

Chapter 2

Comparative Analysis of the Higher-Order Structure of RNA

Robin R. Gutell

Abstract

“If you want to understand function, study structure”

The functions of many RNA molecules are directly associated with their higher-order structure, and given the vast abundance of their functions in a cell, the determination of their structures should contribute significantly to our understanding of the cell. A variety of methods are used to determine their higher-order structure. A multitude of experimental methods are discussed elsewhere in this book. Computational methods are also used. The first type, considered one of the grand challenges in biology, utilizes different fundamental principles of RNA structure to predict their secondary and three-dimensional structure. While the accuracies of these methods have been improving, generally speaking, higher-quality structure information is obtained with experimental methods. In contrast to the computational methods that utilize first principles to predict the RNA's higher-order structure, comparative methods are utilized to infer structure, function, and evolution from the patterns of sequence and structure conservation and variation. The primary objective of this chapter is to briefly review the use of comparative analysis to deduce information about RNA structure, using an evolutionary framework.

Keywords RNA structure • Computational comparative analysis • RNA structure motifs • Covariation analysis

R.R. Gutell (✉)

Integrative Biology, Institute for Cellular and Molecular Biology, and the Center for Computational Biology and Bioinformatics, University of Texas at Austin, PAT 141, 2401 Speedway, Austin, TX 78712, USA
e-mail: robin.gutell@mail.utexas.edu

2.1 Fundamental Changes in Our Understanding of RNA Structure and Function

The central dogma, enunciated by Crick in 1958 and the keystone of molecular biology ever since, is likely to prove a considerable oversimplification (Anonymous 1970).

Since the central dogma in molecular biology was established (Crick 1958), the primary role for RNA has been associated with protein synthesis. Molecular biologists, at the onset of the discovery of replication, transcription, and translation focused primarily on proteins and DNA, since it was already known that proteins can form three-dimensional structures that catalyze reactions, and DNA was known to contain the instructions to make the proteins. Within this process, the primary role for RNA was thought to be the messenger RNA, which carries the information from the DNA to the ribosome to code for the proteins. In addition, it was known that transfer RNAs assign amino acids to their proper codon assignment, and ribosomal RNA is part of the ribosome. Still, these RNAs were initially perceived as simply coding and structural, not dynamically involved in catalytic functions, and while this central dogma is still correct, it primarily reveals only the protein's and DNA's role in the metabolism and regulation of the cell. The significance of RNA structure and function in the cell had been minimal with the central dogma as articulated in 1958 (Crick 1958).

Dennis Overbye stated in the New York Times (July 27, 2011, (http://www.nytimes.com/2011/07/28/science/28life.html?_r=4&ref=science) "... RNA, or ribonucleic acid, ... plays Robin to DNA's Batman in Life As We Do Know It, assembling proteins in accordance with the blueprint encoded in DNA." While our understanding of RNA's structure and function did not change for the first 20 or so years after the central dogma was proposed, it was postulated, based on theoretical considerations, that RNA came before DNA and proteins (Woese 1967; Crick 1968; Orgel 1968). RNA has characteristics of DNA and protein. RNA, like DNA, has similar rules for base pairing - adenine pairs with uracil (thymine) and guanine pairs with cytosine. 'Canonical' base pairs that are consecutive and antiparallel on an RNA sequence form standard helices, and like proteins, RNA forms three-dimensional structures, which for RNA are composed of helices, hairpin, internal, and multistem loops, and other structural motifs (Moore 1999).

Experimental evidence, beginning in the 1970s, started to suggest that rRNA was directly involved in protein synthesis (Noller and Chaires 1972). During the early 1980s a series of studies revealed that the group I intron and RNase P were directly involved in the chemical catalysis of RNA (Kruger et al. 1982; Guerrier-Takada et al. 1983). Subsequently, other RNAs were identified and characterized that catalyze chemical reactions, including riboswitches (Haller et al. 2011; Breaker 2012), while it was determined that RNA has the capacity of catalyzing many different types of chemical reactions (Hiller and Strobel 2011), including the primary steps in decoding and peptidyl transferase during protein synthesis (Moore and Steitz

2011) (Ogle et al. 2001; Noller 2006). Beyond the many functions of RNA's three-dimensional structures, small and large RNAs are being implicated in the regulation of nearly all of the cell's functions. Accordingly different RNAs are associated with many diseases and other anomalies in the cell.

2.2 Comparative Analysis: An Introduction

In the 1830s Darwin used comparative analysis to identify patterns in the anatomical features of some animals and in the process determined fundamental principles about the evolution of biological species (Darwin 1859). More recently, comparative analysis has been used to study macromolecular structure. Once the first few transfer RNA sequences were determined in the early 1960s, it was appreciated that the three-dimensional structures of tRNA would be very similar although their nucleic acid sequences could share little identity with one another. The cloverleaf secondary structure of tRNA, with approximately 76 nucleotides, was determined to be common to all of the known tRNA sequences (Holley et al. 1965; Madison et al. 1966; RajBhandary et al. 1966). Subsequent analysis revealed that the probability that 14 tRNA sequences could all form the same cloverleaf secondary structure by coincidence is 1 in 10^{20} (Levitt 1969). This latter analysis also revealed a few tertiary-structure interactions. This approach to the determination of RNA's secondary structure was substantiated when the proposed tRNA secondary structure and a few of the tertiary-structure base pairs were confirmed with crystallography (Kim et al. 1974; Robertus et al. 1974). This success with tRNA was the foundation for comparative methods to be utilized for the identification of higher-order structures that are conserved in different RNA families. In 1975, the secondary structure for 5S ribosomal RNA, a molecule approximately 120 nucleotides long, was initially proposed with comparative methods (Fox and Woese 1975). Subsequently, once a few 16S and 23S rRNA sequences were determined in the late 1970s and early 1980s, the minimal secondary-structure models were determined for these RNAs that are approximately 1,540 and 2,900 nucleotides long in bacteria (Woese et al. 1980; Branlant et al. 1981; Glotz et al. 1981; Noller et al. 1981; Stiegler et al. 1981; Zwieb et al. 1981). During the 1980s, other RNA molecules were studied with comparative methods, including group I (Cech 1988; Michel and Westhof 1990) and II (Michel et al. 1989) introns, ribonuclease (RNase) P RNA (James et al. 1988), U-RNAs (U1, U2, U4, U5, and U6) (Guthrie and Patterson 1988), 7S SRP RNA (Zwieb 1989), and telomerase RNA (Romero and Blackburn 1991). More recently, the secondary structures for many other RNA types have been elucidated with comparative analysis (Gardner et al. 2009) due to our current appreciation that RNA is directly involved in many, if not all, of the regulations in the cell, and the advent of ultrarapid nucleic acid sequencing that is providing us with the genetic blueprints for a very large number of organisms that span across the entire tree of life.

2.3 Covariation Analysis: Identification of Canonical and Noncanonical Base Pairs

The primary method for the identification of a common structure is based on a very simple principle. While the primary structure (or sequence) of RNAs within the same family can have significant variation with one another, base pairing, the dominant element in RNA structure, can be conserved in the secondary and three-dimensional structure of RNAs. As a consequence, a very large number of RNA sequences can be mapped to the same secondary and three-dimensional structure. In practice, the most common means to determine this common structure is from the analysis of the patterns of variation in an alignment of the sequences. Initially, when the number of sequences in an alignment was small, sub-sequences that had the potential to form G:C, A:U, and G:U base pairs within a helix were identified. Those potential helices with at least two exchange (or covariation) of one canonical base pair with another were considered a possible helix (Noller et al. 1981). As the number of sequences in an alignment increased, covariation algorithms were developed to identify those positions with similar patterns of variation (Olsen 1983; Gutell et al. 1985, 1992; Gautheret et al. 1995). These latter methods did not specifically search for G:C, A:U, and G:U base pairs that occur within a potential canonical helix.

The most recent comparative structure model for *Escherichia coli* 16S rRNA (Fig. 2.1) is the culmination of approximately 30 years of comparative analysis. The coloring of the base pair symbols reveals our confidence for every proposed base pair shortly prior to the determination of the high-resolution crystal structures (see below). Red indicates the base pairs with the most significant covariation (strongest confidence), followed by green and black. Black indicates a minimal amount of covariation and/or variation at one of the paired positions but no corresponding variation at the other paired position, and gray and blue indicate nucleotide conservation greater than 98% for G:C, A:U, and G:U base pairs within a canonical or compound helix that has strong support of other base pairs. The same coloring for each of the proposed base pairs is associated with the base-pair-frequency tables at the Gutell lab's Comparative RNA Web (CRW) Site [http://www.rna.cccb.utexas.edu/SAE/2A/nt_Frequency/BP/16S_Model]. Note that the vast majority of the proposed base pairs in 16S rRNA have a red base-pair symbol. The small number of black, gray, and blue usually occurs at the ends of the helices.

The results from these covariation methods were very profound. While the majority of the sets of positions with similar patterns of variation in the rRNAs contained G:C, A:U, and G:U base pairs that occur within a helix, a small number of covariations contained base pairs that were irregular (Gutell 1993; Gutell et al. 1994). Thus, covariation analysis, a specific type of comparative analysis, has independently identified two of the most fundamental principles of nucleic acid structure: (1) base pairings that are composed of G:C, A:U, and G:U, and (2) these base pairs are arranged adjacent and antiparallel with one another to form a helix. Given this recapitulation of these two canonical structural elements in RNA, we are compelled to accept, or at least seriously consider, the noncanonical structural

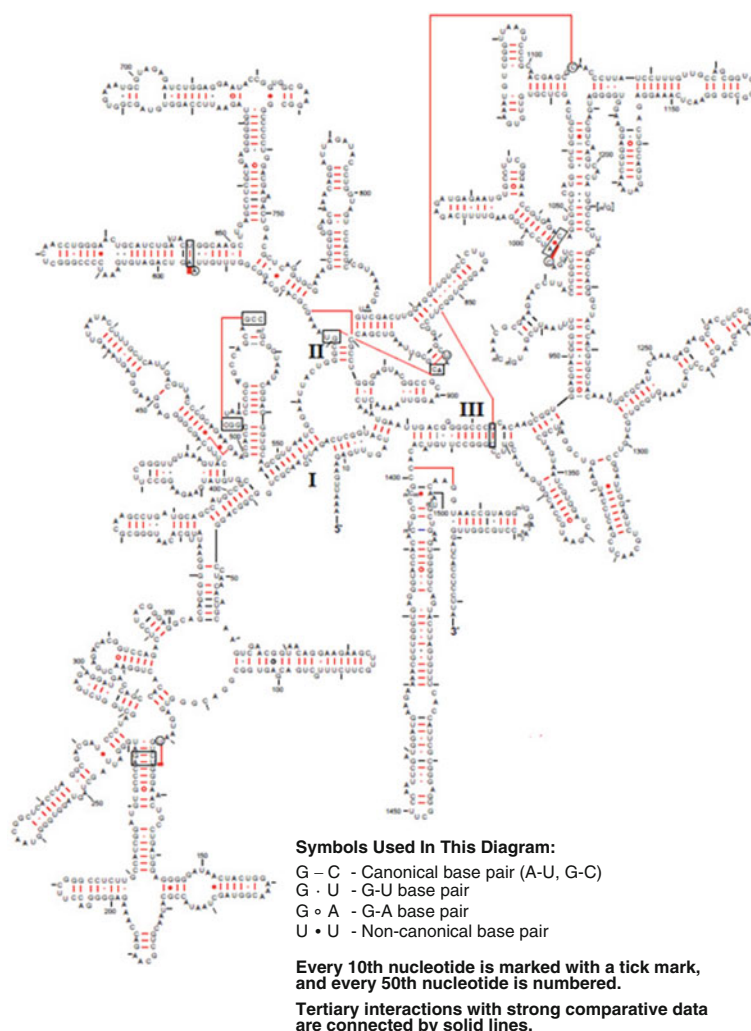


Fig. 2.1 Comparative *Escherichia coli* 16S ribosomal RNA secondary-structure model (Cannone, Subramanian et al. 2002)[<http://www.rna.cccb.utexas.edu/>]

elements identified from the covariation analysis of the rRNAs. These include the following:

- *Non-Canonical base pairs:* Several types of noncanonical base-pair exchanges have been identified. The most common exchanges are A:A \leftrightarrow G:G (i.e., exchanges between A:A and G:G pairs), G:U \leftrightarrow A:C, C:C \leftrightarrow U:U, A:G \leftrightarrow G:A, and G:U \leftrightarrow A:C. These noncanonical base pairs usually occur at the end of a regular canonical helix or as a lone pair not flanked by other base pairs.

- *Lone Pairs*: A Lone pair (individual or isolated base pairs) is not flanked by a base pair on its 5' or 3' end. Lone pairs are not stable enough to occur without additional interactions, such as base stacking or being flanked by nucleotides involved in a tertiary interaction. They occur in several structural environments, including internal loops, multistem loops, and between two hairpin loops (one form of a pseudoknot).
- *Lone-pair tri-loops*: A special class of lone pairs contains a single base pair capped by a hairpin loop with only three nucleotides. Several of these that were identified with covariation analysis occur in the rRNAs (Gutell 1996). All of these are immediately 3' to a secondary-structure helix, suggesting that they form a coaxial stack with the 5' helix. An analysis of the high-resolution crystal structure of the rRNAs revealed that this motif occurs frequently in the rRNAs, and all of them are 3' to an existing helix and all of them are coaxially stacked onto this helix (Lee et al. 2003).
- *Pseudoknots*: Pseudoknots are defined as at least one base pair that crosses a secondary-structure helix. These usually vary from one to three base pairs in length in the rRNAs, and are usually always immediately adjacent to a secondary-structure helix, suggesting that they can form a coaxial stack with these adjacent helices. Nearly 20 pseudoknots were identified with covariation analysis in the rRNAs (Gutell et al. 1986; Gutell and Woese 1990; Alkemer and Nygard 2003).
- *Parallel arrangement of base pairs*: While nearly all of the adjacent base pairs are arranged antiparallel with one another, a few of the base pairs identified with covariation analysis are parallel. The most prominent example occurs in domain V of 23S rRNA. Here positions 2112:2169, 2113:2170, and 2117:2172 form base pairs. While 2112:2169 contains A:G and G:A base pair types, the latter two base pairs exchange primarily between C:G \leftrightarrow U:A and G:C \leftrightarrow A:U, respectively (Gutell 1993).
- *Base triples*: Covariation analysis has identified several base pairs that covary with a third "unpaired" nucleotide. The best candidates include the following: (1) in the 16S rRNA - position 121 with either the 124:237 or 125:236 base pair, 863 with the 570:866 base pair, position 595 with the 596:644 base pair, and (2) in the 23S rRNA: position 2011 with the 2144:2147 base pair and between positions 1072 and the 1092:1099 base pair (Gautheret et al. 1995; Conn et al. 1998).
- *Non-base-pair constraints*: While all of the previous constraints (or dependencies between the evolution of different positions) are associated with a base pair, covariation analysis has also identified weaker albeit significant covariations between positions that are not base paired. One of the first examples, initially published in 1992 and elaborated on thereafter (Gutell et al. 1992; Gutell 1993; Gautheret et al. 1995), revealed that eight of the nucleotides in the D helix and the variable loop co-evolve in the type-I tRNAs. Our rationale for this set of eight co-evolving nucleotides is associated with the structural and evolutionary dynamics of several base triples with several consecutive base pairs. This structural constraint restricts the types of changes that can occur at other positions that are in close proximity in three dimensions. Other examples of non-base-pair constraints have been identified in the group I intron and the rRNAs (Shang et al. 2012).

2.4 Accuracy of the Covariation-Based Higher-Order Structure

As noted earlier, the comparative secondary-structure model for tRNA was substantiated with its high-resolution crystal structure (Kim 1976; Rich and RajBhandary 1976). While all of the secondary-structure base pairs and a few of the tertiary interactions predicted with comparative analysis were in the crystal structure, several tertiary-structure interactions in the crystal structure were not identified with comparative analysis. The comparative structure models for 5S, 16S, and 23S ribosomal RNA, the culmination of more than 30 years of an initial predicted structure followed by refinements as the number of sequences and the diversity of the sequences increased, and multiple improvements in the covariation algorithms (Cannone et al. 2002) were compared with the high-resolution crystal structures that were determined in 2000 (Ban et al. 2000; Wimberly et al. 2000). Of the 476 base pairs in the predicted 16S rRNA secondary-structure model, including a small number of tertiary-structure interactions (e.g., noncanonical base pairs, base triples, base pairs not in canonical helices), 461 (or 97%) were in the 30S ribosomal crystal structure. Of the 797 base pairs in the 23S rRNA secondary-structure model (including the small number of tertiary-structure interactions), 779 (or 98%) were in the 50S ribosomal crystal structure. Nearly all of the base pairs that were predicted with comparative methods, but not in the high-resolution crystal structure, were base pairs with minimal or no covariation, and, accordingly, those base pairs with sufficiently large amounts of covariation were present in the crystal structure (Gutell et al. 2002). Our analysis of the crystal structures of the ribosome also revealed 56 and 425 base-base interactions in the 16S and 23S rRNA, respectively, that were not predicted with comparative analysis. An analysis did not reveal any significant covariation in nearly all of these base pairs first identified in the crystal structures (Shang et al. 2012).

2.5 Structural Motifs

Comparative analysis reveals more than just the base pairs that have covariation at two paired positions. Comparative analysis has been used to identify structural motifs that are the basic building blocks of RNA structure. Earlier it was noted that covariation analysis has independently determined two of the most fundamental principles of RNA structure: (1) the base pair and the most frequent pairing types, G:C, A:U, and G:U, and (2) the arrangement of consecutive and antiparallel base pairs into a helix. Covariation analysis also revealed several other types of noncanonical base-pair types and noncanonical arrangements of base pairs in context with other structural elements. All of these “non-canonical” base pairs were present in the high-resolution crystal structure. We now question if comparative analysis can be utilized to identify structural elements that are present in similar structural environments

that do not have a covariation signal. Below are a few of the many structural motifs that have been identified.

Unpaired Adenosines: An analysis in 1985 revealed that approximately 66% of the adenosines in the *Escherichia coli* 16S rRNA comparative secondary-structure model were unpaired, while only approximately 30% of the G's, C's, and U's were unpaired (Gutell et al. 1985). A more comprehensive analysis in 2000 revealed that these biases in the distribution of the four nucleotides in a large sampling of bacterial 16S and 23S rRNAs were approximately the same as for *E. coli* 16S rRNA. This study revealed many other biases in the distribution of nucleotides in the paired (helices) and unpaired (loops) regions. A few biases worthy of mention are as follows: (1) more than 50% of the 3' ends of a loop contain an A that is conserved at that location in more than 95% of the sequences, (2) G and A are the two most frequent nucleotides at the 5' ends of a loop, and (3) the most frequent consecutive nucleotides are GG, GA, AG, and AA, with ~70% of the GG occurring within a helix, ~70% of the AA occurring in loops. These and other observations from this analysis are consistent with the distribution of nucleotides in GNRA tetraloops (see below), adenosine platforms (Cate et al. 1996), E and E-like loops, and AA and AG juxtapositions flanking the ends of a helix (see below) (Gutell et al. 2000).

Tandem GA & AA.AG@helix.ends: Studies have revealed that tandem G:A juxtapositions occur frequently within helices. The most frequent orientation of the tandem G:A juxtapositions is 5' *N:N'* G:A A:G *M:M'* 3', where *N* and *N'*, and *M* and *M'* can be any set of canonical base pairs flanking the tandem GAs. The G is 3' to a nucleotide that is base paired within a helix and the A is 5' to a nucleotide that is base paired within a helix. The G is frequently exchanged with an A. The G:A and A:A juxtapositions usually form the sheared conformation when this tandem is within a helix. The G is rarely 5' to a nucleotide that is base paired within a helix (SantaLucia et al. 1990) (Gautheret et al. 1994). It was also observed that in the ribosomal RNAs, helices are frequently flanked by a G:A juxtaposition on the loop side of a canonical helix (Traub and Sussman 1982) (Elgavish et al. 2001). With a significantly larger dataset of comparative rRNA secondary structures, it was observed that the Gs in the G:A juxtaposition are replaced with an A. The G in the G:A juxtaposition is nearly always at the 3' end of the helix. The majority of these A:A and A:G at the ends of helices form a base pair (sheared conformation) in the 16S and 23S rRNA crystal structures. The AA & AG at helix end motif occurs within several larger structural motifs—GNRA tetraloops (see below), E and E-like loops, tandem G:A base pairs, U-turns (see below), and adenosine platforms (Cate et al. 1996).

Tetraloops: It was observed that the sequence for the hairpin loop with four nucleotides is frequently GNRA in the group I introns (Michel and Westhof 1990). The majority of the hairpin loops in the rRNAs have four nucleotides, and the majority of these have the GNRA tetraloop sequence (where *N* is any nucleotide and *R* is a purine, either A or G) (Woese et al. 1990). While this GNRA tetraloop is the most frequent hairpin loop with four nucleotides, tetraloops with the sequences UUCG and CUUG also occur frequently in the rRNAs (Woese et al. 1990). Nearly all of the

GNRA tetraloops in the ribosomal RNAs are involved in a tertiary interaction. Comparative analysis reveals the conservation and variation at each of the tetraloops for any portion of the phylogenetic tree. While some tetraloops are invariant, others exchange primarily between the different sequences within the GNRA family, and others exchange between the GNRA, UUCG, CUUG, and possibly other sequences. Notable is the tetraloop at position 83–86 in 16S rRNA. The primary sequences observed are GCAA, UUCG, and CUUG. The rate of exchanges between these sequences is high, as gauged by mapping these tetraloop sequences onto the phylogenetic tree. All three sequences are present in all of the major phylogenetic groups in the bacteria [see Table 1 in (Woese et al. 1990)]. The different rates of evolution of the tetraloops and the different compositions present at each tetraloop location suggest that tetraloops have different functions. While the primary tetraloop sequences are known to be more stable than other hairpin loops, UUCG tetraloops are known to be particularly stable (Tuerk et al. 1988) and are likely to be nucleating the formation of a helix during RNA folding. In contrast, nearly all of the GNRA tetraloops in the rRNA crystal structures form tertiary-structure interactions.

2.6 Future Prospects

These examples of structural motifs in RNA are only a partial list. They reveal some of the utilities and latitude that comparative analysis offers. At this stage we address a variety of questions to assess the full potential of comparative analysis.

The operational premise for comparative analysis is based on one of the major discoveries in molecular biology since the elucidation of the double helix. The three-dimensional structure of proteins and nucleic acids can remain relatively constant during significant evolutionary changes in the macromolecule's primary structure. While this premise is widely used in the study of many macromolecules and their functions in the cell, its full extent has not been fully explored.

What is the relationship between RNA's sequence variation and the variation in its secondary structure and its three-dimensional structure? What is the maximum amount of variation that is possible between two sequences in the same RNA family?

While the positions that form secondary-structure base pairs generally covary with one another, the majority of the positions that form tertiary-structure base pairs do not have a simple covariation. The pattern of covariation in secondary-structure base pairs is simple. Do the tertiary-structure base pairs have patterns of variation that can be deciphered and utilized to predict these base pairs with comparative sequence and structure information?

Can comparative methods be used to identify more structural motifs that do not have any obvious positional covariation? Ultimately we wonder how much of the higher-order structure for an RNA can be inferred with comparative methods?

Acknowledgements The author appreciates the funding from the National Institutes of Health (GM067317) and the Welch Foundation (F-1427) to support the research in the Gutell lab.

References

- Alkemar G, Nygard O (2003) A possible tertiary rRNA interaction between expansion segments ES3 and ES6 in eukaryotic 40S ribosomal subunits. *RNA* 9(1):20–24
- Anonymous (1970). “Central Dogma Reversed.” *Nature* 226(5252): 1198–1199
- Ban N, Nissen P et al (2000) The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289(5481):905–920
- Branlant C, Krol A et al (1981) Primary and secondary structures of *Escherichia coli* MRE 600 23S ribosomal RNA. Comparison with models of secondary structure for maize chloroplast 23S rRNA and for large portions of mouse and human 16S mitochondrial rRNAs. *Nucleic Acids Res* 9(17):4303–4324
- Breaker RR (2012) Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* 4(2)
- Cannone JJ, Subramanian S et al (2002) The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 3(1):2
- Cate JH, Gooding AR et al (1996) RNA tertiary structure mediation by adenosine platforms. *Science* 273(5282):1696–1699
- Cech TR (1988) Conserved sequences and structures of group I introns: building an active site for RNA catalysis—a review. *Gene* 73(2):259–271
- Conn GL, Gutell RR et al (1998) A functional ribosomal RNA tertiary structure involves a base triple interaction. *Biochem* 37(34):11980–11988
- Crick F (1988) *What Mad Pursuit: A Personal View of Scientific Discovery*. Basic Books, New York
- Crick FH (1958) On protein synthesis. *Symp Soc Exp Biol* 12:138–163
- Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38(3):367–379
- Darwin C (1859) *On the Origin of Species*. John Murray, London, England
- Elgavish T, Cannone JJ et al (2001) AA.AG@helix.ends: A:A and A:G base-pairs at the ends of 16S and 23S rRNA helices. *J Mol Biol* 310(4):735–753
- Fox GE, Woese CR (1975) 5S RNA secondary structure. *Nature* 256(5517):505–507
- Gardner PP, Daub J et al (2009) Rfam: updates to the RNA families database. *Nucleic Acids Res* 37(Database issue):D136–D140
- Gautheret D, Damberger SH et al (1995) Identification of base-triples in RNA using comparative sequence analysis. *J Mol Biol* 248(1):27–43
- Gautheret D, Konings D et al (1994) A major family of motifs involving G.A mismatches in ribosomal RNA. *J Mol Biol* 242(1):1–8
- Glitz C, Zwieb C et al (1981) Secondary structure of the large subunit ribosomal RNA from *Escherichia coli*, *Zea mays* chloroplast, and human and mouse mitochondrial ribosomes. *Nucleic Acids Res* 9(14):3287–3306
- Guerrier-Takada C, Gardiner K et al (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* 35(3 Pt 2):849–857
- Gutell R (1993) Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. *Curr Opin Struct Biol* 3:313–322
- Gutell RR (1996) Comparative sequence analysis and the structure of 16S and 23S rRNA. In: Zimmermann RA, Dahlberg AE (eds) *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis*. CRC Press, Boca Raton, FL, pp 111–128
- Gutell RR, Cannone JJ et al (2000) A story: unpaired adenosine bases in ribosomal RNAs. *J Mol Biol* 304(3):335–354
- Gutell RR, Larsen N et al (1994) Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. *Microbiol Rev* 58(1):10–26
- Gutell RR, Lee JC et al (2002) The accuracy of ribosomal RNA comparative structure models. *Curr Opin Struct Biol* 12(3):301–310
- Gutell RR, Noller HF et al (1986) Higher order structure in ribosomal RNA. *EMBO J* 5(5):1111–1113

- Gutell RR, Power A et al (1992) Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. *Nucleic Acids Res* 20(21):5785–5795
- Gutell RR, Weiser B et al (1985) Comparative anatomy of 16-S-like ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* 32:155–216
- Gutell RR, Woese CR (1990) Higher order structural elements in ribosomal RNAs: pseudo-knots and the use of noncanonical pairs. *Proc Natl Acad Sci USA* 87(2):663–667
- Guthrie C, Patterson B (1988) Spliceosomal snRNAs. *Annu Rev Genet* 22:387–419
- Haller A, Souliere MF et al (2011) The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc Chem Res* 44(12):1339–1348
- Hillier DA, Strobel SA (2011) The chemical versatility of RNA. *Philos Trans R Soc Lond B Biol Sci* 366(1580):2929–2935
- Holley RW, Apgar J et al (1965) Structure of a ribonucleic acid. *Science* 147:1462–1465
- James BD, Olsen GJ et al (1988) The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell* 52(1):19–26
- Kim SH (1976) Three-dimensional structure of transfer RNA. *Prog Nucleic Acid Res Mol Biol* 17:181–216
- Kim SH, Suddath FL et al (1974) Three-dimensional tertiary structure of yeast phenylalanine transfer RNA. *Science* 185(419):435–440
- Kruger K, Grabowski PJ et al (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* 31(1):147–157
- Lee JC, Cannone JJ et al (2003) The lonepair triloop: a new motif in RNA structure. *J Mol Biol* 325(1):65–83
- Levitt M (1969) Detailed molecular model for transfer ribonucleic acid. *Nature* 224(221):759–763
- Madison JT, Everett GA et al (1966) On the nucleotide sequence of yeast tyrosine transfer RNA. *Cold Spring Harb Symp Quant Biol* 31:409–416
- Michel F, Umesono K et al (1989) Comparative and functional anatomy of group II catalytic introns—a review. *Gene* 82(1):5–30
- Michel F, Westhof E (1990) Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216(3):585–610
- Moore PB (1999) Structural motifs in RNA. *Annu Rev Biochem* 68:287–300
- Moore PB, Steitz TA (2011) The Roles of RNA in the Synthesis of Protein. *Cold Spring Harb. Perspect, Biol*, 2
- Noller HF (2006) Biochemical characterization of the ribosomal decoding site. *Biochimie* 88(8):935–941
- Noller HF, Chaires JB (1972) Functional modification of 16S ribosomal RNA by kethoxal. *Proc Natl Acad Sci USA* 69(11):3115–3118
- Noller HF, Kop J et al (1981) Secondary structure model for 23S ribosomal RNA. *Nucleic Acids Res* 9(22):6167–6189
- Ogle JM, Brodersen DE et al (2001) Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* 292(5518):897–902
- Olsen GJ (1983) Comparative analysis of nucleotide sequence data Ph.D., University of Colorado.
- Orgel LE (1968) Evolution of the genetic apparatus. *J Mol Biol* 38(3):381–393
- RajBhandary UL, Stuart A et al (1966) Nucleotide sequence studies on yeast phenylalanine sRNA. *Cold Spring Harb Symp Quant Biol* 31:425–434
- Rich A, RajBhandary UL (1976) Transfer RNA: molecular structure, sequence, and properties. *Annu Rev Biochem* 45:805–860
- Robertus JD, Ladner JE et al (1974) Structure of yeast phenylalanine tRNA at 3 Å resolution. *Nature* 250(5467):546–551
- Romero DP, Blackburn EH (1991) A conserved secondary structure for telomerase RNA. *Cell* 67(2):343–353
- SantaLucia J Jr, Kierzek R et al (1990) Effects of GA mismatches on the structure and thermodynamics of RNA internal loops. *Biochem* 29(37):8813–8819

- Shang L, Xu W et al (2012) Structural constraints identified with covariation analysis in ribosomal RNA. *PLoS ONE*, 7(6):e39383
- Stiegler P, Carbon P et al (1981) A general secondary-structure model for procaryotic and eucaryotic RNAs from the small ribosomal subunits. *Eur J Biochem* 120(3):487–495
- Traub W, Sussman JL (1982) Adenine-guanine base pairing ribosomal RNA. *Nucleic Acids Res* 10(8):2701–2708
- Tuerk C, Gauss P et al (1988) CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci USA* 85(5):1364–1368
- Wimberly BT, Brodersen DE et al (2000) Structure of the 30S ribosomal subunit. *Nature* 407(6802):327–339
- Woese CR (1967) The genetic code: the molecular basis for genetic expression (Modern perspectives in biology). Harper & Row, New York
- Woese CR, Magrum LJ et al (1980) Secondary structure model for bacterial 16S ribosomal RNA: phylogenetic, enzymatic and chemical evidence. *Nucleic Acids Res* 8(10):2275–2293
- Woese CR, Winker S et al (1990) Architecture of ribosomal RNA: constraints on the sequence of “tetra-loops”. *Proc Natl Acad Sci USA* 87(21):8467–8471
- Zwieb C (1989) Structure and function of signal recognition particle RNA. *Prog Nucleic Acid Res Mol Biol* 37:207–234
- Zwieb C, Glotz C et al (1981) Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. *Nucleic Acids Res* 9(15):3621–3640



<http://www.springer.com/978-1-4614-4954-6>

Biophysics of RNA Folding

Russell, R. (Ed.)

2013, VI, 238 p.,

ISBN: 978-1-4614-4954-6