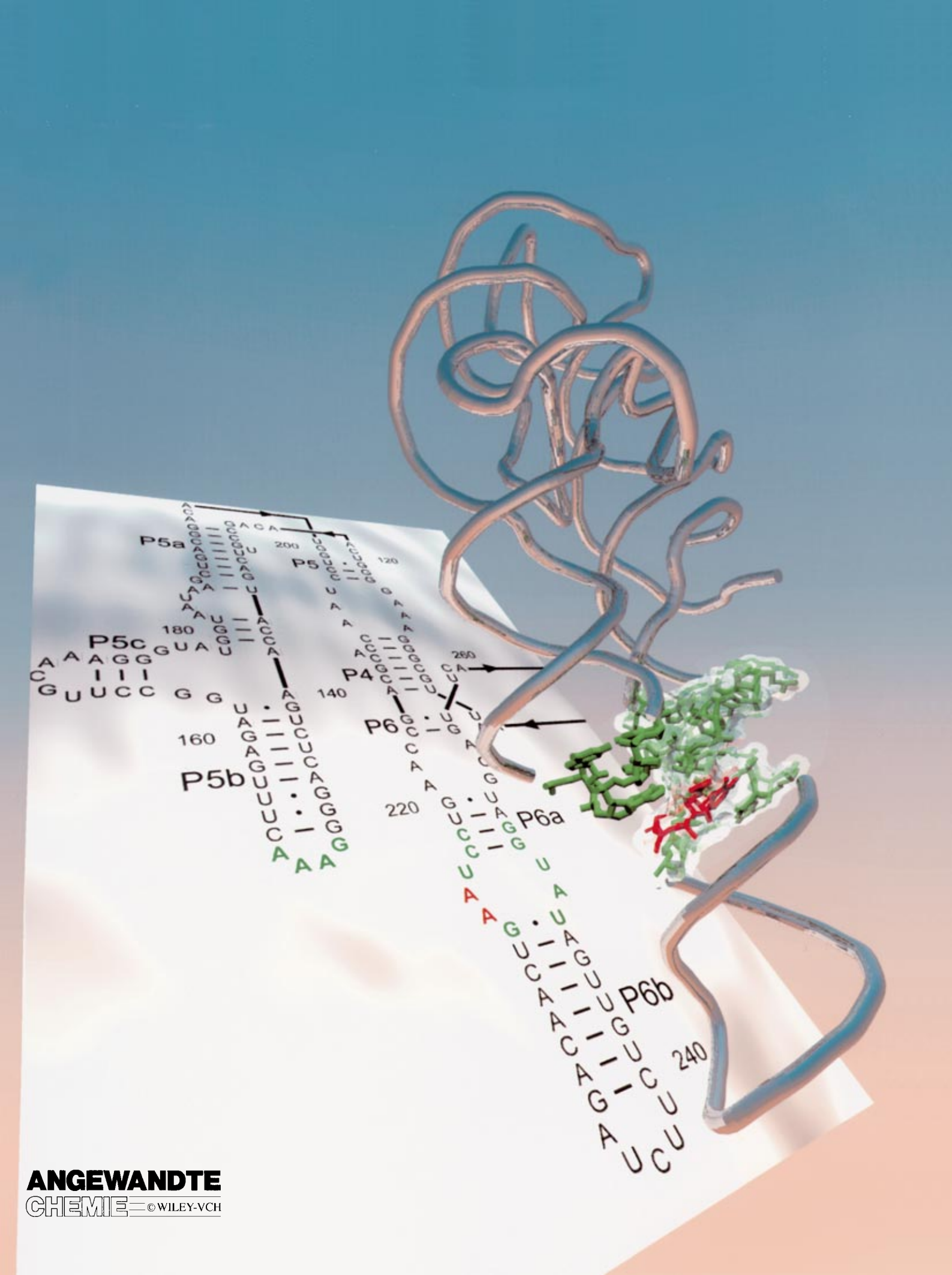



ANGEWANDTE
CHEMIE

# Tertiary Motifs in RNA Structure and Folding

**Robert T. Batey, Robert P. Rambo, and Jennifer A. Doudna***

RNA plays a critical role in mediating every step of the cellular information transfer pathway from DNA-encoded genes to functional proteins. Its diversity of biological functions stems from the ability of RNA to act as a carrier of genetic information and to adopt complex three-dimensional folds that create sites for chemical catalysis. Atomic-resolution structures of several large RNA molecules, determined by X-ray crystallography, have elucidated some of the means by which a global fold is achieved. Within these RNAs are tertiary structural motifs that enable the highly anionic double-stranded helices to tightly pack together to create a globular architecture. In this article we present an overview of the structures of these motifs and their contribution to the organization of large, biologically active RNAs. Base stacking, participation of the ribose 2′-hydroxyl groups in hydrogen-bonding interactions, binding of divalent metal cations, noncanonical base pairing, and backbone topology all serve to stabilize the global structure of RNA and play critical roles in guiding the folding process. Studies of the RNA-folding problem, which is conceptually analogous to the protein-folding problem, have demonstrated that folding primarily proceeds through a hierarchical pathway in which domains assemble sequentially. Formation of the proper tertiary interactions between these domains leads to discrete intermediates along this pathway. Advances in the understanding of RNA structure have facilitated improvements in the techniques that are utilized in modeling the global architecture of biologically interesting RNAs that have been resistant to atomic-resolution structural analysis.

**Keywords:** molecular modeling · nucleic acids · ribozymes · RNA · structure elucidation

## 1. Introduction

The discovery that RNA is capable of catalyzing chemical reactions was a watershed event that led to the realization that in many ways RNA is more akin to proteins than to its chemical cousin DNA.[1, 2] Like proteins, RNAs adopt complex three-dimensional folds for the precise presentation of chemical moieties that is essential for its function as a biological catalyst, translator of genetic information, and structural scaffold. To understand RNA in this light we must begin to address the same fundamental questions that are central to the study of protein structure and enzymology. First, how are three-dimensional structure and function related? This generally entails detailed mechanistic studies of the chemical reaction catalyzed by the enzyme coupled with structural characterization of the ground state, transition state, and functional groups involved in catalysis. Second, how is the macromolecule capable of rapidly folding into the complex three-dimensional shape necessary for catalytic activity in spite of the nearly infinite number of potential available conformations—the famous "Levinthal paradox"?[3] As will be detailed in this review recent insights into the "tertiary" level of RNA structure have revealed some of the means by which RNA is capable of achieving a complex global fold and how these interactions serve to direct the folding pathway.

[*] Dr. J. A. Doudna, Dr. R. T. Batey, R. P. Rambo
Department of Molecular Biophysics and Biochemistry
and
Howard Hughes Medical Institute
Yale University
266 Whitney Ave., New Haven, CT 06520-8114 (USA)
Fax: (+1) 203-432-3104
E-mail: doudna@csb.yale.edu

## 2. RNA Structure

### 2.1. Elements of RNA Structure

RNA structure is divided into three fundamental levels of organization: primary, secondary, and tertiary structure. Primary structure refers to the nucleotide sequence of an RNA, which can be obtained to a first approximation from the DNA sequence of the gene encoding the RNA. Since many biologically active RNAs are post-transcriptionally modified the DNA sequence often does not reveal the true primary structure. These modifications include the methylation of nucleotide bases and 2′-hydroxyl groups of ribose sugars,

  1433-7851/99/3816-2327 $ 17.50+.50/0 

formation of unusual bases such as pseudouracil (Ψ) and dihydrouridine (D), insertion or deletion of nucleotides in messenger RNA, and splicing of internal sequences (introns) from pre-messenger RNA. Thus, to completely determine the primary sequence the RNA must be purified from its native source and characterized by using a combination of sequencing methods[4] and mass spectrometry.[5]

The secondary structure of RNA is presented as a two-dimensional representation of its Watson–Crick base pairs (Figure 1a) and intervening "unpaired" regions. Commonly described secondary structural elements are: duplexes, single-stranded regions, hairpins, bulges, internal loops, and junctions, as illustrated in Figure 1b (a comprehensive review has been written by Chastain and Tinoco[6]). For example, the secondary structure of transfer RNA (tRNA) is organized into three hairpin structures, the D-, T-, and anticodon arms, and the acceptor stem, which is represented as the classic cloverleaf structure shown in Figure 2a.

This level of RNA structure is determined from several methods. The simplest technique, but most fraught with uncertainty, is to use secondary structure prediction programs that fold the input primary sequence into potential secondary structures. These programs are all based upon finding the secondary structure with the total lowest free energy by calculating the free energy of a number of base-pairing schemes, and returning the lowest energy potential secondary structures as the most probable.[7] A more reliable means for determining the secondary structure of biological RNAs is comparative sequence analysis,[8] which exploits the tendency for the global architecture of biological RNAs to be conserved. The phylogenetic covariance of two or more nucleotides that are distant in the primary sequence implies that they interact at some level. In helical regions this covariation manifests itself as the exchange of one type of Watson–Crick base pair for another (for example, a CG to an AU pair). For many RNAs there is a sufficient database of sequences from a broad evolutionary spectrum to allow for very accurate secondary structures to be derived from newly sequenced RNA genes from almost any organism.

A secondary structure generated from the above methods must be experimentally verified by biochemical techniques that probe the solution structure of the RNA. The most common technique involves probing the RNA with ribonu-

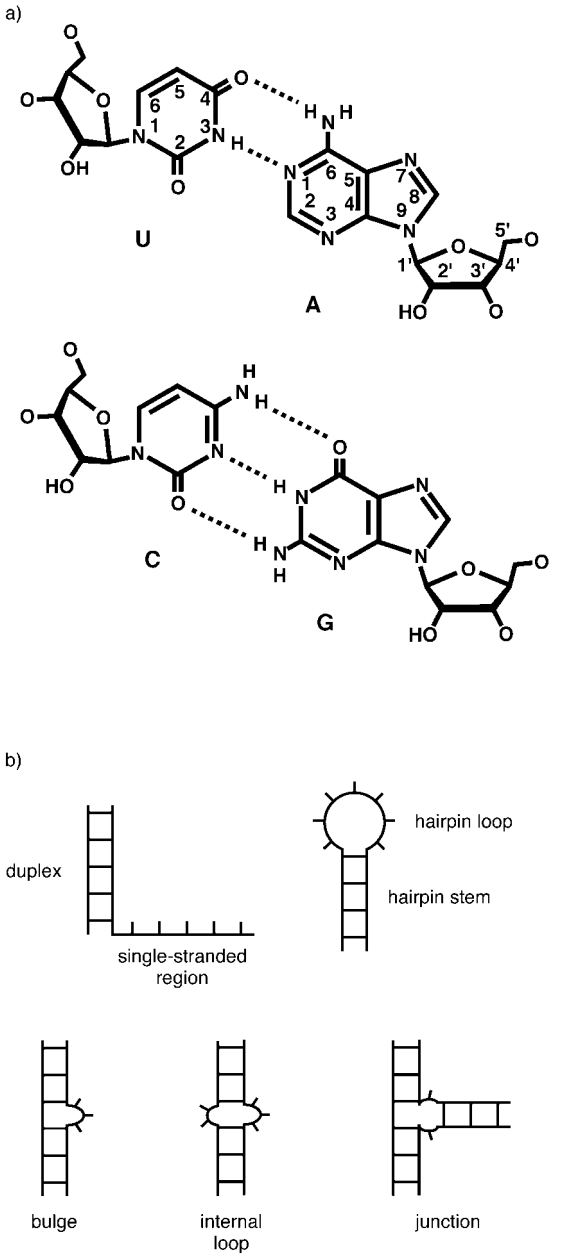



Figure 1. a) The standard Watson–Crick base pairs found in RNA, along with the numbering scheme for the nucleotide bases and ribose sugar. b) Common RNA secondary structural elements.

*Jennifer A. Doudna was born in 1964 in Washington DC (USA). She received her B.A. in Chemistry with honors from Pomona College in Claremont, CA, in 1985, and her Ph.D. in Biochemistry from Harvard University in 1989. At Harvard she worked with Prof. Jack Szostak on the design of self-replicating RNA, which led to her interest in RNA structure and folding. She was awarded a Lucille P. Markey Scholar fellowship in 1991 to study with Prof. Tom Cech at the University fo Colorado, Boulder. During the next three years she developed methods for the crystallization of RNA in collaboration with Prof. Craig Kundrot. She joined the Yale faculty in the Departement of Molecular Biophysics and Biochemistry in 1994, and she was promoted to Associate Professor in 1997 and to Professor in 1999. She received the Johnson Foundation Prize for Innovative Research in 1996 for her work on the crystal structure of the P4–P6 domain of the Tetrahymena self-splicing intron. She is currently an Assistant Investigator at the Howard Hughes Medical Institute.*

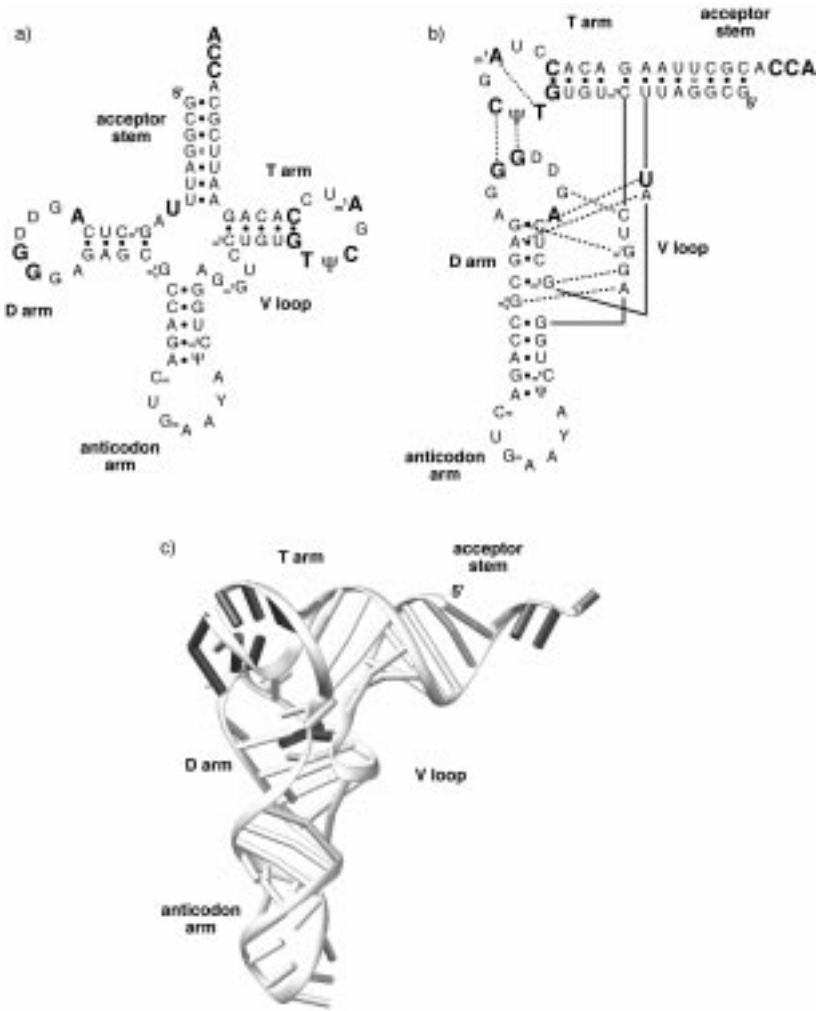



Figure 2. The levels of structure in yeast $tRNA^{Phe}$. a) Nucleotide sequence, modifications, and secondary structure of $tRNA^{Phe}$ as represented in the classic cloverleaf structure. Watson–Crick base pairs are indicated as solid dots and the GU wobble pair with an open circle. Nucleotide positions that are invariant among class I tRNAs are highlighted in boldface. (Adapted from reference [135].) b) The tertiary interactions of $tRNA^{Phe}$ as represented in a modified secondary structure representation that reflects its three-dimensional structure (the dashed lines denote tertiary interactions between bases). The invariant nucleotides are primarily involved in tertiary interactions, clustered in the region of the tRNA responsible for formation of the global architecture. c) Representation of the X-ray crystallographic structure of $tRNA^{Phe}$ with invariant nucleotides represented as black bars (Protein Data Bank (PDB) accession number 6tna).[136] Figures of RNA structures in this review were created with RIBBONS 2.0.[137]

cleases (RNases) and chemicals that target specific features of the RNA.[9] Most of these reagents react at sites in the RNA that are not involved in Watson–Crick base pairing and are solvent accessible; sites of modification are subsequently revealed by sequencing $^{32}$P-end-labeled RNA or by reverse transcription of the modified RNA with a $^{32}$P-end-labeled DNA primer. Comparing the results of biochemical probing with predicted RNA secondary structures generally yields an accurate map of the Watson–Crick pairing scheme.

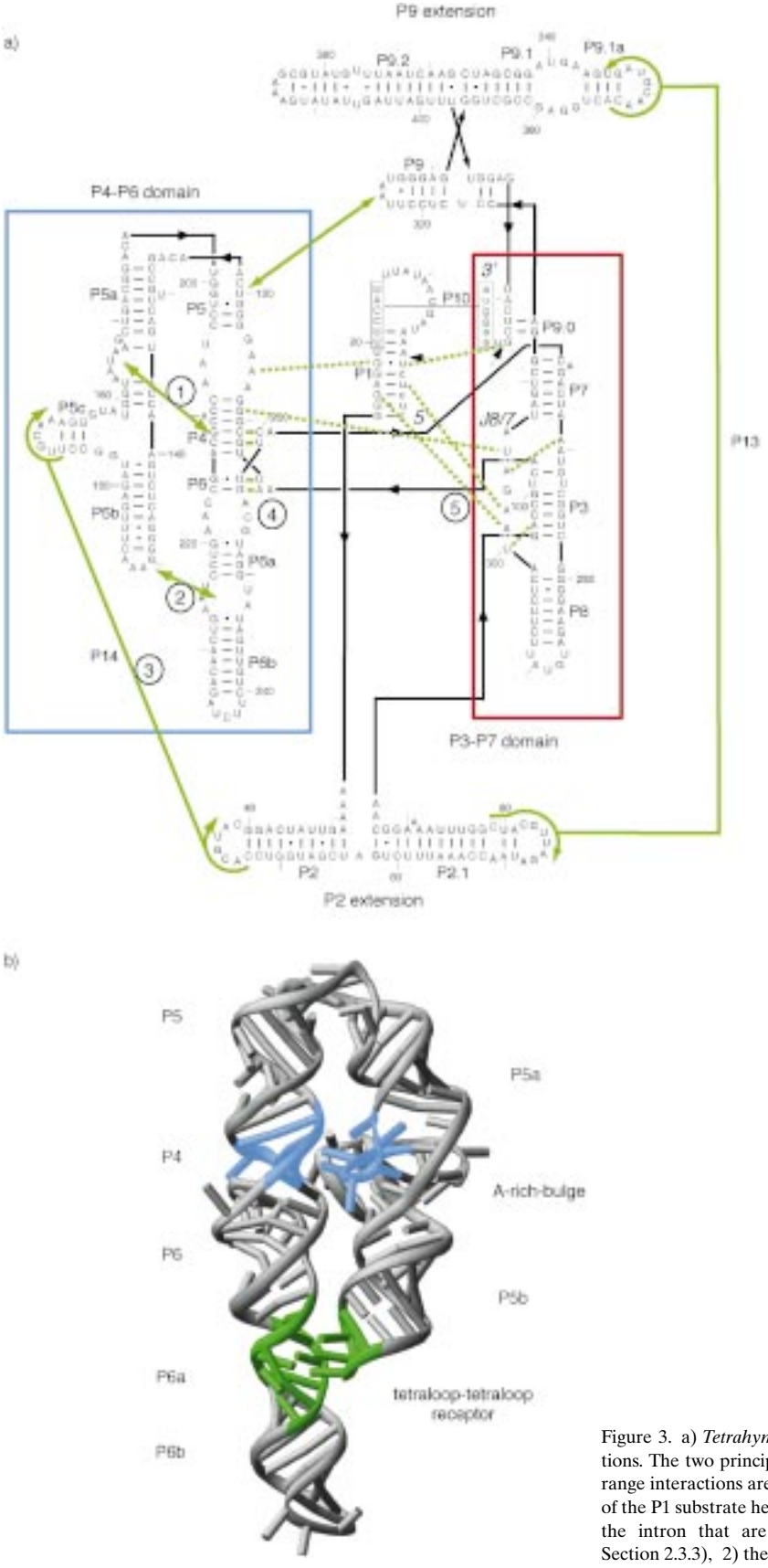
Tertiary structural elements primarily involve an interaction between distinct secondary structural elements. Beyond tRNA, not much was known about this level of RNA architecture until very recently because these types of interactions are extremely difficult to predict or experimentally determine. Progress has been made possible by improvements in the synthesis of milligram quantities of large RNAs[10] and the techniques for structural analysis such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (for excellent reviews see references [11, 12] respectively). Within the high resolution structures of large RNAs solved to date (tRNA,[13, 14] the hammerhead ribozyme,[15, 16] the P4–P6 domain of the *Tetrahymena thermophila* self-splicing intron,[17] and the hepatitis delta virus ribozyme[18]) it is tertiary interactions that play a dominant role in establishing the global fold of the molecule. The tertiary interactions observed in phenylalanine–tRNA ($tRNA^{Phe}$), along with a representation of the secondary structure that reflects the three-dimensional structure, are shown in Figure 2b. This view of tRNA illustrates a fundamental feature of biological RNAs. The invariant nucleotides in class I tRNAs (shown in boldface in Figure 2a, b and as black bars in Figure 2c) fall into two distinct types: biologically functional nucleotides, such as those found at the 3′-terminus of the aminoacyl acceptor arm, and those that are important for the establishment of the global architecture of the RNA, such as the conserved nucleotides in the D- and T-loops. The conservation of nucleotides involved in the formation of tertiary structure indicates that all tRNAs of this class have the same basic fold.

In the following section the structures of the elements of tertiary structure that have been elucidated by X-ray crystallography and NMR spectroscopy will be described. For the purposes of this review these motifs will be classified into three general categories: interactions between two double-stranded helical regions, between a helical region and a non-double-stranded region, and between two non-helical regions (this classification scheme has been adopted from Westhof and Michel[19]). In this review we will focus most of our attention on two RNA molecules whose structure and folding has been characterized in great detail: transfer RNA (Figure 2a–c) and the *Tetrahymena thermophila* self-splicing group I intron (Th-intron; Figure 3a, b). We refer the interested reader to Wedekind and McKay for a comprehensive review of the structure and function of the hammerhead ribozyme.[20]

## 2.2. Interactions Between Helical Motifs

### *2.2.1. Coaxial Stacking*

Coaxial stacking of helical regions, the most fundamental method by which RNA achieves higher order organization, is a consequence of the highly favorable energetic contributions




of stacking interactions between the π-electron system of the nucleotide bases to the overall stability of nucleic structure.[21] The dominance of base stacking in stabilizing RNA structure is convincingly demonstrated by tRNA, in which only 41 of 76 bases are involved in the classic helical structure, yet 72 bases are involved in stacking interactions.[22] All of the other known large RNA structures display a similar high degree of base stacking.

The contribution of coaxial stacking to the global fold of an RNA was first observed in the crystal structure of $tRNA^{Phe}$.[13, 14, 23] In the three-dimensional structure the stems of the D- and anticodon arms stack upon one another as do the stems of the T-arm and aminoacyl acceptor arm (Figure 2b, c).[23] These two coaxial stacks are oriented perpendicularly with respect to one another by tertiary interactions between the D- and T-loops to yield the overall L-shape of the molecule. The predominance of coaxial stacking in the organization of RNA structure is also evident in the structures of the P4–P6 domain (Figure 3b) and the hepatitis delta ribozyme (see Figure 8); each of these structures can be described as two sets of coaxially stacked helices packed against one another.

The organization of junctions, in which three or more helices intersect, by coaxial stacking is often achieved through the binding of divalent metals near the site of the stack. The direct influence of metal-ion binding on the folding of this secondary structural motif is clearly demonstrated in studies of the three-way junction at the catalytic center of the hammerhead ribozyme. In the crystal structure two of the helices are seen to coaxially stack, and the third is oriented relative to the coaxial stack by both tertiary contacts and hydrated magnesium ions specifically bound to the RNA (Figure 4a, b).[15, 16, 24] The orientation of the helices with respect to one another is extremely sensitive to the concentration of divalent cations, as demonstrated by native gel electrophoresis,[25, 26] fluorescence resonance energy transfer,[27] and transient electric birefringence[28] studies. In the absence of $Mg^{2+}$ this junction forms an extended structure in which none of the helices are stacked (Figure 4c).[25] At

Figure 3. a) *Tetrahymena* group I intron secondary structure and tertiary interactions. The two principal structural domains P4–P6 and P3–P7 are boxed, the long-range interactions are shown as green arrows, and interactions involving the binding of the P1 substrate helix are shown as green dashed lines. The tertiary interactions in the intron that are mentioned in this review are: 1) the A-rich bulge (see Section 2.3.3), 2) the tetraloop–tetraloop receptor (see Section 2.3.2), 3) loop–loop interactions (see Section 2.4.1), 4) the triple helical scaffold, and 5) the J8/7 triplex (see Section 4.2). In the nomenclature of the group I introns, P refers to a paired region, J refers to a "joining", or single-stranded, region, and L refers to a hairpin loop. This figure was created using data from the group I intron comparative database[138] and Lehnert et al.[62] b) The three-dimensional structure of the P4–P6 domain from this intron (PDB accession number 1gid).[17] The tertiary interactions that hold this domain together, the A-rich bulge and the tetraloop–tetraloop receptor, are highlighted in blue and green, respectively.

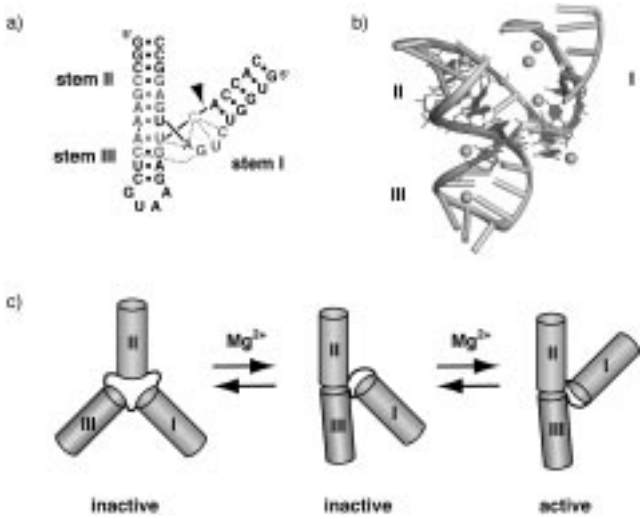



Figure 4. The hammerhead ribozyme. a) The secondary structure of the hammerhead ribozyme that was utilized for structure determination. Nucleotides critical for catalytic activity are highlighted in gray, the base at the cleavage site in light gray, and the observed tertiary interactions denoted as dashed lines. b) The three-dimensional structure of the hammerhead ribozyme with associated cobalt ions (shown as spheres; PDB accession number 379D)[24]. c) Schematic representation of the magnesium-dependent conformation of the hammerhead ribozyme. (This figure was adapted from reference [26].)

low concentrations of divalent cations (500 μM $Mg^{2+}$) helices II and III coaxially stack and helix I forms an acute angle to helix III. At higher magnesium concentrations (5 mM), helix I forms an acute angle with respect to helix II; this form corresponds to the catalytically active form of this ribozyme. Although extremely high concentrations of monovalent ions (4 M $LiSO_4$) can drive the hammerhead ribozyme into an active structure,[29] under physiological conditions the conformational rearrangement of the junction is dependent upon site-specific binding of divalent ions. Studies of other junction elements indicate that the binding of divalent cations generally mediates their folding.[30–33] This mechanism is also observed in other tertiary motifs requiring coaxial stacking (see Sections 2.4.1 and 2.4.2).

### *2.2.2. The Adenosine Platform*

The crystal structure of the P4–P6 domain of the Th-intron reveals a motif that facilitates the interaction between helices by creating a site for a base-stacking interaction.[17, 34] The adenosine platform (A-platform), which occurs in three separate locations in this RNA, consists of two sequential adenosine groups arranged side-by-side to create a "pseudo base pair" (Figure 5a). The 3′-adenosine position of the platform can also accommodate a cytidine base, as observed in the solution structure of the theophylline-binding RNA aptamer by NMR spectroscopy,[35] and in an in vitro selection for variants of the tetraloop–tetraloop receptor interaction (see Section 2.3.2).[36] Each adenosine platform has a noncanonical base pair (a pairing interaction that is not Watson–Crick) immediately below it, either a GU wobble pair or a Hoogsteen-type A·U pair, which creates a local conformation in the helix that enhances the stacking between the

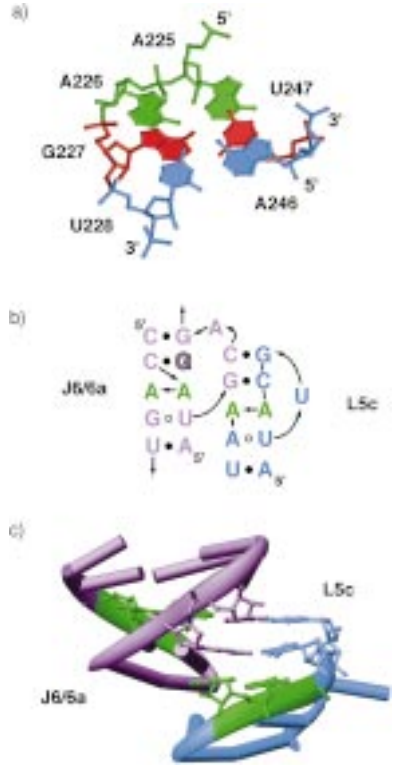



Figure 5. a) The structure of the adenosine platform motif. Two adjacent adenosines (green) are placed in a side-by-side arrangement immediately below a GU wobble base pair (red). b) Secondary and c) tertiary structure of an interaction between the J6/6a internal loop (purple) and L5c (blue) of two P4–P6 molecules in the asymmetric unit mediated by two adenosine platforms (green).

platform and the noncanonical pair. The other face of the 3′-adenosine of the platform remains free to participate in tertiary stacking interactions, as occurs in all three platforms.

Two of the adenosine platforms in the P4–P6 domain are involved in an intermolecular RNA–RNA contact between two adjacent molecules in the crystal lattice (Figure 5b, c). The A-platform within the internal loop J6/6a (the internal loop between helices P6 and P6a) is formed by two adjacent adenosines on the 5′-side of the loop, with the Watson–Crick CG base pair on the 5′-side of the loop stacking upon the platform and a wobble GU pair on the 3′-side. This allows the platform to stack within a helix connecting P6 and P6a and causes the three nucleotides on the 3′-side of the internal loop to be extruded from the helix. Two of the extruded nucleotides form Watson–Crick pairs with nucleotides in L5c of an adjacent RNA, which creates a small two base pair helix that stacks upon the A-platform capping L5c. The other platform in the P4–P6 domain is involved in an intramolecular tertiary interaction involving a four nucleotide loop and an internal loop (see Section 2.3.2). Since all three A-platforms observed in P4–P6 promote tertiary interactions it is likely that this motif will be found in many other large RNAs.

### *2.2.3. 2′-Hydroxy-Mediated Helical Interactions*

Small duplex RNAs face an extreme problem of how to pack into a three-dimensional lattice in crystals. While natural

RNAs have a variety of specialized interhelical packing motifs, the strictly helical nature of duplexes does not present much chemical diversity with which to create these tertiary interactions. The interhelical contacts created in these lattices, though, may provide insights in how helices pack in some natural RNAs. In most of the crystals observed thus far, the duplex oligonucleotides coaxially stack to create "pseudo-infinite" helices that pack into a three-dimensional lattice in two distinct ways.

In the case of the RNA dodecamer $^{5'}$GGCGCUUGCGUC$^{3'}$ the quasi-continuous helices pack such that the backbone of one helix contacts the shallow minor groove of a perpendicular helix.[37] At each site of contact, extensive direct hydrogen bonding occurs between 2′-hydroxyl groups, the 3′-oxygen atom, and phosphate oxygen atoms of the backbone of one helix and the pyrimidine O2 atom, the exocyclic amine of guanosine, and 2′-hydroxyl groups in the minor groove of the adjacent helix (Figure 6). The majority of the hydrogen bonds are mediated by the 2′-hydroxyl groups: 8 out of 13 in the first contact site and 5 out of 8 in the second contact site between helices. Similar sets of contacts are observed in crystals with the same mode of helical packing.[38, 39]

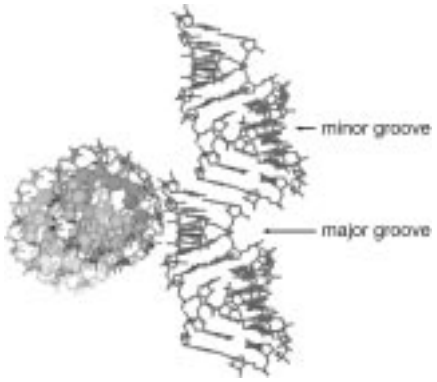



Figure 6. The perpendicular packing arrangement of pseudo-infinite RNA helices. The backbone of one helix interacts with nucleotide bases in the wide, shallow minor groove of an adjacent helix in the crystal lattice. This figure was adapted from the crystal structure of the oligonucleotide $^{5'}$GGCGCUUGCGUC (Lietzke et al.;[37] PDB accession number 280D).

This set of intimate contacts between helices is in stark contrast to crystal structures in which the pseudo-infinite helices are oriented parallel with respect to one another. In these crystals the backbones of the RNA duplexes make a limited number of interhelical contacts, primarily through water-mediated backbone – backbone contacts and 2′-hydroxyl groups of the phosphate contacts.[40, 41] The dominance of 2′-hydroxyl groups in mediating helical packing in both arrangements is reflected in many of the tertiary interactions found in natural RNAs, such as the ribose zipper motif (see Section 2.3.4).

## 2.3. Interactions Between Helical and Unpaired Motifs

### *2.3.1. Base Triples and Triplexes*

Early biophysical studies of heteroduplexes of poly(A) (homopolymeric adenosine) and poly(U) (homopolymeric uridine) demonstrated that the duplexes converted into poly(U)-poly(A)-poly(U) triple helices and single-stranded poly(A) at high ionic strength.[42, 43] Strands of poly(A) and poly(U) form standard Watson – Crick base pairs in the triplex, while the other poly(U) strand is placed in the major groove of the RNA duplex to form Hoogsteen-type pairs with the poly(A) strand (Figure 7a). Similarly, a protonated cytosine can hydrogen bond with the Hoogsteen face of a guanosine involved in a Watson – Crick GC pair to create an isosteric $(C \cdot G\text{-}C)^+$ triple (Figure 7b). Structural characterization of a model RNA triplex containing both types of triple base pairs by NMR spectroscopy revealed that the third strand can be placed into the deep major groove of RNA without significant conformational distortion from ideal A-form helical geometry.[44]

Despite the stability of the extended pyrimidine-purine-pyrimidine triplex, this motif has not been observed in any naturally occurring RNA. Instead, many RNAs employ this motif in a limited fashion as isolated major and minor groove base triples. Structural examples exist for both types of pairing. Major-groove triples are observed within the crystal structure of $tRNA^{Phe}$, where the D-arm forms two consecutive triple base pairs.[13, 14, 23] Unlike the model triplexes, in which only pyrimidines are used in the Hoogsteen pairing strand, both triple base pairs involve a purine interacting with the Hoogsteen face of a purine involved in a Watson – Crick pair (Figure 7c). Minor groove triples are also observed in $tRNA^{Phe}$, where A21, which is coplanar to a $U8 \cdot A14$ reverse-Hoogsteen base pair, interacts through the minor groove by forming a hydrogen bond between its N1 and the 2′-hydroxyl group of U8 (Figure 7d). Minor groove triple base pairs are found within the Th-intron in the GAAA tetraloop – tetraloop interaction[17] (see Section 2.3.2) and the triple-helical scaffold, a structure critical for proper alignment of the P4–P6 and P3–P7 domains. Triple base pairs also play a critical role in the interactions of many small molecule ligands and peptides with RNA.[45]

The diverse morphology of major- and minor-groove triples found in biologically active RNAs suggests that the nature of triplexes in these RNAs is quite different from that of the model pyrimidine-purine-pyrimidine triple helices. Instead, extended single stranded regions appear to tie helical regions together through the formation of base – triple interactions. The variable loop of $tRNA^{Phe}$ (Figure 2b, c) forms two base triples with the major groove of the D-arm, as well as pairing interactions with the D-loop and the coaxial stack between the D- and anticodon arms. In the hepatitis delta ribozyme (Figure 8) a stretch of four single-stranded nucleotides (J4/2) form extensive interactions with the two sets of coaxially stacked helices in the region that is the proposed catalytic pocket.[18] Two invariant adenosines in this single-stranded region interact with a helix through a minor groove triple with a Watson – Crick GC pair and a ribose zipper motif (see Section 2.3.4). Cytosine 75 (shown in red in Figure 8), which is critical for the catalytic activity of this ribozyme, forms a hydrogen bond between its N4 atom and a phosphate group within the other set of coaxially stacked helices. There is extensive biochemical evidence that a minor-groove triple helix formed by J8/7 within the catalytic core of the Th-intron

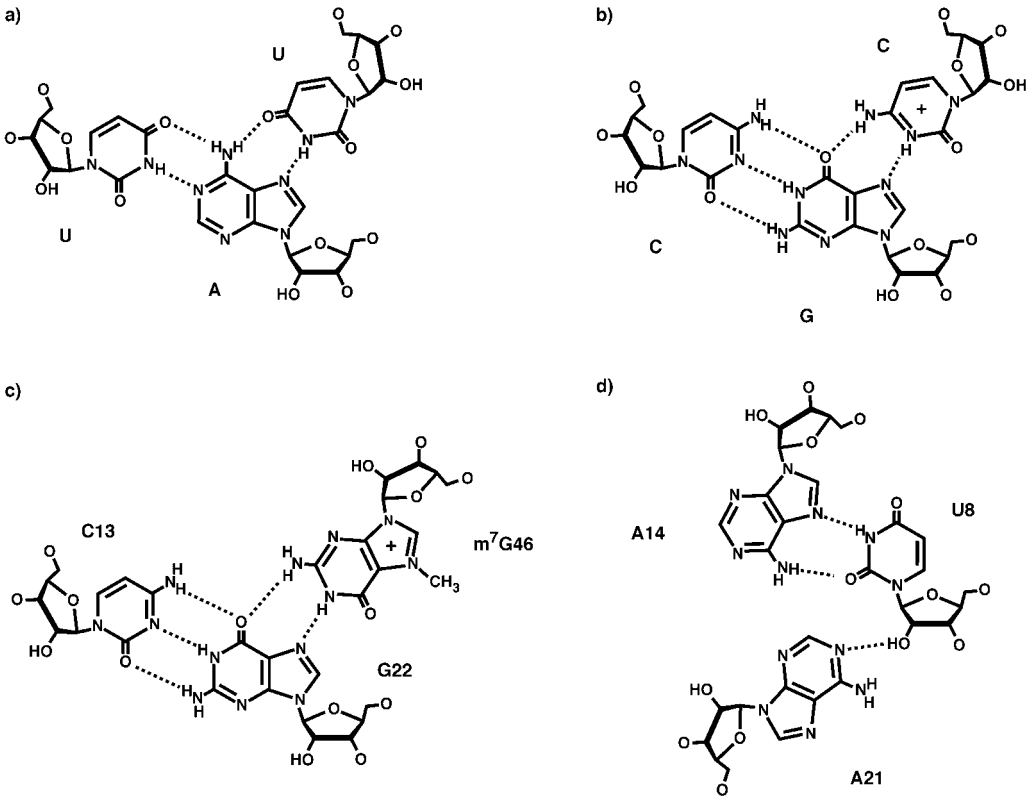



Figure 7. Examples of triple base pairs found in RNA. a) Hoogsteen-type U · A-U triple base pair, with the second uracil hydrogen atom bonding to the major-groove face of a standard Watson – Crick AU pair. b) A protonated Hoogsteen-type $C^+$ · G-C pair. c) A purine triple base pair found in the major groove of $tRNA^{Phe}$. d) A triple base pair found in $tRNA^{Phe}$ in which an adenosine hydrogen atom bonds with a 2′-hydroxyl group on the minor groove face of a reverse-Hoogsteen A · U base pair.

is critical for bridging the P1 substrate helix and the P3 and P4 helices that form the heart of the catalytic core (see Figure 17).[46] Thus, it appears from these examples that single-stranded regions often serve to bind helical regions together through the formation of triple interactions between bases in many large RNAs.

### *2.3.2. The Tetraloop Motif*

Among the most prevalent motifs observed in natural RNAs are several four nucleotide loop sequences (tetraloops) found in the 16S- and 23S-ribosomal RNAs (rRNAs), groups I and II self-splicing introns, ribonuclease P, and bacteriophage T4 messenger RNA.[47] For instance, 16S-rRNA tetraloops account for about 55% of all hairpin loops in eubacterial, while five nucleotide loops, the next most prevalent loop size, account for 13% of the loops.[48] The majority of 16S-rRNA tetraloops fall into two sequence categories: the "UNCG" motif and the "GNRA" motif. The UNCG-tetraloop motif (N indicates that the second nucleotide position in this sequence motif may be any nucleotide) has been shown to confer unusually high thermodynamic stability to an RNA hairpin,[47, 49] but has not yet been implicated in the formation of tertiary interactions.

The GNRA sequence motif, in which any nucleotide is accommodated in the second position and the third position is a purine base (R = A or G), is the most prevalent tetraloop motif in naturally occurring RNAs.[48] Structural analyses of several of these tetraloops (GAGA, GCAA, and GAAA) by NMR spectroscopy reveal a network of hydrogen-bonding and base-stacking interactions that create a stable structure (Figure 9a).[50, 51] In all of the tetraloops the position 1 guanosine ($G^1$) and the adenosine in position 4 ($A^4$) form a sheared base pair in which $G^1$–N3 hydrogen bonds with $A^4$–N6, and $G^1$–N2 forms a hydrogen bond with $A^4$–N7. In the GCAA and GAAA tetraloops the $G^1$–N3 to $A^4$–N6 distance is too long to be a direct hydrogen bond, and is thus likely to be water mediated.[51] In the GAAA tetraloop this pair is surrounded by hydrogen bonds between the phosphate of $A^4$ and the Watson – Crick face of $G^1$, and the 2′-hydroxyl group of $G^1$ interacting with the Hoogsteen face of G/$A^3$ and $A^4$–N6. In the GAGA and GAAA tetraloops the nucleotide in position 2 is stacked upon the nucleotide in position 3, whereas in the GCAA loop the cytosine in position 2 is disordered. This variability in the hydrogen-bonding and stacking interactions between different members of the GNRA family does not significantly alter their thermodynamic stability, but may be critically important for the ability of this motif to form tertiary interactions with a diverse array of RNA motifs.

Covariation analysis of the groups I and II introns and RNase P have revealed several RNA motifs that form tertiary interactions with the GNRA-type tetraloop. The structure of an internal loop motif, termed a tetraloop receptor, which interacts with the GAAA tetraloop with high affinity and specificity[52–54] was elucidated as part of the Th-intron P4–P6

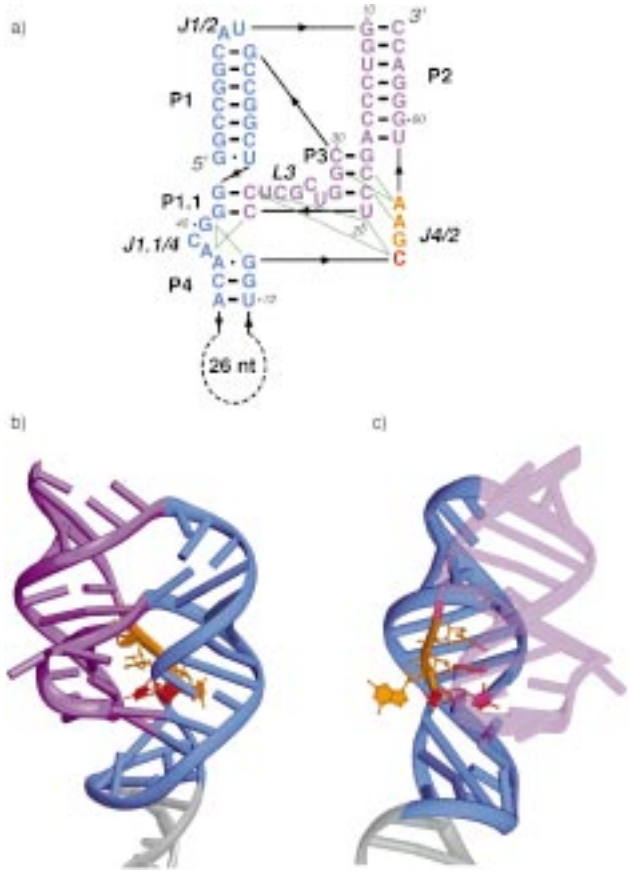



Figure 8. The hepatitis delta virus ribozyme. a) A representation of the secondary structure of the RNA construct used for an X-ray structure determination which reflects its three dimensional fold. This fold divides the ribozyme into two coaxially stacked helices, (blue and magenta), along with the single-stranded region J4/2 (orange) that contains a cytosine (red) essential for catalytic activity. b) The three-dimensional structure of the hepatitis delta ribozyme showing the overall double-pseudoknotted topology of the backbone (PDB accession number 1drz).[18] The gray region at the bottom of the ribozyme was part of an engineered U1A protein binding site that was utilized to facilitate the crystallization of the RNA (the protein is not shown). c) A 90° rotation of the ribozyme with the P2/P3 coaxial stack in front. The highly conserved nucleotides C75, A77, and A78 of J4/2 are thrust deep within the catalytic cleft of the ribozyme between the two sets of coaxially stacked helices.

domain (Figure 9b, c).[17] The tetraloop receptor adopts a conformation consisting of two Watson–Crick GC base pairs, a reverse-Hoogsteen AU base pair, an adenosine platform, and a wobble GU base pair. The GAAA tetraloop bound to the receptor maintains a conformation that is almost identical with the unbound form, with the three adenosines stacking upon the adenosine platform in the receptor. This allows for a network of hydrogen bonds between the adenosines of the tetraloop and the minor groove of the receptor to stabilize this tertiary interaction.

A second mode of tetraloop–RNA interaction involves a GNRA tetraloop binding to the minor groove face of two tandem base pairs. This interaction involves a GNAA tetraloop contacting the minor groove of two tandem Watson–Crick GC base pairs or a GNGA tetraloop and interacting with a [5′CU…AG3′] dinucleotide. This type of interaction was observed in the crystal structure of the hammerhead ribozyme, in which the GAAA tetraloop of one ribozyme is observed to dock against two tandem GC pairs in an adjacent molecule.[15, 55] Biochemical studies of the interactions between GNRA tetraloops and RNA indicate that both the loop and its recognition site can tolerate a suprisingly high degree of variability without sacrificing binding affinity or specificity.[36, 56] This suggests that there are as yet to be identified motifs that form intramolecular contacts with the tetraloop motif within biological RNAs.

The importance of the minor groove in mediating tertiary interactions, such as in tetraloop docking, may be a fundamental reason why RNA, but not DNA, folds into complex three-dimensional structures. The minor groove in A-form helical RNA is wide and shallow (10–11 Å), which allows for easy access to this face of the helix, unlike B-form helical DNA, which has a much narrower minor groove (5.8 Å). This allows for functional groups, particularly the 2′-hydroxyl group, of the ribose sugar to participate in an interaction. Although the major groove of RNA can be sufficiently widened at internal loops containing purine–purine pairs or at the end of helices to allow for recognition, the importance of the major groove appears to be significantly less than that of the minor groove.

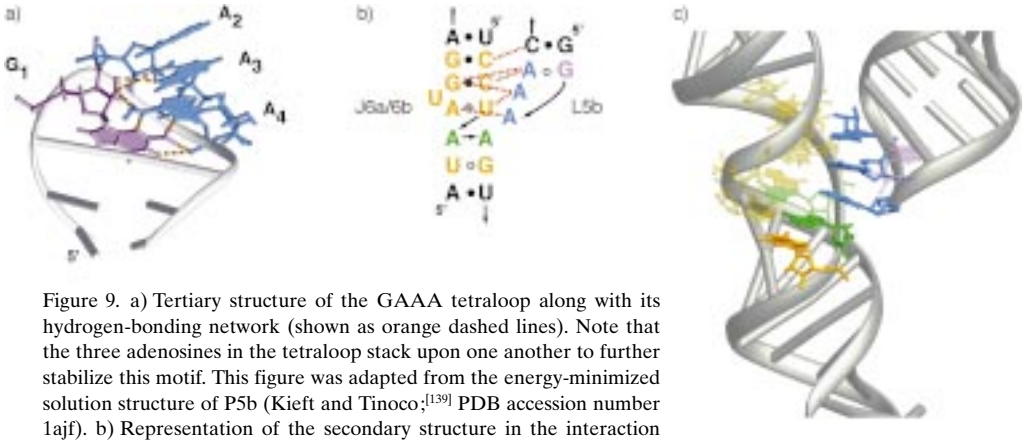



Figure 9. a) Tertiary structure of the GAAA tetraloop along with its hydrogen-bonding network (shown as orange dashed lines). Note that the three adenosines in the tetraloop stack upon one another to further stabilize this motif. This figure was adapted from the energy-minimized solution structure of P5b (Kieft and Tinoco;[139] PDB accession number 1ajf). b) Representation of the secondary structure in the interaction between the GAAA tetraloop (purple/blue) and the tetraloop receptor (gold, with the A-platform shown in green). Long range hydrogen-bonding contacts between the tetraloop and its receptor are shown in red. c) Representation of the structure of the interaction between the GAAA tetraloop and its receptor. The conformation of the receptor-bound tetraloop is nearly identical to that in the free form, with the three stacked adenosines forming a contiguous stack with the adensoine platform within the receptor.

### *2.3.3. The Metal-Core Motif*

Within the P4–P6 domain of several subclasses of group I introns there is a bulge motif in helix P5 whose nucleotide composition and distance from helix P4 are conserved. This bulge, which contains two invariant adenosine residues (A184 and A186 of the Th-intron) and another that is highly conserved (A183), is critical for the folding of the intron. The ribose–phosphate backbone of this bulge forms a corkscrew turn in which the phosphates are placed towards

the interior and the nucleotide bases are flipped outward (Figure 10).[17] Potentially unfavorable electrostatic interactions created by the close packing of the phosphate backbone are relieved by two bound magnesium ions, with the RNA providing many of the inner sphere coordinating ligands for

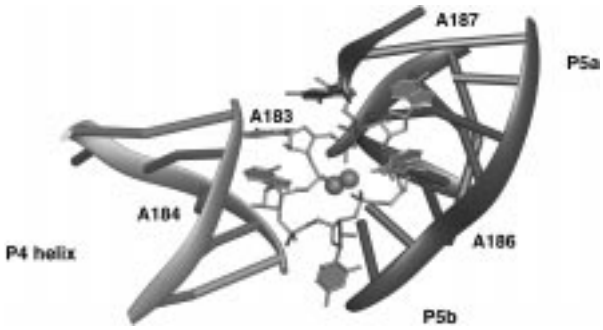



Figure 10. The adenosine-rich bulge of the *Tetrahymena* group I intron. Highly conserved adenosine residues (numbered) within the A-rich bulge interact with the P4 helix and the P5abc subdomain. The corkscrew structure of the A-rich bulge is stabilized by two specifically bound magnesium ions (shown as spheres).

each ion.[57] The nucleotide bases splayed out from the bulge form tertiary contacts to two regions of the RNA. A183 and A184 interact with functional groups within the minor groove of the P4 helix by using the ribose zipper motif (see Section 2.3.4). A186 extensively hydrogen bonds with the minor groove face of the C137–G181 base pair of P5a and G164 in the sheared A139–G164 pair found at the top of P5b. Stacking interactions between A186 and the A139–G164 pair facilitate the P5a–5b coaxial stack at this junction. The fourth adenosine of the bulge, A187, forms a noncanonical base pair with U135. Thus, the adenosine-rich bulge presents nucleotide bases in a conformation that allows helix P4 to be tightly docked against the P5a–5b helical coaxial stack, which, along with the GAAA tetraloop – tetraloop receptor, firmly locks the two sets of helices comprising the P4–P6 domain together to form the overall fold of this RNA.

The crystal structure of the trans-activation response region (TAR) of the human immunodeficiency virus (HIV) displays a similar motif.[58] In this RNA a three nucleotide bulge with the sequence UCU binds three calcium ions. These ions coordinate phosphate oxygen atoms in the backbone of the RNA in and around the bulged nucleotides. As a consequence the three pyrimidine bases face away from the RNA helix and allows the stems flanking the bulge to coaxially stack upon one another. It has been demonstrated by using transient electric birefringence measurements that $Mg^{2+}$ also binds this pyrimidine-rich bulge, which causes the two helices to coaxially stack. If the bulge sequence is changed to three adenosines then magnesium ions no longer effect this change, which indicates that the nucleotide sequence affects the ability of the backbone to specifically bind metal ions.[59, 60] This motif has also been observed in the recognition of single-stranded DNA of the canine parvovirus by the viral capsid protein through chelation of a divalent metal ion by the DNA phosphate backbone, which organizes the nucleotide bases for recognition by the protein.[61] Thus, this motif appears to be a general means by which nucleotide bases can be presented for recognition by nucleic acids and proteins.

#### *2.3.4. The Ribose Zipper*

Within the tertiary interactions mediated by the adenosine-rich bulge and the GAAA tetraloop – tetraloop receptor interaction, the ribose – phosphate backbone is brought into close contact with itself as two helices are docked against one another. At these sites of close contact the ribose sugars of antiparallel strands become interdigitated. Bifurcated hydrogen bonds between the 2′-hydroxyl group of a ribose from one helix and the 2′-hydroxyl group and the N3 atom of a purine or the O2 atom of a pyrimidine in the opposite helix create a "zipper" of ribose sugars in the minor grooves of two helices (Figure 11). Similar interactions have been observed in the

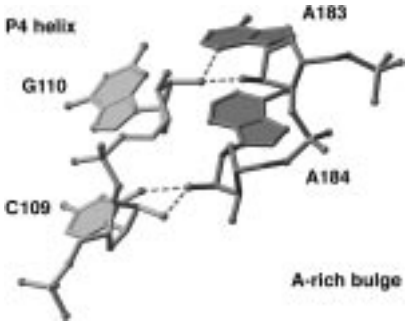



Figure 11. The hydrogen-bond network between 2′-hydroxyl groups and nucleotide bases within the ribose zipper motif found between two adenosines in the A-rich bulge and the minor-groove face of the P4 helix of the *Tetrahymena* group I intron P4–P6 domain.

intermolecular interaction of a GAAA tetraloop and the minor groove in the hammerhead crystal structure[55] and in the crystal structure of the hepatitis delta virus ribozyme.[18] This interaction might be quite pervasive within many biological RNAs because it can potentially pack RNA strands and helices together with few sequence-specific requirements. However, the lack of sequence specificity of the ribose zipper makes it extremely difficult to predict by covariation analysis or many biochemical probing techniques.

### 2.4. Tertiary Interactions Between Unpaired Regions

#### *2.4.1. Loop – Loop Interactions*

Hairpin loops present rich potential for the formation of tertiary contacts through pairing interactions between their nucleotide bases to create new helices. This type of interaction is represented in a limited form in the structure of $tRNA^{Phe}$, in which nucleotides in the D- and T-loops form two base pairs as part of the extensive tertiary contacts that form the elbow of the molecule (see Figure 2b, c). In the Th-intron helices P13 and P14 are formed through the formation of five to seven contiguous Watson – Crick base pairs between complimentary hairpin loops, which tie together distinct structural domains (see Figure 3a).[62]

The structural nature of complimentary loop–loop interactions has been investigated in detail with NMR spectroscopy, using the "kiss" complex formed between two RNAs involved in the regulation of ColE1 plasmid replication[63, 64] and the TAR loop sequence of HIV-2[65, 66] as model systems. Each complex is a single composite coaxially stacked helix comprising the two original hairpin stems and a new helix formed between them created by the Watson–Crick base pairing of the nucleotides in the complimentary loops (Figure 12a, b). Notably, all of the nucleotides in each loop are

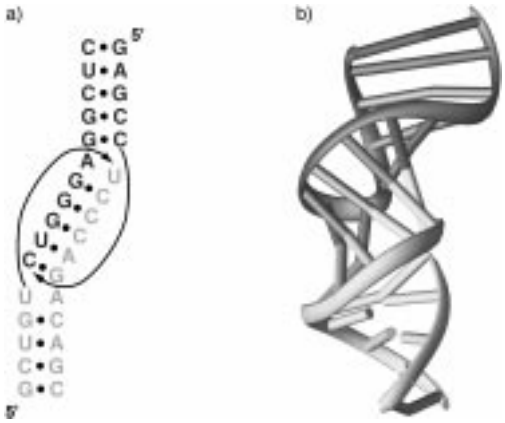



Figure 12. a) Secondary and b) tertiary structure of the kissing complex between stem loops found in the genomic RNA of the HIV-2 transactivation region (PDB accession number 1kis).[66]

stacked on the 3′-side of the central helix and are involved in pairing interactions. Stable formation of this stacked structure, like the coaxial stacks at junction motifs, requires magnesium ions.[67] The ColE1 kiss complex is further stabilized by a purine–purine cross-strand stack,[64] an RNA motif, which has also been found in 5S rRNA,[68] the sarcin–ricin loop,[69] the hammerhead ribozyme,[15, 16] and the P4–P6 domain.[17] This motif involves the six-membered ring of a purine stacking upon that of a purine in the opposite strand of the duplex, as opposed to stacking upon its same-strand neighbor, as is found in a regular A-form RNA duplex.

### *2.4.2. The Pseudoknot*

A pseudoknot is defined as a motif in which nucleotides of a hairpin loop base pair with a complementary single-stranded sequence. The classic pseudoknot, whose structure has been well characterized by NMR spectroscopy, consists of a hairpin loop paring with a complementary sequence next to the hairpin stem to form a contiguous coaxially stacked helix (Figure 13a).[70] Similar to junctions and kissing loops, the coaxial stacking observed in the pseudoknot requires either $Mg^{2+}$ ions or a high concentration of $Na^{+}$ ions for complete stabilization.[71] Additionally, many pseudoknots have bends at the coaxial stack caused by unpaired nucleotides intercalating between the two helices,[72] which is an essential feature for the in vivo function of one retroviral mRNA pseudoknot.[73]

The loops (L1 and L2) that span the two helices of the pseudoknot are inequivalent in that L1 crosses the major groove of the 3′-stem 2 (S2) and L2 crosses the minor groove of the 5′-stem (S1). In a recent structure of the pseudoknot from the tRNA-like genomic TYMV RNA, nucleotides from both loops were observed to interact with the helices (Figure 13b, c).[74] Loop 2, which crosses the minor groove, displays hydrogen-bonding interactions between some of its nucleotides and functional groups within the minor groove of stem 1. An adenosine within this loop, which is highly conserved among plant viral RNAs, is proposed to make hydrogen-bonding contacts between its N1 atom and the exocyclic amino group of a stem 2 guanosine and between the adenosine exocyclic amino group and the N3 atom of an adjacent guanosine in the helix. The adjacent cytosine in loop 2 also makes a hydrogen-bonding contact between its exocyclic amine and a 2′-hydroxyl group in the minor groove of stem 2. This suggests that pseudoknots may employ triple-helical buttressing by the loops to further stabilize the structure. A very recent high-resolution crystal structure of a pseudoknot from the beet western yellow virus mRNA demonstrates extensive tertiary interactions in both the major and the minor grooves.[75]

A remarkable example of the pseudoknot motif being utilized to direct the global architecture of an RNA is demonstrated in the structure of the hepatitis delta ribozyme.[18] This ribozyme is exceptionally stable; both the

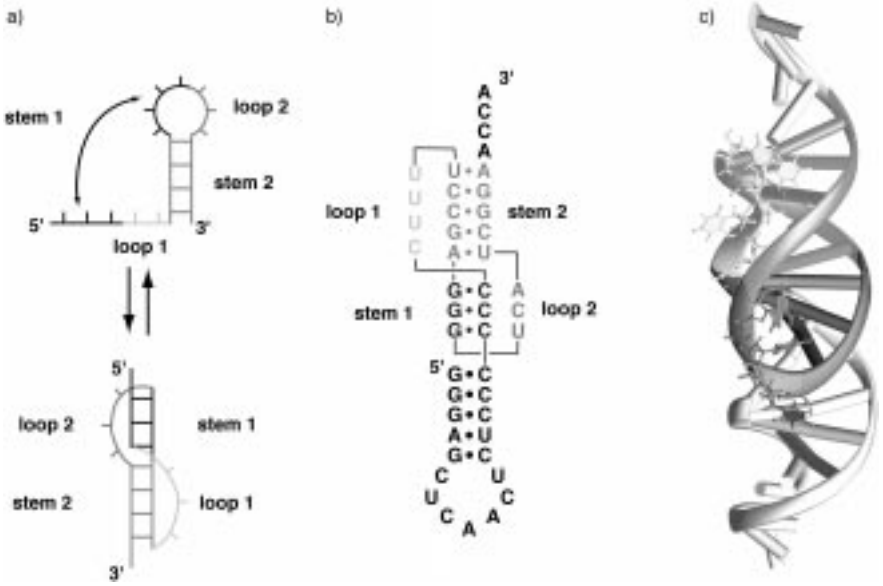



Figure 13. The pseudoknot motif. a) The classical pseudoknot is formed through a base pairing interaction between nucleotides in the loop of a stem loop and an adjacent single-stranded region. b) Secondary structure of the pseudoknot found in the genomic RNA of the turnip yellow mosaic virus (TYMV). c) The three-dimensional structure of the TYMV pseudoknot (PDB accession number 1a60).[74] Nucleotides in both loops are situated in the major and minor grooves of the adjacent helices, which allows them to form triple interactions with the base pairs.

genomic and antigenomic variants exhibit enhanced stability and activity in chemicals that typically denature nucleic acid structures (5–8 M urea or 10–18 M formamide) as well as elevated temperatures (65–70 °C).[76, 77] No divalent cations or unusual base pairs were found in the crystal structure; instead the structure is stabilized entirely through Watson–Crick base pairing and the highly convoluted double pseudoknot topology of the backbone (see Figure 8).[18] Thus, an RNA fold can be created by packing helices using either specific tertiary docking motifs, as in the P4–P6 domain, or by backbone topology, as in the hepatitis delta virus ribozyme.

## 3. The Folding of RNA into Higher Order Structures

Analogous to the "protein-folding problem", there is the question of how RNA establishes a three-dimensional fold from its primary sequence. In studies of protein folding the unfolded state is characterized as a random coil, such that the folding process encompasses the acquisition of secondary structure ($\alpha$-helices, $\beta$-sheets, etc.) and tertiary structure. With RNA, however, the unfolded state is regarded as having a majority of the native secondary structure formed, with the folding process being concerned with the sequential formation of tertiary interactions to establish the native structure. Thus, the tertiary interactions described in the previous section are principal players in the folding landscape of large RNAs, as discussed in this section.

### 3.1. The Domain Architecture of RNA

Most large biological RNAs are organized into domains that are generally reflected in their secondary structure. The term domain does not have a strict definition with respect to RNA, but generally refers to a region of secondary structure that folds as a single unit. For instance, all group I introns have two fundamental domains necessary for formation of their catalytic core: the P4–P6 domain and the P3–P7 domain (see Figure 3a). In the Th-intron these two domains are supplemented by the P2–P2.1 and P9 peripheral extensions, as well as the P5abc subdomain. Extensive evidence indicates that these domains and extensions on the active ribozyme interact with each other almost exclusively through tertiary interactions rather than base pairing.

The transition between the unfolded and folded states of the Th-intron requires the specific binding of at least three magnesium ions to assemble into a fully functional catalytic ribozyme; in the absence of magnesium or other divalent ions only the secondary structure of the ribozyme is formed. This magnesium-facilitated folding can be monitored wih $Fe^{II}$–EDTA (EDTA = ethylenediamine tetraacetate), a reagent that generates free hydroxyl radicals that are capable of cleaving the backbone of single- or double-stranded RNA at any point where it is freely accessible to bulk solvent.[78, 79] Regions of the RNA become solvent inaccessible upon higher order folding, and hence are protected from free radical mediated strand scission; regions of protection are subsequently revealed by sequencing $^{32}$P-end-labeled RNA.[78] Most of the intron is accessible to solvent below 0.75 mM $Mg^{2+}$, which indicates that no significant tertiary structure has formed. Above this concentration there are two highly cooperative transitions, the first corresponding to the formation of the P5abc subdomain and the second to the folding of the catalytic core. Subsequent studies after removing the stabilizing peripheral extension P9.1–9.2 demonstrated that the entire P4–P6 domain becomes stably folded at lower magnesium concentrations than the catalytic core requires for stable folding.[80]

The ability of the P4–P6 domain to fold at lower divalent ion concentrations than the rest of the intron indicates that it is capable of forming stable tertiary structure independently of the rest of the intron. By probing an isolated 160-nucleotide fragment corresponding to the P4–P6 domain with $Fe^{II}$–EDTA and dimethyl sulfate, a chemical that modifies the imino nitrogen atom of adenine and cytosine bases, revealed that in the presence of magnesium the solvent-inaccessible region of this RNA is a large subset of that observed in P4–P6 in the intact intron.[81] This magnesium-dependent folding of the P4–P6 domain is critically dependent upon the GAAA tetraloop of P5b; mutations in this region significantly increase the amount of magnesium that is required to produce the pattern of $Fe^{II}$–EDTA protection observed in the wild-type domain,[53] which is consistent with the tertiary interactions observed in the crystal structure of this domain.[17] Similarly, mutations in the tetraloop receptor and the J5/5a internal loop, the flexible hinge between the two sets of coaxially stacked set of helices in P4–P6, are detrimental to the formation of the solvent-accessible core.

While these mutations destroy most of the tertiary structure of the P4–P6 domain, the adenosine-rich bulge and the three-way junction J5abc remain solvent inaccessible, which suggests that the P5abc subdomain folds independently of the rest of P4–P6. The isolated P5abc subdomain displays the same pattern of $Fe^{II}$–EDTA protections as in the context of the intact intron,[81] as well as a similar pattern of magnesium ion binding as the P4–P6 domain.[57] Substitution of the *pro*-$R_p$ oxygen atoms in the phosphate backbone of an RNA with sulfur has been shown to interfere with inner-sphere coordination by magnesium ions.[82, 83] Single atom substitution of any of the four *pro*-$R_p$ oxygens in the A-rich bulge and L5c that directly coordinate magnesium ions in the crystal structure of the P4–P6 domain significantly destabilize the folding of the isolated P5abc domain.[57]

Other domains of the Th-intron are not capable of achieving an independently stable fold outside the context of the intact molecule. Structural probing of an RNA construct corresponding to the P3–P7 domain and the P9 peripheral extension demonstrated that while its secondary structure is similar to that observed in the intact group I intron, tertiary interactions within this domain are absent.[84] Only upon association with the P4–P6 domain is the P3–P7 region capable of forming a stable set of intradomain tertiary interactions, which indicates that the P4–P6 domain provides a structural scaffold for higher order folding of P3–P7.

Association of the secondary structural domains of the Th-intron are primarily mediated by tertiary interactions. Dele-

tion of P5abc from the intron (→ΔP5abc-intron) significantly impairs its activity under standard splicing conditions (5 mM $MgCl_2$, pH 7.5), but can be rescued at higher ionic strengths (15 mM $MgCl_2$, 2 mM spermine, pH 7.5).[85] Wild-type activity can be restored by supplying the P5abc subdomain to the ΔP5abc intron in a *trans* arrangement (a *trans* association involves the binding of two separate RNA molecules, in contrast to a *cis* association in which two interacting domains are on the same RNA chain). Formation of this bimolecular complex is mediated entirely through three tertiary interactions: the tetraloop–tetraloop receptor, the A-rich bulge and the minor groove of P4, and the kissing-loop interaction between L5c and L2. Formation of this intermolecular complex is magnesium dependent and extraordinarily tight; at 10 mM $MgCl_2$ the apparent dissociation constant ($K_d$) is 100 pM.[86] The Th-intron can also be divided into three separate pieces consisting of the P4–P6 domain, P1–P3 substrate, and the P3–P7/P9 domain, which are all capable of associating solely through tertiary interactions to form a functional ribozyme.[87]

It appears that a general feature of most large biological RNAs is that they are constructed from independently folding secondary structural domains that can associate in the *trans* form, primarily through tertiary-type interactions rather than base pairing, to form functionally active molecules. The catalytic component of eubacterial RNase P, a ribonucleoprotein enzyme responsible for post-transcriptional processing of small cellular RNAs, resides entirely within the RNA.[2, 88] In the absence of the protein subunit RNase P RNA (P RNA) binds the precursor tRNA substrate by specifically recognizing the coaxially stacked T-stem/acceptor stem of tRNA through tertiary interactions involving a number of 2′-hydroxyl and phosphate groups in both the P RNA[89, 90] and tRNA substrate.[91, 92] P RNA consists of two independently folding domains that show similar patterns of protection when separated from each other as they do in the context of the intact ribozyme, as evident from probing experiments with $Fe^{II}$–EDTA in the presence of magnesium ions.[93, 94] Although these domains are individually unable to catalyze tRNA processing, they associate through tertiary interactions to form a catalytically competent complex. Thus, interdomain interactions and substrate recognition occur through tertiary interactions in the same manner as the Th-intron. Similarly, some of the six phylogenetically conserved domains of group II introns are capable of efficiently binding through a *trans* association to form a functional RNA (for an excellent review see reference [95]).

### 3.2. The Hierarchical Folding Model of RNA

Pioneering studies on tRNA in the 1970s elucidated some of the fundamental features of RNA folding pathways. By using temperature-jump relaxation measurements and NMR spectroscopy it was shown that under conditions of moderate ionic strength (174 mM $Na^+$, no $Mg^{2+}$) *E. coli* $tRNA^{fMet}$ undergoes five distinct transitions during thermal unfolding.[96] The lowest temperature transitions involve the disruption of tertiary interactions between the D- and T-loops followed by the weak secondary structural elements in the D-stem.[97] Higher temperature transitions involve the melting of the secondary structural elements corresponding to the T-, anticodon, and acceptor stems (Figure 14). Since thermal unfolding is a reversible reaction in tRNA it follows that its structure forms hierarchically in the folding pathway, with almost all of the secondary structure forming prior to the tertiary structure. Similar behavior is observed in the thermal unfolding of two other tRNAs, which indicates that this is a general mechanism for tRNA folding.[98, 99]

Under conditions containing 3 mM $Mg^{2+}$ *E. coli* $tRNA^{fMet}$ displays a single, cooperative unfolding transition at high temperature such that tertiary and secondary structure are disrupted simultaneously.[100] This transition reflects the binding of a single magnesium ion to tRNA with high affinity ($K_d = 33$ mM) along with a number of weakly bound ions[101] associated with the formation of tertiary structure.[100] This uptake of magnesium ions by RNA during the folding process primarily affects the stabilization of tertiary structure rather than the formation of secondary structure. Since tertiary structure formation involves the close juxtaposition of the highly negatively charged backbone in several regions in tRNA it is not surprising that the formation of tertiary structure creates regions for high affinity multivalent cation binding sites. The structure of $tRNA^{Phe}$ reveals several specifically bound magnesium ions at sites of tertiary interaction, although no specific cation can be attributed to the stablization observed in the thermal melting profile.

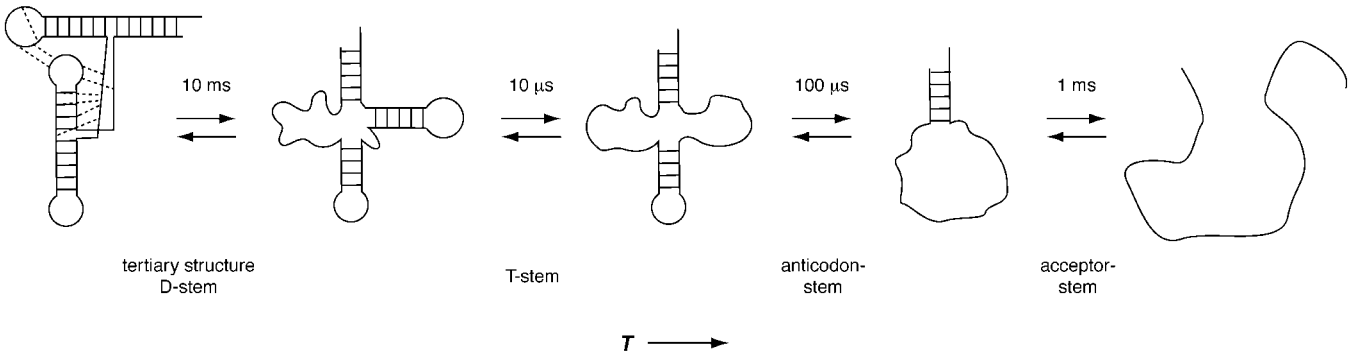



Figure 14. The unfolding pathway of transfer RNA as determined from thermal denaturation studies with approximate time constants for each step of the folding reaction. (Adapted from references [97, 140].)

Thus, these studies illustrate two fundamental features of RNA folding. First, tertiary structure formation is highly dependent upon the prior formation of secondary structure. Sequential formation of secondary and tertiary structures in RNA is the basis of the "hierarchical model" of RNA folding. Secondly, the binding of multivalent ions to RNA during the folding process generally serves to stabilize the formation of tertiary structure rather than secondary structure. This property of RNA folding is particularly evident in light of the structures of tertiary interactions, in which metal ions are often found intimately involved.

### 3.3. Kinetic Folding of the *Tetrahymena* Group I Intron

Recent work on RNA folding has primarily utilized the Th-intron as a model system. Similar to tRNA, this group I intron exhibits multiple transitions in its thermal denaturation, as measured by UV absorbance.[102] Also, chemical probing of the RNA at varying temperatures indicates that the lower temperature transition involves the disruption of tertiary structural elements rather than the opening of base pairs in the secondary structure.[102]

In order to observe folding on the millisecond timescale a variation of footprinting was developed in which a beam of high energy photons from synchrotron radiation is used to produce a burst of hydroxyl radicals in an aqueous sample that is capable of probing an RNA in a similar fashion as the $Fe^{II}$–EDTA reaction, with exposure times as short as 10 milliseconds.[103, 104] By equipping the synchrotron beamline with a stopped-flow apparatus, the magnesium-dependent folding of any RNA can be monitored starting at 20 ms after the induction of folding. This is fast enough to potentially monitor the tertiary folding of a tRNA, whose tertiary interactions form in approximately 100 ms.[97]

These experiments revealed that the P5abc domain folds most rapidly within the P4–P6 domain, where nucleotides in the magnesium core of the adenosine-rich bulge and three-way junction are protected from solvent, with a rate of 2 $s^{-1}$ (Figure 15) The remainder of the tertiary interactions in the P4–P6 domain form in a concerted fashion at slightly slower rates. Nucleotides in the peripheral extensions P2–2.1 and P9, which had been previously shown to assist in the formation of the catalytic core[105] are protected from cleavage at a rate of about 0.3 $s^{-1}$. These two peripheral extensions are proposed to wrap around the ribozyme core, and be stabilized by the formation of the P13 and P14 tertiary interactions. At this point in the folding pathway, the exterior of the ribozyme, which consists of the P4–P6 domain and the peripheral extensions, is completely folded, but the interior of the ribozyme, which contains the catalytic core, remains disordered. Formation of the active structure with a solvent inaccessible P3–P7 domain is very slow and requires minutes ($k_{obs} = 0.02\ s^{-1}$) to fold. Thus, the two structural domains P4–P6 and P3–P7 are also kinetic folding domains. This model of the kinetic folding pathway of the Th-intron, which was originally proposed by Zarrinkar and Williamson from studies performed with a time-dependent oligonucleotide hybridization assay,[106, 107] has been further supported by other time-dependant techniques, including chemical modification,[108] UV cross-linking,[109] and $Fe^{II}$–EDTA footprinting.[103]

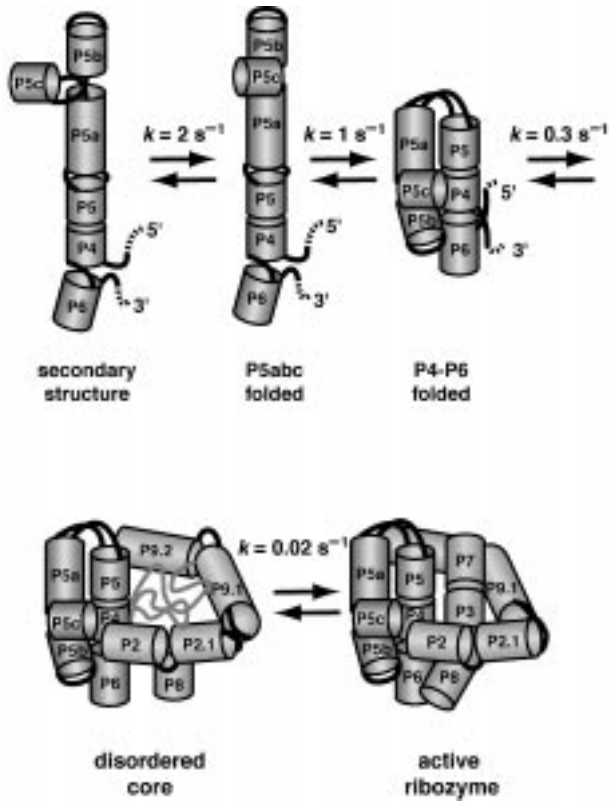



Figure 15. A model of the $Mg^{2+}$-dependent folding of the *Tetrahymena* group I intron, based upon time-dependent synchrotron hydroxyl radical footprinting. (Adapted from reference [104].)

By using the kinetic oligonucleotide hybidization assay to monitor folding it has been shown that the limiting step of the folding reaction is a magnesium-independent step involving the formation of tertiary structure in the triple-helical scaffold that is responsible for properly orienting the two structural domains.[107] In order to understand the nature of this rate-limiting step a selection scheme was developed to find mutations in the intron that accelerate the RNA's passage through this step in the folding pathway.[110] This selection yielded five variants with a faster-folding phenotype and wild-type catalytic activity.[110] Four of the variants contained single-point mutations in the P5abc subdomain that confered the fast-folding phenotype, which suggests that they affect the rate-limiting step of folding through a common mechanism, and importantly, each mutation does not affect the catalysis or the stability of the intron. The characterized mutations cluster around the adenosine-rich bulge of P5abc, likely disrupting the stable formation of this structure, just as single phosphorothioate substitutions in this same region disrupt the stable folding of the P4–P6 domain.[57] Even though P5abc does not directly contact the P3–P7 domain in models of the group I intron,[62] formation of this domain has a significant affect upon the rate of folding of P3–P7. Treiber et al.[110] conclude that since the entire P4–P6 domain folds prior to the rate-limiting step and destabilization of the metal-core motif in

P5abc accelerates the rate of formation of the P3–P7 domain that a tertiary interaction which is also present in the native state causes a kinetic trap in the folding pathway.

Further demonstration that native tertiary interactions within the intron create kinetic barriers in the folding pathway was demonstrated by measuring the rate of folding of wild-type and the fast-folding mutants in the presence of urea.[110, 111] Chemical denaturants act to increase the folding rate of proteins[112] and RNA[113, 114] by disrupting the formation of interactions that disfavor rapid progression towards the fully folded, native state. The folding of the wild-type intron is accelerated by increasing the concentrations of urea up to 3 M, but the folding rates of the selected faster folding mutants are affected to a significantly lesser extent.[110, 111] A detailed study of the effects of urea and mutations on the temperature dependence of the folding rate revealed that they both act to reduce the activation enthalpy for P3–P7 folding, which further supports the theory that a primary kinetic trap in the folding of the wild-type intron involves native tertiary interactions.[111]

Rather than folding being a simple process involving the sequential formation of intermediates as implied by the hierarchical folding model, it is more accurately described as an ensemble of molecules following parallel folding pathways, as represented in a "folding energy landscape".[115, 116] This is indicated by the observation that in a given population of group I RNAs, there is a small fraction capable of rapidly folding to reach the native state, while the rest of the molecules slowly fold, kinetically trapped in various intermediates.[113] Moderate amounts of chemical denaturant alter the distribution of molecules following parallel folding pathways by increasing the fraction of the population capable of evading kinetic traps to rapidly fold.[113] Thus, the kinetic intermediates described in the hierarchical folding model represent the most populated pathway in the folding energy landscape under standard folding conditions. Studies of the folding of the Th-intron have revealed that both native structures (such as the metal core of P4–P6)[110, 111] and non-native interactions (such as incorrect base-pair formation)[113, 117] play a significant role in shaping this landscape. Changes in temperature, solvent conditions, and nucleotide sequence have a profound affect upon the energy landscape, and have revealed new kinetic barriers and traps and allow alternative pathways of folding to become significantly populated.[111]

## 4. Developing Working Models of Large RNAs

### 4.1. Modeling the Tertiary Folding of RNA

Despite significant advances in the ability to prepare and crystallize RNA,[118–121] it remains difficult to obtain crystals that yield high-resolution structural information. In the absence of these structures, modeling the three-dimensional architecture of RNA has been a useful alternative. By far the most successful RNA modeling effort to date has been that for the Th-intron developed by Michel and Westhof.[62, 122] Comparative sequence analysis of 35 sequences of subgroup IC1 introns (to which the *Tetrahymena* intron belongs), provided crucial information for generating a consensus secondary structure and some types of tertiary interactions, and revealed invariant nucleotides likely to be essential for the formation of the correct fold or for biochemical function. These data were augmented by constraints provided by chemical and enzymatic probing of the intron RNA, and site-directed mutagenesis experiments verified the existence of base-specific tertiary interactions such as base-triples[123] and loop–loop interactions.[62] From these data a three-dimensional model of the *Tetrahymena* intron was constructed by using an approach based upon the hierarchical nature of RNA folding. Elementary substructures and motifs were built based upon known structures (such as double-stranded helices, hairpin loops, and pseudoknots), then hooked together by using interactive computer graphics, and finally the model was geometrically and stereochemically refined with a least-squares refinement routine.[124] The resulting model (Figure 16) has provided a framework for the interpretation of chemical activity and folding, has withstood extensive biochemical testing, and agrees with the overall fold of a group I intron catalytic core determined at low resolution (5–6 Å) by X-ray crystallography.[125]

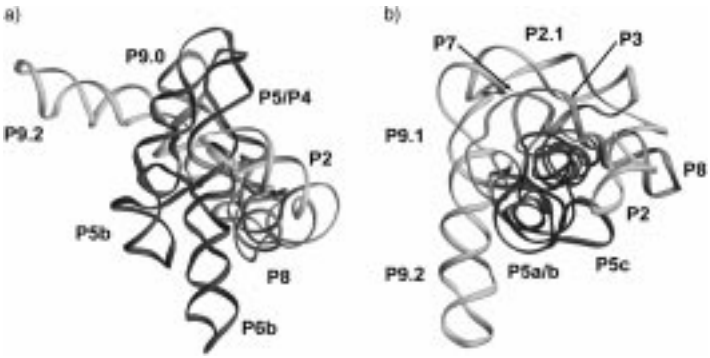



Figure 16. a) Model of the *Tetrahymena* group I intron developed by Lehnert et al.[62] The ribbon traces the backbone of the molecule, with the P4–P6 domain shown in black, the P3–P7 domain in light gray, and the P2 and P9 peripheral extensions in gray. b) A 90° rotation of the model, highlighting the packing of the P3–P7 domain between the P4–P6 and the peripheral extensions.

Producing accurate RNA models requires prior identification of as many tertiary interactions as possible, since these provide critical constraints on the global fold of the RNA that cannot be obtained from the secondary structure. For example, the model of the hepatitis delta virus ribozyme[126] has a global fold different from that seen in the 2.3 Å resolution crystal structure.[18] This difference stems mainly from the failure to predict the helix P1.1, which is essential for establishing the double-pseudoknot fold of the ribozyme. This interaction was hard to predict from the secondary structure of the ribozyme, as there are only two sequences (the genomic and antigenomic variants) with which to perform comparative analysis. Therefore, the key to better models of large RNAs lie in developing biochemical techniques for identifying tertiary contacts.

### 4.2. High-Resolution Probing of the Tertiary Structure of RNA

A clever method for biochemically determining tertiary interactions within RNAs relies on the use of phosphorothioate-tagged nucleotide analogues randomly incorporated into RNA transcripts. Following the separation of functional molecules from inactive (or unfolded) ones, the nucleotide positions where analogue substitution interferes with RNA structure or activity can be mapped by cleavage of the RNA with $I_2$.[127–129] The primary caveat of this technique is that if a phosphorothioate linkage interferes with RNA function, then the effect of nucleotide analogues will be masked. By using the interference mapping of nucleotide analogues the 2′-deoxy and 2′-methoxy analogues were utilized to identify the essential 2′-hydroxyl groups in tRNA for recognition by RNase P,[92, 128] and inosine analogues revealed the exocyclic amines in the Th-intron that are important for catalysis.[129] This approach has been greatly extended by Strobel and coworkers through the synthesis of a large library of nucleotide analogues (for a comprehensive review see reference [130]); these analogues have been used to probe the chemical and structural contribution of almost every chemical moiety within every adensoine in the Th-intron.[131]

A second experimental approach needed to be developed for the "chemogenetic" technique outlined above to mirror biological genetics. In genetics, after identification of a mutation that is deleterious to function, subsequent genetic screens are performed to find secondary mutations that restore function. If the mutation occurs at a site distinct from the original mutation, it is classified as a suppressor and may reveal an intramolecular or intermolecular interaction. Similarly, once interesting functional groups or atoms are determined from analogue interference mapping, potential tertiary interactions involving that position can be identified by synthesizing an RNA molecule that contains the original functional group mutation along with another randomly incorporated analogue.[132] This pool of molecules is again subjected to a folding or activity selection, followed by mapping of the sites of interference by $I_2$ cleavage of the RNA. In this experiment all of the analogues that were identified in the original interference mapping would still interfere except for the analogue that interacts with the site-specific mutation. Since the energetic cost of disruption of the tertiary interaction was already incurred by the site-specific mutation that is incorporated into every RNA molecule, the analogue that disrupts its partner does not further disrupt RNA function, that is, the deleterious effect of the analogue is suppressed by the presence of the site-specific mutation. Multiple tertiary interactions involving hydrogen-bonding pairs can potentially be revealed by testing a number of site-specific mutants, which then provide constraints for modeling the structure of an RNA.

With this methodology a structurally detailed model has been built for the catalytic core of the Th-intron.[46] This model includes the tertiary interactions necessary for the docking of the substrate helix (P1) into the catalytic cleft of this ribozyme, which is primarily mediated through hydrogen bonding of 2′-hydroxyl groups[133] and an exocyclic amine of a critical GU wobble base pair[134] in P1 that defines the cleavage site. It was determined from interference suppression analysis that a 2′-hydroxyl group and exocyclic amine of the guanosine in the GU pair interact with the minor groove of two noncanonical AA pairs in J4-5.[132] The modeled tertiary interaction resembles interactions observed between adjacent P4–P6 domain molecules in the crystal lattice, and between minor grooves observed in crystal structures of perpendicularly packed RNA duplexes (see Section 2.2.3). A second set of tertiary interactions that have been modeled in the catalytic core involve the formation of a triple helix in the minor groove between the J8/7 strand and P1, P3, and P4 helices (Figure 17). This model suggests that the 2′-hydroxyl

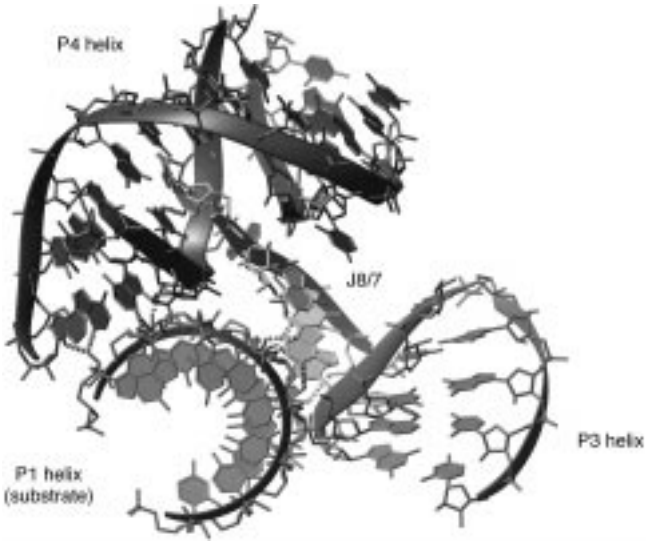



Figure 17. A model of the catalytic core of the *Tetrahymena* group I intron based upon pairwise constraints (shown by dotted lines) provided by nucleotide analogue interference mapping.[46] The single-stranded region J8/7 forms a minor groove-triple helix with the P1 substrate helix as well as minor-groove triple interactions with the P4 helix and a major-groove triple interaction with the P3 helix, which organizes the catalytic core of this ribozyme.

groups play a central role in the establishment of a minor groove triplex, as observed in a $tRNA^{Phe}$ and a model triplex (see Section 2.3.1). Since this model is predicated from biochemical experiments that probe structure and function at the atomic level, it may reveal details at higher resolution than models based upon covariation analysis and site-directed mutagenesis.

## 5. Summary and Outlook

RNA, being both a catalyst and a carrier of heritable information, provides an excellent opportunity to probe structure–function relationships and the macromolecular folding problem through the use of biochemical techniques not available to the protein chemist. Advances in our understanding of RNA structure provide a basis for exploring the RNA-folding problem as well as developing robust protocols for modeling large RNA structures in the absence of crystallographic or NMR spectroscopic data. The challenge

that lies ahead is to extend the lessons learned from the well studied systems of tRNA, the *Tetrahymena* group I intron, the hammerhead ribozyme, and RNase P to other more difficult systems such as the group II self-splicing intron. Furthermore, most RNA in the cellular environment interacts with numerous proteins to create ribonucleoprotein (RNP) enzymes such as the ribosome, spliceosome, telomerase, and signal recognition particle. Our understanding of structure, folding, and assembly, and functional mechanisms of these RNPs lags far behind that of the ribozymes discussed here. Future work will focus increasingly on RNA – protein interactions, and will be guided by themes that have emerged from these studies of RNA structure and function.

*The authors would like to thank Elizabeth Doherty, Patrick Zarrinkar, and Scott Strobel for helpful discussions. Support for this work was provided by a postdoctoral fellowship to R.T.B. from the Jane Coffin Childs Memorial Medical Research Fund and through grants from the Beckman Foundation, the Packard Foundation, the National Institutes, and the Howard Hughes Medical Institute (J.A.D.).*

Received: December 8, 1998 [A317IE]
German version: *Angew. Chem.* **1999**, *111*, 2472 – 2491

[1] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *Cell* **1982**, *31*, 147 – 157.
[2] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849 – 857.
[3] C. Levinthal, *J. Chim. Phys. Phys. Chim. Biol.* **1968**, *65*, 44 – 45.
[4] D. J. Lane, B. Pace, G. J. Olsen, D. A. Stahl, M. Sogin, N. Pace, *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 6955 – 6959.
[5] J. A. Kowalak, S. C. Pomerantz, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1993**, *21*, 4577 – 4585.
[6] M. Chastain, I. Tinoco, Jr., *Prog. Nucleic Acids Res. Mol. Biol.* **1991**, *41*, 131 – 177.
[7] Review: D. H. Turner, N. Sugimoto, S. M. Freir, *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167 – 192.
[8] review: C. R. Woese, N. R. Pace, *The RNA World* (Eds.: R. Gesteland, J. F. Atkins), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1993**, pp. 91 – 117.
[9] Review: C. Ehresmann, F. Baudin, M. Mougel, P. Romby, J.-P. Ebel, B. Ehresmann, *Nucleic Acids Res.* **1987**, *15*, 53 – 72.
[10] J. F. Milligan, D. R. Groebe, G. W. Witherell, O. C. Uhlenbeck, *Nucleic Acids Res.* **1987**, *15*, 8783 – 8798.
[11] S. R. Holbrook, *RNA structure and function* (Eds.: R. W. Simons, M. Grunberg-Manago), Cold Spring Harbor Press, Cold Spring Harbor, NY, **1998**, pp. 147 – 174.
[12] E. V. Puglisi, J. D. Puglisi, *RNA structure and function* (Eds.: R. W. Simons, M. Grunberg-Manago), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1998**, pp. 117 – 146.
[13] S. H. Kim, F. L. Suddath, G. J. Quigley, A. McPherson, J. L. Sussman, A. Wang, N. C. Seeman, A. Rich, *Science* **1974**, *185*, 435 – 440.
[14] J. D. Robertus, J. E. Ladner, J. T. Finch, D. Rhodes, R. D. Brown, B. F. C. Clark, A. Klug, *Nature* **1974**, *250*, 546 – 551.
[15] H. W. Pley, K. M. Flaherty, D. B. McKay, *Nature* **1994**, *372*, 68 – 74.
[16] W. G. Scott, J. T. Finch, A. Klug, *Cell* **1995**, *81*, 991 – 1002.
[17] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, J. A. Doudna, *Science* **1996**, *273*, 1678 – 1685.
[18] A. R. Ferré-D'Amaré, K. Zhou, J. A. Doudna, *Nature* **1998**, *395*, 567 – 574.
[19] E. Westhof, F. Michel, *RNA-Protein interactions* (Eds.: K. Nagai, I. W. Mattaj), IRL Press, New York, **1994**, pp. 25 – 51.
[20] J. E. Wedekind, D. B. McKay, *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 475 – 502.
[21] W. Saenger, *Principles of Nucleic Acid Structure*, Springer, New York, **1984**.
[22] S. H. Kim, *Prog. Nucleic Acid Res. Mol. Biol.* **1976**, *17*, 181 – 216.
[23] A. Jack, J. E. Lander, A. Klug, *J. Mol. Biol.* **1976**, *108*, 619 – 649.
[24] J. B. Murray, D. P. Terwey, L. Maloney, A. Karpeisky, N. Usman, L. Beigelman, W. G. Scott, *Cell* **1998**, *92*, 665 – 673.
[25] G. S. Bassi, N.-E. Mollegaard, A. I. H. Murchie, E. von Kitzing, D. M. J. Lilley, *Nat. Struct. Biol.* **1995**, *2*, 45 – 55.
[26] G. S. Bassi, A. I. H. Murchie, D. M. J. Lilley, *RNA* **1996**, *2*, 756 – 768.
[27] T. Tuschl, C. Gohlke, T. M. Jovin, E. Westhof, F. Eckstein, *Science* **1994**, *266*, 785 – 789.
[28] K. M. A. Amiri, P. J. Hagerman, *Biochemistry* **1994**, *33*, 13172 – 13177.
[29] J. B. Murray, A. A. Seyhan, N. G. Walter, J. M. Burke, W. G. Scott, *Chem. Biol.* **1998**, *5*, 587 – 595.
[30] Z. Shen, P. J. Hagerman, *J. Mol. Biol.* **1994**, *241*, 415 – 430.
[31] D. R. Duckett, A. I. H. Murchie, D. M. Lilley, *Cell* **1995**, *83*, 1027 – 1036.
[32] J. W. Orr, P. J. Hagerman, J. R. Williamson, *J. Mol. Biol.* **1997**, *275*, 453 – 464.
[33] R. T. Batey, J. R. Williamson, *RNA* **1998**, *4*, 984 – 997.
[34] J. H. Cate, A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, A. A. Szewczak, C. E. Kundrot, T. R. Cech, J. A. Doudna, *Science* **1996**, *273*, 1696 – 1699.
[35] G. R. Zimmerman, R. D. Jenison, C. L. Wick, J.-P. Simorre, A. Pardi, *Nat. Struct. Biol.* **1997**, *4*, 644 – 649.
[36] M. Costa, F. Michel, *EMBO J.* **1997**, *16*, 3289 – 3302.
[37] S. E. Lietzke, C. L. Barnes, J. A. Berglund, C. E. Kundrot, *Structure* **1996**, *4*, 917 – 930.
[38] S. R. Holbrook, C. Cheong, I. Tinoco, Jr., S.-H. Kim, *Nature* **1991**, *353*, 579 – 581.
[39] K. J. Baeyens, H. L. De Bondt, S. R. Holbrook, *Nat. Struct. Biol.* **1995**, *2*, 56 – 62.
[40] G. A. Leonard, K. E. McAuley-Hecht, S. Ebel, D. M. Lough, T. Brown, W. N. Hanter, *Structure* **1994**, *2*, 483 – 494.
[41] M. C. Wahl, S. T. Rao, M. Sundaralingam, *Nat. Struct. Biol.* **1996**, *3*, 24 – 31.
[42] R. D. Blake, J. Massoulié, J. R. Fresco, *J. Mol. Biol.* **1967**, *30*, 291 – 308.
[43] J. Massoulié, *Eur. J. Biochem.* **1968**, *3*, 439 – 447.
[44] R. Klinck, J. Liquier, E. Taillandier, C. Gouyette, T. Huynh-Dinh, E. Guittet, *Eur. J. Biochem.* **1995**, *233*, 544 – 553.
[45] Review: J. D. Puglisi, J. R. Williamson, *The RNA World*, *2nd ed.* (Eds.: R. F. Gesteland, T. R. Cech, J. F. Atkins), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1999**, pp. 403 – 425.
[46] A. A. Szewczak, L. Ortoleva-Donnelly, S. P. Ryder, E. Moncoeur, S. A. Strobel, *Nat. Struct. Biol.* **1998**, *5*, 1037 – 1042.
[47] C. Turek, 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, L. Gold, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1364 – 1368.
[48] C. R. Woese, S. Winker, R. R. Gutell, *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 8467 – 8471.
[49] M. Molinaro, I. Tinoco, Jr., *Nucleic Acids Res.* **1995**, *23*, 3056 – 3063.
[50] H. Heus, A. Pardi, *Science* **1991**, *253*, 191 – 194.
[51] F. M. Jucker, H. A. Heus, P. F. Yip, E. H. Moors, A. Pardi, *J. Mol. Biol.* **1996**, *264*, 968 – 980.
[52] L. Jaeger, F. Michel, E. Westhof, *J. Mol. Biol.* **1994**, *236*, 1271 – 1276.
[53] F. L. Murphy, T. R. Cech, *J. Mol. Biol.* **1994**, *236*, 49 – 63.
[54] M. Costa, F. Michel, *EMBO J.* **1995**, *14*, 1276 – 1285.
[55] H. Pley, K. Flaherty, D. McKay, *Nature* **1994**, *372*, 111 – 113.
[56] D. L. Abramovitz, A. M. Pyle, *J. Mol. Biol.* **1997**, *266*, 493 – 506.
[57] J. H. Cate, R. L. Hanna, J. A. Doudna, *Nat. Struct. Biol.* **1997**, *4*, 553 – 558.
[58] J. A. Ippolito, T. A. Steitz, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 9819 – 9824.
[59] M. Zacharias, P. J. Hagerman, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 6052 – 6056.
[60] M. Zacharias, P. J. Hagerman, *J. Mol. Biol.* **1995**, *247*, 486 – 500.
[61] M. S. Chapman, M. G. Rossmann, *Structure* **1995**, *3*, 151 – 162.

[62] V. Lehnert, L. Jaeger, F. Michel, E. Westhof, *Chem. Biol.* **1996**, *3*, 993 – 1009.
[63] J. P. Marino, R. S. Gregorian, Jr., G. Csankovszki, D. M. Crothers, *Science* **1995**, *268*, 1448 – 1454.
[64] A. J. Lee, D. M. Crothers, *Structure* **1998**, *6*, 993 – 1005.
[65] K.-Y. Chang, I. Tinoco, Jr., *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 8705 – 8709.
[66] K.-Y. Chang, I. Tinoco, Jr., *J. Mol. Biol.* **1997**, *269*, 52 – 66.
[67] R. S. Gregorian, Jr., D. M. Crothers, *J. Mol. Biol.* **1995**, *248*, 968 – 984.
[68] C. C. Correll, B. Freeborn, P. B. Moore, T. A. Steitz, *Cell* **1997**, *91*, 705 – 712.
[69] A. A. Szewczak, P. B. Moore, *J. Mol. Biol.* **1995**, *247*, 81 – 98.
[70] J. D. Puglisi, J. R. Wyatt, I. Tinoco, Jr., *J. Mol. Biol.* **1990**, *214*, 437 – 453.
[71] J. R. Wyatt, J. D. Puglisi, I. Tinoco, Jr., *J. Mol. Biol.* **1990**, *214*, 455 – 470.
[72] L. X. Shen, I. Tinoco, Jr., *J. Mol. Biol.* **1995**, *247*, 963 – 978.
[73] X. Chen, H. Kang, L. X. Shen, M. Chamorro, H. E. Varmus, I. Tinoco, Jr., *J. Mol. Biol.* **1996**, *260*, 479 – 483.
[74] M. H. Kolk, M. van der Graaf, S. S. Wijmenga, C. W. A. Pleij, H. A. Heus, C. W. Hilbers, *Science* **1998**, *280*, 434 – 438.
[75] L. Su, L. Chen, M. Egli, J. M. Berger, A. Rich, *Nat. Struct. Biol.* **1999**, *6*, 285 – 292.
[76] S. P. Rosenstein, M. D. Been, *Biochemistry* **1990**, *29*, 8011 – 8016.
[77] J. B. Smith, G. Dinter-Gottlieb, *Nucleic Acids Res.* **1991**, *19*, 1285 – 1289.
[78] J. A. Latham, T. R. Cech, *Science* **1989**, *245*, 276 – 282.
[79] D. C. Celander, T. R. Cech, *Biochemistry* **1990**, *29*, 1355 – 1361.
[80] B. Laggerbauer, F. L. Murphy, T. R. Cech, *EMBO J.* **1994**, *13*, 2669 – 2676.
[81] F. L. Murphy, T. R. Cech, *Biochemistry* **1993**, *32*, 5291 – 5300.
[82] E. L. Christian, M. Yarus, *J. Mol. Biol.* **1992**, *228*, 743 – 758.
[83] E. L. Christian, M. Yarus, *Biochemistry* **1993**, *32*, 4477 – 4480.
[84] E. A. Doherty, J. A. Doudna, *Biochemistry* **1997**, *36*, 3159 – 3169.
[85] G. Van Der Horst, A. Christian, T. Inoue, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 184 – 188.
[86] E. A. Doherty, D. Herschlag, J. A. Doudna, *Biochemistry* **1999**, *38*, 2982 – 2990.
[87] J. A. Doudna, T. R. Cech, *RNA* **1995**, *1*, 36 – 45.
[88] C. Guerrier-Takada, S. Altman, *Science* **1984**, *223*, 285 – 286.
[89] W. D. Hardt, J. M. Warnecke, V. A. Erdmann, R. K. Hartmann, *EMBO J.* **1995**, *14*, 2935 – 2944.
[90] W. D. Hardt, V. A. Erdmann, R. K. Hartmann, *RNA* **1996**, *2*, 1189 – 1198.
[91] T. Pan, A. Loria, K. Zhong, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 12510 – 12514.
[92] F. Conrad, A. Hanne, R. K. Gaur, G. Krupp, *Nucleic Acids Res.* **1995**, *23*, 1845 – 1853.
[93] T. Pan, *Biochemistry* **1995**, *34*, 902 – 909.
[94] A. Loria, T. Pan, *RNA* **1996**, *2*, 551 – 563.
[95] P. Z. Qin, A. M. Pyle, *Curr. Opin. Struct. Biol.* **1998**, *8*, 301 – 308.
[96] P. E. Cole, D. M. Crothers, *Biochemistry* **1972**, *11*, 4368 – 4374.
[97] D. M. Crothers, P. E. Cole, C. W. Hilbers, R. G. Shulman, *J. Mol. Biol.* **1974**, *87*, 63 – 88.
[98] C. W. Hilbers, G. T. Robillard, R. G. Shulman, R. D. Blake, P. K. Webb, R. Fresco, D. Riesner, *Biochemistry* **1976**, *15*, 1874 – 1882.
[99] E. R. Hawkins, S. H. Chang, W. L. Mattice, *Biopolymers*, **1977**, *16*, 1557 – 1566.
[100] A. Stein, D. M. Crothers, *Biochemistry* **1976**, *15*, 160 – 168.
[101] A. Stein, D. M. Crothers, *Biochemistry* **1976**, *15*, 157 – 160.
[102] A. R. Banerjee, J. A. Jaeger, D. H. Turner, *Biochemistry* **1993**, *32*, 153 – 163.
[103] B. Sclavi, S. Woodson, M. Sullivan, M. R. Chance, M. Brenowitz, *J. Mol. Biol.* **1997**, *266*, 144 – 159.
[104] B. Sclavi, M. Sullivan, M. R. Chance, M. Brenowitz, S. A. Woodson, *Science* **1998**, *279*, 1940 – 1943.
[105] P. P. Zarrinkar, J. R. Williamson, *Nucleic Acids Res.* **1996**, *24*, 854 – 858.
[106] P. P. Zarrinkar, J. R. Williamson, *Science* **1994**, *265*, 918 – 924.
[107] P. P. Zarrinkar, J. R. Williamson, *Nat. Struct. Biol.* **1996**, *3*, 432 – 438.
[108] A. R. Banerjee, D. H. Turner, *Biochemistry* **1995**, *34*, 6504 – 6512.
[109] W. D. Downs, T. R. Cech, *RNA* **1996**, *2*, 718 – 732.
[110] D. K. Treiber, M. S. Rook, P. P. Zarrinkar, J. R. Williamson, *Science* **1998**, *279*, 1943 – 1946.
[111] M. S. Rook, D. K. Treiber, J. R. Williamson, *J. Mol. Biol.* **1998**, *281*, 609 – 620.
[112] J. S. Weissman, P. S. Kim, *Science* **1991**, *253*, 1386 – 1393.
[113] J. Pan, D. Thirumalai, S. A. Woodson, *J. Mol. Biol.* **1997**, *273*, 7 – 13.
[114] T. Pan, T. R. Sosnick, *Nat. Struct. Biol.* **1997**, *4*, 931 – 938.
[115] D. Thirumalai, S. A. Woodson, *Acc. Chem. Res.* **1996**, *29*, 433 – 439.
[116] K. A. Dill, H. S. Chan, *Nat. Struct. Biol.* **1997**, *4*, 10 – 19.
[117] J. Pan, S. A. Woodson, *J. Mol. Biol.* **1998**, *280*, 597 – 609.
[118] S. R. Price, N. Ito, C. Oubridge, J. M. Avis, K. Nagai, *J. Mol. Biol.* **1995**, *249*, 398 – 408.
[119] A. R. Ferré-D'Amaré, J. A. Doudna, *Nucleic Acids Res.* **1996**, *24*, 977 – 978.
[120] J. A. Doudna, C. Grosshans, A. Gooding, C. E. Kundrot, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 7829 – 7833.
[121] W. G. Scott, J. T. Finch, F. Grenfell, J. Fogg, T. Smith, M. J. Gait, A. Klug, *J. Mol. Biol.* **1995**, *250*, 327 – 332.
[122] F. Michel, E. Westhof, *J. Mol. Biol.* **1990**, *216*, 585 – 610.
[123] M. A. Tanner, T. R. Cech, *Science* **1997**, *275*, 847 – 849.
[124] E. Westhof, *J. Mol. Struct. (THEOCHEM)* **1993**, *286*, 203 – 210.
[125] B. L. Golden, A. R. Gooding, E. R. Podell, T. R. Cech, *Science* **1998**, *282*, 259 – 264.
[126] N. K. Tanner, S. Schaff, G. Thill, E. Petit-Koskas, A. Crain-Denoyelle, E. Westhof, *Curr. Biol.* **1994**, *4*, 488 – 498.
[127] G. Gish, F. Eckstein, *Science* **1988**, *240*, 1520 – 1522.
[128] R. K. Gaur, G. Krupp, *Nucleic Acids Res.* **1993**, *21*, 21 – 26.
[129] S. A. Strobel, K. Shetty, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2903 – 2908.
[130] S. A. Strobel, *Biopolymers*, in press.
[131] L. Ortoleva-Donnelly, A. A. Szewczak, R. R. Gutell, S. A. Strobel, *RNA* **1998**, *4*, 498 – 519.
[132] S. A. Strobel, L. Ortoleva-Donnelly, S. P. Ryder, J. H. Cate, E. Moncoeur, *Nat. Struct. Biol.* **1998**, *5*, 60 – 66.
[133] S. A. Strobel, T. R. Cech, *Biochemistry* **1993**, *32*, 13593 – 13604.
[134] S. A. Strobel, T. R. Cech, *Science* **1995**, *267*, 675 – 679.
[135] D. Söll, *The RNA world* (Eds.: R. Gesteland, J. F. Atkins), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, **1993**, pp. 157 – 184.
[136] J. L. Sussman, S. R. Holbrook, R. W. Warrant, G. M. Church, S. H. Kim, *J. Mol. Biol.* **1978**, *123*, 607 – 630.
[137] M. Carson, *J. Appl. Crystallogr.* **1991**, *47*, 110.
[138] S. H. Damberger, R. R. Gutell, *Nucleic Acids Res.* **1994**, *22*, 3508 – 3510.
[139] J. S. Kieft, I. J. Tinoco, *Structure* **1997**, *5*, 713 – 721.
[140] D. E. Draper, *Nat. Struct. Biol.* **1996**, *3*, 397 – 400.