Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Structural Elements in RNA

MICHAEL CHASTAIN
AND IGNACIO TINOCO, JR.

University of California
Berkeley, California 94720

The discovery that RNA molecules can catalyze phosphodiester bond cleavage and formation (*1*, *2*) dramatically changed our view of RNA function and greatly increased interest in RNA structure. There are several examples of RNA molecules that catalyze the cleavage, and in some cases the formation, of phosphodiester linkages. The first discovery of catalytic RNA was the generation of active rRNA from pre-rRNA in *Tetrahymena* by self-splicing—the excision of a region of RNA from the middle of a precursor RNA and ligation of the two ends to give an active rRNA and an intron (*1*). Similar reactions have been observed in a number of organisms for other introns (*3*, *4*). Several other catalytic RNA structures that undergo self-cleavage, but not ligation, have been discovered. These include structures within the minus strand of tobacco ringspot virus (*5*, *6*), the human δ virus (*7*), and the "hammerhead" structure found in several organisms (*8*, *9*). In an intermolecular reaction, the RNA component of RNase P (EC 3.1.26.5) is sufficient to cleave mature tRNA from a tRNA precursor (*2*).

There is growing evidence that mRNA structure regulates processes such as transcription (*10*, *11*), splicing (*12*), translation (*13*, *14*), and mRNA decay

Progress in Nucleic Acid Research
and Molecular Biology, Vol. 41

(*15–17*). For example, protein synthesis is reduced in prokaryotic systems if the nucleotides within the Shine–Dalgarno region upstream from the initiation codon form a stable structure, since these nucleotides must pair with the 3′ end of the ribosome to initiate translation (*18–21*). The structure of the mRNA for topoisomerase gene 60 in bacteriophage T4 has a more dramatic effect on protein synthesis (*22*). A sequence of 50 nucleotides in the mRNA is skipped by the ribosome during translation so that amino acids corresponding to these nucleotides are not incorporated into the protein.

Different aspects of RNA structure influence biological processes. The three-dimensional structure of self-splicing RNAs arranges functional groups to catalyze specific phosphodiester bond cleavage and formation. The global folded structure of an mRNA is less likely to regulate translation since the ribosome unfolds the mRNA as it is translated; however, the local structure of the mRNA can influence the ability of the ribosome to bind and translate the mRNA.

The increased interest in RNA function has led to a corresponding interest in RNA structure. Two problems have slowed the characterization of RNA structure. One is that techniques for synthesizing RNA were limited in their ability to generate large quantities of many RNA sequences. The second problem is that biological RNA molecules, with the exception of tRNA, have not formed crystals suitable for X-ray diffraction. The development of enzymatic (*23*) and chemical (*24, 25*) methods for synthesizing large amounts of RNA oligonucleotides, plus the development of new techniques for determining RNA structure, have now led to the characterization of several RNA molecules.

Here we describe the RNA structural characteristics that have emerged so far. In those cases in which RNA studies are incomplete, studies of DNA are described with the rationalization that RNA structures may be analogous to DNA structures, or that the techniques used to study DNA could be applied to the analogous RNA structures. We focus on aspects of RNA structure that affect the three-dimensional shape of RNA and that affect its ability to interact with other molecules.

I. Secondary Structure

Folded RNA molecules are stabilized by a variety of interactions, the most prevalent of which are stacking and hydrogen bonding between bases. Watson–Crick base-pairing is usually thought of first, but the importance of base stacking is seen in the crystal structure of $tRNA^{Phe}$, where 71 of the 76 bases are stacked (*26*). Many interactions between backbone atoms also occur in the structure of tRNA, although they are often ignored when considering RNA structure since they are not as well-characterized as interactions be-

tween bases. Backbone interactions include hydrogen bonding and stacking of sugar or phosphate groups with bases or with other sugar and phosphate groups.

The interactions found in a three-dimensional RNA structure can be divided into two categories: secondary interactions and tertiary interactions. This division is useful for several reasons. Secondary structures are routinely determined by a combination of techniques discussed in Section II, whereas tertiary interactions are more difficult to determine. Computer algorithms that generate RNA structures can search completely through possible secondary structures, but inclusion of tertiary interactions makes a complete search of possible structures impractical for RNA molecules even as small as tRNA. Finally, the division of RNA structure into building blocks consisting of secondary or tertiary interactions makes it easier to describe RNA structures.

To distinguish tertiary interactions from secondary structure, the sequence of the RNA is drawn on a plane; the backbone forms a continuous closed boundary if a line is drawn joining the 5′ end to the 3′ end. Hydrogen bonding between bases in the sequence may be depicted by lines joining the bases; these lines must remain within the closed boundary. A secondary structure can be represented without any lines crossing. Tertiary interactions occur when lines cross; this is called chord crossing. Secondary and tertiary structures of $tRNA^{Phe}$ are shown in Fig. 1.

The secondary structure of RNA consists of duplex and loop regions that can be divided into six different types, as shown in Fig. 2: duplexes, single-

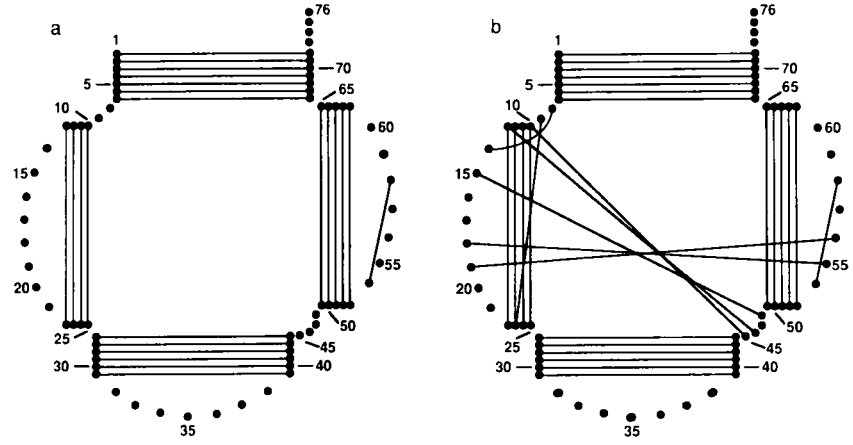

FIG. 1. The definition of secondary structure and tertiary structure in terms of chord crossing. The lines between points represent base-pairs. (a) The secondary structure of $tRNA^{Phe}$. (b) The secondary and tertiary structures of $tRNA^{Phe}$.

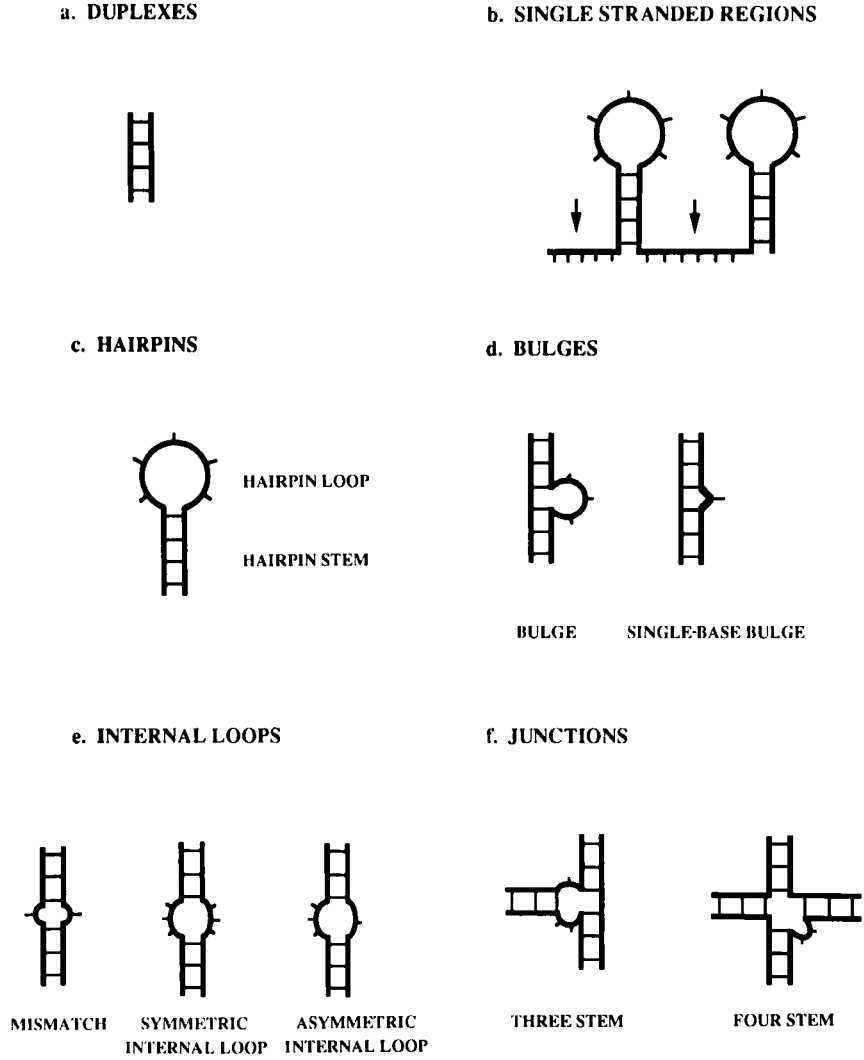

FIG. 2. Secondary structures of RNA.

stranded regions, hairpins, internal loops or bubbles, bulge loops or bulges, and junctions.

A. Duplexes

Duplex RNA forms a right-handed double helix stabilized by stacking between adjoining bases and by hydrogen bonds between bases on opposite, antiparallel strands. The conformations of double-helical regions in RNA

have been determined by X-ray diffraction studies of fibers and single crystals (*26–28*). The structures are all similar to that found for DNA fibers at low humidity; the geometry of these helices is termed "A-form." Proton NMR measurements on a variety of RNA oligonucleotides in solution are also consistent with A-form geometry (*24, 29–32*).

The A form of duplex RNA differs in several ways from DNA duplexes, which typically have B-form geometry in aqueous solution. The riboses in A-form RNA adopt a 3′-*endo* conformation, whereas B-form DNA deoxyriboses are in the 2′-*endo* conformation; with the usual phosphodiester backbone angles, the distance between phosphorus atoms in RNA is 5.9 Å, compared to a distance of 7.0 Å in B-form DNA (*26*). The base-pairs in A-form helices are tilted with respect to the helix axis and displaced from it by about 4 Å; this causes the minor groove in RNA to be wide and shallow, and the major groove to be very narrow and deep (Fig. 3). In crystals, the A-form RNA helix has 11 bp per turn, as opposed to 10 bp per turn of B-form DNA helix. In both RNA and DNA, the helices wind tighter in aqueous solution. The helical repeat of RNA in solution is reported (*34*) as 11.3 bp per turn or as 11.6 bp (*34a*), as compared to 10.6 bp per turn for DNA (*35–37*).

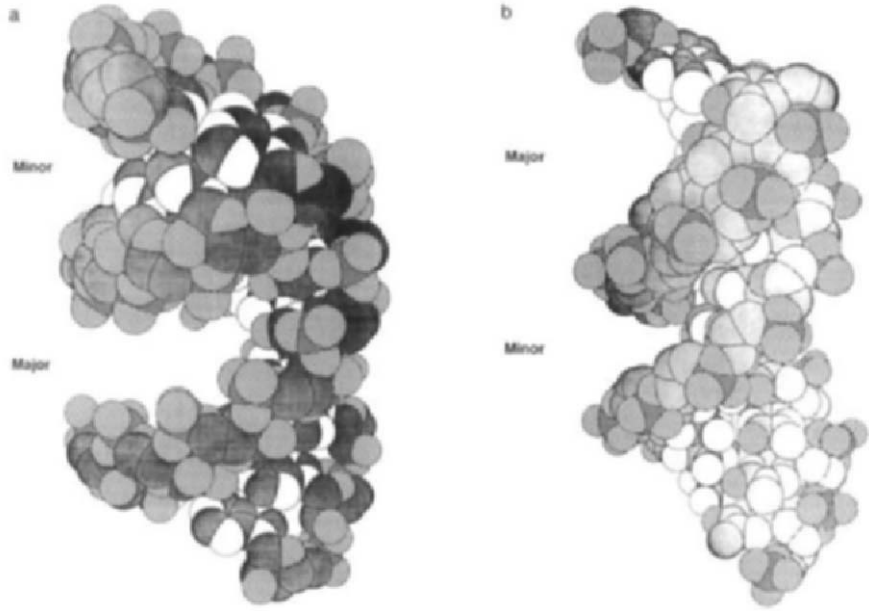

FIG. 3. A- and B-form helices tilted from the helix axis so that the groove sizes are visible. Fiber-diffraction coordinates are used (*27, 33*). Note (a) the deep and narrow major groove of A-form geometry compared to (b) the wide major groove of B-form geometry.

Although there are local variations in conformation depending on base sequence, RNA duplex regions are, in general, in standard A-form geometry. Unusual sequences may adopt other conformations. For example, Z-RNA is formed by alternating C·G base-pairs in the presence of high salt concentrations (*29*, *38*).

B. Single-stranded Regions

Single-stranded regions consist of unpaired nucleotides at the 5′ or 3′ end or between duplex regions of an RNA secondary structure. The conformation of these nucleotides differs from that of nucleotides in loop regions of secondary structure, since the ends of loop regions such as hairpins are constrained by the secondary structure. In the absence of tertiary interactions to constrain single-stranded regions, these regions are assumed to be roughly ordered by base stacking in a helical geometry similar to the structures of single-stranded RNAs that have no potential for base-pairing (*26*).

C. Hairpins

A hairpin consists of a duplex bridged by a loop of unpaired nucleotides. Hairpin loops are known to bind proteins, form tertiary interactions, and serve as nucleation sites for RNA folding. The conformation of the backbone in hairpin loops must differ from the conformation of the backbone in helical regions in order to reverse the direction of the RNA strand. The goal of studies on hairpin loops is to understand how the conformation of the loop nucleotides depends on the size and sequence of the loop.

The smallest loop capable of bridging a duplex was originally thought to be three nucleotides (*39*), but there is growing evidence that in some sequences, two unpaired nucleotides suffice to form hairpins in both RNA (*40*) and DNA (*41*). Thermodynamic studies of hairpins with loop sequences $(U)_n$, $(C)_n$, or $(A)_n$ (n ranging from 3 to 9) showed that loops containing four or five nucleotides are the most stable (*42*). Loops containing four unpaired bases are the most prevalent in 16-S rRNA (*43*). Eight different four-base-loop sequences (tetraloops) account for over 60% of the ribosomal tetraloop sequences (*44;* R. Gutell and O. C. Uhlenbeck, personal communication). Two of these sequences form unusually stable hairpins: UUCG (*45*) and GAAA (*44*, *46;* O. C. Uhlenbeck, personal communication). The interactions that stabilize particular loop sequences can be determined by examining the structures of hairpins determined by X-ray crystallography or NMR.

NMR studies of the hairpin GGAC(UUCG)GUCC show that interactions between loop bases and between loop bases and the sugar–phosphate backbone contribute to the unusual stability of the UUCG loop sequence (*40*, *40a;* see Fig. 4). U_5 and G_8 form a reverse wobble U·G pair (shown in Fig. 5)

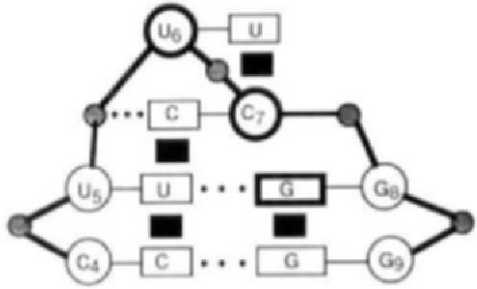

FIG. 4. The structure of an unusually stable hairpin with the loop sequence UUCG as determined by NMR. Interactions stabilizing the loop include a reverse wobble G·U base-pair with a *syn* G and a hydrogen bond between a cytosine amino proton and a phosphate oxygen. Thin circles represent C3′-*endo* sugars; thick circles represent C2′-*endo* sugars. Thin rectangles indicate *anti* bases; thick rectangles indicate *syn* bases. Dotted circles represent phosphates. Ellipses (. . .) indicate hydrogen bonds. Solid boxes represent stacking. (Reprinted from *40*.)

with the guanosine in the *syn* conformation. This leaves a loop of only two unpaired nucleotides: U_6 and C_7. The U_6 base stacks on the sugar of C_7, and the C_7 base stacks on the U_5 base. The C_7 amino proton is close to a 3′ phosphate oxygen of U_5, implying the formation of a base–phosphate hydrogen bond. The presence of this hydrogen bond is consistent with the 1.4 kcal/mol destabilization of the loop (in 10 m*M* buffer, pH 6.5, at 25°C) when the cytidine at position 7 is changed to uridine. NMR studies of the U_7 mutant show little change in loop conformation. G_8 is still *syn*, and most of the chemical shifts are the same as in the original molecule. The destabilization can be explained by the replacement of the favorable contact between the C_7 amino and a 3′ phosphate oxygen of U_5 by an unfavorable interaction of the U_7 keto with the phosphate oxygen. The loop is also destabilized by switching the closing base-pair from $C_4 \cdot G_9$ to $G_4 \cdot C_9$ (*45*); NMR studies have not yet been done to try to understand this effect.

NMR studies of three hairpins with small loops indicate that the conformation of the sugar–phosphate backbone throughout the loop is very different from A-form geometry. In the UUCG loop just described, U_6 and C_7 have 2′-*endo* sugar conformations different from the 3′-*endo* conformations of the A-form stem. Furthermore, the presence of two phosphate resonances in the downfield region of the spectrum indicates that some of the loop backbone torsion angles are different from the backbone conformation found in duplex regions. These changes enable the sugar–phosphate backbone of the small loop to bridge the duplex region. NMR studies of a hairpin with loop sequence UCU (*31*) and a hairpin with loop sequence UUU (P. Davis, personal communication) show similar behavior. The three nucleotides in these loops have 2′-*endo* sugar puckers, and some of the phosphate resonances are in the

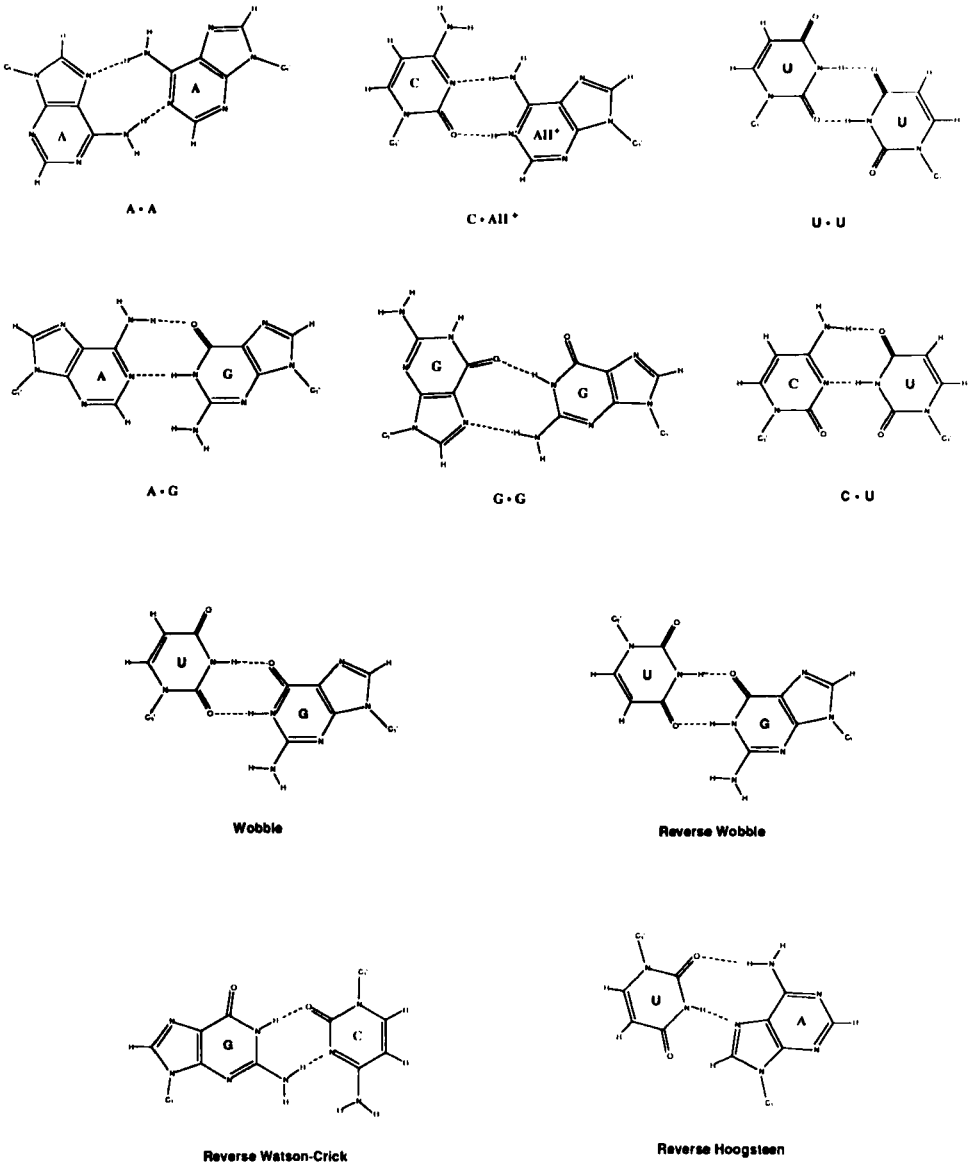

FIG. 5. Non-Watson–Crick hydrogen-bonding schemes. Most mismatches have more than one possible hydrogen-bonding scheme; a complete list is found in 26.

downfield region of the spectrum, indicating non-A-form backbone conformations. In the case of the hairpin with loop sequence UUU, the deviation from A-form geometry extended 1 bp into the stem region. We suspect that changes in the sugar–phosphate backbone, including 2′-*endo* sugar puckers, are a general property of small hairpin loops.

A-form geometry is preserved in portions of larger hairpin loops. As proposed prior to the crystal structure determination (*47*), five of the seven nucleotides in the anticodon loop of yeast $tRNA^{Phe}$ continue stacking on the 3′ strand (*26*). The stacking of the anticodon bases in A-form helix geometry leaves these nucleotides accessible for base-pairing with the mRNA. The five stacked nucleotides are followed by a turn in the sugar–phosphate backbone created by a change in the phosphodiester torsion angles between G_{34} and U_{33}. The turn structure is stabilized by a hydrogen bond between the imino proton of U_{33} and a 3′ phosphate oxygen of A_{35}, as well as stacking between the phosphate of G_{34} and the base of U_{33}. One-dimensional NMR studies of the anticodon loop of yeast $tRNA^{Phe}$ showed that the structure in solution is consistent with the crystal structure described (*48*), although two-dimensional NMR methods are needed to give unambiguous assignments.

A hairpin that occurs in wheat germ 5-S rRNA containing a loop of 12 nucleotides has been studied by NMR (*49*, *50*). A-form stacking does continue into the loop from the stem, but, unlike the anticodon loop, the stacking continues for several nucleotides along the 5′ strand. The loop is also stabilized by C·U mismatch hydrogen bonding (see Fig. 5) between the first two loop nucleotides.

The hairpins studied so far show that the stability of a hairpin loop changes with different loop sequences and sizes. The structures of the tRNA anticodon loop and the UUCG loop suggest that the specific loop sequences adopt conformations that are more stable because they contain more hydrogen bonding and stacking interactions—particularly interactions with the sugar–phosphate backbone. It has been suggested (*51*) that loop nucleotides stacked in A-form geometry along the 3′ strand are a common feature of RNA hairpins. The examples discussed here show that there is a range of loop structures. The anticodon loop contains nucleotides stacked along the 3′ strand, but the hairpin from 5-S rRNA has nucleotides stacked along the 5′ strand. In the three small hairpin loops—UUCG, UCU, and UUU—the backbone angles of each loop nucleotide differ significantly from A-form geometry. The conformations of the nucleotides in these small loops suggest that these nucleotides are not very accessible for base-pairing with other regions of RNA. Five of the nucleotides in the anticodon loop, on the other hand, are stacked in normal A-form geometry which facilitates pairing with other nucleotides.

The structures of hairpin loops determine how the loop nucleotides can

interact with regions of the same RNA molecule, with other RNA molecules, and with proteins. Further studies of RNA hairpins are needed to determine more about the interactions that stabilize hairpin loops. Helical regions are stabilized by hydrogen bonding and stacking between bases. The nucleotides in hairpin loops are stabilized by hydrogen bonding between base protons and phosphate oxygens, and by stacking between bases, sugars, and phosphates. The effect that the closing base-pair of the stem has on the stability and structure of hairpin loops must be determined.

D. Bulge Loops

Bulge loops (or bulges) are defined as unpaired nucleotides on one strand of a double-stranded region; the other strand has continuous base-pairing. Bulges range in size from one to many nucleotides. The main structural questions of interest are: (1) Does one unpaired base intercalate into the helix, or is it extrahelical, with the base-pairs on either side stacked on each other? (2) How much does a bulge bend or kink the helix? (3) What is the effect of very large bulge loops? Do the surrounding base-pairs break to form an internal loop instead of a bulge?

The local conformation of single-base bulges in several DNA oligonucleotides has been studied, but there has been very little work on RNA bulges. The equilibrium between a nucleotide bulge intercalating in or looping out of the helix depends on temperature, the identity of the bulge nucleotide, and the sequence of base-pairs in the duplex surrounding the bulge. This equilibrium is demonstrated by NMR studies on DNA oligonucleotides containing a thymidine bulge that loops out of the helix at 0°C and intercalates into the helix at 35°C when located between two guanosines, whereas a thymidine bulge located between two cytidines remains looped out of the helix independent of the temperature (52). NMR studies on an RNA duplex with a uridine bulge between two guanosines show that it loops out of the helix (53).

The local conformation of nucleotides in bulge loops containing more than one nucleotide have not been studied by high-resolution structural techniques. Determining the conformation of these loops in RNA is important for understanding how these nucleotides interact with other elements of secondary structure to form tertiary interactions and for understanding how proteins bind to bulge loops. The chemical reactivity of bulge loops in 16-S rRNA (54) has been studied. A bulge loop containing six nucleotides appears to exist without breaking the base-pairs on either side, since these pairs are protected from chemical modification. Some of the nucleotides within the bulge loop were also protected from modification, which suggests either that

the loop nucleotides form a structure involving stacking and hydrogen bonding, or that the nucleotides form tertiary interactions.

Bulge loops can affect the long-range structure of nucleic acid helices by creating a bend in the double helix. Bending has been detected by the altered mobility in nondenaturing gel electrophoresis for both RNA and DNA helices containing bulge loops (*55–57*). Bulge loops in DNA or RNA helices containing five adenosines were found to alter the gel mobility more than bulge loops containing five thymidines or uridines (*34, 55*), which shows that the structure of the bend depends on the identity of the bulge nucleotides. The bending was also shown to be affected by the base-pairs surrounding the bulge loop. The gel mobility for bulge loops of adenosines in DNA differed when the sequence flaking the bulge was d[AGG$(A)_n$TCG]·d[CGACCT] rather than d[CGA$(A)_n$CCT]·d[AGGTCG] (*55*).

One of the assumptions made when studying RNA secondary structure is that the structure is independent of the surrounding conformations. For example, a short region of duplex RNA is assumed to be the same whether it occurs next to a hairpin or next to a junction. Several studies suggest that bulge nucleotides affect the structure of the duplex surrounding them for several base-pairs. Thermodynamic studies showed that the free energies of bulge loops containing adenosines were 2 kcal/mol more stable (at 37°C) for the sequences GCG$(A)_n$GCGA·CGCCGCA than for GCG$(A)_n$GUCA·GACCGCA. Helices containing dangling nucleotides were more destabilized by bulge loops than were helices without dangling nucleotides (*58*). Nuclease V_1 (EC 3.1.22.3 or 3.1.27.8) cutting in duplex regions containing bulges and intercalated ethidium showed that ethidium intercalation into a double helix is affected for several base-pairs around a bulge nucleotide (*59*). A distortion in the duplex region near an adenosine bulge (in boldface) in the sequence d[CGCGAAATTTACGCG]$_2$ was observed by NMR. The presence of a phosphate resonance in the downfield region of the spectrum indicated an unusual backbone conformation. The downfield resonance was assigned to the 5′ phosphate of the guanosine one nucleotide removed from the bulge (*60*). Thus, the distortions due to bulges are not necessarily localized at the bulge, but may extend into the surrounding duplex region.

Many questions remain about bulge loop structures. Bulge loops bend RNA, but a correlation between the number and sequence of bases in the bulge and the extent of bending has not been established. Does a single base bulge that is looped out of the helix create a bend? Bulge loops containing more than one nucleotide have not been structurally characterized at high resolution. The structures of these loops determine the ability of bulge loop nucleotides to form tertiary pairs. Finally, the effect bulge loops have on the surrounding duplex structures must be determined.

E. Internal Loops

1. Mismatches

Two apposed nucleotides that cannot form a Watson–Crick pair are called mismatches. The two mismatched bases can engage in some other form of hydrogen-bonding, or they can form an open loop of two nonbonded nucleotides. Figure 5 shows some of the hydrogen-bonding schemes proposed for mismatches with two hydrogen bonds, although mismatches containing only one hydrogen bond have been found in tRNA (*26*). Mismatches occur frequently in proposed secondary structures. The stability, the contribution of the surrounding sequences to the stability, the hydrogen bonding, and the effect on the sugar–phosphate backbone of the various mismatches has not been determined systematically in RNAs.

The most-characterized mismatch is the wobble G·U pair (Fig. 5), which forms two hydrogen bonds and is virtually as stable as an A·U pair (*61*). Crystal structures of $tRNA^{Phe}$ and $tRNA^{Asp}$ containing G·U mismatches show that they are usually incorporated into the helix without creating distortions in the backbone. However, one of the G·U pairs in the crystal structure of yeast $tRNA^{Asp}$ does have an irregularity in the backbone. Two of the sugar–phosphate backbone angles (α and γ in Fig. 6) at the G·U pair in the anticodon stem are in the *trans* conformation instead of the normal *gauche* conformation found in helical regions (*62*). The distortion in the helix backbone is presumably caused by the difference in the width of a mismatch pair (C1′–C1′ distance) versus that of Watson–Crick base-pairs. The difference in width could also explain why G·U mismatches are slightly less stable when they are in the middle of a Watson–Crick region of a helix than at the end (*61*) and why many of the mismatches in proposed rRNA secondary structures are found at helix termini (*43*, *63*). A change in backbone

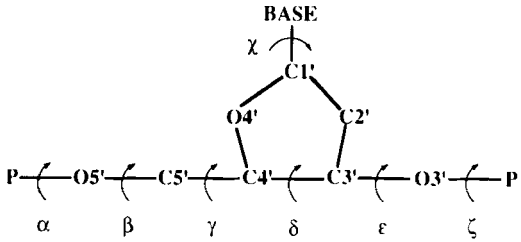

FIG. 6. The nomenclature describing the torsion angles that define the conformation of a nucleotide. Another parameter describing the sugar conformation is needed to completely define the conformation, but this parameter, the sugar puckering amplitude, has not varied in crystal structures of RNA and is usually assumed to be constant (*26*).

conformation may be a recurring feature of mismatches, although these distortions seem to depend on the sequence surrounding the mismatch.

The hairpin $GCGA_4UU_6(UCU)G_{10}AC_{12}CGCC$ has been studied by NMR (*31*). It has a stem of 6 bp, including $A_4^+ \cdot C_{12}$ and $U_6 \cdot G_{10}$ mismatches. Interproton distances by NMR were consistent with A-form stem geometry, even at the mismatches. The $A^+ \cdot C$ pair forms by protonation (at pH 6.5) of the adenine at the imino nitrogen, and is proposed to result in a pair containing two hydrogen bonds whose geometry is very similar to that of a G·U pair (Fig. 5). The free energy of formation of the $A^+ \cdot C$ pair, is 2 kcal/mol less stable than an A·U pair at 37°C.

Two hydrogen bonds form in the G·A mismatch (Fig. 5) found at the end of the anticodon helix in $tRNA^{Phe}$. G·A mismatches occur frequently at helix termini in rRNA (*63*), and hydrogen bonding in G·A mismatches has been proposed (on the basis of chemical modification) to occur in several internal loops in rRNA (*64, 65*). A symmetrical RNA duplex containing two G·A mismatches (in boldface) and 3′ dangling guanosines, $[GCGAGCG]_2$, is 1.9 kcal/mol more stable at 37°C than predicted by the nearest-neighbor parameters, suggesting that G·A mismatches from stable structures (*58*).

More data are available on mismatches in DNA than in RNA. The gel mobilities of DNA restriction fragments containing all 12 possible mismatches are unchanged compared to a fully Watson–Crick helix; this shows that mismatches do not cause significant bends in the helix axis (*66*). Thermodynamic studies on mismatches in the DNA oligonucleotides dCAAAXAAAG + dGTTTYTTTC showed that mismatches involving guanine (G·T, G·G, G·A) are the most stable, and mismatches involving cytosine (C·C, C·A) are the least stable (*67*).

2. Internal Loops

Internal loops contain three or more nucleotides not capable of forming Watson–Crick base-pairs; there is at least one unpaired nucleotide on each strand. Internal loops containing equal numbers of unpaired nucleotides on each strand are symmetrical, and those containing unequal numbers are asymmetrical. Comparison of computer-predicted secondary structures to phylogenetically derived ones suggests that asymmetrical internal loops are thermodynamically less stable than symmetrical ones (*68*). It is not known what determines whether internal loops are open or whether they close by forming non-Watson–Crick hydrogen bonding. The effect that internal loops, particularly asymmetrical ones, have on the helical backbone of RNA has not been determined.

Chemical modification of two asymmetrical internal loops in the S8 protein binding site of 16-S rRNA suggests that one is open and the other is closed by the formation of mismatch hydrogen bonding (*69*). The accessi-

bility of all of the nucleotides in a loop to either chemical or enzymatic modification is good evidence that the loop remains open. Uridines in the second loop were protected from modification, suggesting the formation of hydrogen bonds in a U·U mismatch (Fig. 5); the U·U pair, together with the accessibility of three adenosines in this loop to chemical modification, suggests that the loop closes with the three adenosines looped out of the helix.

NMR studies have been done on a symmetrical internal loop of *Escherichia coli* 5-S rRNA, known as loop E (*70*). The chemical shift and exchange behavior of the imino protons assigned to nucleotides in the loop indicate that the loop does not close by forming hydrogen bonds in G·G or G·A (Fig. 5) mismatch pairs. An open conformation was also found for a symmetrical internal loop similar to loop E of *Xenopus laevis* 5-S rRNA by NMR studies (*30*). This internal loop (Fig. 7) has the potential to close by the formation of hydrogen bonds in G·A mismatches, but the imino proton spectrum indicates that the nucleotides in the loop are not involved in hydrogen bonding. NMR studies on the nonexchangeable protons showed that the bases within this open loop stack extensively with only minor distortions from A-form geometry.

A uridine was added to this internal loop to make the sequence identical to that within loop E of *X. laevis* 5-S rRNA (B. Wimberly, unpublished

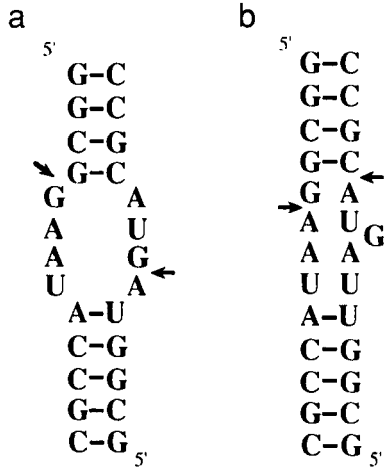

FIG. 7. Conformations of two internal loops, as determined by NMR. Breaks in NOE connectivities between adjacent nucleotides are marked by arrows. (a) An open internal loop structure is formed by an eight-base symmetrical internal loop (*30*). (b) Adding a single nucleotide to form an asymmetrical nine-base internal loop results in a closed structure with mismatch hydrogen bonding and a guanosine looped out of the helix (B. Wimberly, personal communication).

results). The structure of the asymmetrical internal loop which contains four nucleotides on one strand and five on the other (Fig. 7) is very different from the structure of the symmetrical one, despite the addition of only one nucleotide. NMR studies of the asymmetrical loop show that it closes with the guanosine on the longer strand looped out of the helix and hydrogen-bond formation in at least one of the mismatches in the loop. A closed structure for this loop was also proposed on the basis of chemical modification (*64*), although the guanosine thought to be extrahelical on the basis of NMR was proposed to be involved in a G·A pair (Fig. 5).

Determining where bends occur in RNA structures is important for understanding which regions of secondary structure are close together in the tertiary structure. It has been proposed in a model of the 16-S rRNA structure that a large bend occurs in a six-nucleotide symmetrical internal loop (*71*), but the effect of internal loops on the sugar–phosphate backbone in RNA has not been studied experimentally. Gel-mobility studies show that symmetrical internal loops of six or ten adenosines or thymidines in DNA bend the helix axis very much less than bulge loops (*55*).

The structures of internal loops in RNA remain poorly defined. The influence of loop sequence and size on the ability of internal loops to close by the formation of non-Watson–Crick hydrogen bonding must be determined. The accessibility of nucleotides in internal loops to tertiary pairing depends on the loop conformation. Extensive stacking of the bases within open internal loops may facilitate tertiary pairing, whereas bases in closed internal loops would be unavailable. Finally, the effects of symmetrical and asymmetrical internal loops on the sugar–phosphate backbone must be explored: Do bends occur at internal loops?

F. Junctions

Junctions, or multibranched loops, contain three or more double-helical regions with a variable number of unpaired nucleotides where the helical regions come together. Junction regions are important because helical regions can stack coaxially at these regions, and because the alignment of helical regions at junctions gives these regions characteristic shapes.

The only RNA junction whose conformation has been established is the four-stem junction in tRNA. For each tRNA for which a crystal structure has been determined, the acceptor stem stacks coaxially on the T stem, and the D stem stacks on the anticodon stem forming two long helical regions. These two helical regions are oriented roughly perpendicular to each other, creating the overall L shape of the molecule. Several of the unpaired nucleotides in the junction are involved in tertiary interactions, which are discussed in Section III.

An example of an RNA junction whose conformation plays a role in

protein binding is found in the 5-S rRNA from *X. laevis*. This three-stem junction contains four unpaired nucleotides critical for binding transcription factor IIIA. Mutations of the unpaired nucleotides in the junction greatly reduce protein binding (72), even though the protein does not directly contact these nucleotides (73). The implication of these studies is that the unpaired nucleotides in the junction act as a hinge controlling how the helical regions stack on each other, thus determining the overall three-dimensional conformation of the 5-S rRNA.

The conformation of a four-stem DNA junction, a model of recombination intermediates, has been determined (74). The conformation of the four-stem DNA junction was determined by measuring fluorescence energy transfer between acceptor and donor groups located on the four different helical regions. The four stems of the junction stack coaxially in pairs to form two long helical regions oriented in an X shape. The difference between the shapes of four stem-junctions in tRNA and in the four-stem DNA junction could be caused by the presence of unpaired nucleotides in the junction of tRNA.

On the basis of the tRNA structure, 5-S rRNA studies, and the DNA four-stem junction structure, it seems to be a general feature of junction regions that different stems stack coaxially to form longer helical regions. Coaxial stacking between helical regions in the 16-S rRNA has been proposed on the basis of phylogenetic comparison (75). One criterion for deciding whether two helical regions stack is that their combined length in base-pairs is conserved between different organisms, although the lengths of individual helices are not conserved. It has also been suggested (76) that the helical regions separated by the fewest unpaired nucleotides will stack coaxially. Both of these methods for predicting coaxial stacking predict more coaxially stacked helical regions than are found in a 16-S rRNA model based on chemical crosslinking (77). Experimental studies must be done to determine which factors, such as the number of stems and unpaired nucleotides in the junction, govern the stacking of different helices. In addition to stacking helical regions together, junction regions act as hinges that orient the different helical regions in space. The four-stem junction in tRNAs adopt an L shape and the four-stem DNA junction adopts an X shape. The conformations of more RNA junctions should be characterized in order to learn the various shapes that junctions can adopt.

Junction regions appear to constitute catalytic RNA sites. The hammerhead self-cleaving RNA region is coded for by newt satellite DNA (78) and is found in plant viruses (8), virusoids (79), and viroids (9). The consensus secondary structure for the catalytic domain contains a three-stem junction containing 11 unpaired nucleotides. Most of the unpaired nucleotides of the junction region are conserved (Fig. 8) and are required for the self-cleavage

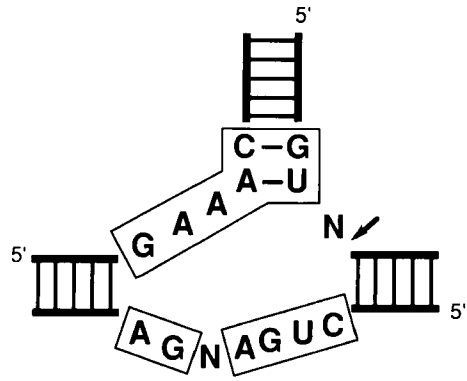

FIG. 8. The "hammerhead" self-cleavage RNA structure found in several organisms is shown, with the position of specific cleavage marked by an arrow and the conserved nucleotides boxed.

reaction (*80, 81*). The specific cleavage takes place between one of the unpaired nucleotides and a stem region (Fig. 8); a 2′,3′-cyclic phosphate and a 5′-OH group are formed (*9*). The catalytic activity must involve the conformation of the unpaired nucleotides in the junction since these comprise almost all of the required nucleotides.

Growing evidence implies that the unpaired nucleotides in a five-stem junction of the 23-S rRNA form the peptidyl transferase site of the ribosome, where peptide bonds are formed when amino acids are transferred from tRNAs to the nascent protein (*82*). Chemical "footprinting" experiments have established that unpaired nucleotides in this junction are protected from chemical modification when tRNA is bound to the peptidyl transferase site (*83*). Direct chemical crosslinking of the tRNA to the five-stem junction has also been seen (*84, 85*).

The examples of the hammerhead catalytic domain and the peptidyl transferase junction of 23-S rRNA illustrate the importance of determining the conformation of unpaired nucleotides in junction regions. Not only do the conformations of these nucleotides have a great impact on the three-dimensional structure by orienting the stem regions that meet at the junction, but also these nucleotides can be positioned to catalyze specific reactions.

II. Predicting Secondary Structure

Two techniques that are commonly used either separately or together to predict RNA secondary structure are phylogenetic comparison (*43, 86*) and thermodynamic stability (*39, 61*). Phylogenetic comparison is the method

currently used as the standard for determining secondary structure. The underlying principle is that mutations that do not alter function will be preserved. Since function is assumed to depend on structure, the preservation of a structure between different organisms, despite changes in the base sequence, is good evidence for the structure's existence.

A. Phylogenetic Comparison

Phylogenetic comparison requires that RNA sequences from several different organisms be known. The sequences are first aligned and then searched for regions capable of base-pairing. since helical regions are maintained if G·C pairs are replaced by A·U pairs or vice versa, covariance of nucleotides establishes which regions are involved in base-pairing and which are not. A helix is usually considered to exist if two or more covariations are found in it (*43*, *86*). In practice, organisms whose primary sequences differ by 20–40% give the best results with phylogenetic comparison (*86*). Sequences that are too dissimilar are difficult to align, and sequences that are too similar do not have enough compensatory base changes to establish the existence of helical regions. A limitation of the phylogenetic method is that it cannot provide any information about regions of secondary structure that contain conserved nucleotides. Thus, phylogenetic methods tend to predict fewer helices than actually exist in the molecule.

B. Thermodynamic Stability

Thermodynamic stabilities are used routinely to predict secondary structure. Computer algorithms predict structures by calculating the free energies for all possible base-pairing schemes and finding the secondary structure of lowest free energy. Computer algorithms that find several different secondary structures whose calculated free energies are close to the lowest free energy (*87*, *88*) are important for several reasons. One reason is that the calculated energies are based on incomplete experimental data and thus have significant uncertainties. Another reason is that biological RNA molecules begin folding as they are synthesized; they could become kinetically trapped in a structure that is not the structure of lowest energy. Furthermore, tertiary interactions may stabilize a secondary structure that is not calculated to have the lowest free energy. The most fundamental reason for calculating alternate secondary structures is that biological RNA molecules may not form a single secondary structure, but may instead have several structures in equilibrium. This has been suggested, for example, for the *cIII* gene mRNA of bacteriophage λ (*19*).

Free energies are calculated from experimentally determined parameters by the computer algorithms using a nearest-neighbor model. Since stacking interactions are short-range, it is reasonable to assume that the free energy of a helical region depends on the sequence of dinucleotide steps it

contains (39). This nearest-neighbor model has been tested for short duplexes with different sequences, but with the same set of dinucleotide steps; with only a few exceptions, the average agreement between the predicted and measured free energies is within 6% (*61*).

The free energies of loop regions (hairpins, internal loops, and bulges) are more difficult to quantitate. Originally, loop free energy was assumed to depend only on the number of unbonded nucleotides within the loop. This was based on free energies of loop formation of a small set of molecules with limited sequence variation (*89, 90*). There are now examples of loop regions whose free energy of formation depends markedly on the sequence of nucleotides within the loop. For example, the hairpin loop UUCG discussed above is much more stable than would be predicted by the current free-energy parameters. So is the internal loop containing two G·A mismatches (in boldface) formed by the duplex with 3′ dangling guanosines $[GC\mathbf{GA}GCG]_2$. As more thermodynamic data are collected for loop regions, more accurate formulas for predicting the free energy of loop regions will become possible.

Junction regions present bigger problems than other loop regions. No thermodynamic data have been measured for the free energy of junction formation in RNA. The present computer programs typically assume that the free energy of a junction region depends on the number of stems and the number of unpaired nucleotide within the junction (*88*). The values of the parameters are empirically derived to best fit the structures of RNA sequences whose secondary structures are well established.

Despite the problems with free energy of loop formation, current computer algorithms do remarkably well in predicting RNA secondary structure. The lowest free-energy structures calculated by the Zuker program for over 200 molecules predicted 70% of the helices deduced from phylogeny (*44*). Furthermore, the best structure within 10% of the lowest free energy predicted 90% of the helices correctly. The number of correctly predicted helices will surely increase as more free energies are measured for loop formation.

In practice, both phylogenetic comparison and computer algorithms are used to predict secondary structure. If only a few sequences that are not homologous enough to align for phylogenetic comparison are available, they can be folded with computer algorithms. A model secondary structure can be determined by searching for a common secondary structure among the various suboptimal foldings generated for the different sequences (*91*). Determining a secondary structure by either method is usually an iterative process. The secondary structure model is refined as more sequences are determined or as the model is compared to the results of chemical and enzymatic modification procedures that map the accessibilities of different nucleotides.

III. Tertiary Interactions

RNA molecules fold into compact structures stabilized by tertiary interactions not included in the secondary structure. Secondary and tertiary interactions are distinguished in terms of chord crossing (Fig. 1). Tertiary interactions are considered separately from secondary structure, since the formation of tertiary interactions depends not only on the nucleotides that form the tertiary interaction but also on the rest of the secondary structure. For example, the conformation about the four-stem junction in tRNA brings the T loop and the D loop close together; this allows tertiary pairing between nucleotides in these loops. The nucleotides in the anticodon loop cannot interact with either of the other loops, since the anticodon loop is fixed at one end of the molecule by the junction conformation.

The biological functions of RNA molecules—in particular, catalytic behavior—are determined by their three-dimensional structures. Characterizing all of the tertiary interactions that occur in an RNA molecule does not completely describe the three-dimensional structure, but tertiary interactions do indicate regions of the secondary structure that are close together in space. Most of what we know about tertiary interactions comes from a few tRNA crystal structures, but we are beginning to learn more about tertiary interactions from a variety of methods (discussed in Section IV) to study an increasing number of RNA molecules. The goal is to find recurring tertiary elements: specific arrangements of secondary structure elements stabilized by tertiary contacts. Here we discuss the three types of tertiary interactions so far characterized: tertiary base-pairing, single-strand–helix interactions, and helix–helix interactions.

A. Tertiary Base-pairing

1. Pseudoknots

Nucleotides that are unpaired in the secondary structure of an RNA molecule can form tertiary contacts by hydrogen bonding to other nucleotides unpaired in the secondary structure. In general, this could occur between any of the secondary structure regions containing unpaired nucleotides (single-stranded regions, hairpin loops, bulge loops, internal loops, and junction loops). All of these interactions were originally termed knots or pseudoknots (92); here, we use the term "pseudoknot" to describe only the structure in which nucleotides in a loop (hairpin, internal, or bulge) pair with nucleotides in a single-stranded region. Different types of pseudoknots are shown in Fig. 9 (93).

Pseudoknots have been found in an increasing number of biological sys-

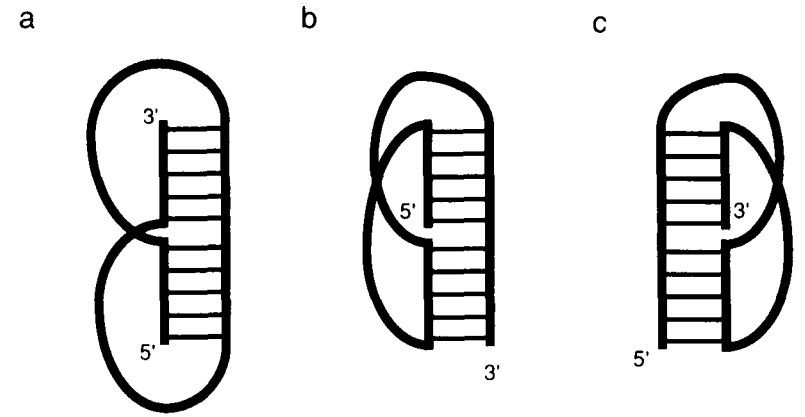

FIG. 9. The different types of pseudoknots are diagrammed for hairpin loops pairing with an adjacent single-stranded region. (a) In the best-characterized form of pseudoknot, the loop at the top crosses the major groove, and the loop at the bottom crosses the minor groove. (b) This form of pseudoknot, in which one loop crosses the major groove and the other loop bridges the whole helix, has not been found. (c) This type of pseudoknot was proposed to occur in the α-mRNA of *E. coli* (*94*). One loop crosses the minor groove, and the other loop bridges the whole helix.

tems since their discovery at the 3′ end of several plant RNAs (*95*, *96*). The viral RNAs were recognized by tRNA-specific enzymes, but the secondary structures predicted for these sequences did not include an amino-acid acceptor stem. The formation of a pseudoknot allows these molecules to fold into a tertiary structure functionally similar to tRNA, even though the secondary structures are different. Pseudoknot formation also enhances frameshifting during translation in the coronavirus IBV (*97*). The mechanism by which pseudoknots contribute to frameshifting has not been determined, but the combination of an (A+U)-rich sequence and a pseudoknot results in a frameshift. One possible explanation is that the pseudoknot causes the ribosome to pause, allowing slippage to occur at the A+U-rich sequence. This mechanism seems to be a general one, since 14 of 22 sequences from a variety of viruses, known or suggested to contain frameshifting sites, have the potential to form pseudoknots (*97*).

A pseudoknot of the type diagrammed in Fig. 9a formed by a short oligonucleotide has been studied by NMR (*98*). The interproton distances, determined by NMR, were consistent with the two stem regions, one containing 5 bp and the other containing 3 bp, with A-form helix geometry. The distances between protons located in the two different stem regions indicate that the two stem regions stack coaxially to form one continuous helical

region. The pseudoknot appears to be a normal duplex from one side, but on the other side, two loops bridge the duplex, one crossing the major groove and one crossing the minor groove. The size of the loops was varied without changing the stem sizes to show that the minimum loop size for crossing the minor groove of the 3-bp stem is three nucleotides, and the minimum loop size for crossing the major groove of the 5-bp stem is two nucleotides. The pseudoknot is only marginally more stable than either of the two potential hairpin structures that the sequence could form. The equilibrium between pseudoknot and hairpins depends on salt concentration, temperature, and nucleotide sequence (*99*).

2. Loop–Loop Interactions

There are several examples of RNA molecules containing tertiary contacts between nucleotides that are in loop regions of secondary structure. Phylogenetic comparison suggests that the RNA component of RNase P, which catalyzes the processing of tRNA precursors, makes Watson–Crick pairs between four nucleotides in one junction loop and four nucleotides in a second junction loop (*100*). These are interactions between secondary structure loop regions in the crystal structure of yeast $tRNA^{Phe}$. Two of the nucleotides in the D loop form parallel pairs with two nucleotides in the T loop: $G_{18}\cdot\psi_{55}$ and $G_{19}\cdot C_{56}$. Two additional tertiary base-pairs are formed between unpaired nucleotides in the central four-stem junction and nucleotides in the D loop: a reverse-Hoogsteen pair (Fig. 5), $U_8\cdot A_{14}$, and a parallel-stranded reverse-Watson–Crick pair (Fig. 5), $G_{15}\cdot C_{48}$ (*26*). These tertiary pairs are stabilized by hydrogen bonding and stacking interactions with adjacent nucleotides; they are partially responsible for stabilizing the L shape of tRNA. Tertiary pairs between loop regions have also been proposed on the basis of phylogenetic comparison in the 16-S and 23-S rRNAs (*101*, *102*).

The high frequency with which known RNA structures contain tertiary pairing between nucleotides that are unpaired in the secondary structure stresses the importance of learning more about such interactions. Can unpaired nucleotides in all of the secondary structure loop types engage in tertiary pairing? There are several examples of pairing involving nucleotides in hairpin loops and junction loops, but it is not known whether nucleotides in bulge loops or internal loops form tertiary interactions. The fact that tRNA has tertiary pairing between parallel strands shows that the rules for forming tertiary pairs can be different from the rules for secondary structure. The thermodynamics of tertiary pairs is of critical importance for predicting their existence. In the tRNAs and pseudoknots, tertiary pairs contribute less to the free energy than secondary structure does, but it may be possible for very stable tertiary interactions to replace secondary structure.

B. Single Strand–Helix Interactions

1. Intercalation

Base stacking is one of the most important factors stabilizing RNA structures. One of the ways in which stacking can stabilize the tertiary structure is for nucleotides that are unpaired to intercalate between base-pairs. An example of intercalation is seen in the crystal structure of yeast $tRNA^{Phe}$, where G_{57}, an unpaired nucleotide in the T loop, intercalates between two nucleotides of the D loop, G_{18} and G_{19}, which form tertiary pairs with nucleotides in the T loop. To accommodate the intercalated guanine between the two tertiary pairs, the sugar–phosphate backbone is extended by a change in the G_{19} sugar conformation to 2′-*endo* (*26*).

2. Base-triples

Base-triples occur when an unpaired nucleotide forms hydrogen bonds with a nucleotide that is already base-paired. The third base of a base-triple may bind to the Watson–Crick pair in either the major or minor groove and be stabilized by the formation of one or two hydrogen bonds as well as stacking interactions. Several biological functions have been proposed for base-triples. Base-triples stabilize the three-dimensional shape of tRNA and several other RNA molecules discussed below. The formation of DNA triples inhibits transcription *in vitro* (*103*), and it has been suggested that a small RNA molecule may inhibit transcription by forming a triplex 115 bp upstream from the transcription site of the human c-*myc* gene (*104*, *105*). The self-splicing intron from *Tetrahymena* binds guanosines via base-triple formation during the self-splicing reaction (*106*).

Three base-triples occur in $tRNA^{Phe}$, all of which involve nucleotides in the junction loop binding in the major groove to Watson–Crick pairs of the D stem. In two of the triples, $A_9 \cdot A_{23} \cdot U_{12}$ and $G_{46} \cdot G_{22} \cdot C_{13}$ (Fig. 10), the third base forms two hydrogen bonds with the purine of the Watson–Crick pair, while only one hydrogen bond is formed by the third base in $G_{45} \cdot G_{10} \cdot C_{25}$ (*26*). Different base-triples are found in the crystal structure of $tRNA^{Asp}$ (*62*); the $A_{46} \cdot G_{22} \cdot \psi_{13}$ triple replaces the $G_{46} \cdot G_{22} \cdot C_{13}$ triple of $tRNA^{Phe}$, with the adenosine forming one hydrogen bond with the $G \cdot \psi$ pair in the major groove. An unusual base-triple occurs at the beginning of the D loop. A_{21} binds to the tertiary reverse-Hoogsteen pair (Fig. 5) $A_{14} \cdot U_8$ by forming base–base and base–sugar hydrogen bonds. A_{21} forms one hydrogen bond with an amino proton of A_{14} and one hydrogen bond with the 2′hydroxyl of U_8; the 2′ hydroxyl of A_{21} forms a hydrogen bond with a base nitrogen of A_{14}. These base-triples in tRNA occur at the hinge region between the two helical domains and help stabilize the tRNA in its characteristic L shape.

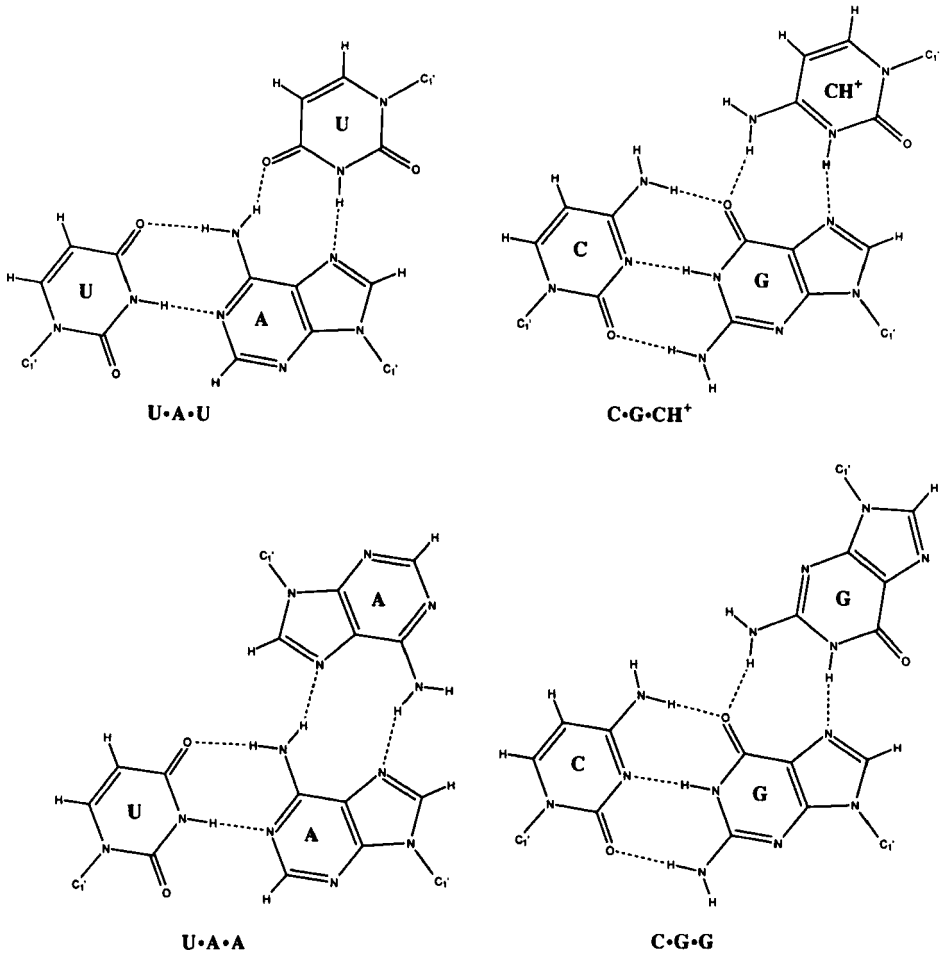

FIG. 10. Proposed hydrogen-bonding schemes for base triples. U·A·A and C·G·G base triples have been proposed with several different hydrogen-bonding schemes. For other schemes, see *26* and *108*. (Reprinted from *107*.)

Another RNA molecule that forms base-triples is the self-splicing intron from *Tetrahymena*. The intron binds a free guanosine during cleavage at the 5′ exon and binds an internal guanosine during cleavage of the 3′ exon. It is postulated that these guanosines bind to a base-pair, G_{264}·C_{311}, in the P7 stem by forming a base-triple (*106*). Replacing the G·C pair with an A·U pair abolished splicing activity, but cleavage at the 5′ exon was restored by adding 2-aminopurine instead of guanine. 2-Aminopurine can form a base-triple with the A·U pair that is isomorphic to the wild-type G·G·C triple.

This is strong evidence for the existence of the G·G·C triple in the splicing reaction of the wild-type intron.

A base-triple has also been proposed to occur between a nucleotide in a junction loop and an adjacent helix in *Xenopus laevis* 5-S rRNA (*109*). On the basis of chemical modification and model building, it is proposed that an adenosine in the junction loop binds to a G·U pair in the minor groove by forming two hydrogen bonds, one between the adenosine N7 and the 2′ hydroxyl of the guanosine, and one between the adenosine amino and a base nitrogen of the guanosine.

The formation of base-triples in junction regions where helices stack coaxially may be a recurring RNA structural element. If two helical regions in a junction stack coaxially (Fig. 11), it is a consequence of A-form helix geometry that the 3′ strand cannot reach the major groove of the adjacent helix, and the 5′ strand cannot reach the minor groove of its adjacent helix.

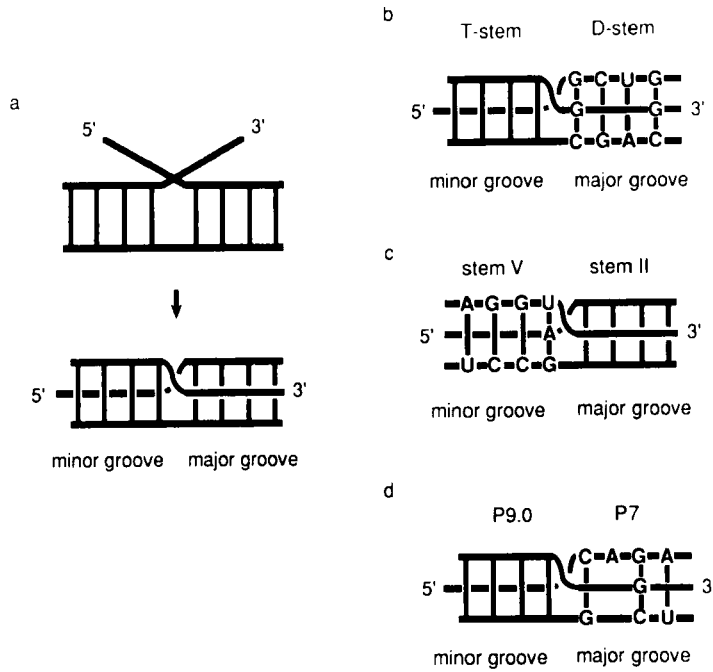

FIG. 11. Base-triple formation at regions where helices coaxially stack may be a recurring RNA structure. (a) The 3′ strand enters the major groove, and the 5′ strand enters the minor groove. (b) Two of the three base-triples in $tRNA^{Phe}$ are consistent with this model. The 5′ strand does not enter the minor groove as expected, but instead loops back into the major groove to form a third triple. (c) The proposed minor-groove triple in *Xenopus laevis* 5-S rRNA (*109*). (d) The major-groove triple that occurs in the self-splicing intron from *Tetrahymena* (*106*).

The 3′ strand of unpaired nucleotides can follow the minor groove of the adjacent helix, while the 5′ strand can follow the major groove of its adjacent helix. Although oversimplified, this picture is consistent with two of the three base-triples that form in $tRNA^{Phe}$, and with the base-triple proposed to occur in the 5-S rRNA from *X. laevis* (Fig. 11). The base-triple formed during the 3′ splicing reaction of the intron from *Tetrahymena* also fits this model if the recently identified P9.0 helix near the 3′ splice site of the *Tetrahymena* intron (*106, 110*) is stacked on the P7 helix (Fig. 11).

To predict the existence of base-triples, the sequences that can form triples must be determined, and their structures and thermodynamic stabilities must be characterized. Triple helices were originally found in polynucleotides with simple repeating sequences. Fiber diffraction studies on the poly(rU)·poly(rA)·poly(rU) system showed that the third strand, poly(rU), bound parallel to the purine strand of the Watson–Crick helix in the major groove with two hydrogen bonds between the adenine and uridine (*111*). In addition to rU·rA·rU, the following polynucleotides have been shown to form triple helices: rA·rA·rU, rC^+·rG·rC (at pH7.0), and rG·rG·rC (*112*) (Fig. 10). Recent NMR studies of DNA triples show that dC^+·dG·dC forms a structure isomorphic with dT·dA·dT (*113, 114*); the corresponding RNA triples are probably isomorphic as well. Structural characterization has not been done on rG·rG·rC or rA·rA·rU triple helices yet, but these base-triples can form isomorphic structures. Replacing the rA·rA·rU base-triple in $tRNA^{Phe}$ with a sequence capable of forming an rG·rG·rC base-triple resulted in no loss of aminoacylation activity, whereas sequences that could not form base-triples did lose activity (*108*).

Unpublished results from our laboratory, including UV absorption mixing curves and circular dichroism spectra, show that poly(rA)·poly(rG)·poly(rC) forms a triple helix. A hydrogen-bonding scheme has not yet been determined for this structure, but the poly(rA) strand appears to bind in the minor groove of the Watson–Crick duplex. The evidence for minor groove binding is that a triple helix did not form when the poly(rG) strand was replaced by poly(rI), which lacks the minor groove amino group capable of hydrogen bonding to poly(rA).

C. Helix–Helix Interactions

We have already described structures in which helical regions stack coaxially end to end. Helix–helix contacts can form between the grooves of different helices when RNA molecules fold into compact tertiary structures. The negatively charged sugar–phosphate backbones repel each other, but a variety of interactions found in crystal structures of nucleic acid duplexes could stabilize the structure. The importance of the 2′ hydroxyl groups in stabilizing helix–helix contacts is seen in the crystal structure

of an RNA duplex $[U(UA)_6A]_2$. There are 12 intermolecular hydrogen bonds between the 2′ hydroxyl groups in one helix and either uracil carbonyl groups or sugars in the minor groove of another helix (*28*). Another type of helix–helix contact is found in the crystal structure of a DNA duplex, d[ACCGGCGCCACA]·d[TGTGGCGCCGGT]. Cytosine amino protons in the major groove of one helix form hydrogen bonds with phosphate oxygens of another helix (*115*). In general, helix–helix interactions could include base–phosphate, base–sugar, sugar–sugar, and sugar–phosphate hydrogen bonding.

Helix–helix contacts have been implicated in the function of one intriguing biological system. The *Tetrahymena* intron is capable of binding a nicked duplex RNA containing three oligonucleotides and then ligating the nick (*116*). Since the nicked duplex is base-paired and the reaction is independent of the duplex sequence, the intron must bind the duplex substrate through the formation of helix–helix contacts. The weakness of the contacts between the intron and the duplex substrate is evidenced by a Michaelis constant greater than 0.1 m*M*. This system illustrates the complexity of the interactions stabilizing RNA structure. Determining the interactions between bases is not enough to understand the structure and function of RNA; the backbone interactions must be determined as well.

Another system in which helical regions may interact involves sequences rich in guanosine. Guanosine-rich DNA sequences have been proposed to form duplex structures (*117*), and there is evidence that two of these duplexes in solution dimerize to form four-stranded DNA complexes (*118–120*). It is not known whether similar sequences in RNA can form these structures, but X-ray diffraction studies have shown that poly(rG) forms a structure containing four parallel strands with the four equivalent guanosines hydrogen-bonded in a coplanar arrangement (*121*).

IV. Predicting Tertiary Interactions

The secondary structures for a wide variety of biological RNA molecules have been established by a combination of techniques such as phylogenetic comparison, chemical and enzymatic modification, and computer prediction algorithms. Some biological functions of RNA can be understood once the secondary structure is known, but understanding most biological functions, particularly catalytic RNA activity, requires the determination of RNA three-dimensional structure. The next step toward the prediction of RNA three-dimensional structure is to develop methods to predict its tertiary interactions. The same two methods used to predict secondary structure (phylogenetic comparison and computer algorithms) can be used to predict tertiary interactions. As currently used, both of these methods are limited because

they predict tertiary interactions only between bases. The tertiary interactions in the crystal structures of tRNA contain many examples of base–sugar, base–phosphate, sugar–sugar, and sugar–phosphate interactions (*62*, *122*).

A. Phylogenetic Comparison

Establishing the presence of tertiary structure from phylogeny relies on the replacement of a specific tertiary pairing by an equivalent one. Phylogenetic comparison has been used to predict the existence of Watson–Crick tertiary pairing in the RNA component of RNase P (*100*), the *Tetrahymena* intron (*123*, *124*), and both the 16-S (*125*) and 23-S (*101*, *102*) rRNAs. Phylogeny can also be used to predict the existence of base triples, as was done for tRNA (*126*).

There are several limitations on predicting tertiary pairs using phylogenetic comparison. All but one of the tertiary pairs formed in tRNA is non-Watson–Crick; these include reverse-Hoogsteen, reverse-Watson–Crick, and parallel-stranded pairing. If this is true for other RNA molecules, phylogenetic comparison will have trouble predicting tertiary pairings. Although phylogenetic evidence has been used to suggest the presence of non-Watson–Crick pairs (*101*), in general, we do not know which non-Watson–Crick pairs are equivalent. Since phylogenetic comparison depends on sequence variation, the structures of conserved regions cannot be predicted. This is a problem in predicting secondary structure and could be a much greater limitation in the case of tertiary structure. The nucleotides in tRNA that are engaged in tertiary pairing are much more highly conserved than are those involved in secondary structure. If this is true in general, it will be difficult to predict tertiary pairing by phylogeny.

B. Thermodynamic Stabilities

The use of thermodynamic stabilities to predict secondary structures has already been discussed. Extending this approach to the prediction of tertiary interactions poses several problems. We discuss the problems inherent in predicting tertiary motifs and then describe an algorithm that predicts the pseudoknot tertiary structure as well as secondary structure. The prediction of tertiary interactions using computer algorithms based on thermodynamic stabilities poses three problems: (1) evaluating free energies for all of the possible secondary and tertiary structures requires prohibitive amounts of computer time; (2) rules governing which regions of secondary structure are sterically allowed to form tertiary interactions have not been established; and (3) the free energies of most tertiary structures have not been determined.

The distinction between secondary and tertiary structures was originally made so that an algorithm could rigorously consider all of the possible secondary structures. For an RNA of n nucleotides, the number of different

base-pairs possible is $n(n-1)/2$. If tertiary pairs are allowed, the number of possible structures increases proportional to n factorial instead of n^2; this renders impractical a rigorous examination of all possible secondary and tertiary structures for a biological RNA.

Since we cannot search every possible combination of secondary and tertiary structures, we must choose criteria that limit the number of structures to be evaluated, but still find the biologically relevant structures. Although the actual path by which RNA molecules fold is a kinetic property, we assume that the structure that forms is the structure of lowest free energy. This means that we can fold the RNA by any path we choose as long as the free energy for each step is known. One way to restrict the number of structures evaluated is first to find the low-free-energy secondary structures by standard programs (*88*). Then tertiary interactions are added to a small number of calculated secondary structures to obtain the tertiary structure of lowest free energy.

We cannot use the approach just outlined for predicting tertiary structures until we know how to look for secondary structure elements that can form tertiary interactions. For example, from the cloverleaf secondary structure of tRNA, how would the algorithm know that the junction conformation brings the D loop close to the T loop, but not to the anticodon loop? Until we learn more about the conformation around the junction, internal, and bulge loops, assumptions must be made about which loop regions can interact. The simplest assumption is that any pair of loop regions can interact.

Prediction of tertiary interactions also requires knowledge of the free energy of forming tertiary interactions. Judging from tRNA and pseudoknots, secondary structure is more stable than tertiary structure. If this is true, the calculation of the free energy for tertiary pairs after the calculation of the free energy for secondary structure is justified. Errors in this assumption can be accounted for by predicting tertiary interactions for several secondary structures. Unfortunately, free energies have not been measured for most tertiary structures. The free energy of pseudoknot formation has been determined for a very limited set of molecules (*99*), but the free energies of other tertiary interactions, such as loop–loop pairing, base-triples, or intercalation, have not been determined.

Despite the difficulties of predicting tertiary structure, an algorithm capable of predicting pseudoknots as well as secondary structure has been developed (*93*). The algorithm calculates the free energy of formation for all of the stem–loops that could be formed by the structure. The predicted structure starts with the stem–loop with the lowest free energy of formation; new stem–loops consistent with those already incorporated are added in order of their stability. The algorithm predicts stem–loops due to pseudoknots as well as those due to normal hairpins. The biggest limitation of this

algorithm is that it only predicts a single structure. The free-energy contribution of the base-pairs in a pseudoknot were assumed to be the same as in a hairpin stem; the free-energy contribution of the two loops of the pseudoknot was empirically determined so that known pseudoknots were predicted. The program also predicted pseudoknots in sequences that had not previously been shown to contain pseudoknots.

The most important information needed to improve tertiary structure prediction are the free-energy parameters for tertiary interactions such as pseudoknots, tertiary base-pairs, and base-triples. Furthermore, rules must be developed indicating which tertiary contacts are sterically possible for a given structure. Realistic predictions of tertiary interactions and three-dimensional structure will not be possible until the conformations around junctions, bulge loops, and internal loops are known.

V. Three-dimensional Structure

Once the secondary structure and tertiary interactions contained in an RNA molecule are known, the next step in understanding its structure is the determination of a three-dimensional structure. The concept of one three-dimensional structure may be misleading, since it implies that the molecule exists in a single structure and ignores the changes the RNA can undergo. Ultimately, we would like to know all of the conformations an RNA molecule can adopt and the dynamics of their interconversion. The first step toward this goal is determination of the three-dimensional structure of one conformation. Since no three-dimensional structures of RNA are known, with the exception of tRNA, which is discussed extensively elsewhere (*26*, *62*, *122*), we discuss the three-dimensional models which have been built for other RNAs.

Three-dimensional models are built in an attempt to understand the functions of RNAs and to guide the development of further experiments to refine the models. These models are built from phylogenetically proven secondary structures plus information about the accessibility of nucleotides to chemical probes, the positions of crosslinks, and any phylogenetically suggested tertiary pairing. Models have been built for 16-S rRNA (*71*, *77*, *127–129*), 5-S rRNA (*109*), the 3′ end of turnip yellow mosaic virus RNA (*130*), and the self-splicing intron of *Tetrahymena* (*76*).

Although the details differ, in general, the following assumptions about RNA structure are used. Model builders first assume that helical regions adopt standard A-form geometry; this is supported by NMR experiments on RNA in solution as well as the known crystal structures of RNAs. Only loop regions of RNA are now left with any degrees of freedom. The conformations of loop regions are varied without bringing atoms closer than van der Waals

contact, so that regions of secondary structure satisfy the crosslinking and tertiary pairing constraints. Helical regions around junction regions are allowed to stack coaxially to form longer helical regions.

The process of model building shows the importance of secondary structure loop regions (internal loops, bulge loops, and junctions) in the three-dimensional structure of RNA. The ability to predict the conformation of loop regions or even the ability to rule out certain conformations would significantly improve the process of building three-dimensional structures of RNA molecules.

The positions of the helical regions in models depend on different types of experimental data. For example, the three-dimensional locations of the proteins bound to 16-S rRNA (*131*) were combined with protein–RNA crosslinking and footprinting data to generate constraints on 75% of the helices in one of the models of 16-S rRNA (*77*). Another model, based on much of the same data, is in substantial agreement with this model (*71*), as is a model built from the accessibility of the 16-S rRNA to DNA oligonucleotide probes, which constrains 40% of the helices in the 16-S rRNA (*128*). Chemical and enzymatic modification data are useful, but these data are insufficient to determine the relative positioning of helical regions.

VI. Determining RNA Structure

RNA structure can be determined at several levels of resolution. The experimental method giving the highest resolution is single crystal X-ray diffraction. In principle, it can provide the coordinates of all of the atoms, although for the large molecules of biological interest the positions of the protons are only inferred. X-Ray diffraction thus reveals the secondary, tertiary, and three-dimensional structures. Unfortunately, RNA molecules often do not form crystals suitable for X-ray analysis, and thus only a few three-dimensional structures of RNA molecules have been solved: $tRNA^{Phe}$ (*132*, *133*), $tRNA^{Asp}$ (*62*), and $tRNA_f^{Met}$ (*134*), as well as an oligonucleotide duplex, $[U(U-A)_6A]_2$ (*28*).

The method with the next highest level of resolution is NMR. It complements the X-ray method in that it provides distances between nearby protons (distances less than 5 Å); it can also determine backbone torsion angles. Thus, NMR provides details about local conformation and can be used to determine secondary, tertiary, and, in principle, three-dimensional structures.

Information about the arrangement of the secondary structure elements in three dimensions can also be obtained from crosslinking and fluorescence energy transfer experiments. Chemical and enzymatic modifications are used to determine the accessibility of functional groups within nucleotides,

and the effect of mutations introduced into RNAs on biological activity can also determine RNA structure.

A. Nuclear Magnetic Resonance

NMR experiments provide a powerful method for determining the structures of proteins and nucleic acids in solution. Detailed explanations of NMR methodology applied to nucleic acids have been published (*135*, *136*), so we only outline the principles of the NMR method here. Then we discuss how well the information determined by NMR defines the three-dimensional structures of RNA molecules.

NMR methods provide three types of information that can be used in structure determination: nuclear Overhauser effects (NOEs), scalar coupling constants, and chemical shifts. NOE is the transfer of magnetization due to magnetic dipole–dipole coupling between nuclei. The effect is directly proportional to the magnetic moments of the nuclei and depends on the inverse sixth power of the distance between them. NOEs can be measured between protons up to 5 Å apart. Both the exchangeable and nonexchangeable protons in RNA are used for NOE measurements.

Information regarding base-pairing and stacking can be obtained from the measurement of NOEs between imino resonances assigned to specific nucleotides. The imino protons resonate in a separate region of the spectrum from other protons, and each Watson–Crick base-pair contains one hydrogen-bonded imino proton. Only imino protons that are hydrogen-bonded, or whose rates of solvent exchange are otherwise decreased, are seen in the NMR spectrum. NOEs between imino protons have been measured in molecules as large as tRNA (*137*). The secondary and tertiary base-pairing as well as the coaxial stacking of helical regions of tRNA in solution have been confirmed by this method (*138*, *139*).

NOEs between the nonexchangeable base and sugar proton resonances give much more detailed information about RNA structure than exchangeable-proton NMR. About 21 intranucleotide distances (9 base–sugar and 12 intra-sugar) can be measured between base and sugar protons. These distances are sufficient to define the conformation of a nucleoside. Up to 11 additional internucleotide (base–base, base–sugar, and sugar–sugar) distances can be measured, depending on the RNA structure.

NMR studies on the nonexchangeable protons are currently limited to oligonucleotides containing no more than 30 or 40 nucleotides. New NMR techniques such as isotopic labeling (*70*, *140*) and three-dimensional NMR methods (*141*) are being developed, which will allow NMR studies on larger RNA molecules. Oligonucleotides used for NMR studies of nonexchangeable protons are designed to adopt structures found in larger RNA molecules. Lower-resolution studies such as chemical modification can be done to check

that the structure adopted by the oligonucleotide is similar to the structure within the larger RNA molecule.

Scalar coupling constants, also called spin–spin splittings, can be measured for two nuclei separated by two, three, or sometimes four bonds. For a three-bond splitting, for example, H–C–C–H, the value of the coupling constant depends on the torsion angle for rotation around the central bond. Coupling constants are related to torsion angles by Karplus-type equations (*142*); these relationships have been determined for the sugar–phosphate backbone in RNA by studies of model compounds (*143*). Coupling constants can be measured between ribose protons as well as between protons and phosphorus atoms. These coupling constants can be used to determine four of the seven torsion angles that completely define the conformation of a nucleotide unit (Fig. 6). Information about two of the three remaining torsion angles can be estimated from phosphorus chemical shift information as described below, and the remaining torsion angle can be determined from the intranucleotide NOEs. In principle, all of the torsion angles can be determined by NMR methods.

The chemical shift of a resonance depends on the local magnetic field at the nucleus. The local magnetic field is extremely sensitive to the bonding, and to the proximities and types, of nearby atoms. Unfortunately, no simple correlation between proton chemical shift and structure has been established so far, although chemical shifts have been calculated for protons in several molecules (*144*). Phosphorus chemical shifts have been correlated with the two phosphodiester torsion angles O–P–O (*145*). Normally, both phosphodiester torsion angles are in the *gauche* conformation. Both theory (*146*) and experiment (*14*, *148*) suggest that if either of the two angles changes to the *trans* conformation, the phosphorus resonance moves to the downfield region of the spectrum. The two adjacent C–O torsion angles and the O–P–O bond angle have also been shown to effect the phosphorus chemical shift (*149*). If the information inherent in the chemical shift could be tapped, NMR structure determination would become much more powerful.

In principle, NMR experiments on RNA molecules could determine their three-dimensional structures. Accurate measurement of seven torsion angles per nucleotide suffices to specify an RNA structure. If all of the torsion angles and a large number of intra- and internucleotide distances are determined (within experimental uncertainties), the atomic coordinates of the RNA molecule can be generated by distance-geometry algorithms (*150–152*). In practice, not all of the torsion angles are determined, and spectral overlap can preclude the measurement of many NOEs. An important question not yet satisfactorily answered is: Which distances and torsion angles must be measured with what accuracy to specify a three-dimensional structure for RNA molecules?

The first step in answering this question was made by testing the distance-geometry algorithm on a known DNA duplex structure (*153*). A set of 117 distances that could be measured easily by NMR was taken from the crystal structure of a 6-bp DNA duplex. The established base-pairing in the duplex was used to add 14 more constraints: the distances between hydrogen-bond donors and acceptors in the base-pairs. Since each nucleotide has seven degrees of freedom, 84 independent constraints suffice to define the structure completely. Although there are more than 84 NMR constraints, the fact that they are not all independent may result in an underdefined structure. The structure generated from these constraints by the distance-geometry algorithm was compared to the starting structure. The results showed that areas of the structure where many interproton distances were used (the sugars and the bases) were well defined. However, the phosphate backbone was not very well defined. The root-mean-square deviation for the generated base hydrogen atoms was 0.43 Å, whereas the deviation was 1.99 Å for the phosphorus atoms. For all of the atoms in the structure, the deviation was 1.29 Å, comparable to the deviation found for the atoms in a protein structure when a similar procedure is undertaken (*154*).

NMR data can determine structures more accurate than that of the DNA duplex just described. The constraints used to generate the DNA structure did not include distances to the H4′, H5′, and H5″ sugar protons because they cannot always be assigned. More importantly, no torsion angles were used to constrain the structure, although many of these angles can be determined from proton–proton and proton–phosphorus coupling constants (*30*, *155*, *156*).

The accuracy of structures determined by NMR ultimately depends on the nucleic acid structure itself. The example of the DNA duplex gives us the lower bound for the accuracy of the structures of duplex RNA determined by NMR, but it is the loop regions of RNA that are of most interest. Can the loop regions of RNA molecules be as well-defined as the duplex regions? Compact loop structures are much better defined than more open ones. The hairpin loop UUCG, discussed above, forms a compact structure in which the sugar protons of the cytosine in the loop give NOE effects to all other nucleotides in the loop. This results in a structure that is well defined by the NMR constraints. The structures of loop regions in which the nucleotides are in extended conformations, or in multiple conformations, can be difficult to determine precisely. If the nucleotides in the loop are not stacked, there probably will be very small NOE effects between them, and their resonances will be difficult to assign. In principle, however, these resonances can be assigned by specific isotope labeling, and a structure can be obtained by measuring coupling constants to determine the backbone torsion angles.

B. Long-range Constraints

1. Crosslinking

Crosslinks—covalent bonds between different bases—can be introduced into RNA molecules by irradiation with UV light (*157*) or by using chemical reagents such as psoralen (*158*) or nitrogen mustard (*77*). The positions of such crosslinks within the structure can be determined by sequencing the RNA on either side of the crosslink. The crosslinks are of two types. The first type occurs between nucleotides adjacent in the secondary structure. The more interesting crosslinks occur between nucleotides that are not adjacent in the secondary structure. The positions of these crosslinks reveal something of the three-dimensional structure of the molecule, since they show the proximity of two secondary structure elements.

Caution must be used when interpreting the results obtained from crosslinking studies. Crosslinks generated by chemical reagents that bind to one base and then crosslink to a second base may not reflect the native structure of the RNA, since the binding of the reagent to the first base may alter the RNA structure.[1] Some of the molecules studied by crosslinking include the 16-S rRNA (*77*), the 23-S rRNA (*77*), tRNA binding to the catalytic subunit of RNase P (*159*), tRNA binding to ribosomes (*84*), and protein binding to ribosomes (*160*).

2. Fluorescence Energy Transfer

Fluorescence energy can be transferred from a donor to an acceptor group by an electric dipole–dipole mechanism. The transfer depends on the inverse sixth power of the distance between these groups. In practice, the transfer can be measured for groups separated by roughly 10–70 Å (*161*). Measuring fluorescence energy transfer in RNA requires adding acceptor and donor groups to the molecule. Some of the applications of this technique include the study of tRNA (*162*), ribosome assembly (*163*), and the conformation about a four-stem DNA junction, as described above (*74*).

C. Chemical Modification

Chemical and enzymatic modification methods determine the accessibility of the nucleotides within an RNA molecule to modification by

[1] This hazard is discussed by Budowsky and Abdurashidova in their article on UV crosslinking (*160*). [Eds.]

chemical reagents or enzymes (*164*). The reactivity of a nucleotide to chemical reagents is a complicated function of solvent accessibility and electrostatic environment (*165*, *166*). The reactivity of nucleotides to chemical modification is used to confirm predicted secondary structures and to learn about tertiary interactions. The advantages of these methods are that they can be used to probe the structure of very large RNAs, and that they require only picomoles of RNA.

Chemicals that react with each of the four bases at Watson–Crick hydrogen-bonding positions can reveal which nucleotides are involved in base-pairing. Conditions are used in which each RNA is modified at most only once, so that the structural information deduced is not an artifact of the modification procedure. Enzymes that cleave specifically in base-paired or unpaired regions are used to determine base-pairing as well, but the large size of enzymes makes them generally less useful than chemicals, since the compact structure of a large folded RNA yields very few enzymatic cleavage sites (*167*).

Chemical modification of the nucleotides within the RNA is detected by one of two methods. The simplest method induces strand scission at the site of modification; this is most useful for short RNA molecules. Sites of modification within large RNA molecules are located by synthesis of a DNA complementary to the RNA using reverse transcriptase (*54*). Modified residues in the RNA cause the reverse transcriptase to stop. Separation of the synthesized DNAs by gel electrophoresis determines the positions of modification. Clear interpretation of the results of chemical modification is often difficult. The amount of modification ranges from essentially none through various degrees of partial modification to strong modification. Strong modification of a nucleotide is good evidence that it is not involved in secondary or tertiary pairing. Weak or partial modification can result from nucleotides engaged in either secondary or tertiary interactions.

D. Mutational Analysis

Mutations are often introduced into RNA sequences in order to determine RNA structure or protein–RNA interactions (*94*, *97*, *108*, *168–172*). The effect of these mutations is often assayed by measuring the ability of the mutated sequences to bind a protein that specifically recognizes the wild-type RNA. Although mutational analysis is a powerful technique for determining interactions in RNA structure, caution is necessary with this approach. The results of such experiments can be unclear, since loss of protein binding can result either from a change in RNA structure or from an RNA sequence that maintains the same structure but is not recognized efficiently by the protein.

VII. Protein–RNA Interactions

Since most cellular RNA molecules are complexed by proteins, understanding protein–RNA interactions is vital to understanding cellular RNA functions. Unfortunately, even less is known about the structure of RNA binding proteins than about RNA structure. The protein–RNA interactions studied show that proteins recognize specific secondary structural features of RNA as well as its three-dimensional shape.

A. Protein–Duplex Interactions

The interactions between the helix–turn–helix proteins and duplex DNA are well-characterized (*173–175*). An α-helix lies in the major groove with amino acids forming specific hydrogen bonds to the DNA sequence. As shown in Fig. 3, the major groove in typical A-form RNA is much narrower than in typical B-form DNA. It has been suggested that the narrower major groove in RNA prevents protein structure elements from binding to the bases in the major groove (*176–178*). However, the fact that a nucleotide strand is capable of binding to duplex RNA in the major groove to form a triple helix, and the fact that there is considerable polymorphism in the size of the A-form major groove in X-ray studies of fibers, suggest that the bases may be accessible to protein structures through the major groove (*179*).

Transcription factor IIIA is a "zinc finger" protein that binds the DNA gene for the 5-S rRNA in *X. laevis* as well as the 5-S rRNA. There is evidence that it binds to the DNA gene in the major groove (*180*), and it is proposed that it binds to the 5-S rRNA in the major groove (*73*). Although it could bind to the DNA and RNA sites by different mechanisms, the possibility that proteins bind in the major groove of RNA should not be ruled out.

B. Protein–Loop Binding

Most of the protein binding sites in RNA that have been characterized are loop regions: hairpins, bulges, and internal loops. Unpaired nucleotides are more conserved than base-paired nucleotides in 16-S rRNA sequences, suggesting that these nucleotides are involved in either tertiary interactions or protein contacts. Unpaired adenosines occur more frequently than the other nucleosides (*181*). Direct evidence for the role of unpaired adenosines in protein binding of the 16-S rRNA comes from the decrease in chemical reactivity of these adenosines when the ribosomal proteins are bound to the 16-S rRNA (*54*).

Both bulge and internal loops bind proteins. A purine bulge is required for the coat protein to bind to the R17 viral RNA (*182*), and an adenosine bulge is part of the L18 protein binding site on the 5-S rRNA (*183*). An

asymmetrical internal loop is required for the binding of S8 ribosomal protein to the 16-S rRNA in *E. coli* (*69, 184, 185*). Ribosomal protein L3 binds to the 23-S rRNA at a large asymmetrical internal loop as well (*186*).

Hairpin structures are commonly found to bind proteins specifically (*187–194*). Hairpins with loop sequences CAGUGN bind to the iron-responsive element (IRE)-binding protein (*195*). As proposed in other protein–RNA interactions (*196–198*), free cysteines in the IRE-binding protein are thought to form transient covalent bonds to the RNA hairpin by a nucleophilic attack by the sulfhydryl group on a uracil in the RNA. Iron regulates this system by altering the equilibrium between reduced and oxidized sulfhydryl groups. This in turn alters the amount of hairpin bound by protein and alters the translation of the iron receptor mRNA (*199*).

The most-characterized example of a hairpin protein binding site is the R17 coat protein binding site (*182*). Binding requires a hairpin, whose loop sequence must be ANYA, plus a purine bulge three nucleotides away on the 5′ side of the loop. It is not clear what role the purine bulge plays in protein binding, since adding substituents such as methyl groups to an adenosine bulge reduces binding 1000-fold, whereas changing the bulge from adenosine to guanosine leaves binding essentially unchanged. These data are insufficient to determine whether the protein forms specific contacts with the purine bulge, or whether intercalation of the purine alters the structure of the RNA. This study shows the limitation of mutagenesis experiments in which specific nucleotides are changed and the effect on protein binding is measured. Without doing structural studies of the RNA molecules, it is not clear whether the different binding affinities that result from substituting specific nucleotides are due to disruption of protein contacts to that nucleotide, or whether the overall structure of the RNA has changed.

C. Protein Recognition of Three-dimensional Structure

The first high-resolution crystal structure of a protein–RNA complex to be solved was that of *E. coli* $tRNA^{Gln}$ and its synthetase (*178*). Sequence-specific contacts occur at the anticodon loop and at the end of the acceptor stem of the tRNA in addition to contacts along one side of the tRNA. Three bases in the anticodon loop, which form specific contacts with the protein, are unstacked compared to the $tRNA^{Phe}$ structure. The first base-pair in the acceptor stem is unpaired, and the protein contacts the acceptor stem in the minor groove, forming hydrogen bonds with exocyclic amino protons of two guanines. Overall, the crystall structure shows that the synthetase recognizes the tRNA shape, but distinguishes it from other tRNAs by forming two types of sequence-specific contacts: binding to anticodon loop nucleotides and binding to guanines in the minor groove of the acceptor stem.

Three-dimensional structure is not always recognized by tRNA syn-

thetases, since a hairpin containing the major recognition feature of $tRNA^{Ala}$ is efficiently aminoacylated by *E. coli* $tRNA^{Ala}$ synthetase (*200*). The major recognition feature is a single G·U mismatch in the acceptor stem (*201*). Replacing the G·U mismatch by G·A, C·A, or U·U mismatches (Fig. 5) results in only small losses in aminoacylation activity, although replacing the mismatch by a Watson–Crick pair completely abolishes activity. The fact that other mismatches are almost as efficient as G·U suggests that the protein recognizes a change in the sugar–phosphate backbone caused by the mismatch rather than specific groups on the G·U pair. Many of the tRNA synthetases probably recognize three-dimensional structure, since the nucleotides required for binding are often scattered throughout the tRNA (*202–204*).

Although the structure of only one protein–RNA complex has been determined, several features of protein binding sites in RNA molecules have been characterized. The most frequently found protein binding sites in RNA are specific nucleotides within loop regions. Hairpin loops are the best-characterized protein binding sites, but bulge loops and internal loops have been implicated in protein binding as well. Proteins recognize primary sequence [poly(A) binding protein] (*205*), secondary structure (hairpin, bulge, and internal loops), and three-dimensional shape (tRNA synthetase binding). The specific interactions between proteins and RNA include transient covalent sulfhydryl–uracil interactions, positively charged amino acids binding to negatively charged phosphates, hydrogen-bond formation with the exocyclic amino group of guanine in the minor groove, and specific hydrogen bonds and stacking interactions with bases in loop regions.

The RNA structures that proteins recognize are better characterized than the protein structural elements that bind to RNA. An 80-amino-acid consensus RNA binding domain, which contains a sequence of eight highly conserved amino acids (*205*), has been identified in several RNA binding proteins, including human U1A protein as well as the poly(A) binding protein (*206–208*). The structure of a peptide portion of the gag polyprotein which binds HIV viral RNA has been determined by NMR (*209*) and is similar to the structure of a zinc finger protein that binds to DNA (*210*).

VIII. RNA–RNA Interactions

Intermolecular RNA interactions occur in a wide range of biological processes, including a "spliceosome" assembly (*211, 212*), RNA "editing" (*213*), and protein synthesis (*214–216*). Intermolecular contacts include base-pairing and backbone interactions. Backbone interactions occur between regions of RNA already base-paired. Although these interactions are not well-characterized, they probably involve the backbone of one helical region interacting

with the groove of another helical region. These interactions include base–sugar, base–phosphate, sugar–sugar, and sugar–phosphate hydrogen bonds.

The stabilities of complexes formed between tRNA anticodon loops and short oligonucleotides (*217*) or between complementary anticodon loops (*218*) are much greater than the stabilities of short duplexes. The enthalpy changes during formation of anticodon loop complexes are similar to that of duplex formation, but the entropy of formation is significantly more favorable for the anticodons than for the duplexes (*217*). The favorable entropy change for the codon–anticodon interaction is probably due to the stacked conformation of the nucleotides within the anticodon loop. Often, the nucleotides in small hairpin loops are not stacked in A-form geometry, presumably making base-paired complexes with these loops much less stable than complexes with anticodon loops.

The importance of backbone interactions in the formation of RNA–RNA complexes is demonstrated by the *Tetrahymena* intron. As previously discussed, the intron is capable of binding an RNA duplex through backbone interactions (*116*). A separate study showed the importance of the 2′ hydroxyl when the intron binds a single-stranded oligonucleotide (*219*). A single-stranded RNA oligonucleotide base-pairs to the intron, forming a complex 10^4 times (6 kal/mol) more stable than would be expected for RNA duplex formation; however, the stability of a complex between the intron and a DNA oligonucleotide is only as stable as the formation of short RNA–DNA duplexes. This suggests that either the 2′ hydroxyl groups of the oligonucleotide form specific interactions with the intron or that the intron forms interactions with the sugar or phosphate groups along the backbone of an A-form duplex.

The complex formed between M1 RNA, the catalytic component of RNase P, and its tRNA substrate is probably stabilized by the formation of backbone interactions. This suggestion is supported by a chemical crosslink that forms in the M1 RNA–tRNA complex between base-paired regions of the M1 RNA and the precursor tRNA (*159*). As previously noted (*159*), there is striking sequence and secondary structure homology between the region of the M1 RNA that crosslinks to the tRNA and the region of the 23-S rRNA which was found to bind tRNA by chemical footprinting (*83*). Complexes stabilized by backbone interactions may be a general feature of catalytic RNA–substrate complexes since these interactions are weak and allow dissociation of the enzyme–substrate complex after the reaction.

IX. RNA–DNA Interactions

Hybrid duplexes form between RNA and DNA during transcription and reverse transcription. The stability of these hybrids is believed to play a role in transcription termination (*220*). The combination of a stable hairpin forma-

tion in the nascent RNA followed by a repeating dA sequence in the DNA leads to termination. The explanation for this is that the stable hairpin causes the polymerase to pause and disrupts part of the hybrid duplex (*221*). The polymerase complex is held together by the remaining hybrid duplex. Since rU·dA hybrid duplexes are much less stable than other hybrids (*222*, *223*), the polymerase falls off and transcription terminates.

The discovery that a ribonucleoprotein complex adds DNA sequences to chromosomal ends (*224*) presents the possibility that an RNA molecule can function as a reverse transcriptase. *Tetrahymena* telomerase[2] adds the DNA sequence TTGGGG to the 5′ end of chromosomes; it contains a protein component and a 159-nucleotide RNA component, including the sequence CAACCCCAA complementary to the synthesized DNA (*225*). The role of the CAACCCAAA sequence as a template for DNA synthesis was proven by the discovery that mutating this sequence changes the DNA sequence at the ends of chromosomes *in vivo* (*226*). Whether the RNA alone can act as a reverse transcriptase or whether it only serves as a template for DNA synthesis by the protein component of telomerase has not been established.

ACKNOWLEDGMENTS

We thank Peter Davis, John Jaeger, and Gabriele Varani for their reading of the manuscript, and we particularly thank Jacqueline Wyatt for her reading of the manuscript and for many useful discussions. M.C. is a Howard Hughes Medical Institute Doctoral Fellow. This work was supported in part by National Institutes of Health Grant GM 10840, and by the Department of Energy, Office of Energy Research, Office of Health and Environmental Research under Grant DE-FG03-86ER60406.

REFERENCES

1. K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling and T. R. Cech, *Cell* **31**, 147 (1982).
2. C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace and S. Altman, *Cell* **35**, 849 (1983).
3. R. van der Veen, A. C. Arnberg, G. van der Horst, L. Bonen, H. F. Tabak and L. A. Grivell, *Cell* **44**, 225 (1986).
4. C. L. Peebles, P. S. Perlman, K. L. Mecklenburg, M. L. Petrillo, J. H. Tabor, K. A. Jarrell and H.-L. Cheng, *Cell* **44**, 213 (1986).
5. A. Hampel and R. Tritz, *Bchem* **28**, 4929 (1989).
6. A. Hampel, R. Tritz, M. Hicks and P. Cruz, *NARes* **18**, 299 (1990).
7. H.-N. Wu, Y.-J. Lin, F.-P. Lin, S. Makino, M.-F. Chang and M. M. C. Lai, *PNAS* **86**, 1831 (1989).
8. G. A. Prody, J. T. Bakos, J. M. Buzayan, I. R. Schneider and G. Bruening, *Science* **231**, 1577 (1986).
9. C. H. Hutchins, P. D. Rathjen, A. C. Forster and R. H. Symons, *NARes* **14**, 3627 (1986).
10. C. Yanofsky, *Nature* **289**, 751 (1981).
11. C. L. Chan and R. Landick, *JBC* **264**, 20796 (1989).
12. L. P. Eperon, I. R. Graham, A. D. Griffiths and I. C. Eperon, *Cell* **54**, 393 (1988).

[2] Formerly "telomere-adding enzyme," "T_mG_n-adding enzyme," or "telomere terminal transferase." [Eds.]

13. P. Schimmel, *Cell* **58,** 9 (1989).
14. M. Kozak, *PNAS* **83,** 2850 (1986).
15. K.-O. Cho and C. Yanofsky, *JMB* **204,** 51 (1988).
16. G. Brawerman, *Cell* **57,** 9 (1989).
17. V. J. Cannistraro, M. N. Subbarao and D. Kennell, *JMB* **192,** 257 (1986).
18. U. Blasi, K. Nam, D. Hartz, L. Gold and R. Young, *EMBO J.* **8,** 3501 (1989).
19. S. Altuvia, D. Kornitzer, D. Teff and A. B. Oppenheim, *JMB* **210,** 265 (1989).
20. M. H. de Smit and J. van Duin, *This Series* **38,** 1 (1990).
21. L. Gold, *ARB* **57,** 199 (1988).
22. W. M. Huang, S.-Z. Ao, S. Casjens, R. Orlandi, R. Zeikus, R. Weiss, D. Winge and M. Fang, *Science* **239,** 1005 (1988).
23. J. F. Milligan, D. R. Groebe, G. W. Witherell and O. C. Uhlenbeck, *NARes* **15,** 8783 (1987).
24. S.-H. Chou, P. Flynn and B. Reid, *Bchem* **28,** 2422 (1989).
25. N. Usman, K. K. Ogilvie, M.-Y. Jiang and R. J. Cedergren, *JACS* **109,** 7845 (1987).
26. W. Saenger, "Principles of Nucleic Acid Structure." Springer-Verlag, New York, 1984.
27. S. Arnott, D. W. L. Hukins and S. D. Dover, *BBRC* **48,** 1392 (1972).
28. A. C. Dock-Bregeon, B. Chevrier, A. Podjarny, J. Johnson, J. S. de Bear, G. R. Gough, P. T. Gilham and D. Moras, *JMB* **209,** 459 (1989).
29. P. W. Davis, R. W. Adamiak and I. Tinoco, Jr., *Biopolymers* **29,** 109 (1990).
30. G. Varani, B. Wimberly and I. Tinoco, Jr., *Bchem* **28,** 7760 (1989).
31. J. D. Puglisi, J. R. Wyatt and I. Tinoco, Jr., *Bchem* **29,** 4215 (1990).
32. C. S. Happ, E. Happ, N. Nilges, A. M. Gronenborn and G. M. Clore, *Bchem* **27,** 1735 (1988).
33. S. Arnott and D. W. L. Hukins, *BBRC* **47,** 1504 (1972).
34. A. Bhattacharyya, A. I. H. Murchie and D. M. J. Lilley, *Nature* **343,** 484 (1990).
34a. R. S. Tang and D. E. Draper, *Bchem* **29,** 5232 (1990).
35. D. Rhodes and A. Klug, *Nature* **292,** 378 (1981).
36. L. J. Peck and J. C. Wang, *Nature* **292,** 375 (1981).
37. T. D. Tullius and B. A. Dombroski, *Science* **230,** 679 (1985).
38. K. Hall, P. Cruz, I. Tinoco, Jr., T. M. Jovin and J. H. van de Sande, *Nature* **311,** 584 (1984).
39. I. Tinoco, Jr., O. C. Uhlenbeck and M. D. Levine, *Nature* **230,** 362 (1971).
40. C. Cheong, G. Varani and I. Tinoco, Jr., *Nature* **346,** 680 (1990).
40a. T. Sakata, H. Hiroaki, Y. Oda, T. Tanaka, M. Ikehara and S. Uesugi, *NARes* **18,** 3831 (1990).
41. M. J. J. Blommers, J. A. L. I. Walters, C. A. G. Haasnoot, J. M. A. Aelen, G. A. van der Marel, J. H. van Bloom and C. W. Hilbers, *Bchem* **28,** 7491 (1989).
42. D. R. Groebe and O. C. Uhlenbeck, *NARes* **16,** 11725 (1988).
43. H. F. Noller, *ARB* **53,** 119 (1984).
44. J. A. Jaeger, D. H. Turner and M. Zuker, *PNAS* **86,** 7706 (1989).
45. C. Tuerk, P. Gauss, C. Thermes, D. R. Groebe, M. Gayle, N. Guild, G. Stormo, Y. d'Aubenton-Carafa, O. C. Uhlenbeck, I. Tinoco, Jr., E. N. Brody and L. Gold, *PNAS* **85,** 1364 (1988).
46. I. Hirao, Y. Nishimura, T. Naraoka, K. Watanabe, Y. Arata and K. Miura, *NARes* **17,** 2223 (1989).
47. W. Fuller and A. Hodgson, *Nature* **215,** 817 (1967).
48. G. M. Clore, A. M. Gronenborn, E. A. Piper, L. W. McLaughlin, E. Graeser and J. H. van Boom, *BJ* **221,** 737 (1984).
49. J. Wu and A. G. Marshall, *Bchem* **29,** 1722 (1990).

50. J. Wu and A. G. Marshall, *Bchem* **29,** 1730 (1990).
51. C. A. G. Haasnoot, C. W. Hilbers, G. A. van der Marel, J. H. van Boom, U. C. Singh, N. Pattabiraman and P. A. Kollman, *J. Biomol. Struct. Dyn.* **3,** 843 (1986).
52. M. W. Kalnik, D. G. Norman, B. F. Li, P. F. Swann and D. J. Patel, *JBC* **265,** 636 (1990).
53. Y. T. van den Hoogen, A. A. van Beuzekom, E. de Vroom, G. A. van der Marel, J. H. van Boom and C. Altona, *NARes* **16,** 5013 (1988).
54. D. Moazed, S. Stern and H. F. Noller, *JMB* **187,** 399 (1986).
55. A. Bhattacharyya and D. M. J. Lilley, *NARes* **17,** 6821 (1989).
56. C.-H. Hsieh and J. D. Griffith, *PNAS* **86,** 4833 (1989).
57. J. A. Rice and D. M. Crothers, *Bchem* **28,** 4512 (1989).
58. C. E. Longfellow, R. Kierzek and D. H. Turner, *Bchem* **29,** 278 (1990).
59. S. A. White and D. E. Draper, *Bchem* **28,** 1892 (1989).
60. S. Roy, V. Sklenar, E. Appella and J. S. Cohen, *Biopolymers* **26,** 2041 (1987).
61. D. H. Turner, N. Sugimoto and S. M. Freier, *Annu. Rev. Biophys. Biophys. Chem.* **17,** 167 (1988).
62. E. Westhof, P. Dumas and D. Moras, *JMB* **184,** 119 (1985).
63. W. Traub and J. L. Sussman, *NARes* **10,** 2701 (1982).
64. P. J. Romaniuk, I. L. de Stevenson, C. Ehresmann, P. Romby and B. Ehresmann, *NARes* **16,** 2295 (1988).
65. P. Romby, E. Westhof, R. Toukifimpa, R. Mache, J.-P. Ebel, C. Ehresmann and B. Ehresmann, *Bchem* **27,** 4721 (1988).
66. A. Bhattacharyya and D. M. J. Lilley, *JMB* **209,** 583 (1989).
67. F. Aboul-ela, D. Koh and I. Tinoco, Jr., *NARes* **13,** 4811 (1985).
68. C. Papanicolaou, M. Gouy and J. Ninio, *NARes* **12,** 31 (1984).
69. M. Mougel, F. Eyermann, E. Westhof, P. Romby, A. Expert-Bezançon, J.-P. Ebel, B. Ehresmann and C. Ehresmann, *JMB* **198,** 91 (1987).
70. P. Zhang and P. B. Moore, *Bchem* **28,** 4607 (1989).
71. S. Stern, B. Weiser and H. F. Noller, *JMB* **204,** 447 (1988).
72. P. J. Romaniuk, *Bchem* **28,** 1388 (1989).
73. J. Christiansen, R. S. Brown, B. S. Sproat and R. A. Garrett, *EMBO J.* **6,** 453 (1987).
74. A. I. H. Murchie, R. M. Clegg, E. von Kitzing, D. R. Duckett, S. Diekmann and D. M. J. Lilley, *Nature* **341,** 763 (1989).
75. C. R. Woese, R. Gutell, R. Gupta and H. F. Noller, *Microbiol. Rev.* **47,** 621 (1983).
76. S.-H. Kim and T. R. Cech, *PNAS* **84,** 8788 (1987).
77. R. Brimacombe, J. Atmadja, W. Stiege and D. Schüler, *JMB* **199,** 115 (1988).
78. L. M. Epstein and J. G. Gall, *Cell* **48,** 535 (1987).
79. A. C. Forster and R. H. Symons, *Cell* **49,** 211 (1987).
80. M. Koizumi, S. Iwai and E. Ohtsuka, *FEBS Lett.* **228,** 228 (1988).
81. C. C. Sheldon and R. H. Symons, *NARes* **17,** 5679 (1989).
82. A. E. Dahlberg, *Cell* **57,** 525 (1989).
83. D. Moazed and H. F. Noller, *Cell* **57,** 585 (1989).
84. G. Steiner, E. Kuechler and A. Barta, *EMBO J.* **7,** 3949 (1988).
85. C. C. Hall, D. Johnson and B. S. Cooperman, *Bchem* **27,** 3983 (1988).
86. N. R. Pace, D. K. Smith, G. J. Olsen and B. D. James, *Gene* **82,** 65 (1989).
87. A. L. Williams, Jr., and I. Tinoco, Jr., *NARes* **14,** 299 (1986).
88. M. Zuker, *Science* **244,** 48 (1989).
89. J. Gralla and D. M. Crothers, *JMB* **73,** 497 (1973).
90. O. C. Uhlenbeck, P. N. Borer, B. Dengler and I. Tinoco, Jr., *JMB* **73,** 483 (1973).
91. D. A. M. Konings and P. Hogeweg, *JMB* **207,** 597 (1989).
92. G. M. Studnicka, G. M. Rahn, I. W. Cummings and W. A. Salser, *NARes* **5,** 3365 (1978).

93. J. P. Abrahams, M. van den Berg, E. van Batenburg and C. Pleij, *NARes* **18,** 3035 (1990).
94. C. K. Tang and D. E. Draper, *Cell* **57,** 531 (1989).
95. K. Rietveld, R. van Poelgeest, C. W. A. Pleij, J. H. van Boom and L. Bosch, *NARes* **10,** 1929 (1982).
96. C. W. A. Pleij, K. Rietveld and L. Bosch, *NARes* **13,** 1717 (1985).
97. I. Brierley, P. Digard and S. C. Inglis, *Cell* **57,** 537 (1989).
98. J. D. Puglisi, J. R. Wyatt and I. Tinoco, Jr., *JMB* **214,** 437 (1990).
99. J. R. Wyatt, J. D. Puglisi and I. Tinoco, Jr., *JMB* **214,** 455 (1990).
100. B. D. James, G. J. Olsen, J. Liu and N. R. Pace, *Cell* **52,** 19 (1988).
101. R. R. Gutell and C. R. Woese, *PNAS* **87,** 663 (1990).
102. H. Leffers, J. Kjems, L. Østergaard, N. Larsen and R. A. Garrett, *JMB* **195,** 43 (1987).
103. L. J. Maher III, B. Wold and P. B. Dervan, *Science* **245,** 725 (1989).
104. M. Cooney, G. Czernuszewicz, E. H. Postel, S. J. Flint and M. E. Hogan, *Science* **241,** 456 (1988).
105. T. C. Boles and M. E. Hogan, *Bchem* **26,** 367 (1987).
106. F. Michel, M. Hanna, R. Green, D. P. Bartel and J. W. Szostak, *Nature* **342,** 391 (1989).
107. I. Tinoco, Jr., J. D. Puglisi and J. R. Wyatt, *in* "Nucleic Acids and Molecular Biology" (F. Eckstein and D. M. J. Lilley, eds.), p. 205. Springer-Verlag, New York, 1990.
108. J. R. Sampson, A. B. DiRenzo, L. S. Behlen and O. C. Uhlenbeck, *Bchem* **29,** 2523 (1990).
109. E. Westhof, P. Romby, P. J. Romaniuk, J.-P. Ebel, C. Ehresmann and B. Ehresmann, *JMB* **207,** 417 (1989).
110. J. M. Burke, J. S. Esherick, W. R. Burfeind and J. L. King, *Nature* **344,** 80 (1990).
111. S. Arnott and P. J. Bond, *Nature NB* **244,** 99 (1973).
112. A. G. Letai, M. A. Palladino, E. Fromm, V. Rizzo and J. R. Fresco, *Bchem* **27,** 9108 (1988).
113. P. Rajagopal and J. Feigon, *Bchem* **28,** 7859 (1989).
114. C. de los Santos, M. Rosen and D. Patel, *Bchem* **28,** 7282 (1989).
115. Y. Timsit, E. Westhof, R. P. P. Fuchs and D. Moras, *Nature* **341,** 459 (1989).
116. J. A. Doudna and J. W. Szostak, *Nature* **339,** 519 (1989).
117. E. Henderson, C. C. Hardin, S. K. Wolk, I. Tinoco, Jr., and E. H. Blackburn, *Cell* **51,** 899 (1987).
118. W. I. Sundquist and A. Klug, *Nature* **342,** 825 (1989).
119. J. R. Williamson, M. K. Raghuraman and T. R. Cech, *Cell* **59,** 871 (1989).
120. D. Sen and W. Gilbert, *Nature* **344,** 410 (1990).
121. S. B. Zimmerman, G. S. Cohen and D. R. Davies, *JMB* **92,** 181 (1975).
122. J. P. Goddard, *Prog. Biophys. Mol. Biol.* **32,** 233 (1977).
123. R. B. Waring, C. Scazzocchio, T. A. Brown and R. W. Davies, *JMB* **167,** 595 (1983).
124. F. Michel and B. Dujon, *EMBO J.* **2,** 33 (1983).
125. R. R. Gutell, H. F. Noller and C. R. Woese, *EMBO J.* **5,** 1111 (1986).
126. M. Levitt, *Nature* **224,** 759 (1969).
127. A. Expert-Bezançon and P. Wollenzien, *JMB* **184,** 53 (1985).
128. M. I. Oakes, L. Kahan and J. A. Lake, *JMB* **211,** 907 (1990).
129. J. Hubbard, Ph.D. thesis, University of California, Berkeley, 1990.
130. P. Dumas, D. Moras, C. Florentz, R. Giegé, P. Verlaan, A. Van Belkum and C. W. A. Pleij, *J. Biomol. Struct. Dyn.* **4,** 707 (1987).
131. M. S. Capel, D. M. Engelman, B. R. Freeborn, M. Kjeldgaard, J. A. Langer, V. Ramakrishnan, D. G. Schindler, D. K. Schneider, B. P. Schoenborn, I.-Y. Sillers, S. Yabuki and P. B. Moore, *Science* **238,** 1403 (1987).
132. S.-H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. H. J. Wang, N. C. Seeman and A. Rich, *Science* **185,** 435 (1974).
133. J. D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. S. Brown, B. F. C. Clark and A. Klug, *Nature* **250,** 546 (1974).

134. N. H. Woo, B. A. Roe and A. Rich, *Nature* **286,** 346 (1980).
135. K. Wüthrich, "NMR of Proteins and Nucleic Acids." Wiley, New York, 1986.
136. F. J. M. van de Ven and C. W. Hilbers, *EJB* **178,** 1 (1988).
137. S. Roy and A. G. Redfield, *NARes* **9,** 7073 (1981).
138. R. E. Hurd and B. R. Reid, *Bchem* **18,** 4017 (1979).
139. A. Heerschap, J.-R. Mellema, H. G. J. M. Janssen, J. A. L. I. Walters, C. A. G. Haasnoot and C. W. Hilbers, *EJB* **149,** 649 (1985).
140. A. Bax, R. H. Griffey and B. L. Hawkins, *J. Magn. Reson.* **55,** 301 (1983).
141. G. W. Vuister, R. Boelens, A. Padilla, G. J. Kleywegt and R. Kaptein, *Bchem* **29,** 1829 (1990).
142. M. Karplus, *JACS* **85,** 2870 (1963).
143. C. Altona, *Recl. Trav. Chim. Pays-Bas* **101,** 413 (1982).
144. C. Giessner-Prettre and B. Pullman, *Q. Rev. Biophys.* **20,** 113 (1987).
145. D. G. Gorenstein, *Annu. Rev. Biophys. Bioeng.* **10,** 355 (1981).
146. F. R. Prado, C. Giessner-Prettre, B. Pullman and J.-P. Dandley, *JACS* **101,** 1737 (1979).
147. T. M. Jovin, J. H. van de Sande, D. A. Zarling, D. J. Arndt-Jovin, F. Eckstein, H. H. Fuldner, C. Greider, I. Grieger, E. Hamori, B. Kalisch, L. P. McIntosh and M. Robert-Nicoud, *CSHSQB* **47,** 143 (1983).
148. A. H.-J. Wang, G. J. Quigley, F. J. Kolpak, G. van der Marel, J. H. van Boom and A. Rich, *Science* **211,** 171 (1981).
149. C. Giessner-Prettre, B. Pullman, F. R. Prado, D. M. Cheng, V. Iuorno and P. O. P. Ts'o, *Biopolymers* **23,** 377 (1984).
150. T. F. Havel and K. Wuthrich, *Bull. Math. Biol.* **46,** 673 (1984).
151. W. Braun and N. Go, *JMB* **186,** 611 (1985).
152. D. R. Hare, L. Shapiro and D. J. Patel, *Bchem* **25,** 7445 (1986).
153. A. Pardi, D. R. Hare and C. Wang, *PNAS* **85,** 8785 (1988).
154. T. F. Havel and K. Wuthrich, *JMB* **182,** 281 (1985).
155. J. R. Williamson and S. G. Boxer, *Bchem* **28,** 2819 (1989).
156. E. Nikonowicz, V. Roongta, C. R. Jones and D. G. Gorenstein, *Bchem* **28,** 8714 (1989).
157. A. D. Branch, B. J. Benenfeld, C. P. Paul and H. D. Robertson, in "Methods in Enzymology" (James E. Dahlberg and John N. Abelson, eds.), Vol. 180, p. 418. Academic Press, San Diego, 1989.
158. G. D. Cimino, H. B. Gamper, S. T. Isaacs and J. E. Hearst, *ARB* **54,** 1151 (1985).
159. G. Guerrier-Takada, N. Lumelsky and S. Altman, *Science* **246,** 1578 (1990).
160. E. I. Budowsky and G. G. Abdurashidova, *This Series* **37,** 1 (1989).
161. L. Stryer, *ARB* **47,** 819 (1978).
162. K. Beardsley and C. R. Cantor, *PNAS* **65,** 39 (1970).
163. K.-H. Huang, R. H. Fairclough and C. R. Cantor, *JMB* **97,** 443 (1975).
164. C. Ehresmann, F. Baudin, M. Mougel, P. Romby, J.-P. Ebel and B. Ehresmann, *NARes* **15,** 9109 (1987).
165. R. Lavery and A. Pullman, *Biophys. Chem.* **19,** 171 (1984).
166. S. Furois-Corbin and A. Pullman, *Biophys. Chem.* **22,** 1 (1985).
167. G. Knapp, *in* "Methods in Enzymology" (James E. Dahlberg and John N. Abelson, eds.), Vol. 180, p. 193. Academic Press, San Diego, 1989.
168. R. Parker, *in* "Methods in Enzymology" (James E. Dahlberg and John N. Abelson, eds.), Vol. 180, 510. Academic Press, San Diego, 1989.
169. H. S. Olsen, P. Nelbock, A. W. Cochrane and C. A. Rosen, *Science* **247,** 845 (1990).
170. S. Heaphy, C. Dingwall, I. Ernberg, M. J. Gait, S. M. Green, J. Karn, A. D. Lowe, M. Singh and M. A. Skinner, *Cell* **60,** 685 (1990).
171. K. Ehrenman, R. Schroeder, P. S. Chandry, D. H. Hall and M. Belfort, *NARes* **17,** 9147 (1989).

172. C. L. Williamson, W. M. Tierney, B. J. Kerker and J. M. Burke, *JBC* **262,** 14672 (1987).
173. C. Wolberger, Y. Dong, M. Ptashne and S. C. Harrison, *Nature* **335,** 789 (1988).
174. S. R. Jordan and C. O. Pabo, *Science* **242,** 893 (1988).
175. A. K. Aggarwal, D. W. Rodgers, M. Drottar, M. Ptashne and S. C. Harrison, *Science* **242,** 899 (1988).
176. S. Stern, R. C. Wilson and H. F. Noller, *JMB* **192,** 101 (1986).
177. J.-H. Wang, *Nature* **319,** 183 (1986).
178. M. A. Rould, J. J. Perona, D. Soll and T. A. Steitz, *Science* **246,** 1135 (1989).
179. S. Arnott, *Nature* **320,** 313 (1986).
180. L. Fairall, D. Rhodes and A. Klug, *JMB* **192,** 577 (1986).
181. R. R. Gutell, B. Weiser, C. R. Woese and H. F. Noller, *This Series* **32,** 155 (1985).
182. H.-N. Wu and O. C. Uhlenbeck, *Bchem* **26,** 8221 (1987).
183. D. A. Peattie, S. Douthwaite, R. A. Garrett and H. F. Noller, *PNAS* **78,** 7331 (1981).
184. R. J. Gregory and R. A. Zimmerman, *NARes* **14,** 5761 (1986).
185. R. J. Gregory, P. B. F. Cahill, D. L. Thurlow and R. A. Zimmerman, *JMB* **204,** 295 (1988).
186. H. Leffers, J. Egebjerg, A. Anderson, T. Christensen and R. A. Garrett, *JMB* **204,** 507 (1988).
187. S. Feng and E. C. Holland, *Nature* **334,** 165 (1988).
188. C. Dingwall, I. Ernberg, M. J. Gait, S. M. Green, S. Heaphy, J. Karn, A. D. Lowe, M. Singh, M. A. Skinner and R. Valerio, *PNAS* **86,** 6925 (1989).
189. B. Berkhout, R. H. Silverman and K.-T. Jeang, *Cell* **59,** 273 (1989).
190. D. Lazinski, E. Grzadzielska and A. Das, *Cell* **59,** 207 (1989).
191. D. Scherly, W. Boelens, W. J. van Venrooij, N. A. Dathan, J. Hamm and I. W. Mattaj, *EMBO J.* **8,** 4163 (1989).
192. D. R. Turner, L. E. Joyce and P. J. G. Butler, *JMB* **203,** 531 (1988).
193. G. W. Witherell and O. C. Uhlenbeck, *Bchem* **28,** 71 (1989).
194. Y. Endo, A. Glück, Y.-L. Chan, K. Tsurugi and I. G. Wool, *JBC* **265,** 2216 (1990).
195. J. L. Casey, M. W. Hentze, D. M. Koeller, S. W. Caughman, T. A. Rouault, R. D. Klausner and J. B. Harford, *Science* **240,** 924 (1988).
196. Y. Wataya, A. Matsuda and D. V. Santi, *JBC* **255,** 5538 (1980).
197. R. M. Starzyk, S. W. Koontz, and P. Schimmel, *Nature* **298,** 136 (1982).
198. P. J. Romaniuk and O. C. Uhlenbeck, *Bchem* **24,** 4239 (1985).
199. M. W. Hentze, T. A. Rouault, J. B. Harford and R. D. Klausner, *Science* **244,** 357 (1989).
200. C. Francklyn and P. Schimmel, *Nature* **337,** 478 (1989).
201. W. H. McClain, Y.-M. Chen, F. Foss and J. Schneider, *Science* **242,** 1681 (1988).
202. J. Normanly and J. Abelson, *ARB* **58,** 1029 (1989).
203. P. Schimmel, *Bchem* **28,** 2747 (1989).
204. J. R. Sampson, A. B. DiRenzo, L. S. Behlen and O. C. Uhlenbeck, *Science* **243,** 1363 (1989).
205. S. A. Adam, T. Nakagawa, M. S. Swanson, T. K. Woodruff and G. Dreyfuss, *MCBiol* **6,** 2932 (1986).
206. C. C. Query, R. C. Bentley and J. D. Keene, *Cell* **57,** 89 (1989).
207. I. W. Mattaj, *Cell* **57,** 1 (1989).
208. G. Dreyfuss, M. S. Swanson and S. Pinol-Roma, *TIBS* **13,** 86 (1988).
209. M. F. Summers, T. L. South, B. Kim and D. R. Hare, *Bchem* **29,** 329 (1990).
210. M. S. Lee, G. P. Gippert, K. V. Soman, D. A. Case and P. E. Wright, *Science* **245,** 635 (1989).
211. P. A. Sharp, *Science* **235,** 766 (1987).
212. D. A. Brow and C. Guthrie, *Nature* **334,** 213 (1988).
213. B. Blum, N. Bakalara and L. Simpson, *Cell* **60,** 189 (1990).

214. D. Moazed and H. F. Noller, *Nature* **342,** 142 (1989).
215. D. Moras, A.-C. Dock, P. Dumas, E. Westhof, P. Romby, J.-P. Ebel and R. Giegé, *J. Biomol. Struct. Dyn.* **3,** 479 (1985).
216. D. Moras, A.-C. Dock, P. Dumas, E. Westhof, P. Romby, J.-P. Ebel and R. Giegé, *PNAS* **83,** 932 (1986).
217. K. Yoon, D. H. Turner and I. Tinoco, Jr., *JMB* **99,** 507 (1975).
218. P. Romby, R. Giege, C. Houssier and H. Grosjean, *JMB* **184,** 107 (1985).
219. D. Herschlag and T. R. Cech, *Nature* **344,** 405 (1990).
220. P. H. von Hippel, D. G. Bear, W. D. Morgan and J. A. McSwiggen, *ARB* **53,** 389 (1984).
221. T. Platt, *Cell* **24,** 10 (1981).
222. M. Riley, B. Maling and M. J. Chamberlin, *JMB* **20,** 359 (1966).
223. F. H. Martin and I. Tinoco, Jr., *NARes* **8,** 2295 (1980).
224. C. W. Greider and E. H. Blackburn, *Cell* **51,** 887 (1987).
225. C. W. Greider and E. H. Blackburn, *Nature* **337,** 331 (1989).
226. G.-L. Yu, J. D. Bradley, L. D. Attardi and E. H. Blackburn, *Nature* **344,** 126 (1990).