JOURNAL OF MAGNETIC RESONANCE **124,** 210–213 (1997)
ARTICLE NO. MN967490

Identification of Ribose-Base Sequential NOEs According to Base Types in Uniformly ^{13}C-Labeled RNAs

RAMADURAI RAMACHANDRAN, CHRISTIAN SICH, OLIVER OHLENSCHLÄGER, AND LARRY R. BROWN*

Institut für Molekulare Biotechnologie e.V., Abteilung für Molekulare Biophysik/NMR Spektroskopie, Postfach 100 813, D-07708 Jena, Germany

Received August 5, 1996; revised September 23, 1996

Recent progress in molecular biology has facilitated the production of ^{13}C/^{15}N-labeled RNAs (*1–3*), and the availability of such labeled molecules has permitted the development of novel NMR schemes for sequential resonance assignments based on through-bond coherence transfer along the phosphate-ribose backbone (*4–10*). Although these methods are contributing to the determination of detailed NMR solution structures for RNA molecules of up to about 30–35 nucleotides, because of limited spectral dispersion and broad linewidths of the phosphorus resonances, such methods may have limited utility in larger RNA molecules as a result of the long interpulse delays employed in such experiments. In these circumstances, it may be necessary to achieve sequential resonance assignments via NOE effects. A widely used approach uses the fact that in helical RNA structures, ribose H1′ resonances give NOE cross peaks to their own H6/H8 resonance and to that of the next base in the sequence (*11, 12*). Unambiguous assignment of these sequential connectivities to the base type can facilitate sequence specific resonance assignments in helical regions of RNAs.

A potential problem is that NMR spectroscopy of RNA molecules is characterized by limited chemical-shift dispersion of resonances, even in ^{13}C, ^{15}N-labeled molecules, and as NMR studies proceed to larger RNA molecules, problems of spectral overlap are likely to become increasingly acute. This has led to the use of selective deuteration (*13*) and selective ^{13}C-labeling procedures (*14*) to resolve spectral overlaps that are encountered in even moderate-sized RNA molecules and which obscure sequential NOE patterns. Such approaches involve very substantial efforts to prepare nucleotides selectively labeled with isotopes and to subsequently prepare several oligonucleotide samples.

In uniformly ^{13}C-labeled RNA molecules, some distinction between purines and pyrimidines can be achieved via the HMQC–NOESY experiment since the C8 peaks normally tend to resonate below 140 ppm and C6 peaks above 140 ppm. However, assignments based on ^{13}C C8/C6 chemical shifts are not always unambiguous. In principle, the observation of the one-bond carbon–carbon couplings in the carbon dimension for pyrimidines can also be used for distinguishing pyrimidines from purines, provided that longer than usual ^{13}C acquisition times are employed to resolve the ^{13}C–^{13}C couplings and that the resolution between different purines/pyrimidines is adequate. Under the typical ^{13}C acquisition times normally employed in the HMQC–NOESY experiment, it may be difficult to resolve such carbon–carbon couplings and hence to achieve purine/pyrimidine distinction. Here we propose an alternative approach for spectral separation and unambiguous assignment of the sequential NOEs which arise from purines or pyrimidines in helical RNAs or in helical regions of larger RNA molecules. The method is based on subspectral editing of the NOESY data of a uniformly ^{13}C-enriched RNA via the application of homonuclear ^{13}C half-filters. The efficacy of the method for sequence-specific resonance assignments is illustrated with a 0.5 m*M* sample of the 19-mer hairpin ribonucleotide shown in the insert in Fig. 2B.

Figure 1 shows the pulse sequence that has been employed in this work. The pulse scheme utilizes a selective INEPT step to transfer the H6/H8 proton coherences to the attached C6/C8 carbons. The antiphase carbon coherences $I^{\pm}_{C6}I^{z}_{H6}$ and $I^{\pm}_{C8}I^{z}_{H8}$ present at the end of the selective INEPT step are allowed to evolve for a period Δ_2 which is set to $1/(2J_{C6C5})$, where $^1J_{C6C5}$ is about 66 Hz, for maximal creation of the term $I^{\pm}_{C6}I^{z}_{C5}I^{z}_{H6}$ in the case of the pyrimidines. After the period Δ_2, a 180° carbon pulse is applied as follows: for data set 1, a 180° carbon pulse acting only on the C6 and C8 spins is applied, while for data set 2, a 180° carbon pulse acting on all the C6, C8, and C5 spins is applied. The ^{13}C chemical-shift evolution and the antiphase $I^{\pm}_{C6}I^{z}_{C5}I^{z}_{H6}$ term created during the first Δ_2 period is then refocused during the second Δ_2 period. For the two data sets, the phase of the C8 carbon magnetization at the end of the period $2\Delta_2$ remains the same while the phase for the C6 carbons will be 180° out of phase. In-phase H6/H8 coherences are then created by the selective reverse INEPT procedure and subjected to the usual evolution and NOE mixing steps. Addition and subtraction of the two data sets give NOE subspectra that contain NOE cross peaks originating only from purine H8 or pyrimidine H6 resonances respectively.

1090-7807/97 $25.00

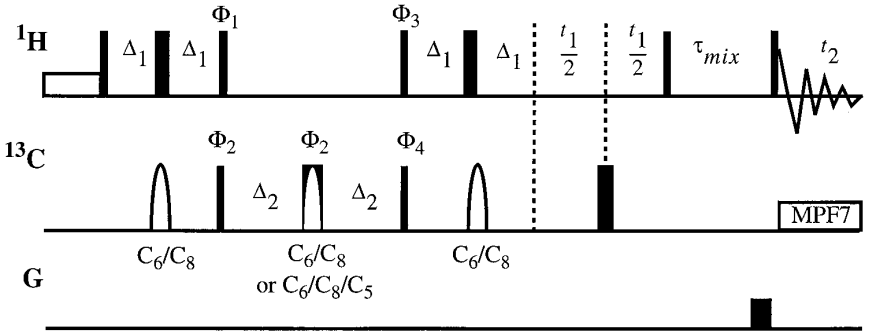

FIG. 1. Two-dimensional pulse sequence for identification of sequential ribose-base NOEs according to base type. The narrow and wide bars represent 90° and 180° pulses, respectively. Selective 180° pulses (G3 Gaussian cascade) with a length of 615 μs are marked with the carbons to which they are applied. The delays are set to $\Delta_1 = 1.4$ ms and $\Delta_2 = 7.5$ ms. The phases were cycled in the following way: $\Phi_1 = (y, y, -y, -y)$, $\Phi_2 = (x, -x)$, $\Phi_3 = 4(x), 4(-x)$, $\Phi_4 = 8(x), 8(-x)$, and the receiver phase $= (x, -x, -x, x, -x, x, x, -x, -x, x, x, -x, x, -x, -x, x)$. Φ_3 is cycled additionally according to the States procedure to obtain quadrature detection in t_1. Two experiments are acquired, one with a selective pulse on C6/C8 during the filter delay $2\Delta_2$ and the other experiment with a hard pulse exciting C6/C8/C5. Adding/subtracting the data sets yields purine/pyrimidine data sets, respectively. A gradient was applied at the end of the mixing period of $\tau_{mix} = 400$ ms in order to select the I^z states. ^{13}C broadband decoupling was applied during acquisition. The proton carrier was on the residual HDO during presaturation of the residual water. Before the first 90° proton pulse, the carrier was switched to 7.5 ppm and back to HDO during the mixing time. The carbon carrier was kept at 139 ppm at the beginning of the experiment and switched to 106 ppm for the carbon decoupling. For the hard pulse during the filtering period, the carrier was switched to 119 ppm before the first carbon 90° pulse and switched back to 139 ppm after the second 90° pulse.

In situations where improved spectral resolution and C6 and C8 chemical-shift information are required, separate 3D spectral data sets for purines and pyrimidines can be generated by allowing ^{13}C chemical-shift evolution after the $2\Delta_2$ period and before the reverse INEPT step. It is very advantageous that, due to the selective INEPT steps employed in the pulse sequence, small spectral widths can be used in both the t_1 (^{13}C) dimension and the t_2 (^{1}H) dimension of the 3D experiment. As it has been reported that the J(C8, C4) and J(C8, C5) in guanosines can be in the range of 8 to 11 Hz, minor subspectral cross talk can arise from these couplings. Compared to a conventional unfiltered, ^{13}C-edited, NOE experiment, one has in the present experiment additional losses of $[1 - \exp(-2\Delta_2/T_{2,C})]$. In our case, with carbon linewidths of 7–10 Hz, this leads to 30–40% in signal loss. Such losses in sensitivity due to fast transverse relaxation are expected to be much more severe in the experiments for sequential assignments based on backbone scalar couplings as one is dealing with small couplings and therefore long interpulse delays in such methods. Hence, it is expected that the present approach will find its utility in the study of larger RNA molecules, where other procedures may fail.

Two-dimensional edited spectra obtained via the procedure outlined in this Note are presented in Fig. 2. Starting from the cross peak marked U_{11} in Fig. 2B (the pyrimidine subspectrum), a sequence of NOE connectivities of the type (YYPYPYPY) is observed. Similarly, starting from the cross peak marked as G_1 in Fig. 2A (the purine subspectrum), a sequence of NOE connectivities of the type (PPYPYPY) is observed, where P and Y represent the purine and pyrimidine base types, respectively. From the RNA sequence (Fig. 2B), the (PPYPYPY) sequential NOE connectivities can be mapped uniquely to the corresponding fragment (G_1–C_7). The (YYPYPYPY) sequential NOE connectivities may correspond to (U_{11}–C_{18}), (C_{19}–C_{12}), or (U_{10}–C_3), since the NOE connectivities do not *a priori* indicate the direction of the oligonucleotide chain. Since (G_1–G_7) has been uniquely assigned, the overlapping segment (U_{10}–C_3) can be excluded. Similarly, the H_1' of the C_{19} residue at the 3′-terminus is not expected to show two sequential NOEs and can also be excluded on this basis.

To verify the assignment of (U_{11}–C_{18}), a few assignments of purines to G or A or a few assignments of pyrimidines to C or U are needed. For the 19-mer stem-loop, H5 frequencies of U and C were distinguished, using the characteristic C5 carbon chemical shifts (*11*) in a ^{13}C–^{1}H correlation spectrum. From the H5 frequencies of the strong H6–H5 NOE cross peaks seen in the pyrimidine subspectrum shown in Fig. 2B, some of the pyrimidine NOEs could be assigned to C or U on the basis of well-resolved H5 frequencies. It was then possible to identify the sequential connectivity (UCPYPCPY) which can only correspond to the fragment U_{11}–C_{18}. This assignment could be confirmed by distinguishing the C8, H8 resonances of G from A by using HCCH-TOCSY transfer to the C2, H2 resonances of A (*15, 16*). It is worth noting at this point that even for the present 19-mer, there exists both a purine and a pyrimidine NOE cross peak which have exactly the same chemical shifts (positions marked with asterisks in the spectra of Figs. 2A and 2B). The ability to resolve such overlapping peaks in the NOE data with the present experiment considerably facilitates the observation of a long stretch of sequential connectivities involving known base types.

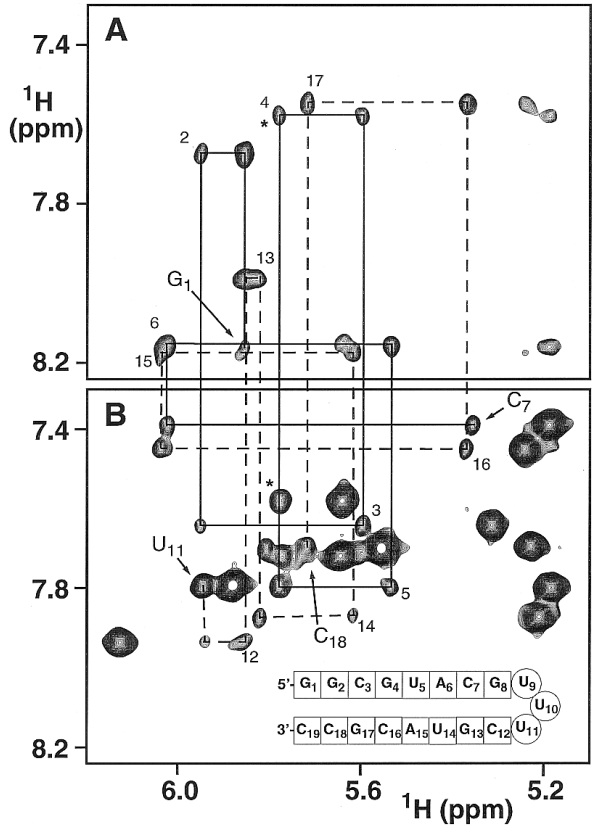

FIG. 2. Subspectra generated via the pulse sequence shown in Fig. 1 for a 0.5 m*M* solution of the $^{13}C/^{15}N$-labeled 19-mer ribonucleotide shown in the inset of (B). (A) and (B) show only NOEs starting at purine H8 and pyrimidine H6 protons, respectively. Sequential walks for both strands of the stem region using purine and pyrimidine base type information are shown. The intranucleotide cross peaks are labeled by their number in the sequence except for C18 for which the H8(C18) to H1′ (G17) cross peak is labeled. Spectral widths employed in t_1 and t_2 are 750 and 4500 Hz, respectively, with 36 and 512 complex points in ω_1 and ω_2. A total of 1024 transients with a recycle time of 1 s were collected per increment.

At the present stage of development, a variety of NMR experiments utilize magnetization transfers involving relatively small scalar couplings to provide effective ways to assign the spectra of $^{13}C/^{15}N$-labeled RNA molecules of up to about 30–35 nucleotides. To study RNA tertiary structure, it will be desirable to determine structures for still larger RNA molecules. Since it completely avoids the use of small scalar couplings and allows spectral simplification via editing, we anticipate that a major application of the present experiment will be in the identification and assignment of helical regions in larger RNA molecules. Identification of base types at the purine/pyrimidine level over long stretches of sequential connectivities can considerably limit the number of possible assignments in the RNA sequence space. The number of possible assignments will, of course, depend on the RNA sequence and the length of the sequential connectivities which can be identified.

Figure 3 shows for several RNA molecules the percentage of nucleotides in the full sequence for which unique sequential assignments are found as a function of the length of the stretches of NOE connectivities identified. For discrimination purely of purine and pyrimidine base types, unique sequential assignments are not always obtained, even with quite long stretches of NOE connectivities and quite small RNA molecules (see 19-mer in Fig. 3A). On the other hand, if complete assignment of the NOEs to G, A, C, and U is possible, then with relatively short stretches of NOE connectivities, unique sequential assignments can be obtained even for sizeable RNA molecules (Fig. 3B). In practice, as with the present experimental results on the 19-mer, it will be relatively easy to assign some of the purine NOEs to G or A and some of the pyrimidine NOEs to C or U; i.e., experimental reality will lie between the extremes represented by Figs. 3A and 3B. Since helical regions in large RNA molecules are on average only 6–8 base pairs in length, the data

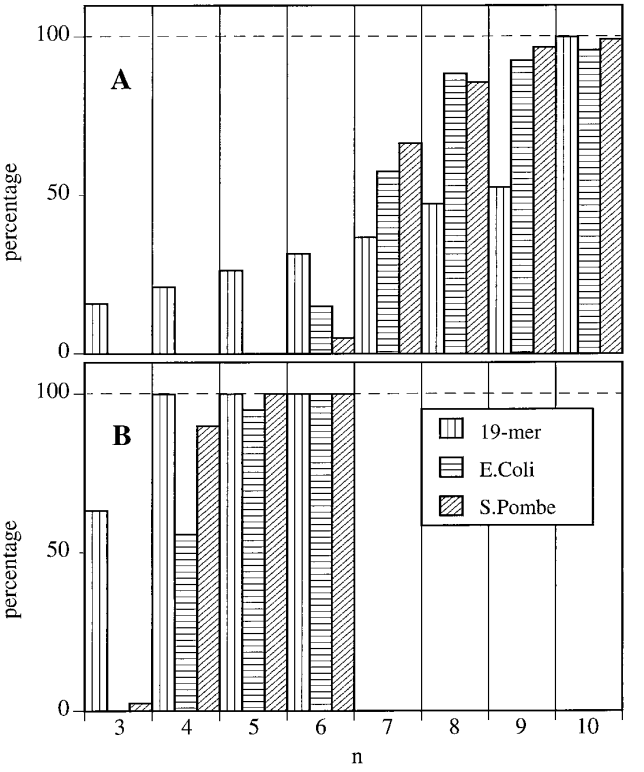

FIG. 3. Histograms showing the percentage of nucleotides in the 19-mer UUU tri-loop, 5S rRNA of *E. coli*, and 5S rRNA of *S. pombe* for which unique sequential assignments are found as a function of the length, n, of the stretch of NOE connectivities identified. (A) Distinction of purine and pyrimidine base types. (B) Distinction of A, G, C, and U base types. For (B), full assignments are obtained with $n = 6$, and the data for $n = 7–10$ are hence omitted. The complete RNA sequence of length N was considered to consist of $N - n + 1$ overlapping sequences of length n. All nucleotides which were contained in a unique n-mer were considered to be uniquely assigned to a specific sequence location.

in Fig. 3 indicate that assignments for such helical regions can be obtained with experiments of the type outlined here.

In conclusion, we have presented an approach for identification of base types in the sequential base-ribose NOE cross peaks in uniformly ^{13}C-enriched RNAs. Where overlap of purine and pyrimidine NOEs occurs, the present experiment is capable of separating the cross peaks. With a few specific identifications of the bases as (G/A) or (C/U) using data from other common NMR experiments, the stretches of sequential NOE connectivities should be uniquely mappable onto the RNA sequence space in helical RNAs or in helical regions of larger RNA molecules.

ACKNOWLEDGMENTS

We thank H.-J. Leppert and M. Semm for technical support. This work was supported by the Fonds der Chemischen Industrie e.V. and the Deutsche Forschungsgemeinschaft.

REFERENCES

1. J. R. Wyatt, M. Chastain, and J. D. Puglisi, *Biotechniques* **11,** 764 (1991).
2. R. T. Batey, M. Inada, E. Kujawinski, J. D. Puglisi, and J. R. Williamson, *Nucl. Acids Res.* **20,** 4515 (1992).
3. E. P. Nikonowicz, A. Sirr, P. Legault, F. M. Jucker, L. M. Baer, and A. Pardi, *Nucl. Acids Res.* **20,** 4507 (1992).
4. H. A. Heus, S. S. Wijmenga, F. J. M. van de Ven, and C. W. Hilbers, *J. Am. Chem. Soc.* **116,** 4983 (1994).
5. J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, and C. Griesinger, *J. Am. Chem. Soc.* **116,** 6472 (1994).
6. S. S. Wijmenga, H. A. Heus, H. A. E. Leeuw, H. Hoppe, M. van der Graaf, and C. W. Hilbers, *J. Biomol. NMR* **5,** 82 (1995).
7. J. P. Marino, H. Schwalbe, C. Anklin, W. Bermel, D. M. Crothers, and C. Griesinger, *J. Biomol. NMR* **5,** 87 (1995).
8. G. Varani, F. Aboul-ela, F. Allain, and C. C. Gubser, *J. Biomol. NMR* **5,** 315 (1995).
9. S. Tate, A. Ono, and M. Kainosho, *J. Magn. Reson. B* **106,** 89 (1995).
10. R. Ramachandran, C. Sich, M. Grüne, V. Soskic, and L. R. Brown, *J. Biomol. NMR* **7,** 251 (1996).
11. G. Varani, and I. Tinoco Jr., *Q. Rev. Biophys.* **24,** 479 (1991).
12. P. B. Moore, *Acc. Chem. Res.* **28,** 251 (1995).
13. C. K. Brush, M. P. Stone, and T. M. Harris, *Biochemistry* **27,** 115 (1988).
14. J. SantaLucia Jr., L. X. Shen, Z. Cai, H. Lewis, and I. Tinoco, Jr., *Nucl. Acid Res.* **23,** 4913 (1995).
15. P. Legault, B. T. Farmer, L. Mueller, and A. Pardi, *J. Am. Chem. Soc.* **116,** 2203 (1994).
16. J. P. Marino, J. H. Prestegard, and D. M. Crothers, *J. Am. Chem. Soc.* **116,** 2205 (1994).