tRNA Structure

Eric Westhof, Institute of Molecular and Cellular Biology, Strasbourg, France
Pascal Auffinger, Institute of Molecular and Cellular Biology, Strasbourg, France

Transfer ribonucleic acid (tRNA) molecules that participate in the elongation step of protein synthesis on the ribosome have a conserved secondary structure, known as the cloverleaf, and fold into a common three-dimensional architecture.

Introduction

Single-stranded RNA (ribonucleic acid) molecules are ubiquitous in cells and play many biological roles. However, three main types of RNA dominate molecular biology: the messenger RNAs (mRNAs), the transfer RNAs (tRNAs) and the ribosomal RNAs (rRNAs). They all participate in the biosynthesis of proteins. The mRNAs contain the sequence of nucleotide triplets or codons which will be read by tRNAs during the translation process on the ribosomes of which the rRNAs constitute central components. In mRNAs, the sequence alone is the primary determinant of biological function. However, for tRNAs and rRNAs, the folding in space of the polynucleotide RNA chain in a native tertiary structure, forms the basis of biological activity.

The existence of some type of ‘adaptor’ RNA molecules acting as go-betweens between the RNA world and the protein universe was predicted by Francis Crick as early as 1955, long before transfer RNA molecules were characterized biochemically. Because there are 61 different codons in the genetic code, one would expect about the same number of different tRNAs, distributed among the 20 different isoacceptor families corresponding to the 20 different amino acids. However, if the wobble hypothesis of Crick (which relaxes the Watson–Crick complementarity and considers not only G≡U but also ι≡U, ι≡C, or ι≡A pairings involving inosine (I), an oxidized form of adenosine) is followed, only 31 different tRNAs are necessary to decode all codons (32 with the additional initiator tRNA) (Crick, 1966). The recent sequencing of whole eubacterial genomes has revealed that the number of tRNA genes varies between 33 (in Mycoplasma genitalium, a parasitic bacterium with a minimal genome) and 88 (in the Gram-positive bacterium Bacillus subtilis) with most being between 44 and 46. It should be added that the copy number of tRNA genes can be as high as four and that some tRNA variants present nucleotide changes outside the anticodon. Thus, in Escherichia coli the 84 tRNA genes code 45 tRNA species with only 41 tRNA species having different anticodons.

As with every biological macromolecule, the molecular evolution of tRNAs is intimately coupled to the structural constraints imposed by the nature of the polymer and its functions. Each tRNA molecule has to evolve under two opposing constraints. On the one hand it needs a three-dimensional architecture that allows it to fit precisely in the ribosome-binding sites for promoting protein synthesis, but on the other hand it needs to contain enough molecular diversity to guarantee specific recognition with a unique cognate tRNA synthetase. Recall that tRNA synthetase aminoaetylates the tRNA 3'-end adenosine with the amino acid specific to that tRNA. Indeed, a tRNA carries an anticodon triplet complementary to a given codon and should consequently be charged by the synthetase solely with the amino acid specified by the anticodon–codon interactions. The signatures of those opposing biological constraints are apparent in the network of invariant residues and interactions that maintain a common architecture on which enough molecular diversity can be coded for specific recognition with the tRNA synthetases. These structural constraints are clearly seen in the invariance in the length of some helices of the secondary structure and in the constant presence of some residues at definite positions. However, tRNAs constitute a set of molecules with which biological evolution has tinkered enormously. Although plant chloroplast and some mitochondrial tRNAs present high sequence homologies to eubacterial tRNAs, other mitochondrial tRNAs display a palette of unusual structural features. tRNAs responsible for the cotranslational insertion of the 21st amino acid, selenocysteine, also contain odd features.

Nature has further exploited the sequence and structural diversity compatible with tRNA folding to adapt tRNA structure to functions unrelated to protein synthesis. For example, a special tRNA charged with a glutamic acid is used during chlorophyll synthesis in plants, while another tRNA charged with a glycine is necessary for peptidoglycan synthesis in bacteria. Retroviruses, like VIH, require a tRNA (in human cells, a tRNA Lys) as a primer for the replication of their genomic RNA. Several plant viruses contain at their 3′-ends, a tRNA-like structure, the integrity of which is necessary for viral replication.
Cloverleaf Structure

The secondary structure of an RNA molecule is generally defined by the set of contiguous Watson–Crick pairs (including the G·U wobble pairs) forming helices between segments of the single-stranded RNA. For a given family of RNA sequences, the ensemble of the so-defined helices is conserved as well as their relative arrangement. This is why secondary structures can be established by sequence comparisons. The basic hypothesis is, indeed, that partly divergent, but nevertheless functionally and historically related, molecules from various biological origins fold into similar secondary and tertiary structures. Helices can occur at homologous positions in the secondary structures of two different sequences if compensatory changes occur on both strands so as to maintain complementarity between the bases in the Watson–Crick sense. For example, when Holley sequenced the first tRNA in 1965, he suggested several possible secondary structures (Holley et al., 1965). However, when a second tRNA was sequenced, the only secondary structure common to both sequences was the one in the form of a planar cloverleaf (Figure 1a). All known cytoplasmic elongator tRNA sequences (of lengths varying between 74 and 98 nucleotides) can be folded in the cloverleaf structure. The cloverleaf structure consists of four helices and three hairpin loops. The four helices are called: the acceptor (AA) helix (or stem) because it will carry the amino acid once charged by the cognate aminoacyl synthetase; the dihydrouridine (D) hairpin because it often contains the modified base dihydrouracil; the anticodon (AC) hairpin because its apical loop presents the anticodon triplet; and the thymine (T) hairpin because its loop often contains the thymine base unusual in RNAs.

Universally Invariant and Semi-Invariant Bases

Alignments of tRNA sequences, equivalent to inserting tRNA sequences on the cloverleaf template, reveal positions that are always occupied by a single type of base (‘invariant’) and positions that are occupied either only by purines or by pyrimidines (‘semi-invariant’). About 22 nucleotides belong to these categories. As is usual with biological sequences, those conservations are never absolute and, for clarity, we will focus the following description on cytoplasmic elongator tRNAs, for which the numbering of yeast tRNA<sup>His</sup> constitutes the reference system, since it was the first tRNA structure solved by X-ray crystallography. Figure 1b shows the distribution of the four nucleotides found in the 932 cytoplasmic tRNA genes listed in the 1998 release of the tRNA database (Sprinzl et al., 1998). In the following, R stands for a purine (A or G), Y for a pyrimidine (U or C) and N for any of the four nucleotides; standard Watson–Crick secondary base pairs will be separated by a dash (e.g. U12–A23); non-Watson–Crick pairs by a dot (e.g. G10·U25 or the <i>trans</i> U8·A14); and the third base of triples will be separated by three dots from the interacting secondary base pair on the side of the contacting base (e.g. G45...G10·U25).

tRNA synthetases attach the cognate amino acid (i.e. the one dictated by the codon–anticodon pairings) on either the 3'-OH or the 2'-OH group of the terminal adenosine. The first residue of all tRNAs, which starts the first base pair of the acceptor helix, carries a 5'-phosphate group. It originates in the biosynthesis of tRNAs. After being transcribed from tRNA genes, the tRNA transcripts are maturated at the 5'-end by the ribonucleoprotein enzyme RNAaseP, which clips off the 5'-end segment of the tRNA gene. The catalysis is effected by the RNA moiety of the RNAaseP, a process which leaves a 3' OH and a 5' phosphate. The maturated tRNAs is very complex and requires a great number of enzymes since tRNAs are characterized by the presence of several modified nucleotides.

The lengths of the helices and loops are either conserved or vary between defined limits. The distribution of the four Watson–Crick pairs within helices is not uniform. For example, the acceptor helix starts in three-quarters of sequences with a G1–C72 pair and ends very rarely with a C7–G66 base pair. The D helix also frequently starts with a R10–Y25 followed by a Y11–R24 base pair and ends in half of the sequences with a C13–G22 pair. The conservation is very strong for the last base pair of the thymine helix which is almost always G53–C61. Only the central base pairs of the acceptor helix as well as the very central one in the anticodon helix present an almost equal distribution of the four Watson–Crick combinations. Interestingly, those two sets of base pairs constitute determinants for recognition by aminoacyl synthetases.

The acceptor helix has seven base pairs, except in tRNA<sup>His</sup> where an eighth residue at the 5' end is added posttranscriptionally. The acceptor helix possesses a 3' dangling strand with four nucleotides, the sequence of which is -RCCA-3'OH (R is most often A and half as frequently G; that position is unfrequently occupied by a pyrimidine). Position 73 is called ‘discriminator’ owing to its role in synthetase selection of tRNAs. In RNA helices, interstrand stacking occurs in 5'-YR-3' steps. The last base pair of the acceptor helix, often a R1–Y72 base pair, is thus probably stabilized by interstrand stacking with R73. The first four base pairs of the acceptor helix are important discriminant elements for synthetase recognition.

The dihydrouridine helix generally possesses four base pairs, while its hairpin loop is more variable with 7–11 residues. The variation occurs in two regions, called α and β, situated on either side of two invariant guanine residues, G18 and G19 (Figure 1a). The dihydrouridine residues occupy either or both of those positions. In the loop, the first (14) and last (21) residues are always a purine, mainly...
an adenine. The first three base pairs of the D helix constitute identity elements for aminoacyl synthetases.

The anticodon hairpin has five base pairs in the helix and seven residues in the loop. The second residue of the anticodon loop is always a uridine (U33); it precedes the three bases of the anticodon, the first position of which is called the wobble base (34). Following the anticodon triplet, a highly modified nucleotide is generally present (position 37). Except for the first two residues (32 and 33),...
the residues of the anticodon loop are important identity elements for synthetases.

After the anticodon helix, there is a variable region which may vary between 4 and 21 residues. According to the size of the variable loop, tRNAs belong to class I or II. Class I tRNAs have 4–5 nucleotides in the variable loop and class II tRNAs from 10 to 24. In class II tRNAs, the variable loop is long enough to form a fifth helix. Only tRNAs specific for the amino acids leucine and serine (with tyrosine in eubacteria and organelles only) belong to class II.

Finally, the thymine hairpin has, like the anticodon hairpin, five base pairs and seven residues in the loop. However, the base conservation in the thymine loop is very different from that in the anticodon loop. The last base pair of the thymine helix, G53–C61, is highly conserved. In the thymine loop, the first three residues are rather well conserved and modified to -TΨC- where Ψ stands for pseudouridine, a modified base in which the link between the base and the sugar is C(1')-C5 instead of C(1')-N1. A pseudouridine is generally present at the second position of the loop in tRNAs but, since pseudouracils represent more than 50% of the modified bases in tRNAs, many more modified positions exist in the cloverleaf (Auffinger and Westhof, 1998).

The Two Domains of the Three-dimensional Architecture

The folding in three dimensions of the cloverleaf secondary structure posed a tantalizing problem for many years. In a remarkable paper of 1968, Levitt proposed a three-dimensional model of tRNA (Levitt, 1969). Although the overall architecture of the model turned out to be wrong, that article laid the foundations for our thinking about the relations between biological evolution and biomolecular structure and function. Since the first correct tracing of a tRNA chain in crystals of yeast tRNA^Phe in 1974 (Suddath et al., 1974), only one other crystal structure of a free tRNA has been solved (yeast tRNA^Asp in 1980) (Moras et al., 1980). In addition, several structures of tRNAs complexed with their cognate aminoacyl synthetases now exist. An examination of the three-dimensional folding revealed by the crystal structures allows a structural basis for all of the base conservations observed to be suggested.

The four helices of the cloverleaf stack on each other coaxially and two by two, forming two main arms or domains (Figure 2a). Thus, the acceptor helix and the thymine hairpin form the acceptor arm and the dihydrouridine together with the anticodon hairpins form the anticodon arm. The two coaxial and contiguous stacks make an angle of about 90° between them, giving to the overall architecture of tRNAs the appearance of the letter L or the Greek Γ. At the two ends, which are about 7.5–8.0 nm apart, are the anticodon triplet and the -CCA 3'-terminus, extremities which, when the tRNA is bound to the acceptor site of the ribosome (A site), should, respectively, bind to the codon triplet of the mRNA and bring the attached amino acid into the peptidyl reaction centre for reacting with the peptide chain carried by the preceding tRNA present in the peptidyl site (P site). The interfaces between the two coaxial helices are different in each domain. The tRNA chain is continuous between the 3' end of the thymine helix and the beginning of the acceptor helix, while the last residue of the 5' strand of the acceptor helix adopts the C2'-endo sugar pucker which facilitates branching off the helix. By contrast, at the interface between the dihydrouridine and the anticodon helices, there is usually a non-Watson–Crick base pair linking residues 26 and 44 (e.g. an imino GtA pair) with a twist angle with the last base pair of the dihydrouridine stem of around 45°, much larger than the standard 33° present between base pairs of RNA helices.

The L-shaped architecture is locked in place by two main structural features. First, the single-stranded regions linking the two domains (residues 8–9 between the acceptor and the dihydrouridine helices, as well as the variable region between the anticodon and the thymine helices) adopt conformations such that their residues form base triples in the deep groove of the dihydrouridine (Figure 2b). Secondly, those base triples position the D and T loops so that extremely precise tertiary interactions can occur between them. Fluorescence and ultraviolet melting experiments have shown that formation of the base triples in the deep groove of the D helix is the rate-limiting step in the tertiary folding of tRNAs.

The stereochemistry of the nucleotides is rather uniform in helices (ribose puckers in the C3'-endo domain, bases in anti orientation with respect to the sugar, helical phosphate torsion angles) and the folding of the chain is accomplished in single-stranded segments by altering the ribose pucker to the C2'-endo domain or by rotating the torsion angles at the phosphorus atoms away from the helical conformations (Sundaralingam, 1973).

Unusual Hydrogen Bonds Maintain the Tertiary Structure

Residues not participating in helices defining the secondary structure, i.e. residues of loops and single-stranded regions, form tertiary interactions either by interacting between themselves or by interacting with base pairs within a helix (leading to base triples). The four natural bases can interact via hydrogen bonding in many different ways. There are 27 pairs with two standard hydrogen bonds. The recent crystal structures of other RNAs expand further the recognition modes, since they reveal pairings with only one interbase hydrogen bond or with a water molecule
bridging the two bases. The fundamental property of the Watson–Crick pairs (the isosteric geometry of all four possible pairs G–C, C–G, A–U, U–A involving complementary bases) is thus even more remarkable. In order to discuss the tertiary interactions in tRNAs, it is necessary to explain two terms. First, the nucleotides in a pair may approach with their sugars on the same sides of the H bonds or on either side of them. Following the chemical literature, in the first case, the pairing is said to be **cis** and, in the second, **trans** (often called ambiguously ‘reverse’).

---

**Figure 2**  Two-dimensional and three-dimensional representation of the tertiary structure of elongator transfer RNAs. (a) Two-dimensional representation of the tertiary structure of tRNAs, as proposed by Kim (1978), which emphasizes the two main domains and the tertiary contacts linking them. Only the secondary structure can be represented in a plane without crossing lines (in other words, mathematically, a secondary structure is equivalent to a planar graph). The representation of a three-dimensional structure and of the underlying tertiary contacts can be drawn in a plane, but with several line crossings. Such schematic drawings are, however, useful for quick assessment and comparisons of tertiary contacts. The Kim representation shows clearly the two domains, the contacts between the T and D loops and the tertiary base pairs and triples between the single-stranded segments and the D hairpin. The contacts represented correspond to those of yeast tRNA<sub>Asp</sub> (Westhof et al., 1985). (b) Stereoview of a schematized representation of the tertiary structure of yeast tRNA<sub>Asp</sub>. The sugar–phosphate backbone is drawn as a ribbon and the base pairs as rods. The colour code is the same as in Figure 2a. Notice the characteristic deep and shallow grooves of an RNA helix in the acceptor and thymine helices, respectively.
Secondly, purine bases can interact via their edge carrying the imidazole ring (involving ring nitrogen N7) instead of only their pyrimidine edge, as in Watson–Crick pairings. Such pairings are called Hoogsteen pairs, after their discovery by Hoogsteen in 1956 in cocrystals of derivatives of U and A. Figure 3a shows examples of a variety of pairings. Thus, elongator tRNAs contain two Hoogsteen U•A pairs, both with a trans orientation of the sugars. The first trans Hoogsteen pair occurs between the invariant U at position 8, the first residue after the acceptor helix, and the invariant A at position 14, the first residue of the D loop. The U8•A14 pair, thus, links the two main domains of tRNAs. The second trans Hoogsteen pair (T54•A58) occurs between two invariant residues within the T loop, the T at position 54 and the A at position 58. There is also a third trans base pair, between the semi-invariant residues R15 and Y48, but involving the Watson–Crick sites. The trans Watson–Crick R15•Y48 base pair, sometimes called after Levitt who first noticed the covariation, stacks with the trans Hoogsteen U8•A14 pair. They are both central for the maintenance of the L-shaped structure of tRNAs (Figure 3a).

Class I tRNAs contain three base triples, all in the deep groove of the D helix. Of the four base pairs of the D helix, only the 11–24 pair does not contact a third base. Both single-stranded regions linking the two domains participate in the network of triple interactions. Residue 9 interacts with base 23 of base pair 12–23 (12–23...9), while residues 45 and 46 of the variable loop interact, respectively, with base pair 10–25 (45...10–25) and base pair 13–22 (13–22...46). The type of the hydrogen bonding sites (Watson–Crick or Hoogsteen) is related to the local orientation of the chains. Thus, cis Watson–Crick pairs (as in usual double helices) lead to antiparallel strands while, with equivalent stereochemistry of the nucleotides, trans Watson–Crick pairs lead to a local parallel orientation of the strands (e.g. R15•Y48). The reverse is true for Hoogsteen pairs between a pyrimidine and a purine (e.g. U8•A14) and for Hoogsteen pairs involving the Watson–Crick of one purine with the Hoogsteen sites of another purine (e.g. G22...A46 in yeast tRNAAsp). However, purine–purine pairs involving only the Hoogsteen sites are locally antiparallel when cis and parallel when trans (e.g. A9...A23 in yeast tRNAAsp). The local orientations of the chains cannot be altered without altering profoundly the overall topology or flipping some bases in the syn orientation with respect to the sugar (a conformation extremely rare in crystal structures of RNA fragments). These structural constraints impose definite covariations between residues involved in triple interactions.

The set of triples in the deep groove of the D helix is such that residue 9 interdigitates between residues 45 and 46 of the variable link and the contacts are made alternatively to the 5' and 3' strands of the D helix. Again, this topology dictates the relative strand orientations: residue R45 interacts with G10 and is parallel to it, but R46 is antiparallel to residue R22 to which it binds, while the inbetween residue R9 is parallel to its binding residue 23. Because the D and anticodon helices stack coaxially with a right-handed twist at the interface, the 3'-dangling variable region faces the deep groove and is antiparallel to the 3' strand of the D helix, which leads to the 5' strand of the anticodon helix. On the other hand, the junctions between the two domains occur at the internucleotide linkages of residues 7–8 and 48 so that residues 7 and 49 could almost be linked as in a continuous helix. In short, the two single-stranded segments linking the two domains run antiparallel to each other, facing the deep groove of the D helix.

Figures 2 and 3 correspond to the molecular structure of the yeast tRNAAsp as determined by X-ray crystallography (Westhof et al., 1985). Figure 4 illustrates the sequence variability in 33 available tRNAAsp genes from other organisms. Invariant residues are clearly seen, e.g. U8, A14, and G18 or G10. Covariations typical of secondary structure are displayed by residues 11 and 24 or 12 and 23, the two central base pairs of the D helix. The flanking base pairs present biases, G10–Y25 and Y13–R22, with some unusual oppositions (for example Y13•Y22). Correlations between the third base and the secondary pair forming triples are not easy to detect. For example (not apparent on Figure 4), any of the four bases is found with the very frequent G10–C25, although a G at position 45 is by far the most frequent situation. In contrast, position 9, either A or G, covaries with base pair 12–23 so that A9...U12–A23 and G9...G12–C23 are most frequently observed, although either A9 (and almost G9) can be found together with any of the four Watson–Crick base pairs (see Table 1). Even the residues involved in triples appear to vary. For example, in the structure of E. coli tRNA^Gln complexed with its cognate synthetase (Rould et al., 1989), it is residue 45 which interacts with the 13–22 base pair, and not residue 46 which is pointing on the exterior of the molecule. In that structure, the 13–22 base pair is an unusual A•A pair (see below). Such adaptability blurs the covariation tables (Klug et al., 1974; Gautheret et al., 1995).

Several isolated additional H bonds contribute to the integrity of the tertiary structure. They involve the O2'...
Comparison gives only a glimpse of the possible base variations.

The anticodon loop should bind precisely the mRNA triplet and the other from the amino group of the invariant C61. Only some of those additional contacts are observed in all crystalline forms of tRNAs and a comparison between different crystal structures reveals a subtle diversity of slight rearrangements and alternative contacts, often including distorted hydrogen bond geometries and the use of the weaker C-H...O/N hydrogen bondings (see Figure 5).

Conservations of Bases and Loop Conformations

As mentioned above, both anticodon and thymine loops always contain seven residues. However, the patterns of conservation are rather different and reflect their respective functions. In the anticodon loop, the only strictly invariant residue is U33, while in the thymine loop almost five residues are highly conserved, -T54PCCAT58-. The anticodon loop should bind precisely the mRNA triplet complementary to its anticodon bases to form a regular RNA helix, since recognition is mainly of the Watson–Crick type. In contrast, structurally the T loop binds intramolecularly to the D loop. The conformation of the T loop might also be important for RNAseP binding to...
tRNAs during maturation and for recognition by protein elongation factors during protein synthesis.

The contact between the D and T loops occurs via the invariant -GG- residues of the D loop and the quasi-invariant -CC- residues of the T loop. The interactions connect the first G residue with the C residue and the second G with the C residue. The G18–C55 contact is locally parallel with a trans orientation of the H bonds which are bifurcated and occur between the O4(C) and N1(G)/N2(G). The G19–56 interaction is a classical cis Watson–Crick pair, antiparallel following a rotation about its 5' phosphate. At the same time, the -GG- segment interdigitates with the R57A58 segment so that R57 of the T loop intercalates between G18 and G19 and G18 of the D loop intercalates between R57 and A58.

The only strictly invariant residue in the anticodon loop is situated where the chain reversal occurs. It is accomplished by a U turn identical to that present in the T loop. It is the phosphate group 3' of U33 (or Ψ55 in the T loop), which rotates about the P-O3' torsion angle. The U turn is stabilized by contacts with residue n + 2, A35 (or G57 in the T loop). The pyridine ring of the U turn stacks over the 5’ phosphate of residue n + 2, the O2' hydroxyl of the pyridine interacts with the N7 of the purine at n + 2, and the N3 imino proton of the uracil H bonds to an anionic phosphate oxygen (the Rp oxygen) of the 3’ phosphate of residue n + 2. Position 37 is occupied invariably by a purine (mainly A) and is often heavily modified. It stabilizes the codon–anticodon triplet interaction by stacking on the very discriminating base pair formed between the first base of the codon and the third base of the anticodon. The structural and functional basis for the chemical modifications on residue 37 are difficult to establish: stabilization by 3'-end dangling purines (as at the discriminator 73 position), blocking of the Watson–Crick sites preventing pairing with the preceding residue on the mRNA. Correlations have also been found between the chemical nature of the modifications and the hydrophobicity/hydrophilicity of the coded amino acids. The first and last bases of the anticodon loop have restricted variations, leading in more than 60% of sequences to a one H bond Y32–A38 base pair between O2(Y32) and N6(A38).

**Unusual tRNAs**

In the preceding sections, the class I tRNAs, the best known structurally, have been extensively described. Only one class II tRNA (or long arm tRNA), the *E. coli* tRNA^Ser^, is known crystallographically in a complex with its cognate aminoacyl tRNA synthetase (Biou et al., 1994). In class II tRNAs, residue 45 base-pairs with a residue in the variable helix and cannot form a triple with base pair 10–25. Also, residue 9 does not interact with 12–23 but with base pair 13–22, which is often a purine–purine G13–A22 pair, or A13–A22 (see Table 2). The R13–R22 base pair belongs to a special type of base pair, the sheared type where the Hoogsteen sites of residue 9 interact with the Watson–Crick sites of residue 13, still available since residue 22 interacts with N3 and N2 on the shallow groove side. The triple A9...A13...A22 is also observed, which implies a slight displacement of A9 so that...
its Hoogsteen sites (N7 and N6) interact with the Watson–Crick sites (N6 and N1) of A13.

The tRNA\textsuperscript{Sec} is a tRNA\textsuperscript{Ser} able to read, in an appropriate context, UGA stop codons for incorporation of selenocysteine and constitutes a variant of class II tRNAs. In archaea and eukaryotes, the acceptor helix is 9 bp long and the T helix is only 4 bp long (9/4 model). On the other hand, in prokaryotes, the acceptor helix is 8 bp long with a 5-bp T helix (8/5 model). Thus, for all phylogenetic groups, the length of the acceptor arm is 13 bp. The D helix is also changed and displays 6 bp. The T loop is shorter with the invariant -GG- residues centrally disposed.

Initiator tRNAs (Schmitt et al., 1998) also display some sequence peculiarities linked to their biological function: they must bind directly into the ribosomal P site and not in the A site where elongator tRNAs are channelled to. Thus, the first base pair of the acceptor helix is never Watson–Crick in prokaryotes, which favours recruitment of initiator factor IF-2. Prokaryotic initiator tRNAs also present an interesting reversal at the level of the 11–24 base pair where it is R11–Y24, instead of the Y11–R24 as in elongators. The T loop, normal in prokaryotes, is different in eukaryotes where positions 54 and 60 are occupied by adenines. The usual \textit{trans} Hoogsteen pair T54\textbullet{}A58 is thus replaced by the similar \textit{trans} A54\textbullet{}G58 pair between the Watson–Crick sites of A54 and the Hoogsteen sites of A58. Eubacterial and eukaryotic initiator tRNAs have an odd distribution of G–C pairs at the beginning of the acceptor helix (alternating G–C/C–G) and at the end of the anticodon helix (a run of three G–C pairs). Eukaryotic initiator tRNAs have other unusual features, like the unique sugar modification on residue 64, a phosphoribosyl group on the 2′ hydroxyl, a modification which hinders binding of the elongation factor EF-1α and, thus, access to the A site.

The structural diversity of mitochondrial tRNAs is enormous. In some of them the T hairpin is missing and replaced by a single-stranded segment, while in others it is the D hairpin. In the remaining stems and loops, there is a systematic loss of base invariance or semi-invariance compared to cytoplasmic tRNAs. In the absence of a reference crystal structure, it is difficult to discuss the structural aspects of such truncated and functional tRNAs.

### Summary

tRNA constitutes a paradigm for RNA structure and folding. It possesses a well-defined consensus secondary structure, the cloverleaf, which folds into a common three-dimensional architecture with a characteristic L shape. The three-dimensional fold can be divided into two domains, similarly built of two coaxial helical stems. The three-dimensional architecture is kept in place by loop–loop tertiary contacts between the thymine and dihydrouridine loops, together with tertiary contacts involving the single-stranded junctions linking the helical domains and the deep groove of one helix, the dihydrouridine helix. The latter contacts involve non-Watson–Crick pairings and formation of base triples.

tRNA structure displays several elements of the RNA folding logic: coaxial stacking of contiguous helices, formation of base triples in a RNA helix deep groove, loop–loop interactions with Watson–Crick and non-Watson–Crick pairs, the U-turn motif for hairpin formation, and the extensive use of additional hydrogen bonding involving the ribose hydroxyl O2′, the anionic phosphate oxygen atoms, and some polar C–H groups. The modular view of RNA architecture, based on a hierarchical assembly of recurrent RNA motifs, first glimpsed at in the tRNA structure, has been corroborated by the recent crystal structures of larger RNA domains and ribozymes.

The increasing number of crystal structures has revealed an amazing and subtle variability in precise atomic contacts. However, the microheterogeneities in the specific atomic contacts between residues important for the stability of the global tertiary fold maintain identical overall topological arrangements. These structural constraints lead to base conservations or base covariations in sequence comparisons and alignments. Comparisons between RNA molecules show that topologically and functionally distinct molecules share quasi-identical three-dimensional motifs which display clear signatures in sequence conservation and variability.

### References


---

### Table 2

Table of covariations between the secondary base pair 13–22 in the D helix and residue 9 in class II tRNAs (analysis is made on 187 sequences of class II elongator tRNAs)

<table>
<thead>
<tr>
<th>Base pair 12–23</th>
<th>A</th>
<th>U</th>
<th>G</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>A\textbullet{}A</td>
<td>32</td>
<td>–</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>G\textbullet{}A</td>
<td>–</td>
<td>4</td>
<td>115</td>
<td>–</td>
</tr>
<tr>
<td>G\textbullet{}U</td>
<td>–</td>
<td>–</td>
<td>11</td>
<td>–</td>
</tr>
<tr>
<td>Others</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>


**Further Reading**


