NMR Studies of RNA

DOI: 10.1002/anie.200906885

High-Resolution Studies of Uniformly ^{13}C,^{15}N-Labeled RNA by Solid-State NMR Spectroscopy**

*Alexey V. Cherepanov, Clemens Glaubitz, and Harald Schwalbe**

Dedicated to Professor Horst Kessler on the occasion of his 70th birthday

Solid-state NMR spectroscopy with magic angle spinning (MAS NMR) is an advanced noninvasive technique to study the structure and dynamics of biologic macromolecules. MAS NMR experiments can be performed in frozen solution, in membranes, and in microcrystalline or freeze-dried proteins. These studies yield information on internuclear distances, torsion angles, molecular orientation, and functional dynamics. MAS NMR spectroscopy provides a unique opportunity to study macromolecules in their own natural environment, in vitro and in vivo, be it a single purified protein,[1] large multiprotein complexes,[2,3] molecular fibrils,[4] cell organelles,[5] the entire cell or tissue,[6,7] or the whole organism.[8] In addition, solid-state NMR spectroscopy emerges as a powerful tool for the time-resolved study of macromolecular folding and catalysis.[9–11] By varying the temperature of the frozen sample, structural and chemical transitions can be selectively trapped or monitored in real time.[11–13]

Herein we apply solid-state NMR spectroscopy for atomic studies on RNA using a cUUCGg tetraloop hairpin as a model. We have recently characterized this hairpin in solution and obtained a refined high-resolution structure (RMSD = 0.3 Å).[14,15] Here, we use ^{13}C MAS NMR spectroscopy to extend our studies to frozen solution, compare the results with solution NMR data, and relate the differences to the hairpin structure. To our knowledge, this is the first high-resolution MAS NMR study on uniformly ^{13}C,^{15}N-labeled RNA. Solid-state ^{2}H NMR spectroscopy was used to describe motion of selected residues in TAR RNA from HIV-1 during protein recognition.[16] NH···N hydrogen bonds in $(CUG)_{97}$ RNA were detected by ^{15}N MAS NMR spectroscopy.[17] ^{1}H correlation was combined with high-resolution spectral dimensions in NHHN, CHHC, and NHHC experiments.[18] Despite the impressive signal intensity, low dispersion precluded assignment of individual spins.

The ^{13}C,^{13}C radio-frequency-driven dipolar recoupling (RFDR) MAS NMR spectrum of the RNA 14-mer is shown in Figure 1. All cross peaks in the spectrum originate from intraresidue correlations. Out of 168 possible short-range direct coherence transfer cross peaks, 158 are found in the solid-state spectrum. For cytidines, the C4C5 correlations could not be detected. For uridines, weak C4C5 cross peaks appeared with a mixing time of 5.74 ms. In total, 116 out of 132 carbon atoms were identified. Ambiguous resonances originate from the C2 and C8 atoms of the bases. Nine of them were assigned because the peaks did not overlap and the solution shifts differed by less than 0.3 ppm (Figure 1, labeled diagonal peaks). In addition to the cross peaks between adjacent carbon atoms, we observed medium-range correlations over a distance of approximately 2.4 Å (C1′C3′, C3′C5′, and C4C6). Cross peaks between the heteroatom-bridged carbons (for example, C1′C4′ and C2C4) were not found, implying a relayed coherence transfer mechanism, for example, C1′→C2′→C3′. Of 112 relayed cross peaks, 30 were observed and 18 assigned (Figure 1, red labels).

Figure 1 shows that 12 relays form four composite cross peaks (Figure 1, peaks 1–4). For purines, the observed relays derived from C4C6 transfer. Similar correlations in pyrimidines were not found. The ^{13}C chemical shift data, the solution NMR data,[14] and calculated differences are summarized in Table S1 in the Supporting Information. Figure 1 and Table S1 indicate that the shifts in the solid state closely correspond to those in solution: 89% of the shifts differ by less than 0.3 ppm. Only six carbon nuclei show differences exceeding 1 ppm.

To obtain structural information, we analyzed the ^{13}C chemical shift data. Such analysis is a common approach in predicting backbone torsion angles and is even used in de novo generation and refinement of protein structures.[19,20] Here, we examine to what extent it can be used for RNA. Each conformation of a D-aldofuranose ring is defined by two steric parameters:[21] P, the phase angle of pseudorotation, and ν_{max}, the degree of pucker. Two torsion angles, χ of the N-glycosidic bond (C1′–N3) and γ of the C5′–C4′ bond define the chemical vicinity of the ring. Ribose chemical shifts are dependent on and, to a large extent, determined by these parameters. However, solving the reverse problem—predicting angles from chemical shifts—is not trivial.

[*] Dr. A. V. Cherepanov, Prof. Dr. H. Schwalbe
Institute of Organic Chemistry and Chemical Biology
Center for Biomolecular Magnetic Resonance (BMRZ)
Johann Wolfgang Goethe-Universität
60438 Frankfurt (Germany)
Fax: (+49) 69-798-29515
E-mail: schwalbe@nmr.uni-frankfurt.de
Prof. Dr. C. Glaubitz
Institute for Biophysical Chemistry
Center for Biomolecular Magnetic Resonance
Johann Wolfgang Goethe-Universität
60438 Frankfurt (Germany)

[**] The work was supported by the EU-funded project EU-NMR. The Center for Biomolecular Magnetic Resonance is funded by the state of Hesse. H.S. and C.G. are members of the DFG-funded Cluster of Excellence "Makromoleculare Komplexe".

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.200906885.

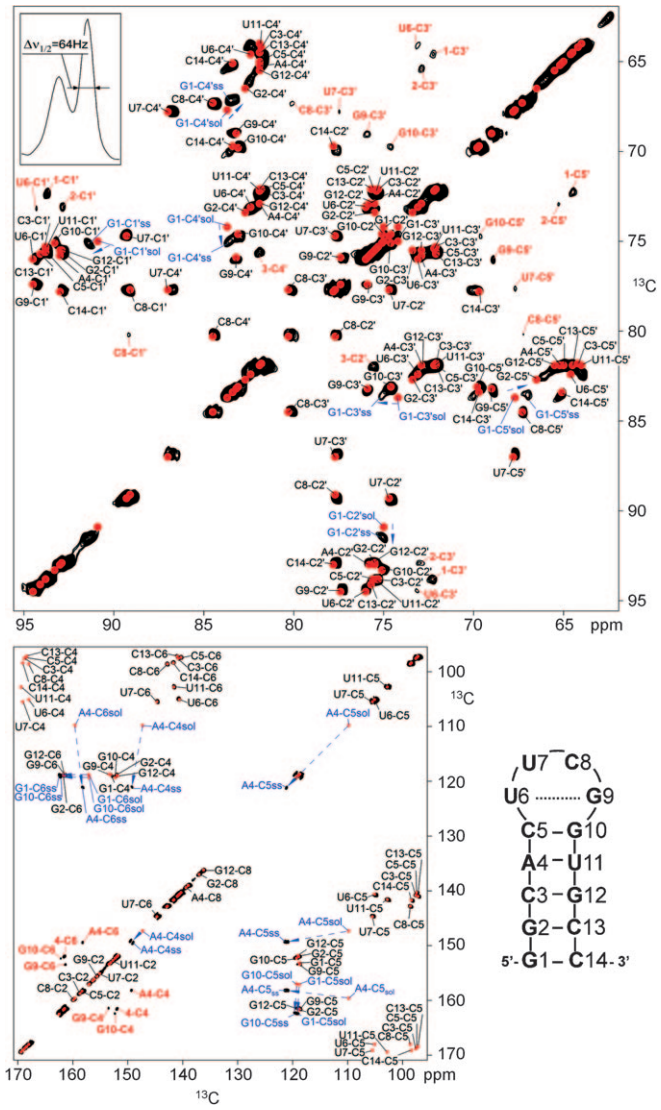

Figure 1. Regions of the $^{13}C,^{13}C$ RFDR MAS NMR spectrum of the 14-mer RNA. Single-bond correlation spectrum (red) is constructed using solution NMR data.[14,15] Relayed cross peaks are labeled in red. The largest differences between the solid-state (ss) and solution (sol) data are marked with arrows, resonances are labeled accordingly. The RFDR linewidth is shown in the inset.

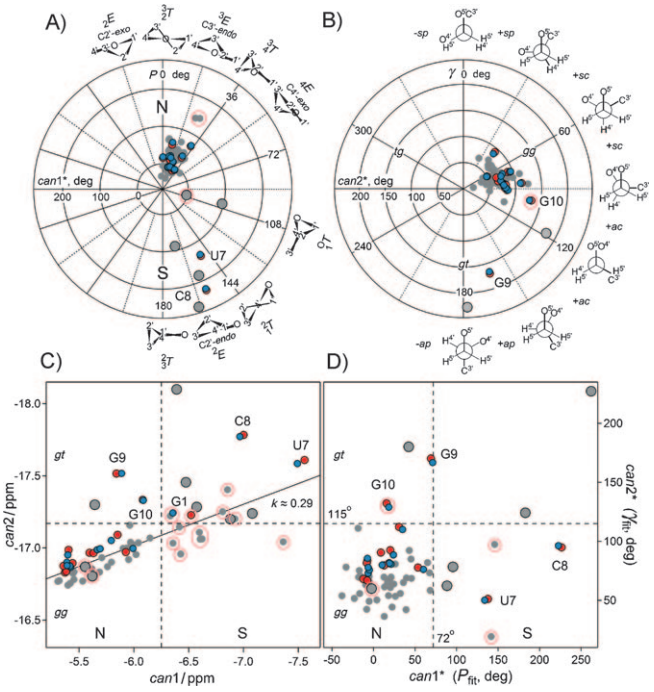

Figure 2. ^{13}C chemical shifts of the ribose C atoms versus pseudorotation phase angle *P* and torsion angle γ. BMRB (biological magnetic resonance bank) entries 4120 (gray) and 5705 (red and blue circles). A) Relation of *can*1* [Eq. (3)] to the pseudoroation phase angle *P*. B) Dependence of *can*2* [Eq. (4)] on the torsion angle γ. C) Canonical coordinates according to Equations (1) and (2). *can*1 values below −6.25 ppm predict ribose conformations with $90 < P < 270°$ (see A). Residues with *can*2 < −17.2 ppm tend to be *gauche–trans (gt)* about the C5′–C4′ bond as shown in (B). D) Canonical coordinates according to Equations (3) and (4). MAS NMR data are shown in blue. Solution NMR data are shown in red. Large gray circles represent residues with *S*-type sugar pucker and/or *+ac*, *+ap* configurations. Residues that are located in the canonical space on the other side of the reaction threshold from the structure-borne values are marked with pink circles. N,S = north, south; *k* = gradient of the solid line calculated by fitting the canonical coordinates using linear regression.

Our approach is based on empiric combinations of the carbon shifts, which correlate with individual conformational parameters. The canonical coordinates *can*1 and *can*2 are defined in Equations (1) and (2).[22] These equations can be used for predicting ribose conformations with different pseudorotation phase and torsion angles *P* and γ (Figure 2C).

$$can1 = 0.179\delta_{C1'} - 0.225\delta_{C2'} - 0.0575\delta_{C5'} \quad (1)$$

$$can2 = -0.0605\delta_{(C2'+C3')} - 0.0556\delta_{C4'} - 0.0524\delta_{C5'} \quad (2)$$

In solution, the ribose moieties of U7 and C8 in the 14-mer RNA show almost exact C2′-*endo* envelope conformation, with $P = 157 \pm 1°$ and $159 \pm 4°$, respectively. G9 has a *gt*-like torsion with $\gamma = 165 \pm 11°$, and G10 has the +anticlinal torsion with $\gamma = 100 \pm 8°$. Figure 2C shows distribution of these conformers in the canonical space. The (*can*1, *can*2) coordinates give a false prediction for the terminal nucleotide G1, which has neither an S-like conformation ($P \approx 12°$) or *gt* torsion ($\gamma \approx 83°$). Liquid- and solid-state data show only minor differences, confirming that the tetraloop structure does not change much upon freezing. The canonical values for G1 in the solid state shift toward the *N/gg* region, indicating that the 5′-end becomes less mobile and thus "more helical", once RNA is trapped in ice.

To test the use of canonical coordinates and to increase sampling, we included data from a 44-mer RNA (BMRB entry 4120, Figure 2, gray circles). With the combined data set, the (*can*1, *can*2) transformation gives 14 false predictions (Figure 2C). The nonorthogonality of (*can*1, *can*2) coordinates is apparent with $can2 \approx 0.3\,can1 - 15$, which can also be seen on a large, structurally diverse RNA data set.[23] To improve separation, we derived a new set of coordinates [Eqs. (3) and (4), see also the Supporting Information].

$$can1^* = -14.7\delta_{C1'} + 22.1\delta_{C2'} + 13.2\delta_{C3'} + 6.5\delta_{C4'} - 2.9\delta_{C5'} - 1595 = P_{FIT} \quad (3)$$

$$can2^* = 9.8\delta_{C1'} + 16.5\delta_{C2'} - 0.5\delta_{C3'} - 1.7\delta_{C4'} 13.5\delta_{C5'} - 2781 = \gamma_{FIT} \quad (4)$$

The coordinate *can*1* spans the angles from −40 to 260°; the actual *P* values are distributed between −10 and 170°; *can*2* covers the range of 20–230° versus 16–180° of the source γ data (Figure 2D). The average angular precision of (*can*1*, *can*2*) is ±30 and ±20°, respectively, which is high enough for the qualitative assessment of an unknown structure. On the 58-residue data set, (*can*1*, *can*2*) coordinates appear more orthogonal than (*can*1, *can*2) and give only four false predictions, which corresponds to a 3.5-fold improvement. The overall prediction fidelity increases from 76 to 93%. Further increase of resolution might be achieved by fitting to the individual torsion angles ν_0–ν_4, γ and χ. Canonical analyses show that the ribose moieties of the 14-mer RNA adopt a very similar conformation in both liquid and frozen states. In Figure 3, the ^{13}C chemical shift differences between these states are mapped on the hairpin structure. Four global trends can be distinguished, suggesting specific modulations of the structure.

The first trend includes the 0.2–0.3 ppm downfield shifts of the C5′, C4′, and C3′ resonances in the sugar backbone solid-state, best seen as the atoms shaded light blue at the edge of the minor groove of the helix. The second trend reflects the opposite 0.1–0.2 ppm upfield shifts of the C1′

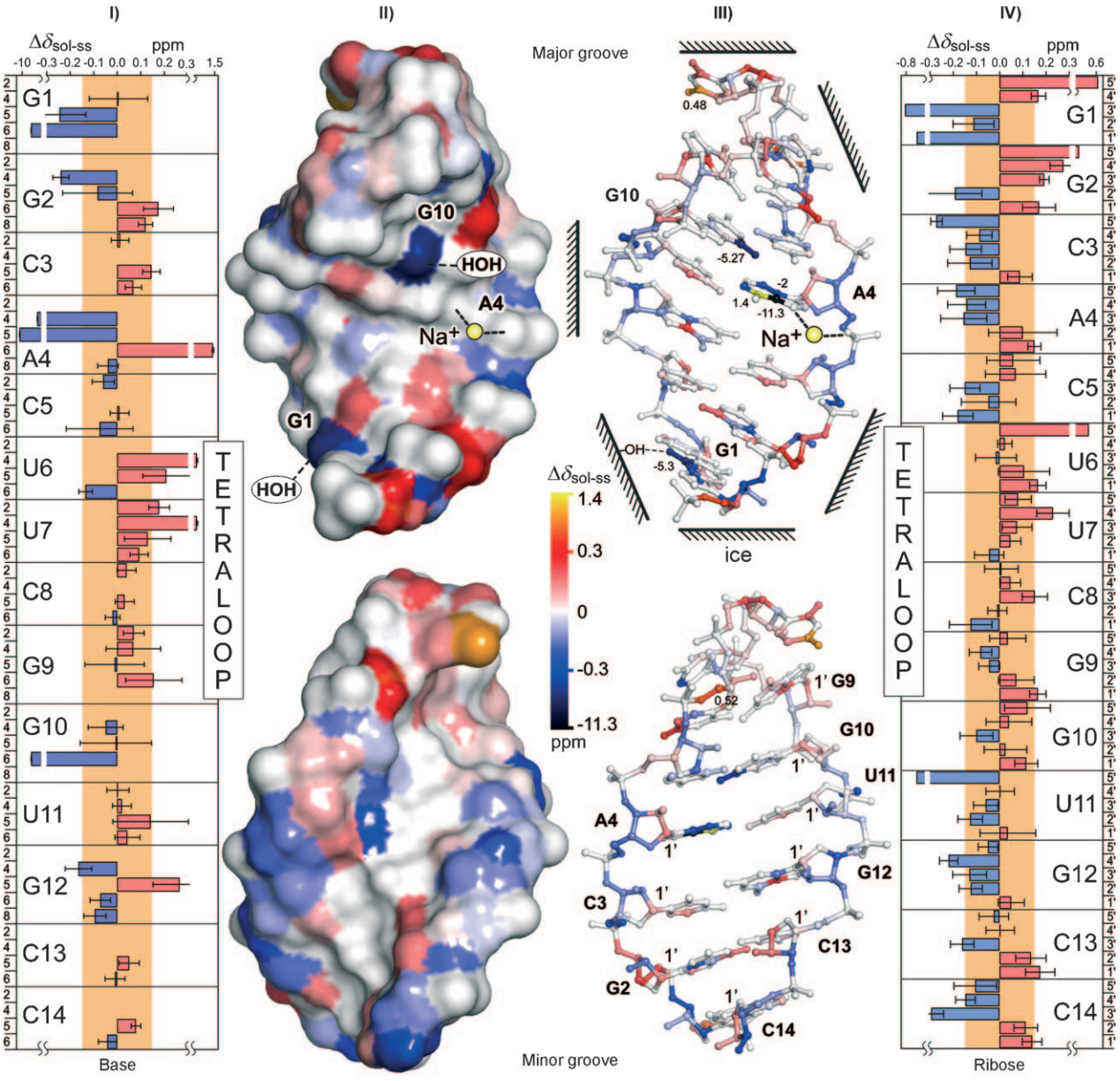

Figure 3. Comparison of solution (sol)[14] and MAS NMR (ss) data: ^{13}C chemical-shift differences. I, IV) Nucleobase and ribose groups, respectively. Error bars show standard deviation. The linewidth in the RFDR experiment is shown in light brown. II, III) Solution minus solid-state chemical-shift difference map for the 14-mer RNA hairpin (PDB entry 2koc).[15] The color bar shows differences in ppm. Positions of the Na^+ ion chelated to A4 and putative hydrogen bonds to external water are indicated. For illustration, oxygen and/or hydrogen atoms are shown in the same color as the carbon atoms to which they are attached.

resonances. The third includes upfield shifts of carbon signals in the tetraloop region, and the fourth relates to the large shifts of G1, A4, and G10 signals. The latter are confined to several nucleobase atoms, while the other signals in the same ring remain unchanged, indicating that the global structure remains the same. The other shifts are at least ten times smaller than the differences between the tetraloop and helix regions (see Table S1 and Figure S3 in the Supporting Information). Hence, the related modulations of the structure should be minor.

Ice lattice has a high affinity to nonfreezing water in the hydration shell of macromolecules.[24,25] The RNA helix in solution is stabilized by water clusters in the major groove and by tandem water molecules in the minor groove.[26] Partial dehydration upon freezing removes some of these molecules.[27] Exclusion of water from the minor groove could account for the upfield shifts of C1′ resonances. The packing of the tetraloop region in the frozen solution could rigidify the structure and thereby lead to the upfield shifts of the carbon signals. Binding of Na^+ ions caused by charge aggregation and formation of Bjerrum ion pairs could shift the signals of the sugar backbone downfield, similar to the case of single nucleotides.[28]

Stacking in nucleic acids is known to shift the ^{13}C resonances of the nucleobases upfield,[29] while hydrogen bonding of the adjacent heteroatoms causes the opposite, downfield shifts.[30,31] In the 14-mer RNA, the C2, C4, C5, and C8 resonances of A4 are shifted upfield relative to free ATP, while the C6 resonance is shifted downfield (see Table S2 in the Supporting Information). Similar shifts of the C6 signals of G1 and G10 could result from hydrogen bonding between O6 and external (solid) water. Chelation of Na^+ ions in the frozen state by N7 and Oα (Figure 3, II and III) could account for the shifts of the carbon signals of the A4 nucleobase.

In summary, we have used ^{13}C MAS NMR spectroscopy to characterize the 14-mer RNA hairpin capped by a cUUCGg tetraloop. Chemical-shift analyses indicate that the structure of the hairpin in ice is highly similar to that in solution. Minor modulation of the structure can be attributed to the partial dehydration of RNA, binding of Na^+ ions, and hydrogen bonding to water molecules at the ice interface. Our results show that biologically relevant RNAs can undergo the water–ice phase transition without significant structural changes and critical loss of NMR resolution and sensitivity. The use of uniformly instead of site-specifically labeled RNA is feasible because correlation experiments reveal remarkably sharp signals and sufficient chemical-shift dispersion. Our findings pioneer freeze-trapping for studying RNA structure and function in folding, ligand recognition, and catalysis. Our results form the basis for the novel molecular analysis of RNAs and their complexes, advocating solid-state NMR spectroscopy to the broader RNA community.

Received: December 7, 2009
Revised: February 16, 2010
Published online: June 8, 2010

Keywords: conformation analysis · freeze-quenching · NMR spectroscopy · RNA · solid-state structures

[1] A. McDermott, *Annu. Rev. Biophys.* **2009**, *38*, 385.
[2] A. Alia, S. Ganapathy, H. J. M. de Groot, *Photosynth. Res.* **2009**, *102*, 415.
[3] I. Kawamura, Y. Ikeda, Y. Sudo, M. Iwamoto, K. Shimono, S. Yamaguchi, S. Tuzi, H. Saito, N. Kamo, A. Naito, *Photochem. Photobiol.* **2007**, *83*, 339.
[4] O. N. Antzutkin in *Modern Magnetic Resonance*, *Vol. 1* (Ed.: G. E. Webb), Springer, Dordrecht, **2006**, pp. 19.
[5] M.-A. Sani, O. Keech, P. Gardestrom, E. J. Dufourc, G. Grobner, *FASEB J.* **2009**, *23*, 2872.
[6] J. Kelly, H. Jarrell, L. Millar, L. Tessier, L. M. Fiori, P. C. Lau, B. Allan, C. M. Szymanski, *J. Bacteriol.* **2006**, *188*, 2427.
[7] T. E. Sjobakk, R. Johansen, T. F. Bathen, U. Sonnewald, R. Juul, S. H. Torp, S. Lundgren, I. S. Gribbestad, *NMR Biomed.* **2008**, *21*, 175.
[8] H. Winning, N. Viereck, B. Wollenweber, F. H. Larsen, S. Jacobsen, I. Sondergaard, S. B. Engelsen, *J. Exp. Bot.* **2009**, *60*, 291.
[9] J. N. S. Evans, R. J. Appleyard, W. A. Shuttleworth, *J. Am. Chem. Soc.* **1993**, *115*, 1588.
[10] A. V. Cherepanov, E. V. Doroshenko, J. Matysik, S. de Vries, H. J. de Groot, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 8563.
[11] A. V. Cherepanov, E. V. Doroshenko, J. Matysik, S. de Vries, H. J. De Groot, *Phys. Chem. Chem. Phys.* **2008**, *10*, 6820.
[12] A. V. Cherepanov, S. De Vries, *Biochim. Biophys. Acta Bioenerg.* **2004**, *1656*, 1.
[13] A. L. Fink, G. A. Petsko in *Advances in Enzymology and Related Areas of Molecular Biology* (Ed.: M. Alton), **2006**, p. 177.
[14] B. Fürtig, C. Richter, W. Bermel, H. Schwalbe, *J. Biomol. NMR* **2004**, *28*, 69.
[15] S. Nozinovic, B. Fürtig, H. R. Jonker, C. Richter, H. Schwalbe, *Nucleic Acids Res.* **2010**, *38*, 683.
[16] G. L. Olsen, M. F. Bardaro, Jr., D. C. Echodu, G. P. Drobny, G. Varani, *J. Am. Chem. Soc.* **2010**, *132*, 303.
[17] J. Leppert, C. R. Urbinati, S. Hafner, O. Ohlenschläger, M. S. Swanson, M. Görlach, R. Ramachandran, *Nucleic Acids Res.* **2004**, *32*, 1177.
[18] C. Herbst, K. Riedel, Y. Ihle, J. Leppert, O. Ohlenschläger, M. Görlach, R. Ramachandran, *J. Biomol. NMR* **2008**, *41*, 121.
[19] Y. Shen, O. Lange, F. Delaglio, P. Rossi, J. M. Aramini, G. Liu, A. Eletsky, Y. Wu, K. K. Singarapu, A. Lemak, A. Ignatchenko, C. H. Arrowsmith, T. Szyperski, G. T. Montelione, D. Baker, A. Bax, *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 4685.
[20] K. Seidel, M. Etzkorn, R. Schneider, C. Ader, M. Baldus, *Solid State Nucl. Magn. Reson.* **2009**, *35*, 235.
[21] *Pure Appl. Chem.* **1983**, *55*, 1273.
[22] M. Ebrahimi, P. Rossi, C. Rogers, G. S. Harbison, *J. Magn. Reson.* **2001**, *150*, 1.
[23] O. Ohlenschläger, S. Haumann, R. Ramachandran, M. Görlach, *J. Biomol. NMR* **2008**, *42*, 139.
[24] A. S. Goryunov, *Gen. Physiol. Biophys.* **2006**, *25*, 303.
[25] M. Falk, A. G. Poole, C. G. Goymour, *Can. J. Chem.* **1970**, *48*, 1536.
[26] M. Egli, S. Portmann, N. Usman, *Biochemistry* **1996**, *35*, 8489.
[27] S. A. Kazakov, S. V. Balatskaya, B. H. Johnston, *RNA* **2006**, *12*, 446.
[28] H. A. Tajmir-Riari, T. Theophanides, *Can. J. Chem.* **1985**, *63*, 2065.
[29] M. P. Stone, S. A. Winkle, P. N. Borer, *J. Biomol. Struct. Dyn.* **1986**, *3*, 767.
[30] P. N. Borer, S. R. LaPlante, N. Zanatta, G. C. Levy, *Nucleic Acids Res.* **1988**, *16*, 2323.
[31] E. A. Boudreau, I. Pelczer, P. N. Borer, G. J. Heffron, S. R. LaPlante, *Biophys. Chem.* **2004**, *109*, 333.

