J Biomol NMR (2008) 42:139–142
DOI 10.1007/s10858-008-9271-1

ARTICLE

Conformational signatures of ^{13}C chemical shifts in RNA ribose

Oliver Ohlenschläger · Sebastian Haumann · Ramadurai Ramachandran · Matthias Görlach

Received: 24 May 2008 / Accepted: 18 August 2008 / Published online: 20 September 2008

Abstract The conformational dependence of ^{13}C chemical shift values of RNA riboses determined by liquid-state NMR spectroscopy was evaluated using data deposited for RNA structures in the RCSD and BMRB data bases. Results derived support the applicability of the canonical coordinates approach of Rossi and Harbison (J Magn Reson 151:1–8, 2001) in liquid-state NMR to assess the sugar pucker of ribose units in RNA.

Keywords NMR spectroscopy · RNA · Chemical shift · Ribose · Sugar pucker

Introduction

The conformational dependence of ^{13}C chemical shift values of ribonucleic acid sugars as determined by liquid-state NMR spectroscopy was evaluated. RNA constitutes a central player in gene expression including regulatory processes. Currently over 600 RNA high resolution structures determined by X-ray crystallography and liquid-state NMR spectroscopy are available.

The utilisation of heteronuclear chemical shift information based upon e.g. the 'chemical shift index' (CSI) as introduced by Wishart and Sykes (1994) is nowadays an accepted supplemental tool in protein NMR structure determination. Although studies to rationalise a shift-structure relationship by solid-state NMR and SCF/DFT approaches exist (Ebrahimi et al. 2001; Rossi and Harbison 2001; Xu and Au-Yeung 2000; Xu et al. 1998), an approach in RNA structure determination by NMR equivalent to the CSI is currently not established, probably also due to the delay in the increase of the number of nucleic acid structures for which heteronuclear chemical shift information became available by using stable isotope labelling techniques (Batey et al. 1992; Nikonowicz et al. 1992).

Based on their work on DNA ^{13}C chemical shifts (Santos et al. 1989), Harbison et al. investigated the conformational dependence of the ^{13}C chemical shifts in ribonucleotides and -nucleosides (Ebrahimi et al. 2001; Rossi and Harbison 2001). As result, an elegant approach to extract conformational information about sugar puckers and exocyclic angles from heteronuclear RNA chemical shift data via two canonical coordinates was proposed (Ebrahimi et al. 2001; Fürtig et al. 2003).

This has prompted us to perform a survey and evaluation of data deposited for RNA structures in the RCSB and BMRB data bases.

Materials and methods

In a first step all entries with ^{13}C assignments for the sugar moieties were identified in the BMRB. The 23 data sets (for a complete list see Table S1 in the Supplementary materials) were extracted and each entry was checked for availability of full assignments for all five backbone carbon atoms (C1′, C2′, C3′, C4′, C5′). Residues with only

Electronic supplementary material The online version of this article (doi:10.1007/s10858-008-9271-1) contains supplementary material, which is available to authorized users.

O. Ohlenschläger (✉) · S. Haumann · R. Ramachandran · M. Görlach
Leibniz Institute for Age Research/Fritz Lipmann Institute, Biomolecular NMR Spectroscopy, Beutenbergstr. 11, 07745 Jena, Germany
e-mail: oliver@fli-leibniz.de
URL: http://nmr.fli-leibniz.de

partially assigned carbon data had to be excluded from further analysis due to the nature of the two formulas for the canonical coordinates (see below). The resulting data set comprised 429 residues with complete sugar/backbone carbon data.

The calculation of the canonical coordinates was performed for the 429 residue data set following (Rossi and Harbison 2001). Essentially, the method is based on the calculation of the two canonical coordinates can1 and can2 as a sum of products of empirically derived constants multiplied by the chemical shift value of the respective sugar backbone carbon atom.

$$\text{can1} = 0.179\delta_{C1'} - 0.225\delta_{C4'} - 0.0585\delta_{C5'} \quad (1)$$

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

The location of a residues coordinates in the plot of can2 (y-axis) versus can1 (x-axis) should then allow a prediction of its sugar pucker and exocyclic torsion angle conformation. can1 values above -6.77 are taken as indication for a C3′-*endo* conformation while can2 below -16.82 predict a γ-torsion angle in gauche-trans (gt) conformation.

In the RCSB data base the PDB entries corresponding to hits in the BMRB (cf. Table S1) were extracted and records relating to RNA solution structures were identified. Subsequently, the coordinates of the respective model representative for the NMR structural ensemble were analysed. In cases where the PDB entry information for the representative conformer was lacking, the mean structure for the ensemble was generated with the program MOLMOL (Koradi et al. 1996) and the structure with the least backbone r.m.s.d. to the mean structure was selected as model for further analysis. In parallel, the corresponding publications were screened for information about conformational averaging of sugar pucker states. The respective residues were also excluded from further analysis.

For the selected models the sugar torsion angles ν_1, ν_2 and δ as well as the exocyclic torsion angle γ were determined using the coordinate files. Torsion angles and literature information were used to classify the two main pucker conformations C3′-*endo* and C2′-*endo* as well as the rotational states gg and gt of the torsion angle γ according to Saenger (1983).

5′- and 3′-Terminal nucleotides were removed from the analysis since—irrespective of the conformations deposited in the PDB data—end fraying effects cannot be ruled out. These dynamic effects are often concomitant with a change in sugar conformation and therefore are prone to obscure the distribution statistics obtained for can1.

From the first ensemble plot of the canonical coordinates (calculated according to (1) and (2)) it became obvious that the chemical shift entries deposited in the BMRB were calibrated according to at least two different reference values (Fig. S1). Consistency was reached only when several of the data sets were recalibrated by subtraction of 1.6 ppm which corresponds to the usage of different reference systems (DSS vs. TMS referenced to 10% dioxane at 67.8 ppm) as outlined by Wishart and Sykes (1994) for the protein case. Thus subsequently, the identified deviating data sets (see remark in Supplementary materials Table S1) were recalibrated and the corresponding canonical coordinates (Fig. S2) were re-evaluated.

Results and discussion

Conformational states in the structure data base

Analysis of the 429 residues for which a full chemical shift data set was available resulted in 377 sugar puckers (88%) in C3′-*endo* conformation and 52 conformers (12%) in C2′-*endo*. A few sugar moieties could not unambiguously be assigned to one of the canonical sugar pucker conformations based on the ν_1, ν_2 and δ torsion angles and were excluded from analysis, even when applying very relaxed limits.

Conformational predictions by the canonical coordinates

Figures 1 and S2 show the plot of the canonical coordinates can1 vs. can2. The highest number of residues (93%) clearly clusters at the C3′-*endo* side of the diagram (can1 > -6.77; left of solid vertical line) as expected for residues in an A-form RNA. In addition, the exocyclic torsion γ is predicted to be mainly populated (75%) in the gg conformation (can2 > -16.82; below horizontal line). The remaining residues show a C2′-*endo* conformation with an exocyclic torsion in gt state. It becomes also evident that the analysed chemical shift data here do rarely (<1%) reflect a C2′-*endo*/gg torsion angle combination. Already this finding indicates a strong correlation of the predictions derived from the chemical shift data with the experimentally determined distribution of conformers (see above).

Comparison on the structural level

Further validation was obtained by correlation of the sugar pucker of individual residues as found in the RNA structure deposited in the RCSB with their respective chemical shift-derived canonical coordinates. This analysis indicates that in 94% of C3′-*endo* cases and 93% of C2′-*endo* cases the predicted and experimentally determined sugar pucker coincide (Fig. 1) using the can1 limit of -6.77. Correct

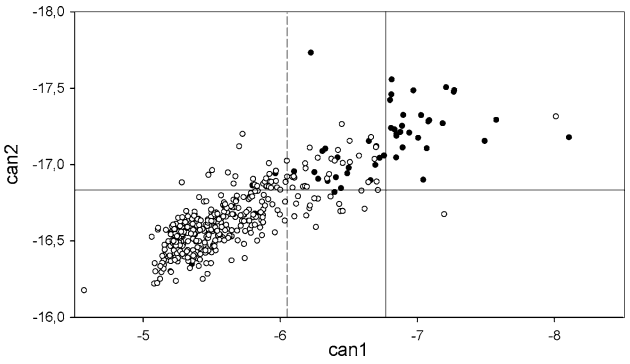

Fig. 1 Canonical coordinate predictions and experimentally observed C3′-*endo* (white) and C2′-*endo* (black) conformations. The dashed line indicates the proposed threshold for pyrimidines

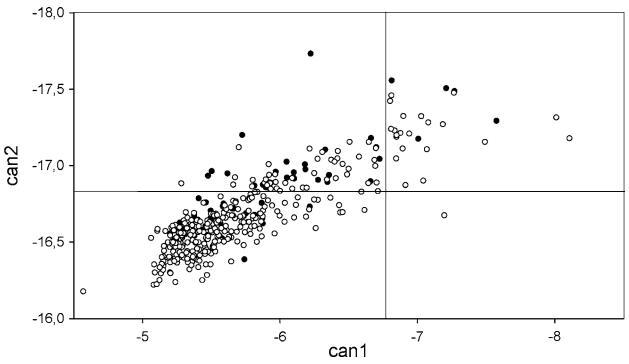

Fig. 3 Canonical coordinate predictions and experimentally observed gg (white) and gt (black) conformations

predictions were also obtained using data from two RNA:protein complexes (cf. Table S1).

However, we noticed a nucleotide-specific difference. While predictions for purines (A, G) match well the can1 limit given (Rossi and Harbison 2001), C2′-*endo* conformations for pyrimidines (C and U) tend to extend to smaller can1 values (see Fig. 2). This suggested that the concept of conformational discrimination based on can1 coordinates could be fine-tuned in a nucleotide-specific manner. Indeed, when limiting to a smaller value of −6.05 for the pyrimidines, percentages for the C2′-*endo* prediction increased from 39 to 77% at the expense of only 6% of correct C3′-*endo* cases (Fig. 1).

In contrast to the clear separation of sugar pucker modes via can1, no clear chemical shift dependence for the conformation of the exocyclic torsion angle γ could be detected. Figure 3 shows the canonical coordinates marked by torsion angle state. It is clearly seen that no significant discrimination for residues in the gg and gt conformations exists. This applies also if the analysis is performed in a nucleotide specific way (data not shown).

For the presented study of the 614 RNA depositions in the RCSB only 324 structures solved by NMR methods were harvested. Based on the analysis of coordinates of 208 RNA structures with NMR data submitted (64%) only 23 sets (11%) could be identified with virtually complete assignments of carbon atoms (C1′, C2′, C3′, C4′ and C5′) as deposited in the BMRB data bank. Hence, the survey presented had to be restricted to a rather small basis set. Still, chemical shift data for 429 residues could be evaluated. These included 134 guanosines (31%), 78 adenosines (18%), 133 cytosines (31%) and 84 uridines (20%).

The results derived serve as an indication for the applicability of the canonical coordinates approach (Rossi and Harbison 2001) in liquid-state NMR. Adjustment of the limits for the C3′-*endo*/C2′-*endo* classification based on the presented empirical findings could increase the reliability of the prediction: We propose to use the can1 limit for discrimination between the C3′-*endo*/C2′-*endo* pucker

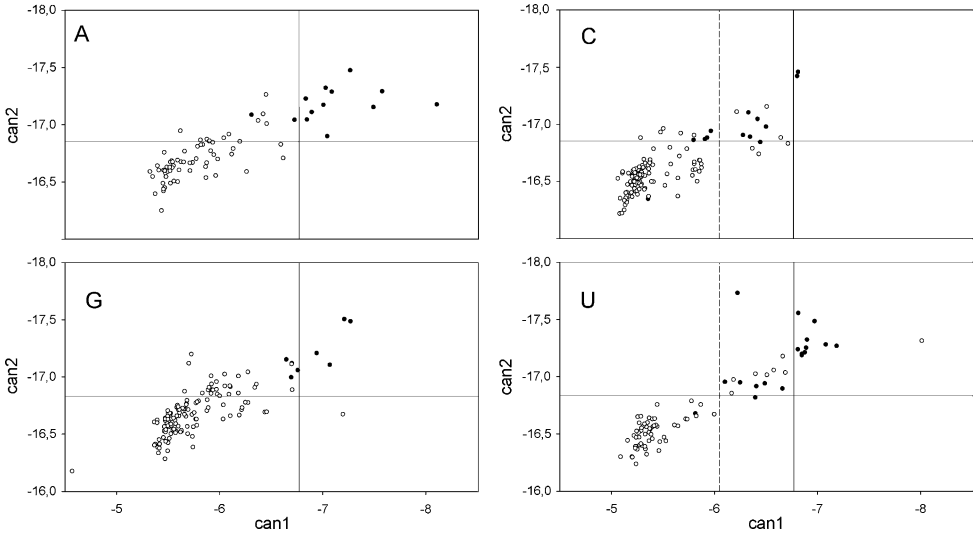

Fig. 2 Canonical coordinate predictions and experimentally observed C3′-*endo* (white) and C2′-*endo* (black) conformations separated by nucleic acid type. The dashed line indicates the proposed new threshold for pyrimidines

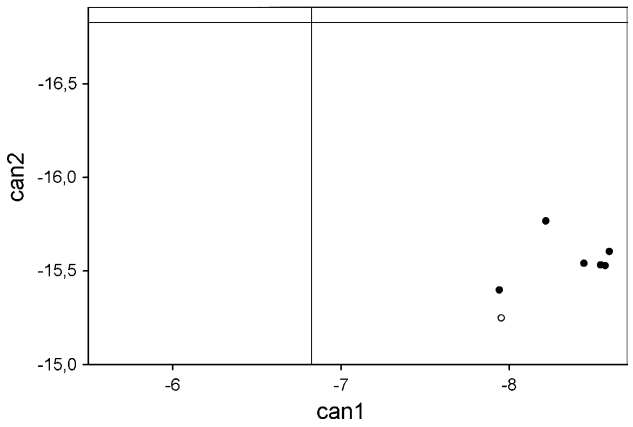

Fig. 4 Canonical coordinate predictions and experimentally observed C3′-*endo* (white) and C2′-*endo* (black) conformations of DNA (1KR8)

modes as introduced by Rossi and Harbison (2001) only for the purine nucleotides A and G. In addition, for pyrimidines a can1 threshold of −6.05 should be applied to assign C3′-*endo* conformations to C and U residues.

As further support for the applicability of the concept the result of a DNA shift evaluation might be used. Of 244 DNA data sets with NMR data deposited in the RCSB only one data set was found to contain carbon chemical shift data (1KR8). This prevented an extension of the full analysis to DNA. However, the structure mentioned describes a short DNA mini-hairpin with a stem of three base pairs in B-form (Padrta et al. 2002). Analysis in terms of canonical coordinates showed that for all residues a C2′-*endo* conformation was predicted (Fig. 4). With the exception of the terminal C7 all sugar pucker modes in this DNA structure exhibited C2′-*endo* conformation.

From the results presented the utilisation of the canonical coordinates concept at two different stages of NMR structure analysis of RNA can be envisaged: (1) Analysis of the carbon chemical shifts in terms of the canonical coordinate can1 will lead to a reliable prediction of C3′-*endo* sugar conformations after the resonance assignment step. Especially for a fast structure determination of smaller RNA, for which the approach should not suffer from resonance overlap in the carbon dimension, this method could provide a useful cross-check tool for sugar conformations estimated by other approaches (e.g. TOCSY or ^{3}J coupling data). (2) In analogy to the Ramachandran plot used to determine anomalous protein backbone conformations, the canonical coordinates can be employed to assess the quality of the experimentally determined sugar conformations for a given NMR structure. Unfortunately, a similar discrimination between the gg and gt conformational states for the exocyclic torsion angle γ was not observed here precluding the use of the canonical coordinates approach for a reliable prediction of this phosphate backbone torsion angle.

Acknowledgements The FLI is financially supported by the State of Thuringia and the Federal Government of Germany.

References

Batey RT, Inada M, Kujawinski E, Puglisi JD, Williamson JR (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR spectroscopy of RNA. Nucleic Acids Res 20:4515–4523

Ebrahimi M, Rossi P, Rogers C, Harbison GS (2001) Dependence of ^{13}C NMR chemical shifts on conformations of RNA nucleosides and nucleotides. J Magn Reson 150:1–9

Fürtig B, Richter C, Wöhnert J, Schwalbe H (2003) NMR spectroscopy of RNA. Chembiochem 4:936–962

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55

Nikonowicz EP, Sirr A, Legault P, Jucker FM, Baer LM, Pardi A (1992) Preparation of ^{13}C and ^{15}N labelled RNAs for heteronuclear multi-dimensional NMR studies. Nucleic Acids Res 20:4507–4513

Padrta P, Stefl R, Kralik L, Zidek L, Sklenar V (2002) Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings. J Biomol NMR 24:1–14

Rossi P, Harbison GS (2001) Calculation of ^{13}C chemical shifts in RNA nucleosides: structure-^{13}C chemical shift relationships. J Magn Reson 151:1–8

Saenger W (1983) Principles of nucleic acid structure. Springer, Berlin

Santos RA, Tang P, Harbison GS (1989) Determination of the DNA sugar pucker using ^{13}C NMR spectroscopy. Biochemistry 28:9372–9378

Wishart DS, Sykes BD (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. J Biomol NMR 4:171–180

Xu X-P, Au-Yeung SCF (2000) Investigation of chemical shift and structure relationships in nucleic acids using NMR and density functional theory methods. J Phys Chem B 104:5641–5650

Xu X-P, Chiu W-LAK, Au-Yeung SCF (1998) Chemical shift and structure relationship in nucleic acids: correlation of backbone torsion angles γ and with ^{13}C chemical shifts. J Am Chem Soc 120:4230–4231