intermediate. Such bifunctional activation plays an important role not only in the hydrolysis of unactivated esters but also in the hydrolysis of nitriles, amides, and phosphate mono-, di-, and triesters. Amazingly, unactivated phosphate monoesters such as methyl phosphate can be hydrolyzed by tetrafunctional activation. A unifying theme for all of the above hydrolysis reactions is that one or more *cis*-diaqua metal complexes that can easily form four-membered rings are required. The ease of octahedral *cis*-diaqua Co(III) complexes in forming four-membered rings is highly sensitive to the tetramine ligand structure. With an appropriate tetramine ligand, a cobalt complex can bring about a 10 billion fold rate enhancement in the hydrolysis of phosphate diesters. Understanding the structure–reactivity relationship of these metal complexes is important for developing artificial restriction proteases[89-91] and nucleases[92] that can hydrolyze proteins and nucleic acids sequence specifically.

I am grateful to my talented students whose names appear in the original references. Financial support was provided by the Natural Sciences and Engineering Research Council of Canada and the U.S. Army Research Office.

(89) Hoyer, D.; Cho, H.; Schultz, P. G. *J. Am. Chem. Soc.* **1990**, *112*, 3249.
(90) Schepartz, A.; Cuenod, B. *J. Am. Chem. Soc.* **1990**, *112*, 3247.
(91) Rana, T. M.; Meares, C. F. *J. Am. Chem. Soc.* **1990**, *112*, 2457.
(92) Mack, D. P.; Dervan, P. B. *J. Am. Chem. Soc.* **1990**, *112*, 4604.

RNA Pseudoknots

JOSEPH D. PUGLISI, JACQUELINE R. WYATT, and IGNACIO TINOCO, JR.*

Department of Chemistry and Laboratory of Chemical Biodynamics, University of California, Berkeley, California 94720

Received November 26, 1990 (Revised Manuscript Received March 25, 1991)

The perception of ribonucleic acid (RNA) has undergone a radical change during the past 10 years. RNA is now known to be integrally involved in each step in the transmission of genetic information. The genetic code is stored in deoxyribonucleic acid (DNA) molecules and ultimately converted to the amino acid sequences of proteins. The DNA strand is first transcribed into a messenger RNA (mRNA) intermediate. The mRNA is then processed: large regions are excised by the spliceosome, a large protein and RNA complex. Finally, the mRNA is translated into a protein by the ribosome, another giant complex of RNA and protein. RNA not only serves as a carrier of the genetic information but also helps regulate and catalyze the reactions involved in protein synthesis.[1] RNA molecules alone can catalyze chemical reactions, such as cleavage and synthesis of phosphodiester bonds.[2]

The enviable specificity of biochemical reactions catalyzed by enzymes results from the specific three-dimensional structure formed by the protein's polypeptide chain. Like protein-mediated reactions, the specific biochemical functions involving RNA are guided by the three-dimensional folding of the polynucleotide. Unfortunately, relatively little is known about RNA conformations. RNA molecules form intramolecular structures; the polynucleotide chain folds back upon itself to form double-helical regions held together by Watson–Crick base pairs. The double helices are connected by single-stranded regions containing unpaired bases. The two-dimensional map of base pairs is called the secondary structure. The secondary structure of an RNA consists of double-helix regions, called stems, and single-stranded regions termed bulge loops (bulges), internal loops (bubbles), hairpin loops, and junctions (Figure 1).[3] Tertiary interactions are those between the elements of the secondary structure; they include tertiary base pairing, single strand–helix interactions, and helix–helix interactions.

A pseudoknot is a type of tertiary interaction that involves base pairing between nucleotides in a loop with nucleotides in a single-stranded region outside the loop.[4] Thus, a pseudoknot has two double-helical stem regions and two loop regions (Figure 2). The term pseudoknot was coined in order to exclude this type of interaction from the possible conformations considered by algorithms used to predict secondary structures. Inclusion of pseudoknots or knots prevents an exhaustive search of possible secondary structures (see Figure 3). If each of the ends of the RNA chain passes through a loop, then an overhand knot is formed; pulling on the two ends of the strand result in a knot. This possibility could arise only if both stems contained full turns of the helix. The biologically relevant interaction is called a pseudoknot; only one or neither of the ends extends through a loop, and a figurative pull on the ends does not lead to a knot. It is important to distinguish this common usage of the term knot from its definition in topology.[5]

Pseudoknots seem to be a widely used motif in RNA structure; they are proposed to have significant roles in a variety of RNA molecules.[6,7] They were originally

Joseph D. Puglisi received a B.A. degree in chemistry (1984) from The Johns Hopkins University and a Ph.D. in chemistry (1989) from the University of California, Berkeley. He is currently a postdoctoral fellow at the Institut de Biologie Moleculaire et Cellulaire in Strasbourg, France.

Jacqueline R. Wyatt received B.S. degrees in chemistry and secondary education (1984) from the University of Wyoming and a Ph.D. in chemistry (1990) from the University of California, Berkeley. She is a postdoctoral fellow in the Department of Molecular Biophysics and Biochemistry at the Yale University School of Medicine.

Ignacio Tinoco, Jr., received a B.S. in chemistry (1951) from the University of New Mexico and a Ph.D. in chemistry (1954) from the University of Wisconsin. He is a professor in the Department of Chemistry at the University of California, Berkeley.

(1) Lewin, B. *Genes IV*; Oxford University Press: Oxford, 1990.
(2) Cech, T. R. *Science* **1987**, *236*, 1532–1539.
(3) Wyatt, J. R.; Puglisi, J. D.; Tinoco, I., Jr. *BioEssays* **1989**, *11*, 100–106.
(4) Studnicka, G. M.; Rahn, G. M.; Cummings, I. W.; Salser, W. A. *Nucleic Acids Res.* **1978**, *5*, 3365–3387.
(5) Wasserman, S. A.; Cozzerelli, N. R. *Science* **1986**, *232*, 951–960.
(6) Schimmel, P. *Cell* **1989**, *58*, 9–12.

0001-4842/91/0124-0152$02.50/0

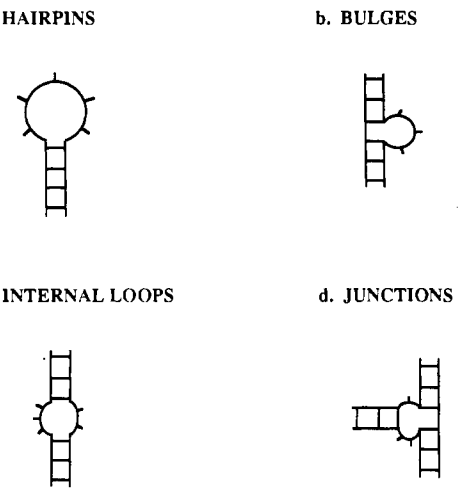

Figure 1. Base-paired and single-stranded regions of RNA form a limited number of structural elements, shown here. In principle, the loops and stems of these structural elements may consist of any number of nucleotides or Watson–Crick base pairs, respectively. A hairpin (a) is formed when the RNA molecule folds back on itself. The loop may contain as few as two nucleotides. A bulge (b) has unpaired nucleotides on only one strand; the other has uninterrupted base pairing. In an internal loop (c), at least one base is unpaired on each strand of the loop. In a mismatch, the smallest type of internal loop, bases that cannot form a Watson–Crick pair are in apposition. A junction (d) occurs when three or more double-stranded regions, separated by any number of unpaired nucleotides, adjoin.

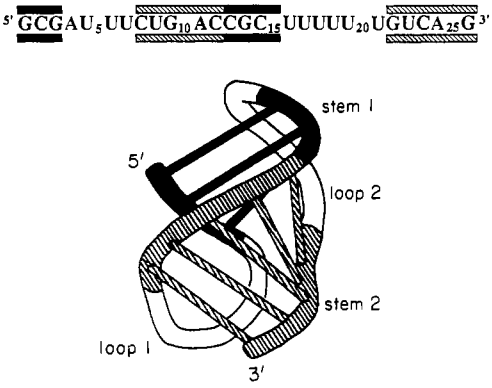

Figure 2. Schematic diagram of the folding of the oligonucleotide discussed in this paper; the sequence is shown at the top. Stem 1 of the pseudoknot is coded by black, stem 2 by hatching. The two stem regions have the potential to stack coaxially as indicated in the three-dimensional drawing. Loop 1, nearest the 5′-end of the molecule, crosses the major groove of stem 2, and loop 2 crosses the minor groove of stem 1. Due to the A-form geometry of the helices, the distance across the five base pairs of stem 2 is shorter than the distance over the three base pairs of stem 1.

discovered in plant viral RNAs. The 3′-ends of these viral RNAs mimic some of the functions of transfer RNA (tRNA). Although the secondary structure of the viral RNA is not the typical tRNA cloverleaf, a pseudoknot structure adopted by the viral RNA results in a three-dimensional structure similar to that of tRNA.[8,9] Pleij and co-workers proposed that the stacking of the two stems in the plant viral pseudoknot leads to a structure in which the separate stems resemble a single continuous helix;[8,10] this structure mimics the 12 base pairs formed by stacking of the amino acid acceptor helix and the TψC-helix in tRNA.

In order to confirm this model of pseudoknot structure, detailed physical studies were required. Biological RNA molecules are quite large; tRNA, one of the smallest, has about 75 nucleotides (molecular weight = 25 000). The plant viral RNAs contain thousands of nucleotides. Physical studies on small fragments of the viral RNA have been attempted.[11] However, even the smallest fragment (44 nucleotides) was still too large for detailed physical study. Our goal has been to design and synthesize RNA oligonucleotides that can fold into pseudoknot structures. This Account describes both the experimental techniques used to characterize pseudoknot structure and the results of our studies (for more details, see Puglisi et al., 1988,[12] 1990;[13] Wyatt et al., 1990[14]). Pseudoknots have proven to be a very useful model system with which to understand the relations among structure, stability, and function of RNA.

Experimental Methods for Studying RNA Structure

Other than X-ray crystallographic methods, no one experimental technique can give a global view of the folding of an RNA molecule. Our approach integrates results from a variety of techniques. Secondary structure at the nucleotide level can be obtained by using enzymes that cut only in single-stranded (for example, nuclease S_1 and RNase T_1) or double-stranded (nuclease V_1) regions of RNA molecules.[15] The RNA strand is labeled at either end with a radioactive phosphate and then subjected to limited hydrolysis by the enzymes; the sites of cleavage give a crude map of secondary structure. Chemical reagents like diethyl pyrocarbonate (DEP) and dimethyl sulfate (DMS) react with the various RNA functional groups that are sterically accessible; further treatment of the RNA results in cleavage of the backbone at the site of modification.[16,17] Despite certain problems with these methodologies, they are good preliminary tests for structure and aid in the interpretation of higher resolution data.[18,19] They are indispensible for probing conformations of large RNA molecules that cannot be studied by biophysical techniques.

Nuclear magnetic resonance provides the most powerful method for the elucidation of RNA structure in solution. A major drawback to NMR studies has been the large quantity of RNA needed (micromoles). The development of new synthetic methods has mitigated this problem.[20-22] NMR spectroscopy provides infor-

(7) Pleij, C. W. A. *Trends Biochem. Sci.* **1990**, *15*, 143–147.
(8) Pleij, C. W. A.; Rietveld, K.; Bosch, L. *Nucleic Acids Res.* **1985**, *13*, 1717–1731.
(9) Pleij, C. W. A.; van Belkum, A.; Rietveld, K.; Bosch, L. In *Structure and Dynamics of RNA*; Plenum Press: New York, 1986; pp 87–98.
(10) Dumas, P.; Moras, D.; Florentz, C.; Giegé, R.; Verlaan, P.; van Belkum, A.; Pleij, C. W. A. *J. Biomol. Struct. Dyn.* **1987**, *4*, 707–728.
(11) van Belkum, A.; Wiersema, P. J.; Joordens, J.; Pleij, C.; Hilbers, C. W.; Bosch, L. *Eur. J. Biochem.* **1989**, *183*, 591–601.
(12) Puglisi, J. D.; Wyatt, J. R.; Tinoco, I., Jr. *Nature* **1988**, *331*, 283–286.
(13) Puglisi, J. D.; Wyatt, J. R.; Tinoco, I., Jr. *J. Mol. Biol.* **1990**, *214*, 437–453.
(14) Wyatt, J. R.; Puglisi, J. D.; Tinoco, I., Jr. *J. Mol. Biol.* **1990**, *214*, 455–470.
(15) Ehresmann, C.; Baudin, F.; Mougel, M.; Romby, P.; Ebel, J.-P.; Ehresmann, B. *Nucleic Acids Res.* **1987**, *15*, 9109–9128.
(16) Peattie, D. A.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **1980**, *77*, 4679–4682.
(17) van Belkum, A.; Verlaan, P.; Kun, J. B.; Pleij, C.; Bosch, L. *Nucleic Acids Res.* **1988**, *16*, 1931–1950.
(18) Auron, P. E.; Weber, L. D.; Rich, A. *Biochemistry* **1982**, *21*, 4700–4706.
(19) Romby, P.; Moras, D.; Dumas, P.; Ebel, J.-P.; Giegé, R. *J. Mol. Biol.* **1987**, *195*, 193–204.

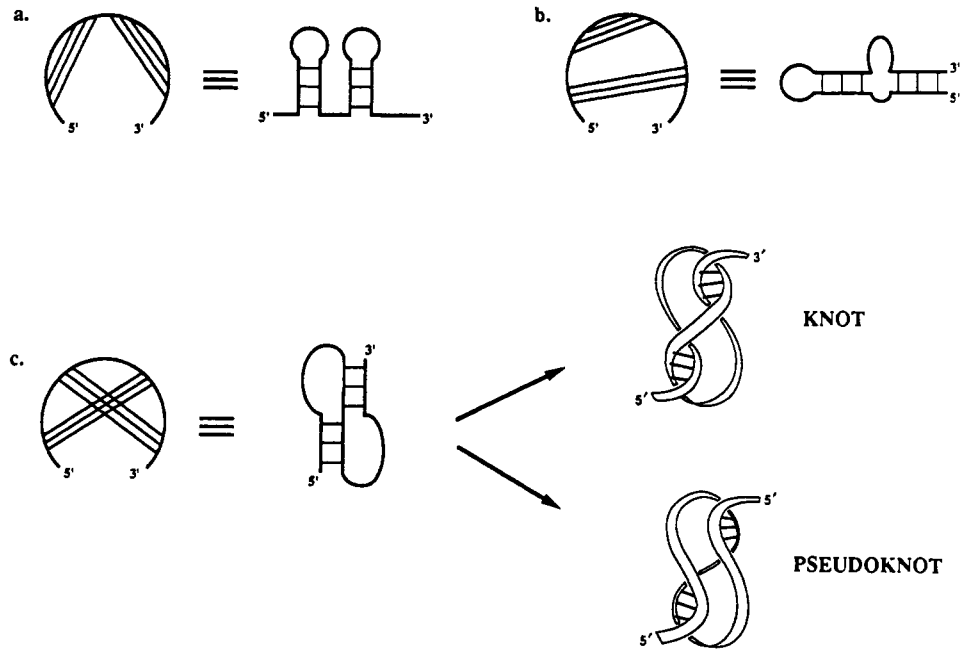

Figure 3. This figure illustrates the interactions allowed and disallowed in most RNA structure prediction algorithms. The RNA primary sequence is depicted as an arc, and base pairing is indicated by chords joining positions on the arc. Interactions demonstrated in parts a and b are allowed. Most algorithms do not consider interactions separated by regions of secondary structure; these interactions involve chord crossing as shown in part c. Chord crossing may result in formation of a "knot" (if each end of the RNA strand passes through a loop) or a "pseudoknot". A "knotted" structure has never been observed. In contrast to parts a and b, prediction of pseudoknots is difficult since the free energies for formation of the two stem and loop regions are not simply the sum of the free energies of the two hairpins.

mation about the interactions between nuclei.[23] For our studies, we relied almost exclusively on protons. Two protons that are close together in space (i.e., less than 5 Å) will interact by a through-space dipolar coupling mechanism; the resulting transfer of magnetization is manifested as a nuclear Overhauser effect (NOE). Intensities of NOEs are proportional to $1/r^6$, where r is the distance between protons.[23,24] NOEs yield short-range distance constraints between pairs of protons. Interaction through bonds (J coupling or scalar coupling) between protons on neighboring atoms leads to a splitting of the proton energy levels. The value of the coupling is dependent on the dihedral angle between the two protons and yields constraints on bond torsion angles.[25] Two-dimensional NMR methods were used to determine NOE interactions (NOESY) and through-bond couplings (COSY). With enough local distance and torsion angle constraints, useful structural models can be generated.

There are two types of protons in an RNA molecule (Figure 4): those attached to electronegative atoms (N, O) that exchange rapidly[26,27] with the aqueous solvent, and those attached to carbon that exchange very slowly (if at all) with solvent. The most useful exchangeable protons are the imino protons of guanine and uracil. When exposed to solvent at neutral pH, these protons exchange rapidly (millisecond time scale) with solvent. However, when hydrogen bonded in a base pair, they are relatively protected from exchange, and a resonance can usually be observed.[28] Imino resonances are observed in a region of the spectrum that is far downfield from nonexchangeable protons (11–15 ppm downfield of trimethylsilyl propionate); even for very large molecules such as tRNA, this part of the spectrum is relatively simple. Resonances corresponding to protons trapped in A·U base pairs resonate further downfield than do those corresponding to G·C pairs (Figure 5). NOE interactions should be observed between imino protons on adjacent base pairs, because the distance between stacked base pairs is only 3.4 Å in a helical structure. From NOE experiments and chemical shifts, one can determine the base-pairing pattern in an RNA oligonucleotide.

There are seven or eight nonexchangeable protons per nucleotide. Each has a resonance that must be assigned to a specific position in the sequence (Figure 4). This process, often complicated by spectral overlap, involves the use of both two-dimensional NOESY and COSY spectra to gain a map of through-space and through-bond interactions among protons. Characteristic patterns of interactions are observed in base-paired regions of RNA, which adopt the A-form helical conformation.[29] The overall helix geometry is determined by the nucleotide conformation. The RNA helix has a repeat of about 11 base pairs per turn. Due to a displacement of the bases away from the helix axis, there is a large difference in size between the two grooves of the helix. The major groove is deep and narrow. The minor

(20) Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res.* **1987**, *15*, 8783–8798.
(21) Davis, P. W.; Adamiak, R. W.; Tinoco, I., Jr. *Biopolymers* **1990**, *29*, 109–122.
(22) Sakata, T.; Hiroaki, H.; Oda, Y.; Tanaka, T.; Ikehara, M.; Uesugi, S. *Nucleic Acids Res.* **1990**, *18*, 3831–3839.
(23) Wüthrich, K. *N.M.R. of Proteins and Nucleic Acids*; John Wiley & Sons: New York, 1986.
(24) Keepers, J. W.; James, T. L. *J. Magn. Reson.* **1984**, *57*, 404–426.
(25) Karplus, M. *J. Am. Chem. Soc.* **1963**, *85*, 2870–2871.
(26) Hore, P. J. *J. Magn, Reson.* **1983**, *55*, 283–300.
(27) Plateau, P.; Guéron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310–7311.
(28) Leroy, J. L.; Kochoyan, M.; Huynh-Dinh, T.; Guéron, M. *J. Mol. Biol.* **1988**, *200*, 223–238.
(29) Haasnoot, C. A. G.; Westernink, H. P.; van der Marel, G. A.; van Boom, J. H. *J. Biomol. Struct. Dyn.* **1984**, *2*, 345–360.

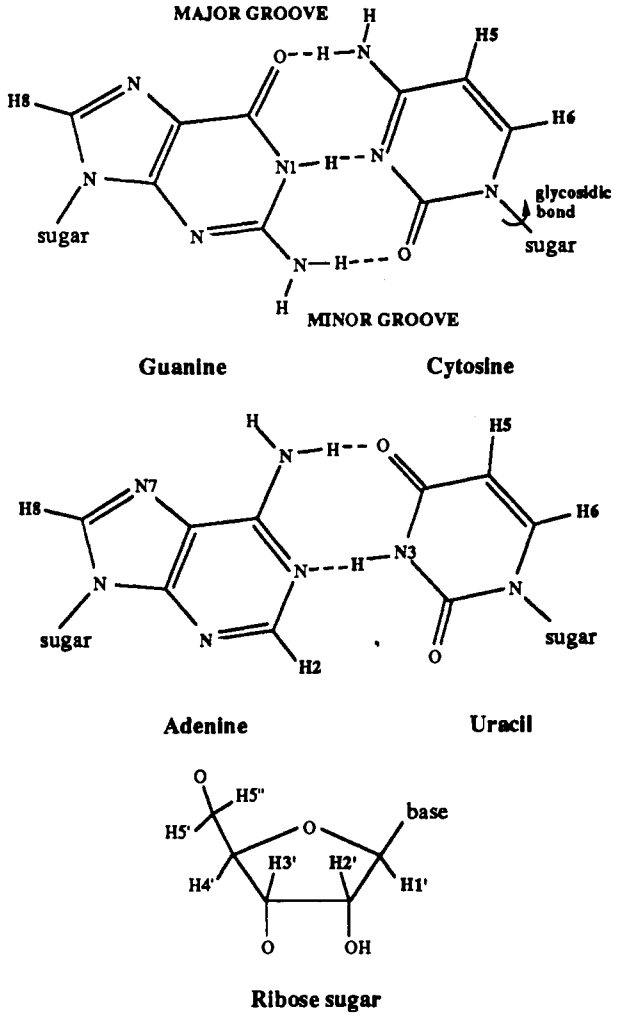

Figure 4. Base pairs in RNA with the protons on the bases and sugars numbered. The sugar protons are designated by primes. The nonexchangeable base protons are the guanine H8, adenine H8 and H2, and pyrimidine H5 and H6. The exchangeable imino protons on guanine (H1) and uracil (H3) are protected from exchange when involved in base pairing. The minor groove of the helix is at the glycosidic-bond side of the base pair. The major groove is opposite.

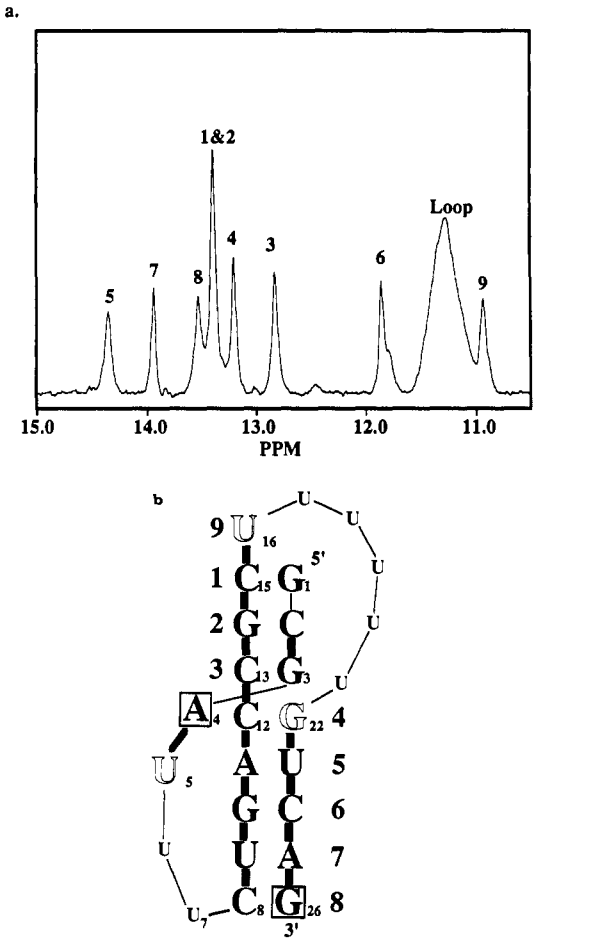

Figure 5. (a) The imino proton NMR spectrum of the pseudoknot. The numbers above the peaks correspond to the large numbers in the schematic in part b. The two resonances furthest downfield (at 13.92 and 14.35 ppm) are at chemical shifts characteristic of imino protons from A·U base pairs. The resonances above 11.5 ppm are due to non-hydrogen-bonded imino protons on the uridines from the loops of the pseudoknot. Resonances were assigned by using one-dimensional NOE measurements. The solution conditions were 5 mM $MgCl_2$, 50 mM NaCl, 10 mM Na phosphate, pH 6.4 at 23 °C. Reprinted with permission from ref 13. Copyright 1990 Academic Press. (b) The NMR data are summarized. Sugar conformation is indicated by letter type: C_2-endo, boldface; C_3-endo, outline; equilibrium between these conformations, box. Each nucleotide in large type adopts an anti glycosidic torsion angle. Thick lines between nucleotides indicate strong NOEs consistent with A-form stacking. Thin lines indicate no observed internucleotide NOE interactions. The conformations of the nucleotides in the loop regions are largely undefined by the NMR data. Reprinted with permission from ref 13. Copyright 1990 Academic Press.

groove, at the glycosidic bond side of the base pair (Figure 5), is shallow and wide.[30]

Two details of RNA nucleotide conformation can be defined fairly precisely by NMR data: the conformation of the ribose ring (sugar pucker) and the orientation of the base with respect to the sugar (glycosidic torsion angle). Sugar pucker generally falls into two low-energy families called C_3-endo or C_2-endo depending on which atoms are puckered in the direction of the aromatic base.[30] Within A-form helical structures, C_3-endo puckers are found. The sugar pucker of a nucleotide or the presence of an equilibrium between conformations can be determined from values of J-coupling interactions between sugar protons, since dihedral angles are different for the two conformers.[31] The distances between sugar protons and base protons depend strongly on the value of the glycosidic torsion angle.[23] From the intensity of NOEs between these protons, one can estimate distances. Even with relatively large experimental errors in distances (±0.5 Å), the glycosidic torsion angle can be determined to ±30°.

The global folding of an RNA molecule is determined by six backbone torsion angles per nucleotide. Four of them can be measured by proton–proton and phosphorus–proton coupling constants; this has been done for relatively short (12 nucleotide) sequences.[32] The phosphodiester torsion angles are undefined by proton and phosphorus coupling constants. In the studies on pseudoknots, we have relied on interproton distances between adjacent nucleotides. There are short internucleotide distances between aromatic H6 or H8 protons and neighboring sugar protons characteristic of

(30) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: Berlin, 1984.

(31) de Leeuw, F. A. A. M.; Altona, C. *J. Chem. Soc., Perkin Trans. 2* **1982**, 375–384.

(32) Cheong, C.; Varani, G.; Tinoco, I., Jr. *Nature* **1990**, *346*, 680–682.

A-form helical geometry; these allowed a qualitative analysis of strand conformation. In nonhelical regions, such NOEs may not be present, but fortunately, unusual internucleotide NOEs may be observed instead.[32,33] The global conformation of the strand can be best defined by using computer techniques such as distance geometry and constrained molecular mechanics. Extensive NMR data, with a large number of distance and torsion angle constraints, are needed to limit the number of structures.[32,34,35]

Pseudoknot Structure

In order to determine the conformation of a pseudoknot, a 26-nucleotide oligonucleotide (Figure 2) was synthesized and characterized.[13,14] This sequence was designed to form a pseudoknot with three base pairs in stem 1 and five base pairs in stem 2; the loop sequences were made sufficiently long to easily bridge the predicted distances across the helices.[8] The results of enzymatic and chemical mapping were consistent with pseudoknot formation; single-strand-specific nuclease S_1 cut only in the two loop regions, whereas double-strand-specific nuclease V_1 cut in the proposed stem regions. Chemical modification with DEP showed that the two adenines in the stem are protected relative to the adenine in loop 1.

NMR studies of the exchangeable imino protons confirmed the formation of the pseudoknot. As shown in Figure 5, eight imino resonances are observed, three corresponding to the base pairs in stem 1 and five corresponding to those in stem 2. An NOE between the imino protons of G_3·C_{13} (resonance 6) and C_{12}·G_{22} (resonance 5) confirms that the two double-helical regions are stacked on each other, unstacked base pairs would likely have an interproton distance greater than 5 Å, and thus no NOE would be observed. The two stems of the pseudoknot form a quasi-continuous helix. Most of the loop uridines have imino protons that exchange relatively rapidly with solvent. This gives rise to the broad lump in the exchangeable proton spectrum at about 11 ppm. However, one imino proton, from the first uridine (U_{16}) in loop 2, is protected from exchange. This nucleotide is stacked at the end of stem 1 on C_{15}.

Many of the nonexchangeable protons in the spectrum were assigned, including the aromatic H8, H2, H6, and H5 resonances and the sugar H1′, H2′, and H3′ resonances for all of the nucleotides in the stems of the pseudoknot. Resonances from protons on loop nucleotides were only partially assigned. Many of the residues in the loops are uridines, and this caused considerable spectral degeneracy.

The NMR results for this molecule are summarized schematically in Figure 5. The pattern of NOEs characteristic of an A-form helix was observed starting at nucleotide U_{16} and continuing, without indication of distortion, until nucleotide C_8. This confirmed the stacking of the two helical stems. In addition, nucleotide U_{16}, the first nucleotide in loop 2, is stacked upon C_{15}, the terminal nucleotide of stem 1, consistent with the exchangeable proton data discussed above. Even though strong NOEs characteristic of the A-form geometry are observed between C_2 and G_3 and from G_{22} through G_{26}, no strong NOEs between the nonexchangeable protons of G_3 and G_{22} are observed. These NOEs should be observed if the two helical stem regions were perfectly continuous. In addition, nucleotide G_{22} shows a majority population of $C_{2'}$-endo sugar pucker instead of the $C_{3'}$-endo pucker of A-form geometry. Thus, there is some distortion in the conformation between the two stems at these nucleotides. This is not surprising, since these nucleotides are the junction between the stem and loop regions and are not covalently joined. Model building has shown that if the two helical stems are stacked in an A-form geometry, then significant repulsive phosphate–phosphate contacts occur at this junction. The minor distortion in structure at this region probably relieves these unfavorable interactions.

The two loop regions of the pseudoknot are not equivalent. Loop 1 crosses the narrow (major) groove of stem 2, whereas loop 2 crosses the wide (minor) groove of stem 1. Unfortunately, the precise structure of these loops was poorly defined by the NMR data. It is clear that many of the loop nucleotides adopt $C_{2'}$-endo conformations. A change in sugar pucker from $C_{3'}$-endo to $C_{2'}$-endo increases the distance between adjacent phosphates from about 5.9 Å to 7 Å.[30] This increase in distance may be especially important for loop 2 nucleotides, which must traverse the wide minor groove. There is evidence of stacking between A_4 and U_5 in loop 1. Loop 1 is more protected from the single-strand-specific enzymatic probe nuclease S_1 than is loop 2. This suggests that loop 1 nucleotides are somewhat protected by the major groove of the helix.

Our results are consistent with the model originally proposed by Pleij and the more detailed model building of Dumas et al.[8,10] The two stem regions are coaxially stacked to form a semicontinuous A-form helix. Experiments have shown that valyl-tRNA synthetase, which aminoacylates the viral pseudoknot structure, interacts with the RNA along the region equivalent to C_8–C_{15} in our pseudoknot.[36] It is this side of the pseudoknot that most resembles a continuous helical structure, mimicking the acceptor stem of tRNA. Other structural features of the pseudoknot may have a biological function. The specific conformation of the loop regions may be sites of protein binding,[37] and the distortion in stacking may serve as a recognition element.

Pseudoknot Stability

The structure of pseudoknots is one feature of biological importance; the thermodynamic stability of the structure may also serve a biological function. The stability of a structure is measured by the free energy difference between folded and unfolded forms. For relatively short sequences, this free energy difference can be determined by using UV absorbance (260 nm) measured as a function of temperature.[38] The assumption is made that there are no intermediates in the transition between folded and denatured forms; thus, the resulting curves (melting curves) of absorbance vs temperature can be converted into equilibrium con-

(33) Puglisi, J. D.; Wyatt, J. R.; Tinoco, I., Jr. *Biochemistry* **1990**, *29*, 4215–4226.
(34) Pardi, A.; Hare, D. R.; Wang, C. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 8785–8789.
(35) Metzler, W. J.; Wang, C.; Kitchen, D. P.; Levy, R. M.; Pardi, A. *J. Mol. Biol.* **1990**, *214*, 711–736.
(36) Florentz, C.; Giegé, R. *J. Mol. Biol.* **1986**, *191*, 117–130.
(37) McPheeters, D. S.; Stormo, G. D.; Gold, L. *J. Mol. Biol.* **1988**, *201*, 517–535.
(38) Puglisi, J. D.; Tinoco, I., Jr. In *Methods in Enzymology, RNA Processing*; Academic Press: New York, 1989; Vol. 180, pp 304–325.

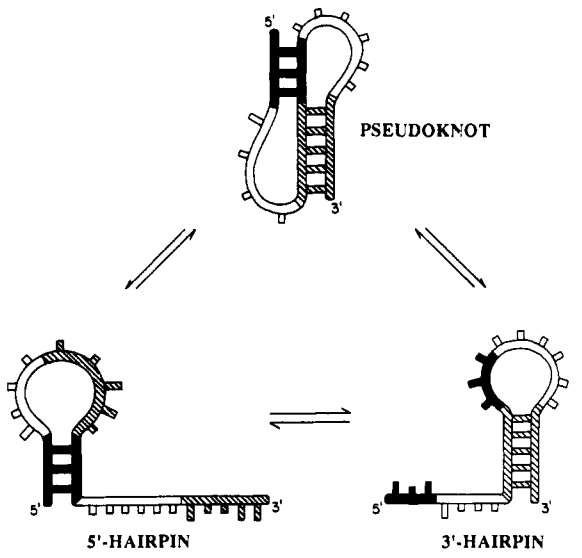

Figure 6. Each sequence capable of forming a pseudoknot structure also has the potential to form two hairpins: the 5′-hairpin, when only stem 1 forms, and the 3′-hairpin, when only stem 2 base pairs form. The equilibria are affected by temperature, cation concentration, and RNA sequence. Reprinted with permission from ref 14. Copyright 1990 Academic Press.

stants and thermodynamic parameters.

RNA sequences can often form alternate structures of similar stability.[39] Pseudoknot sequences have two such alternate structures, as shown in Figure 6. For the pseudoknot to form, it must be more stable than either the 5′- or 3′-hairpin conformation. We have measured the thermodynamic stability of the pseudoknot and compared it to the stability of the structures formed by RNA fragments that form the 5′-hairpin or the 3′-hairpin. In 5 mM Mg^{2+}, 60 mM Na^+ at pH 6.4, the pseudoknot structure is slightly more stable than either alternate conformation, but only by about 2 kcal mol^{-1} at 37 °C. The enthalpy of formation is sensitive to the number of stacking interactions that are present. The enthalpy for the eight base pairs of the pseudoknot stems in a *continuous* helix is predicted to be −72 kcal mol^{-1}.[40] The measured value of the enthalpy of pseudoknot formation is only −50 kcal mol^{-1}. This discrepancy may be explained by the distortion in stacking observed in the structural studies.

RNA tertiary foldings are often stabilized by binding of Mg^{2+} ions.[41] In 60 mM Na^+ at 15 °C in the absence of Mg^{2+}, the pseudoknot is destabilized relative to the 5′-hairpin. The imino proton NMR spectrum of the RNA oligonucleotide under these conditions shows a second set of resonances corresponding to those of the 5′-hairpin conformation.[14,33] At 30 °C, the equilibrium shifts completely to the 5′-hairpin conformation. Addition of Mg^{2+} shifts the equilibrium back toward pseudoknot conformation. Specific binding of Mg^{2+} to the pseudoknot possibly counteracts the unfavorable phosphate repulsions at the junction of the two stems.

Predicting Pseudoknots

One goal of biophysical chemists has been the prediction of the structure of an RNA molecule given only its primary sequence. The prediction of the lowest free energy structure is conceptually simple, since it can be approximated as the sum of the free energies for forming each element of structure.[42] In practice, even prediction of base-pairing interactions can be computationally demanding.[39] Including interactions like pseudoknots would make the algorithms more complex and would require much more computer time. In addition, the free energy is no longer a sum of contributions from separate regions (see Figure 3). For the prediction of pseudoknots, the thermodynamic contributions of pseudoknot loop sizes and sequences and stem sizes and sequences must be known. This is a demanding task since pseudoknot stabilities must be compared with stabilities of the potential hairpin conformations. We have concentrated on one aspect of pseudoknot structure and stability: the effect of changes in loop size and sequence.

If one assumes A-form stacking for the two stem regions of the pseudoknot, one can calculate the distance that each loop must cross to bridge the helical grooves.[8] This distance depends only on the number of base pairs in the stems. Across the major groove, the distance is at a minimum of about 10 Å for six or seven base pairs and remains under 20 Å for stems of two to nine base pairs. In contrast, across the minor groove the distance is a minimum of 18 Å for two or three base pairs and increases sharply as the stem is lengthened. The minimum number of nucleotides is therefore different for each loop and dependent on stem size. If the loop nucleotides adopt $C_{2'}$-endo conformations, then two nucleotides should suffice in loop 1 of the pseudoknot we have studied, but at least three nucleotides will be required in loop 2.

The results of variation in loop size generally confirm the predictions made above. The stem sequences were the same for all molecules, and the loops were shortened independently. Conformations were determined with chemical and enzymatic mapping and, in a few cases, NMR of the exchangeable protons.[14] The pseudoknot was the only conformation present when loop 1 was shortened to as few as three nucleotides, or when loop 2 was shortened to four nucleotides. When either of the loops was further shortened to the predicted minimum, an equilibrium resulted between pseudoknot and hairpin conformations. When the sizes of pseudoknot loops are reduced, the sizes of the hairpin loops are correspondingly reduced. Smaller loops generally lead to more stable hairpin structures and thus an increase in their stability relative to that of the pseudoknot structure.

The conformational equilibria depend strongly on loop sequence. Although a significant population of the molecules form pseudoknots when loop 2 is shortened to three uridines, almost all molecules adopt the 3′-hairpin conformation when the sequence of loop 2 is changed to -UCA-. The A residue in this loop was not accessible to DEP modification, suggesting that it is stacked on the hairpin stem. This stacking results in enthalpic stabilization of the hairpin structure. A major driving force for pseudoknot formation is the enthalpy gained upon formation of additional base pairs. If the loop conformation of the hairpin is well stacked, this driving force is lost.

(39) Zuker, M. *Science* **1989**, *244*, 48–52.
(40) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167–192.
(41) Stein, A.; Crothers, D. M. *Biochemistry* **1976**, *15*, 160–168.
(42) Tinoco, I., Jr.; Borer, P. N.; Dengler, B.; Levine, M. D.; Uhlenbeck, O. C.; Crothers, D. M.; Gralla, J. *Nature, New Biol.* **1973**, *246*, 40–41.

The stability of the pseudoknot structure may also be altered by small sequence changes. If a single nucleotide is removed from loop 1 (A_4), the pseudoknot formed is slightly more stable than the original pseudoknot. The adenine residue may be involved in unfavorable steric contacts with bases in stem 2. These examples illustrate the delicate nature of the conformational equilibria among pseudoknots and alternate hairpins. Small changes in salt concentration are also sufficient to affect these equilibria.[14]

Tertiary interactions are not predicted by the most widely used structure prediction algorithms. For pseudoknots, most of the free energy change (relative to the single strand) upon structure formation results from formation of the secondary structure. The free energy difference between the pseudoknot conformation and the hairpin conformations is small. This means that the structure of an RNA molecule can be predicted in two steps: first, the secondary structure is predicted,[29] then possible tertiary structures (i.e., pseudoknots) are considered. For a pseudoknot to form, the favorable free energy from adding base pairs must outweigh the unfavorable free energy of forming the two loops. Our data yield an approximate value of +10 kcal mol^{-1} for the unfavorable contribution of the two pseudoknot loop regions.

Pleij and co-workers have taken a different approach to predicting RNA secondary structure.[43] Whereas the algorithms referred to above attempt to search all possible foldings for the lowest free energy structure, Pleij's algorithm uses a sequential process. Structures evolve from unfolded to final folded form through addition of the most stable stem regions one at a time. This algorithm can reproduce with some success the pseudoknot structures in certain viral RNAs. Pleij has estimated a value of +4.2 kcal mol^{-1} for the free energy contribution of a single loop region in a pseudoknot; this is close to the value of +5 kcal mol^{-1} obtained from our studies. Thus, progress is being made toward the prediction of pseudoknot structures.

Concluding Remarks

The structure of a pseudoknot, in particular the stacking of the two stem regions, explains its function within the plant viral RNA. The stability of pseudoknots and conformational equilibria involved in their formation are also crucial to understanding their function. Although a pseudoknot may resemble a continuous helical structure, it is much less stable. This may be the reason for its role in the control of translation in some mRNA molecules. A pseudoknot can cause a frame shift in the translation of the genetic code in the mRNA.[44] For control of these processes, it is important to be able to change the conformation of the mRNA;[45,46] ligand binding, by Mg^{2+} or protein, can affect the equilibria of these transitions.

Further studies of pseudoknots are required. In particular, little is known about the conformations of the loop regions. The effect of stem size on pseudoknot structure and stability was not addressed in our studies. We have studied a particular kind of pseudoknot, with short loop regions and adjacent stem regions. Pseudoknots have been proposed with large loop regions, which contain their own secondary structures.[47-49] Unusual pseudoknot structures have been proposed as regulatory elements in certain mRNAs.[50,51] These families of pseudoknots must be studied in detail.

NMR in combination with other physical methods has proven to be a potent means of structure determination for RNA molecules. Application of newer experimental techniques, including use of nuclei other than ^1H, promises further advances in the study of RNA structure, possibly allowing detailed studies of larger RNA molecules. High-resolution studies have been limited to RNA molecules of less than 30 nucleotides. This allows characterization of individual folding domains of RNA molecules, but not the biologically relevant intact molecules. Studies of model systems are slowly yielding a detailed picture of the important elements of RNA structure and will provide a broader basis for our understanding of RNA structural stability and dynamics.

The research described was supported by grants from the National Institutes of Health and the Office of Health Effects Research of the Department of Energy.

(43) Abrahams, J. P.; van den Berg, M.; van Batenburg, E.; Pleij, C. *Nucleic Acids Res.* **1990**, *18*, 3035–3045.
(44) Brierley, I.; Digard, P.; Inglis, S. C. *Cell* **1989**, *57*, 537–547.
(45) Yanofsky, C. *Nature* **1981**, *289*, 751–758.
(46) Altuvia, S.; Kornitzer, D.; Teff, D.; Oppenheim, A. B. *J. Mol. Biol.* **1989**, *210*, 265–280.
(47) Moazed, D.; Noller, H. F. *Nature* **1987**, *327*, 389–394.
(48) James, B. D.; Olsen, G. J.; Liu, J.; Pace, N. R. *Cell* **1988**, *52*, 19–26.
(49) Gutell, R. R.; Woese, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1990**, *87*, 663–667.
(50) Tang, C. K.; Draper, D. E. *Cell* **1989**, *57*, 531–536.
(51) Philippe, C.; Portier, C.; Mougel, M.; Grunberg-Manago, M.; Ebel, J.-P.; Ehresmann, B.; Ehresmann, C. *J. Mol. Biol.* **1990**, *211*, 415–426.