Chapter 4
The versatility of RNA

The final stage in the exaltation of the RNA component of RNase P occurred in 1983 – converting contaminating crud to catalytic component after a decade.


Outline
4.1 Introduction
4.2 Secondary structure of RNA
   Secondary structure motifs in RNA
   Base-paired RNA adopts an A-type double helix
   RNA helices often contain noncanonical base pairs
4.3 Tertiary structure of RNA
   tRNA structure: important insights into RNA structural motifs
   Common tertiary structure motifs in RNA
4.4 Kinetics of RNA folding
4.5 RNA is involved in a wide range of cellular processes
4.6 Historical perspective: the discovery of RNA catalysis
   Tetrahymena group I intron ribozyme
   RNase P ribozyme
   Focus box 4.1: The RNA world
4.7 Ribozymes catalyze a variety of chemical reactions
   Mode of ribozyme action
   Large ribozymes
   Small ribozymes

Chapter summary
Analytical questions
Suggestions for further reading

4.1 Introduction
Initial studies on RNA structure were pursued side by side with that of DNA. What became increasingly apparent is that RNA has a much greater structural and functional versatility compared with DNA. The growing database of RNA structures has led to characterization of numerous RNA secondary and tertiary structural motifs. RNA is now viewed as a modular structure built from a combination of these building blocks and tertiary linkers. RNA chains fold into unique three-dimensional structures which act similarly to globular proteins. The folding patterns provide the basis for their chemical reactivity and specific interactions with other molecules, including proteins, nucleic acids, and small ligands. RNA is involved in a wide range of essential cellular processes from DNA replication to protein synthesis.
As introduced in Chapter 2, RNA is a chain-like molecule composed of subunits called nucleotides joined by phosphodiester bonds (Fig. 4.1). Each nucleotide subunit is composed of three parts: a ribose sugar, a phosphate group, and a nitrogenous base. The common bases found in RNA are adenine (A), guanine (G), cytosine (C), and uracil (U). Single-stranded RNA folds into a variety of secondary structural motifs that are stabilized by both Watson–Crick and unconventional base pairing.

**Secondary structure motifs in RNA**

Secondary structures of RNA can be predicted with good accuracy by computer analysis, based on thermodynamic data for the free energies of various conformations, comparative sequence analysis, and solved crystal structures. Some of the common secondary structures that form the building blocks of RNA architecture are shown in Fig. 4.2. These include bulges, base-paired helices or “stems,” single-stranded hairpin or internal loops, and junctions. RNA structure was once envisioned as a collection of relatively rigid stems comprised of Watson–Crick bases pairs and the single-stranded loops defined by these stems. The first structure of transfer RNA showed otherwise. In fact, as we shall see RNA adopts structures that Harry Noller described in a 2005 *Science* review article as “breathtakingly intricate and graceful.”
Base-paired RNA adopts an A-type double helix

In DNA the double helix forms from two separate DNA strands. In RNA, helix formation occurs by hydrogen bonding between base pairs and base stacking hydrophobic interactions within one single-stranded chain of nucleotides. X-ray crystallography studies have shown that base-paired RNA primarily adopts a right-handed A-type double helix with 11 bp per turn (Fig. 4.2). The 2'-hydroxyl group of the ribose sugar in RNA hinders formation of a B-type helix – the predominant form in double-stranded DNA – but can be accommodated within an A-type helix. Regular A-type RNA helices with Watson–Crick base pairs have a deep, narrow major groove that is not well suited for specific interactions with ligands. On the other hand, although the minor groove does not display sequence specificity, it includes the ribose 2'-OH groups which are good hydrogen bond acceptors, and it is shallow and broad, making it accessible to ligands. Because of these structural features, it is common for RNA to be recognized by RNA-binding proteins in the minor groove.

RNA helices often contain noncanonical base pairs

In addition to conventional Watson–Crick base pairs, RNA double helices often contain noncanonical (non-Watson–Crick) base pairs. There are more than 20 different types of noncanonical base pairs, involving two
or more hydrogen bonds, that have been encountered in RNA structures. The most common are the GU wobble, the sheared GA pair, the reverse Hoogsteen pair, and the GA imino pair (Fig. 4.3). Because the GU pair only has two hydrogen bonds (compared with three for a GC pair), this requires a sideways shift of one base relative to its position in the regular Watson–Crick geometry. Weaker interactions from the reduction in hydrogen bonding may be countered by the improved base stacking that results from each sideways base displacement. In addition, RNA structures frequently involve unconventional base pairing such as base triples (Fig. 4.4). These base triples typically involve one of the standard base pairs, most commonly either a Watson–Crick or a reverse Hoogsteen pair. The third base can interact in a variety of unconventional ways. Noncanonical base pairs and base triples are important mediators of RNA self-assembly and of RNA–protein and RNA–ligand interactions. For example, noncanonical base pairs widen the major groove and make it more accessible to ligands.

### 4.3 Tertiary structure of RNA

RNA chains fold into unique three-dimensional structures that act similarly to globular proteins. Indeed, Francis Crick wrote in his 1966 paper in the *Cold Spring Harbor Symposium on Quantitative Biology* “tRNA looks like Nature’s attempt to make RNA do the job of a protein.” These remarks were made by Crick 2 years after the “cloverleaf” secondary structure of the transfer RNA (tRNA) for alanine in yeast was published by R.W. Holley and colleagues (Fig. 4.5). The actual shape of the functional tRNA in the cell is not an open cloverleaf. X-ray crystallography studies 10 years later showed that tRNA twists into an L-shaped three-dimensional structure. Many basic principles of RNA structure were learned from detailed analysis of both the secondary and tertiary structures of tRNAs. Obtaining crystal structures of larger RNA
molecules has proved to be a challenge. It was not until over 20 years later that structures were solved for larger RNAs, such as the 160 nt P4-P6 domain of a group I ribozyme (see below) and the ribosome subunits that are comprised of over 4500 nt of ribosomal RNA (rRNA) and more than 50 proteins. The structure of the ribosome will be discussed in detail in Chapter 14.

**tRNA structure: important insights into RNA structural motifs**

tRNA is transcribed as a molecule about twice as long as its final form. The pre-tRNA transcript is then processed by various nucleases at both the 5′ and 3′ end (see Section 4.6). After processing, the average tRNA is about 76 nt long, and all of the different tRNAs of a cell fold into the same general shape. One of the important insights into RNA structure came from the observation that the processed tRNA is further altered by the modification of bases.

**Modified bases**

In general, tRNAs contain more than 50 modified bases. Modifications range from simple methylation to complete restructuring of the purine ring (Fig. 4.6). Inosine (I) was the first modified nucleoside in tRNA to be identified. Nucleoside modifications are not unique to tRNA; for example, extensive base modification occurs during maturation of the ribosomal RNAs (see Section 13.9). The first modified nucleoside to be identified in any RNA was the ubiquitous pseudouridine (Ψ). Pseudouridine was discovered over 20 years earlier than inosine, but its role in tRNA function was not characterized until much later.

**tRNA loops each have a separate function**

Certain structural elements, of course, are unique to the function of tRNA. For example, every tRNA so far examined has the sequence ACC on the 3′ end to which the amino acid is attached (see Fig. 4.5).
The versatility of RNA

However, a general principle gleaned from studies of tRNAs is the importance of loop motifs. Each of the three tRNA loops that form the “cloverleaf” secondary structure seems to serve a specific purpose. These functions will be described in detail in Chapter 14 in the context of the mechanism of translation. In brief, the t-loop (or t-ψ-C loop) is involved in recognition by the ribosomes, the D loop (or dihydrouridine loop) is associated with recognition by the aminoacyl tRNA synthetases, and the anticodon loop base pairs with the codon in mRNA. The anticodon loop in all tRNA is bounded by uracil on the 5′ side and a modified purine on the 3′ side. This purine is always modified, but the modification varies widely. Another commonly observed motif in many RNAs is the “U turn.” In the anticodon loop of tRNA, hydrogen bonding of the N3 position of uridine with the phosphate group of a nucleotide three positions downstream causes an abrupt reversal or “U turn” in the direction of the RNA chain.

**Figure 4.5 Secondary and tertiary structure of tRNA.** (A) “Cloverleaf” secondary structure of alanine tRNA from yeast. The key structural features are labeled; note the modified bases in the loops. (B) L-shaped three-dimensional structure of tRNA showing the “arms” formed by coaxial stacking of the acceptor stem with the T stem, and the D stem with the anticodon stem. The arrow points to the U turn motif in the anticodon loop, which causes an abrupt reversal in the direction of the RNA chain.

However, a general principle gleaned from studies of tRNAs is the importance of loop motifs. Each of the three tRNA loops that form the “cloverleaf” secondary structure seems to serve a specific purpose. These functions will be described in detail in Chapter 14 in the context of the mechanism of translation. In brief, the t-loop (or t-ψ-C loop) is involved in recognition by the ribosomes, the D loop (or dihydrouridine loop) is associated with recognition by the aminoacyl tRNA synthetases, and the anticodon loop base pairs with the codon in mRNA. The anticodon loop in all tRNA is bounded by uracil on the 5′ side and a modified purine on the 3′ side. This purine is always modified, but the modification varies widely. Another commonly observed motif in many RNAs is the “U turn.” In the anticodon loop of tRNA, hydrogen bonding of the N3 position of uridine with the phosphate group of a nucleotide three positions downstream causes an abrupt reversal or “U turn” in the direction of the RNA chain (see Fig. 4.5).

**Coaxial stacking of stems**

Another important principle learned from the study of tRNA structure is that base-paired stems often are involved in long-range interactions with other stems by coaxial stacking. In tRNA, the 7 bp acceptor stem stacks on the 5 bp T stem to form one continuous A-type helical arm of 12 bp (see Fig. 4.5). The other two helices, the D stem and anticodon stem, also stack, although imperfectly, to form a second helical arm. The two coaxially stacked arms are what form the familiar L shape of tRNA. Coaxial stacking is a common feature of RNA. It is widespread in ribosomal RNA where continuous coaxial stacking of as many as 70 bp is found, and underpins the formation of pseudoknots (see below).
Common tertiary structure motifs in RNA

Large RNAs are composed of a number of structural domains that assemble and fold independently. RNA folding uses the two principal devices that were first seen in the double-helical structures of DNA and RNA: hydrogen bonding and base stacking. Preformed secondary structural domains of RNA interact to form the tertiary structure. Bases in loops and bulges that are supposedly unpaired are often involved in a variety of long-range interactions, forming noncanonical base pairs. The three-dimensional structure is maintained through these interactions between distant nucleotides and interactions between 2'-OH groups. These long-range interactions are less stable than standard Watson–Crick base pairs and can be broken under mild denaturing conditions. RNA is negatively charged, which makes tertiary structure formation a process that requires charge neutralization, either through binding of basic proteins, or binding of monovalent and/or divalent metal ions. There are a number of highly conserved, complex RNA folding motifs. Common motifs include the pseudoknot, the A-minor motif, tetraloops, ribose zippers, and kink-turns. The examples provided below highlight how these different motifs interact with each other in a modular fashion to form intricate folding patterns.

Figure 4.6 Structure of modified bases found in tRNA. The structures of nucleosides with normal bases and with modified bases are compared. Base modifications are highlighted in red.
**Pseudoknot motif**

A pseudoknot motif forms when a single-stranded loop base pairs with a complementary sequence outside this loop and folds into a three-dimensional structure by coaxial stacking (Fig. 4.7). The first experimental evidence for pseudoknot formation came from studies of a plant RNA virus in 1987. Now, many other pseudoknots have been identified in a wide variety of RNAs. For example, the 5' half of human telomerase RNA consists of the RNA template for telomere synthesis and a highly conserved pseudoknot that is required for telomerase activity. The function of telomerase is described in detail in Section 6.9. Solution structure of the pseudoknot from human telomerase RNA was determined by using nuclear magnetic resonance (NMR) spectroscopy (see Section 9.10 for method). An intricate network of tertiary interactions was shown to form a triple helix structure, stabilized by base triples. Mutations in the pseudoknot region are involved in the human genetic disease, dyskeratosis congenita (see Disease box 6.3). Although vertebrates, ciliates, and yeast have widely divergent telomerase RNAs, they all contain a pseudoknot motif.

**Figure 4.7 RNA pseudoknot motif.** (A) Schematic presentation of a pseudoknot found in the tRNA-like structure of the turnip yellow mosaic virus. S₁ and S₂ represent double helical stem regions. L₁ and L₂ indicate single-stranded loops. (i) Conventional secondary structure. (ii) Formation of stem S₁, simultaneously with S₂. (iii) Coaxial stacking of S₁ and S₂, forming a quasicontinuous double helix. (iv) Schematic three-dimensional representation. (Redrawn from Pleij, C.W.A., Rietveld, K., Bosch, L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Research* 13:1717–1730.) (B) Solution structure of the human telomerase RNA pseudoknot. The phosphate backbone is identified by a gray ribbon. Inset: schematic representation of the pseudoknot junction and tertiary structure, showing details of the extended triple helix surrounding the helical junction. Such multiple base triple interactions between loop 1 and stem 2 are unique among pseudoknot structures determined to date. Nucleotides are colored by structural element: stem 1 (red), stem 2 (blue), loop 1 (orange), and loop 2 (green). (Reproduced from Theimer, C.A., Blois, C.A., Feigon, J. 2005. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Molecular Cell* 17:671–682, Copyright © 2005, with permission from Elsevier.)
A-minor motif
The A-minor motif is one of the most abundant long-range interactions in large RNA molecules. This motif was first observed in the hammerhead ribozyme (see Section 4.7) and the P4-P6 domain of the group I ribozyme, and is found extensively in ribosomal RNAs. In this folding pattern, single-stranded adenosines make tertiary contacts with the minor grooves of RNA double helices by hydrogen bonding and van der Waals contacts (Fig. 4.8). The minor groove interactions have been likened to a “lock and key” because of the precise way in which the adenosines fit into the groove. The motif is stabilized by both base–base interactions and nucleoside–nucleoside interactions. Critical contacts are made within the riboses as well as the bases.

Figure 4.8 The A-minor motif. (A) Examples of the three most important kinds of A-minor motifs from the 23S ribosomal RNA (rRNA) of the archaeon *Haloarcula marismortui*, showing the precise lock-and-key minor groove interactions. Types I and II are adenine (A)-specific. Type III interactions involving other base types are seen, but A is preferred. (B) The interaction between helix 38 of 23S rRNA and 5S rRNA in *H. marismortui*. The only direct contact between these two molecules includes six A-minor interactions, involving three As in 23S rRNA and three As in 5S rRNA. Secondary structure diagrams are provided for the interacting sequences, with the As indicated in orange. (Reproduced from Nissen, P., Ippolito, J.A., Ban, N., Moore, P.B., Steitz, T.A. 2001. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proceedings of the National Academy of Sciences USA* 98:4899–4903. Copyright © 2001 National Academy of Sciences, USA.)
Tetraloop motif
The stability of a stem-loop structure is often enhanced by the special properties of the loop. For example, a stem loop with the "tetraloop" sequence UUUU is particularly stable due to special base-stacking interactions in the loop (Fig. 4.9). Tetraloops often include "G turns" in which a stabilizing hydrogen bond to the backbone phosphate is made from the 1-nitrogen position of a guanine base. Tetraloops are a prominent feature within the P4-P6 group I intron domain (Fig. 4.10).

Ribose zipper motif
Helix–helix interactions are often formed by “ribose zippers” involving hydrogen bonding between the 2'-OH of a ribose in one helix and the 2-oxygen of a pyrimidine base (or the 3-nitrogen of a purine base) of the other helix between their respective minor groove surfaces. Two ribose zippers are found in the P4-P6 group I intron domain (Fig. 4.10). One ribose zipper mediates the interaction between an A-rich bulge and the P4 stem. Another ribose zipper mediates a long-range interaction involving a tetraloop motif.

Kink-turn motif
Another type of motif first found in ribosomal RNA is the kink-turn or “K turn.” Kink-turns are asymmetric internal loops embedded in RNA double helices. The most striking feature is the sharp bend (the “kink”) in the phosphodiester backbone of the three-nucleotide bulge associated with this structure. In a kink-turn from the large ribosomal RNA of the extreme halophilic (salt-tolerant) archaean *Haloarcula marismortui*, each asymmetric loop is flanked by CG base pairs on one side and sheared GA base pairs on the other. Further illustrating how various structural motifs work together to define RNA shape, an A-minor interaction brings together these two helical stems (Fig. 4.11).

4.4 Kinetics of RNA folding
The structural flexibility of the RNA backbone and the propensity of nucleotides to base pair with short stretches of complementary regions can lead to difficulty in defining a single native structure, since there are many possible structures that a particular RNA chain can adopt. Misfolding, for example due to incorrect base pairing, is a problem for both secondary and tertiary structures. This “RNA folding problem” is not just a problem for the molecular biologist trying to determine the significance of predicted RNA secondary structures for function. Since only a single or a few possible structures lead to function, RNA itself must
Figure 4.10 Ribose zipper motif. (A) The secondary structure of the *Tetrahymena thermophila* ribozyme. The phylogenetically conserved catalytic core of the ribozyme is shaded in blue. Arrows indicate the 5′- and 3′-splice sites of this self-splicing group I intron. (B) The secondary structure of the P4–P6 domain is shown in more detail. Helical regions are numbered sequentially through the sequence; J, joining region; P, paired region. Nucleotides are highlighted as follows: blue and red, part of the conserved core; orange, the A-rich bulge; light green, the GAAA tetraloop; dark green, the conserved 11 nt tetraloop receptor; gray, P5c. (Reprinted with permission from Cate, J.H., Gooding, A.R., Podell, E., et al. (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* 273:1678–1685. Copyright © 1996 AAAS.) (C) Structure of the P4–P6 group I intron domain and its two ribose zippers. (i) One ribose zipper mediates the interaction between the A-rich bulge (orange) and the P4 stem (blue). The other ribose zipper mediates the interaction between the tetraloop (light green) and the tetraloop receptor (dark green). (ii) In the ribose zippers, there are two residues on each side (109–110, 184–183 and 152–153, 223–224) in which riboses interact by hydrogen bonding (yellow broken line) between the 2′-hydroxyl groups (O2′) of the two chain segments in an antiparallel orientation. The 2′-hydroxyl groups of the 3′ end residues also form minor groove hydrogen bonds to either the N3 atom of a purine (G110, A152) or the O2 atom of a pyrimidine (G109, C223) of the 5′ end residues on the opposite chain segment. (Reproduced from Tamura, M., Holbrook, S.R. 2002. Sequence and structural conservation in RNA ribose zippers. *Journal of Molecular Biology* 320:455–474, Copyright © 2002, with permission from Elsevier.)
The versatility of RNA

avoid the problem of misfolding into alternative, nonfunctional structures \textit{in vivo}. Specific RNA-binding proteins form tight complexes with their target RNAs and act as chaperones to aid in RNA folding (Fig. 4.12). One example is the family of heterogenous nuclear ribonucleoprotein (hnRNP) proteins. This group of more than 20 different proteins assists in preventing misfolding and aggregation of pre-mRNA. Another example is that of rRNA which folds correctly only by assembly with ribosomal proteins. The ribosomal proteins stage the order of folding of rRNA during ribosome assembly to avoid losing improperly folded RNA in kinetic folding traps.

Kinetic folding profiles were first established for tRNAs. The secondary structure for these small RNA molecules forms first within $10^{-4}$ to $10^{-3}$ seconds, followed by the tertiary structure in $10^{-2}$ to $10^{-1}$ seconds. The folding of the \textit{Tetrahymena thermophila} ribozyme (see Section 4.6) was recently analyzed using a hydroxyl radical footprinting assay (Fig. 4.13). The researchers generated hydroxyl radicals by radiolysis of water with a synchrotron X-ray beam. The short-lived hydroxyl radicals were able to break the ribozyme RNA

---

backbone only in places where it was accessible. As soon as the RNA formed a three-dimensional structure, the backbone region that was located inside the structure became inaccessible and was protected from cleavage. RNA folding could then be monitored by the appearance of the protected regions with time. The most stable domain was shown to form within several seconds, but the catalytic center of this large ribozyme required several minutes to complete folding. A portion of the catalytic center is susceptible to misfolding and the formation of an alternative helix. The resolution of this helix into the correct helix is a slow step.

4.5 RNA is involved in a wide range of cellular processes

Five major types of RNA serve unique roles in mediating the flow of genetic information (Fig. 4.14). Ribosomal RNA (rRNA) is an essential component of the ribosome. Messenger RNA (mRNA) is a copy of the genomic DNA sequence that encodes a gene product and binds to ribosomes in the cytoplasm. Transfer RNAs (tRNAs) are “charged” with an amino acid. They deliver to the ribosome the appropriate amino acid via interaction of the tRNA anticodon with the mRNA codon. Small nuclear RNA (snRNA) has a role in pre-mRNA splicing, a process which prepares the mRNA for translation, and small nucleolar RNA (snoRNA) has a role in rRNA processing. The role of RNA in RNA processing and translation is discussed in detail in Chapters 13 and 14, respectively.

This general way of thinking about the pathway of gene expression from DNA to functional product via an RNA intermediate overemphasizes proteins as the ultimate goal. What came as a surprise early on was
The versatility of RNA

The discovery of the tremendous variety and versatility of functional RNA products (Table 4.1). RNA is involved in a wide range of cellular processes along the pathway of gene expression, including DNA replication, RNA processing, mRNA turnover, protein synthesis, and protein targeting. One of the most important findings in molecular biology in the last 25 years was the discovery that RNA molecules can catalyze chemical reactions in living cells. This led to the hypothesis that the prebiological world was an “RNA world,” populated by RNAs that performed both the informational function of DNA and the catalytic function of proteins (Focus box 4.1).

Contributing to the versatility of RNA function is the ability of RNA to form complementary base pairs with other RNA molecules and with single-stranded DNA. The ability of RNA to make specific base pairs is key to understanding its role in everything from post-transcriptional gene silencing to translation. RNA–protein interactions are also of central importance. Most of the RNA in a eukaryotic cell is associated with protein as part of RNA–protein complexes termed ribonucleoprotein (RNP) particles. In addition, most, if not all, RNA-based catalytic reactions are thought to take place in conjunction with proteins. In other chapters, some of these important RNP complexes, such as the ribosome, are discussed in detail. Functional outcomes of RNA–nucleic acid and RNA–protein interactions are categorized below. Specific examples are highlighted in Table 4.1.
RNA can serve as a “scaffold.” An RNA molecule may act as a scaffold or framework upon which proteins can be assembled in an orderly fashion, as is the case in the signal recognition particle (SRP). Proteins recognize the primary nucleotide sequence of RNA and/or secondary and tertiary structural motifs.

RNA–protein interactions can influence the catalytic activity of proteins. In some catalytic RNPs, the protein functions as the enzyme, but the RNA is required to target or bind the enzyme to the substrate. An example of this is telomerase, where the RNA serves as the template for the addition of deoxynucleoside triphosphates (dNTPs) by the reverse transcriptase protein. In contrast, in other catalytic RNPs, such as ribonuclease (RNase) P and the ribosome, the RNA is catalytic, not the protein.

RNA can be catalytic. RNA molecules termed “ribozymes” can catalyze a number of the chemical reactions that take place in living cells (see Sections 4.6 and 4.7 below).

Small RNAs can directly control gene expression. Examples of how RNA plays a role in gene regulation will be discussed in detail in later chapters. These include differential RNA folding and riboswitches (Section 10.7), and RNA interference and microRNAs (Section 13.10).

RNA can be the hereditary material. Many viruses have RNA genomes and are either self-replicating or replicate through a DNA intermediate (see Section 3.7).
4.6 Historical perspective: the discovery of RNA catalysis

Thousands of different chemical reactions are required to carry out essential processes in living cells. These reactions may take place spontaneously, but they rarely occur at a rate fast enough to support life. Catalysis is necessary for these biochemical reactions to proceed at a useful rate. In the presence of a catalyst, reactions can be accelerated by a factor of a billion or even a trillion under physiological conditions, in a highly specific, regulated manner. For a very long time it was assumed that biological catalysis depended exclusively on protein enzymes. Then, in a landmark discovery at the beginning of the 1980s, two labs demonstrated independently that RNA can also possess catalytic activity.

Thomas Cech and co-workers published a report in 1982 that generated great excitement within the scientific community. In their paper they demonstrated that the single intron of the large ribosomal RNA of *Tetrahymena thermophila* has self-splicing activity *in vitro*. A year later, Sidney Altman and co-workers showed that the RNA component of RNase P from *Escherichia coli* is able to carry out processing of pre-tRNA in the absence of its protein subunit *in vitro*. The discovery of self-splicing RNA was completely unexpected. Needless to say, many control experiments had to be performed to convince all skeptics that RNA itself could possess catalytic activity. In 1989 Cech and Altman were awarded the Nobel Prize in chemistry for this revolutionary discovery. The following sections provide a brief synopsis of the experiments leading to this breakthrough and highlight some of the current research in the field.

**Tetrahymena group I intron ribozyme**

In 1979, Thomas Cech was studying transcription of ribosomal RNA genes from the ciliated protozoan *Tetrahymena thermophila*. After using the “R looping” technique of electron microscopy to hybridize 17S and 26S rRNA with ribosomal DNA, he saw the expected R loops caused by the rRNA hybridizing to the
Molecular biologists who speculate on the origins of life on earth are faced with a classic “chicken and egg problem” – which came first, proteins or nucleic acids? In the modern world, the replication of DNA and RNA is dependent on protein enzymes, and the synthesis of protein enzymes is dependent on DNA and RNA. The term “RNA world” was introduced by Walter Gilbert in 1986 to describe a hypothetical stage in the evolution of life some 4 billion years ago when RNA both carried the genetic information and catalyzed its own replication. The origin and prebiotic chemistry of this RNA world, of course, remains open to speculation.

According to the RNA world hypothesis, “life” first existed in the form of replicating RNA molecules (Fig. 1). In this ancient world neither protein nor DNA existed yet. Evidence in support of this hypothesis is that proteins cannot replicate themselves, except via mechanisms that involve an RNA intermediate. In contrast, RNA has all the structural prerequisites necessary for self-replication. RNA genomes are widespread among viruses and their replication in infected cells proceeds via complementary RNA chains. Compared with DNA or protein, RNA is clearly the most self-sufficient molecule. RNA molecules are capable of doing basically all that proteins can do. They can self-fold into specific three-dimensional structures, recognize other macromolecules and small ligands with precision, and perform catalysis of covalent reactions. Ribozymes can catalyze a diversity of reactions including polymerizing nucleotides, ligating DNA, cleaving DNA phosphodiester bonds, and synthesizing peptides. The later discovery that the ribosome – the catalyst still responsible for synthesizing nearly all proteins in cells – is, in fact, a ribozyme provides strong evidence for an RNA world.

At some point, requirements for enzymes with a greater repertoire of functional groups, more stable tertiary structure, and superior catalytic powers are thought to have favored the transition from RNA-based catalysis to protein-based catalysis that is present in the current DNA/RNA/protein world (Fig. 1). The original RNA world, if it ever existed on earth, is long gone. But a modern RNA world exists that has been vastly underestimated. Each year, more and more new species of noncoding RNAs with important roles in cells are being discovered (see Section 13.10).

**Figure 1 The RNA world and the transition to the present DNA/RNA/protein world.** (A) In the RNA world, RNA functioned as both a carrier of information and an enzyme. It catalyzed its own replication. (B) During the transitional period, RNA catalyzed the synthesis of proteins, and these proteins catalyzed the transition from RNA to DNA. (C) Today, proteins and RNA enzymes catalyze the replication of DNA. They also catalyze the transcription of DNA into RNA, and the reverse transcription of RNA into DNA. The translation of mRNA into proteins is mediated by the ribosome, a large ribozyme.
complementary DNA. He also saw a small loop structure that interrupted the R loop between 26S rRNA and DNA. This looped out stretch of DNA was an intervening sequence (IVS) or intron, which is spliced out in the final RNA product.

To follow up on this observation, Cech and colleagues attempted to develop an in vitro assay in which they could fractionate cell extracts and determine the proteins required for splicing. Completely unexpectedly, splicing of the rRNA intron
occurred in control experiments when the cell extract was left out of the reaction. The startling conclusion (after ruling out human error) was that the RNA was splicing itself. At the time, only proteins were thought to possess catalytic activity. Cech and his team spent a year trying to find alternative explanations. One possibility that had to be ruled out was that residual proteins remained associated with the RNA during its isolation. In 1982, they synthesized the precursor rRNA from a recombinant rDNA gene cloned in *E. coli*. The *in vitro*-generated rRNA was made using pure RNA polymerase in the absence of any other cellular proteins. In the presence of GTP and Mg$^{2+}$ the “naked” rRNA still underwent splicing, demonstrating unequivocally that the RNA was splicing itself. Self-splicing activities were determined by the amount of covalent addition of $^{32}$P-GTP to the 5$^{\prime}$ end of the intron RNA. Reactions that were characterized included the excision of the intervening sequence (intron), attachment of guanosine to the 5$^{\prime}$ end of the intron, covalent cyclization of the intron, and ligation of exons (Fig. 14.15). In 1986 Cech and colleagues engineered a variant ribozyme that worked as a true catalyst. The RNA enzyme was able to catalyze the cleavage and rejoining of oligonucleotide substrates in a sequence–dependent manner, and was regenerated to act again in the reaction.

The *Tetrahymena* ribozyme continues to be a paradigm for the study of RNA catalysis. A goal of Cech and his colleagues is to obtain three-dimensional structural information for each of the multiple steps along the self-splicing pathway. In their 2004 *Molecular Cell* paper Guo, Gooding, and Cech wrote: “Ultimately one would like to see a molecular movie of the entire series of reactions and to understand how group I introns with different secondary structures manage to accomplish the same splicing reactions.” Many molecular biologists will be scrambling for front row seats!

**RNase P ribozyme**

In 1971 Sidney Altman and co-workers began trying to purify and characterize RNase P, the enzyme involved in processing the 5$^{\prime}$-leader sequence of precursor tRNA in *E. coli*. After many attempts to remove the “contaminating” RNA from the preparation, 12 years later they demonstrated that the RNA component was in fact the biological catalyst. The RNase P RNA is a true RNA catalyst, acting on another RNA molecule without undergoing a chemical transformation itself.

*E. coli* RNase P is composed of M1 RNA, the catalytic RNA, and the C5 protein. *In vitro*, the M1 RNA alone can process precursor tRNA in the presence of high concentrations of monovalent and divalent cations. *In vivo*, the C5 protein is required to enhance the catalytic efficiency of M1 RNA and increase its substrate versatility. In contrast, in human cells, H1 RNA associates with at least 10 distinct protein subunits to form RNase P. The proposed tertiary structure of H1 RNA conforms to the catalytic core configuration of *E. coli* M1 RNA. However, H1 RNA shows no catalytic activity *in vitro*, unless associated with protein subunits Rpp21 and Rpp29 (Fig. 4.16). Eukaryotic RNase P is assembled in the nucleolus and shares some subunits with RNase MRP (mitochondrial RNA processing), including Rpp29 (see Table 4.1 and Section 6.7). Thus, while bacterial RNase P is an RNA enzyme, its eukaryotic counterpart acts as a catalytic ribonucleoprotein.

**4.7 Ribozymes catalyze a variety of chemical reactions**

RNA molecules with catalytic activity are called RNA enzymes or “ribozymes.” Naturally occurring ribozymes are often autocatalytic, which leads to their own modification. This characteristic contradicts the classic definition of an enzyme, which is “a substance that increases the rate, or velocity, of a chemical reaction without itself being changed in the overall process.” However, catalytic RNAs have been discovered that are true enzymes. For example, the 23S rRNA in the ribosome catalyzes peptide bond formation without being modified in the process (see Section 14.5).

**Mode of ribozyme action**

Ribozymes catalyze reactions essentially in the same ways that proteins do. They form substrate-binding sites and lower the activation energy of a reaction, thus allowing the reaction to proceed much faster.
Many ribozymes are metalloenzymes. Binding of divalent cations (e.g. Mg$^{2+}$) in the active site is critical for their folding into an active state. Interestingly, even though RNA enzymes and protein enzymes are not evolutionarily related, the active site of a self-splicing group I intron has the same orientation of two metal ions as found in a protein-based DNA polymerase (Fig. 4.17). This observation points to the importance of the two-metal-ion mechanism of catalysis in reactions involving phosphate transfer. However, ribozymes are not limited to using metal ions as functional groups in catalysis. Some ribozymes may use general acid–base chemistry, in which nucleotide bases, sugar hydroxyl groups, and even the phosphate backbone directly contribute to catalysis by donating or accepting protons during the chemical step of the reaction.

Naturally occurring ribozymes are classified into two different groups, the large and small ribozymes, based on differences in size and reaction mechanism.

**Large ribozymes**

The RNA component of RNase P, and members of the group I and group II intron family, belong to the group of large ribozymes. Group I and group II ribozymes are self-splicing introns that are discussed in detail in a later chapter on RNA processing (see Section 13.3). They vary in size from a few hundred nucleotides up to about 3000 nucleotides, and are further distinguished from the small ribozymes by all cleaving RNA to generate 3′-OH termini, as opposed to a product with a 2′,3′-cyclic phosphate and a product with a 5′-OH terminus (Table 4.2). Additional large ribozymes are the RNA components of the spliceosome, which also have enzymatic properties (see Section 13.5), and the ribosomal RNAs, characterized by their ability to catalyze peptide bond formation (see Section 14.5).
**Figure 4.17** Similarity between group I intron and protein-based DNA polymerase active sites. The active sites of a self-splicing group I intron and bacteriophage T7 DNA polymerase are compared. The 3′-exon is analogous to the primer oligonucleotide strand, the 5′-exon to the incoming deoxynucleotide triphosphate (dNTPs), and the ωG (the last nucleotide of the intron) to the pyrophosphate leaving group. Both sites contain two metal ions, M₁ (Metal ion A) and M₂ (Metal ion B), and coordinate those metals in a similar manner. In DNA polymerase, the two metals are held in place by interaction with two highly conserved aspartate residues. The active site Mg²⁺ ions are shown as large blue spheres, the predicted inner and outer sphere ligands are shown as small orange spheres, and the metal-to-metal distance is labeled. Orange lines indicate inner sphere coordinations. (A) Two-metal active site coordination within the group I intron active site. The splicing reaction involving attack on the phosphodiester bond between the exon and intron, with loss of ωG, is shown with curved arrows. (B) Two-metal active site coordination within the T7 DNA polymerase. M₁ (Metal ion A) interacts with the triphosphates of incoming dNTPs to neutralize their negative charge. After catalysis, the pyrophosphate product is stabilized through similar interactions with M₂ (Metal ion B). (Structures reprinted with permission from Stahley, M.R., Strobel, S.A. 2005. Structural evidence for a two-metal-ion mechanism of group I intron splicing. *Science* 309:1587–1590. Copyright © 2005 AAAS.)
Table 4.2 Types of naturally occurring ribozymes.

<table>
<thead>
<tr>
<th>Ribozyme</th>
<th>Source</th>
<th>Function</th>
<th>Reaction products</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Small ribozymes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hammerhead Ribozyme</td>
<td>Plant viroids and newt satellite RNAs</td>
<td>Replication</td>
<td>5′-OH; 2′,3′-cyclic phosphate</td>
</tr>
<tr>
<td>Hairpin Ribozyme</td>
<td>Plant satellite RNAs</td>
<td>Replication</td>
<td>5′-OH; 2′,3′-cyclic phosphate</td>
</tr>
<tr>
<td>HDV</td>
<td>Hepatitis delta virus (human)</td>
<td>Replication</td>
<td>5′-OH; 2′,3′-cyclic phosphate</td>
</tr>
<tr>
<td>VS</td>
<td>Neurospora crassa mitochondria</td>
<td>Replication</td>
<td>5′-OH; 2′,3′-cyclic phosphate</td>
</tr>
<tr>
<td>Riboswitch Ribozyme</td>
<td>Bacillus subtilis</td>
<td>glmS mRNA self-degradation</td>
<td></td>
</tr>
<tr>
<td><strong>Large ribozymes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNase P</td>
<td>Eukaryotes, prokaryotes</td>
<td>tRNA processing</td>
<td>5′-phosphate and 3′-OH</td>
</tr>
<tr>
<td>Group I introns</td>
<td>Eukaryotes (nucleus, organelles), prokaryotes, bacteriophages</td>
<td>Splicing</td>
<td>Intron with 5′-guanosine and 3′-OH, 5′/3′-ligated exons</td>
</tr>
<tr>
<td>Group II introns</td>
<td>Eukaryotes (organelles, prokaryotes)</td>
<td>Splicing</td>
<td>Intron with 2′-5′-lariat and 3′-OH, 5′/3′-ligated exons</td>
</tr>
<tr>
<td>Spliceosome</td>
<td>Eukaryotes (nucleus)</td>
<td>Pre-mRNA splicing</td>
<td>Intron with 2′-5′-lariat and 3′-OH, 5′/3′-ligated exons</td>
</tr>
<tr>
<td>Ribosome</td>
<td>Eukaryotes, prokaryotes</td>
<td>Translation</td>
<td>Peptide bond</td>
</tr>
</tbody>
</table>

**Small ribozymes**

The group of small ribozymes includes the hammerhead and hairpin motif, the hepatitis delta virus (HDV) RNA, the Varkud satellite (VS) RNA, and the glmS riboswitch ribozyme (Table 4.2). These five different ribozymes range in size from about 40 nt up to 154 nt. The hammerhead, so called for its three helices in a T shape, is the most frequently found catalytic motif in plant pathogenic RNAs, such as viroids (Fig. 4.18) (see Section 3.7). The hairpin ribozyme has only been found in some viroid-like satellite viruses of the human hepatitis B virus (HBV) that when present causes an exceptionally strong type of hepatitis in infected patients. The VS ribozyme is part of a larger RNA that is transcribed from a plasmid found in the mitochondria of some strains of *Neurospora crassa*, a filamentous fungus. The glmS riboswitch ribozyme is involved in regulating bacterial gene expression (see Section 10.7 for details).

With the exception of the riboswitch ribozyme, which is involved in gene regulation, the catalytic motifs in the small ribozymes are all involved in their self-replication. Replication of the circular RNAs occurs via a “rolling circle” mechanism (see Section 6.8). This leads to the formation of long linear transcripts consisting of monomers joined in tandem. These are self-cleaved into monomers by the catalytic motifs. The self-cleavage of phosphodiester bonds occurs by “in line” nucleophilic substitution (Fig. 4.18). The internal 2′-OH group of the ribose next to the phosphodiester bond to be cleaved attacks the phosphate, leading to an inversion of the configuration around the phosphorus. The incoming group is “in line” with the hydroxyl group in the transition state leaving the reaction center. The reaction yields a product with a 2′,3′-cyclic phosphate and a product with a 5′-OH terminus. This catalytic property suggests that viroids and other subviral pathogens may have an ancient evolutionary origin independent of viruses, dating back to the RNA world (see Focus box 4.1).

Since their discovery, small ribozymes have received much attention for their potential as tools to combat viral diseases. For example, ribozymes are being tested for their ability to inhibit the replication of human immunodeficiency virus type 1 (HIV-1), the causative agent of acquired immune deficiency syndrome (AIDS) (see Section 17.3).
Chapter summary

RNA is a chain-like molecule composed of subunits called nucleotides joined by phosphodiester bonds. Some of the common secondary structures that form the building blocks of RNA structure are bulges, base-paired A-type double helices (stems), single-stranded hairpin or internal loops, junctions, and turns. Base-paired stems often contain noncanonical base pairs, such as GU pairs or base triples. In addition,
RNA often contains a variety of modified nucleosides, such as inosine or pseudouridine. RNA chains fold into unique three-dimensional structures that act similarly to globular proteins. Important insights in RNA folding motifs have come from X-ray crystallographic studies of the structure of tRNA, group I introns, and rRNA. Preformed secondary structural domains of RNA fold to form a tertiary structure stabilized by many long-range interactions including coaxial stacking of helices, and formation of pseudoknots, A-minor motifs, tetraloops, ribose zippers, and kink-turn motifs. The structural flexibility of the RNA backbone and the tendency of nucleotides to base pair with complementary regions can lead to misfolding of RNA. Specific RNA-binding proteins form tight complexes with their target RNAs in vivo and act as chaperones to aid in proper RNA folding.

In addition to the five major types of RNA – rRNA, mRNA, tRNA, snRNA, and snoRNA – there is a tremendous diversity of functional RNA products. RNA is involved in a wide range of cellular processes along the pathway of gene expression from DNA replication to protein synthesis. Contributing to this versatility is the ability of RNA to form complementary base pairs with other RNAs and with single-stranded DNA, and to interact with proteins as part of RNP.

A landmark discovery in the late 1970s to early 1980s was that RNA can be catalytic. RNA molecules termed ribozymes catalyze a number of chemical reactions that take place in a living cell, ranging from cleavage of phosphodiester bonds to peptide bond formation. The first ribozymes discovered were a self-splicing intron in Tetrahymena thermophila rRNA and the RNA component of RNase P in E. coli. Many other ribozymes have been characterized since that time, including other self-splicing introns, components of the spliceosome, the rRNAs, and small ribozymes such as the hammerhead ribozyme which plays a role in self-replication.

**Analytical questions**

1. Make up an RNA sequence that will form a hairpin with a 9 bp stem and a 7 bp loop. Draw both the primary structure and the secondary structure.

2. What addition(s) would you need to make to the primary sequence in Question 1 to allow pseudoknot formation?

3. You suspect that a tetraloop is critical for the folding of a ribozyme into its active form. Describe an experiment to demonstrate whether the RNA folds into a similar tertiary structure when the tetraloop is deleted.

4. You have discovered a small RNA involved in the removal of a novel type of intron from another RNA transcript. Design an experiment to determine whether the small RNA functions as a catalytic RNA or RNP. Show sample positive results.

**Suggestions for further reading**


