932	48.6 ± 5.2	27.3 ± 7.1	44 ± 4	57 ± 19	56.66
(G9)			69.4 ± 5.2				
G10	-5.7 ± 0.4	0.943	180.0 ± 1.6	191.2 ± 7.7	215.5 ± 13.5	201 ± 2	198.24
U11	-5.0 ± 0.3	0.952	190.9 ± 1.4	199.1 ± 7.8		195	198.2 ± 3.0
G12	-2.1 ± 0.4	0.952	193.0 ± 1.1	188.1 ± 7.9	192 ± 4	207 ± 1	194.8 ± 4.4
C13	-2.6 ± 0.2	0.938	201.4 ± 0.9	200.3 ± 7.2	189.5 ± 3.5	202 ± 1	198.4 ± 3.1
C14	-0.9 ± 0.3	0.867	208.8 ± 1.3	208.9 ± 11.7	198 ± 5	211 ± 1	198.4 ± 4.4

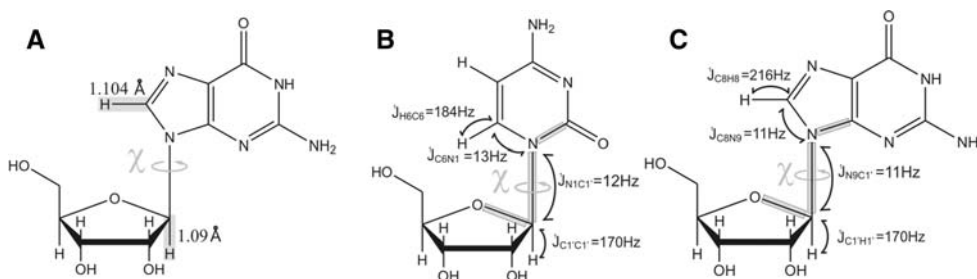
The reference angles for the nucleotides C5–G10 are extracted from the crystal structure (pdb ID 1F7Y) (Ennifar et al. 2000), the other reference values are mean values for the A-Form RNA obtained as described previously (Duchardt et al. 2004). The  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  were determined several times with different mixing times  $T_M$  (20, 30 ms).  $(S_{C6H6,C1'H1'}^{DD,DD})^2$  and  $(S_{C8H8,C1'H1'}^{DD,DD})^2$  were estimated from the autocorrelated order parameters (Ferner and Schwalbe) as a product of  $\sqrt{(\text{auto}S_{C6H6,C6H6})^2} * \sqrt{(\text{auto}S_{C1'H1',C1'H1'})^2}$  and  $\sqrt{(\text{auto}S_{C8H8,C8H8})^2} * \sqrt{(\text{auto}S_{C1'H1',C1'H1'})^2}$  respectively.  $(S_{C6H6,C1'H1'}^{DD,DD})^2$  and  $(S_{C8H8,C1'H1'}^{DD,DD})^2$  were used for determination of the  $\chi$ -angles. There is only one remaining possibility for the  $\chi$ -angle lying close to the reference value (Ennifar et al. 2000; Duchardt et al. 2004) for each nucleotide in the anti conformation. In contrast, two values for the  $\chi$ -angle remain possible for the nucleotide G9 which assumes the syn conformation due to degeneracy of the parametrization curve

account. The mathematical relations for the anisotropic case are given in section S4 of the supplementary material.

#### Pulse sequence

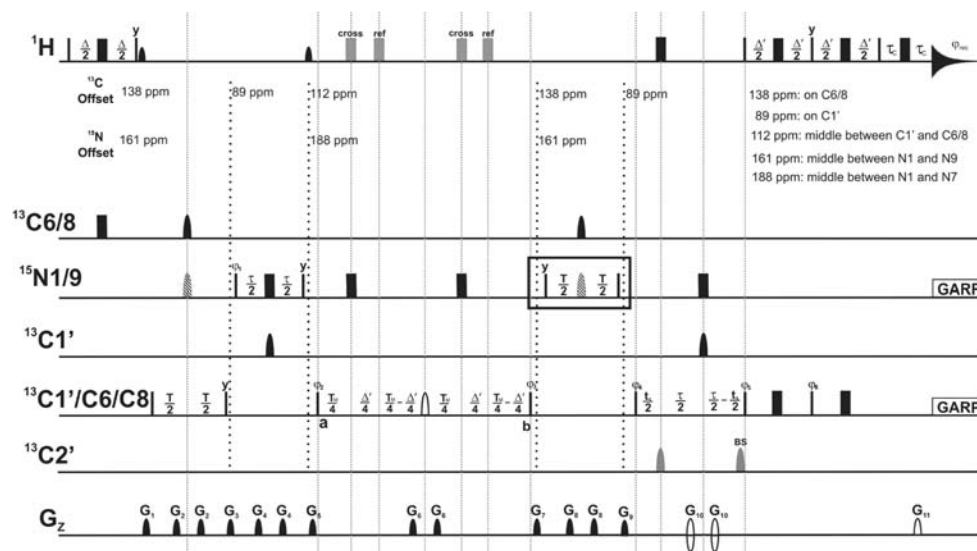
The quantitative  $\Gamma$ -HCNCH experiment (Fig. 4) is derived from the HCN and HCNCH experiments (Sklenar et al.

1993a, b; Fiala et al. 1998; Sklenar et al. 1998). It is designed to transfer coherence in a directed manner from H6 (H8) to H1', where the signal is detected. The scheme is similar to that in the quantitative  $\Gamma$ -HCCH experiment (Felli et al. 1999; Millet et al. 1999). The coherence at time point **a** is created via three successive INEPT steps transferring the coherence from H6 (H8) to C1'. During the



**Fig. 3** In (A), the bond vectors relevant for the derivation of projection angles from  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  are marked in gray. (B) and (C) show the magnitude of the experimentally determined  $^1J$  scalar couplings used to transfer the magnetization from H6(8) to H1' (Fiala

et al. 2004) in the case of pyrimidines and purines in the quantitative  $\Gamma$ -HCNCH experiment. The bonds underlaid with gray define the  $\chi$ -angle



**Fig. 4** Pulse sequence of the quantitative  $\Gamma$ -HCNCH experiment. Narrow and wide filled bars correspond to rectangular  $90^\circ$  and  $180^\circ$  pulses, respectively. Selective pulses and gradients are indicated as semi-ellipses. The default pulse phase is x. The pulse sequences were optimized on a Bruker spectrometer with the Bruker typical phase settings (Roehrl et al. 2005). The reference and the cross experiment are summarized in one pulse sequence scheme. The gray wide filled bars correspond to  $180^\circ$  pulses applied on the proton channel either in the cross or the reference experiment, as indicated. Two water flipback pulses are applied after the INEPT steps before the mixing time  $T_M$  (small black semi-ellipses). Two selective off-resonant pulses on C2' during the carbon constant time chemical shift evolution are shown in gray. Fixed delays are adjusted as follows:  $\Delta' = 2.94 \text{ ms}(1/(2 \cdot J_{H1'C1'}))$ ,  $\Delta = 2.5 \text{ ms}(1/(2 \cdot J_{H6(8)C6(8)}))$ ,  $T = 32 \text{ ms}(1/(2 \cdot J_{C6(8)N1(9)}))$ ,  $\tau = 36 \text{ ms}(1/(2 \cdot J_{N1(9)C1'}))$ .  $T_M$  is the variable relaxation period. The proton carrier frequency is centered at the water frequency (4.7 ppm). The carbon and the nitrogen carrier frequencies change during the course of the experiment as indicated by vertical dashed lines. The values of the  $^{13}\text{C}$  and  $^{15}\text{N}$  offset at every time point are given, respectively. At 700 MHz, band selective pulses are set as follows:  $180^\circ$  Q3 Gaussian cascade (Emsley and Bodenhausen 1992) 2 ms (black semi-ellipse),  $180^\circ$  Q3 Gaussian cascade 1 ms (gray semi-ellipse),  $180^\circ$  Q3 Gaussian cascade 0.5 ms

(hatched semi-ellipse),  $180^\circ$  Reburp (Geen and Freeman 1991) 2.53 ms with maxima on C1' and C6/8 (open semi-ellipse),  $90^\circ$  square pulses 1 ms (small black semi-ellipse, water flipback). Asynchronous GARP decoupling (Shaka et al. 1985) is used to suppress heteronuclear scalar couplings during acquisition. The pulse field gradients of 1 ms length have a smoothed chirp amplitude (Bruker Topspin 2.0, 2006). They are applied along the z-axis and have the following strengths:  $G_1$ :40%,  $G_2$ :25%,  $G_3$ :50%,  $G_4$ :20%,  $G_5$ :40%,  $G_6$ :20%,  $G_7$ :65%,  $G_8$ :20%,  $G_9$ :50%,  $G_{10}$ :40%,  $G_{11}$ :20.1%. 100% of gradient strength corresponds to 55 Gauss/cm. Phase cycling:  $\phi_1 = x, -x$ ;  $\phi_2 = 8(x), 8(-x)$ ;  $\phi_3 = 16(y), 16(-y)$ ;  $\phi_4 = 4(y), 4(-y)$ ;  $\phi_5 = (2x), 2(-x)$ ;  $\phi_6 = -2(y), 2(y)$ ;  $\phi_{\text{rec}} = x, 2(-x), x, -x, 2(x), -x$ .  $G_{10}$  and  $\phi_6$  are modulated according to echo/antiecho modulation in the  $\omega_1$  dimension using sensitivity enhancement (Kay et al. 1992). At time point a, the coherence  $H6(8)_z C6(8)_y C1'(N1(9))_z$  is created and converts during  $T_M$  into the coherence  $H1'_z C6(8)_x C1'_x N1(9)_z$  at time point b. If additional chemical shift modulation is needed, a N1(9) constant time evolution period can be introduced during the first refocusing INEPT step as indicated by the black box. In this case,  $\phi_4$  is incremented in a States-TPPI (Marion et al. 1989) manner to achieve quadrature detection in the N1(N9) dimension

mixing time  $T_M$ , the generated coherence  $8H6(8)_z C6(8)_y C1'_y N1(9)_z$  at time point **a** evolves into the coherence  $8H1'_z C6(8)_x C1'_x N1(9)_z$  at time point **b**. Frequency labeling of  $C1'$  occurs in a constant time manner during the second refocusing INEPT step of the transfer to  $H1'$  after the mixing period  $T_M$ . If additional chemical shift resolution is needed, a constant time  $N1(9)$  chemical shift evolution period can be used during the first refocusing INEPT step alternatively to the  $C1'$  chemical shift evolution period. The coherence transfer could also be performed starting at  $H1'$ . Experimental details for the pulse sequence with reverted coherence transfer are given in the Supplementary Material (S3). The analysis of cross-correlated relaxation rates given later is based on both pulse sequences with different mixing times  $T_M$ , respectively. The stated error is the RMSD calculated from six independent measurements with both pulse sequences using different mixing times  $T_M$  (supplementary material section S1, Fig. S1). The various INEPT delays for magnetization transfer are optimized taking into account the relevant coupling constants which are depicted in Fig. 3.

Two different experiments are recorded, a cross and a reference experiment. In the cross experiment, the coherence selected at time point **b** is  $\sinh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M)$  (or  $\sinh(\Gamma_{C8H8,C1'H1'}^{DD,DD} T_M)$ ) modulated. In the reference experiment, the selected coherence is modulated by  $\cosh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M) \sin(\pi J_{C6H6} \Delta')$  (or  $\cosh(\Gamma_{C8H8,C1'H1'}^{DD,DD} T_M) \sin(\pi J_{C8H8} \Delta')$  for purines).

The measured cross-correlation effect increases at longer mixing times  $T_M$  but is also counteracted by auto-correlated relaxation effects. The optimal sensitivity in the cross experiment was achieved using a mixing time  $T_M$  in the range of 20–30 ms for the 14mer RNA and between 10 ms and 15 ms for the 30mer RNA.

In the case of pyrimidines, the evolution of the excited double and zero quantum coherence  $8H6_z C6_y C1'_y N1_z$  during  $T_M$  is given by

$$\begin{aligned} 8H6_z C6_y C1'_y N1_z &\rightarrow 8H6_z C6_y C1'_y N1_z [\cosh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M) \\ &\cos(\pi J_{C6H6} \Delta') \cos(\pi J_{C1'H1'} \Delta') \\ &+ \sinh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M) \sin(\pi J_{C6H6} \Delta') \sin(\pi J_{C1'H1'} \Delta')] \\ &+ 8H1'_z C6_x C1'_x N1_z [\sinh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M) \cos(\pi J_{C6H6} \Delta') \\ &\cos(\pi J_{C1'H1'} \Delta') + \cosh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M) \\ &\sin(\pi J_{C6H6} \Delta') \sin(\pi J_{C1'H1'} \Delta')] \end{aligned} \quad (3)$$

Note that in the case of purines, the nomenclature of the involved nuclei is different ( $H8 \rightarrow H6$ ,  $C8 \rightarrow C6$ ,  $N9 \rightarrow N1$ ). During the mixing time  $T_M$ , the coherence  $8H6(8)_z C6(8)_y C1'_y N1(9)_z$  converts into the coherence  $8H1'_z C6(8)_x C1'_x N1(9)_z$  that is selected at the end of the mixing time  $T_M$ ; in the reference experiment, this transfer

is accomplished via scalar coupling, in the cross experiment, coherence transfer is accomplished via cross-correlated relaxation.

In the reference experiment with  $\Delta' = 1/(2^* J_{C1'H1'})$ , the intensity of the peak  $I^{\text{ref}}$  is proportional to  $\cosh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M) \sin(\pi J_{C6H6} \Delta')$ . In the cross experiment, the  $J_{CH}$  scalar coupling is completely refocused because of  $\Delta' = 0$  and the intensity of the peak  $I^{\text{cross}}$  is modulated by  $\sinh(\Gamma_{C6H6,C1'H1'}^{DD,DD} T_M)$ . Consequently, division of the two intensities yields an expression for the cross-correlated relaxation rate given by the following equation:

$$\begin{aligned} \Gamma_{C6H6,C1'H1'}^{DD,DD} &= T_M^{-1} \tanh^{-1} \left( \frac{\text{ns}^{\text{ref}} I^{\text{cross}} \sin(\pi J_{C6H6} \Delta')}{\text{ns}^{\text{cross}} I^{\text{ref}}} \right) \\ \Gamma_{C8H8,C1'H1'}^{DD,DD} &= T_M^{-1} \tanh^{-1} \left( \frac{\text{ns}^{\text{ref}} I^{\text{cross}} \sin(\pi J_{C8H8} \Delta')}{\text{ns}^{\text{cross}} I^{\text{ref}}} \right) \end{aligned} \quad (4a)$$

In Eq. 4,  $T_M$  represents the length of the double and zero quantum coherence mixing time,  $\text{ns}^{\text{ref}}$  and  $\text{ns}^{\text{cross}}$  are the number of transients recorded in the cross and the reference experiment.  $I^{\text{cross}}$  and  $I^{\text{ref}}$  are the intensities of the peaks extracted from the cross and the reference experiment and the  $\sin(\pi J_{C6H6} \Delta')$  and  $\sin(\pi J_{C8H8} \Delta')$  terms with  $\Delta' = 2.94$  ms take the different scalar couplings in purines (216 Hz) and pyrimidines (184 Hz) into account. An error of  $\pm 5$  Hz in the  $J_{C8H8}$  of the purines leads to a relative error for the  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  of 1.3%. For purines assuming the syn conformation this error calculates to 0.26 Hz, assuming a  $\tau_c$  of 2.31 ns. For the purines in the anti conformation the error is too small to have any effect since the absolute rate is very small in these cases. An error of  $\pm 5$  Hz in the  $J_{C6H6}$  of the pyrimidines and in the  $J_{C1'H1'}$  are also too small and can therefore be neglected.

There is the need to discuss to which extent cross-correlated relaxation mechanism other than  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  ( $\Gamma_{C8H8,C1'H1'}^{DD,DD}$ ) contribute to the relaxation of the double and zero quantum coherence operators during  $T_M$ . Most other cross-correlated relaxation mechanisms create coherences with a different number of product operators or with different phases. Consequently, they are not selected in the cross experiment and are removed by both phase cycling and pulse field gradients.

Evolution of the cross-correlated relaxation rates  $\Gamma_{C6H6,C1'}^{DD,CSA}$ ,  $\Gamma_{C8H8,C1'}^{DD,CSA}$ ,  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  is suppressed by the  $\pi$ -pulses on  $^1\text{H}$  in the cross experiment. However, for the time the scalar coupling evolves (2.94 ms) in the reference experiment these rates evolve, too. Of course, the desired rates  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  evolve over the entire mixing time, while the  $\Gamma_{C6H6,C1'}^{DD,CSA}$ ,  $\Gamma_{C8H8,C1'}^{DD,CSA}$ ,  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  rates only evolve for 2.94 ms, no matter how long the mixing time is. Therefore the “apparent”  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  rates decrease with longer mixing times  $T_M$ .

The intensity terms of the reference experiment enter into the calculation (Eq. 4b) as cosh-modulated terms. We simulated the curves for the  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  and calculated the error for the syn and anti conformations separately. The maximum values are  $\Gamma_{C1'H1',C6}^{DD,CSA} = 13.4$  Hz and  $\Gamma_{C1'H1',C8}^{DD,CSA} = 11.0$  Hz for nucleotides in the syn conformation. In the anti conformation, these rates are smaller and depend on the  $\chi$ -angle. For the  $\Gamma_{C6H6,C1'}^{DD,CSA}$  and  $\Gamma_{C8H8,C1'}^{DD,CSA}$  rates we took the maximum possible value to estimate the error caused by these rates.  $\Gamma_{C6H6,C1'}^{DD,CSA}$  and  $\Gamma_{C8H8,C1'}^{DD,CSA}$  can assume a maximum value of 3.0 Hz. Since the CSA of the C1' is quite small, this error is negligible small.

Taking the  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  rates into account, this Eq. 4 can be rewritten as:

$$\frac{\sinh(\Gamma_{C6H6,C1'H1'}^{DD,DD} * T_M)}{\cosh((\Gamma_{C6H6,C1'H1'}^{DD,DD} * T_M + \Gamma_{C1'H1',C6}^{DD,CSA} * 2.94 \text{ ms}))} = \left( \frac{n_s^{ref} I^{cross} \sin(\pi J_{C6H6} \Delta')}{n_s^{cross} I^{ref}} \right)$$

$$\frac{\sinh(\Gamma_{C8H8,C1'H1'}^{DD,DD} * T_M)}{\cosh((\Gamma_{C8H8,C1'H1'}^{DD,DD} * T_M + \Gamma_{C1'H1',C8}^{DD,CSA} * 2.94 \text{ ms}))} = \left( \frac{n_s^{ref} I^{cross} \sin(\pi J_{C8H8} \Delta')}{n_s^{cross} I^{ref}} \right) \tag{4b}$$

Equation 4b and simulated  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  rates are the basis for the error calculation. As long as the argument of the cosh function is very small (<0.1) the error is also small since the slope of the cosh function is very small for small arguments.

Because the  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  rates influence the cosh-modulated term of the reference experiment and only evolve for 2.94 ms, the error depends on the mixing time. For a mixing time of 30 ms, the maximum error caused by the  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  rates calculates to 2.2%. This is equal to 0.4 Hz for the G9 residue of the 14mer which assumes syn conformation around  $\chi$ . The error caused by the  $\Gamma_{C1'H1',C6}^{DD,CSA}$  and  $\Gamma_{C1'H1',C8}^{DD,CSA}$  rates for nucleotides assuming the anti conformation is negligible small (<0.03 Hz for the 14mer RNA).

## Results and discussion

### cUUCGg 14mer RNA

The quantitative  $\Gamma$ -HCNCH experiment for the determination of the glycosidic torsion angle  $\chi$  was applied to a 14mer cUUCGg tetraloop RNA (Fig. 1A). For this tetraloop-hairpin, chemical shifts of almost all the nuclei are known (Fürtig et al. 2004). The RNA consists of a stem region with five canonical base pairs and a structured loop region with one nucleotide (G9) assuming the syn conformation (Allain and Varani 1995; Ennifar et al. 2000). The molecule is soluble and monomeric at high concentrations. In addition, the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  rates have been measured before to determine the  $\chi$ -angle for this

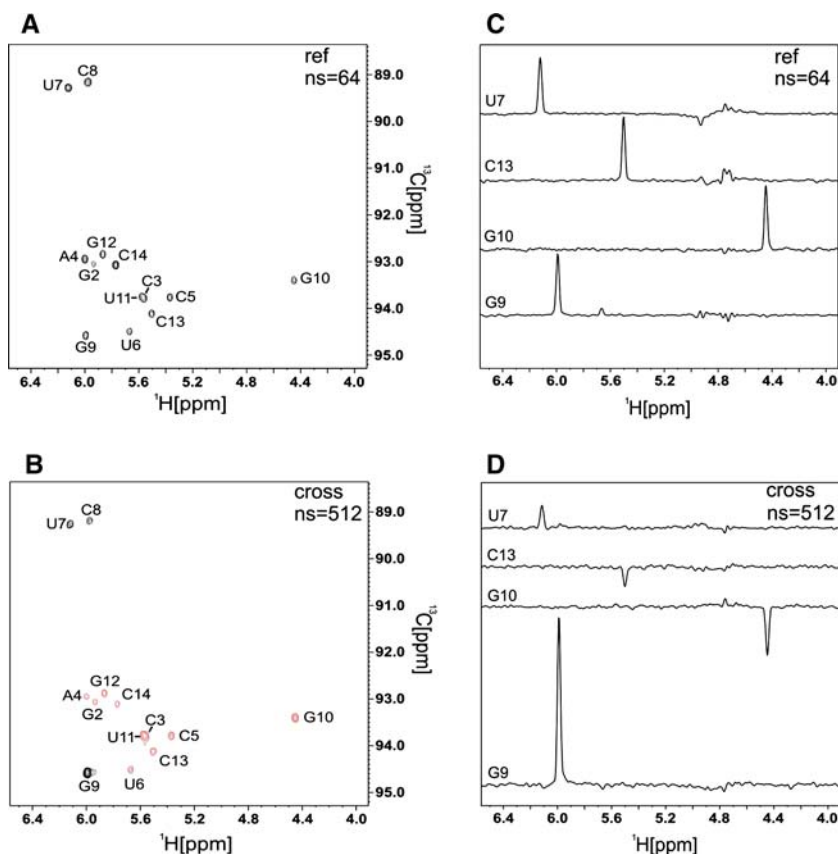
molecule (Duchardt et al. 2004) and the results presented here can be directly compared to these data.

There are two structures of cUUCGg tetraloops published so far, one crystal structure (pdb ID 1F7Y (Ennifar et al. 2000)) and one NMR ensemble (pdb ID 1HLX (Allain and Varani 1995)). The results presented here are compared to the values for  $\chi$  extracted from the crystal structure, to the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$ -rates and to  $^3J(C,H)$  coupling constants determined previously (Duchardt et al. 2004).

Figure 2 shows the dependencies  $\Gamma_{C6H6,C1'H1'}^{DD,DD}(\chi)$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}(\chi)$ . These curves demonstrate that an unambiguous distinction between the syn and the anti conformation is possible for all four types of nucleotides. In addition, the slope of the curve around the anti conformation is steep and accompanied with a sign change. Since the accuracy of the angle determination is determined by the slope of the parametrization curve, the angle determination is very precise for nucleotides assuming the anti conformation. Figure 2 also shows the previously introduced  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$ , these are much smaller, but show a different absolute sign in the interesting conformational regions the angle  $\chi$  commonly adopts.

The spectra of the quantitative  $\Gamma$ -HCNCH reference and cross experiment are shown in Fig. 5.  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  are obtained from the intensities of the peaks from the cross and the reference spectra according to Eq. 4. The rates were determined several times at different mixing times  $T_M$  (Table 2). The margin of error for rates obtained from multiple experiments lies between 0.1 Hz and 0.6 Hz (Table 2), which is very accurate considering the range of values that can be observed for  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  of up to 19 Hz. The correlation of  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  with the  $\chi$ -angles extracted from the X-ray structure (pdb ID 1F7Y) (Ennifar et al. 2000) is shown in Fig. 2. It can be seen that the experimental  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  rates are in remarkable agreement with the predicted rates. The determined  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  rates were translated into  $\chi$ -angles using the theoretically derived  $\Gamma_{C6H6,C1'H1'}^{DD,DD}(\chi)$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}(\chi)$  dependence (Eq. 2, Table 1). The correlation of the  $\chi$ -angles obtained from the X-ray structure (pdb ID 1F7Y) (Ennifar et al. 2000) and the analyzed  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  (Table 2, Fig. 6A) yields an RMSD of 9.2° for all nucleotides. The correlation between  $\chi$ -angles extracted from the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  (Duchardt et al. 2004) with the  $\chi$ -angles obtained by analyzing the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  yields an RMSD of 10.1° (Table 2, Fig. 6B). It appears that the  $\chi$ -angles obtained from the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  tend to be systematically smaller than the ones determined by the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$ . When taking only the cytidines into account the two methods are in very good agreement with a RMSD of

**Fig. 5** Spectra of the quantitative  $\Gamma$ -HCNCH experiment for the 14mer RNA. The reference experiment is shown in (A), the cross experiment in (B). Negative signals are shown in red. (C, D): Selected 1D slices taken from the reference (A) and the cross experiment (B). The spectra were recorded on a 700 MHz spectrometer. The reference and the cross experiment were recorded using 64 increments in the C1' dimension and 64 scans for each increment in the reference experiment. The cross experiment was recorded with 512 scans per increment.  $T_M$  was set to 30 ms. The reference experiment was recorded in 2 h and the cross experiment in 16 h



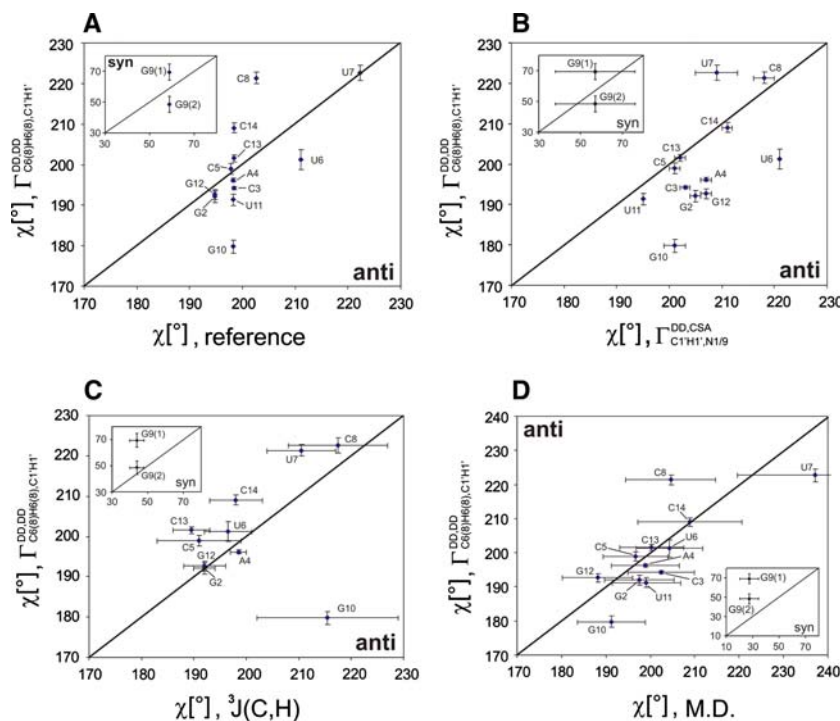
only  $4.7^\circ$ . A possible explanation would be that the CS-tensor for cytidines (Stueber and Grant 2002) is more accurate than the CS-tensors provided for the other types of nucleotides (Duchardt et al. 2004; Ying et al. 2006). The comparison between the  $\chi$ -angles determined by the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  to the  $\chi$ -angles obtained by analyzing the  $^3J(C,H)$  couplings (Duchardt et al. 2004) (Table 2, Fig. 6C) yields an RMSD of  $12.8^\circ$ . Calculating the RMSD without G10, which is far out, reduces the RMSD to  $7.3^\circ$ . Finally, the correlation between the  $\chi$ -angles extracted from a molecular dynamics trajectory and the  $\chi$ -angles obtained by the analysis of the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  yields an RMSD of  $9.9^\circ$ .

The results are in good agreement with the reference structure, the  $\chi$ -angles determined from the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  (Ennifar et al. 2000; Duchardt et al. 2004) and the  $\chi$ -angles extracted from the  $^3J(C,H)$  couplings. Furthermore, it has the following advantages over the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  ( $\Gamma_{N9,C1'H1'}^{CSA,DD}$ )-HCN method: The sensitivity of the experiment is better since the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  rates are considerable larger than the  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  rates (Duchardt et al. 2004). Nucleotides assuming the syn conformation can be identified unambiguously. The main advantage of the new experiment is that it is independent of CSA and possible influences of

other parameters such as the conformation itself on the N1(9) CS tensor do not have any effect on the results. In addition, the method is easy to implement since it is independent from the  $B_0$ -field.

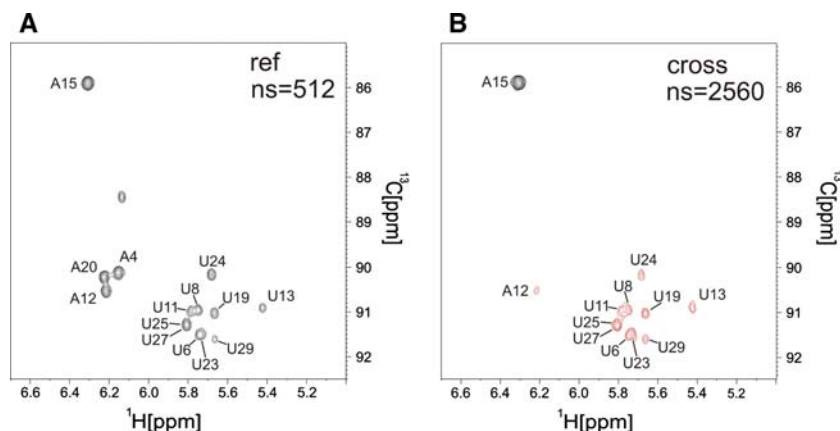
#### SLD 30mer RNA

To test the applicability of the quantitative  $\Gamma$ -HCNCH experiment for larger RNAs it was applied to the SLD 30mer RNA which is part of the CVB3 cloverleaf RNA (Fig. 1B). For this RNA a high resolution solution structure determined by NMR (Ohlenschläger et al. 2004) is available (pdb 1RFR) and the resonance assignment of the C1' and H1' nuclei is known. We tested the method on an adenosine, uridine  $^{13}C$ ,  $^{15}N$  labeled sample. The spectra of the quantitative  $\Gamma$ -HCNCH reference and cross experiment are shown in Fig. 7.  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  are obtained from the intensities of the peaks from the cross and the reference spectra according to Eq. 4. In Fig. 8 the determined  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  rates were compared to the mean values for the  $\chi$ -angle extracted from the NMR ensemble (pdb 1RFR). The black, blue and orange curves depict the predicted  $\Gamma_{C6H6,C1'H1'}^{DD,DD}(\chi)$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}(\chi)$  relations assuming an isotropic  $\tau_c$  of 4.76 ns (Ferner et al. to be published) (Eq. 2, Table 1). Since the



**Fig. 6** (A)  $\chi$ -angles extracted from the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  compared to the  $\chi$ -angles extracted from the X-ray structure (Duchardt et al. 2004). (B) Comparison between the  $\chi$ -angles derived from the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  to the  $\chi$ -angles determined by  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  (Duchardt et al. 2004). (C)  $\chi$ -angles obtained from the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  compared to the  $\chi$ -angles

determined by the  $^3J(C,H)$  couplings (Duchardt et al. 2004). (D) Average of  $\chi$ -angles extracted from the molecular dynamics trajectory. There is only one possibility for the  $\chi$ -angle lying close to the reference value (Ennifar et al. 2000; Duchardt et al. 2004) for each nucleotide in the anti position. In contrast, two values for the  $\chi$ -angle remain possible for the nucleotide G9 which assumes the syn conformation

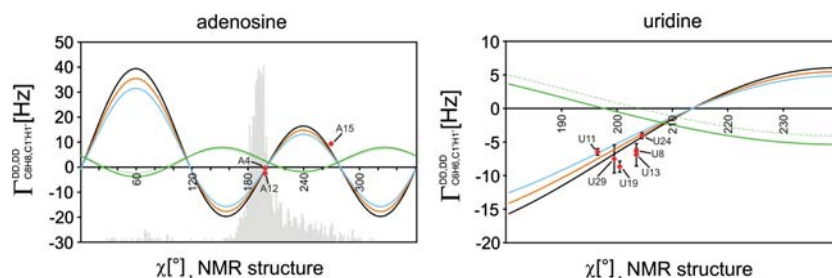


**Fig. 7** Spectra of the quantitative  $\Gamma$ -HCNCH experiment for the 30mer RNA. The reference experiment is shown in (A), the cross experiment in (B). Negative signals are shown in red. The spectra were recorded on a 700 MHz spectrometer. The reference and the cross experiment were recorded using 44 increments in the C1'

dimension and 512 scans for each increment in the reference experiment. The cross experiment was recorded with 2560 scans per increment.  $T_M$  was set to 15 ms. The experiments were recorded within 69 h

30mer RNA has an extended shape, the effect of anisotropic rotational motion was investigated. The size and the orientation of the rotational diffusion tensor were determined with the NMR structure using the program

HYDRONMR7 (Garcia de la Torre et al. 2000). Then, relying on the NMR structure the relative orientations of the bond vectors  $C6H6$  ( $C8H8$ ) and  $C1'H1'$  to the rotational diffusion tensor were determined.

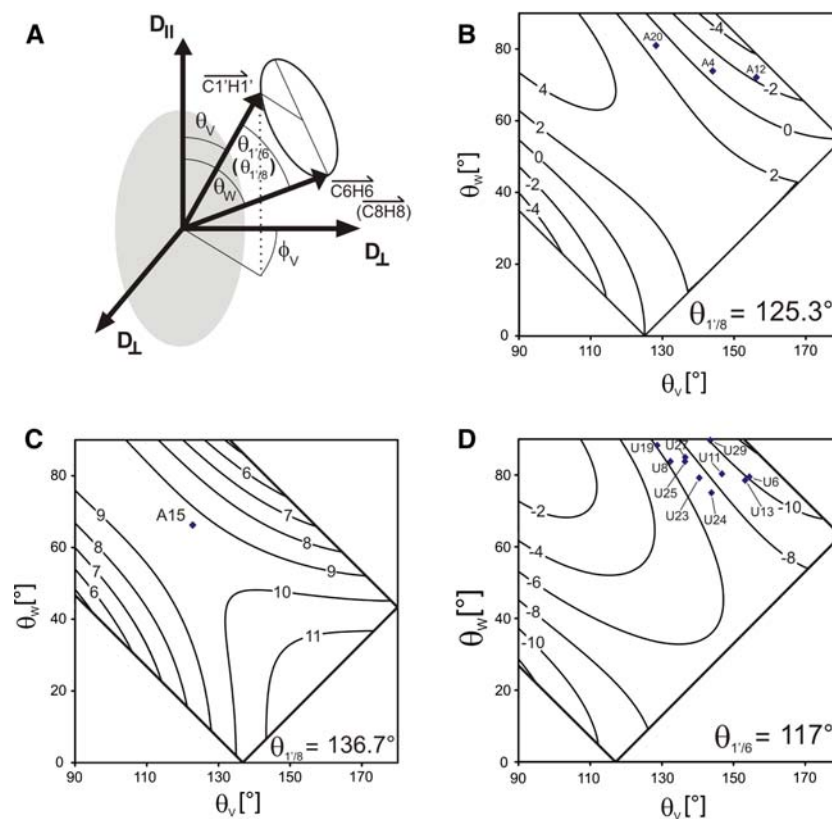


**Fig. 8**  $\Gamma_{C8H8,C1'H1'}^{DD,DD}(\chi)$  for adenosines and  $\Gamma_{C6H6,C1'H1'}^{DD,DD}(\chi)$  for uridines assuming isotropic molecular tumbling, curves are shown for  $(S_{C8H8,C1'H1'})^2$  and  $(S_{C6H6,C1'H1'})^2$  equal to 0.8 (light blue), 0.9 (orange) and 1.0 (black).  $\tau_c$  was set to 4.76 ns (Ferner and Schwalbe to be published) for the SLD 30mer at 310 K. In addition, the experimental  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  from the 30mer RNA are

plotted against the reference  $\chi$ -angles (Ohlenschläger et al. 2004) (red squares).  $\Gamma_{N1,C1'H1'}^{CSA,DD}$  for pyrimidines and  $\Gamma_{N9,C1'H1'}^{CSA,DD}$  for purines are illustrated as green solid (guanosine, cytidine) and dashed (adenosine, uridine) lines, respectively. The gray bars represent the  $\chi$ -angle distribution in the RNA fraction of the large ribosomal subunit (PDB entry 1FFK (Ban et al. 2000))

Figure 9 shows the dependence of the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  on the orientation of the bond vectors relative to the rotational diffusion tensor in the case of anisotropic axially symmetric tumbling (Schneider 1964; Hubbard

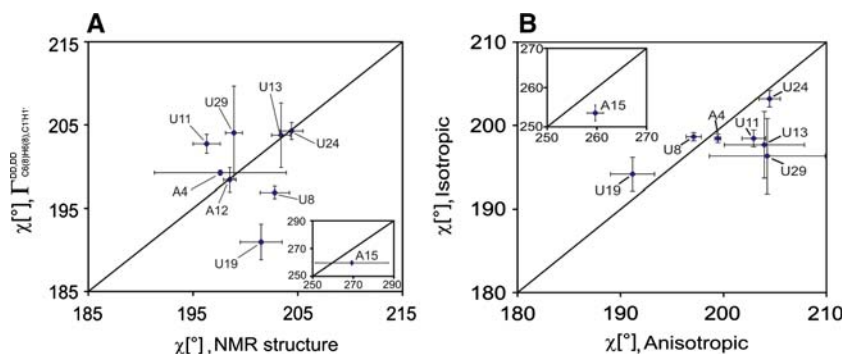
1969). Figure 9A shows the nomenclature that was used. A closer mathematical description is given in the supplementary material (S4). Figure 9B, C and D shows the effect of the anisotropy on the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$



**Fig. 9** Dependence of the  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  on the relative orientation to the axially symmetric rotational diffusion tensor for an anisotropy of 2.05 assuming an overall  $\tau_c$  of 4.76 ns for the SLD 30mer. **(A)** Schematic drawing of the axially symmetric diffusion tensor with the nomenclature of the polar angles in relation to the rotational diffusion tensor. **(B)**  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  dependence on the polar angles  $\theta_V$  and  $\theta_W$  for a projection angle  $\theta_{1/8}$  of 125.3°, which is the mean value of A20( $\theta_{1/8} = 125.3^\circ$ ), A4( $\theta_{1/8} = 125.5^\circ$ ) and A12( $\theta_{1/8} = 124.9^\circ$ ) **(C)**  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  dependence on the polar angles

$\theta_V$  and  $\theta_W$  for a projection angle  $\theta_{1/8}$  of 136.7° (A15). **(D)**  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  dependence on the polar angles  $\theta_V$  and  $\theta_W$  for a projection angle of  $\theta_{1/6}$  of 117° (mean value for all uridines). For **(B, C and D)** the mean values for  $\theta_V$  and  $\theta_W$  for the adenosines and uridines of the SLD 30-mer RNA derived from the NMR structure ensemble (Ohlenschläger et al. 2004) are depicted as blue squares. For the simulated 2D-curves in **(B and D)** the mean values in  $\theta_{1/8}$  and  $\theta_{1/6}$  have been chosen in order to display the results in a compact form. In principle, each nucleotide has its own 2D-curve, which is dependent on  $\theta_{1/6}$  (pyrimidines) or  $\theta_{1/8}$  (purines)





**Fig. 10** (A)  $\chi$ -angles extracted from the  $\Gamma_{\text{C6H6,C1'H1'}}^{\text{DD,DD}}$  and  $\Gamma_{\text{C8H8,C1'H1'}}^{\text{DD,DD}}$  compared to the  $\chi$ -angles extracted from the NMR structure ensemble (Ohlenschläger et al. 2004). (B) Comparison between the  $\chi$ -angles

derived from the  $\Gamma_{\text{C6H6,C1'H1'}}^{\text{DD,DD}}$  and  $\Gamma_{\text{C8H8,C1'H1'}}^{\text{DD,DD}}$  assuming isotropic rotational diffusion and  $\chi$ -angles determined by assuming anisotropic axially symmetric rotational diffusion with an anisotropy of 2.05

at a given projection angle  $\theta_{\text{C6H6,C1'H1'}}$  ( $\theta_{\text{C8H8,C1'H1'}}$ ) between the bond vectors  $\overline{\text{C6H6}}(\overline{\text{C8H8}})$  and  $\overline{\text{C1'H1'}}$ . The extracted angles for  $\theta_V$  and  $\theta_W$  (see Fig. 9A) from the NMR structure show that most of the  $\theta_W$  angles lie between  $70^\circ$  and  $90^\circ$  (blue squares in 9B, C, D). Since the rotational diffusion tensor is oriented almost parallel to the helix axes, it is evident that most of the vectors  $\overline{\text{C6H6}}(\overline{\text{C8H8}})$  are oriented perpendicular to the diffusion tensor. A  $\chi$ -angle of around  $200^\circ$  results in a  $\theta_V$ -angle between  $130^\circ$  and  $160^\circ$  for most of the nucleotides. An exception is the nucleotide A15 that resides within the uCACGg tetraloop (Fig. 9C). Figure 9 shows that the effect of anisotropy on the  $\Gamma_{\text{C6H6,C1'H1'}}^{\text{DD,DD}}$  and  $\Gamma_{\text{C8H8,C1'H1'}}^{\text{DD,DD}}$  rates could be quite large for some  $\theta_V$  and  $\theta_W$  combinations. However, for the observed  $\theta_V$  and  $\theta_W$  combinations in the SLD 30mer this is not the case. Figure 10B shows the effect of rotational anisotropy on the extracted  $\chi$ -angles. The figure shows that the anisotropic tumbling has the largest effect on the nucleotides U29 ( $8^\circ$ ) and A15 ( $6^\circ$ ) but the deviations between anisotropic and isotropic analysis is always  $<9^\circ$ . In fact, the overall RMSD for the isotropic fitting is  $6.9^\circ$ . In conclusion, even for molecules with a large rotational diffusion anisotropy ( $A \approx 2$ ), the assumption of isotropic tumbling still gives reasonable but not absolutely accurate results.

In order to define a high resolution structure the exact orientation of the anisotropy tensor has to be known. However, prediction of the axially symmetric diffusion tensor is quite precise even for a low resolution structure. Therefore, an iterative approach would be possible, in which the data of a molecule with extended shape are first analyzed isotropically and the outcoming structure is used to define a rotational diffusion anisotropy tensor. Subsequently, this tensor can be used to calculate a structure taking anisotropic rotational effects into account. This refinement procedure can be repeated until the structure is converged.

Finally, Fig. 10A gives a comparison between the  $\chi$ -angles taken from the NMR structure and the  $\chi$ -angles determined from the  $\Gamma_{\text{C6H6,C1'H1'}}^{\text{DD,DD}}$  and  $\Gamma_{\text{C8H8,C1'H1'}}^{\text{DD,DD}}$  rates and

an anisotropic but axially symmetric diffusion tensor with an anisotropy of 2.05.

The correlation of the  $\chi$ -angles is in a remarkable agreement with the reference structure with an overall RMSD of  $5.8^\circ$  proving that the method is also applicable for larger RNAs that show anisotropic rotational diffusion. The exact values are given in Table (S3) of the supplementary material.

In fact, the overall RMSD for the isotropic fitting is  $6.9^\circ$ , the RMSD for the anisotropic fitting is slightly better with  $5.8^\circ$ . The interpretation that the precision of the NMR structure might be the limiting factor for the RMSD is likely to be correct. Nevertheless, an RMSD of  $5.8^\circ$  is surprisingly good.

The incorporation of the  $\chi$ -angle data into structural calculations should significantly improve the accuracy of RNA NMR structures since the analysis of the structure of the large ribosomal subunit (pdb ID 1FFK) shows that the  $\chi$ -angles show a considerable spread of around  $50^\circ$  for purines and  $30^\circ$  for pyrimidines (see page 1). Therefore, it is not valid to preset the  $\chi$ -angles to A-form conformation. The  $\chi$ -angles have to be measured. NOEs on their own are most often not able to define the  $\chi$ -angles with the desired accuracy, therefore additional constraints have to be obtained. The analysis of  $\Gamma_{\text{C6H6,C1'H1'}}^{\text{DD,DD}}$  and  $\Gamma_{\text{C8H8,C1'H1'}}^{\text{DD,DD}}$  cross-correlated relaxation rates is an easy to use, robust and precise method to get these constraints. The incorporation of  $\chi$ -angles derived from cross correlated relaxation data into RNA structure calculations has already been described previously (Duchardt et al. 2004).

## Conclusion

We have introduced a new method to determine the glycosidic torsion angle  $\chi$  in RNA oligonucleotides relying on the analysis of CH,CH-dipole–dipole cross-correlated relaxation rates  $\Gamma_{\text{C6H6,C1'H1'}}^{\text{DD,DD}}$  and  $\Gamma_{\text{C8H8,C1'H1'}}^{\text{DD,DD}}$ . This has been

shown exemplarily for the 14mer cUUCGg tetraloop RNA. The method allows to discriminate unambiguously between the syn and the anti conformation and is very accurate for nucleotides in the anti conformation. The provided  $\Gamma^{\text{DD,DD}}(\chi)$  parametrization is directly related to the known geometry of the nucleotides and is independent on the  $B_0$ -field strength. Therefore, the effect is also measurable at lower magnetic fields. Since the measured CH,CH dipolar cross-correlation effect scales up with the rotational correlation time  $\tau_c$  this method is also suited to extract the  $\chi$ -angles of larger RNA molecules as demonstrated for the SLD 30mer RNA.

**Acknowledgements** The paper is dedicated to Christian Griesinger. The work has been supported by the state of Hesse (Center for Biomolecular Magnetic Resonance) and the DFG (Sonderforschungsbereich: RNA-Ligand-Interactions). We wish to thank Oliver Ohlenschläger, Jens Wöhnert and Matthias Görlach for providing us with the 30mer RNA sample.

## References

- Allain FH, Varani G (1995) Structure of the P1 helix from group I self-splicing introns. *J Mol Biol* 250:333–353
- Ban N, Nissen P, Hansen J, Moore PB, Steitz TA (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289:905–920
- Banci L, Bertini I, Felli IC, Hajieva P, Viezzoli MS (2001) Side chain mobility as monitored by CH–CH cross correlation: the example of cytochrome b5. *J Biomol NMR* 20:1–10
- Batey RT, Inada M, Kujawinski E, Puglisi JD, Williamson JR (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. *Nucleic Acids Res* 20:4515–4523
- Batey RT, Battiste JL, Williamson JR (1995) Preparation of isotopically enriched RNAs for heteronuclear NMR. *Methods Enzymol* 261:300–322
- Boisbouvier J, Bax A (2002) Long-range magnetization transfer between uncoupled nuclei by dipole–dipole cross-correlated relaxation: a precise probe of beta-sheet geometry in proteins. *J Am Chem Soc* 124:11038–11045
- Carlomagno T, Felli IC, Czech M, Fischer R, Sprinzl M, Griesinger C (1999) Transferred cross-correlated relaxation: application to the determination of sugar pucker in an aminoacylated tRNA-mimetic weakly bound to EF-Tu. *J Am Chem Soc* 121:1945–1948
- Carlomagno T, Blommers MJ, Meiler J, Cuenoud B, Griesinger C (2001) Determination of aliphatic side-chain conformation using cross-correlated relaxation: application to an extraordinarily stable 2'-aminoethoxy-modified oligonucleotide triplex. *J Am Chem Soc* 123:7364–7370
- Cromsigt J, van Buuren B, Schleucher J, Wijmenga S (2001) Resonance assignment and structure determination for RNA. *Methods Enzymol* 338:371–399
- Duchardt E, Schwalbe H (2005) Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by MNMR <sup>13</sup>C relaxation. *J Biomol NMR* 32:295–308
- Duchardt E, Richter C, Ohlenschläger O, Görlach M, Wöhnert J, Schwalbe H (2004) Determination of the glycosidic bond angle  $\chi$  in RNA from cross-correlated relaxation of CH dipolar coupling and N chemical shift anisotropy. *J Am Chem Soc* 126:1962–1970
- Emsley L, Bodenhausen G (1992) Optimization of shaped selective pulses for NMR using a quaternion description of their overall propagators. *J Magn Reson* 97:135–148
- Ennifar E, Nikulin A, Tishchenko S, Serganov A, Nevskaya N, Garber M, Ehresmann B, Ehresmann C, Nikonov S, Dumas P (2000) The crystal structure of UUCG tetraloop. *J Mol Biol* 304:35–42
- Feller SE, Zhang Y, Pastor RW, Brooks BR (1995) Constant pressure molecular dynamics simulation: the Langevin piston method. *J Chem Phys* 103:4613–4621
- Felli I, Richter C, Griesinger C, Schwalbe H (1999) Determination of RNA sugar pucker mode from cross-correlated relaxation in solution NMR spectroscopy. *J Am Chem Soc* 121:1956–1957
- Ferner J, Duchardt E, Villa A, Stock G, Schwalbe H (to be published)
- Fiala R, Jiang F, Sklenář V (1998) Sensitivity optimized HCN and HCNCH experiments for <sup>13</sup>C/<sup>15</sup>N labeled oligonucleotides. *J Biomol NMR* 12:373–383
- Fiala R, Munzarova ML, Sklenar V (2004) Experiments for correlating quaternary carbons in RNA bases. *J Biomol NMR* 29:477–490
- Foloppe N, MacKerell AD Jr (2000) All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comp Chem* 21:86–104
- Fürtig B, Richter C, Wöhnert J, Schwalbe H (2003) NMR-spectroscopy of RNA. *ChemBioChem* 4:936–962
- Fürtig B, Richter C, Bermel W, Schwalbe H (2004) New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop. *J Biomol NMR* 28:69–79
- García de la Torre J, Huertas ML, Carrasco B (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J Magn Reson* 147:138–146
- Geen HF, Freeman R (1991) Band-selective radiofrequency pulses. *J Magn Reson* 93:93–141
- Hubbard PS (1969) Nonexponential relaxation of three-spin systems in nonspherical molecules. *J Chem Phys* 51:1647
- Ilin S, Bosques C, Turner C, Schwalbe H (2003) Gamma-HMBC: an NMR experiment for the conformational analysis of the o-glycosidic linkage in glycopeptides. *Angew Chem Int Ed* 42:1394–1397
- Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. *J Chem Phys* 79: 926–935
- Kay L, Keifer P, Saarinen T (1992) Pure absorption gradient enhanced heteronuclear single quantum correlation spectroscopy with improved sensitivity. *J Am Chem Soc* 114:10663–10665
- Kloiber K, Schuler W, Konrat R (2002) Automated NMR determination of protein backbone dihedral angles from cross-correlated spin relaxation. *J Biomol NMR* 22:349–363
- MacKerell Jr AD, Banavali NK (2000) All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J Comp Chem* 21:105–120
- Marion D, Ikura M, Tschudin R, Bax AJ (1989) Rapid recording of 2D NMR spectra without phase cycling. Application to the study of hydrogen exchange in proteins. *J Magn Reson* 85:393–399
- Markwick PR, Sprangers R, Sattler M (2005) Local structure and anisotropic backbone dynamics from cross-correlated NMR relaxation in proteins. *Angew Chem Int Ed* 44:3232–3237
- Millet O, Chiarparin E, Pelupessy P, Pons M, Bodenhausen G (1999) Measurement of relaxation rates of NH and Ha backbone protons

- in proteins with tailored initial conditions. *J Magn Reson* 139:434–438
- Munzarova ML, Sklenar V (2003) DFT analysis of NMR scalar interactions across the glycosidic bond in DNA. *J Am Chem Soc* 125:3649–3658
- Nikonowicz EP, Sirr A, Legault P, Jucker FM, Baer LM, Pardi A (1992) Preparation of  $^{13}\text{C}$  and  $^{15}\text{N}$  labeled RNAs for heteronuclear multi-dimensional NMR studies. *Nucleic Acids Res* 20:4507–4513
- Ohlenschläger O, Wöhnert J, Bucci E, Seitz S, Hafner S, Ramachandran R, Zell R, Görlach M (2004) The structure of the stemloop D subdomain of coxsackievirus B3 cloverleaf RNA and its interaction with the proteinase 3C. *Structure* 12:237–248
- Pelupessy P, Chiarparin E, Ghose R, Bodenhausen G (1999) Efficient determination of angles subtended by C( $\alpha$ )-H( $\alpha$ ) and N-H(N) vectors in proteins via dipole-dipole cross-correlation. *J Biomol NMR* 13:375–380
- Quant S, Wechselberger R, Wolter M, Wörner K-H, Schell P, Engels J, Griesinger C, Schwalbe H (1994) Chemical synthesis of  $^{13}\text{C}$ -labelled monomers for solid-phase and template controlled enzymatic synthesis of DNA and RNA oligomers. *Tetrahedron Lett* 35:6649–6652
- Ravindranathan S, Kim CH, Bodenhausen G (2003) Cross correlations between  $^{13}\text{C}$ - $^1\text{H}$  dipolar interactions and  $^{15}\text{N}$  chemical shift anisotropy in nucleic acids. *J Biomol NMR* 27:365–375
- Reif B, Hennig M, Griesinger C (1997) Direct measurement of angles between bond vectors in high-resolution NMR. *Science* 276:1230–1233
- Richter C, Griesinger C, Felli I, Cole PT, Varani G, Schwalbe H (1999) Determination of sugar conformation in large RNA oligonucleotides from analysis of dipole-dipole cross correlated relaxation by solution NMR spectroscopy. *J Biomol NMR* 15:241–250
- Richter C, Reif B, Griesinger C, Schwalbe H (2000) NMR spectroscopic determination of angles  $\alpha$  and  $\zeta$  in RNA from CH-dipolar coupling, P-CSA cross-correlated relaxation. *J Am Chem Soc* 122:12728–12731
- Roehrl MH, Heffron GJ, Wagner G (2005) Correspondence between spin-dynamic phases and pulse program phases of NMR spectrometers. *J Magn Reson* 174:325–330
- Schneider H (1964) *Z Naturforsch Teil A* 19:510
- Schwalbe H, Marino JP, King GC, Wechselberger R, Bermel W, Griesinger C (1994) Determination of a complete set of coupling constants in  $^{13}\text{C}$ -labeled oligonucleotides. *J Biomol NMR* 4:631–644
- Schwalbe H, Carlomagno T, Hennig M, Junker J, Reif B, Richter C, Griesinger C (2001) Cross-correlated relaxation for measurement of angles between tensorial interactions. *Methods Enzymol* 338:35–81
- Shaka AJ, Barker PB, Freeman RJ (1985) Computer-optimized decoupling scheme for wideband applications and low-level operation. *J Magn Reson* 64:547–552
- Sklenar V, Peterson RD, Rejante MR, Feigon J (1993a) Two- and three-dimensional HCN experiments for correlating base and sugar resonances in  $^{15}\text{N}$ ,  $^{13}\text{C}$ -labeled RNA oligonucleotides. *J Biomol NMR* 3:721–727
- Sklenar V, Rejante MR, Peterson RD, Wang E, Feigon J (1993b) Two-dimensional triple-resonance HCNCH experiment for direct correlation of ribose H1' and base H8, H6 protons in  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled RNA oligonucleotides. *J Am Chem Soc* 115:12181–12182
- Sklenar V, Dieckmann T, Butcher SE, Feigon J (1998) Optimization of triple-resonance HCN experiments for application to larger RNA oligonucleotides. *J Magn Reson* 130:119–124
- Stueber D, Grant DM (2002)  $^{13}\text{C}$  and ( $^{15}\text{N}$ ) chemical shift tensors in adenosine, guanosine dihydrate, 2'-deoxythymidine, and cytidine. *J Am Chem Soc* 124:10539–10551
- Sychrovsky V, Muller N, Schneider B, Smrecki V, Spirko V, Sponer J, Trantirek L (2005) Sugar pucker modulates the cross-correlated relaxation rates across the glycosidic bond in DNA. *J Am Chem Soc* 127:14663–14667
- Trantirek L, Stefl R, Masse JE, Feigon J, Sklenar V (2002) Determination of the glycosidic torsion angles in uniformly  $^{13}\text{C}$ -labeled nucleic acids from vicinal coupling constants  $^3\text{J}(\text{C}2)/^4\text{H}1'$  and  $^3\text{J}(\text{C}6)/^8\text{H}1'$ . *J Biomol NMR* 23:1–12
- Tugarinov V, Kay LE (2004)  $^1\text{H}$ ,  $^{13}\text{C}$ - $^1\text{H}$ ,  $^1\text{H}$  dipolar cross-correlated spin relaxation in methyl groups. *J Biomol NMR* 29:369–376
- Varani G, Aboul-ela F, Allain FH-T (1996) NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc* 29:51–127
- Varani G, Tinoco I Jr (1991) RNA structure and NMR spectroscopy. *Q Rev Biophys* 24:479–532
- Vugmeyster L, Pelupessy P, Vugmeister BE, Abergel D, Bodenhausen G (2004) Cross-correlated relaxation in NMR of macromolecules in the presence of fast and slow internal dynamics. *CR Phys* 5:377–386
- Wang T, Frederick KK, Igumenova TI, Wand AJ, Zuiderweg ER (2005) Changes in calmodulin main-chain dynamics upon ligand binding revealed by cross-correlated NMR relaxation measurements. *J Am Chem Soc* 127:828–829
- Wijmenga SS, van Buuren BNM (1998) The use of NMR methods for conformational studies of nucleic acids. *Prog Nucl Magn Reson Spectrosc* 32:287–387
- Ying J, Grishaev A, Bax A (2006) Carbon-13 chemical shift anisotropy in DNA bases from field dependence of solution NMR relaxation rates. *Magn Reson Chem* 44:302–310
- Zwahlen C, Vincent SJ (2002) Determination of ( $^1\text{H}$ ) homonuclear scalar couplings in unlabeled carbohydrates. *J Am Chem Soc* 124:7235–7239

## **Supplementary Material:**

Quantitative  $\Gamma$ -HCNCH: Determination of the glycosidic torsion angle  $\chi$  in RNA oligonucleotides from the analysis of CH dipolar cross-correlated relaxation by solution NMR spectroscopy

Jörg Rinnenthal, Christian Richter, Jan Ferner, Elke Duchardt, and Harald Schwalbe\*

Institute for Organic Chemistry and Chemical Biology, Center for Biomolecular Magnetic Resonance, Max-von-Laue-Strasse 7, D-60438 Frankfurt/Main, Germany.

E-mail: [schwalbe@nmr.uni-frankfurt.de](mailto:schwalbe@nmr.uni-frankfurt.de)

### S1) Results $\Gamma_{C6H6,C1'H1'}^{DD,DD}$ and $\Gamma_{C8H8,C1'H1'}^{DD,DD}$ rates of the 14mer RNA

The  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  rates were measured with the pulse sequence shown in figure 3, with mixing times of  $T_m=20$  ms and of  $T_m=30$  ms. To resolve resonance overlap, the measurements ( $T_m=20$  ms and  $T_m=30$  ms) were also carried out with an evolution period for N1/N9 instead of C1'. In addition, the rates were determined with the alternative pulse sequence shown in this supplementary (S3), in which the magnetization is transferred in the reverse direction from H1' to H6(H8). In the latter case, the rates were determined for  $T_m=20$  ms and  $T_m=30$  ms, too. The error that is given in table 2 takes these 6 different measurements into account. The fact that these different ways of measuring the rates give the same result (small error) show that the measured rates are reproducible.

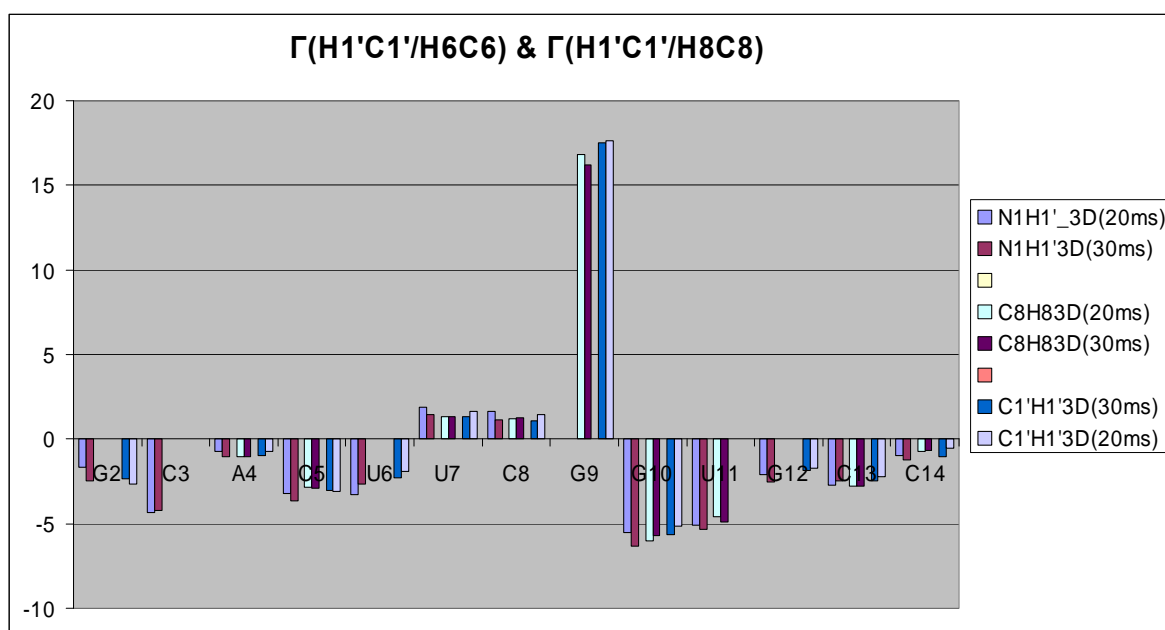


Figure S1: measured  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  [Hz] and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  [Hz] rates for the cUUCGg 14mer RNA

## S2) Deviation of Equation (2)

The geometrical relation is dependent on three bond angles that are constant.

$\alpha$ : H1'-C1'-N9 for purines and H1'-C1'-N1 for pyrimidines

$\beta$ : C1'-N9-C8 for purines and C1'-N1-C8 for pyrimidines

$\gamma$ : N9-C8-H8 for purines and N1-C6-H6 for pyrimidines

The derived equation is:

$$\cos \theta = D \cdot \cos(\alpha) + [\cos(\varepsilon) - D \cdot \cos(\alpha)] \cdot \cos(\chi - 59.01^\circ) \quad (5)$$

$$\text{with } D = \cos(\alpha) \cdot \cos(\varepsilon) + \sin(\alpha) \cdot \sin(\varepsilon) \quad (6)$$

$$\text{and } \varepsilon = \alpha + \beta + \gamma \quad (7)$$

The  $\cos \theta$  term is inserted into equation (1)

$$\Gamma_{C6H6,C1'H1'}^{DD,DD} \equiv \frac{1}{5} \frac{\gamma_H^2 \gamma_C^2}{r_{C6H6}^3 r_{C1'H1'}^3} \left( \frac{\mu_0}{4\pi} \right)^2 \hbar^2 \left( S_{C6H6,C1'H1'}^{DD,DD} \right)^2 \left( 3 \cos^2 \theta_{C6H6,C1'H1'} - 1 \right) \tau_c \quad (1)$$

$$\Gamma_{C8H8,C1'H1'}^{DD,DD} \equiv \frac{1}{5} \frac{\gamma_H^2 \gamma_C^2}{r_{C8H8}^3 r_{C1'H1'}^3} \left( \frac{\mu_0}{4\pi} \right)^2 \hbar^2 \left( S_{C8H8,C1'H1'}^{DD,DD} \right)^2 \left( 3 \cos^2 \theta_{C8H8,C1'H1'} - 1 \right) \tau_c$$

with the trigonometric relation

$$\cos(2 \cdot (\chi - 59.01^\circ)) = 2 \cdot (\cos(\chi - 59.01^\circ))^2 - 1 \quad (8)$$

this leads to equation (2)

$$\Gamma_{C_6H_6,C_1'H_1'}^{DD,DD} = 4.1370\tau_c \left( S_{C_6H_6,C_1'C_1'}^{DD,DD} \right)^2 10^9 [A \cos(\chi - 59.01) + B \cos(2\chi - 118.01) + C] \quad (2)$$

$$\Gamma_{C_8H_8,C_1'H_1'}^{DD,DD} = 4.1370\tau_c \left( S_{C_8H_8,C_1'C_1'}^{DD,DD} \right)^2 10^9 [A \cos(\chi - 59.01) + B \cos(2\chi - 118.01) + C]$$

The parameters A, B and C are given in Table S1 for the 4 different nucleotides. Since the parameters for adenine and guanine are almost identical as well as the parameters for cytosine and uracil, the 4 curves can be summarized to 2 curves, one for purines and one for pyrimidines (Table 1).

Table S1: Parameters for the calculation of the  $\Gamma_{C_6H_6,C_1'H_1'}^{DD,DD}$  and  $\Gamma_{C_8H_8,C_1'H_1'}^{DD,DD}$  according to equation (2)

	A	B	C
adenine	<b>0,575</b>	<b>1,195</b>	<b>0,230</b>
guanine	<b>0,589</b>	<b>1,187</b>	<b>0,224</b>
cytosine	<b>0,802</b>	<b>1,011</b>	<b>0,090</b>
uracil	<b>0,799</b>	<b>1,018</b>	<b>0,096</b>

### S3) Reverse quantitative $\Gamma$ -HCNCH pulse sequence for signal detection on H6 (H8)

Figure S2: Pulse Sequence and experimental parameters for the quantitative  $\Gamma$ -HCNCH experiment employing a  $H1'C1'N9(N1)C8(C6)H8(H6)$  coherence transfer for purines and pyrimidines, resp.

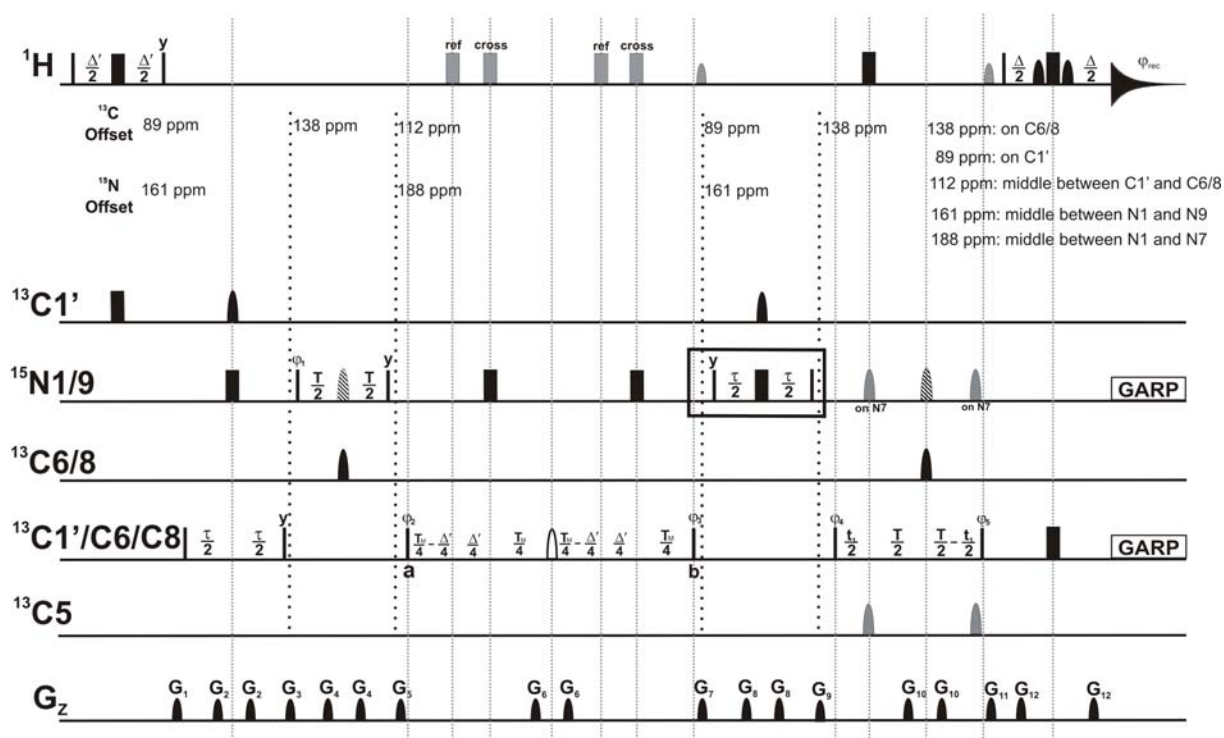


Figure S2: Pulse sequence of the quantitative  $\Gamma$ -HCNCH experiment. Narrow and wide filled bars correspond to rectangular  $90^\circ$  and  $180^\circ$  pulses, respectively. Selective pulses and gradients are indicated as semi-ellipses. The default pulse phase is x. The pulse sequences were optimized on a Bruker spectrometer with the Bruker typical phase settings (Roehrl et al. 2005). The reference and the cross experiment are summarized in one pulse sequence scheme. The gray wide filled bars correspond to  $180^\circ$  pulses applied on the proton channel either in the cross or the reference experiment, as indicated. Two water flipback pulses are applied after the mixing time  $T_M$  and before the last refocusing INEPT step (small gray semi-ellipses). Two selective off-resonant pulses on C5 and two selective off-resonant pulses on N7 during



carbon constant time chemical shift evolution are shown in gray. Fixed delays are adjusted as follows:  $\Delta' = 2.94 \text{ ms } (1/(2*^1J_{H1'C1'}))$ ,  $\Delta = 2.5 \text{ ms } (1/(2*^1J_{H6(8)C6(8)}))$ ,  $T = 32 \text{ ms } (1/(2*^1J_{C6(8)N1(9)}))$ ,  $\tau = 36 \text{ ms } ((1/2*^1J_{N1(9)C1'}))$ .  $T_M$  is the variable relaxation period. The proton carrier frequency is centered at the water frequency (4.7 ppm). The carbon and the nitrogen carrier frequencies change during the course of the experiment as indicated by vertical dashed lines. The values of the  $^{13}\text{C}$  and  $^{15}\text{N}$  offset at every time point are given, respectively. At 700 MHz, band selective pulses are set as follows:  $180^\circ$  Q3 Gaussian cascade (Emsley and Bodenhausen 1992) 2 ms (black semi-ellipse),  $180^\circ$  Q3 Gaussian cascade 1 ms (gray semi-ellipse),  $180^\circ$  Q3 Gaussian cascade 0.5 ms (hatched semi-ellipse),  $180^\circ$  Reburp (Geen and Freeman 1991) 2.53 ms with maxima on  $C1'$  and  $C6/8$  (open semi-ellipse),  $90^\circ$  square pulses 1 ms (small gray semi-ellipse, water flipback),  $90^\circ$  square pulses 0.7 ms (small black semi-ellipse, Watergate soft pulse). Asynchronous GARP decoupling (Shaka et al. 1985) is used to suppress heteronuclear scalar coupling during acquisition. The pulse field gradients of 1 ms length have a smoothed chirp amplitude (Bruker Topspin 2.0, 2006). The pulse field gradient G12 has a length of 0.5 ms. The pulse field gradients are applied along the z-axis and have the following strengths:  $G_1: -40\%$ ,  $G_2: 25\%$ ,  $G_3: 50\%$ ,  $G_4: 20\%$ ,  $G_5: -40\%$ ,  $G_6: 20\%$ ,  $G_7: 65\%$ ,  $G_8: 20\%$ ,  $G_9: 50\%$ ,  $G_{10}: 40\%$ ,  $G_{11}: 20\%$ ,  $G_{12}: 80\%$ . 100% of gradient strength corresponds to 55 Gauss/cm. Phase cycling:  $\phi_1 = x, -x$ ;  $\phi_2 = 8(x), 8(-x)$ ;  $\phi_3 = 16(y), 16(-y)$ ;  $\phi_4 = 4(y), 4(-y)$ ;  $\phi_5 = (2x), 2(-x)$ ;  $\phi_6 = -2(y), 2(y)$ ;  $\phi_{rec} = x, 2(-x), x, -x, 2(x), -x$ . In addition,  $\phi_4$  is incremented in a States-TPPI (Marion et al. 1989) manner to achieve quadrature detection in the  $\omega_1$  dimension. At time point **a**, the coherence  $H1'_z C6(8)_y C1'_y N1(9)_z$  is created and converts during  $T_M$  into the coherence  $H6(8)_z C6(8)_x C1'_x N1(9)_z$  at time point **b**. If additional chemical shift dispersion is needed, a  $N1(9)$  constant time evolution period can be introduced during the first refocusing INEPT step as indicated by the black box.

#### S4) Anisotropic axially symmetric rotational diffusion

For an axially symmetric diffusion tensor, the spectral density of the symmetric top rotator is given by:

$$j_{vw}^q = \frac{1}{20} [(3 \cos^2 \theta_v - 1)(3 \cos^2 \theta_w - 1) * J_{vw}^{q,0} + 12 \cos \theta_v \cos \theta_w \sin \theta_v \sin \theta_w \cos(\phi_v - \phi_w) * J_{vw}^{q,1} + 3 \sin^2 \theta_v \sin^2 \theta_w \cos(2\phi_v - 2\phi_w) * J_{vw}^{q,2}] \quad (9)$$

With the reduced spectral density functions ( $-2 \leq m \leq 2$ )

$$J_{vw}^{q,m} = \frac{2\tau_{c,m}}{(1 + (\omega_q \tau_{c,m})^2)} \quad (10)$$

The rotational correlation time  $\tau_{c,m}$  can be expressed as

$$\frac{1}{\tau_c} = 6D_{vertical} + m^2(D_{parallel} - D_{vertical}) \quad (11)$$

The anisotropy (A) is defined as

$$A = \frac{D_{parallel}}{D_{vertical}} \quad (12)$$

Therefore,  $D_{vertical}$  can be rewritten as

$$D_{vertical} = \frac{1}{2\tau_c(2 + A)} \quad (13)$$

For an axially symmetric diffusion tensor  $\phi_w$  can be set to  $\phi_w = 0$ .

Therefore, the spectral density function (9) simplifies to

$$j_{vw}^q = \frac{1}{20} [(3 \cos^2 \theta_v - 1)(3 \cos^2 \theta_w - 1) * J_{vw}^{q,0} + 12 \cos \theta_v \cos \theta_w \sin \theta_v \sin \theta_w \cos(\phi_v) * J_{vw}^{q,1} + 3 \sin^2 \theta_v \sin^2 \theta_w \cos(2\phi_v) * J_{vw}^{q,2}] \quad (14)$$

To obtain the dependence of the spectral density function from the projection angle between the two bond vectors  $\overrightarrow{C6H6}$  ( $\overrightarrow{C8H8}$ ) and  $\overrightarrow{C1'H1'}$ , the following relation can be used to eliminate  $\phi_v$ :

$$\phi_V = \arccos \left[ \frac{(\cos \theta_{1'6(8)} - \cos \theta_V \cos \theta_W)}{\sin \theta_V \sin \theta_W} \right] \quad (15)$$

Insertion of equation (14) into equation (15) yields:

$$\begin{aligned} j_{VW}^q &= \frac{1}{20} [(3 \cos^2 \theta_V - 1)(3 \cos^2 \theta_W - 1)] * J_{VW}^{q,0} \\ &+ 12 \cos \theta_V \cos \theta_W \sin \theta_V \sin \theta_W \cos \left( \arccos \left[ \frac{(\cos \theta_{1'6(8)} - \cos \theta_V \cos \theta_W)}{\sin \theta_V \sin \theta_W} \right] \right) * J_{VW}^{q,1} \\ &+ 3 \sin^2 \theta_V \sin^2 \theta_W \cos \left( 2 \arccos \left[ \frac{(\cos \theta_{1'6(8)} - \cos \theta_V \cos \theta_W)}{\sin \theta_V \sin \theta_W} \right] \right) * J_{VW}^{q,2} \end{aligned} \quad (16)$$

Finally, the  $\Gamma_{C6H6, C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8, C1'H1'}^{DD,DD}$  rates in the case of anisotropic rotational tumbling are given by insertion of  $j_{VW}^0$  which can be calculated by equation (16) into equation (17)

$$\Gamma_{C6H6, C1'H1'}^{DD,DD} \equiv \frac{\gamma_H^2 \gamma_C^2}{r_{C6H6}^3 r_{C1'H1'}^3} \left( \frac{\mu_0}{4\pi} \right)^2 \hbar^2 (S_{C6H6, C1'H1'}^{DD,DD})^2 * j_{VW}^0 \quad (17)$$

$$\Gamma_{C8H8, C1'H1'}^{DD,DD} \equiv \frac{\gamma_H^2 \gamma_C^2}{r_{C8H8}^3 r_{C1'H1'}^3} \left( \frac{\mu_0}{4\pi} \right)^2 \hbar^2 (S_{C8H8, C1'H1'}^{DD,DD})^2 * j_{VW}^0$$

The used nomenclature for  $\theta_V$ ,  $\theta_W$ ,  $\theta_{1'6}$  and  $\theta_{1'8}$  is indicated in Figure 9A.

## S5) Table of results for the SLD 30mer RNA

**Table S3**  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  and the corresponding  $\chi$ -angles in comparison to the NMR structure.  $(S_{C6H6,C1'H1'}^{DD,DD})^2$  and  $(S_{C8H8,C1'H1'}^{DD,DD})^2$  were used for the  $\chi$ -angle calculation.

Nucleo- tide	$\Gamma_{C6H6,C1'H1'}^{DD,DD}$	$(S_{C6H6,C1'H1'}^{DD,DD})^2$	$\chi[\text{°}]$	$\chi[\text{°}]$	$\chi[\text{°}]$ (NMR- structure)
	[Hz]		$\Gamma_{C6H6,C1'H1'}^{DD,DD}$ $\Gamma_{C8H8,C1'H1'}^{DD,DD}$ (anisotropic)	$\Gamma_{C6H6,C1'H1'}^{DD,DD}$ $\Gamma_{C8H8,C1'H1'}^{DD,DD}$ (isotropic)	
A4	-0.4 ± 0.2	0.955	199.4 ± 0.3	198.4 ± 0.3	197.6 ± 6.3
U8	-6.3 ± 0.3	0.917	197.1 ± 0.7	198.6 ± 0.6	202.8 ± 1.4
U11	-6.5 ± 0.5	0.930	202.9 ± 1.1	198.4 ± 1.0	196.3 ± 1.3
A12	-2.5 ± 0.9	0.933	198.6 ± 1.5	195.2 ± 1.4	198.5 ± 0.6
U13	-6.9 ± 1.7	0.939	204.0 ± 3.9	197.7 ± 3.5	203.4 ± 0.9
A15	9.3 ± 0.5	0.662	259.7 ± 1.6	253.5 ± 2.0	269.3 ± 18
U19	9.3 ± 0.8	0.948	191.2 ± 2.1	194.2 ± 1.7	201.5 ± 2
U24	-4.1 ± 0.4	0.894	204.5 ± 1.0	203.2 ± 1.0	204.4 ± 1.1
U29	-7.5 ± 2.1	0.934	204.2 ± 5.6	196.3 ± 4.4	198.9 ± 0.8

The reference angles are extracted from the NMR structure ensemble (pdb ID 1RFR). The  $\Gamma_{C6H6,C1'H1'}^{DD,DD}$  and  $\Gamma_{C8H8,C1'H1'}^{DD,DD}$  were determined several times with different mixing times  $T_M$  (10ms, 15ms).  $({}^{cross}S_{C6H6,C1'H1'}^{DD,DD})^2$  and  $({}^{cross}S_{C8H8,C1'H1'}^{DD,DD})^2$  were estimated from the autocorrelated order parameters (Ferner, to be published #67)

as a product of  $\sqrt{\left({}^{auto}S_{C6H6,C6H6}\right)^2} * \sqrt{\left({}^{auto}S_{C1'H1',C1'H1'}\right)^2}$  and  $\sqrt{\left({}^{auto}S_{C8H8,C8H8}\right)^2} * \sqrt{\left({}^{auto}S_{C1'H1',C1'H1'}\right)^2}$  respectively.  $\left(S_{C6H6,C1'H1'}^{DD,DD}\right)^2$  and  $\left(S_{C8H8,C1'H1'}^{DD,DD}\right)^2$  were used for determination of the  $\chi$ -angles.

## Literature

- Emsley L, Bodenhausen G (1992). Optimization of Shaped Selective Pulses for NMR Using a Quaternion Description of Their Overall Propagators. *J Magn Reson* 97: 135-148.
- Ferner J, Schwalbe H ( to be published).
- Geen HF, Freeman R (1991). Band-Selective Radiofrequency Pulses. *J Magn Reson* 93: 93-141.
- Marion D, Ikura M, Tschudin R, Bax AJ (1989). Rapid Recording of 2D NMR Spectra without Phase Cycling. Application to the Study of Hydrogen Exchange in Proteins. *J Magn Reson* 85: 393-399.
- Roehrl MH, Heffron GJ, Wagner G (2005). Correspondence between spin-dynamic phases and pulse program phases of NMR spectrometers. *J Magn Reson* 174: 325-30.
- Shaka AJ, Barker PB, Freeman RJ (1985). Computer-optimized decoupling scheme for wideband applications and low-level operation. *J Magn Reson* 64: 547-552.



## **Research Article:** NMR and MD Studies of the Temperature-Dependent Dynamics of RNA YNMG-Tetraloops

Jan Ferner, Alessandra Villa, Elke Duchardt, Elisabeth Widjajakusuma, Jens Wöhnert, Gerhard Stock, and Harald Schwalbe

*Nucleic Acids Res.* **2008**, 36, 1928-1940

The combination of experimental NMR data and MD simulation enabled to extract internal dynamics of RNA tetraloops at distinct temperatures, where the loops are still stable as well as close to the initiation of melting of the less stable variant. This approach revealed the differences in stability of two structurally similar YNMG-tetraloops. To determine the order parameters, a more precise parameterization of the bond lengths and the CSAs were needed for the derivation from autocorrelated relaxation rates as well as a new computational method for the extraction of order parameters from a MD trajectory.

All MD simulation and analysis were performed in the group of Prof. Stock. The author of this thesis rendered the NMR measurements more precisely, incorporated the new parameters in the model-free analysis and optimized this analysis. He was mainly in charge of writing the article.





# NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops

Jan Ferner<sup>1</sup>, Alessandra Villa<sup>2</sup>, Elke Duchardt<sup>1</sup>, Elisabeth Widjajakusuma<sup>2</sup>,  
Jens Wöhnert<sup>1</sup>, Gerhard Stock<sup>2</sup> and Harald Schwalbe<sup>1,\*</sup>

<sup>1</sup>Institut für Organische Chemie und Chemische Biologie, Center for Biomolecular Magnetic Resonance and <sup>2</sup>Institut für Physikalische und Theoretische Chemie, Johann Wolfgang Goethe-Universität Frankfurt, Max-von-Laue-Str. 7, 60438 Frankfurt/M, Germany

Received November 15, 2007; Revised December 27, 2007; Accepted December 28, 2007

## ABSTRACT

In a combined NMR/MD study, the temperature-dependent changes in the conformation of two members of the RNA YNMG-tetraloop motif (cUUCGg and uCACGg) have been investigated at temperatures of 298, 317 and 325 K. The two members have considerable different thermal stability and biological functions. In order to address these differences, the combined NMR/MD study was performed. The large temperature range represents a challenge for both, NMR relaxation analysis (consistent choice of effective bond length and CSA parameter) and all-atom MD simulation with explicit solvent (necessity to rescale the temperature). A convincing agreement of experiment and theory is found. Employing a principle component analysis of the MD trajectories, the conformational distribution of both hairpins at various temperatures is investigated. The ground state conformation and dynamics of the two tetraloops are indeed found to be very similar. Furthermore, both systems are initially destabilized by a loss of the stacking interactions between the first and the third nucleobase in the loop region. While the global fold is still preserved, this initiation of unfolding is already observed at 317 K for the uCACGg hairpin but at a significantly higher temperature for the cUUCGg hairpin.

## INTRODUCTION

Despite the limited number of ribonucleotide residues, RNA molecules are capable to fold into a wide variety of secondary and tertiary structures. This structural diversity is based on the capability of RNA to form networks of stacked nucleobases together with several hydrogen bonds between donor and acceptor functional groups not only

in the nucleobase but also in the ribose moiety and along the phosphodiester backbone. The variability in structure is matched by considerable diversity in biological functions. Besides RNA's encoding and translational activity, enzymatic and regulatory functions have been found in biological systems as well as in artificial systems evolved in SELEX experiments. These functions cannot be explained solely based on characteristics of static structures but dynamic properties appear to play an important role in maintaining the functional diversity of RNA (1). RNA dynamics span a broad range of time scales from picoseconds where vibrations and angular fluctuations occur up to seconds where catalytic function and global refolding take place (2–4).

By NMR heteronuclear relaxation measurements, insight can be obtained into the pico- to nanosecond motions (5) by analyzing the relaxation rates  $R_1$ ,  $R_{1\rho}$  and the steady-state heteronuclear NOE of nitrogen and carbon nuclei following the model-free formalism (6–8). Such information has recently been augmented by data stemming from residual dipolar couplings in RNA, where RNA helices have been specifically elongated to shift the time scale of the overall rotational tumbling (9). While analysis of relaxation data for  $^1\text{H}$ – $^{15}\text{N}$  sites is widely used for proteins to gain information on the backbone dynamics, the dynamical investigation of nucleic acids from protonated  $^{15}\text{N}$  nuclei is quite limited. Only for base-paired guanine and uridine imino groups, relaxation data can be retrieved, which restricts the analysis mostly to the helical stems and to the nucleobase moieties of RNA. In contrast,  $^{13}\text{C}$  relaxation has proven to be a strong alternative since data can be obtained for all nucleobase and ribose moieties by different strategies (10–15) to fulfill the model-free requirement of an isolated I-S spin system. Recently, Bax and co-workers (16) investigated protein dynamics and demonstrated that also for nonisolated spin systems found in uniformly  $^{13}\text{C}$ -labeled proteins or RNA, accurate relaxation data can be obtained and analyzed up to an overall correlation time in the low nanosecond timescale. These reports initiated detailed dynamical

\*To whom correspondence should be addressed. Tel: 69 7982 9737; Fax: 69 7982 9515; Email: schwalbe@nmr.uni-frankfurt.de

analyses of RNA molecules (17,18), which revealed that also parameters such as C–H bond lengths and  $^{13}\text{C}$  chemical shielding anisotropies need to be known accurately to derive correct information about the amplitude of local motions in RNA.

To compare the findings of NMR relaxation experiments with microscopic models, it is instructive to perform complementary molecular dynamics (MD) studies of RNAs. By using an all-atom force field for solute and solvent, classical MD simulations directly provide information at the atomic level on inter- and intramolecular motions (19,20). In protein research, the combination of NMR and MD investigations has been widely used (21–28) and recently, also comprehensive descriptions of fast conformational dynamics of small RNA and DNA systems have been published (29–32, Duchardt, E., Nilsson, L. and Schleucher, J., submitted for publication). While NMR results may serve as benchmark data to study the accuracy of the MD description, the MD trajectory in turn provides a microscopic model for the underlying conformational dynamics.

To investigate the correlation of structure, dynamics and function, RNA hairpins with YNMG loop motifs (Y = C or U, N = any, M = A or C) have been chosen for the present investigation. The most stable RNA tetraloop with a UUCG sequence closed by a Watson–Crick CG base pair has been intensely studied structurally and dynamically both, by NMR (17,18,33,34,86) and MD (30,35–38). This tetraloop motif is highly abundant in nature, stabilizes structures and serves as folding initiation site (secondary structural elements of the cUUCGg tetraloop are summarized in Table S12 in the Supplementary Data). No ligands are known to bind with high affinity to this tetraloop sequence. In contrast, a structurally highly similar hairpin with a CACG loop sequence closed by an UG wobble base pair is more labile with a melting temperature decreased by more than 15 K (39). In Coxsackievirus B3, this tetraloop interacts directly with the protein proteinase 3C as essential component of the viral replication initiation complex. These differing properties cannot be explained by the surprisingly small structural variations (secondary structural elements of the uCACGg tetraloop are summarized in Table S13 in the Supplementary Data). Our approach of investigating temperature-dependent NMR-derived order parameters  $S^2$  has been previously applied to the study of (local) protein folding and unfolding (40–45).

In this work, we investigate the dynamical properties of both hairpin types by NMR  $^{13}\text{C}$  relaxation and MD simulation. We evaluate in detail parameters such as carbon–hydrogen bond length and  $^{13}\text{C}$ -chemical shift anisotropies, which significantly affect the interpretation of subnanosecond dynamics by the model-free analysis. The hairpins dynamics were recorded at a temperature close to the melting start of the less stable uCACGg hairpin to investigate the differences of local and global dynamics of the two tetraloops. MD simulations suggest a similar pathway of temperature-induced unfolding of the two hairpins. Interestingly, at high temperatures, nonnative stacking interactions are revealed in one of the four main conformations.

## METHODS

### Sample preparation

The 14-nt UUCG-hairpin (5'-PO<sub>3</sub><sup>-</sup>-PO<sub>2</sub><sup>-</sup>-PO<sub>2</sub><sup>-</sup>-GGCAC UUCGGUGCC-3'; bold residues indicate the YNMG loop-residues) was purchased from Silantes GmbH (Munich, Germany) as a uniformly  $^{13}\text{C}/^{15}\text{N}$ -labeled RNA. The concentration of the NMR sample was 0.7 mM in 20 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub>, pH 6.4, 0.4 mM EDTA and 10% v/v D<sub>2</sub>O. For the investigation of the 30-nt CACG-hairpin (5'-PO<sub>4</sub><sup>-</sup>-GGCACUCUGGUAU CACGGUACCUUUGUGUC-3'; bold residues indicate the loop-residues) two samples were synthesized by *in vitro* transcription with T7-RNA polymerase with a linearized plasmid DNA as template and purified as described previously (46). One sample contained uniformly  $^{13}\text{C}/^{15}\text{N}$ -labeled guanosine and cytidine residues with a RNA concentration of 0.6 mM, while the second was synthesized with uniformly  $^{13}\text{C}/^{15}\text{N}$ -labeled adenosine and uridine residues at a final concentration of 1.2 mM. The buffer composed 10 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub> adjusted to pH 6.2, 40 mM KCl, 0.2 mM EDTA and 99.99% v/v D<sub>2</sub>O.

### NMR spectroscopy

All NMR experiments were carried out on a 600 MHz Bruker (Rheinstetten, Germany) spectrometer equipped with a 5 mm  $^1\text{H}\{^{13}\text{C}/^{15}\text{N}\}$  Z-Grad TXI probe.  $^{13}\text{C}$   $R_1$ ,  $R_{1\rho}$  and  $^{13}\text{C}\{-^1\text{H}\}$  steady-state NOE data were obtained using  $^{13}\text{C}$  modifications of the Bruker standard pulse sequences for  $^{15}\text{N}$  relaxation measurements (hsqct1etf3gpsi3d, hsqctretf3gpsi3d.2 and invnoef3gpsi) (47–50). For the 14 nt cUUCGg-hairpin, data were collected as published before (17). The 30 nt uCACGg-hairpin data were measured with slight changes compared to the 14 nt RNA. The carrier frequency was set to 137 and 89 p.p.m. for the aromatic and the ribose carbons with spectral widths of 10 and 12 p.p.m., respectively. During the carbon evolution, off-resonant carbon Q3 pulses (512  $\mu\text{s}$ ) were applied with an offset of  $-7000$  Hz and  $-5000$  Hz in order to suppress the  $^1\text{J}(\text{C}5, \text{C}6)$  and  $^1\text{J}(\text{C}1', \text{C}2')$  couplings, respectively. Further, long-range homonuclear couplings were assumed to be negligible. For the  $R_1$  and  $R_{1\rho}$  measurements, a relaxation delay of 2 s was applied between each of the 32 scans for all 104 complex points in the indirect dimension.  $R_1$ - and  $R_{1\rho}$ -subspectra with varying relaxation delays were acquired in one interleaved pseudo-3D experiment. For the acquisition of the  $R_1$  relaxation rates, the relaxation delay  $\tau_M$  was set to 10, 50, 100, 200, 400, 700 ms, 1 and 1.5 s, while spectra with  $\tau_M = 50$  and 400 ms were recorded twice for error determination.  $R_{1\rho}$  rates were determined as described previously (17) with a spin-lock field of 3.6 kHz at an offset of 2000 Hz and relaxation delays  $\tau_M$  of 12, 24, 36, 48, 64, 80, 104, 128 ms. Duplicate measurements were carried out for  $\tau_M = 24$  and 80 ms. The  $^{13}\text{C}\{-^1\text{H}\}$  steady-state NOE experiments were recorded with 8 scans for 128  $t_f$ -increments in an interleaved manner, with alternating proton-presaturated and nonpresaturated spectra. During the premeasurement delay of 6 s, on-resonance presaturation was applied for 3 s subsequent to a 3 s relaxation delay. To avoid heating differences

between the proton-presaturated and the nonpresaturated spectra, an off-resonance presaturation was applied during the blank experiment. The interleaved spectra were separated by a Bruker standard macro. All measurements were repeated four times for error determination.

### Data analysis

Spectra were processed and analyzed with Topspin 1.3 (Bruker, Germany).  $R_1$  and  $R_{1\rho}$  relaxation decays were fit from peak heights to monoexponential two-parameter functions with the program Sparky 3.112 (University of California, San Francisco, USA). The measured  $R_{1\rho}^{\text{meas}}$  rates contain spin-lock offset and spin-lock power dependent contributions of  $R_1$  and  $R_2$  following Equation (1):

$$R_{1\rho}^{\text{meas}} = R_1 \cos^2 \theta + R_2 \sin^2 \theta \quad 1$$

where  $\theta$  is the angle of the effective spin-lock field with the  $B_0$  field for each nucleus defined by Equation (2):

$$\theta = \tan^{-1} \left( \frac{\nu}{\Omega} \right) \quad 2$$

with  $\nu$  being the spin-lock field strength and  $\Omega$  the offset of the nucleus resonance from the spin-lock carrier, both in Hertz. The influence of homonuclear Hartmann–Hahn magnetization transfer on the  $R_{1\rho}$  rates can be neglected for the nuclei investigated (17).

Further, the CC-dipolar contribution to the relaxation of the carbons  $C_6$  and  $C_1'$  from the adjacent carbon nuclei becomes more relevant with an increase of the overall correlation time  $\tau_c$  (51):

$$\frac{\rho_{CC}}{\rho_{CH}} = \frac{\gamma_C^2 r_{CH}^6}{3\gamma_H^2 r_{CC}^6} \omega_C^2 \tau_c^2 \quad 3$$

in which  $\rho_{Ci}$  is the dipolar contribution to the relaxation of the  $^{13}\text{C}$  nucleus arising from the attached  $^1\text{H}$  or  $^{13}\text{C}$  nucleus;  $\gamma_i$  is the gyromagnetic ratio of the attached nucleus;  $r_{Ci}$  is the distance to the attached nucleus;  $\omega_C$  is the carbon Larmor frequency and  $\tau_c$  the overall correlation time. But also for the 30 nt RNA with a  $\tau_c$  at 298 K of about 6 ns the contributions are still below 1% and therefore are neglected in the model-free analysis.

For the smaller 14-mer RNA a minimum error of 2% for the  $R_1$  and  $R_2$  relaxation rates and 3% for the  $^{13}\text{C}\text{--}\{^1\text{H}\}$  steady-state NOE values were used. These minimum errors were increased for the larger 30mer RNA to 3% for the  $R_1$  and  $R_2$  rates and 5% for the  $^{13}\text{C}\text{--}\{^1\text{H}\}$  steady-state NOE. For some of the relaxation rates, the experimentally derived errors were larger and then in these cases used.

For the analysis of the relaxation parameters the program *Modelfree 4.15* by Palmer and co-workers (52) was used. The further residue-specific parameters (C–H bond lengths and  $^{13}\text{C}$  chemical shift anisotropy) are discussed and optimized in this paper. Model selection was according to d’Auvergne and Gooley (53) and Wright and co-workers (54) with a Bayesian Information Criteria (BIC). The procedure is discussed in detail in the Supplementary Data. The global parameters, the overall

rotational correlation time  $\tau_c$  and the diffusion anisotropy were first determined by a hydrodynamic calculation performed with the program *hydromr 5a* (55). These values were then optimized by the program *Modelfree* using the relaxation data.

### MD simulations

The MD simulations were performed using the GROMACS suite of programs (version 3.2) (56,57). The AMBER force field (parm98) (58,59) was employed to describe the 14-mer cUUCGg and uCACGg RNA hairpins. The hairpin was placed in a rhombic dodecahedron box (edge length  $\sim 5$  nm), which was subsequently filled with TIP3P water molecules (60). To neutralize the system, 13 sodium ions were placed randomly in the simulation box.

A twin range cut-off was used for the Lennard–Jones interactions that is interactions between atoms within 1.0 nm were evaluated every step, while interactions between atoms within 1.4 nm were evaluated every five steps. The particle mesh Ewald method (61) was employed to treat Coulomb interactions, using a switching distance of 1.0 nm. Constant pressure  $p$  and temperature  $T$  were maintained by weakly coupling the system to an external bath at one bar and at the selected temperature, using the Berendsen barostat and thermostat, respectively (62). The RNA, the ions and the solvent were independently coupled to the temperature bath with a coupling time of 0.1 ps. The pressure coupling time was 0.5 ps and the isothermal compressibility  $4.5 \times 10^{-5} \text{ bar}^{-1}$ . The bond distances and the bond angle of the solvent water were constrained using the SETTLE algorithm (63). All other bond distances were constrained using the LINCS algorithm (64). A leapfrog integrator with an integration time step of 2 fs was used.

The starting structure of 14-mer cUUCGg hairpin was taken from simulations previously performed at 298 K (32). The starting structure of the 14-mer uCACGg hairpin was taken from the Brookhaven Protein Data Bank (PDB structure 1RFR.pdb) (39). Fifty-nanosecond classical MD simulations have been performed at 298 K, 317 K and 360 K.

Analysis of the trajectories was performed with tools from the GROMACS package and with modified versions of them. To define the presence of a hydrogen bond, an acceptor–donor distance smaller than 0.35 nm and an acceptor–hydrogen–donor angle larger than  $150^\circ$  degrees was requested. To define base–base hydrogen bond interactions, the presence of the N–H...N hydrogen bond between Watson–Crick base pair and N–H...O hydrogen bonds between wobble base pair is used as indicator. Two bases were considered as stacked when their center of mass separation is within 0.15 nm of that seen in the folded structure (simulation at 298 K). Pairs with larger separations are considered as broken (65).

To obtain NMR order parameters, the corresponding internal correlation functions were calculated from Equation (4) (66):

$$C(t) = \langle P_2(\vec{\mu}(0) \cdot \vec{\mu}(t)) \rangle \quad 4$$

where  $\vec{\mu}$  is a unit vector pointing along the C–H bond,  $P_2(x) = (1/2)(3x^2 - 1)$  is the second Legendre polynomial, and  $\langle \dots \rangle$  denotes an equilibrium average over the MD trajectory from which molecular rotation and translation was removed. Subsequently, the first 100 ps of the correlation functions were fit to the Lipari–Szabo expression (6):

$$C_I(t) = S^2 + (1 - S^2)e^{-t/\tau_e} \quad 5$$

where  $S^2$  is the order parameter and  $\tau_e$  is the effective (or internal) correlation time for the C–H dipole. Using the first nanoseconds of  $C(t)$  hardly changes the results (32).

To compare the conformational fluctuations of the hairpin at different temperatures, we performed a principal component analysis (67–69) of the trajectories at 298, 317 and 360 K, where only atoms of residues C5-G10/U13-G18 were considered. Preceding the analysis, each conformation was translated and rotated (through a least root mean square fit) to give the best fit to the averaged structure. For the cUUCGg hairpin, the first two (of in total 579) eigenvectors describe 57 and 47% of the total fluctuations at 298 and 360 K, respectively. For the uCACGg hairpin, 46 and 44% of the total fluctuations are covered.

## RESULTS

### Analysis of relaxation rates and parameter optimization for the model-free analysis

The modulations of the dipolar and the chemical shift anisotropy fields are the two main components influencing the relaxation of NMR-active nuclei. The model-free analysis of  $^{13}\text{C}$  relaxation data, in turn, depends critically on the value of the C–H bond length and the carbon chemical shielding anisotropy (5,17). These two parameters are fundamental, since they weigh the two dominant contributions to the  $^{13}\text{C}$  relaxation and their determination is often mutually dependent. In addition, it is often assumed that the chemical shift anisotropies do not depend on conformation. Studies on proteins, however, have shown a clear conformational dependence of the  $^{15}\text{N}$  CSAs (70–72).

Recently, *ab initio* calculations were performed by Case to incorporate zero-point motion averaging of the bonds in proteins (73) and in nucleic acids (Case, D., personal communication). The corresponding bond lengths for nucleotides are summarized in Table 1. The incorporation of the zero-point motion averaged bond lengths into the model-free analysis results in increased order parameters  $S^2$  close to or even exceeding the maximum value of unity. These unphysical results ( $S^2 > 1$ ) can be remedied for the base vectors C6H6 and C8H8 by correction of the carbon chemical shielding anisotropy ( $^{13}\text{C}$  CSA).

In 2006,  $^{13}\text{C}$ -CSA values have been reported both, for a helical A-form RNA segment and a helical B-form DNA dodecamer (74). These CSA values differ significantly from the previous values (Table 2) with changes up to 16%. Trantirek and co-workers (75) calculated the conformational dependence of cross-correlated relaxation

**Table 1.** C–H bond lengths used in earlier studies for relaxation analysis (R) and zero-point motion averaged bond lengths which give an effective bond length  $r_{\text{eff}}$

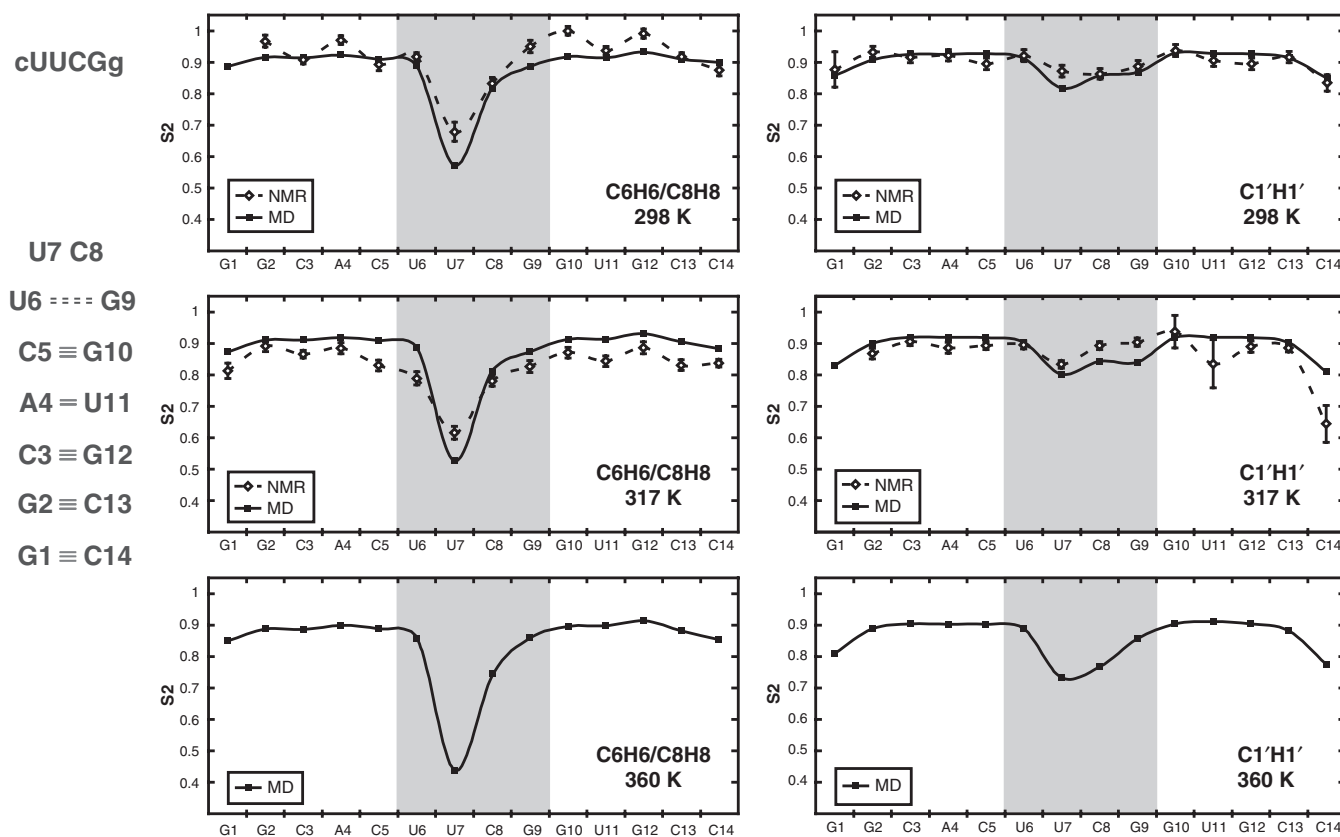
	bond	R [Å]	$r_{\text{eff}}$ [Å]
Ribose	C1'–H1'	1.095	1.118
Pyrimidine	C6–H6	1.082	1.102
Purine	C8–H8	1.079	1.102
	C2–H2	1.079	1.102

**Table 2.** Chemical shift anisotropies (CSA) of aromatic carbon nuclei in nucleobases

Carbon nucleus	Nucleobase	CSA [p.p.m.]	
		Formerly used	New
C8	Adenine	–134	–144
	Guanine (anti-conformation)	–134	–133
	Guanine (syn-conformation)		–122
C6	Cytidine	–179	–208
	Uridine	–179	–208
C2	Adenine		–168

rates across the glycosidic torsion angle  $\chi$  of deoxy-nucleosides. From their analysis, a dependence of the  $^{13}\text{C}$ -CSA values on the sugar pucker mode was proposed (Brumovska, E., Sychrovsky, V., Vokacova, Z., Sponer, J., Schneider, B. and Trantirek, L., manuscript in preparation). While for the aromatic carbons the influence of the sugar pucker mode on the CSA is negligible, the dependence on the torsion angle  $\chi$  is significant. For a purine nucleotide in syn-conformation ( $\chi$ -angle of around  $60^\circ$ ), the absolute value of the carbon C8 CSA is reduced by 10 to 13 p.p.m. For pyrimidines, the difference of 20 to 30 p.p.m. for the C6 CSA is even more pronounced. These results are of particular interest for the analysis of the hairpins studied here, since the guanosine residue in the fourth position of an YNMG-loop adopts a syn-conformation. In our previous analysis, this guanosine residue appeared as the second most flexible moiety while the residue to which guanosine is base paired to at the first position of the loop was as rigid as the stem residues (17). In addition, the order parameters derived from  $^{13}\text{C}$  relaxation rates were different from those derived from a  $^{15}\text{N}$  relaxation analysis (34). Taking into account the new predictions of a  $\chi$ -value dependence of the CSA (Table 2) alleviates the differences between the previous analyses and yields fully consistent data for the guanine residue (Figure S1 in the Supplementary Data).

For the ribose carbon C1', no CSA values have been published so far that take a zero-point motion averaged bond length into account. The published CSA range from 29 to 60 p.p.m. and were determined by DFT calculations on ribose and deoxyribose model compounds (76,77), solid-state measurements on isolated nucleotides (78) or liquid-crystal NMR on an A-form helical RNA (79). These measurements showed no differences for the individual nucleotides. For our analysis, we used the previous



**Figure 1.** Order Parameters  $S^2$  of the 14 nt cUUCGg hairpin (left of the diagrams) extracted from NMR  $^{13}\text{C}$  relaxation data (open diamonds) and from a MD trajectory (squares) for the nucleobase vectors C6H6/C8H8 (left) and the ribose vectors C1'H1' (right) at temperatures of 298 K (top), 317 K (middle) and 360 K (bottom). The loop region is shaded in gray. For the highest temperature of 360 K, NMR relaxation data are not available due to stability reasons.

parameters with a bond length of 1.09 Å and an absolute value of the chemical shielding anisotropy of 45 p.p.m.

Hydrodynamic calculations using the program *hydroNMR* (80) of the 14 mer RNA resulted in starting values of the global correlation time  $\tau_c \approx 2.30$  and a diffusion anisotropy  $D_{\parallel/\perp} \approx 1.50$ . The parameters were further optimized using the program *Modelfree* (52) during the fitting procedure. This optimization results in a correlation time of 2.27 ns, which lies between our previously published values of 2.18 and 2.44 ns, as a result of separately fitting the carbons in the nucleobases and the ribose moieties. The diffusion anisotropy of 1.48 is close to the value calculated by *hydroNMR*.

### Temperature dependence of $S^2$

In order to study the temperature dependence of the conformational dynamics of both RNA hairpins, NMR relaxation measurements have been performed at room temperature (298 K), at a temperature (325 K) close to the melting point of the uCACGg hairpin, and at an intermediate temperature (317 K). To compare these results with corresponding MD simulations, one needs to take into account that MD melting studies typically overestimate melting temperatures (81–84). Besides sampling problems, this temperature shift is most likely related to deficiencies of the force field, since standard

biomolecular force fields are parameterized to reproduce properties at room temperature. In fact, a recent replica exchange MD simulation study on 14-mer RNA hairpins (85) resulted in melting temperatures of 425 K and 397 K for cUUCGg and uCACGg hairpins, respectively. To achieve premelting conditions for uCACGg hairpin in the MD simulation, we therefore employed a temperature of 360 K instead of the experimental temperature of 325 K. As to our knowledge, all current empirical force fields are plagued by this problem; it remains to be shown by future theoretical models whether the simple temperature rescaling yields the correct structures and dynamics.

### Experiment

Figure 1 shows the NMR-derived order parameters  $S^2_{\text{NMR}}$  of the cUUCGg 14-mer hairpin for the nucleobase vectors C6H6/C8H8 (left) and the ribose vectors C1'H1' (right). The results of the model-free analysis at 298 K are given in Table 3 and in Figure 1 (top). The nucleobase of loop residue U7 is by far the most flexible moiety in the entire RNA molecule. This observation is in agreement with the high-resolution structures (33; Nozinovic, S., Fürtig, B., Jonker, H., Richter, C. and Schwalbe, H., manuscript in preparation) that reveal no stabilizing interactions like hydrogen bonds or stacking for this nucleotide. Also, the loop riboses of U7, C8 and G9 display slightly increased

**Table 3.** Results of the model-free analysis of the 14 nt cUUCGg RNA at 298 K extracted from the NMR relaxation data and the MD trajectory

	C6H6/C8H8/C2H2 <sup>a</sup>				C1'H1' <sup>b</sup>				
	NMR		MD		NMR			MD	
	$S^2$	$S_f^2$	$\tau_c$ [ps]	$R_{ex}$ [ $s^{-1}$ ]	$S^2$	$S^2$	$\tau_c$ [ps]	$R_{ex}$ [ $s^{-1}$ ]	$S^2$
G1					0.89	$0.88 \pm 0.06$		$13.97 \pm 1.24$	0.86
G2	$0.97 \pm 0.02$			$0.59 \pm 0.38$	0.92	$0.93 \pm 0.02$		$2.62 \pm 0.38$	0.91
C3	$0.91 \pm 0.01$				0.91	$0.92 \pm 0.02$		$0.87 \pm 0.33$	0.93
A4	$0.97 \pm 0.01$				0.92	$0.92 \pm 0.02$		$1.12 \pm 0.33$	0.93
	<i><math>0.93 \pm 0.02</math></i>			<i><math>2.8 \pm 1.10</math></i>	<i>0.92</i>				
C5	$0.89 \pm 0.02$		$82 \pm 44$		0.91	$0.90 \pm 0.02$		$1.28 \pm 0.31$	0.93
U6	$0.92 \pm 0.01$				0.89	$0.92 \pm 0.02$		$0.51 \pm 0.31$	0.91
U7	$0.68 \pm 0.03$	$0.76 \pm 0.02$	$311 \pm 204$		0.57	$0.87 \pm 0.02$		$3.67 \pm 0.36$	0.82
C8	$0.83 \pm 0.02$			$0.81 \pm 0.47$	0.82	$0.86 \pm 0.02$		$4.81 \pm 0.38$	0.86
G9	$0.95 \pm 0.02$			$1.50 \pm 0.37$	0.89	$0.89 \pm 0.02$		$1.36 \pm 0.33$	0.87
G10	$1.00 \pm 0.01$				0.92	$0.94 \pm 0.02$		$0.56 \pm 0.35$	0.93
U11	$0.94 \pm 0.01$				0.91	$0.91 \pm 0.02$		$1.26 \pm 0.32$	0.93
G12	$0.99 \pm 0.01$				0.93	$0.90 \pm 0.02$		$2.95 \pm 0.34$	0.93
C13	$0.92 \pm 0.01$				0.91	$0.92 \pm 0.02$		$0.80 \pm 0.31$	0.92
C14	$0.87 \pm 0.02$		$36 \pm 29$		0.90	$0.84 \pm 0.03$	$36 \pm 15$	$0.72 \pm 0.33$	0.85

<sup>a</sup>The vector C6H6 of the pyrimidines, the vector C8H8 of the purines and the adenine C2H2 vectors (italics) were analyzed with a bond length of 1.102 Å and a residue-specific carbon CSA (Table 2).

<sup>b</sup>For the C1'H1'-vector a bond length of 1.09 Å and a carbon CSA of 45 p.p.m. were used.

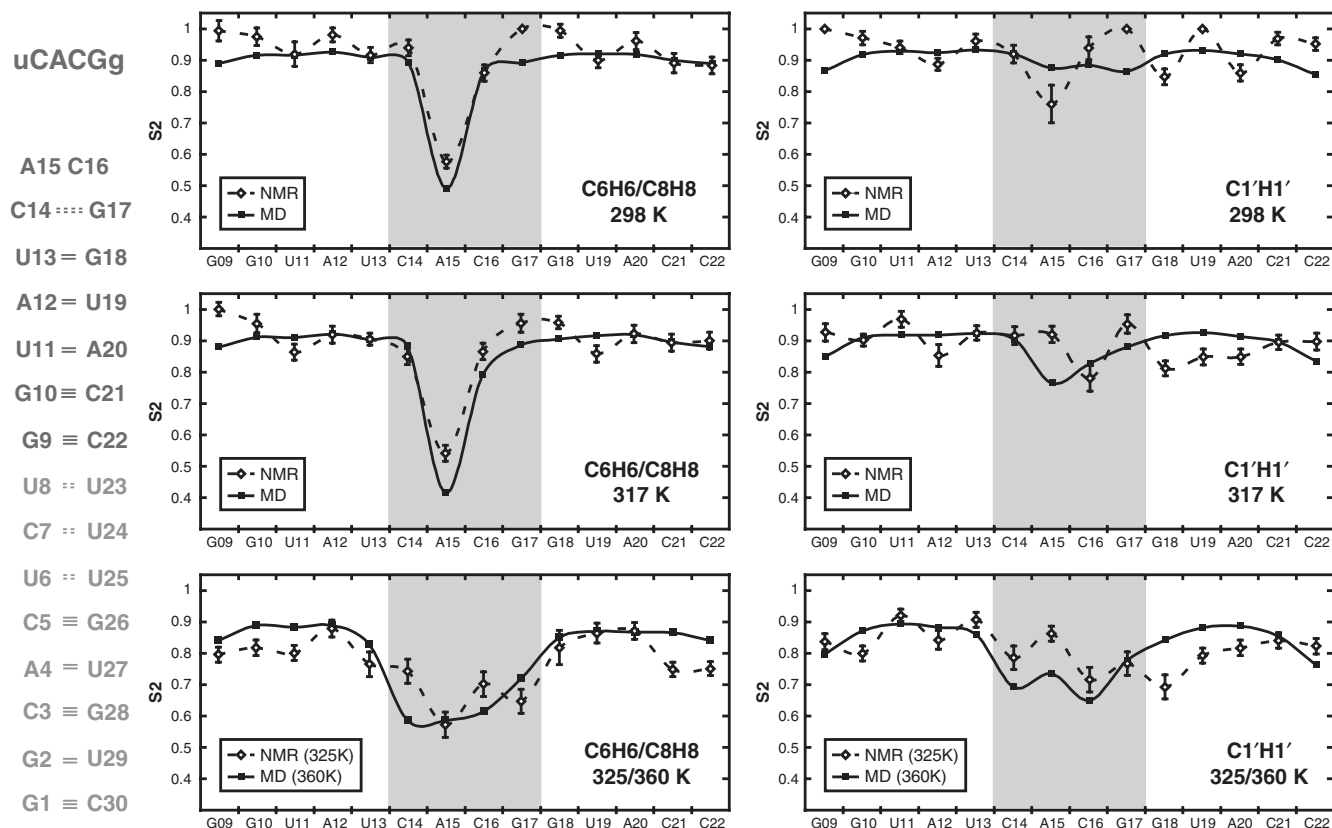
dynamics compared to the rigid structured helical stem. In the high-resolution NMR solution structure (Nozinovic, S., Fürtig, B., Jonker, H., Richter, C. and Schwalbe, H., manuscript in preparation), the riboses U7 and C8 are in C2'-endo conformations [pseudorotation phase P (U7) = 163° and P (C8) = 176°] and the ribose of G9 has a syn-conformation of its nucleobase.

An increase of the temperature to 317 K affects only the global overall tumbling of the RNA, whereas the internal motions do not change significantly. The diffusion anisotropy of  $D_{\parallel/\perp} = 1.47$  is maintained and the global correlation time  $\tau_c$  is decreased to 1.53 ns as expected (the *hydroNMR* derived value is 1.70 ns). In addition to the faster global motion, also the order parameters decrease slightly at a higher temperature. For the nucleobase moieties, the average  $S_{NMR}^2$  decrease is larger and more uniform (0.98 to 0.87 for C8H8 and 0.87 to 0.80 for C6H6) than for the ribose moieties (0.90 to 0.87) where the loop riboses of C8 and G9 even have a slight increase of  $S^2$  (Figure 1 (mid right) and Table S8 in the Supplementary Data). Of all nucleobase moieties, only U7 is flexible with a fast internal motion and a very small  $R_{ex}$ -term.

Figure 2 shows the order parameters  $S_{NMR}^2$  of the uCACGg 30-mer RNA hairpin. In the uCACGg hairpin, the nucleobase of residue A15 at the second loop position is the only flexible moiety (Figure 2 top left). The dynamics of the nucleobase of A15 can be described with a fast motion (<10 ps) and a slow motion with an internal correlation time of  $441 \pm 121$  ps (Table 4). The analyses of the C1' relaxation rates are more difficult. A number of residues around the loop can hardly be fit, especially at the temperature of 298 K. The order parameters of these residues fluctuate between 0.8 and 1.0. In the loop, the sugar moiety of A15 is the most flexible with a slow motion of  $827 \pm 445$  ps. The other C1'H1' ribose sites for which fits can be obtained have  $R_{ex}$ -terms between 1.9 and  $11.8 s^{-1}$  (Table 4).

For the uCACGg hairpin, the changes at higher temperatures become more pronounced. At a temperature of 317 K at which the imino signals of the non-Watson-Crick base pairs can no longer be observed in a <sup>15</sup>N-HSQC (39), no significant changes for the order parameters of the aromatic and anomeric carbons can be observed in the NMR analysis. The global correlation time decreases from 6.55 to 4.15 ns, which is comparable to the predicted decrease in global correlation time by *hydroNMR* of over 2 ns. For the internal dynamics of the nucleobase moieties, a decrease of the order parameter by <0.03 is observed while the overall profile remains unchanged (Figure 2 mid left). This result suggests that the disappearance of the imino resonances as detected in the <sup>15</sup>N-HSQC is due to an increase of exchange rates with the solvent water at elevated temperatures but not due to an increase in dynamics. The doubling of the  $R_{ex}$ -term to  $15.3 s^{-1}$  of the C8 atom of the G17 nucleobase in the syn-conformation is significant and in agreement with the time range of the water exchange. The other loop nucleobases C14 and C16 display  $R_{ex}$ -terms of  $7.4 s^{-1}$  and  $3.1 s^{-1}$ , indicating the appearance of a slow motion. The data for nucleobase of A15 is described by the same motional model with similar order parameters  $S_f^2$ ,  $S_s^2$  and an analogous internal correlation time of  $404 \pm 136$  ps (Table S9 in the Supplementary Data).

A further temperature increase to 325 K does not change the <sup>13</sup>C-HSQC of the aromatic and of the anomeric region significantly, revealing that the averaged conformation is similar to the determined structure at 298 K. Therefore, a model-free analysis of the relaxation data can be performed (Table S10 in the Supplementary Data). The model-free analysis of the relaxation data at 325 K results in a global correlation time of 3.09 ns, which is exactly the same value as derived from *hydroNMR*. The average order parameters of all nucleobases drop to  $\sim 0.77$  and the values obtained for the stem residues are



**Figure 2.** Order Parameters  $S^2$  of the 30-nt (NMR) and 14-nt (MD) uCACGg hairpin (left of the diagrams) extracted from NMR  $^{13}\text{C}$  relaxation data (open diamonds) and from a MD trajectory (squares) for the nucleobase vectors C6H6/C8H8 (left) and the ribose vectors C1'H1' (right) at temperatures of 298 K (top), of 317 K (middle) and at the onset of the hairpin melting at 325 K in NMR and 360 K in MD, respectively. The loop region is shaded in gray.

**Table 4.** Results of the model-free analysis of the 30 nt uCACGg RNA hairpin part G9 to C22 at 298 K extracted from the NMR relaxation data and of the 14 nt uCACGg RNA extracted from the MD trajectory

	C6H6/C8H8/C2H2				$S^2$	C1'H1'				$S^2$
	NMR		MD			NMR		MD		
	$S^2$	$S_r^2$	$\tau_c$ [ps]	$R_{ex}$ [ $s^{-1}$ ]	$S^2$	$S^2$	$S_r^2$	$\tau_c$ [ps]	$R_{ex}$ [ $s^{-1}$ ]	$S^2$
G9	$0.99 \pm 0.03$			$5.5 \pm 1.8$	0.89	$1.0 \pm 0.0$			$11.8 \pm 0.0$	0.87
G10	$0.97 \pm 0.03$				0.92	$0.97 \pm 0.02$				0.92
U11	$0.92 \pm 0.04$			$2.9 \pm 2.8$	0.92	$0.94 \pm 0.02$				0.93
A12	$0.98 \pm 0.02$				0.93	$0.89 \pm 0.02$				0.93
	$0.93 \pm 0.02$				0.92					
U13	$0.92 \pm 0.02$				0.91	$0.96 \pm 0.02$				0.93
C14	$0.94 \pm 0.03$				0.89	$0.92 \pm 0.03$			$4.4 \pm 1.6$	0.92
A15	$0.58 \pm 0.02$	$0.70 \pm 0.02$	$441 \pm 121$		0.49	$0.76 \pm 0.06$	$0.86 \pm 0.06$	$827 \pm 445$		0.88
	$0.57 \pm 0.02$	$0.68 \pm 0.02$	$635 \pm 233$		0.43					
C16	$0.86 \pm 0.03$			$6.7 \pm 2.2$	0.87	$0.94 \pm 0.04$			$2.8 \pm 1.6$	0.89
G17	$1.0 \pm 0.0$			$8.9 \pm 0.0$	0.89	$1.0 \pm 0.0$			$2.3 \pm 0.0$	0.87
G18	$0.99 \pm 0.02$				0.92	$0.85 \pm 0.03$			$7.4 \pm 1.2$	0.92
U19	$0.90 \pm 0.02$				0.92	$1.0 \pm 0.0$			$1.9 \pm 0.0$	0.93
A20	$0.96 \pm 0.03$				0.92	$0.86 \pm 0.03$			$3.9 \pm 1.2$	0.92
	$0.93 \pm 0.02$				0.91					
C21	$0.89 \pm 0.03$			$5.8 \pm 2.4$	0.90	$0.97 \pm 0.02$				0.90
C22	$0.88 \pm 0.03$			$7.4 \pm 2.9$	0.89	$0.95 \pm 0.02$				0.85

Parameters used are the same as in Table 3.

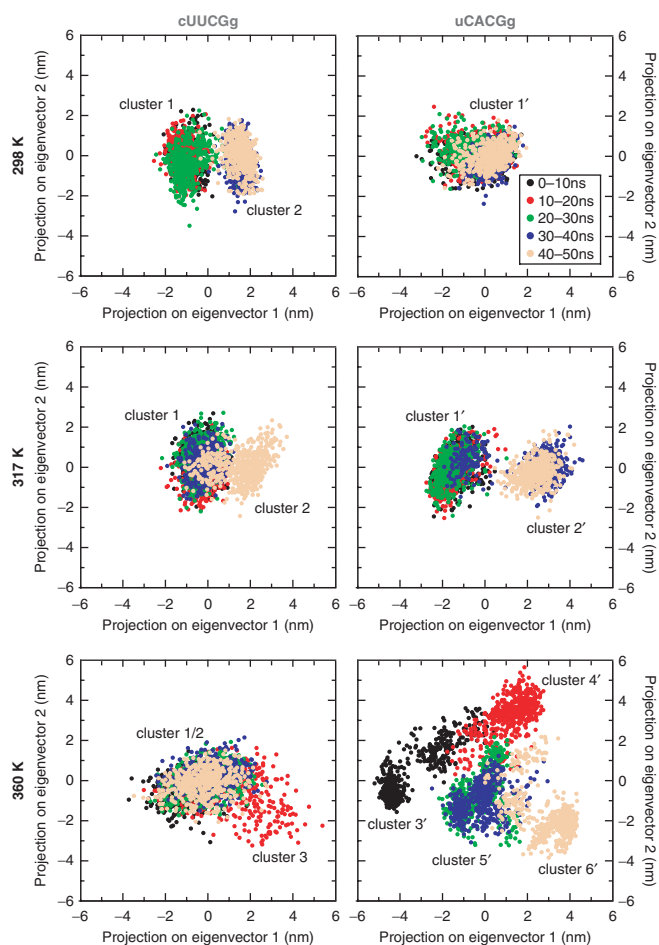
only slightly higher (between 0.78 for the pyrimidine C6H6 and 0.81 for the purine C8H8). For the nucleobase of A15, the order parameters of the slow and the fast motion are very close to the values which are gained at lower temperatures. Only the internal correlation time of the slow motion is reduced by a factor of 2 to  $224 \pm 162$  ps. The surrounding nucleobases of the loop and the loop closing base pair show the largest and most significant changes (Figure 2 bottom). For nucleobases U13, C14 and C16 involved in stacking interaction in the ground state conformation, the order parameters decrease and a similar internal correlation time of  $\sim 25$  ps is obtained. The nucleobase of G17 exhibits a more dramatic increase of dynamics and becomes the second most flexible moiety. In addition, also a significant  $R_{ex}$ -term ( $20.9 \pm 2.8 \text{ s}^{-1}$ ) and a fast internal correlation time ( $13 \pm 11$  ps) are obtained. The adjacent G18 shows a slower motion of  $64 \pm 29$  ps and has a large decrease in  $S_{\text{NMR}}^2$ . These observations indicate that the 3'-end of the loop is the primary melting site, while the 5'-end is more stabilized by the stacking of bases U13, C14 and C16.

The model-free analysis of the adenine C2 relaxation rates results in very similar parameters compared to the corresponding adenine C8 data in both RNAs and at all measured temperatures (Tables 3, 4 and S5, S6 and S8–S11 in the Supplementary Data). Although the NMR relaxation data of the C1' can be fit at 325 K, a wave-like modulation of the order parameters is observed which we do not further investigate here.

### MD simulation

The comparison of measured and calculated order parameters of the cUUCGg 14 mer RNA hairpin shown in Figure 1 reveals an excellent overall agreement of experiment and theory. At both the temperatures, 298 and 317 K, the deviations between  $S_{\text{MD}}^2$  and  $S_{\text{NMR}}^2$  are within the estimated systematic errors of  $\pm 0.05$ . When going from 298 to 317 K, the experimental order parameters show an average decrease of  $\approx 0.1$ , while the calculated values of  $S^2$  change only marginally. Even at 360 K, only a modest decrease of the MD-derived order parameters  $S_{\text{MD}}^2$  is observed. However, particularly the central loop residues U7 and C8 become more flexible. Figure 2 compares the order parameters obtained for the 30 nt (NMR) and 14 nt (MD) uCACGg hairpin. Again, a convincing overall agreement of experiment and MD calculation is observed. In particular, the calculations nicely match the temperature-dependent changes of the nucleobase measurements.

The good overall agreement between experimental and calculated order parameters indicates that the underlying theoretical model is suitable to reproduce the temperature dependence of the RNA structure and dynamics and suggests that valid microscopic dynamic information can be extracted from the MD simulations. The MD trajectories were analyzed by a principal component analysis on the coordinates of residues around the loop region to identify the principal motions of the RNA hairpins (see Methods section). Figure 3 presents the resulting population probability along the first two principal



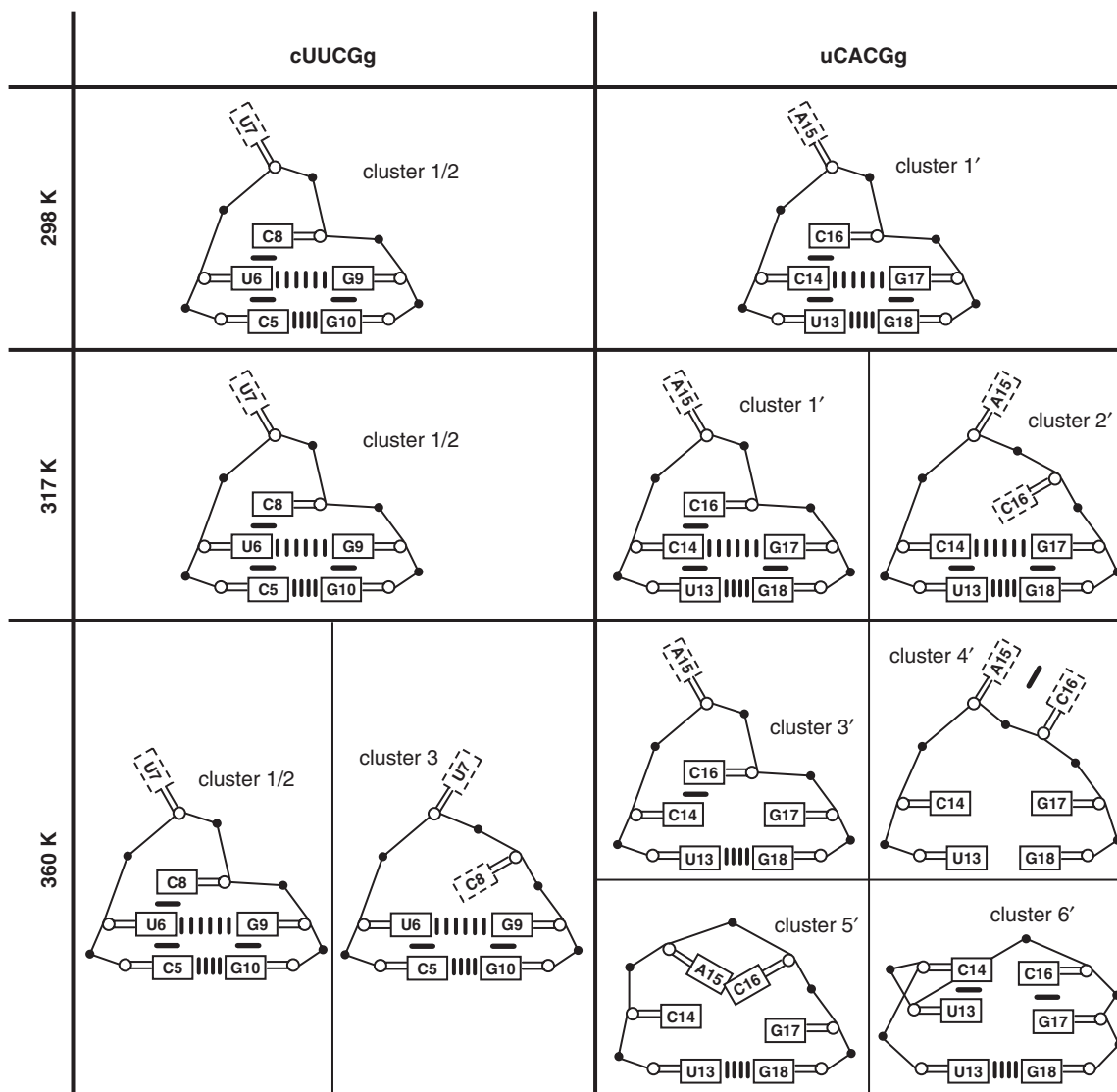
**Figure 3.** The cUUCGg (left) and uCACGg (right) hairpin configurations at 298 K (top), 317 K (middle) and at 360 K (bottom) projected on the first two PCA eigenvectors. The PCA analysis was performed on the atoms coordinates of residues C5–G10/U13–G18. Shown are the projections every 20 ps for five different time intervals: 0–10 ns (black); 10–20 (red); 20–30 ns (green); 30–40 ns (blue) and 40–50 (beige).

components at the three temperatures under consideration, which reveal the coexistence of several conformational states. Figure 4 shows representative conformations of the hairpin for these states.

For the cUUCGg hairpin, the trajectories at 298 and 317 K are seen to consist of two conformational states (cluster 1 and 2). They reflect a rearrangement of the backbone involving residues U6 and U7. A closer analysis reveals that the observed dihedral transition correlates with the increase of the syn-oriented population of the nucleobase U7 (Figure 5). In contrast, the secondary structural interactions in the loop region are not affected. Upon increase of the temperature to 360 K, the principal component analysis reveals an additional conformational state (cluster 3), which is characterized by an unstacked nucleobase C8 (see Figure 4 bottom left). As expected, the unusually stable cUUCGg RNA hairpin generally does not exhibit major changes of structure and dynamics in the temperature range under consideration.

The situation is different for the uCACGg hairpin. Due to its significantly lower melting temperature,



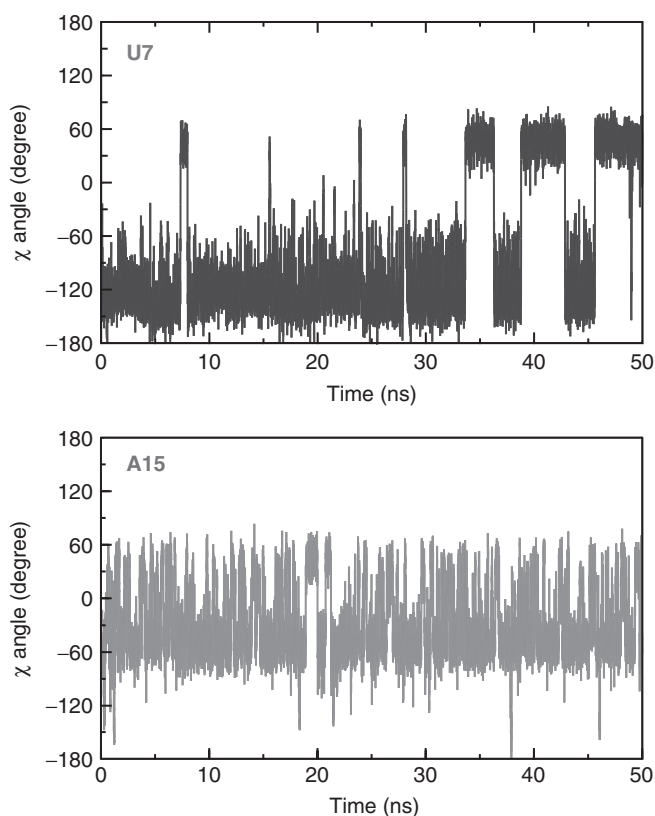


**Figure 4.** Representative conformations of the PCA clusters (Figure 3) from the MD simulations at 298 K, 317 K and 360 K. Dashed lines around residue labels indicate that the residues point into the solvent. Parallel lines between residues indicate base pair interactions and thick lines stacking interactions between nucleobases.

this RNA loop exhibits substantial changes of the local dynamics on the pico- to nanosecond timescale upon an increase of the temperature from 298 to 360 K. While only one conformation is observed at 298 K in the principal component analysis in Figure 3 (cluster 1'), two and four widely spread clusters are populated at 317 K and 360 K, respectively. A first conformational change is observed at 317 K (cluster 2'), where a backbone conformational rearrangement causes the loss of the stacking interactions between the residues C16 and C14, see Figure 4 (mid right).

At 360 K, the secondary structure elements in the loop region are gradually lost and the closing base pair of U13 and G18 is also disturbed. The time evolution of the torsion angle  $\chi$  and the pseudorotational angle  $P$  reveal that the motions of ribose and nucleobase are mostly uncorrelated (Supplementary Figure S2 in the

Supplementary Data). Only the nucleobase and the ribose moiety of G18 have a high correlation value (0.84). Figure 4 (bottom right) comprises the main conformational states found in the 360 K trajectory of the uCACGg hairpin. In the cluster 3', the hairpin mostly satisfies the structural characteristics of the YNMG-motif with only the hydrogen bond between C14 and G17 missing. The cluster 4' displays that the nucleobase of C16 is not stacked anymore on top of C14 but loops out and potentially has stacking interactions with the neighboring A15. A mostly unfolded structure is observed in the cluster 5'. This cluster dominates the simulation with a population of  $\sim 40\%$ . In the cluster 6', the nucleobases seem rigid since stacking interactions between the neighboring residues C14 and A15 as well as C16 and G17 stabilize the loop. Snapshots are in Figure S3 in the Supplementary Data.



**Figure 5.** Dynamic behavior of the loop-out residues U7 (top) and A15 (bottom) at 298 K extracted from the molecular dynamics simulation trajectory revealed by the torsional angle  $\chi$ .

The MD calculated order parameters contain contributions arising from the characteristic fast dynamics for each of these conformations, while the conformational change between the clusters occur at 10 ns timescale, thus are not detected by  $^{13}\text{C}$  NMR relaxation measurement. The presence of several nonnative stacking interactions involving residues of the loop regions explain the similar flexibility observed for these residues ( $S_{\text{MD}}^2 \approx 0.7\text{--}0.6$  and  $S_{\text{NMR}}^2 \approx 0.75\text{--}0.6$ ). In particular, nonnative stacking interactions involving the loop-out residue A15 at the premelting temperature may explain the reduced flexibility of this nucleobase with respect to lower temperatures ( $S_{\text{MD}}^2 \approx 0.5\text{--}0.4$  at 298–317 K) and ( $S_{\text{MD}}^2 \approx 0.7$  at 360 K).

## DISCUSSION

Compared to protein structures, RNA structural architecture appears to be considerably more modular. Tetraloops are an important structural modular element; they play an important role in RNA structure since they connect single-stranded segments involved in regular A-form helices. Loops are therefore often exposed and involved in biomolecular interactions and have also been postulated to form sites where folding and melting are initiated. The two tetraloops investigated here have similar ground state conformations (Figure 4 top) but differ

in their thermal stability. At a temperature of 298 K, the two hairpins also reveal similar conformational dynamics on the pico- to nanosecond timescale (Figures 1 and 2). In both hairpins, increased motions are only found for the looped-out nucleobase at the second loop position (U7 in the cUUCGg hairpin and A15 in the uCACGg hairpin). This observation is reflected in the relative differences of the order parameters between stem and the looped-out nucleobase, which clearly differ for the two hairpins:  $S_{\text{U11}}^2 - S_{\text{U7}}^2 \approx 0.26$  (NMR) and 0.34 (MD) versus  $S_{\text{A12}}^2 - S_{\text{A15}}^2 \approx 0.40$  (NMR) and 0.44 (MD). The finding is a consequence of the nucleobase motion around the glycosidic bond angle  $\chi$  (Figure 5). While U7 exhibits frequent anti/syn transitions, its fluctuations in these conformations are rather small ( $\pm 30^\circ$ ). The anti/syn transitions are on a nanosecond timescale, longer than the overall correlation time of the molecule and therefore cannot be detected by NMR. The glycosidic angle of A15 varies continuously around  $0^\circ$  with an amplitude of  $60^\circ$  and only a slight preference for  $-60^\circ$ . These fluctuations are fast enough to be detected by NMR  $^{13}\text{C}$  relaxation and therefore yield a lower order parameter.

By comparison of MD and NMR data, we can delineate the overall thermal unfolding pathway of the two hairpins and identify hot spots where unfolding is initiated at a temperature where the global fold is still maintained. When the temperature is increased, the loss of the stacking interactions between the first and the third nucleobase in the loop region is the first event that destabilizes the hairpin. In the MD simulations of the cUUCGg hairpin, it takes a temperature of 360 K for this to happen. At this temperature, the C8 nucleobase partially unstacks from the nucleobase U6 and extrudes into the solution (see cluster 3 in Figure 4), which is reflected in a  $S^2$  decrease of U7 and C8. For the uCACGg hairpin, these changes already occur at a temperature of 317 K (see cluster 2' in Figure 4). In both sequences, it is a cytidine stacking on the 5'-end of the loop, which has an intrasidial hydrogen bond to its ribose and a cross-strand hydrogen bond to the phosphate backbone between the first and second loop residue. Less favorable stacking interactions of the nucleobase C16 on the cytidine C14 and/or a slight deviation of the backbone along the top loop residues A15 and C16 which is observable in the structural calculations of the uCACGg hairpin (39) and the cUUCGg hairpin (Nozinovic, S., Fürtig, B., Jonker, H., Richter, C. and Schwalbe, H., manuscript in preparation) may contribute to the decreased stability of the uCACGg hairpin.

A further increase of the temperature to 325 K in the experiment and to 360 K in the simulation destabilizes all residues of the uCACGg hairpin (Figure 4 bottom right). None of the hydrogen bonds are detectable anymore in the NMR spectra (data not shown). The MD trajectory, too, displays enlarged distances and unfavorable arrangements between the hydrogen bond donors and acceptors in the loop. Moreover, the adjacent wobble base pair (U13–G18) is strongly destabilized in comparison to the canonical base pair (C5–G10) in the cUUCGg hairpin at this elevated temperature.

## CONCLUSIONS

In summary, we have presented a detailed investigation of the thermal unfolding of two RNA tetraloops, which exhibit different stabilities although they are closely related in their ground state conformation. The large temperature range considered represents a challenge for both NMR and MD. On one hand, MD simulations are known to yield too high melting temperatures, as standard biomolecular force fields have been parameterized to reproduce room-temperature properties. To obtain the physically correct simulation temperature, the temperature was therefore rescaled according to a previously performed replica exchange MD simulation. For the NMR, on the other hand, the large temperature range represents a stringent test whether the parameters (effective bond length, CSA) have been chosen correctly to facilitate an appropriate and physically meaningful model-free analysis over the entire temperature range. It has been found that more recent data suggest a general increase for  $^{13}\text{C}$  CSAs of the aromatic sites in the nucleobase and in addition the conformation around the glycosidic torsion angle  $\chi$  (syn versus anti) needs to be considered. Such correction has a pronounced effect on the CSA that propagates into the order parameter analysis. Since the experimental determination of  $^{13}\text{C}$  CSA is dependent on the choice of the C–H bond length, the consistent treatment of zero-motion averaged bond lengths is necessary. Following the recommendations in the protein field (73), we and others (74) utilize longer bond lengths. These adjustments now allow determination and interpretation of experimental data that translates into a precision for  $S_{\text{NMR}}^2$  of a few percentages when using the model-free formalism.

Employing the thus established methods, we have presented a detailed temperature dependent NMR/MD study of the melting of the most abundant loop-motif in RNA. It has been found for both systems that the loss of the stacking interactions between the first and the third nucleobase in the loop region is the first event that destabilizes the hairpin. While the global fold is still preserved, this hot spot initiating unfolding was observed at 317 K for the uCACGg hairpin but only at an MD temperature of 360 K for the more stable cUUCGg hairpin. Increasing the temperature further, both NMR and MD elucidate the premelting of the uCACGg hairpin, which is accompanied by a loss of specific hydrogen bonds and stacking interactions. The study demonstrates the potential of joint MD/NMR studies to assess the dynamic properties of RNA secondary structure, which is ultimately required for a microscopic understanding of the various functions of RNA systems.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We would like to thank Oliver Ohlenschläger, Matthias Görlach for providing us with a sample of the 30mer RNA

and David Case, Ad Bax, Lucas Trantirek for sharing their data prior to publication. The work was supported by the Deutsche Forschungsgemeinschaft (DFG) (SFB 579: 'RNA-Ligand-Interactions'), the state of Hesse [Center for Biomolecular Magnetic Resonance (BMRZ)] and the Fonds der Chemischen Industrie (H.S. and G.S.). Funding to pay the Open Access publication charges for this article was provided by the DFG-funded SFB 579: 'RNA-Ligand-Interactions'.

*Conflict of interest statement.* None declared.

## REFERENCES

- Al-Hashimi, H.M. (2007) Beyond static structures of RNA by NMR: folding, refolding, and dynamics at atomic resolution. *Biopolymers*, **86**, 345–347.
- Wenter, P., Fürtig, B., Hainard, A., Schwalbe, H. and Pitsch, S. (2006) A caged uridine for the selective preparation of an RNA fold and determination of its refolding kinetics by real-time NMR. *Chem. Biochem.*, **7**, 417–420.
- Fürtig, B., Buck, J., Manoharan, V., Bermel, W., Jäschke, A., Wenter, P., Pitsch, S. and Schwalbe, H. (2007) Time-resolved NMR studies of RNA folding. *Biopolymers*, **27**, 360–383.
- Al-Hashimi, H.M. (2005) Dynamics-based amplification of RNA function and its characterization by using NMR spectroscopy. *Chem. Biochem.*, **6**, 1506–1519.
- Shajani, Z. and Varani, G. (2007) NMR studies of dynamics in RNA and DNA by  $^{13}\text{C}$  relaxation. *Biopolymers*, **86**, 348–359.
- Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity. *J. Am. Chem. Soc.*, **104**, 4546–4559.
- Lipari, G. and Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results. *J. Am. Chem. Soc.*, **104**, 4559–4570.
- Clore, G.M., Szabo, A., Bax, A., Kay, L.E., Driscoll, P.C. and Gronenborn, A.M. (1990) Deviations from the simple two-parameter model-free approach to the interpretation of nitrogen-15 nuclear magnetic relaxation of proteins. *J. Am. Chem. Soc.*, **112**, 4989–4991.
- Zhang, Q., Sun, X., Watt, E.D. and Al-Hashimi, H.M. (2006) Resolving the motional modes that code for RNA adaptation. *Science*, **311**, 653–656.
- Williamson, J.R. and Boxer, S.G. (1989) Multinuclear NMR studies of DNA hairpins. 1. Structure and dynamics of d(CGCGTTGTCGCG). *Biochemistry*, **28**, 2819–2831.
- Borer, P.N., LaPlante, S.R., Kumar, A., Zanatta, N., Martin, A., Hakkinen, A. and Levy, G.C. (1994)  $^{13}\text{C}$ -NMR relaxation in three DNA oligonucleotide duplexes: model-free analysis of internal and overall motion. *Biochemistry*, **33**, 2441–2450.
- Spielmann, H.P. (1998) Dynamics of a bis-intercalator DNA complex by  $^1\text{H}$ -detected natural abundance  $^{13}\text{C}$  NMR spectroscopy. *Biochemistry*, **37**, 16863–16876.
- Kojima, C., Ono, A., Kainosho, M. and James, T. (1998) DNA duplex dynamics: NMR relaxation studies of a decamer with uniformly  $^{13}\text{C}$ -labeled purine nucleotides. *J. Magn. Res.*, **135**, 310–333.
- Boisbouvier, J., Brutscher, B., Simorre, J.-P. and Marion, D. (1999)  $^{13}\text{C}$  spin relaxation measurements in RNA: sensitivity and resolution improvement using spin-state selective correlation experiments. *J. Biomol. NMR*, **14**, 241–252.
- Hall, K.B. and Tang, C. (1998)  $^{13}\text{C}$  Relaxation and dynamics of the purine bases in the iron responsive element RNA hairpin. *Biochemistry*, **37**, 9323–9332.
- Boisbouvier, J., Wu, Z., Ono, A., Kainosho, M. and Bax, A. (2003) Rotational diffusion tensor of nucleic acids from  $^{13}\text{C}$  NMR relaxation. *J. Biomol. NMR*, **27**, 133–142.
- Duchardt, E. and Schwalbe, H. (2005) Residue specific ribose and nucleobase dynamics of the cUUCGg RNA tetraloop motif by NMR  $^{13}\text{C}$  relaxation. *J. Biomol. NMR*, **32**, 295–308.

18. Shajani,Z. and Varani,G. (2005)  $^{13}\text{C}$  NMR relaxation studies of RNA base and ribose nuclei reveal a complex pattern of motions in the RNA binding site for human U1A protein. *J. Mol. Biol.*, **349**, 699–715.
19. Frenkel,D. and Smit,B. (1996) *Understanding Molecular Simulations: From Algorithms to Applications*. Academic Press, London.
20. van Gunsteren,W.F., Bakowies,D., Baron,R., Chandrasekhar,I., Christen,M., Daura,X., Gee,P., Geerke,D.P., Glättli,A. *et al.* (2006) Biomolecular modeling: goals, problems, perspectives. *Angew. Chem. Int. Ed. Engl.*, **45**, 4064–4092.
21. Levy,R.M., Karplus,M. and Wolynes,P.G. (1981) NMR relaxation parameters in molecules with internal motion: exact Langevin trajectory results compared with simplified relaxation models. *J. Am. Chem. Soc.*, **103**, 5998–6011.
22. Bruschweiler,R., Roux,B., Blackledge,M., Griesinger,C., Karplus,M. and Ernst,R.R. (1992) Influence of rapid intramolecular motion on NMR cross-relaxation rates – a molecular-dynamics study of antamanide in solution. *J. Am. Chem. Soc.*, **114**, 2289–2302.
23. Palmer,A.G. III. and Case,D. (1992) Molecular dynamics analysis of NMR relaxation in a zinc-finger peptide. *J. Am. Chem. Soc.*, **114**, 9059–9067.
24. Chatfield,D.C., Szabo,A. and Brooks,B.R. (1998) Molecular dynamics of staphylococcal nuclease: Comparison of simulation with  $^{15}\text{N}$  and  $^{13}\text{C}$  NMR relaxation data. *J. Am. Chem. Soc.*, **120**, 5301–5311.
25. Prompers,J.J. and Bruschweiler,R. (2001) Reorientational eigenmode dynamics: A combined MD/NMR relaxation analysis method for flexible parts in globular proteins. *J. Am. Chem. Soc.*, **123**, 7305–7313.
26. Peter,C., Daura,X. and van Gunsteren,W.F. (2001) Calculation of NMR-relaxation parameters for flexible molecules from molecular dynamics simulations. *J. Biomol. NMR*, **20**, 297–310.
27. Case,D.A. (2002) Molecular dynamics and NMR spin relaxation in proteins. *Acc. Chem. Res.*, **35**, 325–331.
28. Lange,O.F., Grubmüller,H. and de Groot,B.L. (2005) Molecular dynamics simulations of protein G challenge NMR-derived correlated backbone motions. *Angew. Chem. Int. Ed. Engl.*, **44**, 3394–3399.
29. Koplín,J., Mu,Y., Richter,C., Schwalbe,H. and Stock,G. (2005) Structure and dynamics of a RNA tetraloop: a joint molecular dynamics and NMR study. *Structure*, **13**, 1255–1267.
30. Showalter,S.A., Baker,N.A., Tang,C. and Hall,K.B. (2005) Iron responsive element RNA flexibility described by NMR anisotropic reorientational eigenmode dynamics. *J. Biomol. NMR*, **32**, 179–193.
31. Deng,H.-J. and Cieplak,P. (2007) Molecular dynamics and free energy study of the conformational equilibria in the UUUU RNA hairpin. *J. Chem. Theory Comput.*, **3**, 1435–1450.
32. Villa,A. and Stock,G. (2006) What NMR relaxation can tell us about the internal motion of an RNA hairpin: a molecular dynamics simulation study. *J. Chem. Theory Comput.*, **2**, 1228–1236.
33. Allain,F.H.-T. and Varani,G. (1995) Structure of the P1 helix m group I self-splicing intron. *J. Mol. Biol.*, **50**, 333–353.
34. Akke,M., Fiala,R., Jiang,F., Patel,D. and Palmer,A.G. III. (1997) Base dynamics in a UUCG tetraloop RNA hairpin characterized by  $^{15}\text{N}$  spin relaxation: correlations with structure and stability. *RNA*, **3**, 702–709.
35. Williams,J.D. and Hall,K.B. (2000) Experimental and theoretical studies of the effects of deoxyribose substitutions on the stability of the UUCG tetraloop. *J. Mol. Biol.*, **297**, 251–265.
36. Williams,J.D. and Hall,K.B. (2000) Experimental and computational studies of the g(UUCG)c RNA tetraloop. *J. Mol. Biol.*, **297**, 1045–1061.
37. Miller,J.L. and Kollman,P.A. (1997) Theoretically studies of an exceptionally stable RNA tetraloop: observation of convergence from an incorrect NMR structure to the correct one using unrestrained molecular dynamics. *J. Mol. Biol.*, **270**, 436–450.
38. Nina,M. and Simonson,T. (2002) Molecular dynamics of the tRNA(Ala) acceptor stem: comparison between continuum reaction field and particle-mesh ewald electrostatic treatments. *J. Phys. Chem. B*, **106**, 3696–3705.
39. Ohlenschläger,O., Wöhnert,J., Bucci,E., Seitz,S., Hafner,S., Ramachandran,R., Zell,R. and Görlach,M. (2004) . The structure of the stemloop D subdomain of Coxsackievirus B3 sloverleaf RNA and its interaction with the proteinase 3C. *Structure (Camb.)*, **12**, 237–248.
40. Yang,D. and Kay,L.E. (1996) Contributions to conformational entropy arising from bond vector fluctuations measured from NMR-derived order parameters: application to protein folding. *J. Mol. Biol.*, **263**, 369–382.
41. Mandel,A.M., Akke,M. and Palmer,A.G. III. (1996) Dynamics of ribonuclease H: temperature dependence of motions on multiple time scales. *Biochemistry*, **35**, 16009–16023.
42. Idiyatullin,D., Nesmelova,I., Daragan,V.A. and Mayo,K.H. (2003) Heat capacities and a snapshot of the energy landscape in protein GB1 from the pre-denaturation temperature dependence of backbone NH nanosecond fluctuations. *J. Mol. Biol.*, **325**, 149–162.
43. Wang,T., Cai,S. and Zwietering,E.R.P. (2003) Temperature dependence of anisotropic protein backbone dynamics. *J. Am. Chem. Soc.*, **125**, 8639–8643.
44. Massi,F. and Palmer,A.G. III. (2003) Temperature dependence of NMR order parameters and protein dynamics. *J. Am. Chem. Soc.*, **125**, 11158–11159.
45. Johnson,E., Palmer,A.G. III. and Rance,M. (2007) Temperature dependence of the NMR generalized order parameter. *Proteins*, **66**, 796–803.
46. Stoldt,M., Wöhnert,J., Ohlenschläger,O., Görlach,M. and Brown,L.R. (1999) The NMR structure of the 5S rRNA E-domain-protein L25 complex shows preformed and induced recognition. *EMBO J.*, **18**, 6508–6521.
47. Kay,L.E., Nicholson,L.K., Delaglio,F., Bax,A. and Torchia,D.A. (1992) Pulse sequences for the removal of the effects of cross-correlation between dipolar and chemical shift anisotropy relaxation mechanism on the measurement of heteronuclear  $T_1$  and  $T_2$  values in proteins. *J. Magn. Reson.*, **97**, 359–375.
48. Farrow,N.A., Muhandiram,R., Singer,A.U., Pascal,S.M., Kay,C.M., Gish,G., Shoelson,S.E., Pawson,T., Forman-Kay,J.D. and Kay,L.E. (1994) Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by  $^{15}\text{N}$  NMR relaxation. *Biochemistry*, **33**, 5984–6003.
49. Mulder,F.A.A., de Graaf,R.A., Kaptein,R. and Boelens,R. (1998) An off-resonance rotating frame relaxation experiment for the investigation of macromolecular dynamics using adiabatic rotations. *J. Magn. Reson.*, **131**, 351–357.
50. Korzhnev,D.M., Skrynnikov,N.R., Millet,O., Torchia,D.A. and Kay,L.E. (2002) An NMR experiment for the accurate measurement of heteronuclear spin-lock relaxation rates. *J. Am. Chem. Soc.*, **124**, 10743–10753.
51. Yamazaki,T., Muhandiram,R. and Kay,L.E. (1994) NMR experiments for the measurement of carbon relaxation properties in highly enriched, uniformly  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled proteins: application to  $^{13}\text{C}_\alpha$  carbons. *J. Am. Chem. Soc.*, **116**, 8266–8278.
52. Mandel,A.M., Akke,M. and Palmer,A.G. III (1995) Backbone dynamics of Escherichia coli ribonuclease HI: correlations with structure and function in an active enzyme. *J. Mol. Biol.*, **246**, 144–163.
53. d’Auvergne,E.J. and Gooley,P.R. (2003) The use of model selection in the model-free analysis of protein dynamics. *J. Biomol. NMR*, **25**, 25–39.
54. Chen,J., Brooks,C.L. III. and Wright,P.E. (2004) Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation. *J. Biomol. NMR*, **29**, 243–257.
55. Garcia de la Torre,J., Huertas,M.L. and Carrasco,B. (2000) HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. *J. Magn. Reson.*, **147**, 138–146.
56. Berendsen,H.J.C., van der Spoel,D. and van Drunen,R. (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Comm.*, **91**, 43–56.
57. Lindahl,E., Hess,B. and van der Spoel,D. (2001) GROMACS 3.0: A package for molecular simulation and trajectory analysis. *J. Mol. Model.*, **7**, 306–317.
58. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **117**, 5179–5197.

59. Cheatham, T., Cieplak, P. and Kollman, P.A. (1999) A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.*, **16**, 845–861.
60. Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L. (1983) Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **79**, 926–935.
61. Darden, T., York, D. and Pedersen, L. (1993) Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **98**, 10089–10092.
62. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, **81**, 3684–3690.
63. Miyamoto, S. and Kollman, P.A. (1992) SETTLE: an analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.*, **13**, 952–962.
64. Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.*, **18**, 1463–1472.
65. Sorin, E.J., Engelhardt, M.A., Herschlag, D. and Pande, V.S. (2002) RNA simulations: probing hairpin unfolding and the dynamics of a GNRA tetraloop. *J. Mol. Biol.*, **317**, 493–506.
66. Ernst, R.R., Bodenhausen, G. and Wokaun, A. (2004) *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Oxford University Press Inc., New York.
67. Ichiye, T. and Karplus, M. (1991) Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Struct. Funct. Genet.*, **11**, 205–217.
68. Garcia, A.E. (1992) Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.*, **68**, 2696–2699.
69. Amadei, A., Linssen, A.B.M. and Berendsen, H.J.C. (1993) Essential dynamics of proteins. *Proteins: Struct. Funct. Genet.*, **17**, 412–425.
70. Kroenke, C.D., Rance, M. and Palmer, A.G. III. (1999) Variability of the  $^{15}\text{N}$  chemical shift anisotropy in *Escherichia coli* ribonuclease H in solution. *J. Am. Chem. Soc.*, **121**, 10119–10125.
71. Loth, K., Pelupessy, P. and Bodenhausen, G. (2005) Chemical shift anisotropy tensors of carbonyl, nitrogen and amide proton nuclei in proteins through cross-correlated relaxation in NMR spectroscopy. *J. Am. Chem. Soc.*, **127**, 6062–6068.
72. Hall, J.B. and Fushman, D. (2006) Variability of the  $^{15}\text{N}$  chemical shielding tensors in the B3 domain of protein G from  $^{15}\text{N}$  relaxation measurements at several fields. Implications for backbone order parameters. *J. Am. Chem. Soc.*, **128**, 7855–7870.
73. Case, D.A. (1999) Calculations of NMR dipolar coupling strengths in model peptides. *J. Biomol. NMR*, **15**, 95–102.
74. Ying, J., Grishaev, A., Bryce, D.L. and Bax, A. (2006) Chemical shift tensors of protonated base carbons in helical RNA and DNA from NMR relaxation and liquid crystal measurements. *J. Am. Chem. Soc.*, **128**, 11443–11454.
75. Sychrovsky, V., Muller, N., Schneider, B., Smrecki, V., Spirko, V., Sponer, J. and Trantirek, L. (2005) Sugar pucker modulates the cross-correlated relaxation rates across the glycosidic bond in DNA. *J. Am. Chem. Soc.*, **127**, 14663–14667.
76. Dejaegere, A.P. and Case, D.A. (1998) Density functional study of ribose and deoxyribose chemical shifts. *J. Phys. Chem. A*, **102**, 5280–5289.
77. Fiala, R., Czernek, J. and Sklenar, V. (2000) Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *J. Biomol. NMR*, **16**, 291–302.
78. Stueber, D. and Grant, D.M. (2002)  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shift tensors in adenosine, guanosine dihydrate, 2'-deoxythymidine, and cytidine. *J. Am. Chem. Soc.*, **124**, 10539–10551.
79. Bryce, D.L., Grishaev, A. and Bax, A. (2005) Measurement of ribose carbon chemical shift tensors for A-form RNA by liquid crystal NMR spectroscopy. *J. Am. Chem. Soc.*, **127**, 7387–7396.
80. Fernandes, M.X., Ortega, A., Martinez, M.C.L. and Garcia de la Torre, J. (2002) Calculation of hydrodynamic properties of small nucleic acids from their atomic structure. *Nucleic Acids Res.*, **30**, 1782–1788.
81. Garcia, A.E. and Onuchic, J.N. (2003) Folding a protein in a computer: an atomic description of the folding/unfolding of protein A. *Proc. Natl Acad. Sci. USA*, **100**, 13898–13903.
82. Sorin, E.J., Rhe, Y.M., Nakatani, B.J. and Pande, V.S. (2003) Insights into nucleic acid conformational dynamics from massively parallel stochastic simulations. *Biophys. J.*, **85**, 790–803.
83. Zhou, R. (2003) Trp-cage: folding free energy landscape in explicit water. *Proc. Natl Acad. Sci. USA*, **100**, 13280–13285.
84. Nguyen, P.H., Stock, G., Mittag, E., Hu, C.-K. and Li, M.-S. (2005) Free energy landscape and folding mechanism of a  $\beta$ -hairpin in explicit water: A replica exchange molecular dynamic study. *Proteins*, **61**, 795–808.
85. Villa, A., Widjajakusuma, E. and Stock, G. (2008) Molecular dynamics simulation of the structure, dynamics and thermostability of the RNA hairpins uCACGg and cUUCGg. *J. Phys. Chem. B*, **112**, 134–142.
86. Rinnenthal, J., Richter, C., Ferner, J., Duchardt, E. and Schwalbe, H. (2007) Quantitative  $\Gamma$ -HCNCH: Determination of the glycosidic torsion angle  $\chi$  in RNA oligonucleotides from the analysis of CH dipolar cross-correlated relaxation in solution NMR spectroscopy. *J. Biomol. NMR*, **39**, 17–29.

**Supplementary Material to Manuscript:**

# NMR and MD Studies of the Temperature- Dependent Dynamics of RNA YNMG-Tetraloops

*Jan Ferner<sup>1</sup>, Alessandra Villa<sup>2</sup>, Elke Duchardt<sup>1</sup>, Elisabeth Widjajakusuma<sup>2</sup>, Gerhard Stock<sup>2</sup>,  
Harald Schwalbe<sup>1\*</sup>*

<sup>1</sup>Institut für Organische Chemie und Chemische Biologie, Center for Biomolecular Magnetic Resonance, <sup>2</sup>Institut für Physikalische und Theoretische Chemie, Johann Wolfgang Goethe-Universität Frankfurt, Max-von-Laue-Str. 7, 60438 Frankfurt/M, Germany.

AUTHOR E-MAIL: [schwalbe@nmr.uni-frankfurt.de](mailto:schwalbe@nmr.uni-frankfurt.de)

## METHODS

### Selection of Motional Model for Analysis of Heteronuclear Relaxation Rates

The model-free analysis by Lipari and Szabo (Lipari 1982a and 1982b, Clore 1990) allows to convert the autocorrelated relaxation rates  $R_1$ ,  $R_2$  ( $R_{1\rho}$ ) and the steady-state heteronuclear NOE into parameters representing dynamical information. These parameters are the order parameter  $S^2$  of the investigated  $^{13}\text{C}^1\text{H}$ -vector which can be separated in a fast ( $S_f^2$ ) and a slow ( $S_s^2$ ) contribution, the internal correlation time  $\tau_e$  and an  $R_{ex}$ -term. There are 5 different combinations, called models, possible to fit the relaxation data.

Model 1 only includes the order parameter  $S^2$ . Internal motions are assumed to be faster than about 10 ps and no exchange term contributes to the relaxation rates. Two parameters are implied in the Models 2 and 3. Model 2 yields an internal correlation time  $\tau_e$  on the picosecond timescale, slower than 10 ps but faster than the global correlation time  $\tau_c$  on an order of nanoseconds. For Model 3, an exchange term  $R_{ex}$  is extracted besides the order parameter. This additional parameter can indicate motions on the micro- to millisecond timescale. Both of these parameters,  $\tau_e$  and  $R_{ex}$ , are implemented in the Model 4. The Model 5 results in a separation of the order parameter in a fast and a slow component. In addition an internal correlation time of the slower motion is obtained.

Which model should be selected for each measured vector is decided by criteria. The mostly used criterion in the past for the model-free analysis (Mandel 1995) bases on a critical value  $\alpha$ , which defines the percentage of the Monte-Carlo numerical simulations, the fitted results have to accord to. The critical value  $\alpha$  is compared to a  $\chi^2$  value which is determined from the deviation of experimental and back-calculated relaxation rates and their error (equation S1). A F-statistic test is performed to distinguish between easier and more complex models where the added parameter has to be significantly different from 0. The three parameter models (Model 4 and 5) need to result in a  $\chi^2$  value of 0 since there are no degrees of freedom left if three relaxation rates are inputted.  $\chi^2$  is calculated with this equation:

$$\chi^2 = \sum_{i=1}^N \sum_{j=1}^{M_i} \frac{(R_{ij} - R'_{ij})^2}{\sigma_{ij}^2} \quad \text{equation S1,}$$

where  $N$  is the total number of nuclear spins,  $M_i$  is the number of experimental relaxation parameters of the  $i^{\text{th}}$  spin,  $R_{ij}$  is the  $j^{\text{th}}$  experimental relaxation parameter,  $R'_{ij}$  is the  $j^{\text{th}}$  theoretical relaxation parameter and  $\sigma_{ij}$  is the experimental uncertainty in the  $j^{\text{th}}$  relaxation parameter.

d'Auvergne and Gooley pointed out in 2003 (d'Auvergne 2003) that the primarily used model selection method in a model-free analysis by Palmer and co-workers (Mandel 1995) may underestimate the influence of the possible fitting model. The result of their analysis was that probabilistic approaches like the Akaike's Information Criteria (AIC) and the Bayesian Information Criteria (BIC) yield a more realistic model selection. The old method based on step-up hypothesis testing often oversimplifies and underfits the model that could be extracted from the relaxation data. Wright and co-workers also investigated the model selection protocols and arrived at the conclusion that the BIC provides the best model-free results (Chen 2004). The Bayesian Information Criteria (BIC) is calculated with the following equation:

$$BIC = \chi^2 + k \ln M \quad \text{equation S2,}$$

where  $\chi^2$  is determined by equation S2,  $k$  is the number of independent parameters in the model and  $M$  is the dimension of the sample data set. This dimension accords to the relaxation data input which is in our analysis 3 ( $R_1$ ,  $R_{1\rho}$  and *hetNOE*). For Model 1 the value of  $k$  is 1, for the Models 2 and 3 it is a value of 2 and for the Models 4 and 5 it is 3.

A comparison of the model selection techniques, the commonly used by Palmer and co-workers (Mandel 1995), AIC and BIC, with the relaxation data of the 14mer cUUCGg hairpin at 298 K reveals only small differences in the resultant order parameters and underlying models. While AIC and BIC mostly yield in the same selected models, the original method only has 1 or 2 data sets out of 14 selected with a lower number of parameters.



TABLES

**Table S1:** Autocorrelated relaxation parameters of the 14-nt cUUCGg RNA of the carbon nuclei C6/C8/C2 and C1' at 317 K and 600 MHz.

	$R_1$ [ $s^{-1}$ ]		$R_2$ [ $s^{-1}$ ] <sup>a</sup>		<i>hetNOE</i>	
	C6/C8	C1'	C6/C8	C1'	C6/C8	C1'
G1	$3.778 \pm 0.076^b$	-	$10.62 \pm 0.21$	-	$1.219 \pm 0.037$	-
G2	$3.921 \pm 0.078$	$3.370 \pm 0.067$	$21.08 \pm 1.05$	$9.70 \pm 0.20$	$1.161 \pm 0.035$	$1.161 \pm 0.035$
C3	$5.082 \pm 0.102$	$3.518 \pm 0.070$	$12.95 \pm 0.26$	$8.02 \pm 0.16$	$1.149 \pm 0.034$	$1.177 \pm 0.035$
A4 <sup>c</sup>	$4.110 \pm 0.082$	$3.479 \pm 0.070$	$10.96 \pm 0.22$	$8.05 \pm 0.16$	$1.170 \pm 0.035$	$1.190 \pm 0.036$
	<i><math>4.408 \pm 0.071</math></i>		<i><math>15.92 \pm 0.41</math></i>		<i><math>1.136 \pm 0.019</math></i>	
C5	$5.044 \pm 0.101$	$3.533 \pm 0.071$	$12.89 \pm 0.26$	$7.79 \pm 0.16$	$1.165 \pm 0.035$	$1.212 \pm 0.036$
U6	$4.685 \pm 0.124$	$3.522 \pm 0.070$	$13.10 \pm 0.26$	$7.81 \pm 0.16$	$1.160 \pm 0.035$	$1.205 \pm 0.036$
U7	$3.607 \pm 0.072$	$3.222 \pm 0.064$	$10.90 \pm 0.22$	$7.50 \pm 0.15$	$1.180 \pm 0.035$	$1.207 \pm 0.036$
C8	$4.552 \pm 0.091$	$3.300 \pm 0.066$	$12.35 \pm 0.25$	$8.52 \pm 0.20$	$1.161 \pm 0.035$	$1.195 \pm 0.036$
G9	$3.576 \pm 0.084$	$3.289 \pm 0.066$	$10.82 \pm 0.22$	$8.80 \pm 0.18$	$1.187 \pm 0.036$	$1.230 \pm 0.037$
G10	$3.955 \pm 0.079$	$3.456 \pm 0.192$	$11.30 \pm 0.23$	$10.58 \pm 0.92$	$1.156 \pm 0.035$	$1.137 \pm 0.100$
U11	$5.091 \pm 0.102$	$3.497 \pm 0.070$	$12.78 \pm 0.26$	$8.08 \pm 0.16$	$1.155 \pm 0.035$	$1.229 \pm 0.037$
G12	$3.996 \pm 0.088$	$3.449 \pm 0.069$	$15.95 \pm 0.62$	$8.32 \pm 0.17$	$1.174 \pm 0.035$	$1.188 \pm 0.036$
C13	$4.860 \pm 0.097$	$3.516 \pm 0.070$	$13.25 \pm 0.27$	$7.68 \pm 0.15$	$1.152 \pm 0.035$	$1.208 \pm 0.036$
C14	$4.922 \pm 0.098$	$3.208 \pm 0.064$	$12.38 \pm 0.25$	$6.55 \pm 0.13$	$1.165 \pm 0.035$	$1.291 \pm 0.082$

<sup>a</sup>  $R_2$ -rates have been extracted from experimental  $R_{1\rho}$ -data as described in equation 1. <sup>b</sup> Errors have been derived from the curve-fitting procedure using the program Sparky 3.112 (University of California, San Francisco, USA) with a minimum error set to 2 % for the  $R_1$ - and  $R_2$ -rates and 3 % for the  $^{13}\text{C}\{-^1\text{H}\}$  steady-state NOE. <sup>c</sup> The adenine C2 relaxation rates are written in italics.

**Table S2:** Autocorrelated relaxation parameters of the 30-nt uCACGg RNA of the carbon nuclei C6/C8/C2 and C1' at 298 K and 600 MHz.

	$R_1$ [s <sup>-1</sup> ]		$R_2$ [s <sup>-1</sup> ] <sup>a</sup>		<i>hetNOE</i>	
	C6/C8/C2	C1'	C6/C8/C2	C1'	C6/C8/C2	C1'
G1	1.461 ± 0.077		34.24 ± 3.12		1.170 ± 0.061	1.139 ± 0.057
G2	1.586 ± 0.060	1.110 ± 0.033	37.70 ± 1.40	31.52 ± 1.22	1.094 ± 0.055	1.204 ± 0.060
C3	1.972 ± 0.059	1.064 ± 0.032	52.51 ± 3.48	31.96 ± 0.96	1.108 ± 0.055	1.128 ± 0.056
A4 <sup>c</sup>	1.585 ± 0.048	1.070 ± 0.032	36.38 ± 2.44	32.31 ± 0.98	1.066 ± 0.053	1.109 ± 0.055
	<i>1.719 ± 0.052</i>		<i>37.01 ± 1.62</i>		<i>1.078 ± 0.054</i>	
C5	1.820 ± 0.055	1.060 ± 0.044	50.64 ± 2.48	32.82 ± 0.98	1.116 ± 0.056	1.122 ± 0.056
U6						
C7	1.894 ± 0.062	1.004 ± 0.030	53.56 ± 2.78	33.82 ± 1.01	1.160 ± 0.058	1.137 ± 0.057
U8	1.993 ± 0.060	1.098 ± 0.033	47.39 ± 1.81	29.21 ± 0.88	1.103 ± 0.055	1.083 ± 0.054
G9	1.596 ± 0.053	1.155 ± 0.050	40.32 ± 1.39	44.37 ± 2.74	1.123 ± 0.056	1.086 ± 0.054
G10	1.533 ± 0.057	1.053 ± 0.032	35.22 ± 1.65	33.41 ± 1.09	1.091 ± 0.055	1.114 ± 0.056
U11	1.995 ± 0.085	1.110 ± 0.033	46.80 ± 2.03	30.52 ± 0.92	1.087 ± 0.054	1.130 ± 0.056
A12	1.625 ± 0.049	1.130 ± 0.034	36.01 ± 1.19	28.49 ± 0.86	1.093 ± 0.055	1.145 ± 0.057
	<i>1.712 ± 0.051</i>		<i>36.03 ± 1.29</i>		<i>1.079 ± 0.054</i>	
U13	1.991 ± 0.060	1.149 ± 0.034	44.48 ± 2.46	30.94 ± 0.94	1.105 ± 0.055	1.142 ± 0.057
C14	2.034 ± 0.071	1.133 ± 0.034	45.21 ± 1.93	32.56 ± 1.35	1.149 ± 0.057	1.189 ± 0.059
A15	1.488 ± 0.045	1.381 ± 0.041	22.06 ± 0.66	23.74 ± 2.24	1.300 ± 0.065	1.229 ± 0.061
	<i>1.461 ± 0.044</i>		<i>27.03 ± 0.56</i>		<i>1.212 ± 0.061</i>	
C16	1.878 ± 0.056	1.202 ± 0.046	47.54 ± 1.80	30.59 ± 1.22	1.149 ± 0.057	1.185 ± 0.059
G17	1.616 ± 0.048	1.273 ± 0.038	42.75 ± 2.65	32.11 ± 0.96		1.179 ± 0.059
G18	1.624 ± 0.049	1.100 ± 0.033	34.24 ± 1.03	32.19 ± 0.98	1.131 ± 0.057	1.179 ± 0.059
U19	1.968 ± 0.059	1.250 ± 0.038	43.05 ± 2.33	31.56 ± 0.95	1.077 ± 0.054	1.098 ± 0.055
A20	1.624 ± 0.049	1.061 ± 0.032	35.19 ± 2.51	30.21 ± 0.91	1.119 ± 0.056	1.113 ± 0.056
	<i>1.709 ± 0.051</i>		<i>36.85 ± 1.53</i>		<i>1.078 ± 0.054</i>	
C21	1.922 ± 0.067	1.044 ± 0.031	48.56 ± 1.94	33.67 ± 1.01	1.100 ± 0.055	1.130 ± 0.057
C22	1.903 ± 0.057	1.023 ± 0.031	49.79 ± 2.65	33.07 ± 0.99	1.134 ± 0.057	1.127 ± 0.059
U23						
U24	1.954 ± 0.059	1.163 ± 0.035	45.50 ± 2.23	27.74 ± 0.84	1.057 ± 0.053	1.121 ± 0.056
U25	1.922 ± 0.058		44.84 ± 1.35		1.116 ± 0.056	
G26	1.499 ± 0.045	1.067 ± 0.058	36.30 ± 1.09	65.51 ± 6.52	1.019 ± 0.051	1.205 ± 0.060
U27	1.932 ± 0.058		47.89 ± 2.12		1.072 ± 0.054	
G28	1.523 ± 0.074	1.043 ± 0.031	37.47 ± 2.44	35.97 ± 1.08	1.117 ± 0.056	1.190 ± 0.060
U29	1.985 ± 0.060	1.222 ± 0.037	42.71 ± 1.41	30.24 ± 1.08	1.113 ± 0.056	1.176 ± 0.059
C30	1.856 ± 0.084	1.100 ± 0.033	61.63 ± 4.13	26.00 ± 1.24	1.184 ± 0.074	1.233 ± 0.073

<sup>a</sup>  $R_2$ -rates have been extracted from experimental  $R_{1\rho}$ -data as described in equation 1. <sup>b</sup> Errors have been derived from the curve-fitting procedure using the program Sparky 3.112 (University of California, San Francisco, USA) with a minimum error set to 3 % for the  $R_1$ - and  $R_2$ -rates and 5 % for the <sup>13</sup>C-<sup>1</sup>H} steady-state NOE. <sup>c</sup> The adenine C2 relaxation rates are written in italics.

**Table S3:** Autocorrelated relaxation parameters of the 30-nt uCACGg RNA of the carbon nuclei C6/C8/C2 and C1' at 317 K and 600 MHz.

	$R_1$ [s <sup>-1</sup> ]		$R_2$ [s <sup>-1</sup> ] <sup>a</sup>		<i>hetNOE</i>	
	C6/C8/C2	C1'	C6/C8/C2	C1'	C6/C8/C2	C1'
G1	2.280 ± 0.068		26.23 ± 0.94		1.246 ± 0.062	1.355 ± 0.068
G2	2.284 ± 0.069	1.611 ± 0.048	24.39 ± 0.73	19.16 ± 0.57	1.118 ± 0.056	1.193 ± 0.060
C3	2.776 ± 0.083	1.653 ± 0.050	32.44 ± 0.97	20.40 ± 0.61	1.094 ± 0.055	1.172 ± 0.070
A4 <sup>c</sup>	2.240 ± 0.067	1.497 ± 0.045	25.04 ± 0.75	20.80 ± 0.62	1.131 ± 0.057	1.159 ± 0.058
	<i>2.361 ± 0.071</i>		<i>26.77 ± 0.80</i>		<i>1.116 ± 0.056</i>	
C5	2.870 ± 0.086	1.532 ± 0.046	30.45 ± 0.91	20.88 ± 0.65	1.141 ± 0.057	1.138 ± 0.057
U6						
C7	2.814 ± 0.084	1.584 ± 0.048	29.97 ± 1.04	19.71 ± 0.59	1.160 ± 0.058	1.185 ± 0.059
U8	2.845 ± 0.085	1.554 ± 0.047	27.78 ± 0.83	19.16 ± 0.57	1.110 ± 0.055	1.203 ± 0.060
G9	2.405 ± 0.072	1.665 ± 0.050	23.28 ± 0.71	23.46 ± 0.70	1.206 ± 0.060	1.162 ± 0.066
G10	2.308 ± 0.069	1.545 ± 0.046	23.37 ± 0.72	19.70 ± 0.59	1.104 ± 0.055	1.132 ± 0.057
U11	2.809 ± 0.084	1.628 ± 0.049	29.28 ± 0.88	20.66 ± 0.98	1.048 ± 0.052	1.157 ± 0.058
A12	2.272 ± 0.068	1.613 ± 0.048	24.06 ± 0.72	20.91 ± 0.79	1.146 ± 0.057	1.181 ± 0.059
	<i>2.461 ± 0.074</i>		<i>25.24 ± 0.76</i>		<i>1.058 ± 0.053</i>	
U13	2.908 ± 0.087	1.661 ± 0.050	28.88 ± 0.87	20.91 ± 0.63	1.143 ± 0.057	1.197 ± 0.060
C14	2.803 ± 0.084	1.694 ± 0.051	33.78 ± 1.18	20.48 ± 0.77	1.127 ± 0.056	1.204 ± 0.080
A15	1.917 ± 0.058	1.864 ± 0.056	13.74 ± 0.41	18.09 ± 0.58	1.293 ± 0.065	1.238 ± 0.062
	<i>1.896 ± 0.057</i>		<i>18.24 ± 0.55</i>		<i>1.308 ± 0.065</i>	
C16	2.833 ± 0.085	1.702 ± 0.051	30.12 ± 1.07	19.37 ± 0.58	1.136 ± 0.057	1.336 ± 0.067
G17	2.233 ± 0.067	1.867 ± 0.056	36.35 ± 1.42	24.83 ± 0.75		1.195 ± 0.100
G18	2.341 ± 0.070	1.603 ± 0.048	21.84 ± 0.66	19.71 ± 0.64	1.166 ± 0.058	1.268 ± 0.126
U19	2.840 ± 0.085	1.579 ± 0.047	28.00 ± 0.84	20.77 ± 0.62	1.108 ± 0.055	1.150 ± 0.058
A20	2.338 ± 0.070	1.573 ± 0.047	22.99 ± 0.69	20.76 ± 0.62	1.104 ± 0.055	1.126 ± 0.056
	<i>2.469 ± 0.074</i>		<i>26.41 ± 0.79</i>		<i>1.073 ± 0.054</i>	
C21	2.892 ± 0.087	1.590 ± 0.048	29.99 ± 0.90	19.62 ± 0.59	1.101 ± 0.055	1.134 ± 0.057
C22	2.925 ± 0.088	1.555 ± 0.047	35.30 ± 1.06	20.21 ± 0.70	1.125 ± 0.056	1.171 ± 0.082
U23						
U24	2.807 ± 0.084	1.666 ± 0.050	27.55 ± 0.83	19.35 ± 0.70	1.130 ± 0.057	1.145 ± 0.057
U25						
G26	2.290 ± 0.069	1.673 ± 0.050	23.82 ± 0.71	29.63 ± 0.89	1.174 ± 0.059	1.185 ± 0.059
U27						
G28	2.243 ± 0.067	1.542 ± 0.046	25.02 ± 0.75	19.66 ± 0.59	1.081 ± 0.054	1.189 ± 0.059
U29	2.862 ± 0.086	1.537 ± 0.046	28.70 ± 0.86	21.12 ± 0.63	1.037 ± 0.052	1.171 ± 0.059
C30	2.670 ± 0.080	1.515 ± 0.045	32.42 ± 0.98	16.30 ± 0.61	1.163 ± 0.058	1.302 ± 0.065

<sup>a</sup>  $R_2$ -rates have been extracted from experimental  $R_{1\rho}$ -data as described in equation 1. <sup>b</sup> Errors have been derived from the curve-fitting procedure using the program Sparky 3.112 (University of California, San Francisco, USA) with a minimum error set to 3 % for the  $R_1$ - and  $R_2$ -rates and 5 % for the <sup>13</sup>C-<sup>1</sup>H} steady-state NOE. <sup>c</sup> The adenine C2 relaxation rates are written in italics.

**Table S4:** Autocorrelated relaxation parameters of the 30-nt uCACGg RNA of the carbon nuclei C6/C8/C2 and C1' at 325 K and 600 MHz.

	$R_1$ [s <sup>-1</sup> ]		$R_2$ [s <sup>-1</sup> ] <sup>a</sup>		<i>hetNOE</i>	
	C6/C8/C2	C1'	C6/C8/C2	C1'	C6/C8/C2	C1'
G1	2.384 ± 0.148	1.754 ± 0.053	23.94 ± 3.73	16.97 ± 0.51	1.200 ± 0.062	1.348 ± 0.067
G2	2.455 ± 0.076	1.750 ± 0.052	19.01 ± 1.05	18.29 ± 0.55	1.207 ± 0.079	1.221 ± 0.061
C3	3.299 ± 0.099	1.716 ± 0.051	25.69 ± 0.81	19.04 ± 0.57	1.112 ± 0.056	1.222 ± 0.061
A4 <sup>c</sup>	2.734 ± 0.082	1.943 ± 0.058	20.60 ± 0.62	16.39 ± 0.49	1.183 ± 0.059	1.207 ± 0.060
	<i>2.946 ± 0.088</i>		<i>20.63 ± 0.62</i>		<i>1.114 ± 0.056</i>	
C5	2.915 ± 0.087	1.594 ± 0.048	28.33 ± 1.43	18.22 ± 0.55	1.132 ± 0.057	1.212 ± 0.061
U6	3.262 ± 0.098	1.892 ± 0.057	26.87 ± 0.81	17.36 ± 0.52	1.124 ± 0.056	1.200 ± 0.060
C7	3.054 ± 0.092	1.836 ± 0.063	25.93 ± 0.78	17.26 ± 0.52	1.182 ± 0.071	1.209 ± 0.060
U8	3.413 ± 0.102	2.018 ± 0.061	24.45 ± 0.73	15.37 ± 0.46	1.146 ± 0.057	1.224 ± 0.061
G9	2.471 ± 0.074	1.935 ± 0.058	21.05 ± 0.82	19.50 ± 0.58	1.164 ± 0.098	1.157 ± 0.058
G10	2.527 ± 0.076	1.733 ± 0.052	21.18 ± 0.82	17.01 ± 0.87	1.097 ± 0.055	1.180 ± 0.059
U11	3.329 ± 0.100	1.950 ± 0.059	25.32 ± 0.76	16.18 ± 0.49	1.124 ± 0.056	1.183 ± 0.059
A12	2.787 ± 0.084	1.941 ± 0.058	19.93 ± 0.60	15.70 ± 0.47	1.136 ± 0.057	1.200 ± 0.060
	<i>2.939 ± 0.088</i>		<i>21.09 ± 0.63</i>		<i>1.121 ± 0.056</i>	
U13	3.349 ± 0.100	2.010 ± 0.060	27.00 ± 0.81	16.10 ± 0.48	1.166 ± 0.058	1.227 ± 0.061
C14	3.259 ± 0.098	1.973 ± 0.059	21.82 ± 0.65	16.63 ± 0.55	1.162 ± 0.058	1.251 ± 0.063
A15	2.166 ± 0.065	2.171 ± 0.065	11.10 ± 0.33	13.43 ± 0.40	1.293 ± 0.065	1.268 ± 0.063
	<i>2.188 ± 0.066</i>		<i>11.76 ± 0.35</i>		<i>1.262 ± 0.063</i>	
C16	3.178 ± 0.095	1.967 ± 0.067	19.52 ± 0.59	16.04 ± 1.48	1.212 ± 0.061	1.325 ± 0.072
G17	2.028 ± 0.071	2.060 ± 0.062	32.12 ± 2.73	23.39 ± 0.81	1.223 ± 0.078	1.288 ± 0.064
G18	2.793 ± 0.105	1.835 ± 0.055	16.58 ± 1.00	15.53 ± 1.00	1.271 ± 0.078	1.251 ± 0.079
U19	3.638 ± 0.132	1.899 ± 0.057	25.01 ± 1.20	16.33 ± 0.49	1.137 ± 0.057	1.171 ± 0.059
A20	2.824 ± 0.085	1.956 ± 0.059	18.98 ± 0.57	15.61 ± 0.47	1.168 ± 0.058	1.179 ± 0.059
	<i>2.939 ± 0.088</i>		<i>21.50 ± 0.64</i>		<i>1.106 ± 0.055</i>	
C21	3.097 ± 0.093	1.807 ± 0.054	26.73 ± 1.06	16.80 ± 0.53	1.084 ± 0.054	1.191 ± 0.060
C22	3.113 ± 0.093	1.753 ± 0.053	26.69 ± 1.01	17.01 ± 0.54	1.132 ± 0.057	1.138 ± 0.057
U23	3.287 ± 0.099	1.892 ± 0.057	26.00 ± 0.78	17.36 ± 0.52	1.173 ± 0.059	1.200 ± 0.060
U24	3.349 ± 0.100	2.057 ± 0.062	24.32 ± 0.73	14.94 ± 0.45	1.194 ± 0.060	1.228 ± 0.061
U25	3.277 ± 0.098	2.026 ± 0.061	25.20 ± 0.76	16.86 ± 0.51	1.154 ± 0.058	1.218 ± 0.061
G26	2.152 ± 0.078	1.777 ± 0.053	19.95 ± 2.16	21.33 ± 0.77	1.114 ± 0.093	1.209 ± 0.060
U27	3.277 ± 0.098	1.864 ± 0.056	25.20 ± 0.76	18.65 ± 0.56	1.154 ± 0.058	1.248 ± 0.062
G28	2.478 ± 0.074	1.637 ± 0.066	22.23 ± 1.62	16.90 ± 0.73	1.142 ± 0.057	1.264 ± 0.063
U29	3.316 ± 0.099	1.880 ± 0.056	26.83 ± 0.80	19.71 ± 0.59	1.139 ± 0.057	1.272 ± 0.064
C30	2.806 ± 0.084	1.533 ± 0.071	26.29 ± 1.00	22.72 ± 0.84	1.194 ± 0.060	1.280 ± 0.104

<sup>a</sup>  $R_2$ -rates have been extracted from experimental  $R_{1\rho}$ -data as described in equation 1. <sup>b</sup> Errors have been derived from the curve-fitting procedure using the program Sparky 3.112 (University of California, San Francisco, USA) with a minimum error set to 3 % for the  $R_1$ - and  $R_2$ -rates and 5 % for the <sup>13</sup>C-<sup>1</sup>H} steady-state NOE. <sup>c</sup> The adenine C2 relaxation rates are written in italics.

**Table S5:** Results of the model-free analysis of the 14-nt cUUCGg hairpin RNA at 317 K. The vector C6H6 of the pyrimidines, the vector C8H8 of the purines and the adenine C2H2 vectors (*italics*) were analyzed with a bond length of 1.102 Å and a residue specific carbon CSA. For the C1'H1'-vector a bond length of 1.09 Å and a carbon CSA of 45 ppm were used.

	C6H6/C8H8/C2H2			C1'H1'			
	$S^2$	$\tau_e$ [ps]	$R_{ex}$ [ $s^{-1}$ ]	$S^2$	$S_f^2$	$\tau_e$ [ps]	$R_{ex}$ [ $s^{-1}$ ]
G1	0.81 ± 0.03	27 ± 20	1.96 ± 0.30				
G2	0.89 ± 0.02		11.31 ± 1.06	0.87 ± 0.02			1.99 ± 0.25
C3	0.86 ± 0.01			0.91 ± 0.01			
A4	0.88 ± 0.02		1.10 ± 0.29	0.89 ± 0.02			0.28 ± 0.22
	<i>0.88 ± 0.02</i>		<i>4.85 ± 0.47</i>				
C5	0.83 ± 0.02		1.21 ± 0.35	0.90 ± 0.01			
U6	0.79 ± 0.02		1.52 ± 0.40	0.90 ± 0.01			
U7	0.61 ± 0.02	17 ± 9	0.85 ± 0.34	0.84 ± 0.01			
C8	0.78 ± 0.02	22 ± 17		0.90 ± 0.01			
G9	0.82 ± 0.02		2.41 ± 0.29	0.91 ± 0.01			
G10	0.87 ± 0.02		2.25 ± 0.29	0.94 ± 0.05			1.61 ± 1.05
U11	0.84 ± 0.02		0.74 ± 0.35	0.84 ± 0.08	0.91 ± 0.02	1300 ± 3001	
G12	0.88 ± 0.02		6.64 ± 0.65	0.89 ± 0.02			0.38 ± 0.23
C13	0.83 ± 0.02		0.75 ± 0.36	0.89 ± 0.01			
C14	0.83 ± 0.01			0.65 ± 0.06	0.81 ± 0.03	873 ± 1052	

**Table S6:** Order parameter  $S^2$  from the 14-nt cUUCGg hairpin RNA for C6H6/C8H8/C2H2 (left) and C1'H1' (right) as a function of the residue number for three different temperatures extracted from the MD trajectory.

	$S^2$ of C6H6/C8H8/C2H2			$S^2$ of C1'H1'		
	298 K	317 K	360 K	298 K	317 K	360 K
G1	0.88	0.87	0.85	0.86	0.83	0.81
G2	0.92	0.91	0.89	0.91	0.90	0.89
C3	0.91	0.91	0.89	0.93	0.92	0.91
A4 <sup>a</sup>	0.92	0.91	0.90	0.93	0.92	0.91
	<i>0.92</i>	<i>0.91</i>	<i>0.90</i>			
C5	0.91	0.91	0.89	0.93	0.92	0.91
U6	0.89	0.88	0.86	0.91	0.91	0.89
U7	0.57	0.52	0.44	0.82	0.80	0.73
C8	0.82	0.81	0.75	0.86	0.85	0.77
G9	0.89	0.87	0.86	0.87	0.84	0.86
G10	0.92	0.91	0.90	0.93	0.92	0.91
U11	0.91	0.91	0.90	0.93	0.92	0.91
G12	0.93	0.93	0.91	0.93	0.92	0.91
C13	0.91	0.90	0.88	0.92	0.90	0.89
C14	0.90	0.88	0.86	0.85	0.81	0.78

<sup>a</sup> results for adenine C2H2 in italics

**Table S7:** Results of the model-free analysis of the 14-nt RNA at 317 K. The vector N1H1 of the guanines and the vector N3H3 of the uridine were analyzed with a bond length of 1.04 Å and a nitrogen CSA of -130 ppm for guanine N1 and -100 ppm for uridine N3.

$S^2$ of N1H1(G) / N3H3(U)	
G1	
G2	$0.95 \pm 0.02$
C3	
A4	
C5	
U6	
U7	
C8	
G9	$0.93 \pm 0.02$
G10	$0.92 \pm 0.02$
U11	$0.94 \pm 0.02$
G12	$0.94 \pm 0.02$
C13	
C14	

**Table S8:** Results of the model-free analysis of the 30-nt uCACGg hairpin RNA at 298 K. The parameters are the same as with the 14-nt cUUCGg RNA.

	C6H6/C8H8/C2H2				C1'H1'			
	$S^2$	$S_f^2$	$\tau_e$ [ps]	$R_{ex}$ [s <sup>-1</sup> ]	$S^2$	$S_f^2$	$\tau_e$ [ps]	$R_{ex}$ [s <sup>-1</sup> ]
G1	0.91 ± 0.04							
G2	0.97 ± 0.04			4.3 ± 1.9	0.94 ± 0.03			1.7 ± 1.5
C3	0.90 ± 0.03			9.5 ± 3.7	0.95 ± 0.02			
A4 <sup>a</sup>	0.95 ± 0.03				0.96 ± 0.02			
	<i>0.94 ± 0.02</i>							
C5	0.86 ± 0.03			8.9 ± 2.8	0.97 ± 0.02			
U6								
C7	0.83 ± 0.04		20 ± 14	14.3 ± 3.4	0.84 ± 0.03			7.1 ± 1.3
U8	0.91 ± 0.03			4.1 ± 2.2	0.92 ± 0.02			
G9	0.99 ± 0.03			5.5 ± 1.8	1.0 ± 0.0			11.8 ± 0.0
G10	0.97 ± 0.03				0.97 ± 0.02			
U11	0.92 ± 0.04			2.9 ± 2.8	0.94 ± 0.02			
A12	0.98 ± 0.02				0.89 ± 0.02			
	<i>0.93 ± 0.02</i>							
U13	0.92 ± 0.02				0.96 ± 0.02			
C14	0.94 ± 0.03				0.92 ± 0.03			4.4 ± 1.6
A15	0.58 ± 0.02	0.70 ± 0.02	441 ± 121		0.76 ± 0.06	0.86 ± 0.06	827 ± 445	
	<i>0.57 ± 0.02</i>	<i>0.68 ± 0.02</i>	<i>635 ± 233</i>					
C16	0.86 ± 0.03			6.7 ± 2.2	0.94 ± 0.04			2.8 ± 1.6
G17	1.0 ± 0.0			8.9 ± 0.0	1.0 ± 0.0			2.3 ± 0.0
G18	0.99 ± 0.02				0.85 ± 0.03			7.4 ± 1.2
U19	0.90 ± 0.02				1.0 ± 0.0			1.9 ± 0.0
A20	0.96 ± 0.03				0.86 ± 0.03			3.9 ± 1.2
	<i>0.93 ± 0.02</i>							
C21	0.89 ± 0.03			5.8 ± 2.4	0.97 ± 0.02			
C22	0.88 ± 0.03			7.4 ± 2.9	0.95 ± 0.02			
U23								
U24	0.92 ± 0.02				0.87 ± 0.02		37 ± 10	
U25	0.88 ± 0.03			3.4 ± 1.8				
G26	0.92 ± 0.03			4.7 ± 1.4	0.91 ± 0.05			36.5 ± 6.7
U27	0.88 ± 0.03			6.4 ± 2.5				
G28	0.93 ± 0.05			5.2 ± 2.9	1.00 ± 0.02			
U29	0.90 ± 0.02				0.97 ± 0.01		554 ± 452	
C30	0.79 ± 0.06		22 ± 14	23.9 ± 4.9	0.85 ± 0.03		23 ± 9	

<sup>a</sup> results for adenine C2H2 in italics

**Table S9:** Results of the model-free analysis of the 30-nt uCACGg hairpin RNA at 317 K. The parameters are the same as with the 14-nt cUUCGg RNA.

	C6H6/C8H8/C2H2				C1'H1'			
	$S^2$	$S_f^2$	$\tau_e$ [ps]	$R_{ex}$ [ $s^{-1}$ ]	$S^2$	$S_f^2$	$\tau_e$ [ps]	$R_{ex}$ [ $s^{-1}$ ]
G1	0.85 ± 0.04		54 ± 21	6.7 ± 1.3				
G2	0.92 ± 0.03			3.5 ± 1.0	0.91 ± 0.02			
C3	0.85 ± 0.03			6.0 ± 1.3	0.93 ± 0.03		48 ± 29	
A4 <sup>a</sup>	0.88 ± 0.03			4.0 ± 1.0	0.89 ± 0.03			1.0 ± 0.9
	<i>0.86 ± 0.03</i>			<i>4.5 ± 1.0</i>				
C5	0.89 ± 0.03			2.2 ± 1.2	0.93 ± 0.02			
U6								
C7	0.81 ± 0.04		27 ± 20	4.6 ± 1.6	0.86 ± 0.03			2.2 ± 0.8
U8	0.88 ± 0.02				0.84 ± 0.03			2.2 ± 0.8
G9	1.00 ± 0.02				0.93 ± 0.03			4.1 ± 0.9
G10	0.95 ± 0.03			1.4 ± 1.0	0.90 ± 0.02			
U11	0.86 ± 0.03			2.2 ± 1.2	0.97 ± 0.03			
A12	0.92 ± 0.03			1.7 ± 1.0	0.85 ± 0.03	0.92 ± 0.03	1168 ± 1732	
	<i>0.94 ± 0.02</i>							
U13	0.90 ± 0.02				0.93 ± 0.02		94 ± 41	
C14	0.85 ± 0.03			7.4 ± 1.4	0.92 ± 0.03			1.9 ± 1.0
A15	0.54 ± 0.03	0.67 ± 0.02	404 ± 136		0.92 ± 0.03		56 ± 30	
	<i>0.60 ± 0.03</i>	<i>0.72 ± 0.03</i>	<i>266 ± 108</i>					
C16	0.87 ± 0.03			3.1 ± 1.3	0.78 ± 0.04		37 ± 10	4.1 ± 0.9
G17	0.96 ± 0.03			15.3 ± 1.6	0.95 ± 0.03			6.5 ± 0.9
G18	0.96 ± 0.02				0.81 ± 0.02			4.1 ± 0.8
U19	0.86 ± 0.03			1.4 ± 1.2	0.85 ± 0.03			4.2 ± 0.8
A20	0.92 ± 0.03			1.0 ± 1.0	0.85 ± 0.03			3.5 ± 0.8
	<i>0.90 ± 0.03</i>			<i>3.2 ± 1.1</i>				
C21	0.89 ± 0.03			1.8 ± 1.2	0.90 ± 0.02		22 ± 16	
C22	0.90 ± 0.03			7.0 ± 1.4	0.90 ± 0.03		32 ± 20	
U23								
U24	0.87 ± 0.02				0.92 ± 0.03		34 ± 22	
U25								
G26	0.93 ± 0.03			2.8 ± 1.0	0.91 ± 0.03			10.9 ± 1.1
U27								
G28	0.91 ± 0.03			4.3 ± 1.0	0.88 ± 0.02		28 ± 14	
U29	0.86 ± 0.03			1.9 ± 1.2	0.89 ± 0.03			1.9 ± 0.9
C30	0.77 ± 0.04		22 ± 16	8.2 ± 1.5	0.78 ± 0.02		25 ± 8	

<sup>a</sup> results for adenine C2H2 in italics



**Table S10:** Results of the model-free analysis of the 30-nt uCACGg hairpin RNA at 325 K. The parameters are the same as with the 14-nt cUUCGg RNA.

	C6H6/C8H8/C2H2				C1'H1'			
	$S^2$	$S_f^2$	$\tau_e$ [ps]	$R_{ex}$ [s <sup>-1</sup> ]	$S^2$	$S_f^2$	$\tau_e$ [ps]	$R_{ex}$ [s <sup>-1</sup> ]
G1	0.73 ± 0.05		16 ± 14	11.1 ± 3.8	0.66 ± 0.03		25 ± 8	6.5 ± 0.7
G2	0.78 ± 0.02			5.3 ± 1.1	0.76 ± 0.02			6.1 ± 0.7
C3	0.79 ± 0.02			6.6 ± 1.0	0.79 ± 0.02			5.6 ± 0.7
A4 <sup>a</sup>	0.85 ± 0.03			5.0 ± 0.8	0.89 ± 0.03			1.3 ± 0.7
	<i>0.84 ± 0.03</i>			<i>3.8 ± 0.8</i>				
C5	0.71 ± 0.02			10.9 ± 1.5	0.73 ± 0.02			5.9 ± 0.7
U6	0.78 ± 0.02			8.2 ± 1.0	0.89 ± 0.03			2.0 ± 0.7
C7	0.69 ± 0.04		21 ± 17	9.3 ± 1.2	0.78 ± 0.03			5.0 ± 0.7
U8	0.81 ± 0.02			4.9 ± 0.9	0.85 ± 0.03			2.0 ± 0.6
G9	0.80 ± 0.02			6.9 ± 0.9	0.84 ± 0.03			6.0 ± 0.7
G10	0.82 ± 0.03			6.6 ± 0.9	0.80 ± 0.02			3.3 ± 1.0
U11	0.80 ± 0.02			5.9 ± 1.0	0.92 ± 0.02			
A12	0.88 ± 0.03			3.4 ± 0.8	0.84 ± 0.03	0.90 ± 0.02	692 ± 767	
	<i>0.86 ± 0.03</i>			<i>3.4 ± 0.8</i>				
U13	0.77 ± 0.04		26 ± 20	8.4 ± 1.2	0.91 ± 0.02		52 ± 28	
C14	0.74 ± 0.04		21 ± 18	3.8 ± 1.0	0.79 ± 0.04		21 ± 13	4.3 ± 0.8
A15	0.57 ± 0.04	0.68 ± 0.03	224 ± 162		0.86 ± 0.02		29 ± 20	
	<i>0.48 ± 0.03</i>	<i>0.63 ± 0.02</i>	<i>466 ± 162</i>					
C16	0.70 ± 0.04		33 ± 16	2.3 ± 1.0	0.72 ± 0.04		29 ± 11	5.2 ± 1.6
G17	0.65 ± 0.04		13 ± 11	20.9 ± 2.8	0.77 ± 0.04		28 ± 13	11.8 ± 1.0
G18	0.82 ± 0.06		64 ± 29	1.9 ± 1.3	0.69 ± 0.04		13 ± 11	5.1 ± 1.1
U19	0.87 ± 0.03			4.3 ± 1.4	0.79 ± 0.02			4.4 ± 0.6
A20	0.87 ± 0.03			2.9 ± 0.7	0.82 ± 0.03			2.8 ± 0.6
	<i>0.84 ± 0.03</i>			<i>4.8 ± 0.8</i>				
C21	0.75 ± 0.02			8.5 ± 1.2	0.84 ± 0.03			2.2 ± 0.7
C22	0.75 ± 0.02			8.4 ± 1.1	0.82 ± 0.03			2.6 ± 0.7
U23	0.75 ± 0.04		26 ± 19	7.7 ± 1.1	0.86 ± 0.03			2.7 ± 0.7
U24	0.75 ± 0.04		35 ± 20	5.8 ± 1.1	0.86 ± 0.03	0.91 ± 0.02	375 ± 380	
U25	0.78 ± 0.02			6.5 ± 0.9	0.88 ± 0.03			2.5 ± 0.7
G26	0.69 ± 0.03			7.9 ± 2.2	0.76 ± 0.02			9.3 ± 0.8
U27	0.78 ± 0.02			6.5 ± 0.9	0.76 ± 0.04		17 ± 11	6.5 ± 0.8
G28	0.79 ± 0.02			8.3 ± 1.7	0.73 ± 0.04		16 ± 9	4.0 ± 1.0
U29	0.79 ± 0.02			7.9 ± 1.0	0.79 ± 0.04		25 ± 12	6.5 ± 0.8
C30	0.63 ± 0.03		19 ± 11	11.0 ± 1.2	0.62 ± 0.05		12 ± 10	12.8 ± 1.1

<sup>a</sup> results for adenine C2H2 in italics

**Table S11:** Order parameter  $S^2$  from the 14-nt uCACGg hairpin RNA for C6H6/C8H8/C2H2 (left) and C1'H1' (right) as a function of the residue number for three different temperatures extracted from the MD trajectory.

	$S^2$ of C6H6/C8H8			$S^2$ of C1'H1'		
	298 K	317 K	360 K	298 K	317 K	360 K
G9	0.89	0.88	0.84	0.87	0.85	0.80
G10	0.92	0.91	0.89	0.92	0.91	0.87
U11	0.92	0.91	0.88	0.93	0.92	0.89
A12 <sup>a</sup>	0.93	0.92	0.89	0.93	0.92	0.89
	<i>0.92</i>	<i>0.91</i>	<i>0.87</i>			
U13	0.91	0.90	0.83	0.93	0.92	0.86
C14	0.89	0.88	0.59	0.92	0.90	0.69
A15	0.49	0.42	0.59	0.88	0.77	0.74
	0.43	0.39	0.52			
C16	0.87	0.79	0.62	0.89	0.83	0.65
G17	0.89	0.89	0.72	0.87	0.88	0.78
G18	0.92	0.91	0.85	0.92	0.92	0.84
U19	0.92	0.92	0.87	0.93	0.93	0.88
A20	0.92	0.92	0.87	0.92	0.91	0.89
	<i>0.91</i>	<i>0.91</i>	<i>0.85</i>			
C21	0.90	0.89	0.87	0.90	0.90	0.86
C22	0.89	0.88	0.84	0.85	0.84	0.76

<sup>a</sup> results for adenine C2H2 in italics

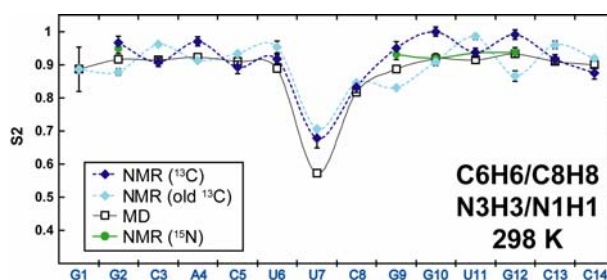
**Table S12:** Secondary structural elements (stacking and hydrogen bonds) of the cUUCGg tetraloop. The hydrogen bond between the 2'-OH of U7 and the O6 of G9 is possible with respect to the distances of the heavy atoms but in NMR there is no hydroxyl proton detectable and also the orientation is not favorable.

Base Stacking	Hydrogen Bonds
U6 on C5 G10 on G9 C8 on U6	G9 (N2) to U6 (O2) G9 (N1) to U6 (O2) U6 (O2') to G9 (O6) C8 (N4) to U7 (PO <sup>proR</sup> ) U7 (O2') to G9 (O6) (?)

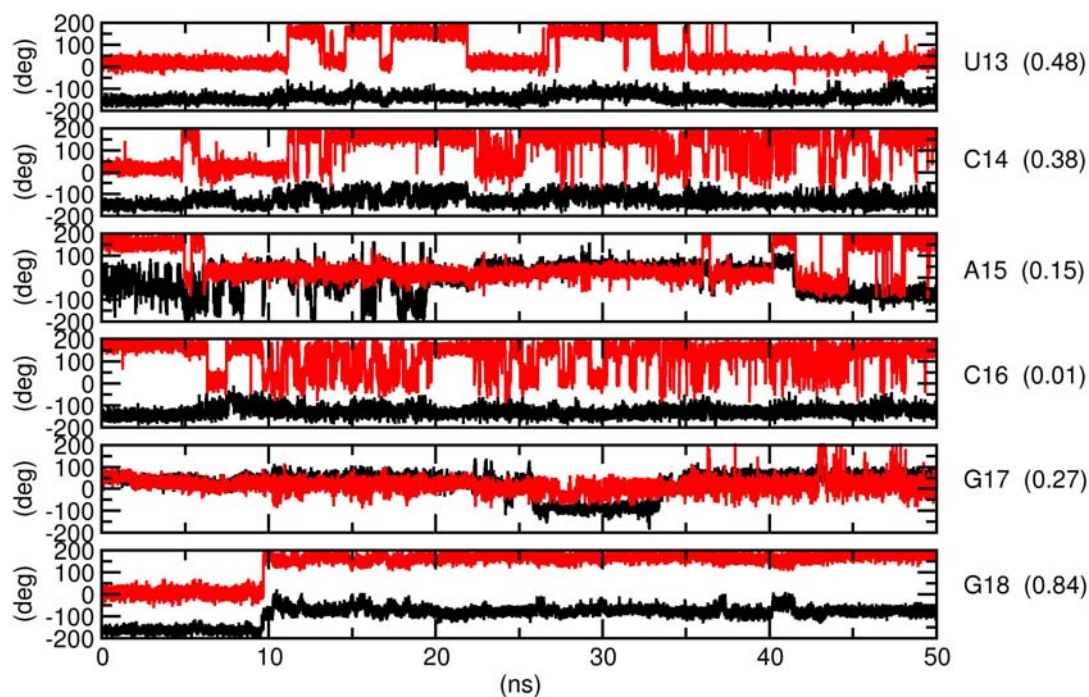
**Table S13:** Secondary structural elements (stacking and hydrogen bonds) of the uCACGg tetraloop.

Base Stacking	Hydrogen Bonds
C14 on U13 G18 on G17 C16 on C14	G17 (N2) to C14 (O2) G17 (N1) to C14 (O2) C14 (O2') to G17 (O6) C16 (N4) to A15 (PO <sup>proR</sup> )

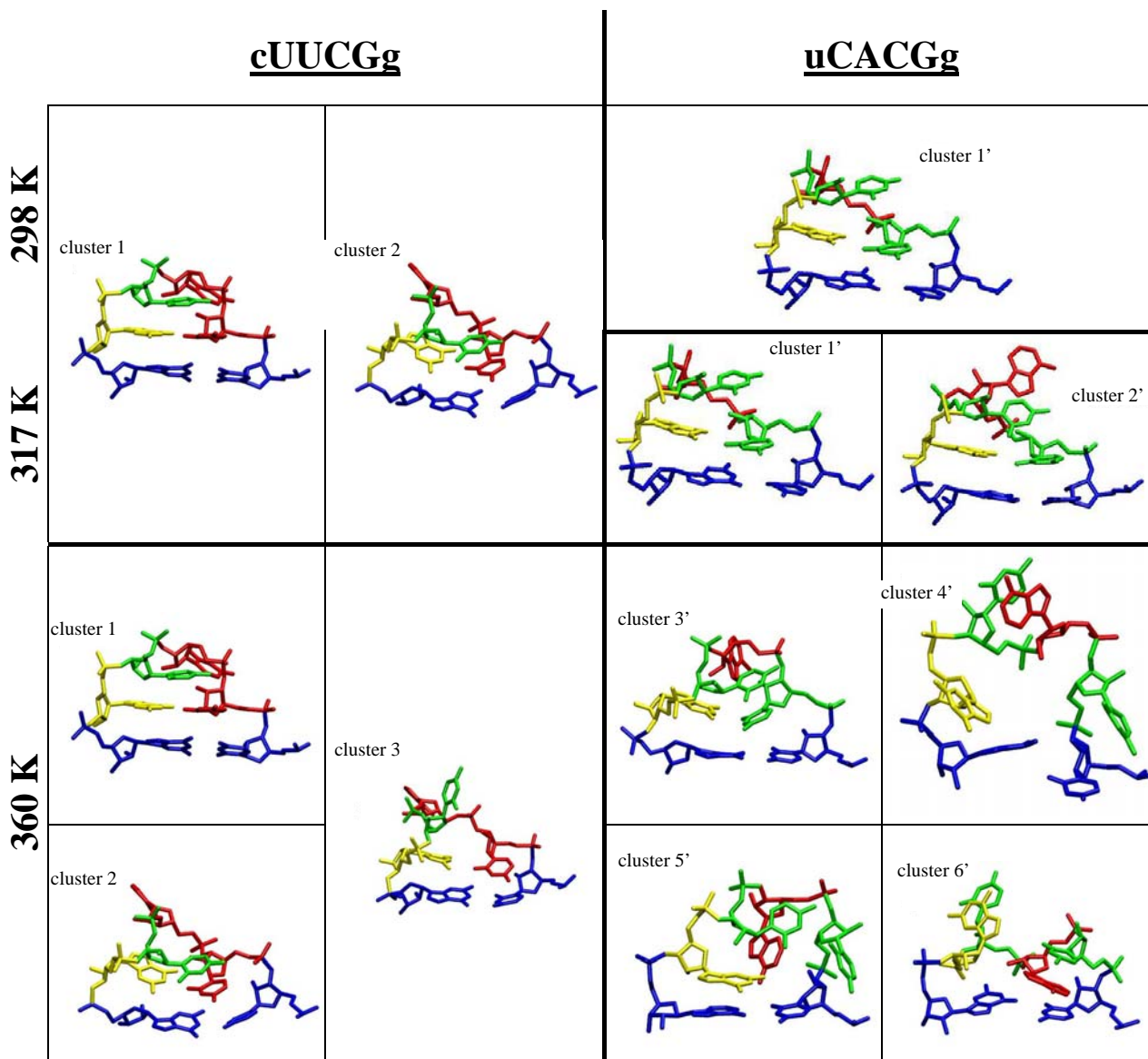
## FIGURES



**Figure S1:** Order Parameters for the nucleobase vectors C6H6/C8H8 of the previous analysis (light blue) (Duchardt 2005), the new analysis (dark blue), extracted from the MD simulation (black) (Villa 2006) and for imino vectors N3H3 of uridine (green) and N1H1 of guanine (Akke 1994) sorted by the sequence of the cUUCGg hairpin and measured at 298 K.



**Figure S2:** Conformational dynamics of the uCACGg hairpin as observed in the time evolution at 360 K. Torsion angle  $\chi$  (black) and pseudorotational angle P (red) for residue U13, C14, A15, C16, G17, G18. Correlation values between the dynamics of the sugar and the base are reported together with the residue names.



**Figure S3:** MD conformational snapshots of the cUUCGg RNA hairpin (residues C5-G10, left) and of the uCACGg hairpin RNA (residues U13-G18, right) from the MD simulations at 298 K (top), 317 K (middle) and 360 K (bottom). Shown are the closing residues in blue, the cytidines in green, the guanines in yellow and the uridines and adenines in red. The clusters correspond to the classification in the PCA analysis (Figure 3 in the manuscript).

**ZUSAMMENFASSUNG****Spektroskopische Untersuchungen zur Bestimmung  
von RNA-Ligand Wechselwirkungen und RNA  
Dynamiken**

Die vorliegende Doktorarbeit behandelt die strukturelle Aufklärung von Wechselwirkungen zwischen biologisch relevanten Ribonukleinsäuren (RNA) und ausgewählten Liganden, um zur Methodenentwicklung für den Entwurf neuer Pharmaka beizutragen. Dafür sind nicht nur genaue Kenntnisse der Strukturen der RNA, der Liganden und der Komplexe von Nöten, sondern auch Informationen der inhärenten Dynamik insbesondere der RNA. Zur Bestimmung sowohl der Strukturen, als auch der Dynamiken stellt die Flüssig-Kernspinresonanz-Spektroskopie (NMR) eine ideale biophysikalische Methode dar.

Die Notwendigkeit zur Untersuchung von RNA-Ligand-Wechselwirkungen ergibt sich aus der wachsenden Erkenntnis der Vielfalt von RNA-Funktionen in biologischen Systemen.<sup>[8]</sup> Besonders die Retroviren, z.B. das HI-Virus (humane immunodeficiency virus), stellen die medizinische und pharmakologische Forschung weltweit vor immer neue Probleme. Entweder gibt es noch keine geeigneten Zielstrukturen oder die Viren entwickeln Resistenzen, vor allem gegen Wirkstoffe, die virale Proteine als Ziel haben. Durch die Entdeckung der funktionellen Vielfalt von RNA bieten sich diese als neue Zielstrukturen für orthogonale Therapien an.<sup>[24]</sup>

Die ersten Kapitel dieser Doktorarbeit geben zum einen eine Einleitung in die RNA-Forschung mit besonderem Fokus auf den allgemeinen strukturellen und dynamischen Eigenschaften von Ribonukleinsäuren (Chapter II), stellen zweitens die ausgewählten RNA-Zielstrukturen und deren mit verschiedenen Methoden bestimmten Liganden vor (Chapter III), und erklären drittens die zugrundeliegende NMR-Theorie und die verwendeten Methoden zur

Untersuchung der Bindungscharakteristika, zur Strukturbestimmung der RNA und der Liganden und zur Ableitung dynamischer Parameter aus experimentellen Daten (Chapter IV).

In Chapter II werden die Eigenschaften, Strukturen und Funktionen von RNA vorgestellt.<sup>[39]</sup> Dabei wird kurz die hierarchische Einteilung räumlicher Konformationen von RNA Biomakromolekülen erklärt: primäre, sekundäre und tertiäre Strukturelemente.<sup>[59]</sup> Des Weiteren werden die diversen Funktionen aufgelistet, die RNA-Moleküle in lebenden Zellen übernehmen.<sup>[8]</sup>

Im nächsten Kapitel (Chapter III) werden die biologischen Systeme und ihre Sequenzen vorgestellt, an denen die Untersuchungen durchgeführt wurden: HIV  $\Psi$ -RNA<sup>[35]</sup> und HIV TAR RNA<sup>[36]</sup>, sowie der Stemloop D im Cloverleaf-Element aus der 5'-untranslatierten Region des Coxsackievirus B3.<sup>[37, 38]</sup> Dabei werden auch die Liganden präsentiert und die Methoden, durch die sie ausgewählt und ihre Affinität bestimmt wurden.<sup>[1, 4]</sup>

Abgeschlossen wird die Einleitung durch das Chapter IV, worin die NMR-spektroskopische Theorie und Methoden beschrieben werden, die für die durchgeführten Untersuchungen relevant waren. Zunächst wird die Thermodynamik und Kinetik der Wechselwirkungen von RNA und Liganden unter dem Gesichtspunkt der NMR zusammengefasst. Darauf folgt eine ausführliche Beschreibung der Vorgehensweise bei der Zuordnung der NMR Resonanzen zu den Wasserstoff-, Stickstoff-, Kohlenstoff- und Phosphorkernen der RNAs und der Liganden. Hierbei stehen vor allem die Protonenresonanzen im Vordergrund, die für die spätere Strukturrechnung essentiell sind. Das Verfahren der Strukturbestimmung der RNA, der Liganden und deren Komplexe wird mit einer kurzen Beschreibung der Grundlagen und der verwendeten Programme, ARIA<sup>[152]</sup> und HADDOCK<sup>[178, 179]</sup>, geschildert. Das Kapitel wird mit einer Zusammenfassung der Relaxationstheorie<sup>[180]</sup>, sowie einer methodischen Erklärung der Ableitung dynamischer Parameter

aus autokorrelierten Relaxationsraten mit dem Model-Free Formalismus<sup>[182, 183]</sup> abgeschlossen.

Die Chapter V bis VII dieser Arbeit enthalten die Originalpublikationen, die in drei Themenbereiche eingeteilt sind. Chapter V gruppiert die drei Publikationen, in denen die Bestimmung und Charakterisierung peptidischer Liganden der HIV  $\Psi$ -RNA und deren Wechselwirkungen miteinander behandelt werden. Durch einen Phage-Display Assay wurde zunächst eine Konsensussequenz identifiziert, die sich durch einen hohen Anteil an der Aminosäure Tryptophan auszeichnet (HWWPWW).<sup>[1]</sup> Die Bindungsaffinitäten dieses Hexapeptids für die gesamte  $\Psi$ -RNA, als auch für ihre vier einzelnen Stemloops wurden durch Zirkulare Dichroismus Spektroskopie (CD-Spektroskopie) untersucht. Zur Verbesserung der Bindungseigenschaften, als auch für eine verbesserte Handhabung in den spektroskopischen Untersuchungen, z.B. durch verbesserte Löslichkeit der stark hydrophoben Sequenz, wurde das Hexapeptid mittels einer Sequenzvariierung auf einer Membranoberfläche (SPOT-Assay) weiter optimiert.<sup>[3]</sup> Dabei konnte die RNA-Affinität bei Austausch einer der sechs Aminosäure gegen ein Lysin beibehalten, aber die Löslichkeit signifikant verbessert werden. Die weiteren strukturellen Untersuchungen der RNA-Ligand-Wechselwirkungen wurden per Fluoreszenz- und NMR-Spektroskopie durchgeführt, wobei die NMR-Spektroskopie aufzeigen konnte, dass das Peptid HKWPWW in zwei Konformationen der zentralen Prolinpeptidbindung zu beinahe gleichen Anteilen vorliegt.<sup>[2]</sup> Beide Konformere binden an die  $\Psi$ -RNA. Auch die selektive Bindung des Liganden HKWPWW an weitere Nukleinsäuresequenzen aus dem HI-Virus, z. B. TAR-RNA und PBS-DNA, wurden bestimmt. Diese untersuchten Nukleinsäuresequenzen haben gemein, dass sie Bindungspartner des gleichen viralen Proteins, NCp7, sind. Durch Synthese weitere Mutanten des HKWPWW-Liganden soll die konformationelle Freiheit der Prolinpeptidbindung eingeschränkt werden, wodurch die spezifischen Interaktionen der beiden Konformere mit der Ziel-RNA weiter charakterisiert werden können.

Die zwei Publikationen in Chapter VI beschreiben die Ligandselektion gegen die Zielstruktur TAR in den HI-Viren des Typs 1 und Typs 2 und der Strukturaufklärung des Komplexes mittels NMR-Spektroskopie. Als Liganden wurden Tripeptide synthetisiert, in denen zwei Arginine eine synthetische Aminosäure flankieren.<sup>[4]</sup> Diese synthetischen Aminosäuren trugen aromatische oder heteroaromatische Gruppierungen in ihren Seitenketten, die zu einer Verbesserung der Affinität und Selektivität beisteuern sollten. Eine Vorauswahl der Liganden wurde durch eine Fluoreszenz-Resonanzenergietransfersichtung (FRET-Assay) vorgenommen und die Interaktionen der ausgewählten Liganden mit der RNA per NMR-Spektroskopie konkretisiert. Eine intensive strukturelle Untersuchung des Liganden mit einer Pyrimidinylgruppe in der Seitenkette der zentralen Aminosäure in Komplex mit der HIV-2 TAR RNA ergab eine 2:1 Bindungsstöchiometrie des Liganden.<sup>[5]</sup> Die erste stärkere Bindungsstelle im Bulge der RNA war bereits weitgehend bekannt als Ziel von Arginin-tragenden Liganden.<sup>[116]</sup> Das Tripeptid umklammert die Nukleinbase U23, wobei die Seitenkette des N-terminalen Arginins in Wasserstoffbrückenbindungsnähe der Nukleinbase von G26 liegt, ähnlich der Bindungsstelle von Argininamid.<sup>[58, 114]</sup> Das C-terminale Arginin ist oberhalb der Nukleinbase U23 positioniert und könnte mit der Nukleinbase A27 interagieren. Die zentrale Aminosäure mit der synthetischen Seitenkette geht keine direkten Interaktionen mit der RNA ein. Durch Substitution der Arginine mit Lysin konnten weitere Informationen über die Bindungscharakteristika abgeleitet werden.

Die strukturellen Untersuchungen konnten auch die zweite Bindungsstelle des Tripeptids unterhalb des Bulges lokalisieren. Das Peptid bindet schwächer und weniger spezifisch entlang des Streifens der Pyrimidine U40-C41-U42 am 3'-Ende des Bulges. Diese Ergebnisse können zur weiteren Optimierung von TAR-Liganden beitragen.

Im letzten Kapitel (Chapter VII) sind die zwei Publikationen zur Untersuchung der RNA-Dynamik zusammengefasst. Aus autokorrelierten Relaxationsraten



der Kerne C1' und C8 (für Purine) bzw. C6 (für Pyrimidine) in Nukleotiden der RNA Tetraloopsequenzen UUCG und CACG wurden mittels des Model-Free Formalismus<sup>[182, 183]</sup> Parameter abgeleitet, die über Dynamiken auf der Zeitskala von Pico- bis Nanosekunden der C-H Vektoren berichten: dem Orderparameter  $S^2$ , der internen Korrelationszeit  $\tau_e$  und der Rate  $R_{ex}$ , die von chemischen oder konformationellen Austauschprozessen abhängig sein kann.<sup>[66, 191]</sup> Für diese Ableitung ist ein detailliertes Wissen weiterer Parameter von Nöten: der Bindungslänge des C-H Vektors<sup>[193]</sup> und der Anisotropie der chemischen Verschiebung des  $^{13}\text{C}$ -Kerns ( $^{13}\text{C}$ -CSA).<sup>[197, 200]</sup> Die Verwendung dieser erst vor kurzem und exakter bestimmten Werte ermöglichte eine genauere Ableitung der inhärenten Dynamiken dieser RNA Moleküle.<sup>[7]</sup> Diese Informationen konnten in die strukturellen Untersuchungen der glykosidischen Bindung durch kreuzkorrelierte Relaxationsraten eingebaut werden. Dadurch wurde eine Parametrisierung ermöglicht, die die Bestimmung des Winkels der glykosidischen Bindung auch in anderen RNA-Molekülen erlaubt.<sup>[6]</sup>

Des Weiteren konnten die dynamischen Parameter bei verschiedenen Temperaturen mit Parametern abgeglichen werden, die aus Molekulardynamischen (MD) Trajektorien abgeleitet wurden.<sup>[7]</sup> Dies ermöglichte die Visualisierung der internen Bewegungen zweier strukturell ähnlicher Tetraloops aus der YNMG-Familie, die sich aber in ihrer Stabilität unterscheiden. Bei Temperaturen nahe dem Schmelzpunkt des weniger stabilen CACG-Tetraloops offenbarten sich die Änderungen in der Dynamik, die zum Aufschmelzen des Loops führen. Dabei verliert als erstes die Nukleinbase in dritter Loopposition ihre stabilisierenden Interaktionen. Bei weiterer Erhöhung der Temperatur löst sich auch das Loopbasenpaar zwischen der ersten und der letzten Nukleinbase und andere stabilisierende Stapelungen der Nukleinbasen treten temporär auf. Weitere Untersuchungen mit vergleichbaren Tetraloops sollen nun helfen, die besondere Stabilität der Loopsequenz UUCG zu verstehen. Dabei spielt sehr wahrscheinlich auch das Basenpaar direkt unterhalb des Loops eine wesentliche Rolle.



## REFERENCES

- [1] A. Pustowka, J. Dietz, J. Ferner, M. Baumann, M. Landersz, C. Konigs, H. Schwalbe, U. Dietrich, Identification of Peptide Ligands for Target RNA Structures Derived from the HIV-1 Packaging Signal  $\Psi$  by Screening Phage-Displayed Peptide Libraries. *ChemBioChem* **2003**, 4, 1093-1097.
- [2] C. Raja, J. Ferner, U. Dietrich, S. Avilov, D. Ficheux, J. L. Darlix, H. de Rocquigny, H. Schwalbe, Y. Mely, A Tryptophan-Rich Hexapeptide Inhibits Nucleic Acid Destabilization Chaperoned by the HIV-1 Nucleocapsid Protein. *Biochemistry* **2006**, 45, 9254-9265.
- [3] J. Dietz, J. Koch, A. Kaur, C. Raja, S. Stein, M. Grez, A. Pustowka, S. Mensch, J. Ferner, L. Moller, N. Bannert, R. Tampe, G. Divita, Y. Mely, H. Schwalbe, U. Dietrich, Inhibition of HIV-1 by a Peptide Ligand of the Genomic RNA Packaging Signal  $\Psi$ . *ChemMedChem* **2008**, 3, 749-755.
- [4] V. Ludwig, A. Krebs, M. Stoll, U. Dietrich, J. Ferner, H. Schwalbe, U. Scheffer, G. Durner, M. W. Gobel, Tripeptides from Synthetic Amino Acids Block the Tat-TAR Association and Slow Down HIV Spread in Cell Cultures. *ChemBioChem* **2007**, 8, 1850-1856.
- [5] J. Ferner, M. Suhartono, S. Breitung, H. R. Jonker, M. Hennig, J. Wohnert, M. Gobel, H. Schwalbe, Structures of HIV TAR RNA-Ligand Complexes Reveal Higher Binding Stoichiometries. *ChemBioChem* **2009**, 10, 1490-1494.
- [6] J. Rinnenthal, C. Richter, J. Ferner, E. Duchardt, H. Schwalbe, Quantitative  $\Gamma$ -HCNCH: Determination of the Glycosidic Torsion Angle  $\chi$  in RNA Oligonucleotides from the Analysis of CH Dipolar Cross-Correlated Relaxation by Solution NMR Spectroscopy. *J. Biomol. NMR* **2007**, 39, 17-29.
- [7] J. Ferner, A. Villa, E. Duchardt, E. Widjajakusuma, J. Wohnert, G. Stock, H. Schwalbe, NMR and MD Studies of the Temperature-Dependent Dynamics of RNA YNMG-Tetraloops. *Nucleic Acids Res.* **2008**, 36, 1928-1940.
- [8] G. Storz, An Expanding Universe of Noncoding RNAs. *Science* **2002**, 296, 1260-1263.
- [9] T. M. Tarasow, S. L. Tarasow, B. E. Eaton, RNA-Catalysed Carbon-Carbon Bond Formation. *Nature* **1997**, 389, 54-57.
- [10] B. Seelig, A. Jaschke, A Small Catalytic RNA Motif with Diels-Alderase Activity. *Chem. Biol.* **1999**, 6, 167-176.

- 
- [11] F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, A. Yonath, Structure of Functionally Activated Small Ribosomal Subunit at 3.3 Å Resolution. *Cell* **2000**, 102, 615-623.
- [12] M. M. Yusupov, G. Z. Yusupova, A. Baucom, K. Lieberman, T. N. Earnest, J. H. Cate, H. F. Noller, Crystal Structure of the Ribosome at 5.5 Å Resolution. *Science* **2001**, 292, 883-896.
- [13] N. Ban, P. Nissen, J. Hansen, P. B. Moore, T. A. Steitz, The Complete Atomic Structure of the Large Ribosomal Subunit at 2.4 Å Resolution. *Science* **2000**, 289, 905-920.
- [14] T. R. Cech, Structural Biology. The Ribosome Is a Ribozyme. *Science* **2000**, 289, 878-879.
- [15] C. M. Spahn, M. G. Gomez-Lorenzo, R. A. Grassucci, R. Jorgensen, G. R. Andersen, R. Beckmann, P. A. Penczek, J. P. Ballesta, J. Frank, Domain Movements of Elongation Factor eEF2 and the Eukaryotic 80S Ribosome Facilitate tRNA Translocation. *EMBO J.* **2004**, 23, 1008-1019.
- [16] W. Gilbert, Origin of Life - the RNA World. *Nature* **1986**, 319, 618-618.
- [17] F. S. Collins, M. Morgan, A. Patrinos, The Human Genome Project: Lessons from Large-Scale Biology. *Science* **2003**, 300, 286-290.
- [18] M. A. Batzer, P. L. Deininger, Alu Repeats and Human Genomic Diversity. *Nat. Rev. Genet.* **2002**, 3, 370-379.
- [19] G. J. Hannon, F. V. Rivas, E. P. Murchison, J. A. Steitz, The Expanding Universe of Noncoding RNAs. *Cold Spring Harb. Symp. Quant. Biol.* **2006**, 71, 551-564.
- [20] S. L. Lin, J. D. Miller, S. Y. Ying, Intronic MicroRNA (miRNA). *J. Biomed. Biotechnol.* **2006**, 2006, 26818.
- [21] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, C. C. Mello, Potent and Specific Genetic Interference by Double-Stranded RNA in *Caenorhabditis Elegans*. *Nature* **1998**, 391, 806-811.
- [22] Y. Zhao, D. Srivastava, A Developmental View of microRNA Function. *Trends Biochem. Sci.* **2007**, 32, 189-197.
- [23] A. J. Hamilton, D. C. Baulcombe, A Species of Small Antisense RNA in Posttranscriptional Gene Silencing in Plants. *Science* **1999**, 286, 950-952.
- [24] P. J. Paddison, A. A. Caudy, G. J. Hannon, Stable Suppression of Gene Expression by RNAi in Mammalian Cells. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 1443-1448.

- 
- [25] A. S. Mironov, I. Gusarov, R. Rafikov, L. E. Lopez, K. Shatalin, R. A. Kreneva, D. A. Perumov, E. Nudler, Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **2002**, 111, 747-756.
- [26] A. Nahvi, N. Sudarsan, M. S. Ebert, X. Zou, K. L. Brown, R. R. Breaker, Genetic Control by a Metabolite Binding mRNA. *Chem. Biol.* **2002**, 9, 1043.
- [27] W. C. Winkler, S. Cohen-Chalamish, R. R. Breaker, An mRNA Structure That Controls Gene Expression by Binding Fmn. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, 99, 15908-15913.
- [28] W. C. Winkler, A. Nahvi, R. R. Breaker, Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression. *Nature* **2002**, 419, 952-956.
- [29] R. T. Batey, S. D. Gilbert, R. K. Montange, Structure of a Natural Guanine-Responsive Riboswitch Complexed with the Metabolite Hypoxanthine. *Nature* **2004**, 432, 411-415.
- [30] J. Noeske, C. Richter, M. A. Grundl, H. R. Nasiri, H. Schwalbe, J. Wohnert, An Intermolecular Base Triple as the Basis of Ligand Specificity and Affinity in the Guanine- and Adenine-Sensing Riboswitch RNAs. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, 102, 1372-1377.
- [31] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, R. R. Breaker, Riboswitches Control Fundamental Biochemical Pathways in *Bacillus Subtilis* and Other Bacteria. *Cell* **2003**, 113, 577-586.
- [32] J. N. Kim, A. Roth, R. R. Breaker, Guanine Riboswitch Variants from *Mesoplasma Florum* Selectively Recognize 2'-Deoxyguanosine. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, 104, 16092-16097.
- [33] M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, R. R. Breaker, A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science* **2004**, 306, 275-279.
- [34] N. Sudarsan, J. K. Wickiser, S. Nakamura, M. S. Ebert, R. R. Breaker, An mRNA Structure in Bacteria That Controls Gene Expression by Binding Lysine. *Genes Dev.* **2003**, 17, 2688-2697.
- [35] B. Berkhout, Structure and Function of the Human Immunodeficiency Virus Leader RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **1996**, 54, 1-34.
- [36] B. Berkhout, R. H. Silverman, K. T. Jeang, Tat Trans-Activates the Human Immunodeficiency Virus through a Nascent RNA Target. *Cell* **1989**, 59, 273-282.

- [37] E. Ehrenfeld, N. L. Teterina, in *Molecular Biology of Picornaviruses* (Eds.: B. L. Semler, E. Wimmer), ASM Press, Washington, D.C., **2002**, pp. 159-170.
- [38] A. V. Paul, in *Molecular Biology of Picornaviruses* (Eds.: B. L. Semler, E. Wimmer), ASM Press, Washington, D.C., **2002**, pp. 227-246.
- [39] W. Saenger, *Principles of Nucleic Acid Structure*, Springer-Verlag, New York, **1984**.
- [40] J. Fohrer, M. Hennig, T. Carlomagno, Influence of the 2'-Hydroxyl Group Conformation on the Stability of A-Form Helices in RNA. *J. Mol. Biol.* **2006**, 356, 280-287.
- [41] D. J. Proctor, J. E. Schaak, J. M. Bevilacqua, C. J. Falzone, P. C. Bevilacqua, Isolation and Characterization of a Family of Stable RNA Tetraloops with the Motif YNMG That Participate in Tertiary Interactions. *Biochemistry* **2002**, 41, 12062-12075.
- [42] G. Varani, Exceptionally Stable Nucleic Acid Hairpins. *Annu. Rev. Biophys. Biomol. Struct.* **1995**, 24, 379-404.
- [43] C. C. Correll, K. Swinger, Common and Distinctive Features of GNRA Tetraloops Based on a GUAA Tetraloop Structure at 1.4 Å Resolution. *RNA* **2003**, 9, 355-363.
- [44] H. A. Heus, A. Pardi, Structural Features That Give Rise to the Unusual Stability of RNA Hairpins Containing GNRA Loops. *Science* **1991**, 253, 191-194.
- [45] F. M. Jucker, H. A. Heus, P. F. Yip, E. H. M. Moors, A. Pardi, A Network of Heterogeneous Hydrogen Bonds in GNRA Tetraloops. *J. Mol. Biol.* **1996**, 264, 968-980.
- [46] C. C. Correll, B. Freeborn, P. B. Moore, T. A. Steitz, Metals, Motifs, and Recognition in the Crystal Structure of a 5S rRNA Domain. *Cell* **1997**, 91, 705-712.
- [47] A. Dallas, P. B. Moore, The Loop E Loop D Region of Escherichia Coli 5S rRNA: The Solution Structure Reveals an Unusual Loop That May Be Important for Binding Ribosomal Proteins. *Structure* **1997**, 5, 1639-1653.
- [48] D. K. Hendrix, S. E. Brenner, S. R. Holbrook, RNA Structural Motifs: Building Blocks of a Modular Biomolecule. *Q. Rev. Biophys.* **2005**, 38, 221-243.
- [49] A. Roy, S. Panigrahi, M. Bhattacharyya, D. Bhattacharyya, Structure, Stability, and Dynamics of Canonical and Noncanonical Base Pairs: Quantum Chemical Studies. *J. Phys. Chem. B* **2008**, 112, 3786-3796.

- [50] O. Ohlenschlager, J. Wohnert, E. Bucci, S. Seitz, S. Hafner, R. Ramachandran, R. Zell, M. Gorch, The Structure of the Stemloop D Subdomain of Coxsackievirus B3 Cloverleaf RNA and Its Interaction with the Proteinase 3C. *Structure* **2004**, 12, 237-248.
- [51] M. Stoldt, J. Wohnert, O. Ohlenschlager, M. Gorch, L. R. Brown, The NMR Structure of the 5S rRNA E-Domain-Protein L25 Complex Shows Preformed and Induced Recognition. *EMBO J.* **1999**, 18, 6508-6521.
- [52] S. E. Butcher, T. Dieckmann, J. Feigon, Solution Structure of a GAAA Tetraloop Receptor RNA. *Embo Journal* **1997**, 16, 7490-7499.
- [53] J. H. Cate, A. R. Gooding, E. Podell, K. H. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, J. A. Doudna, Crystal Structure of a Group I Ribozyme Domain: Principles of RNA Packing. *Science* **1996**, 273, 1678-1685.
- [54] R. C. Davis, I. Tinoco, Jr., Temperature-Dependent Properties of Dinucleoside Phosphates. *Biopolymers* **1968**, 6, 223-242.
- [55] P. N. Borer, Y. Lin, S. Wang, M. W. Roggenbuck, J. M. Gott, O. C. Uhlenbeck, I. Pelczer, Proton NMR and Structural Features of a 24-Nucleotide RNA Hairpin. *Biochemistry* **1995**, 34, 6488-6503.
- [56] Q. Zhang, A. C. Stelzer, C. K. Fisher, H. M. Al-Hashimi, Visualizing Spatially Correlated Dynamics That Directs RNA Conformational Transitions. *Nature* **2007**, 450, 1263-1267.
- [57] A. S. Brodsky, J. R. Williamson, Solution Structure of the HIV-2 TAR-Arginamide Complex. *J. Mol. Biol.* **1997**, 267, 624-639.
- [58] J. D. Puglisi, R. Tan, B. J. Calnan, A. D. Frankel, J. R. Williamson, Conformation of the TAR RNA-Arginine Complex by NMR Spectroscopy. *Science* **1992**, 257, 76-80.
- [59] E. Westhof, P. Auffinger, in *Encyclopedia of Analytical Chemistry : Applications, Theory and Instrumentation* (Ed.: R. A. Meyers), Wiley, New York, **2000**, pp. 5222-5232.
- [60] P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, T. A. Steitz, RNA Tertiary Interactions in the Large Ribosomal Subunit: The A-Minor Motif. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, 98, 4899-4903.
- [61] J. D. Puglisi, J. R. Wyatt, I. Tinoco, Jr., A Pseudoknotted RNA Oligonucleotide. *Nature* **1988**, 331, 283-286.
- [62] A. Serganov, S. Keiper, L. Malinina, V. Tereshko, E. Skripkin, C. Hobartner, A. Polonskaia, A. T. Phan, R. Wombacher, R. Micura, Z. Dauter, A. Jaschke, D. J. Patel, Structural Basis for Diels-Alder

- Ribozyme-Catalyzed Carbon-Carbon Bond Formation. *Nat. Struct. Mol. Biol.* **2005**, 12, 218-224.
- [63] J. Noeske, J. Buck, B. Furtig, H. R. Nasiri, H. Schwalbe, J. Wohnert, Interplay of 'Induced Fit' and Preorganization in the Ligand Induced Folding of the Aptamer Domain of the Guanine Binding Riboswitch. *Nucleic Acids Res.* **2007**, 35, 572-583.
- [64] E. Duchardt, L. Nilsson, J. Schleucher, Cytosine Ribose Flexibility in DNA: A Combined NMR <sup>13</sup>C Spin Relaxation and Molecular Dynamics Simulation Study. *Nucleic Acids Res.* **2008**, 36, 4211-4219.
- [65] Z. Shajani, G. Varani, <sup>13</sup>C Relaxation Studies of the DNA Target Sequence for HhaI Methyltransferase Reveal Unique Motional Properties. *Biochemistry* **2008**, 47, 7617-7625.
- [66] E. Duchardt, H. Schwalbe, Residue Specific Ribose and Nucleobase Dynamics of the cUUCGg RNA Tetraloop Motif by NMR <sup>13</sup>C Relaxation. *J. Biomol. NMR* **2005**, 32, 295-308.
- [67] E. J. Sorin, M. A. Engelhardt, D. Herschlag, V. S. Pande, RNA Simulations: Probing Hairpin Unfolding and the Dynamics of a GNRA Tetraloop. *J. Mol. Biol.* **2002**, 317, 493-506.
- [68] Q. Zhang, X. Sun, E. D. Watt, H. M. Al-Hashimi, Resolving the Motional Modes That Code for RNA Adaptation. *Science* **2006**, 311, 653-656.
- [69] J. R. Williamson, Induced Fit in RNA-Protein Recognition. *Nat. Struct. Mol. Biol.* **2000**, 7, 834-837.
- [70] W. D. Wilson, L. Ratmeyer, M. Zhao, L. Strekowski, D. Boykin, The Search for Structure-Specific Nucleic Acid-Interactive Drugs: Effects of Compound Structure on RNA Versus DNA Interaction Strength. *Biochemistry* **1993**, 32, 4098-4104.
- [71] T. Hermann, Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes. *Angew. Chem., Int. Ed.* **2000**, 39, 1890-1904.
- [72] A. Krebs, V. Ludwig, O. Boden, M. W. Gobel, Targeting the HIV Trans-Activation Responsive Region - Approaches Towards RNA-Binding Drugs. *ChemBioChem* **2003**, 4, 972-978.
- [73] F. J. Palella, Jr., K. M. Delaney, A. C. Moorman, M. O. Loveless, J. Fuhrer, G. A. Satten, D. J. Aschman, S. D. Holmberg, Declining Morbidity and Mortality among Patients with Advanced Human Immunodeficiency Virus Infection. HIV Outpatient Study Investigators. *N. Engl. J. Med.* **1998**, 338, 853-860.
- [74] J. Martinez-Picado, M. P. DePasquale, N. Kartsonis, G. J. Hanna, J. Wong, D. Finzi, E. Rosenberg, H. F. Gunthard, L. Sutton, A. Savara, C.



- J. Petropoulos, N. Hellmann, B. D. Walker, D. D. Richman, R. Siliciano, R. T. D'Aquila, Antiretroviral Resistance During Successful Therapy of HIV Type 1 Infection. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 10948-10953.
- [75] T. Hermann, E. Westhof, RNA as a Drug Target: Chemical, Modelling, and Evolutionary Tools. *Curr. Opin. Biotechnol.* **1998**, *9*, 66-73.
- [76] M. Laughrea, L. Jette, A 19-Nucleotide Sequence Upstream of the 5' Major Splice Donor Is Part of the Dimerization Domain of Human Immunodeficiency Virus 1 Genomic RNA. *Biochemistry* **1994**, *33*, 13464-13474.
- [77] E. Skripkin, J. C. Paillart, R. Marquet, B. Ehresmann, C. Ehresmann, Identification of the Primary Site of the Human Immunodeficiency Virus Type 1 RNA Dimerization in Vitro. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 4945-4949.
- [78] B. Berkhout, J. L. van Wamel, Role of the DIS Hairpin in Replication of Human Immunodeficiency Virus Type 1. *J. Virol.* **1996**, *70*, 6723-6732.
- [79] D. F. Purcell, M. A. Martin, Alternative Splicing of Human Immunodeficiency Virus Type 1 mRNA Modulates Viral Protein Expression, Replication, and Infectivity. *J. Virol.* **1993**, *67*, 6365-6378.
- [80] G. K. Amarasinghe, R. N. De Guzman, R. B. Turner, K. J. Chancellor, Z. R. Wu, M. F. Summers, NMR Structure of the HIV-1 Nucleocapsid Protein Bound to Stem-Loop SL2 of the  $\Psi$ -RNA Packaging Signal. Implications for Genome Recognition. *J. Mol. Biol.* **2000**, *301*, 491-511.
- [81] N. Hagan, D. Fabris, Direct Mass Spectrometric Determination of the Stoichiometry and Binding Affinity of the Complexes between Nucleocapsid Protein and RNA Stem-Loop Hairpins of the HIV-1  $\Psi$ -Recognition Element. *Biochemistry* **2003**, *42*, 10736-10745.
- [82] T. Hayashi, T. Shioda, Y. Iwakura, H. Shibuta, RNA Packaging Signal of Human Immunodeficiency Virus Type 1. *Virology* **1992**, *188*, 590-599.
- [83] H. Huthoff, B. Berkhout, Two Alternating Structures of the HIV-1 Leader RNA. *RNA* **2001**, *7*, 143-157.
- [84] S. Baba, K. Takahashi, S. Noguchi, H. Takaku, Y. Koyanagi, N. Yamamoto, G. Kawai, Solution RNA Structures of the HIV-1 Dimerization Initiation Site in the Kissing-Loop and Extended-Duplex Dimers. *J. Biochem.* **2005**, *138*, 583-592.
- [85] G. K. Amarasinghe, R. N. De Guzman, R. B. Turner, M. F. Summers, NMR Structure of Stem-Loop SL2 of the HIV-1  $\Psi$ -RNA Packaging Signal Reveals a Novel A-U-A Base-Triple Platform. *J. Mol. Biol.* **2000**, *299*, 145-156.

- [86] L. Pappalardo, D. J. Kerwood, I. Pelczer, P. N. Borer, Three-Dimensional Folding of an RNA Hairpin Required for Packaging HIV-1. *J. Mol. Biol.* **1998**, 282, 801-818.
- [87] D. J. Kerwood, M. J. Cavaluzzi, P. N. Borer, Structure of SL4 RNA from the HIV-1 Packaging Signal. *Biochemistry* **2001**, 40, 14518-14529.
- [88] E. T. Yu, A. Hawkins, J. Eaton, D. Fabris, MS3D Structural Elucidation of the HIV-1 Packaging Signal. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, 105, 12248-12253.
- [89] H. M. Azzazy, W. E. Highsmith, Jr., Phage Display Technology: Clinical Applications and Recent Innovations. *Clin. Biochem.* **2002**, 35, 425-445.
- [90] W. J. Friesen, M. K. Darby, Specific RNA Binding Proteins Constructed from Zinc Fingers. *Nat. Struct. Mol. Biol.* **1998**, 5, 543-546.
- [91] S. Hoffmann, D. Willbold, A Selection System to Study Protein-RNA Interactions: Functional Display of HIV-1 Tat Protein on Filamentous Bacteriophage M13. *Biochem. Biophys. Res. Commun.* **1997**, 235, 806-811.
- [92] R. N. De Guzman, Z. R. Wu, C. C. Stalling, L. Pappalardo, P. N. Borer, M. F. Summers, Structure of the HIV-1 Nucleocapsid Protein Bound to the SL3  $\Psi$ -RNA Recognition Element. *Science* **1998**, 279, 384-388.
- [93] A. C. Paoletti, M. F. Shubsda, B. S. Hudson, P. N. Borer, Affinities of the Nucleocapsid Protein for Variants of SL3 RNA in HIV-1. *Biochemistry* **2002**, 41, 15423-15428.
- [94] M. F. Shubsda, A. C. Paoletti, B. S. Hudson, P. N. Borer, Affinities of Packaging Domain Loops in HIV-1 RNA for the Nucleocapsid Protein. *Biochemistry* **2002**, 41, 5276-5282.
- [95] D. J. Patel, Adaptive Recognition in RNA Complexes with Peptides and Protein Modules. *Curr. Opin. Struct. Biol.* **1999**, 9, 74-87.
- [96] J. Karn, Tackling Tat. *J. Mol. Biol.* **1999**, 293, 235-254.
- [97] F. Aboul-ela, J. Karn, G. Varani, Structure of HIV-1 TAR RNA in the Absence of Ligands Reveals a Novel Conformation of the Trinucleotide Bulge. *Nucleic Acids Res.* **1996**, 24, 3974-3981.
- [98] K. M. Weeks, C. Ampe, S. C. Schultz, T. A. Steitz, D. M. Crothers, Fragments of the HIV-1 Tat Protein Specifically Bind TAR RNA. *Science* **1990**, 249, 1281-1285.
- [99] K. M. Weeks, D. M. Crothers, RNA Recognition by Tat-Derived Peptides: Interaction in the Major Groove? *Cell* **1991**, 66, 577-588.

- [100] M. J. Churcher, C. Lamont, F. Hamy, C. Dingwall, S. M. Green, A. D. Lowe, J. G. Butler, M. J. Gait, J. Karn, High Affinity Binding of TAR RNA by the Human Immunodeficiency Virus Type-1 Tat Protein Requires Base-Pairs in the RNA Stem and Amino Acid Residues Flanking the Basic Region. *J. Mol. Biol.* **1993**, 230, 90-110.
- [101] Z. Athanassiou, R. L. A. Dias, K. Moehle, N. Dobson, G. Varani, J. A. Robinson, Structural Mimicry of Retroviral Tat Proteins by Constrained - Hairpin Peptidomimetics: Ligands with High Affinity and Selectivity for Viral TAR RNA Regulatory Elements. *J. Am. Chem. Soc.* **2004**, ASAP Article.
- [102] L. Dassonneville, F. Hamy, P. Colson, C. Houssier, C. Bailly, Binding of Hoechst 33258 to the TAR RNA of HIV-1. Recognition of a Pyrimidine Bulge-Dependent Structure. *Nucleic Acids Res.* **1997**, 25, 4487-4492.
- [103] N. Gelus, F. Hamy, C. Bailly, Molecular Basis of HIV-1 TAR RNA Specific Recognition by an Acridine Tat-Antagonist. *Bioorg. Med. Chem.* **1999**, 7, 1075-1079.
- [104] K. E. Lind, Z. Du, K. Fujinaga, B. M. Peterlin, T. L. James, Structure-Based Computational Database Screening, in Vitro Assay, and NMR Assessment of Compounds That Target TAR RNA. *Chemistry & Biology* **2002**, 9, 185-193.
- [105] B. Davis, M. Afshar, G. Varani, A. I. Murchie, J. Karn, G. Lentzen, M. Drysdale, J. Bower, A. J. Potter, I. D. Starkey, T. Swarbrick, F. Aboul-ela, Rational Design of Inhibitors of HIV-1 TAR RNA through the Stabilisation of Electrostatic "Hot Spots". *J. Mol. Biol.* **2004**, 336, 343-356.
- [106] A. Lapidot, E. Ben-Asher, M. Eisenstein, Tetrahydropyrimidine Derivatives Inhibit Binding of a Tat-Like, Arginine-Containing Peptide, to HIV TAR RNA in Vitro. *FEBS Lett.* **1995**, 367, 33-38.
- [107] C. Faber, H. Sticht, K. Schweimer, P. Rosch, Structural Rearrangements of HIV-1 Tat-Responsive RNA Upon Binding of Neomycin B. *J. Biol. Chem.* **2000**, 275, 20660-20666.
- [108] D. Raghunathan, V. M. Sanchez-Pedregal, J. Junker, C. Schwiegk, M. Kalesse, A. Kirschning, T. Carlomagno, TAR-RNA Recognition by a Novel Cyclic Aminoglycoside Analogue. *Nucleic Acids Res.* **2006**, 34, 3599-3608.
- [109] J. A. Ippolito, T. A. Steitz, A 1.3 Å Resolution Crystal Structure of the HIV-1 Trans-Activation Response Region RNA Stem Reveals a Metal Ion-Dependent Bulge Conformation. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, 95, 9819-9824.

- [110] T. E. Edwards, S. T. Sigurdsson, EPR Spectroscopic Analysis of TAR RNA-Metal Ion Interactions. *Biochem. Biophys. Res. Commun.* **2003**, 303, 721-725.
- [111] R. A. Colvin, S. W. White, M. A. Garcia-Blanco, D. W. Hoffman, Structural Features of an RNA Containing the CUGGGA Loop of the Human Immunodeficiency Virus Type 1 Trans-Activation Response Element. *Biochemistry* **1993**, 32, 1105-1112.
- [112] T. Kulinski, M. Olejniczak, H. Huthoff, L. Bielecki, K. Pachulska-Wieczorek, A. T. Das, B. Berkhout, R. W. Adamiak, The Apical Loop of the HIV-1 TAR RNA Hairpin Is Stabilized by a Cross-Loop Base Pair. *J. Biol. Chem.* **2003**, 278, 38892-38901.
- [113] M. J. Michnicka, J. W. Harper, G. C. King, Selective Isotopic Enrichment of Synthetic RNA: Application to the HIV-1 TAR Element. *Biochemistry* **1993**, 32, 395-400.
- [114] A. S. Brodsky, H. A. Erlacher, J. R. Williamson, NMR Evidence for a Base Triple in the HIV-2 TAR C-G.C+ Mutant-Argininamide Complex. *Nucleic Acids Res.* **1998**, 26, 1991-1995.
- [115] K. T. Dayie, A. S. Brodsky, J. R. Williamson, Base Flexibility in HIV-2 TAR RNA Mapped by Solution (15)N, (13)C NMR Relaxation. *J. Mol. Biol.* **2002**, 317, 263-278.
- [116] J. Tao, A. D. Frankel, Specific Binding of Arginine to TAR RNA. *Proceedings of the National Academy of Sciences of the USA* **1992**, 89, 2723-2726.
- [117] M. Zacharias, P. J. Hagerman, The Bend in RNA Created by the Trans-Activation Response Element Bulge of Human Immunodeficiency Virus Is Straightened by Arginine and by Tat-Derived Peptide. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 6052-6056.
- [118] F. Aboul-ela, J. Karn, G. Varani, The Structure of the Human Immunodeficiency Virus Type-1 TAR RNA Reveals Principles of RNA Recognition by Tat Protein. *J. Mol. Biol.* **1995**, 253, 313-332.
- [119] A. Krebs, V. Ludwig, J. Pfizer, G. Durner, M. W. Gobel, Enantioselective Synthesis of Non-Natural Aromatic Alpha-Amino Acids. *Chemistry* **2004**, 10, 544-553.
- [120] I. Huq, Y. H. Ping, N. Tamilarasu, T. M. Rana, Controlling Human Immunodeficiency Virus Type 1 Gene Expression by Unnatural Peptides. *Biochemistry* **1999**, 38, 5172-5177.
- [121] E. P. Nikonowicz, A. Pardi, An Efficient Procedure for Assignment of the Proton, Carbon and Nitrogen Resonances in <sup>13</sup>C/<sup>15</sup>N Labeled Nucleic Acids. *J. Mol. Biol.* **1993**, 232, 1141-1156.

- [122] S. S. Wijmenga, B. N. M. van Buuren, The Use of NMR Methods for Conformational Studies of Nucleic Acids. *Progress in NMR Spectroscopy* **1998**, 32, 287-387.
- [123] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, Oligoribonucleotide Synthesis Using T7 RNA Polymerase and Synthetic DNA Templates. *Nucleic Acids Res.* **1987**, 15, 8783-8798.
- [124] J. F. Milligan, O. C. Uhlenbeck, Synthesis of Small RNAs Using T7 RNA Polymerase. *Methods Enzymol.* **1989**, 180, 51-62.
- [125] B. Furtig, C. Richter, J. Wohnert, H. Schwalbe, NMR Spectroscopy of RNA. *ChemBioChem* **2003**, 4, 936-962.
- [126] J. Xu, J. Lapham, D. M. Crothers, Determining RNA Solution Structure by Segmental Isotopic Labeling and NMR: Application to Caenorhabditis Elegans Spliced Leader RNA 1. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, 93, 44-48.
- [127] T. Ohtsuki, G. Kawai, K. Watanabe, Stable Isotope-Edited NMR Analysis of Ascaris Suum Mitochondrial tRNAMet Having a TV-Replacement Loop. *J. Biochem.* **1998**, 124, 28-34.
- [128] G. Varani, I. Tinoco, Jr., RNA Structure and NMR Spectroscopy. *Q. Rev. Biophys.* **1991**, 24, 479-532.
- [129] K. Wüthrich, *NMR of Proteins and Nucleic Acids*, Wiley, New York, **1986**.
- [130] D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J. Mol. Biol.* **1999**, 288, 911-940.
- [131] M. Zuker, Mfold Web Server for Nucleic Acid Folding and Hybridization Prediction. *Nucleic Acids Res.* **2003**, 31, 3406-3415.
- [132] A. J. Dingley, S. Grzesiek, Direct Observation of Hydrogen Bonds in Nucleic Acid Base Pairs by Internucleotide (2)J(NN) Couplings. *J. Am. Chem. Soc.* **1998**, 120, 8293-8297.
- [133] S. Meiboom, D. Gill, Modified Spin-Echo Method for Measuring Nuclear Relaxation Times. *Rev. Sci. Instrum.* **1958**, 29, 688-691.
- [134] L. Mueller, P. Legault, A. Pardi, Improved RNA Structure Determination by Detection of NOE Contacts to Exchange-Broadened Amino Protons. *J. Am. Chem. Soc.* **1995**, 117, 11043-11048.
- [135] A. Pardi, E. P. Nikonowicz, Simple Procedure for Resonance Assignment of the Sugar Protons in <sup>13</sup>C-Labeled RNAs. *J. Am. Chem. Soc.* **1992**, 114, 9202-9203.

- [136] H. Schwalbe, J. P. Marino, S. J. Glaser, C. Griesinger, Measurement of H,H-Coupling Constants Associated with  $\nu-1$ ,  $\nu-2$ , and  $\nu-3$  in Uniformly C-13-Labeled RNA by HCC-TOCSY-CCH-E.COSY. *J. Am. Chem. Soc.* **1995**, 117, 7251-7252.
- [137] S. J. Glaser, H. Schwalbe, J. P. Marino, C. Griesinger, Directed TOCSY, a Method for Selection of Directed Correlations by Optimal Combinations of Isotropic and Longitudinal Mixing. *Journal of Magnetic Resonance Series B* **1996**, 112, 160-180.
- [138] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, 2-Dimensional and 3-Dimensional HCN Experiments for Correlating Base and Sugar Resonances in N-15,C-13-Labeled RNA Oligonucleotides. *J. Biomol. NMR* **1993**, 3, 721-727.
- [139] V. Sklenar, M. R. Rejante, R. D. Peterson, E. Wang, J. Feigon, Two-Dimensional Triple-Resonance HCNCH Experiment for Direct Correlation of Ribose H1' and Base H8, H6 Protons in <sup>13</sup>C,<sup>15</sup>N-Labeled RNA Oligonucleotides. *J. Am. Chem. Soc.* **1993**, 115, 12181-12182.
- [140] B. T. Farmer, 2nd, L. Muller, E. P. Nikonowicz, A. Pardi, Unambiguous Resonance Assignments in Carbon-13, Nitrogen-15-Labeled Nucleic Acids by 3D Triple-Resonance NMR. *J. Am. Chem. Soc.* **1993**, 115, 11040-11041.
- [141] J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, C. Griesinger, Sequential Correlation of Anomeric Ribose Protons and Intervening Phosphorus in RNA Oligonucleotides by a H-1,C-13,P-31 Triple-Resonance Experiment - HCP-CCH-TOCSY. *J. Biomol. NMR* **1995**, 5, 87-92.
- [142] M. Hennig, J. R. Williamson, Detection of N-H...N Hydrogen Bonding in RNA Via Scalar Couplings in the Absence of Observable Imino Proton Resonances. *Nucleic Acids Res.* **2000**, 28, 1585-1593.
- [143] V. Sklenar, R. D. Peterson, M. R. Rejante, J. Feigon, Correlation of Nucleotide Base and Sugar Protons in a N-15-Labeled HIV-1 RNA Oligonucleotide by H-1-N-15 HSQC Experiments. *J. Biomol. NMR* **1994**, 4, 117-122.
- [144] A. J. Dingley, J. E. Masse, R. D. Peterson, M. Barfield, J. Feigon, S. Grzesiek, Internucleotide Scalar Couplings across Hydrogen Bonds in Watson-Crick and Hoogsteen Base Pairs of a DNA Triplex. *J. Am. Chem. Soc.* **1999**, 121, 6019-6027.
- [145] J. Wohnert, A. J. Dingley, M. Stoldt, M. Gorchach, S. Grzesiek, L. R. Brown, Direct Identification of NH...N Hydrogen Bonds in Non-Canonical Base Pairs of RNA by NMR Spectroscopy. *Nucleic Acids Res.* **1999**, 27, 3104-3110.

- [146] J. P. Marino, J. H. Prestegard, D. M. Crothers, Correlation of Adenine H2/H8 Resonances in Uniformly C-13 Labeled RNAs by 2D HCCH-TOCSY - a New Tool for H-1 Assignment. *J. Am. Chem. Soc.* **1994**, 116, 2205-2206.
- [147] A. Bax, D. G. Davis, Practical Aspects of Two-Dimensional Transverse NOE Spectroscopy. *J. Magn. Reson.* **1985**, 63, 207-213.
- [148] G. Otting, K. Wuthrich, Studies of Protein Hydration in Aqueous-Solution by Direct NMR Observation of Individual Protein-Bound Water-Molecules. *J. Am. Chem. Soc.* **1989**, 111, 1871-1875.
- [149] A. J. Dingley, J. E. Masse, J. Feigon, S. Grzesiek, Characterization of the Hydrogen Bond Network in Guanosine Quartets by Internucleotide (3h)J(NC') and (2h)J(NN) Scalar Couplings. *J. Biomol. NMR* **2000**, 16, 279-289.
- [150] A. Liu, A. Majumdar, F. Jiang, N. Chernichenko, E. Skripkin, D. J. Patel, NMR Detection of Intermolecular N-H...N Hydrogen Bonds in the Human T Cell Leukemia Virus-1 Rex Peptide-RNA Aptamer Complex. *J. Am. Chem. Soc.* **2000**, 122, 11226-11227.
- [151] C. A. Theimer, L. D. Finger, L. Trantirek, J. Feigon, Mutations Linked to Dyskeratosis Congenita Cause Changes in the Structural Equilibrium in Telomerase RNA. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, 100, 449-454.
- [152] J. P. Linge, M. Habeck, W. Rieping, M. Nilges, ARIA: Automated NOE Assignment and NMR Structure Calculation. *Bioinformatics* **2003**, 19, 315-316.
- [153] H. Schwalbe, J. P. Marino, G. C. King, R. Wechselberger, W. Bermel, C. Griesinger, Determination of a Complete Set of Coupling-Constants in C-13-Labeled Oligonucleotides. *J. Biomol. NMR* **1994**, 4, 631-644.
- [154] L. Trantirek, R. Stefl, J. E. Masse, J. Feigon, V. Sklenar, Determination of the Glycosidic Torsion Angles in Uniformly C-13-Labeled Nucleic Acids from Vicinal Coupling Constants (3)J(C2/4-H1') and (3)J(C6/8-H1'). *J. Biomol. NMR* **2002**, 23, 1-12.
- [155] M. L. Munzarova, V. Sklenar, DFT Analysis of NMR Scalar Interactions across the Glycosidic Bond in DNA. *J. Am. Chem. Soc.* **2003**, 125, 3649-3658.
- [156] E. Duchardt, C. Richter, O. Ohlenschlager, M. Gorlach, J. Wohnert, H. Schwalbe, Determination of the Glycosidic Bond Angle  $\chi$  in RNA from Cross-Correlated Relaxation of CH Dipolar Coupling and N Chemical Shift Anisotropy. *J. Am. Chem. Soc.* **2004**, 126, 1962-1970.
- [157] C. Richter, B. Reif, C. Griesinger, H. Schwalbe, NMR Spectroscopic Determination of Angles  $\alpha$  and  $\xi$  in RNA from CH-Dipolar Coupling, P-

- CSA Cross-Correlated Relaxation. *J. Am. Chem. Soc.* **2000**, 122, 12728-12731.
- [158] G. C. K. Roberts, *NMR of Macromolecules : A Practical Approach*, IRL Press at Oxford University Press, Oxford ; New York, **1993**.
- [159] J. R. Tolman, J. M. Flanagan, M. A. Kennedy, J. H. Prestegard, Nuclear Magnetic Dipole Interactions in Field-Oriented Proteins - Information for Structure Determination in Solution. *Proc. Natl. Acad. Sci. U. S. A.* **1995**, 92, 9279-9283.
- [160] N. Tjandra, J. G. Omichinski, A. M. Gronenborn, G. M. Clore, A. Bax, Use of Dipolar H-1-N-15 and H-1-C-13 Couplings in the Structure Determination of Magnetically Oriented Macromolecules in Solution. *Nat. Struct. Mol. Biol.* **1997**, 4, 732-738.
- [161] J. H. Prestegard, J. R. Tolman, H. M. Al-Hashimi, M. Andrec, in *Biological Magnetic Resonance, Vol. 17* (Eds.: N. R. Krishna, L. J. Berliner), Kluwer Academic, New York ; London, **1999**, pp. 311-355.
- [162] P. Bayer, L. Varani, G. Varani, Refinement of the Structure of Protein-RNA Complexes by Residual Dipolar Coupling Analysis. *J. Biomol. NMR* **1999**, 14, 149-155.
- [163] M. Zweckstetter, A. Bax, Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **2000**, 122, 3791-3792.
- [164] M. R. Hansen, L. Mueller, A. Pardi, Tunable Alignment of Macromolecules by Filamentous Phage Yields Dipolar Coupling Interactions. *Nat. Struct. Mol. Biol.* **1998**, 5, 1065-1074.
- [165] G. M. Clore, M. R. Starich, A. M. Gronenborn, Measurement of Residual Dipolar Couplings of Macromolecules Aligned in the Nematic Phase of a Colloidal Suspension of Rod-Shaped Viruses. *J. Am. Chem. Soc.* **1998**, 120, 10571-10572.
- [166] M. R. Hansen, P. Hanson, A. Pardi, Pf1 Filamentous Phage as an Alignment Tool for Generating Local and Global Structural Information in Nucleic Acids. *J. Biomol. Struct. Dyn.* **2000**, 365-369.
- [167] J. F. Ying, A. Grishaev, M. P. Latham, A. Pardi, A. Bax, Magnetic Field Induced Residual Dipolar Couplings of Imino Groups in Nucleic Acids from Measurements at a Single Magnetic Field. *J. Biomol. NMR* **2007**, 39, 91-96.
- [168] T. F. Havel, An Evaluation of Computational Strategies for Use in the Determination of Protein-Structure from Distance Constraints Obtained by Nuclear-Magnetic-Resonance. *Progress in Biophysics & Molecular Biology* **1991**, 56, 43-78.



- [169] P. Guntert, K. Wuthrich, Improved Efficiency of Protein Structure Calculations from NMR Data Using the Program DIANA with Redundant Dihedral Angle Constraints. *J. Biomol. NMR* **1991**, 1, 447-456.
- [170] D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, R. J. Woods, The AMBER Biomolecular Simulation Programs. *J. Comput. Chem.* **2005**, 26, 1668-1688.
- [171] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, CHARMM - a Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, 4, 187-217.
- [172] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, 54, 905-921.
- [173] T. E. Cheatham, M. A. Young, Molecular Dynamics Simulation of Nucleic Acids: Successes, Limitations, and Promise. *Biopolymers* **2000**, 56, 232-256.
- [174] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, 4, 435-447.
- [175] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, 26, 1701-1718.
- [176] W. L. Jorgensen, J. Tiradorives, The OPLS Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, 110, 1657-1666.
- [177] J. Pranata, S. G. Wierschke, W. L. Jorgensen, OPLS Potential Functions for Nucleotide Bases - Relative Association Constants of Hydrogen-Bonded Base-Pairs in Chloroform. *J. Am. Chem. Soc.* **1991**, 113, 2810-2819.
- [178] C. Dominguez, R. Boelens, A. M. J. J. Bonvin, HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**, 125, 1731-1737.
- [179] S. J. De Vries, A. D. J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, T. Wassenaar, A. M. J. J. Bonvin, HADDOCK Versus HADDOCK: New Features and Performance of Haddock2.0 on the CAPRI Targets. *Proteins: Struct., Funct., Bioinf.* **2007**, 69, 726-733.

- [180] A. Abragam, *The Principles of Nuclear Magnetism*, Clarendon Press, Oxford, **1961**.
- [181] E. L. Hahn, Spin Echoes. *Physical Review* **1950**, 80, 580-594.
- [182] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **1982**, 104, 4546-4559.
- [183] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *J. Am. Chem. Soc.* **1982**, 104, 4559-4570.
- [184] L. E. Kay, D. A. Torchia, A. Bax, Backbone Dynamics of Proteins as Studied by N-15 Inverse Detected Heteronuclear NMR-Spectroscopy - Application to Staphylococcal Nuclease. *Biochemistry* **1989**, 28, 8972-8979.
- [185] A. G. Palmer, M. Rance, P. E. Wright, Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance C-13 Heteronuclear NMR-Spectroscopy. *J. Am. Chem. Soc.* **1991**, 113, 4371-4380.
- [186] G. Wagner, NMR Relaxation and Protein Mobility. *Curr. Opin. Struct. Biol.* **1993**, 3, 748-754.
- [187] R. Bruschweiler, D. A. Case, Characterization of Biomolecular Structure and Dynamics by NMR Cross-Relaxation. *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, 26, 27-58.
- [188] J. Garcia de la Torre, M. L. Huertas, B. Carrasco, HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. *J. Magn. Reson.* **2000**, 147, 138-146.
- [189] G. M. Clore, A. Szabo, A. Bax, L. E. Kay, P. C. Driscoll, A. M. Gronenborn, Deviations from the Simple Two-Parameter Model-Free Approach to the Interpretation of Nitrogen-15 Nuclear Magnetic Resonance of Proteins. *J. Am. Chem. Soc.* **1990**, 112, 4989-4991.
- [190] A. M. Mandel, M. Akke, A. G. Palmer, Backbone Dynamics of Escherichia-Coli Ribonuclease-H1 - Correlations with Structure and Function in an Active Enzyme. *J. Cell. Biochem.* **1995**, 29-29.
- [191] A. M. Mandel, M. Akke, A. G. Palmer, Backbone Dynamics of Escherichia-Coli Ribonuclease HI - Correlations with Structure and Function in an Active Enzyme. *J. Mol. Biol.* **1995**, 246, 144-163.

- [192] E. J. d'Auvergne, P. R. Gooley, The Use of Model Selection in the Model-Free Analysis of Protein Dynamics. *J. Biomol. NMR* **2003**, 25, 25-39.
- [193] D. A. Case, Calculations of NMR Dipolar Coupling Strengths in Model Peptides. *J. Biomol. NMR* **1999**, 15, 95-102.
- [194] S. Tang, D. A. Case, Vibrational Averaging of Chemical Shift Anisotropies in Model Peptides. *J. Biomol. NMR* **2007**, 38, 255-266.
- [195] Z. Shajani, G. Varani, <sup>13</sup>C NMR Relaxation Studies of RNA Base and Ribose Nuclei Reveal a Complex Pattern of Motions in the RNA Binding Site for Human U1A Protein. *J. Mol. Biol.* **2005**, 349, 699-715.
- [196] J. B. Hall, D. Fushman, Variability of the N-15 Chemical Shielding Tensors in the B3 Domain of Protein G from N-15 Relaxation Measurements at Several Fields. Implications for Backbone Order Parameters. *J. Am. Chem. Soc.* **2006**, 128, 7855-7870.
- [197] J. F. Ying, A. Grishaev, D. L. Bryce, A. Bax, Chemical Shift Tensors of Protonated Base Carbons in Helical RNA and DNA from NMR Relaxation and Liquid Crystal Measurements. *J. Am. Chem. Soc.* **2006**, 128, 11443-11454.
- [198] R. Fiala, J. Czernek, V. Sklenar, Transverse Relaxation Optimized Triple-Resonance NMR Experiments for Nucleic Acids. *J. Biomol. NMR* **2000**, 16, 291-302.
- [199] D. Stueber, D. M. Grant, C-13 and N-15 Chemical Shift Tensors in Adenosine, Guanosine Dihydrate, 2'-Deoxythymidine, and Cytidine. *J. Am. Chem. Soc.* **2002**, 124, 10539-10551.
- [200] E. Brumovska, V. Sychrovsky, Z. Vokacova, J. Sponer, B. Schneider, L. Trantirek, Effect of Local Sugar and Base Geometry on C-13 and N-15 Magnetic Shielding Anisotropy in DNA Nucleosides. *J. Biomol. NMR* **2008**, 42, 209-223.
- [201] A. P. Dejaegere, D. A. Case, Density Functional Study of Ribose and Deoxyribose Chemical Shifts. *J. Phys. Chem. A* **1998**, 102, 5280-5289.
- [202] D. L. Bryce, A. Grishaev, A. Bax, Measurement of Ribose Carbon Chemical Shift Tensors for A-Form RNA by Liquid Crystal NMR Spectroscopy. *J. Am. Chem. Soc.* **2005**, 127, 7387-7396.



**CURRICULUM VITAE****Jan-Peter M. Ferner**

Falkstrasse 58  
60487 Frankfurt

**Personal Data**

Date of birth January 25<sup>th</sup>, 1977  
Place of birth New Rochelle, NY, USA  
Citizenship German  
Marital status married to Elke Duchardt-Ferner nee Duchardt

**University Education**

## Graduation

11/2002-07/2009 Graduation at the Institute for Organic Chemistry and Chemical Biology under supervision of Prof. Dr. Harald J. Schwalbe

## Course of studies

10/1996-07/2002 Study of Chemistry at the Goethe-University Frankfurt am Main, Certificate Diploma with the diploma thesis title „*Synthese von Galactosederivaten zur Darstellung von unnatürlichen Kohlenhydrat-Aminosäure-Konjugaten*“ under supervision of Prof. Dr. Joachim W. Engels

10/1998 Prediploma in Chemistry

**Schooling**

08/1987-06/1996 Altkönigschule Kronberg (Taunus), Gymnasium (secondary school), Certificate Abitur

---

08/1984-07/1987	Geschwister-Scholl-Schule Steinbach (Taunus), (primary school)
08/1983-07/1984	Siwanoy School Pelham, New York, USA, (primary school)

### Teaching Activities

09/2007-to date	Main assistant of the organic-chemistry practical course for chemistry and biochemistry students under supervision of Prof. Dr. Michael Göbel and Prof. Dr. Harald J. Schwalbe
11/2002-03/2007	Assistant lecturer in the organic-chemistry basis lecture by Prof. Dr. Harald J. Schwalbe, assistant in the organic-chemistry practical course and mentoring of students in workgroup internships

### Further Activities

08/2006	International Conference on Magnetic Resonance in Biological Systems (ICMRBS), Göttingen, promoted talk with the title „ <i>RNA-Tetraloop Dynamics Investigated by <sup>13</sup>C Relaxation Measurements</i> “
11/2005	Meeting of the SFB 579 RNA-Liganden-Wechselwirkungen, Weilburg, talk in cooperation with Julia Dietz with the title „ <i>Selection and functional analysis of peptide ligands for the HIV-1 packaging signal <math>\Psi</math></i> “
01/1999-01/2002	Elected student representative in the faculty council of the faculty 14, Goethe-University Frankfurt am Main

### Academic Teachers

Prof. Dr. W. Aßmus, Prof. Dr. N. Aunuer, Prof. Dr. B. Brutschy, Prof. Dr. E. Egert, Prof. Dr. J. W. Engels, Prof. Dr. Dr. H. Fasold, Prof. Dr. M. Göbel, Prof.

Dr. C. Griesinger, Prof. Dr. K. Hensen, Prof. Dr. G. Kohlmaier, Prof. Dr. B. O. Kolbesen, Prof. Dr. B. Ludwig, Prof. Dr. B. Lüthi, Prof. Dr. T. Prisner, Prof. Dr. D. Rehm, Prof. Dr. E. Reinsch, Prof. Dr. H. Schmidt-Böcking, Prof. Dr. F. Schüth, Prof. Dr. H. J. Schwalbe, Prof. Dr. G. Stock, Prof. Dr. M. Wagner, Prof. Dr. V. Wittmann, Prof. Dr. J. Wöhnert

## ACKNOWLEDGEMENT

First of all, I would like to thank my supervisor Prof. Dr. **Harald J. Schwalbe** to give me the opportunity to work in the field of NMR spectroscopy. His support and encouragement to work self-sufficiently on complex and challenging projects with highly complex NMR machines helped me significantly in my scientific education. Also his large general knowledge of chemistry and biology and his quick grasp of presented topics with fast conclusions and suggestions impressed me over and over.

Special thanks go to **Jens Wöhnert** who taught me the highly complex biochemistry and biophysics of nucleic acids. His change from a critical assessment of my biochemical skills to a high appreciation, helped to build up my self-confidence in the biochemistry lab.

Listed clockwise surrounding my workstation, I would like to thank the first generation of the Boyz Room consisting of **Sergey Ilin, Kai Schlepckow, Jürgen Graf, Steffen Grimm, Boris Fürtig** and **Christian Schlörb**, as well as the second generation consisting of **Martin Hähnke, Friedi Heinicke, Fabian Hiller, Anna Wacker** (nee Fröhlich), **Melanie Koschinat** und **Daniel Mathieu**. Both generations had a significant influence on the high standard and elaborateness of my work. In addition, they all entertained me with discussions about computers, networks, sport, politics etc.

The concerted appearance of the Schwalbe group at several occasions impressed me again and again. This sense of community is a considerable factor why I am happy to be a member of the Schwalbe group. In particular, I want to acknowledge the group members who organized the several group trips enabling me to visit miscellaneous regions near and far from Frankfurt and to participate in memorable evenings.



I am grateful to my collaborators in Frankfurt, Dr. **Ursula Dietrich**, Prof. Dr. **Gerhard Stock** and Prof. Dr. **Michael Göbel** and around the world, Prof. Dr. **Hashim Al-Hashimi** (Ann Arbor, USA), Prof. Dr. **Mirko Hennig** (Charleston, USA) and Prof. Dr. **Yves Mély** (Strasbourg, France) for the outcome of our interdisciplinary projects.

What is a workgroup without a staff? Nothing! Therefore, I am deeply grateful to **Elke Stirnal**, **Sarah Mensch** and **Marie Anders-Maurer** for the efficient HPLC purifications and peptide syntheses despite the amount and complexity of requests from all-around.

I also want to thank Dr. **Christian Richter** and Dr. **Gottfried Zimmermann** aka Ric and Zim for their outstanding work on the maintenance of and help with the NMR spectrometers. I am looking forward to learn more and more about these machines from both of them.

For the big help in the battle against German bureaucracy, I want to send a big thank you to **Anna Paulus** and **Elena Hartmann**. I always like to visit their office, albeit only to make some copies or to grab a cookie or two ;-)

A special thank you goes to my fellow students and colleagues in the Fachschaftsrat. They laid the foundation for my interest in and endurance at the Goethe-University.

For the enormous and never fading support throughout my whole study I want to gratefully thank my family. Hopefully, I can provide with this thesis something to demonstrate my work and help to explain what I have been doing the past years.

Last, but clearly not least I want to acknowledge and thank my wife **Elke**. Our marriage is the most outstanding result of this dissertation. Thank You for everything!

