Determination of RNA Conformation by Nuclear Magnetic Resonance

PETER B. MOORE

Department of Chemistry, Yale University, New Haven, Connecticut 06520

Received November 2, 1994

Introduction

Spectroscopic determination of RNA conformation has recently emerged as a distinct subdiscipline of biophysical chemistry, and as befits its youth, it is in flux. In the past, RNA spectroscopy depended on ^{1}H and ^{31}P experiments. In the future, experiments that use samples labeled with ^{13}C and ^{15}N will dominate. This account comments on both approaches and on the nature of the information they provide.

Historical Background

Although it has been recognized for decades that the function of many RNAs depends on their conformations, it is still true that only a handful of RNA structures are known at atomic resolution. The reason is that, until the late 1980s, no RNAs except those available in nature could be characterized structurally, and only a few can be obtained that way in the amounts required: the transfer RNAs (tRNAs), the ribosomal RNAs, and viral RNAs. With the exception of the 5S RNA from ribosomes ($M_r = 40\,000$), only the tRNAs had molecular weights small enough to appear workable ($M_r = 25\,000$).

The systematic application of nuclear magnetic resonance (NMR) to RNA began in 1971 when it was discovered that resonances due to hydrogen-bonded imino protons in RNA base pairs, GN1 and UN3 protons (Figure 1), can be detected between 10 and 15 ppm, downfield of all other RNA proton resonances.[1,2] Imino proton spectra tend to be well-resolved. Those given by tRNAs, for example, contain roughly 25 resonances, about one per base pair, and they are dispersed over a spectral region as wide as that which contains the molecule's 600 nonexchangeable proton resonances (see Figure 3). Furthermore, downfield spectra are worth studying because assigning an RNA's downfield spectrum is approximately equivalent to determining how it is base paired.

Pure tRNAs were available in 1971, and tRNA conformation was not understood. Thus, provided resonances could be assigned, a significant problem could be solved. By the time it was discovered that imino proton spectra can be assigned by NOE spectroscopy,[3–9] however, the tRNA problem had been solved crystallographically.

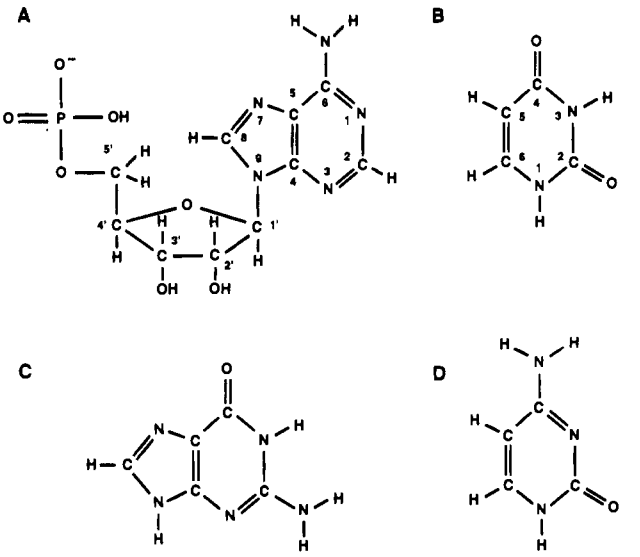

Figure 1. Structures of the RNA nucleotides: (A) adenosine 5′-monophosphate, with the number conventions indicated for its purine and ribose rings; (B) uracil; (C) guanine; (D) cytosine. The other 5′ monophosphates are obtained by replacing the purine ring of adenosine with uracil, guanine, and cytosine, respectively. Purines are linked to riboses by C1′–N9 glycosidic bonds. The linkage for pyrimidines is C1′–N1.

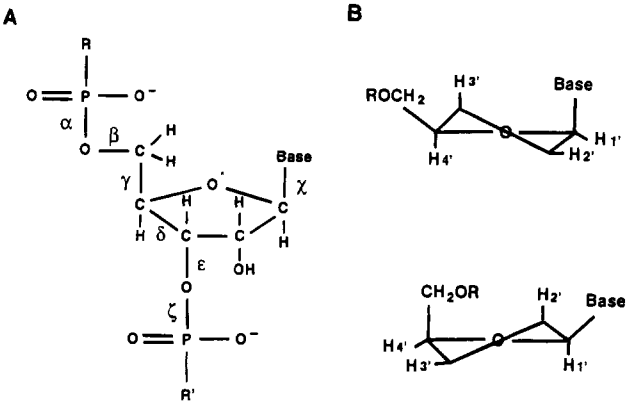

Figure 2. Nucleotide conformation: (A) The torsion angles associated with each nucleotide are indicated. R and R′ are the nucleotides on the 5′ and 3′ sides of the one shown, respectively. "Base" is any one of the four heterocyclic bases shown in Figure 1. (B) The conformations of ribose when its puckering is the C3′-endo (north) (top) and C2′-endo (south) (bottom) are shown.

Emphasis shifted from imino proton spectroscopy to full conformational characterization in the late 1980s because of two developments: (1) the discovery of chemical[10] and enzymatic[11] methods for synthesizing RNAs, which liberated the field from its preparative

Peter B. Moore, who was born in Boston, MA, in 1939, got his B.S. from Yale University in 1961 and his Ph.D. from Harvard University in 1966, both in biophysics. After three years as a postdoctoral fellow at the University of Geneva (Institut de Biologie Moleculaire) in Switzerland and the Medical Research Council Laboratory of Molecular Biology, Cambridge (U.K.), he joined the faculty of the Department of Molecular Biophysics and Biochemistry at Yale. He is now professor of chemistry at Yale, with a joint appointment in molecular biophysics and biochemistry. The focus of his research is the structural characterization of the RNAs and ribonucleoproteins involved in protein synthesis.

(1) Kearns, D. R.; Patel, D.; Shulman, R. G. *Nature* **1971,** *229,* 338–339.

0001-4842/95/0128-0251$09.00/0

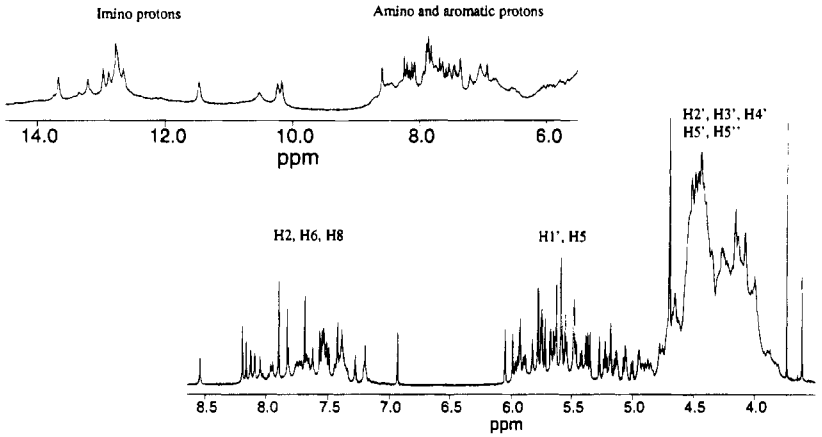

Figure 3. Typical RNA proton spectra. Top: the downfield part of an RNA spectrum taken in H_2O. Bottom: an entire RNA spectrum taken in D_2O. The sharp resonance at 4.7 ppm is residual HDO, and the sharp resonances at about 3.8 ppm are buffer components. The types of resonances expected in each region are indicated. (Reprinted with permission from the Ph.D. thesis of A. Szewczak, Yale University, 1994. Copyright 1994, A. Szewczak.)

bind, and (2) the evolution of methods for determining macromolecular conformations by NMR.[12] The discovery of a host of new, stable RNAs to study in the 1970s and 1980s and the discovery of RNA catalysis[13,14] contributed mightily to the field's vigor.

The Problem To Be Solved

Four nucleotide monomers predominate in RNA: two purines, A (adenosine 5′-monophosphate) and G (guanosine 5′-monophosphate), and two pyrimidines, U (uridine 5′-monophosphate) and C (cytidine 5′-monophosphate) (Figure 1). (Modified nucleotides do occur in nature, but so far, none of the RNAs characterized in detail spectroscopically has included any.) Adjacent monomers are linked 5′ to 3′ by phosphodiester bonds.

RNAs adopt folded conformations, which are stabilized by base stacking and by hydrogen bonding between the donors and acceptors, which abound in RNAs. Their conformations can be described by specifying seven torsion angles per nucleotide monomer (Figure 2). Six of the seven describe the trajectory of the molecule's backbone, and the seventh, χ, specifies the orientation of bases relative to ribose rings. The anti conformation, in which purine N1s and pyrimidine N3s point away from ribose rings, predominates in RNA, but occasional residues are found in the syn conformation, where purine N1s and pyrimidine N3s lie over their ribose rings. The ribose ring puckers that predominate are C3′-endo (abundant) and C2′-endo (rare)[15] (Figure 2). Sugar pucker fixes one of the backbone torsion angles, δ, as well as all the other ribose torsion angles.

In principle, the most efficient way to work out an RNA's conformation would be to determine its torsion angles directly: by measuring three-bond coupling constants, for example. But, even if all the relevant couplings could be measured, unrealistically accurate torsion angle estimates would have to be derived from them to obtain reasonable structures. In practice, heavy reliance is placed on proton–proton distances derived from nuclear Overhauser effect (NOE) intensities. Whatever the mixture of coupling constants and NOEs used, as Figure 3 shows, the spectroscopic information required for conformation determination must be extracted from spectra that are intrinsically poorly resolved.

Molecular Weight and the Choice of RNAs for Study

For any class of macromolecules there is always a molecular weight beyond which the techniques of the day cannot assign spectra because resonances are too numerous and too broad. The limit is in the neighborhood of 10 000 (30 nucleotides) for RNA today, but the new heteronuclear approaches discussed below, combined with increases in spectrometer field strength, could double it in the next few years.

Unfortunately, almost no naturally occurring RNAs have chain lengths as short as 60 nucleotides. However, the secondary structure of the typical RNA is dominated by hairpin-like structures that consist of an intramolecular helical stem with a loop that closes its distal end, and its tertiary structure is generated by interactions between these stem/loops and between them and less ordered sequences.[16] Many stem/loops are small enough to analyze spectroscopically, and small RNA motifs like these often retain their conformations in isolation.

While Watson–Crick pairing (A with U, and C with G) dominates in small RNA motifs, they often include other juxtapositions and bases that lack obvious

(2) Kearns, D. R.; Patel, D.; Shulman, R. G.; Yamane, T. *J. Mol. Biol.* **1971**, *61*, 265–270.
(3) Johnston, P. D.; Redfield, A. G. *Nucl. Acids Res.* **1978**, *4*, 3599–3615.
(4) Johnston, P. D.; Redfield, A. G. *Biochemistry* **1981**, *20*, 1147–1156.
(5) Sanchez, V.; Redfield, A. G.; Johnston, P. D.; Tropp, J. *Proc. Nat. Acad. Sci. U.S.A.* **1980**, *77*, 5659–5662.
(6) Roy, S.; Redfield, A. G. *Nucleic Acids Res.* **1981**, *9*, 7073–7083.
(7) Hare, D. R.; Reid, B. R. *Biochemistry* **1982**, *21*, 1835–1842.
(8) Hare, D. R.; Reid, B. R. *Biochemistry* **1982**, *21*, 5129–5135.
(9) Roy, S.; Papastarvos, M. Z.; Redfield, A. G. *Biochemistry* **1982**, *21*, 6081–6088.
(10) Usman, N.; Ogilvie, K. K.; Jiang, M.-Y.; Cedergren, R. J. *J. Am. Chem. Soc.* **1987**, *109*, 7845–7854.
(11) Milligan, J. F.; Groebe, D. R.; Wetherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783–8798.
(12) Wuthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(13) Kruger, K.; Grabowski, P. J.; Zaug, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147–157.
(14) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849–857.
(15) Altona, C. *Rec: J. R. Neth. Chem. Soc.* **1982**, *101*, 413–433.
(16) Gutell, R. R. *Nucl. Acids Res.* **1993**, *21*, 3051.

hydrogen-bonding partners.[17] Thus the conformations of many small RNAs cannot be inferred from sequence alone. Furthermore, many RNA binding proteins interact with single stem/loops, and catalytically active RNAs below the molecular weight limit can sometimes be prepared by trimming natural sequences. There is a lot to be learned from the study of the conformations of RNA oligonucleotides.

Unfortunately, RNAs cannot be selected for study on the basis of size and interest alone. Many sequences aggregate at NMR concentrations (≥1 mM), and others lack unique structures. While these difficulties can sometimes be palliated by changes in temperature and/or ionic conditions, most sequences that show either propensity cannot be analyzed spectroscopically. About 75% of the RNA investigations we undertake are abandoned for these reasons. Once a suitable RNA is identified, the next step is to assign its proton spectrum.

Imino Proton Assignments

Imino proton resonances in RNA helices, which are always A-form,[17] are (usually) easy to assign because the imino protons in successive base pairs are close enough to give through-space, NOE correlations.[6] Thus every helical segment in an RNA engenders a set of NOE-connectable imino proton resonances. Less regularly structured regions often do the same.

UN3 resonances in Watson–Crick AU pairs resonate on the downfield side of the imino proton region (13–15 ppm) and give intense NOEs to the (narrow) AH2 proton resonances of their hydrogen-bonding partners.[5] GN1 protons in Watson–Crick GCs, which usually resonate between 12 and 13 ppm, give strong NOEs to the two (exchangeable) CN4 protons of their hydrogen-bonding partners and weaker, transferred NOEs to CH5s.[12] GUs, the most common non-Watson–Crick pairing in RNA, are also easy to identify.[3] Once imino proton resonances have been sorted by base pair type and ordered on the basis of imino–imino NOEs, sequences likely to form stems[18] can be correlated with imino proton runs and assignments can be made.

Imino proton runs in regions that are irregularly paired are matched with sequences using base type as the criterion, instead of base-pair type. The base type of an imino proton resonance can be determined by experiments that correlate imino proton resonances with the ^{15}N resonances of the nitrogens to which they are bonded. These correlations are best observed using ^{15}N-labeled samples,[19–22] but also can be seen using the ^{15}N present in unlabeled samples.[23] In any case, these experiments work because the ^{15}N chemical shift ranges of UN3s and GN1s do not overlap.[24]

"Traditional" ^1H-driven methods for assigning RNA spectra start with base resonances and work in toward the backbone, i.e., from the "outside" of the molecule, covalently speaking, to its "inside". They are largely NOE-driven. The new heteronuclear methods work from the backbone out to the bases and depend on scalar couplings (see below). The imino proton assignment method just described, which is a fallible (!), bases-in method, is the only one available today. A more robust, heteronuclear method, which depended on correlations between imino proton and nonexchangeable proton resonances generated by scalar couplings, would be welcome indeed.

Why Nonexchangeable Proton Resonances Are Hard To Assign

The assignment of RNA proton resonances is challenging for several reasons. First, base protons are not scalar coupled to ribose protons, and the protons within a single ribose (usually) do not constitute a single, scalar-coupled spin system either.[25] Consequently, proton total correlation spectroscopy (TOCSY) and correlated spectroscopy (COSY) experiments, the through-bond experiments protein spectroscopists use to identify resonances belonging to single residues,[12] are much less useful for RNA. Second, while it is easy to distinguish purines from pyrimidines (pyrimidine H6 protons are coupled to H5 protons, and purine H8 protons are not coupled to anything) there is no easy way to distinguish Us from Cs and Gs from As. Third, the backbone of an RNA is the only part of the molecule where covalent chemistry constrains atoms to relate to each other in a spectroscopically benign way. The backbone-associated protons of an RNA, however, are its H3′, H4′, H5′, and H5″ protons, which resonate around 4 ppm, where resonance overlap is at its worst. Thus the resonances most likely to yield conformation-independent sequential assignment information are the hardest to observe.

The Bases-In Strategy

The bases-in strategy depends on anomeric (H1′)–aromatic (H6, H8) NOEs, which are found in one of the better resolved regions of an RNA's nuclear Overhauser effect spectroscopy (NOESY) spectrum.[12,26–30] The H1′ (anomeric) proton of every ribose is (always) within NOE range of its own aromatic (H8 or H6) proton. In addition, in helical stems, the H1′ of residue n is within NOE range of the H6 or H8 of residue $(n + 1)$. Thus every anomeric proton in a stem, except the one at its 3′ terminus, gives NOE cross peaks to its own (H6, H8) and that of the next base in the sequence, and every (H6, H8) except the one at the 5′ terminus of a sequence gives NOEs to two H1′ proton resonances. The result is that the aromatic–anomeric part of a helical RNA's NOESY spectrum includes rectangular patterns of cross peaks

(17) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984.
(18) Turner, D.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167–192.
(19) Bax, A.; Griffey, R. H.; Hawkins, B. L. *J. Magn. Reson.* **1983**, *55*, 301–335.
(20) Griffey, R. H.; Poulter, C. D.; Bax, A.; Hawkins, B. L.; Yamaizumi, Z.; Nishimura, S. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 5895–5897.
(21) Griffey, R. H.; Redfield, A. G.; Loomis, R. E.; Dahlquist, F. W. *Biochemistry* **1985**, *24*, 817–822.
(22) Kime, M. J. *FEBS Lett.* **1984**, *175*, 259–262.
(23) Szewczak, A. A.; Kellogg, G. W.; Moore, P. B. *FEBS Lett.* **1993**, *327*, 261–264.
(24) Gonnella, N. C.; Birdseye, T. R.; Nee, M.; Roberts, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **1982**, *79*, 4834–4837.
(25) de Leeuw, F. A. A. M.; Altona, C. J. *Chem. Soc., Perkin Trans. 2* **1982**, 375–384.
(26) Feigon, J.; Leupin, W.; Denny, W. A.; Kearns, D. R. *Biochemistry* **1983**, *22*, 5943–5951.
(27) Hare, D. R.; Wemmer, D. E.; Chou, S.-H.; Drobny, G.; Reid, B. R. *J. Mol. Biol.* **1983**, *171*, 319–336.
(28) Haasnoot, C. A. G.; Heershap, A.; Hilbers, C. W. *J. Am. Chem. Soc.* **1983**, *105*, 5483–5484.
(29) Scheek, R. M.; Russo, N.; Boelens, R.; Kaptein, R.; van Boom, I. H. *J. Am. Chem. Soc.* **1983**, *105*, 2914–2916.
(30) Varani, G.; Tinoco, I., Jr. *Q. Rev. Biophys.* **1991**, *24*, 479–532.

called "anomeric–aromatic walks" that can be used for making sequential assignments. Helix-like, anomeric–aromatic walks often occur in irregularly paired regions as well.

Sometimes anomeric–aromatic walks can be assigned to sequences by the ordering of purine and pyrimidine bases implied. More often, anomeric–aromatic walks are placed by correlating them with assigned imino proton resonances. In helical regions, the AH2 proton of each AU base pair, which is identified by its UN3–AH2 NOE, gives NOEs to the H1′ protons of the nucleotides on the 3′ side of both the A and the U.[12] Similarly, at long mixing times, NOEs can be seen between GN1 imino protons in GCs and the H1′ protons of the residues on the 3′ sides of both the G and the C.[31]

Once H1′ assignments have been made, other ribose resonances can be addressed. Most H2′ resonances are assignable because H2′–H1′ cross peaks dominate the anomeric–ribose region of RNA NOESY spectra, and because H2′–aromatic walks exist analogous to the anomeric–aromatic walks just discussed. H3′ assignments are harder to make because most H2′–H3′ NOEs fall in the ribose–ribose region. They can usually be pieced together, however, from H3′–aromatic walks and from ^{31}P–^{1}H coupling information (see below). Bases-in assignment tends to stall at H3′ unless the RNA being examined is so small that its ribose–ribose region is resolvable. Complete proton assignments have been obtained in a few such cases, and even H5′s distinguished from H5″s.[33]

TOCSY and COSY experiments contribute relatively little because H1′–H2′ couplings are very small in ribose rings with C3′-endo puckering.[25] TOCSY data will correlate the H2′, H3′, and H4′ resonances of a C3′-endo ribose, but these correlations are often hard to identify because they fall in the ribose–ribose region. A C2′-endo puckered ribose gives useful TOCSY correlations, but only H1′, H2′, and H3′ chemical shifts are correlated because H3′–H4′ couplings are small. H1′–H2′–H3′–H4′ TOCSY correlations are seen only when sugars are exchanging rapidly between the two puckers. H4′–(H5′,H5″) coupling constants can be appreciable, but their magnitudes are strongly affected by backbone torsion angles and thus cannot be relied upon for assignments.

The risk of assignment error can be minimized by experiments that identify resonances by chemical type. AH2s can be recognized using inversion–recovery experiments because their T_1s are much longer than those of other aromatic protons. Natural abundance ^{13}C–^{1}H correlation experiments are very useful also.[34] Since the ^{13}C chemical shifts of RNA carbon atoms fall into (largely) nonoverlapping ranges determined by chemical type, proton resonances can be sorted by chemical type on the basis of the chemical shifts of the ^{13}Cs to which they are bonded. Selective deuteration can be used to sort H6 and H8 resonances by base type. d5-U, d5-C, d8-A, and d8-G are easy to prepare by exchange starting with protonated nucleosides (or nucleotides)[35–37] and to convert into RNA (see below).

The weaknesses of the bases-in approach are obvious. It seldom yields a complete set of assignments, and it is conformation-dependent. It can fail completely when the conformation of some part of an RNA does not lead to helix-like aromatic–anomeric NOE connectivities or, worse yet, leads to something that looks like the helical pattern but is not. Unhappily, the regions where assignment is difficult are always the most interesting.

^{31}P–^{1}H Spectroscopy: A Step toward Backbone-Out Assignments

Since the predominant isotope of phosphorus, ^{31}P, has a nuclear spin of $^1/_2$, all RNAs have a ^{31}P spectrum that contains (about) one resonance per residue. Assignment of an RNA's ^{31}P spectrum is useful because ^{31}P chemical shifts are sensitive to α and ζ,[38] and so ^{31}P resonances with unusual chemical shifts are associated with regions having unusual conformations. More important, every ^{31}P atom is scalar coupled to the 3′ proton of the ribose on the 5′ side of its phosphate group, and to the 5′ and/or 5″ protons of the ribose on the 3′ side of its phosphate. (3′ side ^{31}P/H4′ couplings are also sometimes large enough to observe.)

The first backbone-out, sequential assignment strategy proposed for nucleic acids was based on these couplings, but required that cross peaks be resolved in the ribose–ribose region, and so was applicable only to small RNAs.[32] A much less restrictive approach became possible when ^{31}P/^{1}H hetero-TOCSY experiments were developed by Kellogg.[39–41] These experiments can generate ^{31}P–H1′ and sequential ^{31}P–aromatic correlations. At worst, they allow one to assign ^{31}P resonances on the basis of anomeric and aromatic correlations. At best, something approaching a complete set of ^{31}P and ^{1}H assignments for a small RNA or DNA can emerge from a single experiment.[40,41]

^{31}P–^{1}H experiments are not the final answer, however. The three-bond couplings they exploit depend on conformation,[41] and the clearest sequential walks they generate in RNA involve phosphate–aromatic correlations that depend on interresidue NOES. Finally, the dispersion in the ^{31}P dimension of ^{31}P–^{1}H spectra tends to be poor because the prevalence of A-form torsion angles in RNA leads to resonance overlap.

The Heteronuclear, Backbone-Out Revolution

Starting in about 1990, scalar coupling methods for assigning protein spectra were developed that use samples uniformly labeled with ^{13}C and ^{15}N.[42–44]

(31) Heus, H.; Pardi, A. *J. Mol. Biol.* **1991**, *217*, 113–124.
(32) Pardi, A.; Walker, R.; Rapoport, H.; Wider, G.; Wuthrich, K. *J. Am. Chem. Soc.* **1983**, *105*, 1652–1653.
(33) See: Hines, J.; Varani, G.; Landry, S. M.; Tinoco, I., Jr. *J. Am. Chem. Soc.* **1993**, *115*, 11002–11003 and references therein.
(34) Varani, G.; Tinoco, I., Jr. *J. Am. Chem. Soc.* **1991**, *113*, 9349–9354.
(35) Benevides, J. M.; Lemeur, D.; Thomas, G. J., Jr. *Biopolymers* **1984**, *23*, 1011–1024.
(36) Brusch, C. K.; Stone, M. K.; Harris, T. M. *Biochemistry* **1988**, *27*, 115–122.
(37) Hayatsu, H. *Prog. Nucleic Acid Res. Mol. Biol.* **1976**, *16*, 75–124.
(38) Gorenstein, D. G., Ed. *Phosphorus-31 NMR*; Academic Press, Inc.: New York, 1984.
(39) Kellogg, G. W. *J. Magn. Reson.* **1992**, *98*, 176–182.
(40) Kellogg, G. W.; Szewczak, A. A.; Moore, P. B. *J. Am. Chem. Soc.* **1992**, *114*, 2727–2728.
(41) Kellogg, G. K.; Schweitzer, B. I. *J. Biomol. NMR* **1993**, *3*, 577–595.
(42) Bax, A.; Clore, G. M.; Gronenborn, A. M. *J. Magn. Reson.* **1990**, *88*, 425–431.

Resolution is routinely improved in these experiments by dispersing proton–proton cross peaks using associated heteroatom chemical shifts. In addition, assignments are conformation-independent because the one- and two-bond couplings these experiments exploit are largely independent of torsion angles. The RNA community was not slow to grasp the potential of this approach.

It is easy to prepare ^{13}C- and/or ^{15}N-labeled RNA samples. Bacterial cells are grown on a minimal medium in which ^{15}N NH_4^+ salts are the sole source of nitrogen and/or ^{13}C glucose[45] or ^{13}C methanol[46] is the sole source of carbon. The RNA recovered from these cells is digested to 5′ nucleoside monophosphates with RNAse P1 and then enzymatically converted into nucleoside triphosphates or nucleosides, the substrates for enzymatic and chemical RNA synthesis, respectively.[10,11,47,48]

A $^{13}C-^{15}N$-labeled nucleotide is a single-spin system. Its ribose protons can be completely assigned using HCCH experiments.[49] Experiments of the HCN and HCNCH class will correlate H1's with H8s and H6s,[50–53] and adenine H8s can be correlated with adenine H2s in much the same way.[54,55] Finally, experiments of the HCP class can generate conformation-independent, through-bond, backbone-based, sequential assignments.[56–58] ^{13}C and ^{15}N assignments are produced almost as a byproduct of this approach, and once those resonances are assigned, $^{13}C-^{13}C$ and $^{13}C-^{1}H$ couplings can be used to obtain torsion angles.

Many of the heteronuclear experiments just alluded to involve pulse trains that consume times that are significant compared to RNA T_2s. They are bound to be less effective for RNAs larger than the ones on which they have been demonstrated because of signal loss resulting from reduced transverse relaxation times. Heteronuclear labeling aggravates this problem by opening additional relaxation pathways in macromolecules. We find, for example, that ^{13}C-edited NOESY experiments done on specifically labeled RNAs have noticeably poorer signal to noise than unedited experiments done on comparable unlabeled samples (Kellogg, G. W., and Moore, P. B., unpublished observations). Also proton chemical shifts have an irritating tendency to correlate with ^{13}C chemical shifts, which reduces the dispersion achieved by labeling $^{1}H,^{1}H$ cross peaks with ^{13}C frequencies. In the end, heteronuclear methods may contribute more to the field by making complete assignment of small RNAs possible than by increasing the molecular weight limit.

Diagnosis of RNA Topology

Once proton resonances have been assigned, two of each nucleotide's seven torsion angles, χ and δ, can be determined. Nucleotides can be classified as anti or syn because intranucleotide H1′–aromatic NOEs are weak in the former case and strong in the latter, and ribose ring puckers can be identified as C3′ endo or C2′ endo (or some mixture of the two) on the basis of H1′–H2′ couplings.

The pairings of imino proton carrying bases in irregular regions can also be diagnosed using imino–other NOEs. Seventeen of the 20 possible base pairings involving Gs and Us put hydrogen-bonded imino protons within NOE distance of protons from partner bases.[17] The other three sometimes give diagnostic imino–other NOEs as well.[59] When more than one pairing is compatible with an imino–other NOE(s), the decision can often be made on the basis of glycosidic torsion angle. Opposing strands in a double helix or a closed loop are necessarily antiparallel, and in antiparallel structures, the ribose of one base cannot be superimposed on the ribose of its partner by rotation about an axis perpendicular to the plane of the pair. Thus pairings that are rotationally symmetric are unlikely unless one of the participants is in the syn conformation. (Exceptions have been found![59,60]) Remaining problems can usually be resolved by model building; one option will fit, and the other(s) will not.

Determination of RNA Conformations

Distance geometry algorithms provide the best route for proceeding from NMR data to macromolecular conformations (see the recent review of Brunger and Nilges[61]). The input includes the sequence of the RNA, a list of standard bond lengths and angles, a description of the molecule's base-pairing pattern, NOE-derived distances, and coupling-constant-derived torsion angles. The data are converted into information about the range of distances possible between all pairs of atoms in the molecule, and models are generated consistent with those ranges, which are then refined.

Refinement invariably involves the minimization of a structure's nominal energy as well as reduction of its deviation from the experimental data. Energy minimization is essential to ensure proper bond lengths and angles, and to prevent close contacts, but it does mean that the model which emerges is "invented" by the minimization program's energy functions to some degree. This would be acceptable if the potentials used were entirely accurate, but they are

(43) Bax, A.; Clore, G. M.; Driscoll, P. C.; Gronenborn, A. M.; Ikura, M.; Kay, L. E. *J. Magn. Reson.* **1990**, *87*, 620–627.
(44) Fesik, S. W.; Eaton, H. L.; Olejniczak, E. T.; Zuiderweg, E. R. P. *J. Am. Chem. Soc.* **1990**, *112*, 886–888.
(45) Nikonowicz, E. P.; Sirr, A.; Legault, P.; Jucker, F. M.; Baer, L. M.; Pardi, A. *Nucleic Acids Res.* **1992**, *20*, 4507–4513.
(46) Batey, R. T.; Inada, M.; Kujawinski, E.; Puglisi, J. D.; Williamson, J. R. *Nucleic Acids Res.* **1992**, *20*, 4515–4523.
(47) Haynie, S. L.; Whitesides, G. M. *Appl. Biochem. Biotechnol.* **1990**, *23*, 205–220.
(48) Batey, R. T.; Battiste, J. L.; Williamson, J. R. *Methods Enzymol.*, in press.
(49) Pardi, A.; Nikonowicz, E. P. *J. Am. Chem. Soc.* **1992**, *114*, 9202–9203.
(50) Farmer, B. T., II; Muller, L.; Nikonowicz, E. P.; Pardi, A. *J. Am. Chem. Soc.* **1993**, *115*, 11040–11041.
(51) Sklenar, V.; Peterson, R. D.; Rejante, M. R.; Feigon, J. *J. Biomol. NMR* **1993**, *3*, 721–727.
(52) Sklenar, V.; Peterson, R. D.; Rejante, M. R.; Wang, E.; Feigon, J. *J. Am. Chem. Soc.* **1993**, *115*, 12181–12182.
(53) Farmer, B. T., II; Muller, L.; Nikonowicz, E. P.; Pardi, A. *J. Biomol. NMR* **1994**, *4*, 129–133.
(54) Marino, J. P.; Prestegard, J. H.; Crothers, D. M. *J. Am. Chem. Soc.* **1994**, *116*, 2205–2206.
(55) Legault, P.; Farmer, B. T., II; Mueller, L.; Pardi, A. *J. Am. Chem. Soc.* **1994**, *116*, 2203–2204.
(56) Marino, J. P.; Schwalbe, H.; Anklin, C.; Bermel, B.; Crothers, D. M.; Griesinger, C. *J. Am. Chem. Soc.* **1994**, *116*, 6472–6473.
(57) Marino, J. P.; Schwalbe, H.; Anklin, C.; Bermel, W.; Crothers, D. M.; Griesinger, C. *J. Biomol. NMR*, in press.
(58) Heus, H. A.; Wijmenga, S. S.; van de Ven, F. J. M.; Hilbers, C. W. *J. Am. Chem. Soc.* **1994**, *116*, 4983–4984.
(59) Szewczak, A. A.; Moore, P. B. *J. Mol. Biol.* **1995**, *247*, 81–98.
(60) Wimberly, B.; Varani, G.; Tinoco, I., Jr. *Biochemistry* **1993**, *32*, 1078–1087.
(61) Brunger, A. T.; Nilges, M. *Q. Rev. Biophys.* **1993**, *26*, 49–125.

not. They seldom (never?) treat electrostatic interactions rigorously, for example.

What impact does energy minimization have on the structures that emerge? No general answer can be given: (1) the amount of experimental data available varies from one structure determination to the next; (2) error propagation is an inexact art in the NMR world, to put it mildly; and (3) there are no conventions about how contributions to molecular energy should be weighted during refinement relative either to experimental data or to each other. It follows that the superpositions of independently computed structures, which are used to illustrate the reliability of structures, cannot be compared from one paper to the next because the variation depicted depends on the variables just mentioned.

These issues are particularly sensitive ones for RNA spectroscopists because the data they obtain seldom determine their structures uniquely.[63] In our experience, 15−20 experimental constraints per nucleotide can be obtained from fully assigned RNA's proton spectra;[62] one often must settle for less. This is more than the minimum required (seven), but the experimental constraints are poorly distributed. They usually determine base placements well, but leave many backbone torsion angles virtually undetermined. In addition, inter-residue constraints that speak to the relative placement of residues that are unrelated by secondary structure are vanishingly few in number.

One symptom of the shortage of information is the high failure rate that accompanies the derivation of RNA models by distance geometry methods. A "failure" is a model that cannot be refined so that it has good covalent geometry, low energy, and good agreement with the NMR data. (Fortunately, in our experience, good geometry, low energy, and good agreement with the data have always gone hand-in-hand.) Failure occurs because the information supplied is insufficient to prevent the construction of models that include conformational discrepancies so serious that they cannot be refined away. Failure rates of 80−90% are common.

In many cases, what should be sought is not so much an accurate representation of an RNA's conformation as a sensible portrayal of the conformational implications of the NMR data. I feel that one should compute NMR-derived RNA structures that are as similar to A-form RNA as they can be, except where the data require otherwise, and represent them to the public as being no more than that. When computations are done this way, for example, those backbone torsion angles for which no direct experimental information is available, but which are unlikely to be other than A-form-like, are constrained to fall within the A-form range, and base pairs are (mildly) restrained to be planar.[17] In our experience, the failure rate drops, and models result that have the topology required and low energies, that violate few, if any, of the quantitative data, and that look the way crystallographically determined RNAs do. Another way to state it is that there have been highly A-like conformations within the universe allowed by the data for every molecule we have analyzed. Faut de mieux, it seems natural to regard them as best estimates for the structures in question.

It should be emphasized, on the other hand, that families of models calculated in this way give the impression that they are better determined experimentally than is actually the case. Clearly, the field must continue to search for ways to better constrain its RNA models experimentally. One would like to be able to arrive at structures that are so well determined that modest changes in the potentials used during refinement do not matter, and the assumption of A-likeness need not be invoked.

Conclusions

While the application of heteronuclear techniques should improve the quality of spectroscopically-determined RNA structures, only a few are likely to be determined with the accuracy protein spectroscopists regularly achieve. That said, it must be recognized that the aspects of an RNA's structure that can be determined definitively by NMR, that is, the way its bases are paired and its topology, are the ones that most concern nucleic acid biochemists and molecular biologists.

The enterprise is validated by the fact that results of general importance have already begun to emerge. It is becoming increasingly evident, for example, that natural RNAs contain irregularly paired, secondary structure motifs that are as characteristic as the double helix. The sarcin/ricin loop from ribosomal RNA, which consists of a GNRA tetraloop joined to a bulged G motif, is a good example.[59] Both motifs occur in many other contexts.[60,64] Our capacity to predict RNA conformations from sequences, which is a major goal of those who investigate RNA structure, will improve as the number of these special secondary structure motifs identified and characterized increases.

What the field needs most at this point, beyond what individuals can achieve by determining more structures and determining them better, are collective decisions on how the errors associated with NMR data should be assessed, on the conventions to be used when computing models, and on how the accuracy of models should be estimated.

I thank John P. Marino for his help with issues related to heteronuclear spectroscopy. This work was supported by a grant from the National Institutes of Health (GM-41651).

AR940073L

(62) E.g.: Cheong, C.; Moore, P. B. *Biochemistry* **1992**, *31*, 8406−8414.
(63) van de Ven, J. M.; Hilbers, C. W. *Eur. J. Biochem* **1988**, *178*, 1−38.
(64) Heus, H. A.; Pardi, A. *Science* **1991**, *253*, 191−194.