Nuclear magnetic resonance studies of ribonucleic acids

Oliver Ohlenschläger a,*, Jens Wöhnert b, Ramadurai Ramachandran a, Christian Sich a,1 and Matthias Görlach a

a Institut für Molekulare Biotechnologie e.V., Beutenbergstr. 11, D-07745 Jena, Germany
b J. W. Goethe-Universität, Marie-Curie-Str. 6-8, D-06110 Frankfurt am Main, Germany

Abstract. Ribonucleic acids (RNA) and RNA–protein complexes are essential components of biological information transfer, catalytic processes and are associated with regulatory functions. This broad range of biological functions is paralleled at the conformational level by a large number of non-canonical structural elements or sequences with non-standard backbone conformations, e.g., loops, bulges, pseudo-knots and complex tertiary folds. NMR spectroscopy has evolved to a powerful tool for the determination of ribonucleic acid structures of up to 20 kDa. Uniform or selective stable isotope labelling aids in solving assignment problems arising from the inherently limited chemical shift dispersion and overlap of resonances for larger nucleotide sequences. Recent developments of multi-dimensional heteronuclear NMR pulse sequences allow e.g., to directly observe the hydrogen bonding pattern of canonical Watson–Crick base pairs as well as of unusual types of base pairs, thereby opening up a fast access to secondary structure screening of RNA. Detailed conformational descriptions are obtained using conventional NOE and J-coupling-derived data, nowadays supplemented by information from residual dipolar couplings. The latter method also provides a new means for the probing of dynamical features of ribonucleic acids.

Keywords: RNA, NMR, structure determination, stable isotope labelling

Abbreviations

COSY, correlation spectroscopy; HSQC, heteronuclear single-quantum correlation; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; nt, nucleotide; TROSY, transverse relaxation-optimized spectroscopy; RDC, residual dipolar couplings; RNA, ribonucleic acid; r.m.s., root mean square.

1. Introduction

Ribonucleic acids are mediators between the genetic information and the expression of this information. They constitute the entire genetic material of some viruses, catalyse important biological reactions or are involved in regulatory processes. This explains the steadily increasing interest in the knowledge of RNA structure at the atomic level. Progress in the structure determination of RNA by NMR can be attributed to the availability of biochemical techniques for isotope labelling of RNA and the development of NMR experiments tailored towards the extraction of relevant structural data. This short review summarises some recent contributions of NMR spectroscopy to the study of RNAs.

* Corresponding author. Tel.: +49 3641 656219; Fax: +49 3641 656225; E-mail: oliver@imb-jena.de.
1 Present address: Volkswagen AG, Brieffach 1617/0, D-38436 Wolfsburg, Germany.
2. Recombinant RNA labelling techniques

NMR spectroscopical methods employing triple-resonance heteronuclear experiments, now routinely used for the structure determination of nucleic acids up to 45–50 bases, require access to stable-isotope labelled samples. RNA labelled uniformly with the NMR-active nuclei $^{13}\text{C}/^{15}\text{N}$ is routinely obtained by \textit{in vitro} transcription methods [1,2] employing T7 RNA polymerase and a synthetic or recombinant DNA template similar to the one depicted in Fig. 1. Due to the limited chemical shift dispersion seen in the $^1\text{H}$ dimension of RNAs (Fig. 2) a reliable extraction of distance restraints can become difficult. In this context, for certain resonance assignment problems and particularly for the study of larger RNA systems, stable isotope labelling with $^{13}\text{C}$ and/or $^{15}\text{N}$ becomes necessary. Segment- and strand-specific labelling techniques have been introduced to simplify spectral analysis [3,4] by reducing spectral overlaps. Furthermore, RNA can also be labelled base-selectively or atom-specifically e.g., by making use of a new $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ labelling scheme relying on the conversion of specifically labelled glucose and bases into nucleotides employing enzymes from the pentose phosphate pathway and subsequent phosphoribosyl transferase treatment [5].

Fig. 1. Labelling scheme including the production of labelled nucleotide-triphosphates (rNTP) and their subsequent utilisation in a T7 polymerase transcription. In the approach depicted, a hammerhead (HH) ribozyme is fused 3' to the sequence of the desired RNA strand. Self-cleavage of the primary transcript yields homogenous 3'-ends of the desired RNA.
Fig. 2. Typical \([1^H-1^3C]\) HSQC spectrum of a RNA illustrating the poor chemical shift dispersion of sugar resonances in the \(^1H\) (A) dimension. The spectrum in (B) corresponds to a 19-mer RNA [9].

3. NMR experiments

Numerous NMR experiments have been developed to perform the three basic steps in the assignment of nucleic acids: (i) the identification of the sugar resonances, (ii) the assignment of the base resonances and (iii) the linking of the bases to their respective sugar moiety. For reviews about the standard pulse
sequences used in this assignment process see Wijmenga & van Buuren or Varani et al. [6,7]. Selective pulses, resulting in an excitation of the selected nuclei, are an integral part of a number of novel NMR experiments. A modified HCCH-COSY experiment optimised for the correlation of H2 and H8 resonances provides the complete $^{13}$C resonance assignment in the adenine base [8]. 2D triple-resonance H6/H5(C4N)H (Fig. 3) and C6/C5(C4N)H experiments separately tuneable for uridines and cytosines were recently developed in our labs (unpublished) and facilitate the assignment of exchangeable protons by simultaneously linking them to all non-exchangeable protons in the base. While resonance assignment in regular A-form helices is often performed via imino/imino and imino/amino NOEs, these experiments are particularly useful for irregular regions. 2D triple-resonance H5(C5C4N)H experiments to correlate the H5 proton to its intra-base amino/imino hydrogens in uridines and cytidines also employ selective excitation steps [10].

In larger molecules the sensitivity of the conventional NMR experiments is drastically reduced due to fast transverse relaxation processes. Two methods allow to overcome this problem: NMR experiments utilizing multiple-quantum coherence states and transverse relaxation-optimised spectroscopy.

![Diagram](image)

Fig. 3. H6H5(C4N)H experiment correlating the resonances of the aromatic H6 and H5 protons with their respective exchangeable base hydrogen resonance.
Marino et al. [11] have demonstrated that the sensitivity of triple-resonance HCN experiments can be considerably improved by the use of multiple-quantum coherence. This multiple-quantum approach was extended by the group of Sklenar [12] and shown to work best for sugar-to-base correlation experiments [13], e.g., a H1′–C1′–N1/N9 transfer. From the same laboratory a new multiple-quantum based 2D \(J\)-correlation experiment was introduced facilitating the precise measurement of \(^3 J_{C2/C4H1′}\) and \(^3 J_{C6/C8H1′}\) coupling constants in uniformly \(^{13}\)C-labelled nucleic acids [14] which define the glycosidic torsion angle \(\chi\). Limitations due to relaxation processes in larger RNAs can be minimised also by TROSY experiments [15]. The selection of the slowly relaxing signal component in the TROSY experiment provides improved spectral resolution and can be combined with the aforementioned multiple-quantum approach [16]. The TROSY pulse sequence module has been incorporated into a number of multi-dimensional NMR experiments [8,13,17,18]. Especially in combination with the HCN experiment optimal performance is achieved for the base moieties (H6/H8–C6/C8–N1/N9 transfer) [13]. It is expected that the application of these techniques will increase the range of RNAs accessible for NMR structure determination.

The 2D HNN-COSY experiment allows the direct detection of NH···N imino hydrogen-mediated H-bonds in RNA. This method, introduced by Grzesiek and colleagues [19], utilizes the two-bond \(J\) coupling between the donor and acceptor nitrogen nuclei of the base pairs. It delivers information about canonical Watson–Crick-type base pairs as well as non-canonical hydrogen bonds occurring in e.g., Hoogsteen base pairs or other mismatches [20]. Figure 4 shows an example of a HNN-COSY experiment of a 30-mer RNA currently studied in our laboratory. Modifications of the original pulse sequence by including e.g., a TROSY-based module [21], the spin-echo difference method [22] or an optimization for the detection of hydrogen bonds in case of non-observable imino resonances have been introduced [23]. An adaptation of the HNCO experiment [24] which was initially developed for protein NMR spectroscopy also yields base pairing information by exploiting the hydrogen-bond mediated \(^{3h}J_{NC}\) coupling (NH···OC) [25].

Fig. 4. HNN-COSY spectrum of a viral 30-mer RNA forming three canonical Watson–Crick A:U and five C:G base pairs as well as a non-canonical C:U base pair (see U24).
4. Structural information

The conventional nuclear Overhauser effect (NOE) contributes – due to its $r^{-6}$ dependence – distance information only for protons separated by up to approx. 6 Å. Thus, this source of information can be used in nucleic acids to deduce the conformation of one base step and one base pair. By the measurement of residual dipolar couplings (RDC; see Fig. 5), which in contrast to NOEs follow a $r^{-3}$ distance dependence, additional orientational restraints are accessible and can be employed for RNA structure refinement. The data from RDC measurements depend upon the orientation of interatomic vectors (e.g., the N–H bond) relative to an external coordinate system (alignment tensor of the molecule in a liquid crystalline medium). For RNA, filamentous phages [26,27] as well as alcohol/polyethylene glycol mixtures [28] were successfully employed to reduce the isotropic rotation and to introduce an alignment relative to the anisotropic medium. Starting from a 3D model and incorporating 27 RDCs obtained from a $^{15}\text{N}$-labelled sample in Pf1 phage solution, the global structure of tRNA$^{\text{Val}}$ was determined by Mollova et al. [29]. RDCs from a Pf1 phage-oriented sample were also used to refine the solution structure of the 29-mer Sarcin-Ricin hairpin-loop from rat 28S rRNA [30]. By inclusion of 100 $^{1}\text{H}–^{13}\text{C}$ RDCs the r.m.s. deviation of the structure family to the averaged structure dropped by approx. 0.5 Å due to the improved definition of its helical bend. Another example is the structure of the theophylline-binding RNA aptamer which was recently refined to an average pairwise r.m.s. deviation of 1.5 Å with data obtained from a sample oriented in filamentous phages [31]. In contrast, a mixture of polyethylene glycol and hexanol was employed to determine the structure of hairpin P5.1 of RNase P RNA in a weakly oriented state [32].

In contrast to protein structure determination, the higher number of torsion angles describing a nucleotide conformation (backbone angles $\alpha, \beta, \gamma, \varepsilon, \zeta$, sugar angles $\delta, \nu_1, \nu_2, \nu_4$, glycosidic torsion angle $\chi$)

![Fig. 5. Residual dipolar couplings as observed in undecoupled $^{15}\text{N,}^{1}\text{H}$-HSQC spectra of a viral 30-mer RNA recorded without (A) and with 7 mg/ml Pf1 phages (B).](image-url)
O. Ohlenschläger et al. / Nuclear magnetic resonance studies of ribonucleic acids

and the reduced density of protons in nucleic acids (67% of a protein) necessitate an increased number of NMR experiments in order to obtain sufficient structural information. Thus, in addition to distance information, the conformational description of the nucleotide units can be considerably improved by including $^{3}J$ coupling constants. For an overview on the different $J$ coupling restraints which can be utilized in RNA structure determination see Marino et al. [33]. The $^{3}J$ coupling constants are parametrized using the Karplus equation coefficients and converted into torsion angle restraints. Due to the high number of torsion angles, efficient grid-search modules such as the FOUND algorithm [34] or AngleSearch [35] are employed to generate torsion angle constraints to be included in distance geometry calculations.

Several programs are available for a subsequent structure calculation, like X-PLOR [36], CNS [37] and DYANA [38]/CYANA [39]. The simulated annealing algorithm has proven to allow for an efficient optimisation towards the global minimum. In the easiest approach the function to be optimised (target function) consists of terms for the deviation of the actual values of the calculated conformer against the experimental data. In the torsion-angle dynamics approach (like DYANA), the covalent geometries are kept fixed and the optimisation proceeds by changing only the torsion angles. A similar approach is implemented in the program CNS. Finally, the generated conformers have to be relaxed by application of force field methods which integrate properties like charges or more complex non-bonded interactions. This energy minimisation and molecular dynamics step can be performed separately, e.g., by the program OPAL [40], or by using integrated programs like XPLOR/CNS, SYBYL (Tripos Inc., St. Louis) or Insight II (Accelrys, San Diego).

Molecular dynamics calculations provide access to dynamical information of nucleic acids like order parameters which can also be deduced from relaxation time measurements. These experiments are sensitive to motions on a pico- to nanosecond time scale. However, in order to analyse the data, an appropriate motional model, e.g., anisotropic or isotropic motion in the model-free formalism [41,42], has to be presumed. By analysing $^{15}$N spin relaxation measurements, the base dynamics of guanine and uracil bases in a UUCG tetraloop were shown to be of similar rigidity to the paired stem bases [43]. For this tetraloop motif Williams & Hall have shown by computational studies that the substitution of a G:C for a C:G closing base pair results in an increase in intrinsic loop flexibility [44]. Cabello-Villegas et al. [45] have recently revealed dynamical changes in the loop of a 17-mer stem-loop RNA corresponding to the anticodon arm of E. coli tRNA$^{Phe}$ introduced by a destabilising dimethylallyl modification. In order to gain insight into dynamical properties of the lead-dependent ribozyme, Hoogstraten et al. have performed rotating frame $^{13}$C relaxation studies [46]. These revealed a variety of motional processes in the active site and the GAAA tetraloop. The broad range of dynamic heterogeneity in the hexanucleotide loop of the HIV-2 TAR RNA was studied by $^{15}$N,$^{13}$C NMR relaxation experiments [47]. For a related 27-mer HIV-1 TAR RNA molecule, consisting of two helical stems separated by a short bulge, information from RDC measurements revealed the different amplitudes and directions of interhelical motions [48].

The HNN-COSY and HNCO experiments already mentioned, provide via the observation of a magnetisation transfer via $^{2}hJ_{NN}$ and $^{3}hJ_{NC}$ coupling constants direct physical evidence for the inclusion of additional hydrogen-bond distance constraints into the structure calculations. However, they also allow for a fast screening of RNA folds and a comparison with predictions from various programs (e.g. mfold [49]; Vienna RNA package [50]) without undertaking the tedious resonance assignment process. The combination of these experiments with a new RDC- and thus structure-based assignment method [51] might represent a new and faster approach to the structure determination of nucleic acids.
5. Recent RNA solution structures

Beside the 3D structures already mentioned in the technical sections, NMR spectroscopy was employed to elucidate the conformations of other RNA molecules. The structure of a 21-mer hexaloop RNA derived from the P5.1 hairpin of *Bacillus* RNase P was studied by heteronuclear NMR and refined with RDCs [32]. The 24 selected solution structures show an improvement of the heavy atom r.m.s. deviation from 1.48 Å to 1.16 Å when including RDC information. The determined hexaloop structure did not show an expected similarity to the GNRA tetraloop motif [52]. For a 29-mer RNA interacting as ribonucleic antiterminator with the bacterial regulatory protein LicT, hydrophobic and stacking interactions were identified to establish the interaction in the distorted minor groove hairpin stem with the protein [53]. While in this case the bound protein did not undergo important conformational changes, for the interaction of the ribosomal protein L25 the formation of an α-helical segment was observed upon binding to its cognate E-domain of 5S rRNA [54,55]. A similar observation was made for loops in the yeast ribosomal protein L30 upon RNA complex formation [56]. A selection of further RNA–protein complexes and of the use of RNA–protein complexes as targets for therapeutic intervention have been reviewed elsewhere [57,58].

A number of RNA structures involving tetraloops, structural motifs abundant in many RNA stem-loops and often connected with important biological functions, were recently studied by NMR spectroscopy: The UUUC tetraloop of the 24-mer mammalian histone mRNA stem-loop, stabilized by stacking of the first and third uridines of the loop and by a U:A closing base pair, was found to be an essential motif for the regulation of the histone mRNA metabolism through the interaction with its cognate hairpin-binding protein [59]. Another study on the 26-nt sequence of the same system reported the nucleotides at the base of the stem not to display a defined conformation [60]. The internal ribosome entry site (IRES) from Hepatitis C virus, a 30-mer RNA, was shown by NMR to accomodate an internal tetraloop adjacent to a mismatched double-helical stem crucial for IRES-mediated translation [61]. A novel family of tetraloops with consensus sequence (U/A)GNN was discovered in the recognition site for *Saccharomyces cerevisiae* RNase III. The two NMR structures with the sequences AGAA and AGUU solved revealed that the backbone turn originates at a syn G in the loop which is essential for RNase III binding [62].

In the solution structure of a 24-mer from the U6 intramolecular stem-loop, a highly conserved key component of the spliceosome’s catalytic core beside a pentaloop, an unpaired uridine in the stem mismatch was observed which is essential for metal ion binding via its phosphate oxygen. This base is not bulged-out and exhibits stacking interactions to a C:A wobble base pair the protonation state of which was followed by changes in the $^{13}$C shift of the adenosine C2 resonance [63]. A base-triple between a G:C base pair and the 3′-terminal guanosine was observed by a NMR study of a 22-nt model RNA including the guanosine-binding site of the *Tetrahymena* group I intron [64]. The recognition is stabilized by stacking interactions to the adjacent cytosine and the resulting bulge-and-twist structure motif on the major groove site introduces a kink between the helix stems.

6. Outlook

NMR spectroscopy has proven to be an important tool for the structure determination of RNA and RNA–protein complexes. New NMR technological developments increase the accessible molecular size and will allow for a quick verification of RNA structure predictions by observation of the H-bond pattern and thereby efficiently contribute to an increase in the knowledge of RNA structure-activity relationships.
Furthermore, with the availability of more RNA structures and their assignments in connection with a complete description of the conformational space [65] it can be expected that a chemical shift–structure relationship will become applicable similar to the chemical shift analysis in proteins [66].

References


[15] K. Pervushin, R. Riek, G. Wider and K. Wuthrich, Attenuated T2 relaxation by mutual cancellation of dipole-dipole interaction will become applicable similar to the chemical shift analysis in proteins [66].


Submit your manuscripts at
http://www.hindawi.com