



RESEARCH ARTICLE

# Refining RNA solution structures with the integrative use of label-free paramagnetic relaxation enhancement NMR

Zhou Gong<sup>1</sup>✉, Shuai Yang<sup>1,2</sup>, Qing-Fen Yang<sup>1</sup>, Yue-Ling Zhu<sup>1,2</sup>,  
Jing Jiang<sup>1</sup>, Chun Tang<sup>1,3</sup>✉

<sup>1</sup> CAS Key Laboratory of Magnetic Resonance in Biological Systems, State Key Laboratory of Magnetic Resonance and Atomic Molecular Physics, National Center for Magnetic Resonance at Wuhan, Wuhan Institute of Physics and Mathematics of the Chinese Academy of Sciences, Wuhan 430071, Hubei, China

<sup>2</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>3</sup> Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, Hubei, China

Received: 1 August 2019 / Accepted: 8 October 2019 / Published online: 15 November 2019

**Abstract** NMR structure calculation is inherently integrative, and can incorporate new experimental data as restraints. As RNAs have lower proton densities and are more conformational heterogenous than proteins, the refinement of RNA structures can benefit from additional types of restraints. Paramagnetic relaxation enhancement (PRE) provides distance information between a paramagnetic probe and protein or RNA nuclei. However, covalent conjugation of a paramagnetic probe is difficult for RNAs, thus limiting the use of PRE NMR for RNA structure characterization. Here, we show that the solvent PRE can be accurately measured for RNA labile imino protons, simply with the addition of an inert paramagnetic cosolute. Demonstrated on three RNAs that have increasingly complex topologies, we show that the incorporation of the solvent PRE restraints can significantly improve the precision and accuracy of RNA structures. Importantly, the solvent PRE data can be collected for RNAs without isotope enrichment. Thus, the solvent PRE method can work integratively with other biophysical techniques for better characterization of RNA structures.

**Keywords** RNA structure, NMR, Paramagnetic relaxation enhancement, Paramagnetic cosolute, Label-free

## INTRODUCTION

RNA is an important class of biological macromolecules. Different from protein-coding messenger RNAs, non-coding RNAs play important structural, signaling, and catalytic roles, and are involved in all aspects of cellular

life (Bhan *et al.* 2017; Eddy 2001; Wilusz *et al.* 2009). Like proteins, the three-dimensional (3D) structures of non-coding RNAs are important for their functions (Butcher and Pyle 2011; Larsen *et al.* 2019). But unlike proteins, the RNA folding is basically hierarchical (Herschlag *et al.* 2018). As a result, though there are programs to predict RNA secondary structures regarding the patterns of base-pairing (Bellaousov *et al.* 2013; Parisien and Major 2008; Zhang *et al.* 2019; Zuker 2003), it remains difficult to predict RNA secondary structures with complex topologies (Schlick and Pyle 2017; Zhao *et al.* 2018).

Experimental characterization of RNA structures has also been lagging behind that of protein structures.

Zhou Gong and Shuai Yang contributed equally to this work.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s41048-019-00099-2>) contains supplementary material, which is available to authorized users.

✉ Correspondence: gongzhou@wipm.ac.cn (Z. Gong), tanglab@wipm.ac.cn (C. Tang)

Indeed, there are much fewer RNA structures in the PDB, especially the structures only containing RNA molecules, in comparison to protein-only structures. RNAs are more dynamic than proteins, and can be recalcitrant to crystallization (Larsen *et al.* 2019; Murray *et al.* 2003; Ponce-Salvatierra *et al.* 2019). As a result, solution NMR spectroscopy is a major tool for RNA structure determination, and the proportion of RNA NMR structures is much higher than the protein NMR structures in the PDB (Barnwal *et al.* 2017).

RNAs have low proton density than proteins. Since nuclear Overhauser effect (NOE) only provides the distance relationship between adjacent protons (<6 Å), cumulative errors can build upon for large RNA structures if using only the NOE restraints. Thus, additional restraints, like residual dipolar coupling (RDC) NMR measurements, have been incorporated into RNA solution structure refinement (Hansen *et al.* 1998, 2000). The RDCs restrain the C–H and N–H bond vectors of nucleotide bases with respect to an overall alignment tensor, and the incorporation of RDC restraints can make the RNA structures more converged. More recently, restraints from other types of biophysical measurement have also been included, such as small angle X-ray scattering (SAXS) (Bhandari *et al.* 2016), cryogenic electron microscopy (Gong *et al.* 2015; Kappel *et al.* 2018), and single-molecule FRET (Suddala and Walter 2014; Warhaut *et al.* 2017). These additional restraints provide long-range information about overall shape and topology of RNA molecules, and complement the standard NMR restraints.

Paramagnetic relaxation enhancement (PRE) also provides long-range distance information between a covalently tagged paramagnetic probe and the nuclei of a biological macromolecule (Clare and Iwahara 2009). The PRE has been successfully used for proteins for the characterization of protein structure and dynamics (Liu *et al.* 2014, 2015, 2016). Though the conjugation methods originated from electron paramagnetic resonance (EPR) and FRET (Edwards and Sigurdsson 2007; Zhang *et al.* 2009), tagging an RNA with a paramagnetic probe remains technically challenging. This is because for NMR studies, the RNA molecule should also be enriched with  $^{13}\text{C}$  and  $^{15}\text{N}$  isotopes in addition to the probe attachment. Thus, an *in vitro* transcribed isotopically labeled RNA molecule needs to be annealed with another chemically synthesized and paramagnetically tagged RNA molecule (Helmling *et al.* 2014), or directly synthesized using isotopically labeled phosphoramidites (Wunderlich *et al.* 2013). The two annealed RNA strands can be further ligated, with the ligation site usually located in an internal loop (Duss *et al.* 2015). As such, preparation such an RNA sample

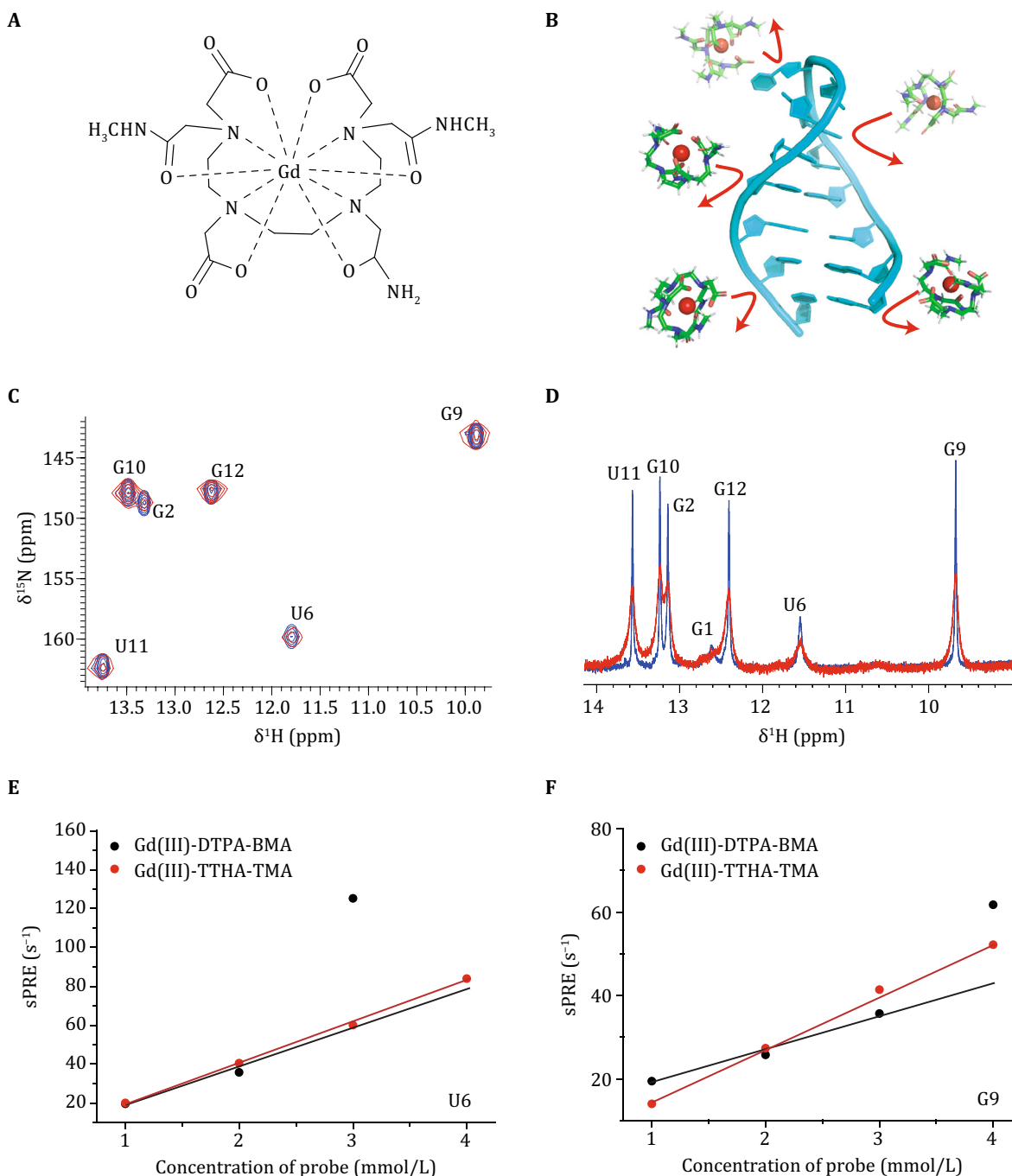
for PRE NMR measurement is cumbersome and costly (Duss *et al.* 2010).

We have previously shown that the PRE measured for a protein in the presence of a freely diffusing and randomly colliding paramagnetic cosolute, termed solvent PRE or sPRE, affords better characterized protein structures (Gong *et al.* 2017, 2018; Gu *et al.* 2014). Thus, without covalent conjugation of the paramagnetic probe, the sPRE gauges the depth of nuclei, namely protons in a protein, which is related to the protein structure. Here we extend the sPRE method to the characterization of RNA structures. When an RNA molecule folds into a specific tertiary structure, the relative solvent exposure of protons in nucleotide bases differs from that in an unfolded single-stranded RNA, and also from that in RNA with only secondary structure formed. We show that the incorporation of sPRE restraints leads to more converged RNA structures, especially for those with complex topologies. Moreover, since the sPRE is measured label-free and an inert paramagnetic cosolute is used, the sPRE term can be incorporated for structure refinement only based on RNA imino proton signals, thus without the need for isotope enrichment. In addition, the imino protons are well isolated and are relatively easy to assign (Furtig *et al.* 2003). Together, the sPRE provides additional structural information about RNA molecules, and complements the other more commonly used NMR restraints.

## RESULTS

### The sPRE can be accurately measured for labile protons in RNA

Previously we have developed a new paramagnetic cosolute. Unlike diethylenetriamine pentaacetate bis-methylamide gadolinium chelate (Gd(III)-DTPA-BMA), the sPRE probe that was commonly used, the  $\text{Gd}^{3+}$  paramagnetic lanthanide ion in this new probe is completely buried by the triethylenetetraamine hexaacetate trimethylamide (TTHA-TMA) molecule, leaving no coordination site for water (Fig. 1A). As such, the paramagnetic cosolute simply gauges the relative solvent exposure of the nuclei in a biological macromolecule (Fig. 1B), when it randomly collides with the biomacromolecule and causes relaxation enhancement through an outer-sphere relaxation mechanism (Gong *et al.* 2017, 2018; Gu *et al.* 2014). In comparison, the Gd(III)-DTPA-BMA probe can cause additional relaxation enhancement to labile protons in biological macromolecules, as these protons exchange with the



**Fig. 1** The sPRE measurement for 14-mer RNA tetraloop hairpin in the presence of Gd(III)-TTHA-TMA paramagnetic cosolute. **A** The structure of triethylenetetraamine hexaacetate trimethylamide gadolinium chelate (Gd(III)-TTHA-TMA). **B** The paramagnetic cosolute randomly collides with the RNA and probes the relative solvent exposure of RNA protons. **C** 2D NMR spectrum of imino protons without (*blue*) or with (*red*) paramagnetic cosolute added. **D**  $^1\text{H}$  NMR spectrum of imino protons without (*blue*) or with (*red*) paramagnetic cosolute added. The addition of paramagnetic cosolute causes line broadening and decrease of peak intensity, which allows the evaluation of sPRE. **E**, **F** The assessment of the sPRE values for the imino protons of nucleotides U6 and G9 in the RNA tetraloop, by adding 1–4 mmol/L concentration of Gd(III)-TTHA-TMA (*red*) or Gd(III)-DTPA-BMA (*black*) paramagnetic cosolute

relaxation-enhanced water molecules. This is especially a problem for the sPRE measurement for RNA molecules. Indeed, the imino protons of nucleotide bases of

guanosine and uridine rapidly exchange with water; unless they are stabilized by base-pairing.

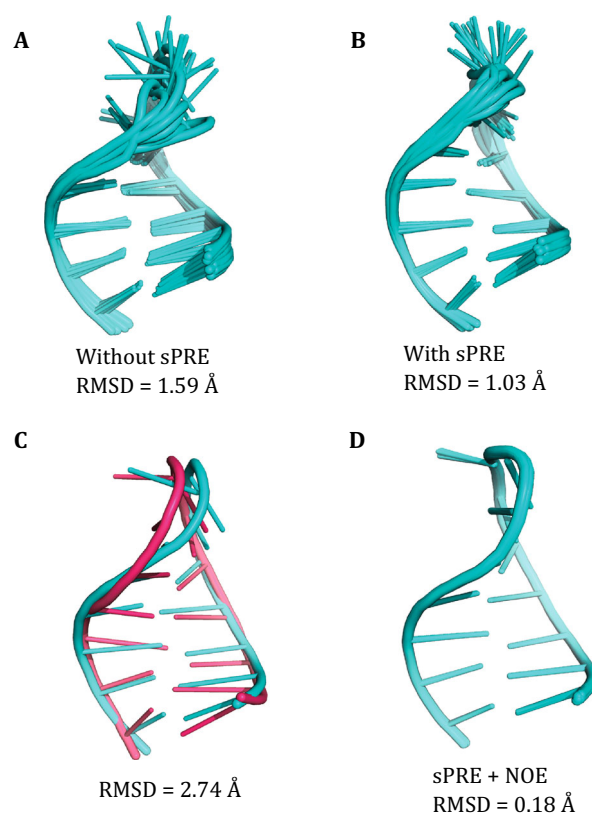
Here we collected two-dimensional (2D)  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum for a U- $^{13}\text{C}$ , $^{15}\text{N}$ -labeled 14-mer UUCG-tetraloop RNA hairpin (Nozinovic *et al.* 2010). A total of six imino protons can be detected on the 2D NMR spectrum that can be assigned to base-paired guanosine and uridines (Fig. 1C). Importantly, although the secondary structure prediction indicates that C5-G10 is the last base pair in the helix, the appearance of U6 and G9 imino protons indicate that these two nucleotides are also base-paired (Duss *et al.* 2010). The largest sPRE  $\Gamma_2$  value (the enhancement of the transverse relaxation rate upon the addition of the probe) was observed for nucleotide U6, indicative of its high solvent exposure (supplementary Table S1). Importantly, the PRE measurement can also be measured for the small RNA without isotope enrichment. Since the imino protons can be easily resolved based on their  $^1\text{H}$  chemical shifts (Fig. 1D), the sPRE values can be determined using a simple spin-echo experiment.

Using the spin-echo experiment, we thus assessed the sPRE  $\Gamma_2$  values for the imino protons of the RNA tetraloop with the two different probes. The sPRE values have a perfect linear relationship with the concentration of the Gd(III)-TTHA-TMA added to the RNA system. In comparison, the sPRE value as function of the concentration of the Gd(III)-DTPA-BMA added deviates upwards from a linear relationship at an increasing concentration of the probe. As such, the Gd(III)-TTHA-TMA probe allows more accurate measurement of the sPRE values of the RNA labile protons (Fig. 1E, F).

### The sPRE restraints afford better refined RNA tetraloop structure

The 14-mer RNA tetraloop hairpin has often been used as a model system (Borkar *et al.* 2016; Duchardt and Schwalbe 2005; Hartmuller *et al.* 2017). For its short length and simple secondary structure, even with only the hydrogen-bonding and knowledge-based restraints for the A-form RNA hairpin, the 14-mer RNA structure can already be calculated to a root-mean-square (RMS) deviation of 1.59 Å for all heavy atoms (Fig. 2A).

Thus, we implemented the sPRE restraints with the PSolPot term in the Xplor-NIH software package, in a similar fashion as we did for proteins (Gong *et al.* 2018). The implementation of the sPRE restraints, in conjunction with the restraints enforcing A-form RNA hairpin geometry, afforded better structural convergence, with an RMS of 1.03 Å for all heavy atoms (Fig. 2B). Accordingly, the back-calculated sPRE values agree well with the experimental sPRE values, with a correlation coefficient of 0.99. The structure also agrees quite well with the known structure (PDB code 6BY4), with an



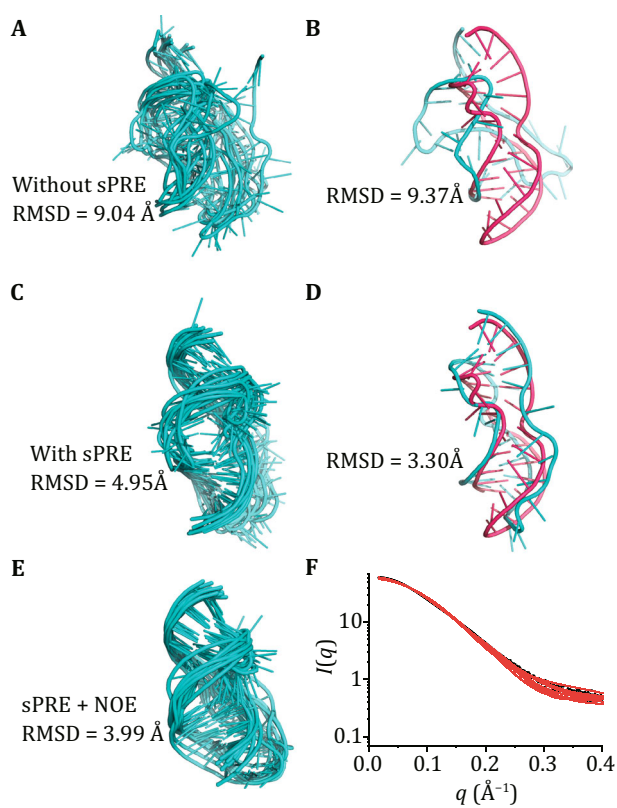
**Fig. 2** Refinement of 14-mer RNA tetraloop hairpin structure. **A** 10 lowest-energy structure refined with only 2D structure and knowledge-based restraints, the RMSD is 1.59 Å for all heavy atoms. **B** 10 lowest-energy structure refined with sPRE in addition with 2D structure and knowledge-based restraints, the RMSD is 1.03 Å for all heavy atoms. **C** Structure comparison between refined with sPRE (cyan) and known structure (red, PDB code 6BY4), the RMSD is 2.70 Å for all heavy atoms. **D** 10 lowest structure refined with sPRE, NOE, 2D structure and knowledge-based restraints, the RMSD is 0.18 Å for all heavy atoms

RMS difference of 2.74 Å (Fig. 2C). Nevertheless, the sPRE and NOE restraints are consistent with each other. Indeed, the joint refinement against both NOE and sPRE restraints afforded even more converged structure, with an RMS deviation of 0.18 Å for all heavy atoms (Fig. 1D and supplementary Table S2). Together, we show that, for this 14-mer RNA hairpin, the incorporation of sPRE restraints for imino protons, just six of them, can afford better defined RNA structure.

### The sPRE restraints afford better refined RNA pseudoknot structure

For the second example to illustrate the strength of sPRE, we used an RNA pseudoknot from mouse mammary tumor virus (MMTV), which is responsible for

frameshifting and of which the structure has been previously characterized. RNA pseudoknot is known for its complex secondary structure and base-pairing. Indeed, the secondary structure prediction using the Mfold server (Zuker 2003) only predicts the first RNA helix, and fails to predict the second RNA helix. The 2D  $^1\text{H}$ - $^{15}\text{N}$  correlation spectrum for the isotope-labeled MMTV pseudoknot structure allows the identification of additional base-pairing involving the 3'-tail. However, using only the base-pairing secondary structure information as restraints, the structure convergence is poor, with an RMS deviation of 9.04 Å for all backbone heavy



**Fig. 3** Structure refinement of RNA pseudoknot from MMTV. **A** The ten lowest-energy structure refined with only 2D structure and knowledge-based restraints, affording an RMS deviation of 9.04 Å for all heavy atoms. **B** The comparison with the known structure (red, PDB code 1KPD); the RMS difference is 9.37 Å for all heavy atoms. **C** The ten lowest-energy structure refined with sPRE restraints in conjunction with 2D structure and knowledge-based restraints, affording an RMS deviation of 4.95 Å for all heavy atoms. **D** The comparison with the known PDB structure; the RMS difference is 3.30 Å for all heavy atoms. **E** The ten lowest-energy structure refined with sPRE, NOE, 2D structure and knowledge-based restraints, affording an RMS deviation of 3.99 Å for all heavy atoms. **F** The theoretical SAXS curves for the ten structure match well with the experimental SAXS curve for the RNA pseudoknot

atoms (Fig. 3A), and an RMS difference of 9.37 Å from the known structure determined with extensive NOE distance restraints (Kang and Tinoco 1997) (Fig. 3B).

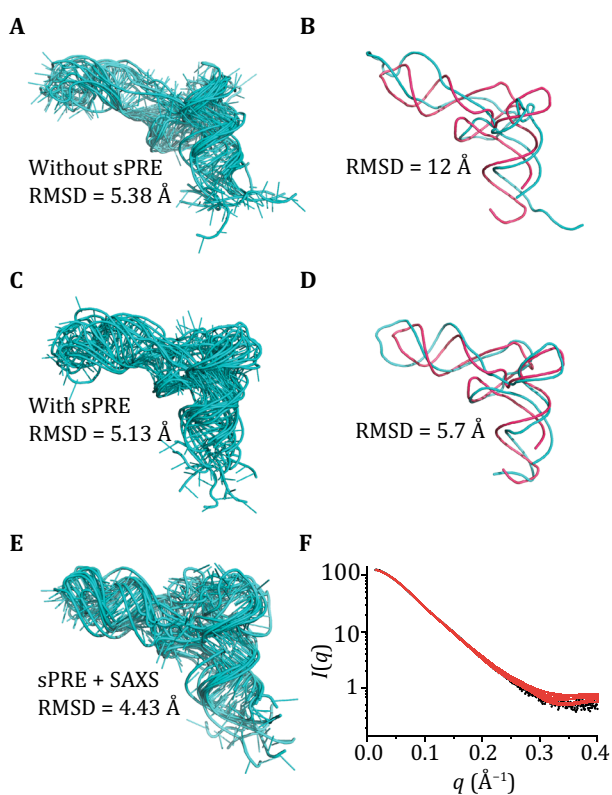
We thus collected the sPRE data for the imino protons of the base-paired guanines and uridines in the MMTV pseudoknot in the presence of Gd(III)-TTHA-TMA paramagnetic cosolute (supplementary Fig. S1A and Table S3). The incorporation of just eight sPRE restraints led to better converged structure, with an RMS deviation of 4.95 Å (Fig. 3C), and an RMS difference from the PDB structure by 3.30 Å (Fig. 3D). Accordingly, the agreement between the observed and calculated sPRE values can be as high as 0.98 in correlation coefficient. The structure convergence of the RNA pseudoknot can be further improved with the incorporation of both NOE and sPRE restraints, to an RMS deviation of 3.99 Å for all heavy atoms (Fig. 3E). The structure can also be cross-validated by the SAXS measurement, as the SAXS profiles predicated for the RNA structure calculated without using the SAXS restraint largely match the experimental SAXS profile (Fig. 3F). Taken together, the incorporation of sparse sPRE restraints greatly improves the precision and accuracy of the RNA pseudoknot structure.

### The sPRE restraints afford better refined tRNA<sup>Val</sup> four-way junction structure

The tRNA<sup>Val</sup> has a four-way junction secondary structure and folds into an L-shaped tertiary structure. Previously, Bax and colleagues incorporated residual dipolar coupling (RDC) restraints and characterized the solution structure the tRNA<sup>Val</sup> to high resolution. In addition to the incorporation of RDC restraints, an important stratagem in their refinement protocol is the use of homology modeling based on the known crystal structure of tRNA<sup>Phe</sup> (Grishaev *et al.* 2008). As a result, the tRNA<sup>Val</sup> structure is already quite converged even before the incorporation of RDC and SAXS restraints.

Here we collected the sPRE values for tRNA<sup>Val</sup> in the presence of Gd(III)-TTHA-TMA paramagnetic cosolute. Out of the 76 nucleotides, 23 base-paired imino protons can be observed, assigned, and measured (supplementary Fig. S1B). Note that with the same concentration of the sPRE probe added, the absolute sPRE values are smaller for tRNA<sup>Val</sup> than for the other two RNAs (supplementary Table S4). This is because the tRNA<sup>Val</sup> is larger and the probabilistic distribution of a paramagnetic probe in the vicinity of each nucleotide is smaller.

The incorporation of these 23 sPRE restraints in conjunction with the restraints enforcing four-way junction secondary structure can afford a rather converged structure of tRNA<sup>Val</sup>, with the RMS deviation of



**Fig. 4** The refinement of tRNA<sup>Val</sup> structure with the incorporation of sPRE restraints. **A, C** The ten lowest-energy structures refined with only the 2D structure and knowledge-based restraints have an RMS deviation of 5.38 Å for all heavy atoms; with the sPRE restraints added, the RMS deviation decreases to 5.13 Å. **B, D** The comparison to the known structure (red, PDB code 2K4C). The RMS difference can be improved from 12 Å (**B**) with only the 2D structure restraints to 5.7 Å with the sPRE restraints added. **E** The ten lowest-energy structure refined with both sPRE and SAXS restraints, affording an RMS deviation of 4.43 Å for all heavy atoms. **F** The calculated SAXS curves are consistent with the experimental curve

5.13 Å for all heavy atoms (Fig. 4A), as compared to 5.38 Å for the RNA structure calculated without the sPRE restraints. Importantly, the RMS difference from the known structure can be improved from 12 Å (Fig. 4B) for the structure refined with only secondary structure restraints to 5.7 Å (Fig. 4C, D). The large improvement of RMSD can be attributed to the correction of the interhelical angle with the use of sPRE restraints. The structure convergence can be further improved with the incorporation of SAXS restraints, to an RMS deviation of 4.43 Å for all heavy atoms (Fig. 4E, F). The synergy between sPRE and SAXS data further indicates that two types of experimental inputs cross-validate each other. Taken together, the sPRE restraints for the imino protons in tRNA<sup>Val</sup> can

greatly improve the precision and accuracy for its structure refinement.

## DISCUSSION

In this report, we show that the integrative use of NMR sPRE restraints greatly improves the precision and accuracy of RNA structures. Unlike the more established PRE measurement, the sPRE is measured without the covalent conjugation of a paramagnetic probe. The incorporation of additional sPRE restraints can afford more converged RNA structure, as the final structure has to satisfy the different types of experimental measurements at the same time.

The accurate measurement of the sPRE data is owing to a paramagnetic probe we have developed. The sPRE probe detects relative solvent exposure of the nuclei in the RNA molecule, and does not cause additional relaxation enhancement from water exchange. As a result, we could accurately measure the sPRE values for the labile imino protons in an RNA molecule. The measurement is also possible for RNA samples without isotope enrichment. Thus, the sPRE restraints can be readily incorporated in the RNA structure refinement.

The sPRE restraints have been previously implemented using an NBtarget term. This term also assesses solvent accessibility of nuclei (Wang *et al.* 2012). However, its use first requires the optimization of two parameters, the slope and intercept, for each set of experimental data, and thus may introduce additional uncertainties. In comparison, the PSolPot term recapitulates the sPRE value without parameter fitting a priori, and therefore we advocate the use of PSolPot term in the Xplor-NIH software package. The PSolPot term refines against the sPRE restraints based on surface integral, and can be rapidly derivatized for the search of global energy minimum, *i.e.*, leading to the identification of the RNA structure that best accounts for the sPRE data.

In the present study, we show that the incorporation of the sPRE restraints, even just a few of them, can make the RNA molecule fold into a correct tertiary structure. Since the secondary and tertiary structures of an RNA molecule are often intertwined, the incorporation of sPRE restraints is particularly important for RNAs with complex topologies. Note that the previous use of sPRE data for RNA structure refinement (Hartmuller *et al.* 2017) was demonstrated with two relatively simple RNA structures, and their secondary structures can be readily predicted. In comparison, the secondary structure prediction does not work for the pseudoknot and four-way junction tRNA, the two RNA systems used in

the present study, as the folding of the RNA secondary structure is intertwined with the tertiary structure.

In conclusion, the implementation of the sparse sPRE restraints and the invocation of the PSolPot term allow better characterization of RNA solution structures. Since the sPRE method is label-free and can be performed on RNA samples without isotope enrichment, the sPRE restraints can also be used integratively with SAXS, cryoEM and other types of biophysical measurements. As RNA molecules are inherently dynamic, the sPRE may also be used to uncover the conformational fluctuation of RNA structure, thus to further improve the agreement between the observed and calculated sPRE values.

## MATERIALS AND METHODS

### Preparation of RNA samples

To demonstrate how the sPRE can be used in RNA structure refinement, we used three model systems. They are 14-mer tetraloop RNA hairpin, an RNA pseudoknot from MMTV responsible for frameshifting, and tRNA<sup>Val</sup>, a four-way RNA junction that folds into a cloverleaf tertiary structure. The 14-mer tetraloop RNA hairpin and RNA pseudoknot were prepared using *in vitro* transcription with T7 RNA polymerase, with uniformly <sup>13</sup>C/<sup>15</sup>N labeled nucleotide triphosphates as the starting material, per standard protocol (Dethoff *et al.* 2008). To fold these two RNAs, the *in vitro* transcribed RNA samples were heated to 95 °C for 5 min and were immediately put on ice for cooling. The tRNA<sup>Val</sup> was transcribed in *E. coli* using the recombinant plasmid, pVAL119-21, containing the wild-type *E. coli* tRNA<sup>Val</sup> gene downstream of a T7 promoter (Chu and Horowitz 1989; Mingsong Liu *et al.* 1997). BL21 star cells were used for the tRNA expression, and were grown in LB medium (for unlabeled RNA) or in M9-minimum medium with U-<sup>15</sup>N-NH<sub>4</sub>Cl as the sole nitrogen source (for <sup>15</sup>N-labeled RNA). The *E. coli* cells were grown at 37 °C and 0.3 mmol/L IPTG were added for induction, and the cells were harvested 4 h later by centrifugation. The cells were resuspended using 10 mmol/L Tris-HCl pH 7.5 buffer with 15 mmol/L MgCl<sub>2</sub>. Tris saturated phenol solution (pH 7.5) was added at equal volume as the pellet, and were stirred for 30 min. Subsequently, the mixture was centrifuged to separate proteins (in precipitation) from nucleic acids (in water phase). The tRNA<sup>Val</sup> was further extracted using phenol chloroform, and was precipitated using isopropanol. The final purification was performed with a Source Q column (GE Healthcare, New Brunswick, NJ).

The purity of RNA was assessed by electrophoresis on a 15% polyacrylamide gel.

### NMR spectroscopy experiments

All RNA samples were prepared in 20 mmol/L sodium phosphate buffer pH 6.5, containing 100 mmol/L NaCl, 5 mmol/L MgCl<sub>2</sub> in 10% D<sub>2</sub>O, with a final concentration of about 1 mmol/L. NMR experiments were performed at 10 °C (for tetraloop RNA hairpin and RNA pseudoknot) or 25 °C (for tRNA<sup>Val</sup>) on a Bruker 600 MHz instrument equipped with a cryogenic probe. The transverse relaxation  $R_2$  rates were measured for the imino protons of the RNA, using the established pulse sequence (Iwahara *et al.* 2007), with a delay of 12 ms between two time points. Alternatively, the  $R_2$  values were measured for well-isolated imino protons in the 1D <sup>1</sup>H NMR spectrum, with multiple spin echoes up to 12 ms. The transverse relaxation  $R_2$  rates were again measured in the presence of 2 and 5 mmol/L triethylene-tetraamine-hexa-acetate tri-methyl-amide gadolinium(III), or Gd(III)-TTHA-TMA paramagnetic cosolute, using the same pulse sequence and same time delay. In this way, the paramagnetic relaxation enhancement or solvent PRE values can be obtained (Gu *et al.* 2014). Importantly, the 5 mmol/L values almost linearly scales with the 2 mmol/L values by 2.5 times, suggesting that the sPRE data manifest the structure of the RNAs and the Gd(III)-TTHA-TMA probe is nearly inert and randomly colliding with the RNAs.

### SAXS data collection and analysis

The RNA sample used for SAXS experiment was solved in the same buffer as used for NMR, but without the D<sub>2</sub>O. The SAXS data were collected at the National Center for Protein Science Shanghai on the BL19U2 beamline. The exposure time was 1 s, and a total of 20 frames were recorded and averaged on 25 °C. The theoretical curve for a given RNA structure was calculated using CRY SOL (Svergun *et al.* 1995) in the ATSAS 2.8 package (Franke *et al.* 2017) from the refined structures.

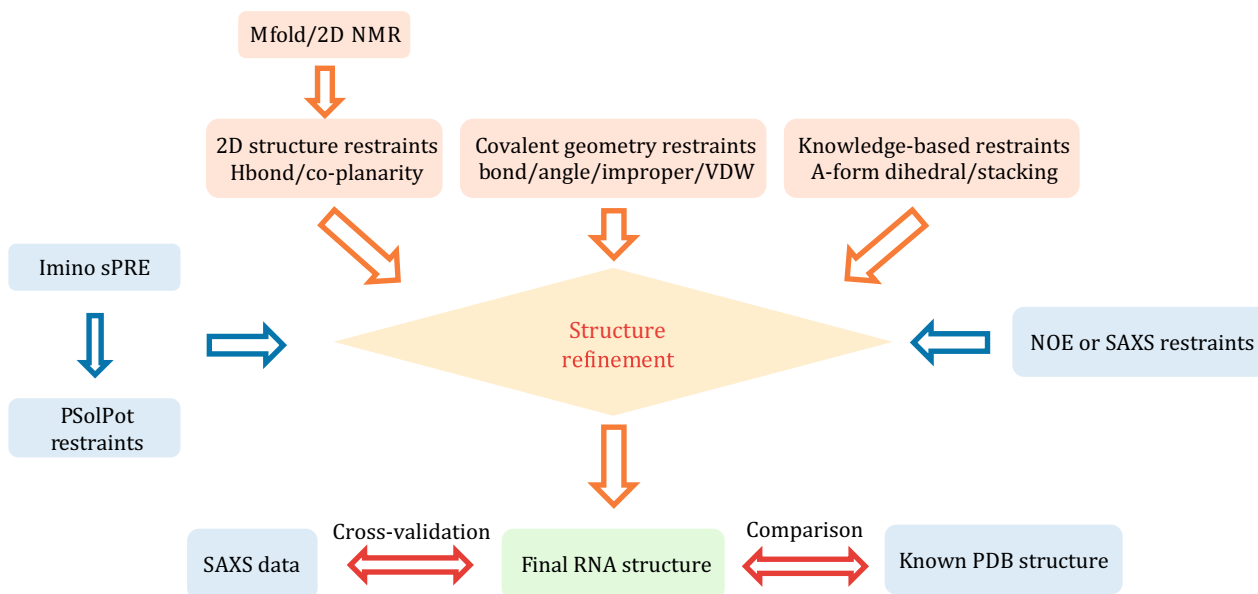
### RNA solution structure refinement

The structure refinement was performed using the Xplor-NIH (Schwieters *et al.* 2003). Secondary structure restraints used are from the secondary structure prediction using the program Mfold (Zuker 2003), which are confirmed by the appearance of imino proton peaks on the 2D <sup>1</sup>H-<sup>15</sup>N correlation spectrum. The restraints include the hydrogen-bonding restraints for the

base-pairing and restraints for the co-planarity of the base pair. In addition to covalent geometry restraints including bond angle, bond length, and improper angles, and van der Waals restraints in the form of repulsive term, knowledge-based restraints that constrain the dihedral angles close to the standard A-form helix (Schwieters *et al.* 2018) and for the stacking between the neighboring bases (Clore and Kuszewski 2003) are also invoked. In some cases, NOE and SAXS restraints were also used, which were taken from respective references (Kang and Tinoco 1997; Nozinovic *et al.* 2010). Besides using for assessing structure quality, the SAXS restraints were also directly implemented as described (Schwieters and Clore 2014). The SAXS data were discretized and only 40 angle points were used to speed up back calculation.

The PSolPot term was invoked for the structural refinement against the sPRE data. The cosolute-excluded RNA surface was covered with tessellation, and the theoretical sPRE based on surface integral was calculated. The theoretical background of PSolPot term has been described in detail elsewhere (Gong *et al.*

2018). The radius of paramagnetic cosolute molecule was set to 4 Å. In comparison, a 3.5 Å radius is used for the slightly smaller Gd(III)-DTPA-BMA probe (Pintacuda and Otting 2002). Since the sPRE values are linearly related to the concentrations of the paramagnetic cosolute, the correlation between experimental and calculated sPRE data was used as the target function. All the refinement starts from an extended RNA single strand, and the system was first heated to 3500 °C and gradually cooled to 25 °C for 15,000 steps. The refinement process was repeated to produce 240 structures. The ten structures with lowest energy were selected to evaluate the structure convergence as well as the cross-validation with the SAXS measurements. All RNA structures were illustrated using PyMOL (The PyMOL Molecular Graphics System, Version 2.2 Schrödinger). The flowchart for the structure calculation is shown in Fig. 5. The statistics of structures without and with the incorporation of the sPRE restraints is shown in supplementary Table S1. The Xplor-NIH scripts for the refinement of the three RNA systems can also be downloaded from the supplementary data.



**Fig. 5** The flowchart of RNA structure refinement with the integrative use of solvent PRE, as presented in the results. The refinement also uses knowledge-based, covalent and secondary structure-based restraints, as well as other types of experimental measurements. Note that SAXS data are directly used either as restraints, or as a way to cross-validate NMR structures



**Acknowledgements** We thank the staff of BL19U2 beamlines at National Facility for Protein Science Shanghai (NFPS) and Shanghai Synchrotron Radiation Facility, Shanghai, People's Republic of China for assistance during SAXS data collection. The work has been supported by the National Key R&D Program of China (2018YFA0507700, 2016YFA0501200) and by the National Natural Science Foundation of China (91753132 and 31770799).

### Compliance with Ethical Standards

**Conflict of interest** Zhou Gong, Shuai Yang, Qing-Fen Yang, Yue-Ling Zhu, Jing Jiang and Chun Tang declare that they have no conflict of interest.

**Human and animal rights and informed consent** This article does not contain any studies with human or animal subjects performed by any of the authors.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided 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.

### References

- Barnwal RP, Yang F, Varani G (2017) Applications of NMR to structure determination of RNAs large and small. *Arch Biochem Biophys.* <https://doi.org/10.1016/j.abb.2017.06.003>
- Bellaousov S, Reuter JS, Seetin MG, Mathews DH (2013) RNAstructure: web servers for RNA secondary structure prediction and analysis. *Nucleic Acids Res* 41(Web Server issue):W471–W474
- Bhan A, Soleimani M, Mandal SS (2017) Long noncoding RNA and cancer: a new paradigm. *Cancer Res* 77(15):3965–3981
- Bhandari YR, Jiang W, Stahlberg EA, Stagno JR, Wang YX (2016) Modeling RNA topological structures using small angle X-ray scattering. *Methods* 103:18–24
- Borkar AN, Vallurupalli P, Camilloni C, Kay L, Vendruscolo M (2016) Simultaneous NMR characterisation of multiple minima in the free energy landscape of an RNA UUCG tetraloop. *Phys Chem Chem Phys.* <https://doi.org/10.1039/c6cp08313g>
- Butcher SE, Pyle AM (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* 44(12):1302–1311
- Chu WC, Horowitz J (1989) 19F NMR of 5-fluorouracil-substituted transfer RNA transcribed *in vitro*: resonance assignment of fluorouracil-guanine base pairs. *Nucleic Acids Res* 17(18):7241–7252
- Clore GM, Iwahara J (2009) 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(9):4108–4139
- Clore GM, Kuszewski J (2003) Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipolar coupling cross-validation. *J Am Chem Soc* 125(6):1518–1525
- Dethoff EA, Hansen AL, Musselman C, Watt ED, Andricioaei I, Al-Hashimi HM (2008) Characterizing complex dynamics in the transactivation response element apical loop and motional correlations with the bulge by NMR, molecular dynamics, and mutagenesis. *Biophys J* 95(8):3906–3915
- Duchardt E, Schwalbe H (2005) Residue specific ribose and nucleobase dynamics of the cUUCG RNA tetraloop motif by MNMR 13C relaxation. *J Biomol NMR* 32(4):295–308
- Duss O, Maris C, von Schroetter C, Allain FH (2010) A fast, efficient and sequence-independent method for flexible multiple segmental isotope labeling of RNA using ribozyme and RNase H cleavage. *Nucleic Acids Res* 38(20):e188
- Duss O, Diarra Dit Konte N, Allain FH (2015) Cut and paste RNA for nuclear magnetic resonance, paramagnetic resonance enhancement, and electron paramagnetic resonance structural studies. *Methods Enzymol* 565:537–562
- Eddy SR (2001) Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2(12):919–929
- Edwards TE, Sigurdsson ST (2007) Site-specific incorporation of nitroxide spin-labels into 2'-positions of nucleic acids. *Nat Protoc* 2(8):1954–1962
- Franke D, Petoukhov MV, Konarev PV, Panjkovich A, Tuukkanen A, Mertens HDT, Kikhney AG, Hajizadeh NR, Franklin JM, Jeffries CM, Svergun DI (2017) ATSAS 28: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr* 50(Pt 4):1212–1225
- Furtig B, Richter C, Wohnert J, Schwalbe H (2003) NMR spectroscopy of RNA. *Chem Bio Chem* 4:936–962
- Gong Z, Schwieters CD, Tang C (2015) Conjoined use of EM and NMR in RNA structure refinement. *PLoS One* 10(3):e0120445
- Gong Z, Gu XH, Guo DC, Wang J, Tang C (2017) Protein structural ensembles visualized by solvent paramagnetic relaxation enhancement. *Angew Chem Int Ed Engl* 56(4):1002–1006
- Gong Z, Schwieters CD, Tang C (2018) Theory and practice of using solvent paramagnetic relaxation enhancement to characterize protein conformational dynamics. *Methods* 148:48–56
- Grishaev A, Ying J, Canny MD, Pardi A, Bax A (2008) Solution structure of tRNAVal from refinement of homology model against residual dipolar coupling and SAXS data. *J Biomol NMR* 42(2):99–109
- Gu XH, Gong Z, Guo DC, Zhang WP, Tang C (2014) A decadentate Gd(III)-coordinating paramagnetic cosolvent for protein relaxation enhancement measurement. *J Biomol NMR* 58(3):149–154
- Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* 5(12):1065–1074
- Hansen MR, Hanson P, Pardi A (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol* 317:220–240
- Hartlmüller C, Gunther JC, Wolter AC, Wohnert J, Sattler M, Madl T (2017) RNA structure refinement using NMR solvent accessibility data. *Sci Rep* 7(1):5393
- Helmling C, Bessi I, Wacker A, Schnorr KA, Jonker HR, Richter C, Wagner D, Kreibich M, Schwalbe H (2014) Noncovalent spin labeling of riboswitch RNAs to obtain long-range structural NMR restraints. *ACS Chem Biol* 9:1330–1339
- Herschlag D, Bonilla S, Bisaria N (2018) The story of RNA folding, as told in epochs. *Cold Spring Harb Perspect Biol.* <https://doi.org/10.1101/cshperspect.a032433>
- Iwahara J, Tang C, Marius Clore G (2007) Practical aspects of (1)H transverse paramagnetic relaxation enhancement measurements on macromolecules. *J Magn Reson* 184(2):185–195
- Kang H Jr, Tinoco I (1997) A mutant RNA pseudoknot that promotes ribosomal frameshifting in mouse mammary tumor virus. *Nucleic Acids Res* 25(10):1943–1949

- Kappel K, Liu SH, Larsen KP, Skiniotis G, Puglisi EV, Puglisi JD, Zhou ZH, Zhao R, Das R (2018) *De novo* computational RNA modeling into cryo-EM maps of large ribonucleoprotein complexes. *Nat Methods* 15(11):947–954
- Larsen KP, Choi J, Prabhakar A, Puglisi EV, Puglisi JD (2019) Relating structure and dynamics in RNA biology. *Cold Spring Harb Perspect Biol*. <https://doi.org/10.1101/cshperspect.a032474>
- Liu Z, Gong Z, Guo DC, Zhang WP, Tang C (2014) Subtle dynamics of holo glutamine binding protein revealed with a rigid paramagnetic probe. *Biochemistry*. <https://doi.org/10.1021/bi4015715>
- Liu Z, Gong Z, Jiang WX, Yang J, Zhu WK, Guo DC, Zhang WP, Liu ML, Tang C (2015) Lys63-linked ubiquitin chain adopts multiple conformational states for specific target recognition. *Elife*. <https://doi.org/10.7554/eLife.05767>
- Liu Z, Gong Z, Dong X, Tang C (2016) Transient protein–protein interactions visualized by solution NMR. *Biochim Biophys Acta Bioenerg* 1864(1):115–122
- Mingsong Liu W-CC, Liu Jack C-H, Horowitz J (1997) Role of acceptor stem conformation in tRNAVal recognition by its cognate synthetase. *Nucleic Acids Res* 25(24):4883–4890
- Murray LJW, Arendall WB, Richardson DC, Richardson JS (2003) RNA backbone is rotameric. *Proc Natl Acad Sci USA* 100(24):13904–13909
- Nozinovic S, Furtig B, Jonker HR, Richter C, Schwalbe H (2010) High-resolution NMR structure of an RNA model system: the 14-mer cUUCGg tetraloop hairpin RNA. *Nucleic Acids Res* 38(2):683–694
- Parisien M, Major F (2008) The MC-fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 452(7183):51–55
- Pintacuda G, Otting G (2002) Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate. *J Am Chem Soc* 124(3):372–373
- Ponce-Salvatierra A, Astha, Merdas K, Nithin C, Ghosh P, Mukherjee S, Bujnicki JM (2019) Computational modeling of RNA 3D structure based on experimental data. *Biosci Rep*. <https://doi.org/10.1042/bsr20180430>
- Schlick T, Pyle AM (2017) Opportunities and challenges in RNA structural modeling and design. *Biophys J* 113(2):225–234
- Schwieters CD, Clore GM (2014) Using small angle solution scattering data in Xplor-NIH structure calculations. *Prog Nucl Magn Reson Spectrosc* 80:1–11
- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73
- Schwieters CD, Bermejo GA, Clore GM (2018) Xplor-NIH for molecular structure determination from NMR and other data sources. *Protein Sci* 27:26–40
- Suddala KC, Walter NG (2014) Riboswitch structure and dynamics by smFRET microscopy. *Riboswitch Discov Struct Funct* 549:343–373
- Svergun D, Barberato C, Koch MHJ (1995) CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773
- Wang Y, Schwieters CD, Tjandra N (2012) Parameterization of solvent–protein interaction and its use on NMR protein structure determination. *J Magn Reson* 221:76–84
- Warhaut S, Mertinkus KR, Hollthaler P, Furtig B, Heilemann M, Hengesbach M, Schwalbe H (2017) Ligand-modulated folding of the full-length adenine riboswitch probed by NMR and single-molecule FRET spectroscopy. *Nucleic Acids Res* 45(9):5512–5522
- Wilusz JE, Sunwoo H, Spector DL (2009) Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23(13):1494–1504
- Wunderlich CH, Huber RG, Spitzer R, Liedl KR, Kloiber K, Kreutz C (2013) A novel paramagnetic relaxation enhancement tag for nucleic acids: a tool to study structure and dynamics of RNA. *ACS Chem Biol*. <https://doi.org/10.1021/cb400589q>
- Zhang XJ, Cekan P, Sigurdsson ST, Qin PZ (2009) Studying RNA using site-directed spin-labeling and continuous-wave electron paramagnetic resonance spectroscopy. *Methods Enzymol* 469:303–328
- Zhang H, Zhang C, Li Z, Li C, Wei X, Zhang B, Liu Y (2019) A new method of RNA secondary structure prediction based on convolutional neural network and dynamic programming. *Front Genet* 10:467
- Zhao Y, Wang J, Zeng C, Xiao Y (2018) Evaluation of RNA secondary structure prediction for both base-pairing and topology. *Biophys Rep* 4(3):123–132
- Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13):3406–3415