Carbon-Proton Scalar Couplings in RNA: 3D Heteronuclear and 2D Isotope-Edited NMR of a ^{13}C-Labeled Extra-Stable Hairpin†

Jennifer V. Hines, Stacy M. Landry, Gabriele Varani,‡ and Ignacio Tinoco, Jr.*

Contribution from the Department of Chemistry, University of California and Division of Structural Biology, Lawrence Berkeley Laboratory, Berkeley, California 94720

Received December 9, 1993®

Abstract: Long range carbon-proton scalar couplings were measured for an RNA hairpin of 12 nucleotides using 3D and ^{13}C-edited 2D NMR. The large one-bond carbon-proton scalar couplings ($^1J_{CH}$) and small *n*-bond couplings ($^nJ_{CH}$) produce ECOSY type cross-peaks, thus facilitating the determination of the sign and magnitude of the smaller $^2J_{CH}$ or $^3J_{CH}$. The UUCG RNA hairpin (5′-rGGACUUCGGUCC-3′), whose structure has been determined by our laboratory, was uniformly ^{13}C-labeled at 30% isotopic enrichment. The observed $^nJ_{CH}$ couplings were then correlated to the known structure. The signs of $^2J_{C4'H5'}$, $^2J_{C4'H5''}$, and $^2J_{C5'H4'}$ can be used for the stereospecific assignment of H5′(pro-*S*)/H5″ (pro-*R*) protons and for determining torsion angle γ(O5′–C5′–C4′–C3′). The sign of any one of the $^2J_{C1'H2'}$, $^2J_{C2'H3'}$, $^2J_{C3'H2'}$, and $^2J_{C4'H3'}$ can be used to determine the ribose sugar conformation. Comparison of the magnitude of $^3J_{C3'H5'}$ and $^3J_{C3'H5''}$ can also be used for the stereospecific assignment of H5′(pro-*S*)/H5″(pro-*R*). The utility and potential limitations of using $^nJ_{CH}$ for RNA structure determination are discussed in-depth. This new information which correlates signs and magnitudes of $^nJ_{CH}$ both to double-stranded A-form and nonhelix RNA structure forms a basis from which strategies for the structure determination of larger RNA and RNA–protein complexes can be derived.

Introduction

Knowledge of the importance of RNA in biological processes has been increasing rapidly; however, the relation between structure and function in RNA needs to be understood more thoroughly.[1] Current two-dimensional NMR methods for determining RNA structure are useful for small molecular weight molecules (<30 nucleotides or 10 000 Da) but become severely limited by spectral overlap and line width with larger molecules.[2] Alternative methods for obtaining structural information which are not limited by these problems are needed. The method of choice for reducing spectral overlap is heteronuclear multidimensional (3D or 4D) NMR.[3] To overcome the interference of line width in the measurement of coupling constants, ECOSY type cross-peaks can be used.[4] Two-dimensional experiments with ECOSY type cross-peaks have been used to determine some heteronuclear couplings in a B-form DNA duplex.[5] Three-dimensional NMR with ECOSY type cross-peaks has been employed in protein studies to determine small heteronuclear couplings[6] but has not been applied to RNA structural studies.

This paper discusses the potential utility and limitations of using carbon-proton long range scalar coupling constants for the structure determination of RNA molecules. These $^nJ_{CH}$ were obtained using 3D NMR with ECOSY type cross-peaks; thus the structure determination methodology could be extended to larger molecules. Carbon-proton coupling constants have been useful in assessing sugar conformation in carbohydrates and mononucleotides.[7,8] In a preliminary report we briefly discussed use of $^2J_{CH}$ for the stereospecific assignment of the H5′(pro-*S*) and H5″(pro-*R*) protons in RNA as well as the determination of torsion angle γ(O5′–C5′–C4′–C3′) and the pseudorotation phase angle P. The full results of this study, including attempts to correlate $^3J_{CH}$ with χ(O4′–C1′–N1–C2 for pyrimidines, O4′–C1′–N–C4 for purines) and $^3J_{CH}$ with stereospecific assignments, are reported here. Only through the careful analysis of $^nJ_{CH}$ in many well-characterized RNA molecules will it be possible to form definitive correlations between $^nJ_{CH}$ values and structure. This study begins to build the basis set of information which can be used for the structure determination of larger RNA and protein/RNA complexes.

The molecule we chose to study was the 12 nucleotide, extrastable, UUCG RNA hairpin (Figure 1) whose structure has been thoroughly characterized with 2D NMR experiments.[10] Uniformly ^{13}C-labeled UUCG RNA was prepared at 30% isotopic

‡ Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, England.

† Abbreviations: NMR, nuclear magnetic resonance, ppm, parts per million, 2D, two-dimensional; 3D, three-dimensional; NOE, nuclear Overhauser effect; NOESY, nuclear Overhauser effect spectroscopy; ECOSY, exclusive correlation spectroscopy; TOCSY, total correlation spectroscopy; HMQC, heteronuclear multiple quantum correlation spectroscopy; HSQC, heteronuclear single quantum correlation spectroscopy; TPPI, time proportional phase incrementation; TSP, (trimethylsilyl)propionic acid; RNA ribonucleic acid; UV, ultraviolet; OD, optical density, rNTP, ribonucleoside triphosphate; rNMP, ribonucleoside monophosphate; PEP, phosphoenolpyruvate (trisodium salt hydrate); 3-PGA, 3-phosphoglyceric acid (barium salt); PK, pyruvate kinase; PM, phosphoglycerate mutase; MK, myokinase; GK, guanylate kinase; NMPK, nucleoside monophosphate kinase; DTT, dithiothreitol; SDS, sodium dodecyl sulfate; rNMP, ribonucleotide monophosphate; dNMP, deoxyribonucleotide monophosphate; DW, dwell time; A_{260}, A_{280}, absorption at 260 and 280 nm, respectively.

® Abstract published in *Advance ACS Abstracts*, May 15, 1994.

(1) Wyatt, J. R.; Tinoco, I., Jr. *RNA Structural Elements and RNA Function in The RNA World*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, 1993; pp 465–496.

(2) Varani, G.; Tinoco, I., Jr. *Quart. Rev. Biophys.* **1991**, *24*, 479–532.

(3) (a) Nikonowicz, E. P.; Pardi, A. *Nature* **1992**, *355*, 184–186. (b) Nikonowicz, E. P.; Pardi, A. *J. Am. Chem. Soc.* **1992**, *114*, 1082–1083.

(4) Griesinger, C.; Sorensen, O. W.; Ernst, R. R. *J. Magn.Reson.* **1987**, *75*, 474–492.

(5) Schmeider, P.; Ippel, J. H.; van den Elst, H.; van der Marel, G. A.; van Boom, J. H.; Altona, C.; Kessler, H. *Nucl. Acids Res.* **1992**, *20*, 4747–4751.

(6) (a) Wider, G.; Neri, D.; Otting, G.; Wüthrich, K. *J. Magn. Reson.* **1989**, *85*, 426–431. (b) Edison, A. S.; Westler, W. M.; Markley, J. L. *J. Magn. Reson.* **1991**, *92*, 434–438.

(7) (a) Kline, P. C.; Serianni, A. S. *J. Am. Chem. Soc.* **1990**, *112*, 7373–7381. (b) Kline, P. C.; Serianni, A. S. *Mag. Res. Chem.* **1988**, *26*, 120–123. (c) Serianni, A. S.; Barker, R. *J. Org. Chem.* **1984**, *49*, 3292–3300. (d) Wu, J.; Serianni, A. S. *Carbohyd. Res.* **1992**, *226*, 209–219. (e) Bandyopadhyay, T.; Wu, J.; Serianni, A. S. *J. Org. Chem.* **1993**, *58*, 5513–5517. (f) Hayes, M. L.; Serianni, A. S.; Barker, R. *Carb. Res.* **1982**, *100*, 87–101.

(8) (a) Block, K.; Pedersen, C. *Acta Chem. Scand.* **1977**, *B31*, 354–358. (b) Cyr, N.; Perlin, A. S. *Can. J. Chem.* **1979**, *57*, 2504–2511. (c) Schwartz, J. A.; Cyr, N.; Perlin, A. S. *Can. J. Chem.* **1975**, *53*, 1872–1875.

(9) Hines, J. V.; Varani, G.; Landry, S. M.; Tinoco, I., Jr. *J. Am. Chem. Soc.* **1993**, *115*, 11002–11003.

(10) (a) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680–682. (b) Varani, G.; Tinoco, I., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 9349–9354. (c) Varani, G.; Cheong, C.; Tinoco, I., Jr. *Biochemistry* **1991**, *30*, 3280–3289.

0002-7863/94/1516-5823$04.50/0

enrichment. The method for preparing the ^{13}C-labeled rNTPs used in the UUCG synthesis is slightly different from those published previously.[11]

The choice of the rigid UUCG RNA hairpin as a model system has the added advantage that for most of the molecule (except G1 and G12) the observed couplings will not be due to an average of very different conformations as found in mononucleosides. In addition, the hairpin contains both A-form helix and nonhelix structural elements. This supplies internal controls for any correlation between heteronuclear coupling values and structure.

Results

A. ^{13}C RNA Isolation. The isolation of isotopically labeled RNA from bacterial sources has been described by several laboratories,[11] each using slightly different methods. In bacteria, the majority of the cellular RNA is ribosomal RNA (80% for *E. coli*), and the total amount of modified bases is very small (<1% of total RNA is modified tRNA bases).[12] We thus chose to isolate total cellular RNA. As has been noted previously,[11a] methylotrophic bacteria provide a less expensive route to incorporation of ^{13}C label into RNA since ^{13}C methanol is cheaper than glucose (the preferred carbon source for *E. coli*). The cells used in our study were *Methylobacterium extorquens* strain AM1 uniformly labeled with ^{13}C at 30% isotopic enrichment. The cells were supplied by the Stable Isotope Resource at Los Alamos National Laboratory. While it is possible for these bacteria to utilize CO_2 as a carbon source, no uptake of atmospheric CO_2 was detected.[13] This result was confirmed by integration of the NMR spectrum of the rNMPs, which indicated uniform labeling with ^{13}C at 30% ± 2% isotopic enrichment.

Under the conditions used to grow the *Methylobacterium extorquens* for this study (0.5% methanol w/v), the doubling time is approximately 8 h. However, when the cells were grown on only 0.25% w/v methanol, the doubling time was 14 h.[13] The cells with a doubling time of 8 h gave approximately 11 mg of rNMPs (NH_4^+ salt) per gram of wet cell based on a hot phenol/SDS, pH 5.5 extraction method of Oelmuller et al.[14] This is comparable to the yield reported by Batey et al.[11a] for *Methylophilus methylotrophus*, 10 mg of rNMPs (Na^+ salt) per gram wet cell. However, the yield for the cells with a slower doubling time of 14 h was less than 1 mg of rNMPs (NH_4^+ salt) per gram of wet cell. This is not surprising since the amount of ribosomal RNA is greatest when a large amount of protein synthesis is occurring, as is the case in the log phase of a rapidly doubling bacterial culture.[15]

One advantage of the extraction method we chose is that the low pH of the initial extraction media causes most of the DNA to precipitate with the proteins at the phenol/aqueous interface. After all of the RNA has been extracted, the DNA can be isolated simply by extracting the protein/DNA/phenol mixture with a high pH buffer.[16] This allows the isolation of both DNA and RNA without requiring a boronate affinity column for the separation of the rNMPs from the dNMPs.[11a] Based on the A_{260}/A_{280} ratio, little (<5%) or no DNA was contaminating the RNA at this stage of the extraction procedure. Following the hot phenol extraction, the crude RNA is treated with RNase-free DNAse I (to remove the last traces of DNA), ethanol precipitated, and then treated with nuclease P1. The rNMPs can then be phosphorylated directly using a slight modification of the procedure reported by Haynie and Whitesides[17] or first separated by ion exchange chromatography and then phosphorylated by the method of Simon et al.[18] In either case, the rNTPs were free of dNTPs based on NMR and UV spectra.

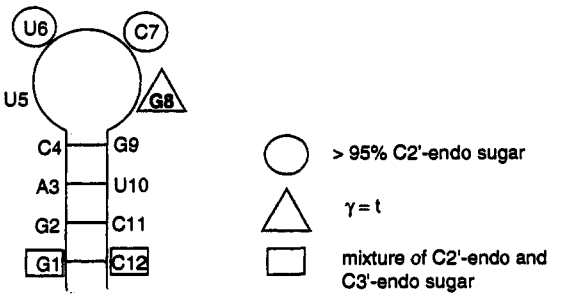

Figure 1. Structural elements of the UUCG RNA hairpin as previously determined by Varani et al.[10c] All bases have the standard A-form geometry of C3′-endo sugar pucker, *anti*-base and γ = g+ except as noted specifically. The boldface **G8** indicates a *syn*-base.

This isolation method also works with *Methylophilus methylotrophus*, but not when the cells are grown in the media described by Batey et al.[11a] When the cells were grown according to the protocol of Batey et al.[11a] and RNA was isolated with the hot phenol/SDS pH 5.5 method, the yield was less than 1 mg of rNMPs (NH_4^+ salt) per gram of wet cell. However, when the cells were grown in the media used for *Methylobacterium extorquens* strain AM1,[19] the yield after hot phenol/SDS pH 5.5 extraction was approximately 10 mg of rNMPs (NH_4^+ salt) per gram of wet cell. The only major difference between the two growth conditions is that iron is present in the media used by Batey et al.[11a] and not in the media reported[19] by Anthony. Residual iron from the growth media appears to be interfering with the extraction process.

B. Heteronuclear Scalar Coupling Measurements. The NMR experiments used to measure the long range carbon-proton scalar couplings were an $\omega 1$ 1/2X-filtered NOESY,[20] a 3D HMQC-TOCSY,[21] and a 3D HMQC-NOESY.[22] In the absence of ^{13}C decoupling, these experiments exhibit ECOSY type cross-peaks which are split by the large $^1J_{CH}$ in one proton dimension and by the smaller long range $^2J_{CH}$ or $^3J_{CH}$ in the directly observed proton dimension. The direction of offset along $\omega 3$ in the 3D experiments and $\omega 2$ in the 2D experiments indicates the relative sign of the coupling constants (Figure 2). The direction for a positive offset was determined by examining the $\omega 3\omega 2$ H3′H1′ cross-peak of U6 and C7, which are both split along $\omega 3$ by the positive $^3J_{C3'H1'}$. Additional splittings within the cross-peaks are due to passive ^1H–^1H and ^1H–^{31}P couplings.

The $^2J_{CH}$ and $^3J_{CH}$ measured from the 3D HMQC-TOCSY are summarized in Tables 1 and 3, respectively. The $^2J_{CH}$ measured from the 3D HMQC-NOESY are summarized in Table 2. The $^3J_{CH}$ for determining the torsion angle χ (O4′–C1′–N1–C2 for pyrimidines, O4′–C1′–N9–C4 for purines) are summarized in Table 5. For all of the heteronuclear scalar couplings, the cross-peak offset was measured at least three times, and the values were averaged. The errors reported in the tables reflect the variability in the multiple measurements. Table 4 contains predicted signs for $^2J_{CH}$ (vide infra).

(11) (a) Batey, R. T.; Inada, M.; Kujawinski, E.; Puglisi, J. D.; Williamson, J. R. *Nucleic Acids Res.* **1992**, *20*, 4515–4523. (b) Nikonowicz, E. P.; Sirr, A.; Legault, P.; Jucker, F. M.; Baer, L. M.; Pardi, A. *Nucleic Acids Res.* **1992**, *20*, 4507–4513. (c) Polson, A. G.; Crain, P. F.; Pomerantz, S. C.; McCloskey, J. A.; Bass, B. L. *Biochemistry* **1991**, *30*, 11507–11514. (d) Michnicka, M. J.; Harper, J. W.; King, G. C. *Biochemistry* **1993**, *32*, 395–400.

(12) Ingraham, J. L.; Maaloe, O.; Neidhardt, F. C. *Growth of the Bacterial Cell*; Sinauer Associates, Inc.: Sunderland, 1983; pp 2–14.

(13) Unkefer, C. Stable Isotopes Resource, Los Alamos National Laboratory, personal communication.

(14) Oelmuller, U.; Kruger, N.; Steinbuchel, A.; Friedrich, C. G. *J. Microbiol. Methods* **1990**, *11*, 73–84.

(15) Hou, C. T. *Methylotrophs: Microbiology, Biochemistry and Genetics*; CRC Press, Inc.: Boca Raton, FL, 1984; pp 123–125.

(16) Majumdar, D.; Avissar, Y. J.; Wyche, J. H. *BioTechniques* **1991**, *11*, 94–101.

(17) Haynie, S.; Whitesides, G. M. *Appl. Bioch. Biotech.* **1990**, *23*, 205–220.

(18) Simon, E. S.; Grabowski, S.; Whitesides, G. M. *J. Org. Chem.* **1990**, *55*, 1834–1841.

(19) Anthony, C. *Biochemistry of Methylotrophs*; Academic Press: New York, 1982; pp 167–186.

(20) Otting, G.; Wüthrich, K. *Quart. Rev. Biophys.* **1990**, *23*, 39–96.

(21) Wijmenga, S. S.; Hallenga, K.; Hilbers, C. W. *J. Magn. Reson.* **1989**, *84*, 634–642.

(22) Fesik, S. W.; Zuiderweg, E. R. P. *J. Magn. Reson.* **1988**, *78*, 588–593.

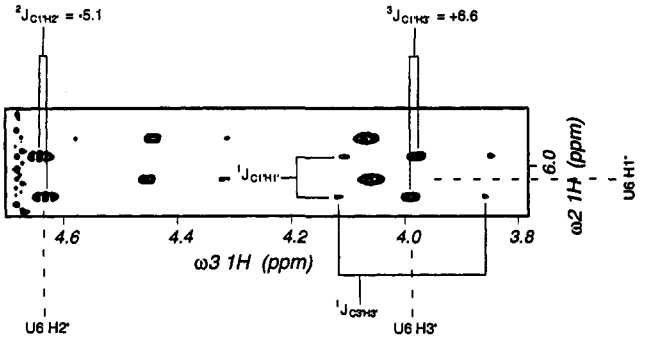

Figure 2. Portion of the 3D HMQC-TOCSY at the C1′ carbon plane (ω1 = 88.3 ppm) of U6 and C7. The labeled cross-peaks correspond to U6 ω3ω2 H2′H1′ and ω3ω2 H3′H1′. The strong unlabeled peaks are from C7. The intense peaks are due to the 70% protons bound to ^{12}C in ω3 for U6 and C7. The weaker peaks are due to the 30% protons bound to ^{13}C in ω3.

C. New Assignments. Due to overlap problems in the 2D experiments done previously on the UUCG hairpin, certain carbon resonances could not be assigned.[10b] While two possible values were reported for each, the C4′ carbon chemical shift of C4, G8, G9, U10, and C11 and the C3′ of G2 and A3 could not be specifically identified. With the additional carbon dimension in the 3D HMQC-TOCSY the following assignments can now be made: ^{13}C chemical shifts for G8 C4′ and G9 C4′ = 82.5 ppm; C4 C4′, U10 C4′, and C11 C4′ = 81.1 ppm; G2 C3′ = 72.4 ppm; A3 C3′ = 72.0 ppm.

Discussion

A. NMR Experiments. In the absence of ^{13}C decoupling, the 3D HMQC-TOCSY, 3D HMQC-NOESY, and the ω1 1/2X-filtered 2D NOESY all exhibit ECOSY type cross-peaks. Measurements based on ECOSY cross-peaks reduce the interference of line width on the measurement of the scalar coupling constant.[23] A similar approach has been used for the study of $^nJ_{NH}$ scalar couplings[6a] and $^nJ_{CH}$ scalar couplings[6b] in proteins.

The rationale for using 30% ^{13}C labeled RNA was twofold. The main reason for using 30% enrichment was to decrease the amount of ^{13}C–^{13}C scalar coupling due to neighboring labeled carbons and therefore decrease the apparent line width in the carbon dimension. In ribonucleosides, the $^1J_{CC}$ are approximately 40 Hz.[7a] Another reason is to reduce the complexity of the spectrum. In Figure 2, the major cross-peaks are due to the 70% ^1H bound to ^{12}C in ω3. The extra peaks along ω3 are due to the protons bound to the residual 30% ^{13}C and split by the large $^1J_{CH}$. These satellites cause minimal problems because they are very weak with only 30% enrichment; however, with 100% ^{13}C enrichment they could cause considerable overlap problems.

The disadvantage of using only 30% ^{13}C label is the obvious loss in sensitivity compared to 100% enrichment. Since the cross-peaks are split in an ECOSY experiment by the heteronuclear CH scalar couplings, sensitivity is reduced further. Consequently, only peaks with sufficiently strong TOCSY or NOESY transfer can be observed in the corresponding 3D or isotope-filtered experiments. In both the 1/2X-filtered NOESY and the HMQC-NOESY a long mixing time (500 ms) was needed in order to observe protons separated by 3 Å or more.

In the 3D HMQC-TOCSY experiments, there were severe overlap problems only when two peaks had similar proton chemical shifts and had carbon chemical shifts differing by 0.2 ppm or less. It was thus possible in the 3D HMQC-TOCSY (Table 1) to distinguish between the ω3ω2 H2′H3′ cross-peaks of G2 and A3 even though the proton chemical shifts are very similar and their carbon C3′ chemical shifts differ only by 0.4 ppm.[10b] There still

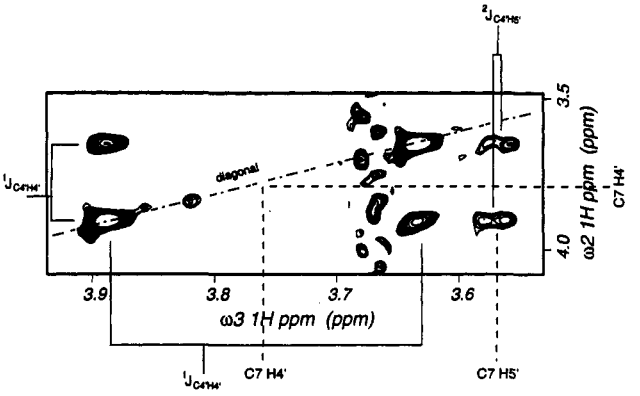

Figure 3. Portion of the 3D HMQC-NOESY (τ_m = 150 ms) at C4′ carbon plane of C7 (ω1 = 79.5 ppm) showing the strong diagonal peaks (ω3 = ω2 = C7 H4′).

was considerable overlap, but a carbon plane for each cross-peak could be found where there was no interference from the other cross-peak. However, the ω3ω2 H3′H2′ cross-peaks for G2 and A3 could not be distinguished because the C2′ chemical shifts differed by only 0.1 ppm.

The 2D and 3D NOESY experiments have several inherent problems that need to be addressed. For most of the protons separated by 3 Å or more, the cross-peaks were too weak to measure accurately in the 3D NOESY experiments. In addition, the diagonal in the 2D and 3D NOESY causes severe overlap problems. The diagonal is very intense and is split in both dimensions by $^1J_{CH}$ (Figure 3). There is also the potential that for a given cross-peak (e.g., ω3ω2 H3′H1′) the observed CH coupling along ω3 ($^3J_{C1'H3'}$) would be affected if part of the NOESY magnetization transfer occurred via spin-diffusion through another proton which was also bound to ^{13}C (e.g., H2′). In this case the offset along ω3 would be determined both by $^3J_{C1'H3'}$ and $^2J_{C1'H2'}$. With only 30% enrichment this likelihood is reduced but still possible.

It should be noted that the magnitude, and in some cases the sign, of the long range coupling for certain protons (e.g., C7 $^2J_{C1'H2'}$) changes with longer NOESY mixing times (τ_m = 500 ms, data not shown). One possible explanation is that the spin of the heteroatom changes during the mixing time due to T1 relaxation. This would result in the offset along ω3 having the opposite sign. Previous T1 studies on a DNA duplex of $(dC\text{-}dG)_3$ indicate that sugar carbon T1s are on the order of several hundred milliseconds,[24] and therefore comparable with the long (500 ms) mixing time. Further studies on selectively ^{13}C-labeled RNA are currently underway to investigate the T1s for the UUCG. T2 relaxation could also affect the observed couplings. The effect of T2 relaxation on the magnitude of homonuclear scalar couplings measured from ECOSY cross-peaks has been investigated.[25] Another possible explanation is that with the longer mixing times there may be more interference with the observed sign and magnitude due to spin-diffusion affects as discussed above. Only further investigations will determine which, if any, of these explanations is correct. Because of the above problems, data from the long mixing time 3D HMQC-NOESY (τ_m = 500 ms) were not used in this analysis.

The 3D HMQC-TOCSY experiment is, by far, the most sensitive method for measuring the long range CH scalar couplings. Since good coherence transfer can be achieved with a short spin lock, there are no difficulties with the spin of the carbon changing during the spin lock. However, the 3D HMQC-TOCSY is inherently limited by the need for a sufficiently strong $^3J_{HH}$ or $^2J_{HH}$ scalar coupling in order to have efficient coherence transfer. In the cases where a TOCSY coherence transfer is not

(23) Griesinger, C.; Otting, G.; Wurthrich, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1988**, *110*, 7870–7872.

(24) Borer, P. N.; Zanatta, N.; Holak, T. A.; Levy, G. C.; Van Boom, J. H.; Wang, A. H.-J. *J. Biomol. Struct. Dynamics.* **1984**, *1*, 1373–1386.

(25) Norwood, T. J. & Jones, K. *J. Magn. Reson.* **1993**, *104*, 106–110.

Table 1. $^2J_{CH}$ (Hz) from 3D HMQC-TOCSY[g]

	C1'H2'	C2'H1'	C2'H3'	C3'H2'	C3'H4'	C4'H3'	C4'H5' [c] (pro-S)	C4'H5'' [c] (pro-R)	C5'H4'
G1	+0.8	−2.1	+1.8	−0.9	*a*	−2.4	*a*	*a*	*a*
G2	*a*	*a*	G2=A3[e] (+2.6 ± 0.6)	−1.7	−6.2 ± 0.6	*b*	*a*	*a*	*a*
A3	*a*	*a*	A3=G2[e]	−1.0	*b*	*b*	*f*	*a*	*a*
C4	*a*	*a*	+3.0	−1.0	−5.3	*b*	*f*	*a*	*a*
U5	*a*	*a*	+2.6	−1.4	−2.4	*b*	*f*	*a*	*a*
U6	**−5.1**	−2.7	**−1.9**	**+2.1**	*f*	*a*	−5.3	+1.9	+3.8
C7	**−5.4**	−2.2	**−2.4**	**+2.5 ± 0.5**	*a*	*a*	−5.4[d]	+1.9[d] ± 0.5	+4.3
G8	*a*	*a*	+1.2	−1.2	G8H5'' [d]=H4' (−2.5)	−2.8	**+1.2**[d]	**−4.8**[d]	**−3.7**
G9	*a*	*a*	+2.4	*b*	*b*	*b*	*f*	*b*	+4.8
U10	*a*	*a*	*b*	*b*	*b*	*b*	*f*	*a*	*a*
C11	*a*	*a*	+2.6	−1.7	−6.6 ± 0.6	*b*	*f*	*a*	*a*
C12	*a*	−2.6	+1.7	−0.6	−5.3 ± 0.7	*b*	*f*	*a*	*a*

[a] ^{1}H–^{1}H TOCSY cross-peak is very weak or not observed. [b] Cross-peak is observed but is partially overlapped with diagonal, satellite, or another peak. [c] H5' is the more downfield proton unless otherwise specified. [d] H5'' is the more downfield proton. [e] C2' of G2 and A3 could not be distinguished; see text. [f] Too much overlap to tell if the peak is present. [g] The cross-peak offset was measured three times, and the values were averaged. When the cross-peak components were split due to passive coupling, the offset for each passive coupling sub-cross-peak was measured. When the cross-peak components were an apparent singlet, the offset of the components was measured for different carbon planes when possible. The indicated errors represent the variability in the measured values. All values are within ±0.4 Hz unless otherwise noted. Couplings in boldface are values uniquely different from couplings observed for standard A-form RNA.

Table 2. $^2J_{CH}$ (Hz) from 3D HMQC-NOESY[g]

	C1'H2'	C2'H1'	C2'H3'	C3'H2'	C3'H4'	C4'H3'	C4'H5' [c] (pro-S)	C4'H5'' [c] (pro-R)	C5'H4'
G1	+1.0	−2.0	*b*	−2.2	*a*	*f*	*b*	*b*	*f*
G2	+2.1	−1.7	G2=A3[e] (+3.5)	*b*	*a*	*f*	*b*	*b*	*b*
A3	+2.4	−2.1	A3=G2[e]	*b*	*a*	*f*	*b*	+0.5	*b*
C4	+0.9	−3.1	+4.1 ± 1	−1.9 ± 0.5	*a*	*a*	*b*	b	*b*
U5	+4.0	−2.2	+4.5 ± 1	−2.2	*a*	*f*	*b*	+0.5	+5.7 ± 1
U6	$(H_2O)^b$	*a*	**−1.1 ± 0.6**	**+3.1 ± 0.5**	−5.7	**+2.0**	−5.2	+2.2 ± 0.6	+4.3 ± 0.5
C7	**−3.0 ± 1**	*a*	**−2.5 ± 0.7**	**+3.7 ± 1**	−5.2	**+5.5 ± 0.5**	**−4.6**[d] ± 1	**+2.3**[d]	+4.9
G8	+1.6	−4.3	+3.4 ± 1	−1.1	*a*	*a*	*b*	*b*	*b*
G9	+3.5	*b*	*b*	*b*	*a*	*f*	*b*	*b*	*b*
U10	+1.2	−2.6	*b*	*b*	*a*	*f*	*b*	+1.3	*b*
C11	+3.6	−3.4 ± 0.5	*b*	−1.4	*a*	*f*	*b*	+0.3	*b*
C12	+0.6	−2.8	*b*	−2.1	*a*	*f*	*b*	b	*b*

[a] ^{1}H–^{1}H NOESY cross-peak is very weak or not observed. [b] Cross-peak is observed but is partially overlapped with diagonal, satellite, or another peak. [c] H5' is the more downfield proton unless otherwise specified. [d] H5'' is the more downfield proton. [e] C2' of G2 and A3 could not be distinguished; see text. [f] Too much overlap to tell if the peak is present. [g] The cross-peak offset was measured three times, and the values were averaged (see Table 1). The indicated errors represent the variability in the measured values. All values are within ±0.4 Hz unless otherwise noted. Couplings in boldface are values uniquely different from couplings observed for standard A-form RNA.

efficient, a short mixing time NOESY can fill in some of the gaps of information.

It should also be noted that where the TOCSY and NOESY (τ_m = 150 ms) experiments can be used to observe the same coupling (e.g., C2'H3' and C3'H2') there is good agreement. In only three cases, out of a total of 24 $^2J_{CH}$ which could be measured by both experiments, were the couplings not within reported experimental error. However, as was discussed earlier, the couplings measured from a NOESY with 500 ms mixing time were very different, and in some cases gave the opposite sign.

B. $^nJ_{CH}$ Couplings. Two-bond and three-bond carbon-proton scalar couplings in furanose ring systems[7c,8b] and ribonucleosides[7a,d] have been studied previously. Proton–phosphorus and carbon–phosphorus scalar couplings have also been reported for a B-form DNA duplex.[5] Carbon–phosphorus couplings in modified di- and triribonucleosides have been studied.[26] However, to date there has been no detailed study of proton-carbon scalar couplings in an RNA oligonucleotide. We have measured ^{13}C–^{1}H scalar coupling constants in the well-characterized UUCG RNA hairpin and correlated these values with the known structure. This information can be used to assign stereospecifically H5'(pro-*S*)/H5''(pro-*R*) and define relationships between $^nJ_{CH}$ and γ-(O5'–C5'–C4'–C3'), the pseudorotation phase angle[27] P and χ(O4'–C1'–N1–C2 for pyrimidines, O4'–C1'–N9–C4 for purines) in RNA and presumably DNA oligonucleotides.

$^2J_{CH}$ Couplings. The $^2J_{CH}$ obtained from the 3D HMQC-TOCSY and the 3D HMQC-NOESY (150 ms mixing time) are shown in Tables 1 and 2 respectively. Neither experiment provides all of the possible two-bond couplings due to the limitations described above. Together, however, they provide most of the desired couplings except in cases of severe spectral overlap.

The sign and magnitude of the two-bond carbon-proton scalar coupling in the system H–C–^{13}C–X depends on the electronegativity of X and the orientation between H and X.[8,28] The magnitude increases with increasing electronegativity of X. Calculations on ethanol[8c] have shown that the magnitude and sign of the coupling depends on the torsion angle (θ) between H and X. For $0° \le \theta < 90°$, $^2J_{CH}$ is negative; it is positive when $90° < \theta \le 180°$ with maximum absolute values at 0° and 180°. At $\theta = 90°$, $^2J_{CH} = 0$. In cases of multiple heteroatom substitution (e.g., H–C–^{13}C(X_1)X_2) a projection rule can be used to explain the sign of the coupling.[8a]

In all of the possible ribose $^2J_{CH}$, except for $^2J_{C1'H2'}$, the sign is relatively easy to predict for a specific ribose conformation

(26) Lankhorst, P. P.; Haasnoot, C. A. G.; Erkelens, C.; Altona, C. *J. Biomol. Struct. Dynamics* **1984**, *1*, 1387–1405.

(27) Altona, C. *Recl. Trav. Chim. Pays-Bas.* **1982**, *101*, 413–433.

(28) For a review of $^nJ_{CH}$ see: (a) Hansen, P. E. *Prog. NMR Spect.* **1981**, *14*, 175–296. (b) Marshall, J. L. *Carbon–Carbon and Carbon-Proton NMR Couplings*. Verlag Chemie International, Inc.: Deerfield, 1983; pp 11–64.

Table 3. $^3J_{CH}$ (Hz) from 3D HMQC-TOCSY[f]

	C1′H3′	C3′H1′	C2′H4′	C4′H2′	C3′H5′ [c] (pro-*S*)	C3′H5″ [c] (pro-*R*)	C5′H3′
G1	*a*	*b*	+0.4	+3.8	+2.2	+1.9	*a*
G2	*a*	*a*	*b*	+4.1	*e*	*e*	*a*
A3	*a*	*a*	*b*	A3 H3′=H2′ (+1.7)	*e*	*e*	*a*
C4	*a*	*a*	+1.6	+4.4	*e*	*e*	*a*
U5	*a*	*a*	+1.3	+5.2	*a*	+1.8	*a*
U6	**+6.6**	**+0.6**	*a*	*a*	*a*	*a*	*a*
C7	**+6.6**	**+0.6**	*a*	*a*	*a*	*a*	*a*
G8	*a*	*a*	+0.3	+4.6	**+0.5**[d]	G8 H5″ [d]=H4′ (−2.5)	+6.2
G9	*a*	*a*	+1.1	+4.3	+2.0	+0.5	+4.7 ± 1.0
U10	*a*	*a*	*b*	+5.8	*e*	+1.5	*a*
C11	*a*	*a*	*b*	+5.8 ± 0.8	*e*	+0.5	*a*
C12	*a*	+2.1	+1.0	+4.7	+4.8 ± 0.8	*b*	*a*

[a] ^{1}H–^{1}H TOCSY cross-peak is very weak or not observed. [b] Cross-peak is observed but is partially overlapped with diagonal, satellite or another peak. [c] H5′ is the more downfield proton unless otherwise specified. [d] H5″ is the more downfield proton. [e] Too much overlap to tell if the peak is present. [f] The cross-peak offset was measured three times, and the values were averaged (see Table 1). The indicated errors represent the variability in the measured values. All values are within ±0.4 Hz unless otherwise noted. Couplings in boldface are values uniquely different from couplings observed for standard A-form RNA.

since they are simple H–C–^{13}C–X systems. Table 4 shows only the predicted signs for the ribose $^2J_{CH}$. No attempt was made to predict the magnitudes. These sign predictions are based only on whether there is a gauche or trans oxygen for the exact ribose geometry specified. For $^2J_{C1'H2'}$, however, the arrangement is H–C–^{13}C(O)N. The resultant vector model[8b,c] and the projection sum[8a] can be used to analyze this system; however, since the effect on magnitude and sign of $^2J_{CH}$ increases with increasing electronegativity of X,[28] the differing electronegativity of N and O needs to be considered. It is reasonable that for a pure 3′-endo conformation the effect of an oxygen trans to H2′ will be greater than a gauche nitrogen, thus leading to a net positive $^2J_{C1'H2'}$. For a pure 2′-endo conformation the H2′ is gauche to both heteroatoms resulting in a negative $^2J_{C1'H2'}$. This dependence of the sign on orientation with respect to a heteroatom (e.g., oxygen) is very useful for the determination of several important parameters in RNA as described below.

$^3J_{CH}$ Couplings. The $^3J_{CH}$ values measured from the HMQC-TOCSY are shown in Table 3. The couplings which could not be measured with the TOCSY due to poor ^{1}H–^{1}H coherence transfer could be observed in a NOESY. However, in the short mixing time 3D HMQC-NOESY these cross-peaks were too weak to measure the coupling accurately. Three-bond carbon-proton scalar coupling constants are expected to follow the general Karplus curve trend, with large values at torsion angles close to 0° or 180° and very small values close to 90°. In general, $^3J_{CH}$ for the system H–C–C–^{13}C–X increases with increasing electronegativity of X, while $^3J_{CH}$ for H–C–C(X)–^{13}C decreases with increasing electronegativity of X.[28] It is also possible for nonbonded interactions to affect the $^3J_{CH}$ value.[28a]

$^2J_{CH}$ for Stereospecific Assignment of H5′(pro-*S*)/H5″(pro-*R*) Protons, Determination of Sugar Conformation and Torsion Angle γ. The sign of the $^2J_{C4'H5'/H5''}$ can be used to assign stereospecifically the H5′(pro-*S*)/H5″(pro-*R*) protons and along with $^2J_{C5'H4'}$ can assign the torsion angle γ (for a preliminary report of this work see Hines et al.[9]). Table 4 shows the predicted signs based on the relation between the orientation of HCCX and the sign of $^2J_{CH}$. Different methods have been used previously for the stereospecific assignment of diastereotopic protons in sugars. Stereoselective deuteration at C5′ has allowed the stereospecific assignment of H5′ and H5″ in ribonucleosides and could be extended to RNA oligonucleotides.[7b] The absolute magnitude of the $^2J_{C3H4S}$ versus $^2J_{C3H4R}$ in tetrahydrofuranosides has been used to stereospecifically assign H4-*S* and H4-*R*.[7c] Comparison of the magnitude of three-bond CH scalar couplings has been used to stereospecifically assign the H-6 protons in β-D-glucopyranoside,[7f] and the same method could be used to assign H5′/H5″ in ribonucleotides using $^3J_{C3'H5'/H5''}$.[2] While it was not discussed by the authors, it appears that $^3J_{HP}$ and $^3J_{CP}$ were used to stereospecifically assign H5′/H5″ in a DNA duplex.[5]

By using the sign of $^2J_{CH}$ a stereospecific assignment is possible even when only one $^2J_{C4'H5'/H5''}$ can be measured, provided that the value for γ is already known: the sign of only one of $^2J_{C4'H5'}$ or $^2J_{C4'H5''}$ is necessary for the stereospecific assignment of H5′/H5″ if γ = g$^+$ or t (see Table 4). Using three-bond carbon-proton scalar couplings for stereospecific assignment requires that both $^3J_{C3'H5'}$ and $^3J_{C3'H5''}$ are known in order to compare

Table 4. Predicted Signs of $^2J_{CH}$ in RNA (See Text)

	γ[a] =		
$^2J_{CH}$	g+	t	g−
C4′H5′(pro-*S*)	−	+	−
C4′H5″(pro-*R*)	+	−	−
C5′H4′	+	−	−

	P[b] =	
$^2J_{CH}$	2′-endo	3′-endo
C1′H2′	−	+
C2′H1′	−	−
C2′H3′	−	+
C3′H2′	+	−
C3′H4′	−	−
C4′H3′	+	−

[a] Torsion angle γ = O5′–C5′–C4′–C3′. [b] The pseudorotation phase angle P characterizes the sugar conformation.[27]

Table 5. $^3J_{C8/6H'}$ from ω1 1/2X-Filtered NOESY[a]

	torsion angle χ	torsion angle C8–N9–C1′–H1′ or C6–N1–C1′–H1′	observed $^3J_{C8/6H1'}$
G1	−177 ± 3	+123	+1.7 ± 0.7
G2	−171 ± 2	+129	+1.8 ± 0.8
A3	−156 ± 4	+144	+2.2
C4	−134 ± 8	+166	+0.4
U5	−126 ± 14	+174	overlap
U6	−164 ± 14	+136	+3.0 ± 0.8
C7	−156 ± 6	+144	+2.4 ± 1.0
G8	+36 ± 4	−24	+1.0
G9	−148 ± 8	+152	overlap
U10	−152 ± 7	+148	+1.5
C11	−163 ± 2	+137	+1.3
C12	−159 ± 2	+141	+1.2

[a] The cross-peak offset was measured three times, and the values were averaged (see Table 1). The indicated errors represent the variability in the measured values. All values are within ±0.4 Hz unless otherwise noted. χ (O4′–C1′–N1–C2 for pyrimidines, O4′–C1′–N9–C4 for purines) is from ref 10c. Torsion angle C8–N9–C1′–H1′ or C6–N1–C1′–H1′ calculated from χ.

their magnitudes (vide infra). Examining two-bond and three-bond carbon-proton scalar couplings for stereospecific assignments are complementary methods. When γ = t, $^3J_{C3'H5'}$ and $^3J_{C3'H5''}$ will be of similar size, whereas $^2J_{C4'H5'}$ and $^2J_{C4'H5''}$ will have different signs. For $\gamma = g^-$ the $^2J_{C4'H5'/H5''}$ will have the same signs, and $^3J_{C3'H5'/H5''}$ will have different magnitudes.

The additional structural information gained by analyzing the $^2J_{CH}$ of UUCG illustrates the potential utility of this structure determination method. The data from the HMQC-TOCSY and HMQC-NOESY (Table 1 and 2) provided the stereo assignment for A3, U5, U6, C7, G8, U10, and C11 in UUCG. Stereospecific assignments could not be made for C4 due to overlap. In A-form RNA, it can be assumed that the H5'(pro-*S*) proton is the more downfield proton.[2] For A3, U5, U6, U10, and C11, all are $\gamma = g^+$ [10c] and have $^2J_{C4'H5'}$ = minus for the downfield proton. The signs of $^2J_{C4'H5'}$ and $^2J_{C4'H5''}$ agree with the predicted signs shown in Table 4 for $\gamma = g^+$ and for H5'(pro-*S*) being the more downfield proton. For C7 and G8, however, the H5''(pro-*R*) is the more downfield proton based on the observed and predicted signs for $^2J_{C5'H4'}$, $^2J_{C4'H5'}$, and $^2J_{C4'H5''}$.[9] This can be explained by the unusual geometry in the loop region and the ring current effects of the syn G8.

Previously, γ for G8 could not be determined due to spectral overlap. In all calculated structures, however, it was always t, presumably due to other less direct constraints.[10c] The signs of $^2J_{C5'H4'}$, $^2J_{C4'H5''}$, and $^2J_{C4'H5'}$ establish conclusively that the conformer γ = t was indeed correctly identified by the structure determination. The sign of $^2J_{C5'H4'}$ for U5, U6, and C7 also confirm the previous observation of $\gamma = g^+$. The H5'/H5'' protons of G9 could not be stereospecifically assigned with $^2J_{C4'H5'}$ and $^2J_{C4'H5''}$ due to overlap problems; however, they were assigned with $^3J_{C3'H5'/H5''}$ couplings (Table 3, *vide infra*).

In the stem region, the TOCSY transfer was not sufficient to see many of the desired peaks for determing $^2J_{C5'H4'}$, $^2J_{C4'H5''}$, and $^2J_{C4'H5'}$. In the loop region, even limited flexibility may have led to larger average ^{1}H–^{1}H scalar couplings and therefore a better coherence transfer. The NOESY experiment allowed access to the desired couplings in the stem region (Table 3). The magnitudes of $^2J_{C4'H5''}$ in the stem region are slightly smaller than in the loop. This could be due to spectral overlap problems. All of the stem $\omega3\omega2$ H5''H4' chemical shifts are very similar, and four of the nucleotides have exactly the same carbon chemical shift for C4' [10b] making it very difficult to have well resolved peaks due to the apparent line width in the proton dimensions. Even though the error between repeat measurements was small, partial spectral overlap could skew the observed coupling. There are also potential problems due to relaxation during the NOESY experiment as discussed above.

The only time when the sign of $^2J_{C4'H5'/H5''}$ cannot be used for stereospecific assignment is when $\gamma = g^-$ since the signs of both $^2J_{C4'H5'}$ and $^2J_{C4'H5''}$ would be negative. However, in structure studies of crystalline tRNA and ribomononucleosides, $\gamma = g^-$ is quite rare.[29] Another factor to consider is that the value of $^2J_{C5'H4'}$ and $^2J_{C4'H5''}$ will be expected to get small, pass through a null, and change sign at γ = –30° and +150°, while the same can be expected for $^2J_{C4'H5'}$ at γ = –90° and +90°. This leads to potential ambiguity in the stereospecific assignment of H5'/H5'' and determination of γ for 90° < γ < 150°, –90° < γ < –150°, and –30 < γ < 30° when only the carbon-proton couplings are examined.

The sign of $^2J_{CH}$ can also be used to determine the sugar conformation. Table 4 shows the predicted signs for the ribose ring $^2J_{CH}$s in 2'- and 3'-endo sugars. The observed couplings in Tables 1 and 2 confirm these predictions. Four out of six of the $^2J_{CH}$s in the ribose ring can be readily used for structure determination since the signs differ in 2'- versus 3'-endo sugar pucker. This allows for redundant information which will be useful when there are overlap difficulties.

The absolute magnitude of $^2J_{C1'H2'}$ and $2J_{C2'H1'}$ for UUCG (Tables 1 and 2) were similar to values reported for ribonucleosides.[7a] However, the absolute magnitudes of $^2J_{C2'H3'}$ (~2.5 Hz) in UUCG were larger than for ribonucleosides (~0 Hz). This is probably because the nucleotides in the rigid UUCG structure are all >95% 2'- or 3'-endo (except for G1 and C12[10c]), whereas the ribonucleoside values[7a] are averaged over different conformers. Cancellation in the observed $^2J_{C2'H3'}$ could occur with conformational averaging since the sign of $^2J_{C2'H3'}$ changes from 2'- to 3'-endo pucker. Consistent with this suggestion, the magnitude of $^2J_{C2'H3'}$, $^2J_{C3'H2'}$, and $^2J_{C1'H2'}$ for G1 and C12 are slightly smaller than the other 3'-endo sugars (Tables 1 and 2). Proton–proton couplings indicated that G1 and C12 are a mixture of 2'- and 3'-endo.[10c]

Both the sign and absolute magnitude of $^2J_{C1'H2'}$, $^2J_{C2'H3'}$, $^2J_{C3'H2'}$, and $^2J_{C4'H3'}$ can be used to estimate the proportion of 2'- vs 3'-endo RNA. A similar approach has been proposed for DNA using $|^2J_{C1'H2'}| - |^2J_{C1'H2''}|$.[7e] Information based on carbon-proton scalar couplings will be very useful for larger molecular weight RNA (>30 nucleotides), when the ^{1}H line width may significantly interfere with the current methods based on ^{1}H–^{1}H scalar couplings for determining the sugar pucker.[27]

$^3J_{CH}$ for Determining Sugar Conformation and Stereospecific Assignment of H5'(pro-*S*)/H5''(pro-*R*) Protons. In the UUCG hairpin all the nucleotides are >95% C2'- or C3'-endo except for G1 and C12 which are a mixture.[10c] The size of $^3J_{C2'H4'}$ shown in Table 3 (~1 Hz) agrees with the $^3J_{C2'H4'}$ for C3'-endo sugar conformation as determined by a previously published Karplus curve.[7a] For G8, $^3J_{C2'H4'}$ may be slightly smaller, perhaps due to a smaller sugar pucker *P* value of 13° ± 7° (corresponding to the angle H4'–C4'–C3'–C2' approximately equal to 90°) as opposed to C4, U5, G9, and C12 which are all in the range of *P* = 23–37° (H4'–C4'–C3'–C2' > 90°) with similar amplitudes, ϕ_m = 38–40°.[10c]

According to the published Karplus curves,[7a] the value for $^3J_{C1'H3'}$ of the C2'-endo sugars should be approximately 4 Hz. However, the observed coupling shown in Table 3 is close to 6.5 Hz for U6 and C7. This difference could be due to the sugars in the loop having a more rigid structure than the previously studied ribomononucleosides which are conformationally flexible.

The $^3J_{C3'H5'}$ and $^3J_{C3'H5''}$ can be used for the stereospecific assignment of the C5' protons. In the UUCG hairpin, however, the only nucleotide where both couplings could be measured was G9 (Table 3). The magnitude of the two couplings indicate that the more downfield proton is H5' since with $\gamma = g^+$ the $^3J_{C3'H5'}$ should be larger than $^3J_{C3'H5''}$.

The stereospecific assignment with the three-bond carbon-proton scalar coupling indicates that G9 H5'(pro-*S*) is the more downfield proton. This information rules out one of the two possible families of structures reported earlier.[10c] The stereospecific assignment of the more downfield proton of G9 as H5'(pro-*S*) now confines β to g^- due to the large phosphorus–proton coupling with the downfield C5' proton.[10c] This only fits the family with G8 ϵ = t, $\zeta = g^-$ and G9 α = t, $\beta = g^-$. In addition, data[30] from a 2D HSQC experiment on the UUCG hairpin indicate that the $^3J_{C2'P3'}$ coupling for G8 is very small, consistent with ϵ = t. Finally, the G8-P-G9 phosphorus resonance is far downfield, which is consistent with α = t.[10c] None of these results supports the other possible family of structures which has G8 $\epsilon = g^-$, $\zeta = g^-$ and G9 $\alpha = g^-$, $\beta = g^+$.

Using $^3J_{C6H1'}$ and $^3J_{C8H1'}$ To Determine Torsion Angle χ. Karplus curves exist for $^3J_{C6H1'}$ in pyrimidine nucleosides,[7a,31] but there are no Karplus curves for $^3J_{C8H1'}$ in purine nucleosides. We measured $^3J_{C6H1'}$ and $^3J_{C8H1'}$ in the UUCG hairpin to relate these values to the torsion angle χ. The coupling constants

(29) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; pp 17–21, 55–59, 79–81, and 88–92.

(30) Santa Lucia, J. unpublished results.

measured with the 2D 1/2X-filtered NOESY are reported in Table 5. The torsion angle χ which was previously determined[10c] is shown, together with torsion angles for C6NC1'H1' and C8NC1'H1' calculated from χ.

Except for U6 and C7, the pyrimidine $^3J_{C6H1'}$ are smaller (0.4–1.8 Hz) than would be expected (4–6 Hz) based on the previously published Karplus curves.[7a,31] A possible explanation is the inherent limitations discussed above of using a NOESY to measure these couplings. Since a long mixing time (τ_m = 500 ms) NOESY was necessary in order to observe the cross-peaks, it is possible that some of the cross-peak is due to spin diffusion through the same sugar's H2' (a reasonable possibility in A-form RNA). If the H2' is then also bound to ^{13}C, the apparent offset of the cross-peak components for $\omega 3\omega 2$ H1'H6 or H1'H8 will be a result of both $^4J_{C6/8H2'}$ and $^3J_{C6/8H1'}$. Even if $^4J_{CH} \sim 0$ Hz, part of the cross-peak will not be offset by $^3J_{CH}$, and therefore the apparent coupling will be smaller than the actual $^3J_{CH}$. Another problem could be due to nonbonded interactions. Calculations have shown that nonbonded interactions can affect the magnitude of three-bond carbon-proton scalar couplings by as much as 30 Hz.[28a] It is possible that in the stem region there are nonbonding interactions, possibly aromatic ring currents, which are affecting the magnitude of $^3J_{C6H1'}$ and $^3J_{C8H1'}$. Due to all of the above potential problems and inconsistent results for the values of $^3J_{C6H1'}$ and $^3J_{C8H1'}$ more studies need to be done before this method can be used to reliably determine χ.

There was no noticeable difference in $^3J_{C8H1'}$ for the syn G8 compared to the other anti purines. However, this is not surprising since the approximate torsion angle of purine $^3J_{C8H1'}$ or pyrimidine $^3J_{C6H1'}$ for an anti base will be 142° and for a syn base will be −24°. Since, for pyrimidines, the Karplus curve does follow a typical U shape, both the −24° and 142° will give relatively large couplings.[7a,31] While the value measured for G8 is not as large as that expected based on the pyrimidine Karplus curves, it is still very similar to the other purines in UUCG.

Conclusion

We have demonstrated the utility and discussed some of the limitations of using carbon-proton scalar coupling constants for the structure determination of ^{13}C labeled RNA molecules at 30% isotopic enrichment. Similar results could be obtained with 100% labeled material, but there will be more interference from extra ^{13}C–^{13}C and ^1H–^{13}C scalar couplings as well as greater line broadening due to dipolar relaxation and passive couplings. The 100% enrichment will be more sensitive but also more expensive. The sensitivity of the 30% enrichment could be enhanced by use of a triple resonance probe in order to eliminate the passive phosphorus couplings. This would also help for 100% enrichment, but the passive ^1H–^{13}C couplings would still present a potential problem, more so than at only 30% enrichment.

The 3D HMQC-TOCSY is an excellent way to obtain the long range ^{13}C–^1H scalar couplings. The 3D HMQC-NOESY and the 1/2X-filtered NOESY are also very useful but can potentially give inaccurate results due to spin diffusion, poor sensitivity, and relaxation. This limitation is important when only the $^nJ_{CH}$ data are going to be used in the analysis of a particular torsion angle. Some of the inaccuracies can be minimized by using shorter mixing times and low isotopic enrichment. The ECOSY type cross-peaks significantly reduce the interference of line width in the coupling determination. The 3D experiments also allowed the assignment of several carbon resonances in the UUCG hairpin which could not be specifically identified by conventional 2D experiments.

The sign of the $^2J_{CH}$ is an excellent method for the stereospecific assignment of H5'(pro-*S*)/H5''(pro-*R*), determining the torsion angle γ(O5'–C5'–C4'–C3') and the pseudorotation phase angle P. Four out of six possible $^2J_{CH}$ in the ribose ring can be used to determine the sugar pucker P, and the absolute magnitudes could be used to determine the proportion of 2'-endo versus 3'-endo. The ability to determine the same structural parameter using different ribose sugar $^2J_{CH}$ will prove invaluable when there are overlap limitations. We have also demonstrated that $^3J_{C3'H5'}$ and $^3J_{C3'H5''}$ can be used to stereospecifically assign the H5'(pro-*S*) and H5''(pro-*R*) protons.

While this paper has mainly discussed the utility of two-bond and three-bond coupling constants for RNA structure determination, we have also been able to use this information to further refine the structure of the UUCG hairpin. With the two methods for stereospecific assignment of H5'/H5'' protons, we have determined that for two nucleotides, C7 and G8 of UUCG, the more downfield proton is H5''(pro-*R*). Analysis of the previously determined UUCG structure shows that these protons are positioned such that they could be affected by the ring current of the syn G8. For all the other nucleotides which could be measured, the data are consistent with the more downfield proton being H5'(pro-*S*) as expected for A-form duplexes. With the stereospecific assignment of G9 H5'/H5'' we can now report that the correct family of structures for the UUCG RNA hairpin is G8 ϵ = t, ζ = g$^-$, G9 α = t, β = g$^-$. The sign of $^2J_{C5'H4'}$ of G8 also confirmed that γ of G8 was trans. Due to spectral overlap in the 2D experiments[10c] this angle could not be determined from $^3J_{H4'H5'/H5''}$.

The refinement of this well-defined structure illustrates the utility of $^nJ_{CH}$ in RNA structure determination. Significantly, this method is most powerful (less spectral overlap, better TOCSY transfer) in the less well determined but structurally more interesting single-stranded regions. For further studies, both the advantages and potential limitations of this new structure determination methodology need to be considered. This study, which correlates the signs and magnitudes of $^nJ_{CH}$ both to double-stranded A-form and non-helix RNA structure, forms a basis from which strategies for the structure determination of larger RNA and RNA–protein complexes can be derived.

Experimental Section

Phenol (pH 4 for crude RNA isolation, pH 8 for all other extractions) was obtained from Amresco. The pH 4 phenol was extracted once with an equal volume of 20 mM sodium acetate, 1 mM EDTA pH 5.5 before use. Sodium dodecyl sulfate (gold label) was from Aldrich. Dowex 50W-X8 (H^+-form, 20–50 mesh) and AG1-X8 resin (acetate-form, 200–400 mesh) were obtained from Bio-Rad. Nuclease P1 was obtained from Calbiochem, and DNaseI (RNase-free) was obtained from Boehringer Mannheim. All the enzymes for the rNMP to rNTP conversion were purchased from Sigma as lyophilized powders and then dissolved in doubly distilled water, except for PM. PM was purchased from Sigma as a crystalline suspension in ammonium sulfate solution and used in that form.

All other reagents were reagent grade and obtained from standard sources. All buffers were prepared from doubly distilled water and autoclaved for sterilization unless otherwise noted.

A. Isolation of Total Cellular ^{13}C RNA from Methylobacterium extorquens. The procedure used was based on the procedure of Oelmuller et al.[14] for the isolation of total prokaryotic RNA. Log phase *Methylobacterium extorquens* strain AM1 cells were supplied as a frozen suspension (1:1 weight/volume in 0.2% saline solution) by the Stable Isotopes Resource at Los Alamos National Laboratory. The cells were grown in a minimal salts/phosphate media containing trace amounts of $ZnCl_2$, NH_4MoO_4, $MnSO_4$, $CuSO_4$, $CaCl_2$, $CoCl_2$, and boric acid.[19] The carbon source, methanol (0.5% w/v), was ^{13}C-labeled at 30% isotopic enrichment. A 64 L fermentation yielded 180 g wet cells.[13]

The frozen cell suspension (80 g) was thawed in ice-cold 20 mM sodium acetate, 1 mM EDTA pH 5.5 buffer (AE buffer, 100 mL), and the cells were pelleted by centrifugation on a GS-3 rotor (7 K rpm, 4 °C, 20 min). In the meantime, a 1-L, round-bottom flask containing AE buffer (250 mL), phenol (equilibrated with AE buffer, 150 mL), and sodium dodecyl sulfate (5 g) was heated in an oil bath at 95 °C. Hot phenol extraction

(31) (a) Davies, D. B.; Rajani, P.; Sadikot, H. *J. Chem. Soc., Perkin Trans. II*, **1985**, 279–285. (b) Davies, D. B.; Rajani, P.; MacCross, M.; Danyluk, S. S. *Mag. Res. Chem.* **1985**, *23*, 72–77. (c) Lemieux, R. U.; Nagabhushan, T. L.; Paul, B. *Can. J. Chem.* **1972**, *50*, 773–776.

at only 65 °C resulted in a lower yield of RNA. Pelleted cells were resuspended in ice-cold AE buffer (150 mL) and then added to the hot phenol mixture. The bright pink mixture was stirred vigorously for 20 min. The lysed cells were then cooled on ice, and chloroform/isoamyl alcohol (a solution of 24:1, 150 mL) was added. The layers were mixed by shaking and then separated by centrifugation (GS-3, 7 K rpm, 20 min, 4 °C).

After separation, sodium acetate (45 mL of a 3 M solution, pH 5.6) was added to the ice-cold aqueous layer. This 0.3 M NaOAc solution was then extracted with phenol (equilibrated with AE buffer, 150 mL) and combined with chloroform/isoamyl alcohol (24:1), 150 mL). The first phenol layer was back extracted with AE buffer (150 mL). The combined aqueous layers were extracted with chloroform/isoamyl alcohol (24:1, 150 mL). DNA can then be isolated from the phenol/protein mixtures by extraction with an equal volume of 1 M Tris (pH 10.5).[16]

The crude RNA solution was frozen and then lyophilized to dryness. The dried material was resuspended in ice cold water (120 mL), and 3 volume equiv of 100% ethanol was added. The resultant solution was stored at −20 °C overnight. After centrifugation (ss-34 rotor, 12 K rpm, 30 min, 4 °C) the RNA pellets were resuspended in 40 mM Tris, 6 mM $MgCl_2$ pH 8.1 buffer (TM buffer, 54 mL). (A_{260} = 10 400 OD units, A_{260}/A_{280} = 2.0.)

The crude RNA solution was treated with DNAse I (RNase-free 200 μL, 40 units/μL) for 2 h at 37 °C. The DNAse I was removed by phenol extraction (40 mL, pH 8 phenol) followed by extraction with chloroform/isoamyl alcohol (24:1, 20 mL). The aqueous layer was made 0.3 M in NaOAc at pH 5.6, and ethanol precipitated overnight with 3 volume equiv of 100% ethanol at −20 °C.

The cellular RNA was pelleted as before, and the pellets were resuspended in 10 mM NaOAc pH 5.6 (24 mL). The resultant solution had an RNA concentration of A_{260} = 9411 OD units and an A_{260}/A_{280} ratio of 2.0 indicating pure RNA with no protein, phenol, or DNA contamination.

B. Preparation of ^{13}C rNMP's from Isolated Cellular RNA. Nuclease P_1 (400 μL of a 400 units/mL solution in 10 mM NaOAc pH 5.7) was added to the cellular RNA prepared above. The solution was incubated at 55 °C for 2 h and periodically shaken gently. The completion of the reaction was monitored by UV and also by the ion-exchange column described below. Incubation with additional enzyme did not increase the yield. Even though the enzyme requires Zn^{2+},[32] no increase in yield was observed when Zn^{2+} was added to a final concentration of 0.1 mM. Presumably, there is enough adventitious metal left over from the cellular extraction for the nuclease P1 to function properly. The solution was then cooled on ice and extracted with phenol (pH 8, 20 mL) followed by chloroform/isoamyl alcohol (24:1, 20 mL). The solution was frozen and lyophilized overnight.

All buffers for the anion exchange column were prepared from sterile double distilled water, the pH was adjusted, and then the solution was filtered through a sterile 0.2 μ filter (Schleicher & Schuell). The rNMPs were then separated on AG1-X8 resin (acetate form, 60 mL in slurry form, final column dimensions 34 cm × 1.7 cm) using a step gradient of 50 mM, 0.3 M, and 0.8 M ammonium acetate pH 4.3.[11c,33] The appropriate fractions for the rNMPs were collected, frozen, and lyophilized to yield CMP 160 mg, UMP 80 mg, AMP 80 mg, and GMP 120 mg as dry powders. The overall yield is approximately 11 mg of rNMPs (NH_4^+ salt) per gram of wet cell.

All of these rNMPs are in the ammonium salt form. When resuspended in sodium phosphate buffer at pH 7, the UV and NMR spectra of all the isolated rNMPs agreed with rNMPs from Sigma. The percent of isotopic enrichment was measured by integrating the area of the ^{13}C-bound proton satellites (split by the large $^1J_{CH}$) versus the ^{12}C-bound proton peak in 1D proton NMR of GMP H1′. To avoid integration errors due to different relaxation characteristics of ^{13}C and ^{12}C, the integration was done both on a single scan of a concentrated GMP sample and on an average of scans with a long (>15 s) recycle delay. In each case, the isotopic enrichment was determined to be 30% ± 2%.

The rNMPs in the ammonium salt form were used directly in the phosphorylation reactions. The rNMP mixture before anion exchange chromatography can also be used directly in the enzymatic phosphorylation step.

C. Enzymatic Conversion of ^{13}C rNMPs to ^{13}C rNTPs. 5′-rNTPs were enzymatically synthesized based on the procedure of Simon et al.[18] This can be done on each nucleoside individually or on all four together. All reaction conditions are for a total volume of 10 mL.

For the UTP and CTP individual preparations, PEP, the phosphate source, was prepared in the reaction mixture from 3-PGA. The barium salt of 3-PGA was converted to the sodium salt by stirring with Dowex 50W-X8 resin in the Na^+ form for 30 min. The resin was then filtered and rinsed. The filtrate was used directly in the reaction mixture to give a concentration of 200 mM 3-PGA for UTP and 100 mM 3-PGA for CTP. The reaction mixture for UTP also contained $MgCl_2$ (40 mM), KCl (40 mM), Tris (32 mM), ATP (0.4 mM), and UMP (45.6 mM). The pH of the above solutions was adjusted to pH 7.6 using NaOH for UTP and KOH for CTP. The solutions were then deaerated with N_2 for 40 min, and dithiothreitol (DTT) was added for a final concentration of 2.5 mM. The enzymes were added in an N_2 glove bag. For the CTP reaction, MK (62 units), PK (38 units), enolase (16 units), PM (35 units), and NMPK (0.024 units) were added. For the UTP reaction, MK (480 units), PK (48 units), enolase (12 units), PM (48 units), and NMPK (0.05 units) were added. Reactions were run in a N_2 glove bag at room temperature. The pH was maintained with a pH controller and peristaltic pump delivering 0.5 M HCl, and 0.5 mL was added over the course of each reaction. The UTP reaction was complete after 16 h, and the CTP reaction after 5 h. These two reactions may be run without NMPK but require longer reaction times (12 and 6 days for UTP and CTP, respectively). Similar results have been reported previously.[11a]

GTP and ATP reactions were carried out at 37 °C using PEP directly. The reaction mixtures contained PEP (180 mM), $MgCl_2$ (9.5 mM), KCl (25.6 mM for GTP and 9.5 mM for ATP), ATP (0.6 mM), Tris (12.5 mM) and GMP or AMP (20 mM). The pH was adjusted to 7.6 using 0.5 M NaOH. The solutions were then deaerated with N_2 for 40 min, and dithiothreitol (DTT) was added for a final concentration of 2.5 mM. The enzymes were added under N_2. For the GTP reaction, GK (2.3 units) and PK (11.6 units) were added. For the ATP reaction, MK (21 units), and PK (21 units) were added. Both reactions were run under N_2-filled balloons at 37 °C for 20 h. The pH was maintained at 7.6 by the addition of 1 mL of 0.5 M HCl over the course of the reaction.

For the phosphorylation of all four rNMPs together, the rNMPs (113 mM total concentration) were added to $MgCl_2$ (80 mM), KCl (80 mM), Tris (140 mM), PEP (400 mM), and ATP (0.8 mM). The pH was adjusted, and the solution was deaerated as described above. GK (2 units), MK (800 units), PK (80 units), and NMPK (6.5 units) were added. DTT (2.6 mM) was also added. The reaction was run at 37 °C under N_2 for 14 h. HCl (2 mL of 0.5 M) was added over the course of the reaction to maintain the pH at 7.6.

All of the enzymatic reactions were monitored by ^{1}H and ^{31}P NMR. All reactions were quantitative as determined by NMR. When complete, the rNTP was precipitated from the reaction by the addition of an equal volume of 100% ethanol and then pelleted by centrifugation. The resulting material (~95% recovery) was used in transcription reactions without further purification.

D. T7 Transcription of UUCG RNA. The rNTPs were dissolved in doubly distilled water to give solutions that were 20 mM in each rNTP. The pH was carefully adjusted to 8.1. In vitro T7 transcription was carried out as previously described.[34] A 40-mL transcription reaction was carried out in a buffer containing Tris (40 mM), spermidine (1 mM), triton X-100 (0.01%), DTT (5 mM), polyethylene glycol (80 mg/mL), rNTPs (6 mM in each), DNA template (800 nM), DNA top strand (800 nM), GMP (5 mM), $MgCl_2$ (12.5 mM), and T7 RNA polymerase. All reactions were 10 mL in volume. The $MgCl_2$ concentration was optimized for the isotopically labeled cellular rNTPs. This concentration was $^1/_4$ the $MgCl_2$ concentration used with commercial rNTPs. After optimization, RNA yields are the same as with commercially available nonlabeled rNTPs. The RNA was then phenol extracted, ethanol precipitated, and purified using 20% polyacrylamide gel electrophoresis.

E. NMR. The UUCG RNA (approximately 5 mg) was dialyzed extensively against 10 mM sodium phosphate/100 μM EDTA (pH 6.5). The purified RNA was then lyophilized several times in 99.8% D_2O and dissolved in 0.4 mL 99.96% D_2O (Aldrich) to a final concentration of approximately 3 mM. All the NMR experiments were done on a Bruker AMX-600 spectrometer operating at 600 MHz for proton and 150.9 MHz for carbon, using an inverse 5-mm probe and referenced against internal TSP. Data were transferred to a Silicon Graphics personal Iris and processed using the program FELIX (Biosym). Both the 2D and 3D

(32) (a) Fugimoto, M.; Kuninaka, A.; Yoshino, H. *Agr. Biol. Chem.* **1975**, *39*, 1991–1997. (b) Fugimoto, M.; Kuninaka, A.; Yoshino, H. *Agr. Biol. Chem.* **1975**, *39*, 2145–2148.

(33) Sinsheimer, R. L., Koerner, J. F. *Science* **1951**, *114*, 42–43.

(34) Milligan, J. F.; Uhlenbeck, O. C. *Methods Enzymol.* **1989**, *180*, 51–62.

experiments were recorded in the phase-sensitive mode using the TPPI method to obtain pure absorption spectra in the indirectly observed dimensions.[35]

For measurement of $^3J_{C6H1'}$ in pyrimidines and $^3J_{C8H1'}$ in purines, the $\omega 1$ 1/2X-filtered NOESY was acquired without carbon decoupling.[20] The delay $1/(4J)$ was tuned to 1.25 ms for efficient excitation/selection of aromatic resonances (J = 200 Hz), and the carbon carrier was set for aromatic carbons. The standard phase cycle was used to observe in $\omega 1$ only protons bound to ^{13}C.[20] In addition, the phase of the first proton pulse was cycled to reduce TPPI diamond artifacts[36] which overlapped with the $\omega 2$ = H1', $\omega 1$ = aromatic cross-peaks. The initial t_1 increment was set to $t_1(0) = DW_{t1} - (4/\pi)\tau_{90H}$ in order to overcome base line distortion and simplify phasing in t_1.[37] The mixing time was 500 ms. The residual HDO peak was reduced with a low decoupler power preirradiation during the relaxation delay and the mixing time. The 400 FIDs of 2 K complex data points were collected with t_2 max = 0.41 s and t_1 max = 40 ms. Each FID was an average of 128 scans with a repetition delay of 2 s. The sweep width in both dimensions was 5000 Hz. Total acquisition time was 2 days. Data were zero-filled to 4 K real points in t_2 and apodized using 90° phase-shifted sine bell skewed by a factor of 0.7. In t_1, data were zero-filled to 2 K real points and apodized using 90° phase-shifted sine bell.

The 3D HMQC-TOCSY was acquired without carbon decoupling.[21] A four-step phase cycle was used,[21] and a 68 ms "clean" TOCSY mixing sequence was used.[23] The delay $1/(2J)$ was tuned to 3.3 ms for efficient excitation/selection of the sugar resonances (J_{CH} ~150 Hz), and the carbon carrier was set in the center of the sugar carbon spectrum. The $t_1/2$ increment was 40 μs ($t_1/2$ max = 1.6 ms), and the t_2 increment was 250 μs (t_2 max = 40 ms). Total number of increments in each dimension were t_1 = 40, t_2 = 160 resulting in 6400 FIDs of 512 complex data points (t_3 max = 0.26 s). An average of 16 scans was collected with 32 dummy scans at the beginning of the experiment. A repetition delay of 2 s resulted in a total acquisition time of 3 days. The sweep width in $\omega 2 = \omega 3$ = 2000 Hz and $\omega 1$ = 6250 Hz with a filter width of 2500 in $\omega 3$ in order to reduce folded aromatic resonances. No irradiation of the residual HDO peak was necessary. A second experiment was done with $\omega 1$ = 6250, $\omega 2$ = 4000 (t_2 increment of 125 μs), and $\omega 3$ = 2000 with all the other parameters the same. The data were processed as follows: $t_3 = t_2$, a sine bell phase shifted by 30° and skewed by a factor of 0.7; t_1, a sine bell phase shifted by 90° and skewed by a factor of 0.7. The dimensions of the final real matrix after zero-filling were $\omega 3$ = 1024 × $\omega 2$ = 256 × $\omega 1$ = 128.

The 3D HMQC-NOESY was obtained without carbon decoupling and without homospoil at the beginning of the mixing time.[22] A 16-step phase cycle selected protons bound to ^{13}C. The delay $1/(2J)$ was tuned to 3.3 ms for efficient excitation/selection of the sugar resonances (J_{CH} ~ 150 Hz), and the carbon carrier was set to the middle of the carbon sugar spectrum. The mixing times were 150 or 500 ms. The initial t_1 increment was set to $2[t_1(0)/2] = DW_{t1} - \tau_{180H} - (4/\pi)\tau_{90C}$ in order to overcome base line distortion and simplify phasing in t_1.[37] The $DW_{t1/2}$ was 40 μs ($t_1/2$max = 1.6 ms) and the DW_{t2} increment was 250 μs (t_2 max = 33.5 ms). Total number of increments in each dimension were t_1 = 40, t_2 = 134 resulting in 5360 FIDs of 256 complex data points. An average of 16 scans and a repetition delay of 1.5 s with 32 dummy scans at the beginning of the experiment resulted in a total acquisition time of 3 days. The sweep width in $\omega 2 = \omega 3$ = 2000 Hz and $\omega 1$ = 6250 Hz with a filter width of 2500 Hz in $\omega 3$. The data were processed as follows: $t_3 = t_2$, a sine bell phase shifted by 30° and skewed by a factor of 0.7; t_1, a sine bell phase shifted by 90° and skewed by a factor of 0.7. After $\omega 3$ processing first the first t_2 data point for each $\omega 3$ spectrum was multiplied by 0.5 in order to reduce "t_1" noise.[38] The dimensions of the final real matrix after zero-filling were $\omega 3$ = 1024 × $\omega 2$ = 256 × $\omega 1$ = 128.

Acknowledgment. Supported by NIH GM 10840, DOE DE-FG03-86ER60406, and instrumentation grants DOE DE-FG05-86ER75281, NSF DMB 86-09305, and NSF BBS 86-20134. J.V.H. was supported in part by NIEHS postdoctoral training grant 2T32 ES 07075. We would like to thank the Stable Isotope Resource at Los Alamos National Laboratory for supplying us with the ^{13}C-labeled bacteria. We would also like to thank Dr. John Santa Lucia and Dr. Jeffrey Pelton for informative NMR discussions, Mr. David Koh for DNA synthesis, Ms. Alison Greene for help with RNA isolation, Mr. Robert Lesch for bacterial fermentation, and Ms. Barbara Dengler for running an efficient laboratory.

(35) Marion, D.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 967–974.

(36) Simorre, J.-P.; Marion, D. *J. Magn. Reson.* **1990**, *89*, 191–197.

(37) Bax, A.; Ikura, M.; Kay, L. E.; Zhu, G. *J. Magn. Reson.* **1991**, *91*, 174–178.

(38) Otting, G.; Widmer, H.; Wagner, G.; Wüthrich, K. *J. Magn. Reson.* **1986**, *66*, 187–193.