Protein-RNA Interactions

DOI: 10.1002/anie.201208024

Protein–RNA Interfaces Probed by ¹H-Detected MAS Solid-State NMR Spectroscopy**

Sam Asami, Magdalena Rakwalska-Bange, Teresa Carlomagno, and Bernd Reif*

In the past decades, solid-state nuclear magnetic resonance (NMR) spectroscopy has evolved to become an important tool, enabling structural investigation of crystalline^[1] and, most notably, noncrystalline biomolecular systems.^[2] So far, mostly protein samples have been studied in the solid state, while only few investigations of RNA or DNA have been reported.[3] However, many fundamental biochemical processes are carried out by protein-RNA complexes, the most prominent of which is the translation machinery. Prior to the assembly of the eukaryotic ribosome, which catalyzes translation of messenger RNA, the RNA of the ribosomal subunits becomes extensively modified.^[4] The modifications consist of the conversion of uridine to pseudouridine and the methylation of nucleobases and 2'-O-ribose units, which cluster predominantly within functionally crucial regions.^[5] It has been suggested that the modifications play an important role in RNA folding. [6] In eukaryotes, the site-specific methylation of 2'-O-ribose units is guided by small nucleolar RNAs (snoRNAs).[7] These snoRNAs are defined by two conserved sequence elements, the C box (RUGAUG, where R is purine) and D box (CUGA).[8] The snoRNAs are complexed with specific proteins and assemble to form ribonucleoprotein particles (snoRNPs). In archaea, the small ribonucleoprotein particle (sRNP) is composed of a box C/D snoRNA-like RNA (sRNA) and three proteins, L7Ae, Nop56/58, and fibrillarin, of which fibrillarin is the methyltransferase.^[9] The assembly of archaeal sRNP is initiated by the binding of L7Ae to box C/D RNA. Upon binding, the sRNA folds into a Kturn motif.[9b,10] Herein we suggest a new approach to characterize protein-RNA interfaces, which is based on proton-detected magic-angle-spinning (MAS) solid-state NMR experiments, using protonated and deuterated samples.

Solid-state NMR spectroscopy is not subject to size limitations and is therefore best suited for the study of large macromolecular complexes that cannot be crystallized. [11] Even using MAS frequencies of about 60 kHz yields only partial averaging of the ¹H, ¹H dipolar coupling network in protonated protein samples. [12] It has been demonstrated recently for various proteins, however, that already at moderate spinning frequencies (20–24 kHz) and external magnetic fields (9–14 T), high sensitivity and resolution can be obtained using deuteration and back-exchange of labile protons to enable ¹H detection. [13] Similarly, non-exchangeable protons are observable with specific labeling schemes. [14]

We apply here a similar approach to study the protein–RNA interactions between the archaeal L7Ae protein (ca. 13.4 kDa) from *Pyrococcus furiosus* (PF) with a 26-mer box C/D RNA (ca. 8.6 kDa; Figure 1). So far, the crystal structure of a homologous complex from *Archaeoglobus fulgidus* (AF) was solved by X-ray diffraction. [9b] The complex from PF could not be crystallized. [3d] For the solid-state NMR

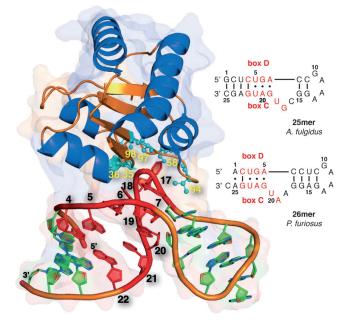


Figure 1. Crystal structure of the L7Ae box C/D RNA complex from Archaeoglobus fulgidus (PDB: 1RLG). In this study, the homologous L7Ae from *Pyrococcus furiosus* in complex with the illustrated 26-mer box C/D RNA from the same organism was investigated. The box C/D elements are highlighted in red, the protein residues at the interaction sites are shown in cyan (Gly35, Thr36, Asp58, Val94, Ala97, Ser98); β-sheets and loops of L7Ae are in gold, helices in marine blue.

[*] S. Asami, Prof. Dr. B. Reif

Munich Center for Integrated Protein Science (CIPS-M) at Department Chemie, Technische Universität München (TUM) Lichtenbergstrasse 4, 85747 Garching (Germany)

Helmholtz-Zentrum München (HMGU)

Deutsches Forschungszentrum für Gesundheit und Umwelt Ingolstädter Landstrasse 1, 85764 Neuherberg (Germany) E-mail: reif@tum.de

M. Rakwalska-Bange, Dr. T. Carlomagno Computational and Structural Biology Unit European Molecular Biology Laboratory (EMBL) Meyerhofstrasse 1, 69117 Heidelberg (Germany)

[**] We thank the Leibniz-Institut für Molekulare Pharmakologie (FMP) for providing measurement time and U. Fink for technical assistance in the lab. This work was performed in the framework of SFB 1035 of the German Research Foundation. MAS = magic-angle-spinning.



Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201208024.

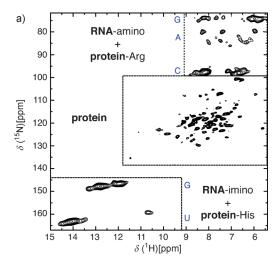


studies, two differently labeled L7Ae box C/D RNA complexes were prepared. In both cases, the protein was uniformly deuterated and isotopically ¹⁵N-enriched. The RNA was synthesized from uniformly ¹H, ¹³C, ¹⁵N- or ²H, ¹³C, ¹⁵N-labeled nucleotides. The samples are referred to as the ¹H- and ²H-RNA samples, respectively. In both samples, labile protons were back-exchanged to a level of about 10%.

In ¹H-detected ¹H, ¹⁵N-correlation experiments at moderate MAS frequencies in the range of 15-20 kHz all expected resonances from the protein and the RNA were detected (Figure 2a). In protein-RNA complexes, the spectral windows for protein and RNA resonances do not significantly overlap.^[15] The imino and amino spectral regions can in principle contain resonances of histidine and arginine side chains, respectively (one histidine and four arginine residues are found in the primary sequence of L7Ae). The average ¹H (15N) linewidth for the protein is on the order of roughly 100 (60) Hz, whereas the linewidths for the amino and imino resonances amount to about 150 (140) and 300 (160) Hz, respectively. The differential linewidths for the protein and the RNA resonances might be a result of chemical exchange, local dynamics, or structural heterogeneity resulting from crystal imperfections. Superposition of the solid-state with a solution-state spectrum reveals a high structural similarity, with comparable resolution in both states (Figure 2b).

Herein we suggest an experimental strategy for the identification of the protein-RNA interface. The method is based on the fact that in the ¹H-RNA sample, protein regions in close proximity to RNA protons show reduced signal intensities because of dipole-mediated line broadening. Amide protein moieties, which are distant from RNA protons (>6 Å), show similar signal intensities for both the ¹H- and ²H-RNA samples $(I[^{1}H]/I[^{2}H] \approx 1)$. On the other hand, protein amide resonances at the interface display a smaller intensity in the ¹H-RNA sample $(I[^{1}H]/I[^{2}H] < 1)$. Since the absolute intensities are also dependent on the amount of the sample, we normalized the intensities for the two samples prior to calculating the $I[^{1}H]/I[^{2}H]$ ratios. For this purpose, we employed Ala48, which is highly resolved (Figure 2b) and quite distant from the nearest RNA ¹H^C proton (14.7 Å). Crystal contacts are taken into account (Figure S4d in the Supporting Information). For the calculation of the relative intensities, only the contacts between the protein ¹H^N backbone and RNA ¹H^C protons were considered, as exchangeable protons are back-exchanged at a level of 10% and, to a first approximation, ¹H^{N,1}H^N dipole interactions can be neglected.

In Figure 3a the smallest distance between a protein backbone ¹H^N proton and an RNA ¹H^C proton is plotted as a function of the protein residue number. The distances are extracted from the AF crystal structure (PDB: 1RLG), which can be used as a model owing to its high homology (Figure S4a in the Supporting Information). The minima in the distance plot (Figure 3a) indicate protein residues at the protein–RNA interface, in particular Gly35, Thr36, Asp58, Val94, Ala97, and Ser98. The proximity of protein residues to the RNA can be obtained experimentally by the comparison of the normalized intensities of the ¹H- and the ²H-RNA samples. The same contour levels are employed for each set of



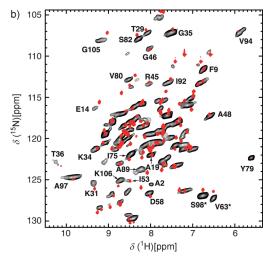


Figure 2. a) ¹H-detected ¹H, ¹⁵N correlation spectrum of the precipitated L7Ae box C/D RNA complex from PF, employing a ²H-RNA sample. Clearly, the characteristic protein and RNA spectral patterns are observed. The processing was optimized for each region, to account for different natural line widths of the biomolecules. The spectrum was recorded at an external field of 16.4 T (700 MHz), 20 kHz MAS frequency, and an effective temperature of roughly 27°C. The imino region showed enhanced sensitivity and resolution at lower temperatures. Therefore this region was extracted from a spectrum recorded at 0°C, 15 kHz rotation frequency, and an external field of 14.1 T (600 MHz). The spectral regions for the nucleobases, as expected from the average chemical shift in the BMRB database, are indicated. b) Protein region of the ¹H-RNA sample with tentative assignments. The solution-state NMR spectrum is shown in red. Ser98 and Val63 (indicated by asterisks) are folded.

cross peaks to allow for comparison (Figure 3b). For example, the intensities of Gly35, Thr36, and Ala97 are significantly attenuated in the ¹H-RNA sample, since these residues are close to non-exchangeable ¹H^C RNA protons (within 3.6–4.3 Å), unlike Thr29 and Ser82, which are far apart (>11 Å).

To quantitatively describe the experimental $I[^1H]/I[^2H]$ intensity ratios, we carried out simulations, employing the SIMPSON software package. To construct an adequate spin system, we considered a spin geometry consisting of four nuclei, resembling the protein $^1H^{N,15}N$ nuclei and the RNA $^1H^{C,13}C$ nuclei (Figure S2 in the Supporting Information).

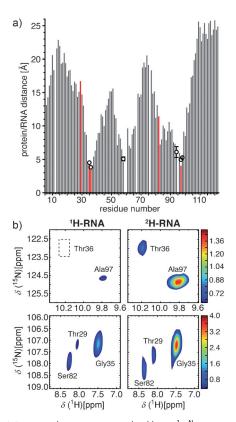


Figure 3. a) Distance between protein backbone ¹H^N proton and the closest RNA ¹H^C proton as a function of the protein primary sequence. Calculated distances are based on the crystal structure of the L7Ae box C/D RNA complex from AF (PDB: 1RLG; protons were added using WHAT IF^[17]). Missing bars are due to proline residues. Experimentally determined distances are plotted as empty circles. b) ¹H, ¹⁵N correlation plots, employing the ¹H- (left) and ²H-RNA (right) samples. Absolute intensities were normalized using Ala48. The illustrated residues (Thr29, Gly35, Thr36, Ser82, Ala97) are highlighted by red bars in (a). As expected, the protein residues, which are in close proximity to the RNA (Gly35, Thr36, Ala97), show a reduced intensity in the ¹H-RNA sample.

Since the signal attenuation in the ¹H, ¹⁵N correlation of the protein in close proximity to the RNA is induced by dipolemediated line broadening, the simulated I[1H]/I[2H] ratio is expected to be dependent on the distance, as well as on the orientation of all considered dipoles. Therefore, we introduced two angles, β_1 and β_2 , to allow sampling of all orientations of the ¹H^{N,15}N and ¹H^{C,13}C dipole tensors with respect to each other. Besides the β_1 and β_2 angles, the only other variable is the ¹H^{N,1}H^C distance. All other dipolar couplings and orientations were calculated accordingly. As expected, we find that the simulated $I[^{1}H]/I[^{2}H]$ ratios depend on the relative orientation of the two dipole-dipole vectors. This dependency decreases with increasing distance. Furthermore, the simulation reveals an up to fourfold degeneracy in the angular space (Figure S2). To reduce the degeneracy, we determined the β_1 and β_2 angle distribution for the L7Ae box C/D RNA complex as well as for two additional protein-RNA complexes and used their common angle interval for constraining the angular space during the fitting procedure (Figure S3). The simulated data matrix, $I[^{1}H]/I[^{2}H]$, as a function of the $^1H^N, ^1H^C$ distance, $\beta_1,$ and β_2 is represented in Figure 4a.

The fit of the experimental data does not assume prior knowledge of the structure of the system. As a first attempt, we assumed a linear arrangement of the ¹H^{N,15}N and ¹H^{C,13}C dipole tensors, which yields a poor correlation (Figure 4b,

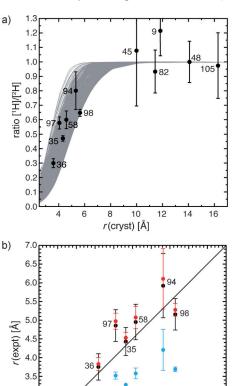


Figure 4. a) Fit of the experimentally determined /[1H]//[2H] intensity ratios. The experimental I[1H]/I[2H] ratios are plotted in black as a function of r(cryst), the ¹H^N(protein), ¹H^C(RNA) distance, which was extracted from the crystal structure of the L7Ae box C/D RNA complex (PDB: 1RLG). The full array of simulated curves employed for the fitting procedure is plotted in gray. The fitted distances are shown in Figure 3 a and given in Table S1 in the Supporting Information. The error of the experimental intensities was set to two times the noise rmsd. The uncertainty of the intensity ratios were then determined by error propagation. b) Correlation plot of ¹H^N(protein), ¹H^C(RNA) distances extracted from the crystal structure, r(cryst), and from the experimental $I[^1H]/I[^2H]$ ratios by simulation, r(expt). For the fitting procedure, three different cases were assumed: 1) $\beta_1 \in [0^\circ, 180^\circ]$, $\beta_2 \in [0^{\circ}, 360^{\circ}]$ (black), 2) $\beta_1 \in [0^{\circ}, 180^{\circ}]$, $\beta_2 \in [70^{\circ}, 100^{\circ}]$ (red), and 3) $\beta_1 = \beta_2 = 0^{\circ}$ (cyan). For (2), the angular space of β_2 was constrained, as described in the Supporting Information (Figure S3). Obviously, introducing the angles β_1 and β_2 significantly improves the fit and yields a linear correlation between the experiment and the distance extracted from the X-ray structure. When the intercept of a linear function was set to zero, the obtained slopes were roughly 1.05, 1.08, and 0.76 for the nonlinear/unrestricted, the nonlinear/restricted, and the linear arrangements, respectively. In this manner, for protein-RNA complexes a robust structural restraint can be obtained that does not rely on prior structural knowledge. The error of the experimental distances r(expt) was estimated by 1000 Monte Carlo simulations using a normal distribution for the intensity ratios I[1H]/I[2H].

4.0 4.5 5.0

r(cryst) [Å]

5.5 6.0 6.5

3.0

2.0

2.5 3.0 3.5



cyan). However, introducing the angles β_1 and β_2 improves the correlation significantly (Figure 4b, black). As expected, restricting the β_2 angular space results in smaller errors; however, it does not significantly change the absolute determined distances (Figure 4b, red). Details concerning the fitting procedure are given in the Supporting Information. The determined parameters are summarized in Table S1 and the obtained distances are plotted in Figure 3a, together with the distances from the crystal structure. The estimated distances and angles are within the error margins. For the experiments, we used only about 2 mg of the 2 H-RNA sample. Obviously, a higher fitting accuracy can be achieved by increasing the experimental sensitivity for the two samples.

It should be noted that the proton density of RNA is naturally much lower than that of a protein. Here, the low proton density simplifies the determination of structural parameters, since, to first approximation, every protein ¹H^N proton is in proximity to only very few RNA ¹H^C protons.

In conclusion, we could show that high-resolution spectra for a protein–RNA complex can be obtained utilizing deuterated samples. To the best of our knowledge, this is the first time that ¹H-detected ¹H, ¹⁵N correlation spectra of a protein–RNA complex are presented in the solid state. Furthermore, by introducing a novel approach for the determination of structural parameters in the solid state, we could identify the protein–RNA interface and quantify the distance and orientation for interacting protein residues.

The fact that the extinction of amide proton resonances in the ¹H-RNA sample depends on the distance and the relative orientation of the atom pairs opens new possibilities for structure calculation protocols. In addition to a distance restraint, an angular restraint can now be employed for systems, for which only sparse data is available. This should make it possible to determine complex structures with improved accuracy. These parameters can be incorporated as restraints into an iterative structure calculation algorithm. We believe that this approach can be applied to other protein-RNA complexes, as well as to protein-ligand complexes in general. The L7Ae box C/D RNA complex served here as a model system. In the future, we aim to investigate fully assembled sRNP[18] complexes which exceed the molecular weight limit imposed by molecular tumbling, a current limitation in solution-state NMR applications.

Experimental Section

The L7Ae protein from *P. furiosus* was expressed and purified as described elsewhere; the 26-mer RNA was transcribed and purified according to established procedures. [3d] For the ¹H- and ²H-RNA samples u-[¹H,¹³C,¹⁵N] and u-[²H,¹³C,¹⁵N] nucleotides were employed, respectively. Both assembled complexes (1:1 molar ratio) were back-exchanged in a buffer with a H₂O/D₂O ratio of 10%/90% for 2–4 weeks to allow for sufficient H/D exchange. Prior to precipitation, the protein concentration was 7 mg mL⁻¹ (in 10%/90% H₂O/D₂O buffer with 40 mm HEPES, 14% PEG 400, and 50 mm magnesium acetate). The precipitation of the complex was induced by vapor diffusion using a SpeedVac at 60 mbar and 30°C. As expected from the crystallization conditions, [9b] the complex precipitated at half volume. The final precipitates of the ¹H- (ca. 10 mg) and ²H-RNA (ca. 2 mg) complexes were filled into 3.2 mm rotors by centrifugation.

The solid-state NMR experiments were carried out using Bruker Biospin Avance spectrometers operating at ^1H Larmor frequencies of 600 and 700 MHz, employing commercial 3.2 mm triple-resonance probes. Details about temperature and MAS settings are given in the Figure legends. The ^1H , ^{15}N magnetization transfer was achieved by a ramped cross polarization step, matching the n=1 Hartmann–Hahn condition. The ^{15}N evolution was ^1H - and ^{13}C -decoupled, whereas the ^1H evolution was ^{15}N -decoupled. For all nuclei, low-power WALTZ16 decoupling with a RF field strength of 2–3 kHz was applied.

Received: October 4, 2012 Published online: January 18, 2013

Keywords: NMR spectroscopy · proteins · RNA · proton detection · structural restraints

- a) F. Castellani, B. van Rossum, A. Diehl, M. Schubert, K. Rehbein, H. Oschkinat, *Nature* 2002, 420, 98–102; b) W. T. Franks, B. J. Wylie, H. L. F. Schmidt, A. J. Nieuwkoop, R. M. Mayrhofer, G. J. Shah, D. T. Graesser, C. M. Rienstra, *Proc. Natl. Acad. Sci. USA* 2008, 105, 4621–4626.
- [2] a) C. Wasmer, A. Lange, H. Van Melckebeke, A. B. Siemer, R. Riek, B. H. Meier, Science 2008, 319, 1523-1526; b) J. M. Lopez Del Amo, U. Fink, M. Dasari, G. Grelle, E. E. Wanker, J. Bieschke, B. Reif, J. Mol. Biol. 2012, 421, 517-524; c) J. M. Lopez del Amo, M. Schmidt, U. Fink, M. Dasari, M. Fandrich, B. Reif, Angew. Chem. 2012, 124, 6240-6243; Angew. Chem. Int. Ed. 2012, 51, 6136-6139; d) A. Loquet, N. G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, C. Griesinger, M. Kolbe, D. Baker, S. Becker, A. Lange, Nature 2012, 486, 276-279
- [3] a) K. Riedel, C. Herbst, S. Hafner, J. Leppert, O. Ohlenschlaeger, M. S. Swanson, M. Gorlach, R. Ramachandran, Angew. Chem. 2006, 118, 5748-5751; Angew. Chem. Int. Ed. 2006, 45, 5620-5623; b) A. V. Cherepanov, C. Glaubitz, H. Schwalbe, Angew. Chem. 2010, 122, 4855-4859; Angew. Chem. Int. Ed. 2010, 49, 4747-4750; c) W. Huang, G. Varani, G. P. Drobny, J. Am. Chem. Soc. 2010, 132, 17643-17645; d) S. Jehle, M. Falb, J. P. Kirkpatrick, H. Oschkinat, B. J. van Rossum, G. Althoff, T. Carlomagno, J. Am. Chem. Soc. 2010, 132, 3842-3846; e) W. Huang, G. Varani, G. P. Drobny, J. Biomol. NMR 2011, 51, 347-356; f) I. V. Sergeyev, L. A. Day, A. Goldbourt, A. E. McDermott, J. Am. Chem. Soc. 2011, 133, 20208-20217.
- [4] B. E. Maden, Prog. Nucleic Acid Res. Mol. Biol. 1990, 39, 241 303.
- [5] W. A. Decatur, M. J. Fournier, Trends Biochem. Sci. 2002, 27, 344-351.
- [6] M. Helm, Nucleic Acids Res. 2006, 34, 721-733.
- [7] a) Z. Kiss-Laszlo, Y. Henry, T. Kiss, EMBO J. 1998, 17, 797 807;
 b) T. Kiss, EMBO J. 2001, 20, 3617 3622.
- [8] a) K. Tyc, J. A. Steitz, EMBO J. 1989, 8, 3113 3119; b) K. Q. Ye,
 R. Jia, J. Z. Lin, M. H. Ju, J. Peng, A. B. Xu, L. M. Zhang, Proc. Natl. Acad. Sci. USA 2009, 106, 13808 13813.
- [9] a) A. D. Omer, S. Ziesche, H. Ebhardt, P. P. Dennis, *Proc. Natl. Acad. Sci. USA* 2002, *99*, 5289-5294; b) T. Moore, Y. M. Zhang, M. O. Fenley, H. Li, *Structure* 2004, *12*, 807-818; c) S. Oruganti, Y. M. Zhang, H. Z. Li, H. Robinson, M. P. Terns, R. M. Terns, W. Yang, H. Li, *J. Mol. Biol.* 2007, *371*, 1141-1150.
- [10] M. Falb, I. Amata, F. Gabel, B. Simon, T. Carlomagno, *Nucleic Acids Res.* 2010, 38, 6274–6285.
- [11] a) A. Mainz, S. Jehle, B. J. van Rossum, H. Oschkinat, B. Reif, J. Am. Chem. Soc. 2009, 131, 15968-15969; b) A. Mainz, B. Bardiaux, F. Kuppler, G. Multhaup, I. C. Felli, R. Pierattelli, B. Reif, J. Biol. Chem. 2012, 287, 1128-1138.



- [12] A. Marchetti, S. Jehle, M. Felletti, M. J. Knight, Y. Wang, Z. Q. Xu, A. Y. Park, G. Otting, A. Lesage, L. Emsley, N. E. Dixon, G. Pintacuda, *Angew. Chem. Int. Ed.* 2012, 51, 10756–10759.
- [13] a) V. Chevelkov, K. Rehbein, A. Diehl, B. Reif, Angew. Chem. 2006, 118, 3963-3966; Angew. Chem. Int. Ed. 2006, 45, 3878-3881; b) M. Huber, S. Hiller, P. Schanda, M. Ernst, A. Bockmann, R. Verel, B. H. Meier, ChemPhysChem 2011, 12, 915-918; c) M. J. Knight, A. L. Webber, A. J. Pell, P. Guerry, E. Barbet-Massin, I. Bertini, I. C. Felli, L. Gonnelli, R. Pierattelli, L. Emsley, A. Lesage, T. Herrmann, G. Pintacuda, Angew. Chem. 2011, 123, 11901-11905; Angew. Chem. Int. Ed. 2011, 50, 11697-11701; d) R. Linser, M. Dasari, M. Hiller, V. Higman, U. Fink, J. M. L. del Amo, S. Markovic, L. Handel, B. Kessler, P. Schmieder, D. Oesterhelt, H. Oschkinat, B. Reif, Angew. Chem. 2011, 123, 4601-4605; Angew. Chem. Int. Ed. 2011, 50, 4508-4512.
- [14] a) V. Agarwal, A. Diehl, N. Skrynnikov, B. Reif, J. Am. Chem. Soc. 2006, 128, 12620-12621; b) S. Asami, P. Schmieder, B. Reif, J. Am. Chem. Soc. 2010, 132, 15133-15135; c) P. Schanda, M. Huber, J. Boisbouvier, B. H. Meier, M. Ernst, Angew. Chem. 2011, 123, 11198-11202; Angew. Chem. Int. Ed. 2011, 50, 11005-11009.
- [15] B. Fürtig, C. Richter, J. Wohnert, H. Schwalbe, *ChemBioChem* 2003, 4, 936–962.
- [16] M. Bak, J. T. Rasmussen, N. C. Nielsen, J. Magn. Reson. 2000, 147, 296–330.
- [17] G. Vriend, J. Mol. Graphics 1990, 8, 52-56.
- [18] J. Z. Lin, S. M. Lai, R. Jia, A. B. Xu, L. M. Zhang, J. Lu, K. Q. Ye, Nature 2011, 469, 559 – U140.
- [19] A. J. Shaka, J. Keeler, T. Frenkiel, R. Freeman, J. Magn. Reson. 1983, 52, 335–338.

2349