

# SCIENTIFIC REPORTS



OPEN

## RNA structure refinement using NMR solvent accessibility data

Christoph Hartmüller<sup>1,2</sup>, Johannes C. Günther<sup>1,2</sup>, Antje C. Wolter<sup>3</sup>, Jens Wöhnert<sup>3</sup>, Michael Sattler<sup>1,2</sup>  & Tobias Madl<sup>1,2,4</sup> 

Received: 13 March 2017

Accepted: 2 June 2017

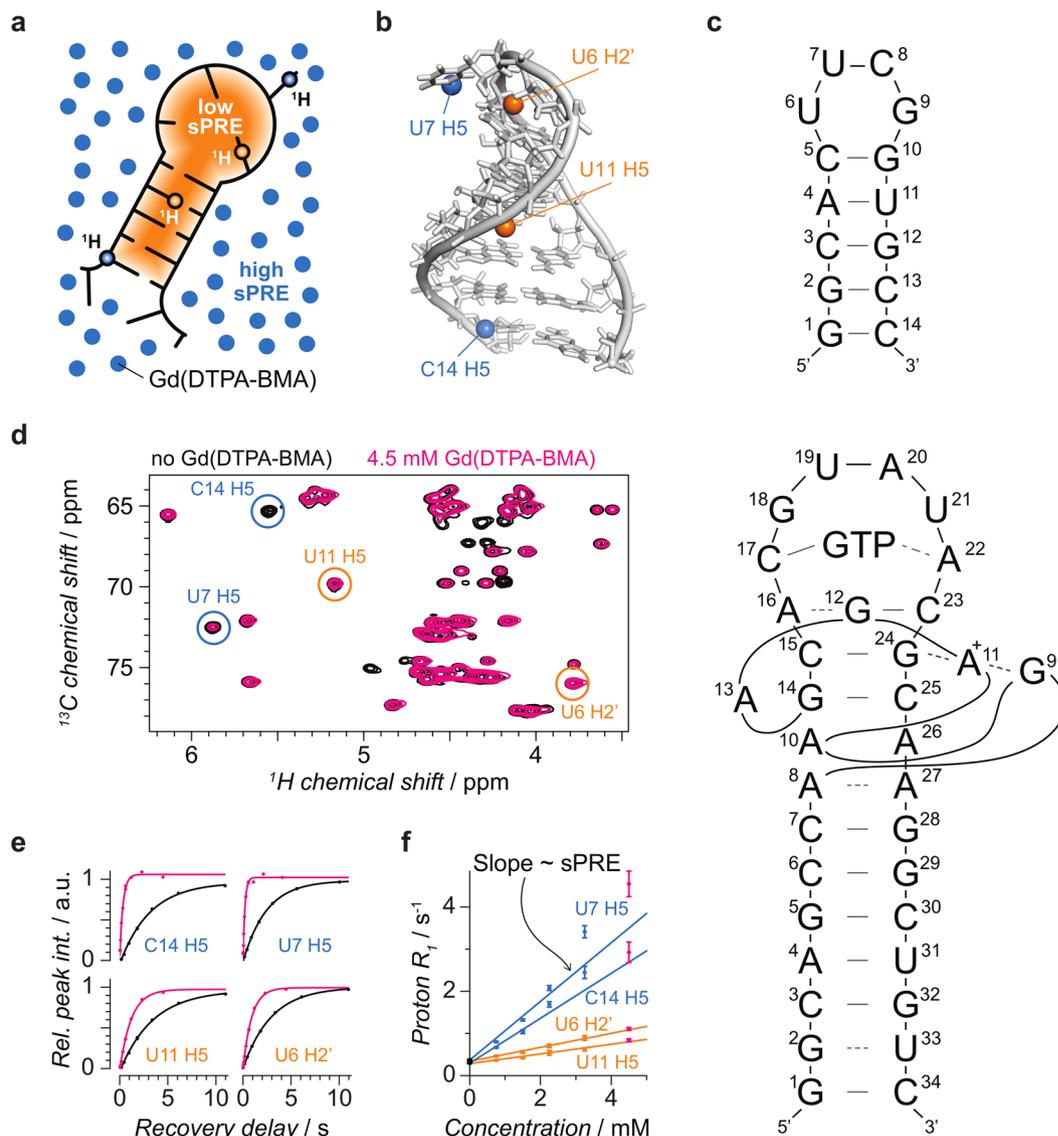
Published online: 14 July 2017

**NMR spectroscopy is a powerful technique to study ribonucleic acids (RNAs) which are key players in a plethora of cellular processes. Although the NMR toolbox for structural studies of RNAs expanded during the last decades, they often remain challenging. Here, we show that solvent paramagnetic relaxation enhancements (sPRE) induced by the soluble, paramagnetic compound Gd(DTPA-BMA) provide a quantitative measure for RNA solvent accessibility and encode distance-to-surface information that correlates well with RNA structure and improves accuracy and convergence of RNA structure determination. Moreover, we show that sPRE data can be easily obtained for RNAs with any isotope labeling scheme and is advantageous regarding sample preparation, stability and recovery. sPRE data show a large dynamic range and reflect the global fold of the RNA suggesting that they are well suited to identify interaction surfaces, to score structural models and as restraints in RNA structure determination.**

In the last decades, ribonucleic acids (RNAs) have been found to be key regulators for numerous important cellular processes including transcription, translation, splicing, cell differentiation as well as cell death and proliferation<sup>1</sup>. Besides regulatory functions, RNAs are essential for catalysis in ribosomes and spliceosomes, sensing environmental parameters such as temperature or metabolite concentration, as well as storing the genomic information of RNA viruses. Understanding the underlying molecular mechanisms of these processes requires insights into the structure and dynamics at an atomic level. In the last decades, NMR spectroscopy has developed to a powerful technique to study RNA structure and dynamics<sup>2–8</sup>. However, technical challenges are caused by poor chemical shift dispersion leading to spectral overlap as well as by the low proton density and the low number of intramolecular interactions limiting the number of observable distance restraints that can be used to define RNA structure. Moreover, the RNA backbone is defined by several torsion angles, requiring a large set of restraints to obtain an accurate structural model. To overcome these limitations, different labeling strategies, including for example specific or uniformly <sup>13</sup>C- and/or <sup>15</sup>N-labeling and segmental labelling of RNAs were developed to reduce spectral complexity and signal overlap<sup>9–22</sup>. Large sets of restraints including NOE-based distances, torsion angles and orientational restraints derived from RDCs and J-couplings, respectively, as well as base pairing contacts obtained from J-couplings across hydrogen bonds are typically collected to provide a sufficient number of restraints for high-resolution models of RNAs<sup>2, 6, 23–26</sup>.

In previous NMR studies, solvent accessibility data were obtained by measuring the NOE between water and protein protons, but the data were found to be dominated by chemical exchange<sup>27, 28</sup>. Here, we demonstrate that solvent accessibility data derived from solvent paramagnetic relaxation enhancements (sPREs) provide valuable information to characterize the conformation of RNAs and will be useful to enable NMR studies of larger RNAs. sPRE data are a quantitative measure for solvent accessibility and thus readily provide distance-to-surface information<sup>29</sup> (Fig. 1). Directly observing the solvent accessible areas is a promising approach not only for mapping interaction surfaces but also for structure determination, as has been shown for proteins<sup>30–32</sup>. Recently, computational protocols using sPRE data were developed for XplorNIH<sup>27</sup> and the Rosetta framework<sup>33</sup>. The implementation and use of sPRE data has several advantages: i) no covalent labeling of RNAs is required, ii) the sample can be recovered, iii) complete chemical shift assignment is not required and iv) any type of NMR spectrum such as

<sup>1</sup>Center for Integrated Protein Science Munich, Department Chemie, Technical University of Munich, Lichtenbergstr. 4, 85748, Garching, Germany. <sup>2</sup>Institute of Structural Biology, Helmholtz Zentrum München, Ingolstadter Landstr. 1, 85764, Neuherberg, Germany. <sup>3</sup>Institut für Molekulare Biowissenschaften and Zentrum für Biomolekulare Magnetische Resonanz (BMRZ), Goethe-Universität Frankfurt, Max-von-Laue Str. 9, 60438, Frankfurt/M, Germany. <sup>4</sup>Institute of Molecular Biology and Biochemistry, Center of Molecular Medicine, Medical University of Graz, Harrachgasse 21, 8010, Graz, Austria. Christoph Hartmüller and Johannes C. Günther contributed equally to this work. Correspondence and requests for materials should be addressed to T.M. (email: [tobias.madl@medunigraz.at](mailto:tobias.madl@medunigraz.at))



**Figure 1.** Concept of sPRE and exemplary sPRE data of the UUCG tetraloop. **(a)** sPRE data provide a quantitative measure for solvent accessibility and encode distance-to-surface information. sPRE data are acquired by titrating the RNA with the paramagnetic compound Gd(DTPA-BMA). **(b)** The NMR solution structure of the UUCG tetraloop (PDB code 2KOC) is shown and for illustration purpose, two solvent-exposed protons (blue spheres) and two buried protons (orange spheres) are highlighted. **(c)** Schematic representation of the UUCG tetraloop (top) and the GTP-bound GTP aptamer. **(d)** NMR spectra of the UUCG tetraloop are shown in the absence (black) and presence (magenta) of Gd(DTPA-BMA) with circles around the peaks corresponding to the protons shown in **(b)**. **(e)** Quantitative sPRE data are obtained by measuring the longitudinal proton  $R_1$  relaxation rate as a function of the concentration of Gd(DTPA-BMA). **(f)** Proton  $R_1$  rates increase linearly with the concentration of the paramagnetic compound and the slopes correspond to the sPRE values of the respective resonances highlighted in **(b)**.

proton 1D,  $^1\text{H}$ ,  $^{15}\text{N}$  HSQC or  $^1\text{H}$ ,  $^{13}\text{C}$  HSQC experiments can be used. Direct measurements of solvent accessibility data of RNAs have been reported previously using the soluble compound TEMPOL<sup>34</sup>. While the correlation of TEMPOL-induced sPREs agreed with RNA structure for some parts, the data were limited to a qualitative validation of models. Moreover, unexpectedly large sPREs were found for certain nucleotides and it was later suggested that TEMPOL forms intermolecular hydrogen bonds leading to preferred binding to specific sites<sup>35</sup>. Here, sPRE data are obtained by titrating the RNA sample with Gd(DTPA-BMA), a soluble, well-characterized  $\text{Gd}^{3+}$  chelating paramagnetic contrast agent that was originally developed for MRI imaging<sup>31,36</sup>.

## Results

**Gd(DTPA-BMA) provides quantitative solvent-accessibility data of RNAs.** To first assess a potential binding to specific RNA moieties, fingerprint spectra in the absence and presence of the paramagnetic compound were recorded for two RNAs exhibiting distinct folds and structural elements. The resulting NMR spectra

of the UUCG tetraloop<sup>37,38</sup> and the GTP class II aptamer<sup>39,40</sup> are shown in Supplementary Figures 1 and 2, respectively. As expected, the presence of Gd(DTPA-BMA) causes line-broadening of NMR signals. However, even in the presence of the paramagnetic compound, the vast majority of peaks are still well observable and the absence of chemical shift perturbations confirms the absence of specific binding of the compound to the RNA (Fig. 1d and Supplementary Figures 1 and 2). Thus Gd(DTPA-BMA) can be used as a paramagnetic probe that screens the accessible surface area of RNA molecules.

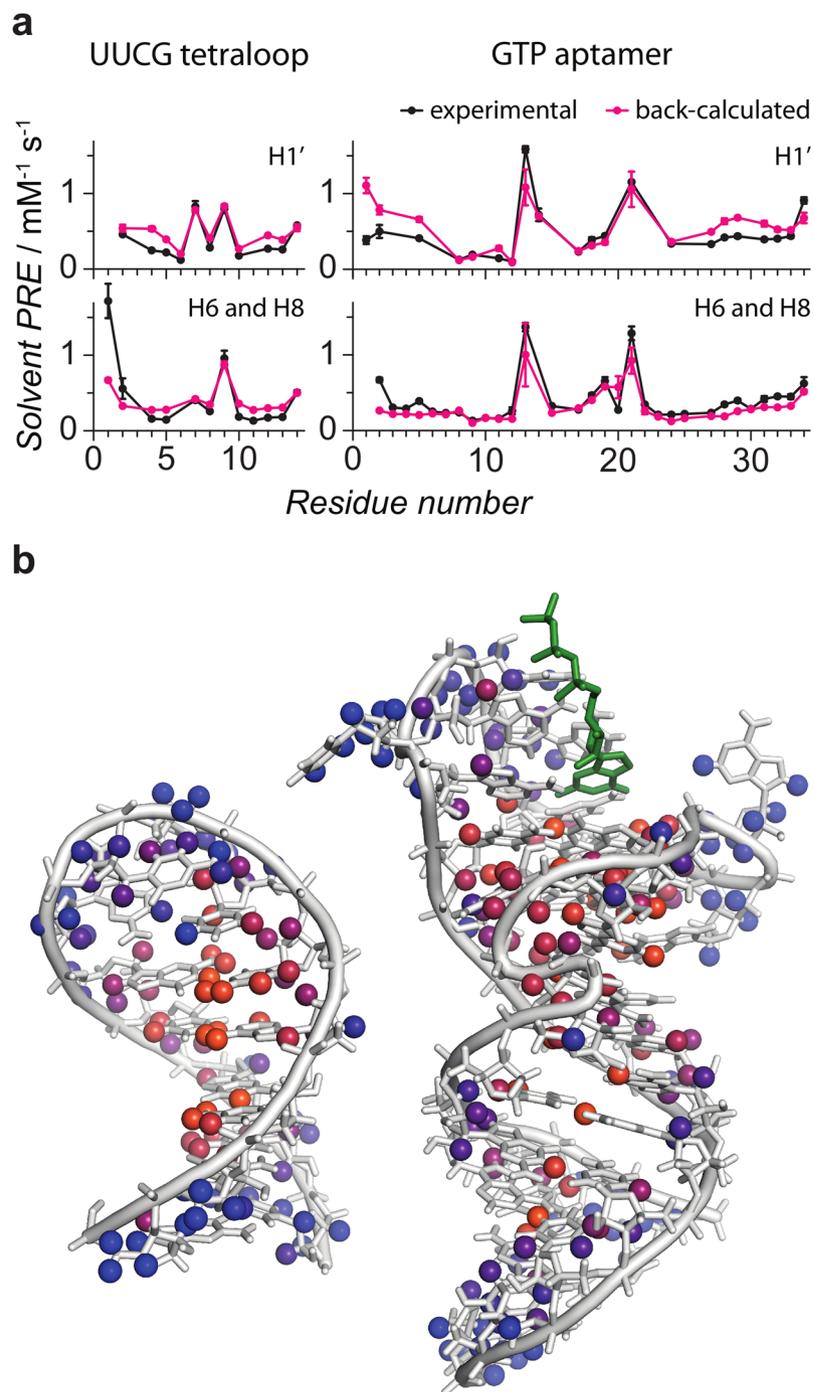
As unchelated lanthanide ions can efficiently hydrolyze RNA<sup>41</sup>, we checked for potential cleavage products of the RNAs in the presence of the Gd<sup>3+</sup> compound. No degradation products were observed for any RNA studied and sPRE data obtained from a single RNA sample were reproducible even after several sPRE measurements, each followed by a removal of Gd(DTPA-BMA) (data not shown). The absence of phosphodiester hydrolysis can be explained by the very high chelating stability of DTPA-BMA<sup>36</sup> and the presence of a slight excess of Ca<sup>2+</sup>-bound chelator.

The sPRE data were acquired by measuring proton longitudinal relaxation rates (<sup>1</sup>H-*R*<sub>1</sub>) as a function of increasing concentrations of Gd(DTPA-BMA). As expected, <sup>1</sup>H-*R*<sub>1</sub> rates were found to increase linearly with the concentration of Gd(DTPA-BMA) in the case of carbon-bound protons. For imino and amino protons, the increase of <sup>1</sup>H-*R*<sub>1</sub> rates is larger at small concentrations of the paramagnetic compound (below about 1 mM), and becomes linear at higher concentrations (Supplementary Figure 3). This observation likely reflects an additional exchange contribution that mixes the relaxation mechanisms of water and RNA protons. Nevertheless, a quantitative measure for the solvent accessibility can be obtained for nitrogen-bound exchangeable protons by determining the minimum concentration of Gd(DTPA-BMA) for which a linear response is observed. Only titration points at this concentration or above are used to fit the linear sPRE model (0.5 mM or above for the UUCG tetraloop and 1.29 mM or above for the GTP aptamer; compare Supplementary Tables 1 and 2). For details about data acquisition and analysis, refer to the methods section in the supplementary information.

**sPRE data correlate well with RNA structure.** Next, we analyzed how sPRE data reflect the structural features of the RNAs studied. To this end, the NMR structures of the UUCG tetraloop and the GTP-bound GTP class II aptamer were used to predict the expected sPRE data using a previously published grid-based approach<sup>30,31</sup> (see method section for details). The experimental sPRE data correlate very well with the predicted data (Fig. 2a and Supplementary Figures 4a and 5). This demonstrates that sPRE data provide a quantitative measure of solvent-accessibility that correlates well with RNA structure. In particular, all buried protons of both RNAs show a small sPRE. Protons for which significant deviations between experimental and back-calculated sPREs are observed are found in loop regions or in terminal nucleotides. These regions are flexible and thus are expected to show a population-weighted average of the sPRE. The structural NMR ensembles used to predict the sPRE data represent the most stable conformations in solution and depend on the algorithm and experimental restraints used for the structure determination. Thus, the NMR ensembles do not necessarily reflect the equilibrium distribution of the different conformations of these solvent-exposed regions of the RNA. Since all conformations of the ensemble were equally weighted in the prediction of the sPRE data, the prediction is expected to deviate for dynamic regions of the RNA. It is further worth noting, that despite the compact fold of the UUCG tetraloop, the experimental sPRE data show a large dynamic range, which renders them a powerful source to characterize the structure of RNAs. The excellent correlation between experimental data and RNA structure in combination with a large dynamic range suggests sPRE data as promising parameters to probe RNA structures by NMR.

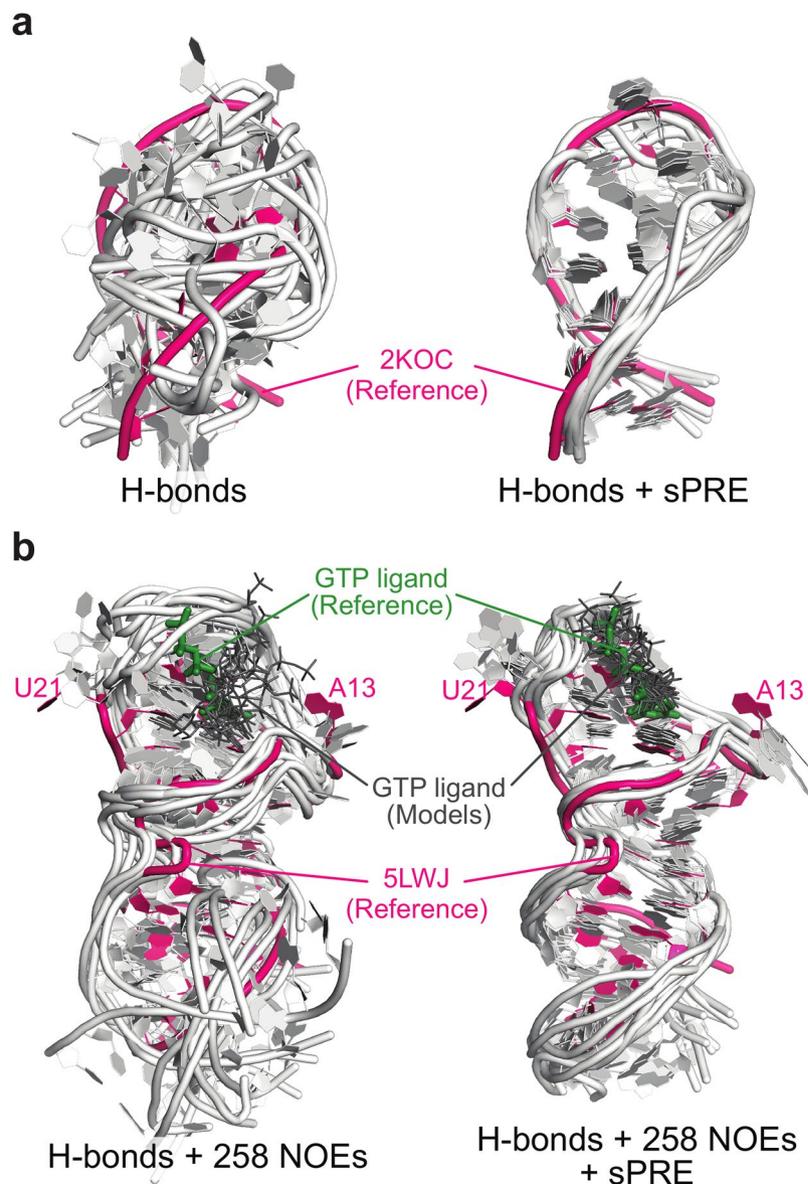
As described above, significant deviations between predicted and measured sPRE data are observed for some terminal nucleotides. Among all 258 observed sPRE values (93 for the UUCG tetraloop and 165 for the GTP aptamer), only two protons that are located in non-terminal nucleotides show significant deviations (9Gua H1 and 6Ura H3 in the UUCG tetraloop). Both protons are located in the loop region of the RNA in close spatial proximity to each other (Supplementary Figure 4b) and the corresponding sPRE values are underestimated based on the NMR solution structure, independent of which structural model of the UUCG loop motif (PDB codes 1HLX, 1K2G, 1TLR, 1Z31, 2KHY, 2KOC, 2KZL, 2LHP, 2LUB and 2N6X) was used. Although the sPRE of these protons are expected to have an additional contribution due to chemical exchange with water, the sPRE values of 2.2 and 0.44 s<sup>-1</sup>mM<sup>-1</sup> of 6Ura H3 and 9Gua H1, respectively, are significantly larger compared to the average of 0.079 ± 0.037 s<sup>-1</sup>mM<sup>-1</sup> for all other imino protons. This suggests that conformational dynamics in this region may cause the unusually high experimental sPREs. Indeed, recent molecular dynamics studies that included a large set of NMR relaxation data of the UUCG tetraloop, suggest the loop region to be more flexible than the RNA stem<sup>42–48</sup>. Moreover, intermediate structural states with the bases of the loop region experiencing higher solvent exposure were observed in these studies. This suggests that the increased sPREs are a result of increased flexibility and dynamics of the loop region. However, future studies are required to fully understand the usability of sPRE data to detect dynamics in RNAs.

**sPRE data provide orthogonal restraints for RNA structure determination.** Next, the potential of sPRE data to facilitate the structure determination of RNAs was investigated (Fig. 3). To this end, structural models of the UUCG tetraloop as well as the GTP-bound aptamer were computed using the XplorNIH framework<sup>27,49–51</sup> in combination with different sets of experimental NOE data (Supplementary Tables 3 and 4, Supplementary Figure 7). The results show that sPRE data significantly improves convergence and accuracy of the structure determination, in particular in cases where only limited sets of NOEs are available. For example, using only 258 experimental NOEs of the GTP-bound aptamer (30% of the all NOEs), the average RMSD of the obtained models was significantly reduced from 7.80 Å to 3.51 Å (Fig. 3b; Random set 2 in Supplementary Table 3). Next, the performance of the sPRE potential in combination with other experimental NMR-based restraints was investigated (Supplementary Table 5). The benchmark revealed that solvent accessibility data is an orthogonal restraint since its usage improves structure determination of the UUCG tetraloop in combination with NOE, RDC and



**Figure 2.** sPRE data correlates well with RNA structure. **(a)** Experimental sPRE data (black) are compared to predicted sPRE data based on NMR solution structures of the UUCG tetraloop (2KOC) and the GTP-bound aptamer (5LWJ). Data are shown for sugar protons (H1') and aromatic protons (H6 and H8) as indicated. **(b)** NMR solution structures of the UUCG tetraloop<sup>37,38</sup> (2KOC, left) and the GTP-bound aptamer<sup>39,40</sup> (5LWJ, right, heavy atoms of GTP ligand shown in green) are shown. All protons for which a sPRE value was obtained are shown as spheres and colored according to the sPRE (blue corresponding to high and orange corresponding to low sPRE values).

torsion angle restraints. Strikingly, using only hydrogen bonds-derived restraints in combination with sPRE data, the structure of the UUCG tetraloop can be determined with an RMSD of 2.8 Å compared to the published NMR-based structure (Fig. 3a). Even sparse sPRE data sets, in some cases down to 25%, improve structural quality of the UUCG tetraloop (Supplementary Table 6).



**Figure 3.** sPRE data improve structure determination of RNAs. Structural models of the UUCG tetraloop (**a**) and the GTP-bound GTP aptamer (**b**) were obtained without (left) and with (right) sPRE data using XplorNIH. The 10 best scored models (light gray) in terms of total energy were selected from a total of 200 models and aligned to the corresponding NMR structure (magenta). All restraints used in the computations are indicated below the respective models. In (**b**) heavy atoms of the GTP ligand are shown as sticks (Reference in green, computed models in dark gray) and the positions of the intrinsically flexible nucleotides A13 and U21<sup>40</sup> are indicated.

## Discussion

In summary, we show that sPRE data are efficient NMR observables to probe RNA structure. sPRE data obtained with the paramagnetic compound Gd(DTPA-BMA) yield quantitative information of solvent accessibility as they correlate well with RNA structure, provide structural information for both buried and surface-exposed atoms and can be used as restraints to drive molecular dynamics-based structure determination of RNAs. Since sPRE data reflect the global fold of a RNA they are well suited to identify tertiary contacts or map interaction surfaces with other molecules<sup>32</sup>, for example, in RNA-protein complexes. As the measurement of sPRE data does not require complete chemical shift assignments, surface accessibility restraints can be recorded for structural studies also in the case of large RNAs, in particular in combination with specific or segmental labelling, where it is difficult to obtain a comprehensive set of experimental restraints for structural analysis. Compared to proteins, several considerations have to be taken into account for sPRE-based studies of RNAs. First, different NMR pulse sequences might be needed, in particular for triple resonance experiments. Second, multiple data sets acquired for different RNA labelling schemes and different solvent conditions need to be combined. In order to compensate for minor differences between these data sets we have proposed an approach in which we normalize the sPRE data using

the sPRE of water protons in each sample. Third, lower concentrations of Gd(DTPA-BMA) are used for RNAs compared to proteins due to the higher sPRE caused by the lack of large hydrophobic cores resulting in smaller distances to the RNA surface. Last, chemical exchange of protons with water is typically more efficient in RNAs and modulates sPREs of nitrogen-bound amino and imino protons. In these cases, only the region of linear response is used for quantitative analysis.

Our results show that sPRE-derived restraints are particularly promising to define the global fold of larger RNAs and provide valuable information for structure determination that is orthogonal to other NMR-based restraints in molecular dynamics-based approaches. Recently, accuracy of chemical-shift based RNA *de novo* structure prediction using the Rosetta framework was significantly enhanced by improving the chemical shift scoring function for RNAs<sup>52</sup>. Combining this improved score with a sampling algorithm driven by sPRE data<sup>33</sup> and specific or segmental labeling could provide powerful tools for structural modeling of large RNAs based on NMR data in the future.

## Methods

**RNA sample preparation.** A uniformly <sup>13</sup>C, <sup>15</sup>N labeled sample of the UUCG tetraloop (5'-GGCACUUCGGUGCC-3') was purchased from Silantes (Munich, Germany). The NMR buffer contained 20 mM K<sub>2</sub>HPO<sub>4</sub> and 0.4 mM EDTA and was adjusted to pH 6.4. Transcription and purification of the GTP class II aptamer in complex with GTP was been described previously<sup>39</sup>. The sample was measured in NMR buffer containing 25 mM KH<sub>2</sub>PO<sub>4</sub>, 25 mM KCl and 2 mM magnesium acetate. The buffer was adjusted to pH 6.3 and uniformly <sup>13</sup>C, <sup>15</sup>N labeled RNA was measured in the presence of a two-fold excess of unlabeled GTP. NMR assignments and a solution NMR structure of the UUCG tetraloop were obtained from literature<sup>37, 38</sup> (PDB code 2KOC, BMRB 5705). For the GTP class II aptamer in complex with GTP, the recently reported NMR assignments and solution NMR structure were used<sup>39, 40</sup> (PDB code 5LWJ, BMRB 25661).

**NMR Experiments.** NMR sPRE data were obtained by measuring proton *R*<sub>1</sub> relaxation rates as a function of the concentration of Gd(DTPA-BMA) using a saturation-recovery approach which has been described previously for protein applications<sup>30</sup>. Briefly, a 7.5 ms proton trim pulse followed by a gradient is used to dephase magnetization. Next, longitudinal magnetization is recovered during a recovery delay of variable length. The recovery delay is followed by a read-out spectrum, such as a <sup>1</sup>H, <sup>15</sup>N or <sup>1</sup>H, <sup>13</sup>C HSQC. This saturation-recovery scheme allows to use a short interscan delay (about 50 ms) which dramatically reduces experimental time compared to measuring proton *R*<sub>2</sub> rates. Pseudo-3D experiments were acquired for 4 to 7 different concentrations of the paramagnetic compound. To recover the NMR RNA sample, the paramagnetic compound was first removed by dialysis against water using a membrane with a 1 kDa molecular weight cut-off and then, the sample was lyophilized and resuspended in the corresponding NMR buffer.

sPRE data for nitrogen-bound protons of the UUCG tetraloop were recorded with a 0.8 mM uniformly <sup>13</sup>C, <sup>15</sup>N labeled sample at 283 K on an Avance III Bruker 900 MHz NMR spectrometer equipped with a cryo-TXI probe head. Proton *R*<sub>1</sub> relaxation rates were measured in the absence and presence of 0.5, 1, 1.75, 2.75 and 4.0 mM Gd(DTPA-BMA) using <sup>1</sup>H, <sup>15</sup>N HSQC-based pseudo-3D experiments. sPRE data for carbon-bound protons were acquired with the same sample at 298 K and *R*<sub>1</sub> relaxation rates were measured in the absence and presence of 0.75, 1.5, 2.25, 3.25 and 4.5 mM Gd(DTPA-BMA) using <sup>1</sup>H, <sup>13</sup>C HSQC-based pseudo-3D experiments. Delay times and other NMR parameters for all experiments as well as the total NMR time are presented in Supplementary Table 1. To obtain sPRE data of the GTP aptamer in complex with GTP, samples containing about 150 μM of (<sup>13</sup>C, <sup>15</sup>N)-GU, (<sup>13</sup>C, <sup>15</sup>N)-C or (<sup>13</sup>C, <sup>15</sup>N)-A labeled RNA and 300 μM unlabeled GTP were measured at 293 K on an Avance III Bruker 900 MHz NMR spectrometer equipped with a cryo-TXI probe head and an Oxford Instruments 600 MHz NMR spectrometer equipped with a Bruker Avance III console and a cryo-TCI probe head. <sup>1</sup>H, <sup>13</sup>C and <sup>1</sup>H, <sup>15</sup>N HSQC-based pseudo-3D experiments were used to acquire proton *R*<sub>1</sub> relaxation rates for 4 to 7 different concentrations of Gd(DTPA-BMA). Delay times and other NMR parameters for all experiments as well as the total NMR time are presented in Supplementary Table 2. For all RNA samples, amino and imino protons were measured in buffer containing 5 to 10% D<sub>2</sub>O whereas carbon-bound protons were acquired in buffer containing ≥99.990% D<sub>2</sub>O (Sigma-Aldrich). For every titration step of Gd(DTPA-BMA), solvent water *R*<sub>1</sub> was measured and the derived sPRE of the water solvent was used to normalize the RNA sPRE values. The normalized sPRE data were then rescaled by the mean sPRE of the solvent, obtained by averaging the sPRE of water protons in all relevant experiments. This procedure allows to combine sPRE measurements of different samples or different magnetic field strengths (e.g. in different buffers or with different labeling schemes; see Supplementary Figure 6). Furthermore, by plotting the solvent water *R*<sub>1</sub> against the concentration of Gd(DTPA-BMA) errors in pipetting can be detected and compensated accordingly if necessary.

In the case of nitrogen-bound protons, plotting the proton *R*<sub>1</sub> against the concentration of Gd(DTPA-BMA) revealed a non-linear correlation with rates at low concentrations of the paramagnetic compound being lower than expected according to a linear model. Consequently, the *R*<sub>1</sub> for low concentrations (0 mM in the case of the UUCG tetraloop, as well as 0 and 0.645 mM in the case of the GTP aptamer) were omitted in the computations of the sPRE values and only the linear correlation was used (compare Supplementary Table 1 and 2 as well as Supplementary Figure 3). In all cases, at least 4 different concentrations were used to derive the sPRE.

For large RNAs, specific isotope labeling schemes are required to reduce signal overlap<sup>9, 10, 53</sup>. In a simple and straight-forward approach, only one or two nucleotide types are enriched in NMR active nuclei. While this approach reduces spectral complexity, the measurement of sPREs with different samples might introduce systematic errors, due to pipetting errors. In addition, the usage of different magnetic field strengths affects the scaling of the sPRE data<sup>31</sup>. To ensure that sPRE data acquired on multiple samples and field strengths are consistent, the sPRE data sets are scaled by referencing to the sPRE of the water solvent in each sample. This referencing notably improves the correlation of calculated and experimental sPRE (Supplementary Figure 6). To normalize the sPRE

data, the sPRE of water protons was acquired by measuring the  $R_1$  rate of the water proton signal. The  $R_1$  rate was measured using a pseudo 2D experiments with simple 1D proton read-out (the proton pulse length was set between 0.3 and 1  $\mu$ s). The change of the  $R_1$  rate of the water protons during the titration Gd(DTPA-BMA) with was used to obtain the sPRE of the water protons. The sPRE of the water solvent was then used to normalize the sPRE values of the RNA signals. The normalized sPRE values of different titration experiments (with different labeling schemes or at different field strengths) were then combined to generate a large sPRE data set for the corresponding RNA. To obtain an absolute quantity, the normalized sPRE data were then multiplied by the mean sPRE of the solvent, obtained by averaging the sPRE of water protons in all relevant experiments. This procedure allows to combine sPRE measurements of different samples or different magnetic field strengths (e.g. in different buffers or with different labeling schemes; see Supplementary Figure 6). Furthermore, by plotting the solvent water  $R_1$  against the concentration of Gd(DTPA-BMA) errors in pipetting can be detected and compensated accordingly if necessary.

The measurement times for the collection of sPRE data can be quite long, as sPRE data are recorded as a series of proton  $R_1$  relaxation experiments. Typical  $^1\text{H}$ - $R_1$  rates in the absence of Gd(DTPA-BMA) are relatively slow with an average  $^1\text{H}$ - $R_1$  rate of  $0.23\text{ s}^{-1}$  ( $\pm 0.10\text{ s}^{-1}$ ) at 900 MHz and  $0.30\text{ s}^{-1}$  ( $\pm 0.14\text{ s}^{-1}$ ) at 600 MHz. As a consequence, long delay times of several seconds (Supplementary Table 2) are required which in turn increase the overall experimental time. To overcome this drawback, the data sets of both RNAs were re-evaluated by determining the sPREs based on  $R_1$  rates measured only in the presence of 0.75, 1.5, 3.25 and 4.5 mM Gd(DTPA-BMA) for the UUCG tetraloop and 0.8, 1.6, 3.2 and 4 mM for the GTP aptamer. Notably, the obtained sPRE data are fully consistent with those derived from a full set of measurements, but the optimized acquisition scheme reduces experimental time by about 40%. It should be noted, that experimental time could be further reduced by applying selective saturation schemes that employ longitudinal relaxation-enhancing techniques as demonstrated in a previous study<sup>54</sup>.

**NMR data analysis.** NMR spectra were processed with the NMRpipe software package<sup>55</sup>. Peak positions from literature were transferred using CcpNmr Analysis<sup>56</sup> and peak intensities of well-resolved peaks were obtained using the nmrglue Python package<sup>57</sup>. Peak intensities of the pseudo-3D experiments were fitted to an exponential recovery function according to equation (1):

$$I(\tau) = I_0 - A \cdot e^{-R_1 \cdot \tau} \quad (1)$$

where  $\tau$  is the recovery delay,  $I(\tau)$  is the peak intensity measured for the recovery delay  $\tau$ ,  $I_0$  is the maximum peak intensity,  $R_1$  is the longitudinal proton relaxation rate and  $A$  is the amplitude of the relaxation.

To estimate the error of the peak intensities  $\varepsilon$ , the recovery delay  $\tau_d$  was measured twice to obtain two intensity values  $I(\tau_d, i, 1)$  and  $I(\tau_d, i, 2)$ , where  $i$  is the index of the peak. For every peak  $i$ , the difference of the duplicates  $\Delta I_i = I(\tau_d, i, 1) - I(\tau_d, i, 2)$  was computed. The overall error of the pseudo-3D experiment  $\varepsilon_{\text{pseudo-3D}}$  was then computed according to equation (2):

$$\varepsilon_{\text{pseudo-3D}} = \sqrt{\frac{1}{2N} \cdot \sum_i^N (\Delta I_i - \overline{\Delta I})^2} \quad (2)$$

where  $N$  is the number of peaks, and  $\overline{\Delta I} = \sum_i^N \Delta I_i$  is the average difference of the duplicates. The subtraction of  $\overline{\Delta I}$  accounts for systematic errors, and in all cases was significantly lower in magnitude than the differences of the duplicates  $\Delta I_i$ .

The error of the peak intensities  $\varepsilon_{\text{pseudo-3D}}$  was then used to estimate the error of the fitted proton relaxation rates  $\Delta R_1$ . To this end, the experimental data was resampled using the following combined Monte Carlo-type bootstrapping approach: For every peak in a given pseudo-3D experiment, a set of delay-intensity data points,  $I(\tau)$ , was measured. This set is used to generate 1000 new random sets  $\widetilde{I}_j(\tau)$ , each having 2.5 times as many data points as the original data set  $I(\tau)$  (Recovery delays measured as duplicates are only counted as one). These sets were created by randomly selecting data points from the original data set  $I(\tau)$  and allowing values to be selected multiple times (Increasing the size of the new sets by a factor of 2.5 allows to generate a more diverse ensemble of sets). For every set  $\widetilde{I}_j(\tau)$ , the intensity values were randomly altered by adding a random number drawn from a normal distribution centered at 0 and with a standard deviation of  $\varepsilon_{\text{pseudo-3D}}$ . Every set  $\widetilde{I}_j(\tau)$  was then fitted according to equation (1) giving rise to 1000 different  $R_1^j$  rates. The error of the proton relaxation rate  $\Delta R_1$  was computed as the standard deviation of all values  $R_1^j$ . This procedure was repeated for every peak and every pseudo-3D experiment.

To obtain the sPRE value, the proton  $R_1$  rates and the corresponding errors  $\Delta R_1$  were collected for all measured concentrations of Gd(DTPA-BMA)  $c_{\text{para}}$  and a weighted linear regression using equation (3) was performed

$$R_1(c_{\text{para}}) = m_{\text{sPRE}} \cdot c_{\text{para}} + R_1^0 \quad (3)$$

where  $R_1(c_{\text{para}})$  is the proton  $R_1$  measured at the concentration  $c_{\text{para}}$  of the paramagnetic compound, the slope  $m_{\text{sPRE}}$  corresponds to the sPRE and  $R_1^0$  is the fitted proton  $R_1$  in the absence of the paramagnetic compound. The errors  $\Delta R_1$  (obtained from the resampling approach described above) were used as weights in a weighted linear regression and the error of the sPRE value  $\Delta m_{\text{sPRE}}$  as well as the error of the relaxation rate  $\Delta R_1^0$  were directly obtained from the weighted linear regression.

**Prediction of sPRE Data.** To predict sPRE data based on published NMR solution structure, a previously published grid-based approach was used<sup>31</sup>. Briefly, the structural model was placed in a regularly-spaced grid representing the uniformly distributed paramagnetic compound and the grid was built with a point-to-point distance of 0.1 Å and a minimum distance of 20 Å between the RNA model and the outer border of the grid. Next, grid points that overlap with the RNA model were removed assuming a molecular radius of 3.5 Å for the paramagnetic compound. To compute the sPRE for a given RNA proton sPRE<sub>predicted</sub><sup>*i*</sup>, the distance-dependent paramagnetic effect<sup>29</sup> was numerically integrated over all remaining grid points according to equation (4):

$$\text{sPRE}_{\text{predicted}}^i = c \cdot \sum_{d_{i,j} < 20 \text{ \AA}} \frac{1}{d_{i,j}^6} \quad (4)$$

where *i* is the index of a proton of the RNA, *j* is the index of the grid point, *d<sub>i,j</sub>* is the distance between the *i*-th proton and the *j*-th grid point and *c* is an arbitrary constant to scale the sPRE values. Here, *c* was chosen such that the sum of all predicted sPRE values is equal to the sum of all experimental sPRE values.

**Structure Determination Benchmark.** To demonstrate the potential of sPRE data for the structure determination of RNAs, structural models of the UUCG tetraloop as well as the GTP class II aptamer in complex with GTP were computed using the XplorNIH 2.40 framework<sup>49,50</sup>. The used protocol is based on the gb1\_rdc example that is included in the XplorNIH package and was adjusted to fold the respective RNA during an annealing procedure starting from an extended structure. The initial temperature was set to 3500 and the simulation at this high temperature was stopped after reaching 1000 ps or 10,000 steps. The temperature was then reduced to a final temperature of 25 in steps of 12.5 degrees and simulations were run for at least 4 ps or 2000 steps at every temperature step. The annealing procedure was repeated to obtain 200 models. The protocol made use of the recently published torsion potential RNA-ff1<sup>51</sup> and the sPRE data was included in the structure determination using the nbTargetPot module<sup>27</sup> of the framework. nbTargetPot is the scaling factor (or weight) of the sPRE potential. The value of the scaling factor mainly depends on the scaling factor of the other energy terms. The weight of the sPRE potential should not be proportional to the size of the RNA. For the mentioned benchmark, the weight of the sPRE potential was determined by testing different weights and the weight that produced the best models was used for all runs. To this end, one weight was determined for every RNA and the obtained weight was used for all computations of the corresponding RNA. The lower value of the UUCG tetraloop can be explained with the fact that the Xplor runs of the UUCG tetraloop included more energy functions than the ones of the GTP aptamer (in the case of the UUCG tetraloop, torsion angles and RDC, both obtained from PDB entry 2KOC, were used). The sPRE data sets of the UUCG tetraloop and GTP aptamer were filtered to remove all data points with an experimental error above 10% and all data points of nitrogen-bound protons. The nbTargetPot energy was initialized using the slope and intercept parameter as returned by the calibrate function of the nbTargetPotTools module in combination with the NMR structure of the corresponding RNA. The weight of the nbTargetPot energy was set to 1000 in the case of the UUCG tetraloop and 3000 for the GTP aptamer. For more details such as weight factors of all potentials, please refer to the python code of the protocols which are shown at the end of this document for both RNAs.

The annealing protocol was then used to benchmark the benefit of the sPRE data for structure prediction. To this end, structural models of the UUCG tetraloop and the GTP-bound aptamer were computed in the absence and presence of the sPRE potential and in combination with experimental restraints derived from hydrogen bond information and experimental NOEs (PDB entries 2KOC and 5LWJ)<sup>37,38,40</sup>. To simulate several NOE assignments, random experimental NOE subsets were created by randomly drawing a certain percentage of restraints from the full NOE set (compare Supplementary Tables 3 and 4). To account for the effect of the random selection process, at least 3 different random sets with the same number of NOE restraints were created and used to benchmark the impact of the sPRE data. It should be noted, that the randomly created NOE subsets not only account for different levels of assignments, but also simulate different qualities of NOE data as random subsets with the same number of NOE restraints perform differently in driving the structure determination to the correct fold. For every given set of restraints, 200 structure models were computed with and without the sPRE data. To quantify the impact of the sPRE data, the 20 best models of each run (10%, scored according to the total energy) were selected and the RMSD to the published NMR structures was computed. Computation of RMSD was performed on all carbon, nitrogen and phosphorus atoms in the case of the UUCG tetraloop. For the GTP-bound aptamer, the RMSD was computed using all carbon, nitrogen and phosphorus atoms of the GTP ligand and all non-terminal nucleotides except the intrinsically flexible nucleotides A13 and U21<sup>40</sup>.

A second benchmark was performed to address the performance of the sPRE potential in the absence of any other experimental restraint as well as in combination with other experimental NMR restraints, such hydrogen bonds, NOEs, RDCs and torsion angles. Experimental restraints for the UUCG tetraloop were obtained from PDB entry 2KOC<sup>37,38</sup>. Two hundred structure models of the UUCG tetraloop were computed with and without the sPRE data for every restraint set (compare Supplementary Tables 5) and the 20 best scored models according to the total energy were selected. The average RMSD to the published NMR structure was computed for the top 20 models and used to quantify the impact of the sPRE data (computation of RMSD was performed on all carbon, nitrogen and phosphorus atoms).

**Data Availability.** The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request. All scripts used for data analysis and back-calculation of sPRE data are openly available (<http://mbbc.medunigraz.at/forschung/forschungseinheiten-und-gruppen/forschungsgruppe-tobias-madl/software/>).

## References

- Morris, K. V. & Mattick, J. S. The rise of regulatory RNA. *Nat Rev Genet* **15**, 423–437, doi:10.1038/nrg3722 (2014).
- Furtig, B., Richter, C., Wohnert, J. & Schwalbe, H. NMR spectroscopy of RNA. *ChemBiochem* **4**, 936–962, doi:10.1002/cbic.200300700 (2003).
- Latham, M. P., Brown, D. J., McCallum, S. A. & Pardi, A. NMR methods for studying the structure and dynamics of RNA. *ChemBiochem* **6**, 1492–1505, doi:10.1002/cbic.200500123 (2005).
- Bothe, J. R. *et al.* Characterizing RNA dynamics at atomic resolution using solution-state NMR spectroscopy. *Nat Methods* **8**, 919–931, doi:10.1038/nmeth.1735 (2011).
- Dominguez, C., Schubert, M., Duss, O., Ravindranathan, S. & Allain, F. H. Structure determination and dynamics of protein-RNA complexes by NMR spectroscopy. *Prog Nucl Magn Reson Spectrosc* **58**, 1–61, doi:10.1016/j.pnmrs.2010.10.001 (2011).
- Wijmenga, S. S. & van Buuren, B. N. M. The use of NMR methods for conformational studies of nucleic acids. *Progress in Nuclear Magnetic Resonance Spectroscopy* **32**, 287–387, doi:10.1016/s0079-6565(97)00023-x (1998).
- Wu, H., Finger, L. D. & Feigon, J. Structure determination of protein/RNA complexes by NMR. *Methods Enzymol* **394**, 525–545, doi:10.1016/S0076-6879(05)94022-6 (2005).
- Varani, G., Aboul-ela, F. & Allain, F. H. NMR investigation of RNA structure. *Prog Nucl Magn Reson Spectrosc* **29**, 51–127, doi:10.1016/0079-6565(96)01028-X (1996).
- Duss, O., Lukavsky, P. J. & Allain, F. H. Isotope labeling and segmental labeling of larger RNAs for NMR structural studies. *Adv Exp Med Biol* **992**, 121–144, doi:10.1007/978-94-007-4954-2\_7 (2012).
- Liu, Y. *et al.* Synthesis and applications of RNAs with position-selective labelling and mosaic composition. *Nature* **522**, 368–372, doi:10.1038/nature14352 (2015).
- Lu, K., Miyazaki, Y. & Summers, M. F. Isotope labeling strategies for NMR studies of RNA. *J Biomol NMR* **46**, 113–125, doi:10.1007/s10858-009-9375-2 (2010).
- Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res* **15**, 8783–8798 (1987).
- Milligan, J. F. & Uhlenbeck, O. C. Synthesis of small RNAs using T7 RNA polymerase. *Methods Enzymol* **180**, 51–62 (1989).
- Hennig, M., Williamson, J. R., Brodsky, A. S. & Battiste, J. L. Recent advances in RNA structure determination by NMR. *Curr Protoc Nucleic Acid Chem* **Chapter 7**, Unit 7.7, doi:10.1002/0471142700.nc0707s02 (2001).
- Wunderlich, C. H. *et al.* Synthesis of (6-(13C)pyrimidine nucleotides as spin-labels for RNA dynamics. *J Am Chem Soc* **134**, 7558–7569, doi:10.1021/ja302148g (2012).
- Johnson, J. E. Jr., Julien, K. R. & Hoogstraten, C. G. Alternate-site isotopic labeling of ribonucleotides for NMR studies of ribose conformational dynamics in RNA. *J Biomol NMR* **35**, 261–274, doi:10.1007/s10858-006-9041-x (2006).
- Alvarado, L. J. *et al.* Regio-selective chemical-enzymatic synthesis of pyrimidine nucleotides facilitates RNA structure and dynamics studies. *ChemBiochem* **15**, 1573–1577, doi:10.1002/cbic.201402130 (2014).
- Longhini, A. P. *et al.* Chemo-enzymatic synthesis of site-specific isotopically labeled nucleotides for use in NMR resonance assignment, dynamics and structural characterizations. *Nucleic Acids Res* **44**, e52, doi:10.1093/nar/gkv1333 (2016).
- Wenter, P., Raymond, L., Auweter, S. D., Allain, F. H. & Pitsch, S. Short, synthetic and selectively 13C-labeled RNA sequences for the NMR structure determination of protein-RNA complexes. *Nucleic Acids Res* **34**, e79, doi:10.1093/nar/gkl427 (2006).
- Tzakos, A. G., Easton, L. E. & Lukavsky, P. J. Complementary segmental labeling of large RNAs: economic preparation and simplified NMR spectra for measurement of more RDCs. *J Am Chem Soc* **128**, 13344–13345, doi:10.1021/ja064807o (2006).
- Xu, J., Lapham, J. & Crothers, D. M. Determining RNA solution structure by segmental isotopic labeling and NMR: application to *Caenorhabditis elegans* spliced leader RNA 1. *Proc Natl Acad Sci USA* **93**, 44–48 (1996).
- Nelissen, F. H. *et al.* Multiple segmental and selective isotope labeling of large RNA for NMR structural studies. *Nucleic Acids Res* **36**, e89, doi:10.1093/nar/gkn397 (2008).
- Clore, G. M. & Iwahara, J. Theory, practice, and applications of paramagnetic relaxation enhancement for the characterization of transient low-population states of biological macromolecules and their complexes. *Chem Rev* **109**, 4108–4139, doi:10.1021/cr900033p (2009).
- Lipsitz, R. S. & Tjandra, N. Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct* **33**, 387–413, doi:10.1146/annurev.biophys.33.110502.140306 (2004).
- Dallmann, A. & Sattler, M. Detection of hydrogen bonds in dynamic regions of RNA by NMR spectroscopy. *Curr Protoc Nucleic Acid Chem* **59**, 7.22.21–19, doi:10.1002/0471142700.nc0722s59 (2014).
- Grzesiek, S., Cordier, F., Jaravine, V. & Barfield, M. Insights into biomolecular hydrogen bonds from hydrogen bond scalar couplings. *Progress in Nuclear Magnetic Resonance Spectroscopy* **45**, 275–300, doi:10.1016/j.pnmrs.2004.08.001 (2004).
- Wang, Y., Schwieters, C. D. & Tjandra, N. Parameterization of solvent-protein interaction and its use on NMR protein structure determination. *J Magn Reson* **221**, 76–84, doi:10.1016/j.jmr.2012.05.020 (2012).
- Modig, K., Liepinsh, E., Otting, G. & Halle, B. Dynamics of protein and peptide hydration. *Journal of the American Chemical Society* **126**, 102–114, doi:10.1021/ja038325d (2004).
- Hocking, H. G., Zangger, K. & Madl, T. Studying the structure and dynamics of biomolecules by using soluble paramagnetic probes. *Chemphyschem* **14**, 3082–3094, doi:10.1002/cphc.201300219 (2013).
- Madl, T., Bermel, W. & Zangger, K. Use of relaxation enhancements in a paramagnetic environment for the structure determination of proteins using NMR spectroscopy. *Angew Chem Int Ed Engl* **48**, 8259–8262, doi:10.1002/anie.200902561 (2009).
- Pintacuda, G. & Otting, G. Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. *J Am Chem Soc* **124**, 372–373 (2002).
- Madl, T., Guttler, T., Gorlich, D. & Sattler, M. Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements. *Angew Chem Int Ed Engl* **50**, 3993–3997, doi:10.1002/anie.201007168 (2011).
- Hartlmüller, C., Göbl, C. & Madl, T. Prediction of Protein Structure Using Surface Accessibility Data. *Angew Chem Int Ed Engl* **55**, 11970–11974, doi:10.1002/anie.201604788 (2016).
- Venditti, V., Nicolai, N. & Butcher, S. E. Measuring the dynamic surface accessibility of RNA with the small paramagnetic molecule TEMPOL. *Nucleic Acids Res* **36**, e20, doi:10.1093/nar/gkm1062 (2008).
- Bernini, A. *et al.* NMR studies on the surface accessibility of the archaeal protein Sso7d by using TEMPOL and Gd(III)(DTPA-BMA) as paramagnetic probes. *Biophys Chem* **137**, 71–75, doi:10.1016/j.bpc.2008.07.003 (2008).
- Caravan, P., Ellison, J. J., McMurry, T. J. & Laufer, R. B. Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications. *Chem Rev* **99**, 2293–2352 (1999).
- Furtig, B., Richter, C., Bermel, W. & Schwalbe, H. New NMR experiments for RNA nucleobase resonance assignment and chemical shift analysis of an RNA UUCG tetraloop. *J Biomol NMR* **28**, 69–79, doi:10.1023/B:JNMR.0000012863.63522.1f (2004).
- Nozinovic, S., Furtig, B., Jonker, H. R., Richter, C. & Schwalbe, H. High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res* **38**, 683–694, doi:10.1093/nar/gkp956 (2010).
- Wolter, A. C. *et al.* NMR resonance assignments for the class II GTP binding RNA aptamer in complex with GTP. *Biomol NMR Assign* **10**, 101–105, doi:10.1007/s12104-015-9646-7 (2016).
- Wolter, A. C. *et al.* A Stably Protonated Adenine Nucleotide with a Highly Shifted pKa Value Stabilizes the Tertiary Structure of a GTP-Binding RNA Aptamer. *Angewandte Chemie*. doi:10.1002/anie.201609184 (2016).

41. Matsumura, K. & Komiyama, M. Enormously fast RNA hydrolysis by lanthanide(III) ions under physiological conditions: eminent candidates for novel tools of biotechnology. *J Biochem* **122**, 387–394 (1997).
42. Deng, N. J. & Cieplak, P. Free energy profile of RNA hairpins: a molecular dynamics simulation study. *Biophys J* **98**, 627–636, doi:10.1016/j.bpj.2009.10.040 (2010).
43. Giambasu, G. M., York, D. M. & Case, D. A. Structural fidelity and NMR relaxation analysis in a prototype RNA hairpin. *RNA* **21**, 963–974, doi:10.1261/rna.047357.114 (2015).
44. Villa, A., Widjajakusuma, E. & Stock, G. Molecular dynamics simulation of the structure, dynamics, and thermostability of the RNA hairpins uCACGg and cUUCGg. *J Phys Chem B* **112**, 134–142, doi:10.1021/jp0764337 (2008).
45. Borkar, A. N., Vallurupalli, P., Camilloni, C., Kay, L. E. & Vendruscolo, M. Simultaneous NMR characterisation of multiple minima in the free energy landscape of an RNA UUCG tetraloop. *Phys Chem Chem Phys* **19**, 2797–2804, doi:10.1039/c6cp08313g (2017).
46. Vallurupalli, P. & Kay, L. E. A suite of 2H NMR spin relaxation experiments for the measurement of RNA dynamics. *J Am Chem Soc* **127**, 6893–6901, doi:10.1021/ja0427799 (2005).
47. Williams, D. J. & Hall, K. B. Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J Mol Biol* **297**, 1045–1061, doi:10.1006/jmbi.2000.3623 (2000).
48. Akke, M., Fiala, R., Jiang, F., Patel, D. & Palmer, A. G. 3rd. Base dynamics in a UUCG tetraloop RNA hairpin characterized by 15N spin relaxation: correlations with structure and stability. *RNA* **3**, 702–709 (1997).
49. Schwieters, C. D., Kuszewski, J. J. & Clore, G. M. Using Xplor-NIH for NMR molecular structure determination. *Progress in nuclear magnetic resonance spectroscopy* **48**, 47–62 (2006).
50. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Clore, G. M. The Xplor-NIH NMR molecular structure determination package. *Journal of magnetic resonance* **160**, 65–73 (2003).
51. Bermejo, G. A., Clore, G. M. & Schwieters, C. D. Improving NMR Structures of RNA. *Structure* **24**, 806–815, doi:10.1016/j.str.2016.03.007 (2016).
52. Sripakdeevong, P. *et al.* Structure determination of noncanonical RNA motifs guided by (1)H NMR chemical shifts. *Nat Methods* **11**, 413–416, doi:10.1038/nmeth.2876 (2014).
53. Dallmann, A. *et al.* Site-Specific Isotope-Labeling of Inosine Phosphoramidites and NMR Analysis of an Inosine-Containing RNA Duplex. *Chemistry*. doi:10.1002/chem.201602784 (2016).
54. Farjon, J. *et al.* Longitudinal-relaxation-enhanced NMR experiments for the study of nucleic acids in solution. *Journal of the American Chemical Society* **131**, 8571–8577, doi:10.1021/ja901633y (2009).
55. Delaglio, F. *et al.* NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **6**, 277–293 (1995).
56. Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696, doi:10.1002/prot.20449 (2005).
57. Helmus, J. J. & Jaroniec, C. P. NmrGlue: an open source Python package for the analysis of multidimensional NMR data. *J Biomol NMR* **55**, 355–367, doi:10.1007/s10858-013-9718-x (2013).

## Acknowledgements

This work was supported by the Integrative Metabolism Research Center Graz (to T.M.), the Austrian infrastructure programme 2016/2017 (to T.M.), BioTechMed/Graz (to T.M.), the Omics Center Graz (to T.M.), the Bavarian Ministry of Sciences, Research and the Arts (Bavarian Molecular Biosystems Research Network, to T.M.), the President's International Fellowship Initiative of CAS (No. 2015VBB045, to T.M.), the National Natural Science Foundation of China (No. 31450110423, to T.M.), the Austrian Science Fund (FWF: P28854 and W1226-B18 to T.M.) as well as the Deutsche Forschungsgemeinschaft (DFG) with the grants Wo901/1-1 (to J.W.), SFB1035 (to M.S.), GRK1721 (to M.S.) and MA5703/1-1 (to T.M.).

## Author Contributions

C.H. and T.M. designed the studies, established and tested the NMR experiments, and performed the prediction of sPRE data. J.G., A.W., J.W. and M.S. prepared the isotope labelled RNA samples. C.H. and T.M., C.H. and J.G. performed the NMR titrations. C.H. processed and fitted the NMR data, performed the structure prediction computations and analyzed the obtained models. C.H. and T.M. wrote the manuscript and all authors have given critical feedback and approved the final version of the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at doi:10.1038/s41598-017-05821-z

**Competing Interests:** The authors declare that they have no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017