A Common Motif Organizes the Structure of Multi-helix Loops in 16 S and 23 S Ribosomal RNAs

Neocles B. Leontis and Eric Westhof

Chemistry Department
Bowling Green State University, Bowling Green OH 43403, USA

Institut de Biologie Moléculaire et Cellulaire du CNRS, Modélisations et Simulations des Acides Nucéiques, UPR 9002 15 rue René Descartes F-67084 Strasbourg Cedex France

Phylogenetic and chemical probing data indicate that a modular RNA motif, common to loop E of eucaryotic 5 S ribosomal RNA (rRNA) and the α-sarcin/ricin loop of 23 S rRNA, organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. The motif occurs in the 3' domain of 16 S rRNA at positions 1345-1350/1372-1376 (Escherichia coli numbering), within the three-way junction loop, which binds ribosomal protein S7, and which contains nucleotides that help to form the binding site for P-site tRNA in the ribosome. The motif also helps to structure a three-way junction within domain I of 23 S, which contains many universally conserved bases and which lies close in the primary and secondary structure to the binding site of r-protein L24. Several other highly conserved hairpin, internal, and multi-helix loops in 16 S and 23 S rRNA contain the motif, including the core junction loop of 23 S and helix 27 in the core of 16 S rRNA. Sequence conservation and range of variation in bacteria, archaea, and eucaryotes as well as chemical probing and cross-linking data, provide support for the recurrent and autonomous existence of the motif in ribosomal RNAs. Besides its presence in the hairpin ribozyme, the loop E motif is also apparent in helix P10 of bacterial RNase P, in domain P7 of one sub-group of group I introns, and in domain 3 of one subgroup of group II introns.

Introduction

Single-stranded RNA molecules fold on themselves to form short double helices consisting of classical Watson-Crick base-pairs frequently interrupted by stretches of “mismatched” bases, comprising so-called internal loops, hairpin loops, and multi-helix junction loops (Fresco et al., 1960). Universally conserved, nominally unpaired nucleotides often occur within these loops. That many of these regions are in fact highly structured was first clearly indicated by chemical probing experiments of 16 S rRNA (Moazed et al., 1986).

Phylogenetically based covariation analysis is the most reliable means for determining secondary structure in homologous RNA molecules (Michel et al., 1982; Woese & Pace, 1993). It can also be used to identify nucleotides involved in tertiary interactions, when these co-vary in a Watson-Crick pairing fashion, but not when these involve universally conserved, single-stranded regions of the already deduced secondary structure (Woese & Gutell, 1989; Larsen, 1992). The modular approach to sequence analysis of RNA is based on the assumption that recurrent, conserved elements of primary and secondary structure share common three-dimensional (3D) structures and therefore constitute building blocks for organizing the tertiary structure of RNA (Michel & Westhof, 1990). Such elements must be viewed as structural units or motifs, since they comprise highly conserved nucleotides, which are paired non-canonically and which therefore do not co-vary in a Watson–Crick fashion. Identification of such motifs within internal loops, followed by structural characterisation using NMR and X-ray crystallography, advances our ability to model the secondary and tertiary architecture of native RNA molecules, by elaborating a “structural vocabulary” applicable in evaluating possible interactions in any structured RNA.

Keywords: RNA; multi-helix loops; recurrent motif; phylogenetic analysis

Abbreviations used: rRNA, ribosomal RNA; r-protein, ribosomal protein; bp, base-pair; PSTV, potato spindle tuber viroid; R, purine; Y, pyrimidine; CMCT, 1-cyclohexyl-3-(2N-morpholinethyl)-carbodiimide-metho-p-toluene sulphonate.

E-mail addresses of the corresponding authors: neocles@rosalind.bgsu.edu and westhof@ibmc.u-strasbg.fr
Potential structural motifs are easiest to identify when the conserved sets of nucleotides occur as internal loops, bracketed on either side by Watson–Crick helices or hairpin loops. Such is the case of the loop E motif first identified in 1985 as a distinctive RNA structural motif when it was found that very similar, conserved, asymmetric internal loops in eucaryotic 5 S rRNA and potato spindle tuber viroid (PSTV) shared an exceptional propensity for UV crosslinking between specific guanosine and uridine residues, as shown in Figure 1A (Branch et al., 1985). The motif was also found in a conserved hairpin loop in domain VI of 23 S-like rRNA molecules (residues 2654–2666 of *Escherichia coli* 23 S and 4218–4330 of rat 28 S rRNA), which is the site of specific RNA modification by the cytotoxic proteins α-sarcin and ricin (Endo et al., 1993) and is directly involved in binding of elongation factors by the ribosome (Moazed et al., 1988). It has been pointed out that the element is likely to be widespread (Shen et al., 1995) and that it may exist in other hairpin/stem loops in 23 S rRNA (Wimberly et al., 1993; Wimberly, 1994). UV crosslinking showed that the

**Figure 1.** A, Molecules in which the “loop E motif” has been established by NMR or UV cross-linking. The conserved nucleotides of the motif are boxed. Asterisks designate the G and U residues cross-linked upon irradiation with UV light. From left to right: loop E from eucaryotic 5 S rRNA (305 sequences superimposed on that of *X. laevis*). Loop E from Euryarchaeal 5 S rRNA (36 sequences superimposed on that of *Methanococcus vannielii*). The 5 S sequences are numbered according to the 5 S of *X. laevis*. sarcin/ricin stem from 23 S rRNA (sequence variations from all three phylogenetic groups superimposed on *E. coli* sequence and numbering). Conserved core from PSTV (Branch et al., 1985). Loop B from the hairpin ribozyme (Butcher & Burke, 1994a). Red letters: conserved nucleotides. Green: base-pairs showing Watson–Crick variation. Blue: residues within motif showing variation (for sarcin/ricin stem see Table 1; for euryarchaeal 5 S rRNA: 13 A–A, 4 G–A, 5 A–G, 15 U–A). Black: Other variable positions. B, Loop E motifs in internal or hairpin loops of 23 S rRNA and 16 S rRNA. From left to right: 23 S rRNA helix 11; 23 S rRNA internal loop between helices 31a and 32; 16 S rRNA helix 27 in core region joining the 5' and the central domains of the molecule. Residues are colour-coded as in A. Variations observed in three main phylogenetic groups superimposed on *E. coli* sequence and numbering system.
motif exists in loop B of the hairpin ribozyme from tobacco ringspot virus (Butcher & Burke, 1994a). Crosslinked molecules retain enzymatic activity (Butcher & Burke, 1994a; Burke et al., 1996). Here we present evidence indicating that the motif is not limited to hairpin stems or internal loops, but also occurs in multi-helix junctions at critical sites in the large ribosomal RNAs.

Results

Structural description of the motif

Both the sarcin/ricin loop from 23 S and the E-loop from 5 S rRNA have been studied as oligonucleotide fragments by high-resolution NMR, from which very similar three-dimensional structures were derived for the conserved nucleotides of the motif (Wimberly et al., 1993; Szewczak & Moore, 1995). The structures also agree with the extensive chemical probing data available for X. laevis 5 S rRNA (Romaniuk et al., 1988; Westhof et al., 1989; Romby et al., 1990). The salient structural features identified by NMR are shown in Figure 2. They are: (1) a sheared A-G pair; (2) a trans-Hoogsteen U-A pair; (3) a bulged G positioned in the deep groove; and (4) a trans (locally parallel)-Hoogsteen-Hoogsteen A-A pair (boxed residues in Figure 1). The sheared pairing of the conserved A-G pair (A77-G99 in the Xenopus sequence, A106-G130 in the universal alignment) involves hydrogen bonding between AN6 and GN3 and between AN7 and GN2. This leaves AN1 very exposed and reactive towards DMS, whereas GN1 is partly exposed and only reactive under semi-denaturing conditions in X. laevis 5 S rRNA. Neither AN7 nor GN7 is reactive under native conditions. In the conserved, trans-Hoogsteen U76-A100 (Xenopus numbering, U104-A132 universal alignment), UN3 is H-bonded to AN7 leaving the Watson–Crick site (N1) of A100 exposed and reactive while UN3 and AN7 are protected in this pair. The N1 of the bulged G75 (G103) is likely to be H-bonded to the phosphate oxygen of A101 (A132), consistent with observed lack of reactivity toward CMCT. Its N7, however, is completely exposed, as shown by the fact that it is the most reactive base in 5 S rRNA towards a sterically bulky nickel-organic reagent.
(Zheng et al., 1998). The direction of the ribose phosphate chain is locally reversed at A74 (A102) resulting in a symmetrical pairing between A74 and A101 (A133), which involves H-bonding of N7 and N6 of one adenine with N6 and N7 of the other. Both bases are found to be reactive at N1 and unreactive at N7 in Xenopus 5 S rRNA.

On the basis of these structural features, one expects to observe the following sequence variation where the motif is evolutionarily conserved (Leontis & Westhof, 1998): the sheared A-G is expected to covary with a sheared A-A (Cate et al., 1996a) but neither with the non-isosteric opposite sheared G-A nor with the G-G opposition, which would require another pairing geometry. The U-A pairing should be strictly conserved. On the other hand, significant variation is expected at the position of the locally parallel A-A pair. One expects to see A-G as well as G-A Hoogsteen-Hoogsteen (but no G-G oppositions again). All R-Y, Y-R, and Y-Y combinations can be accommodated by adopting a \textit{trans}-Watson–Crick geometry, while maintaining the locally parallel strand orientation.

\textbf{Sequence variation in established loop E motifs}

To establish criteria for evaluating candidate loop E motifs more precisely than by consensus sequence comparisons, we examined the sequence variations in the best-characterized molecules known to harbour the motif, the sarcin/ricin stem of 23 S rRNA and loop E of 5 S rRNA (Figure 1A). We found that the motif exists not only in eucaryotic 5 S rRNA but also in the 5 S of the Euryarchaeota subgroup of Archaea. The sarcin/ricin stem exists in the 23 S-like RNA of all three phylogenetic groups and of mitochondria. The bulged G, \textit{trans}-Hoogsteen U-A and sheared A-G pairs are 100\% conserved in both 5 S groups. The only exceptions in the sarcin/ricin stem are two mitochondrial sequences, which have A-A in place of A-G. As discussed above, A-A can also form a sheared base-pair isosteric to a sheared A-G bp. The A-G and U-A pairings are also conserved in the viroid and hairpin-ribozyme versions of the motif. The bulged base is a C in the viroids (Keese & Symons, 1985) and a U in the hairpin ribozyme. The symmetrical A-A pairing is conserved in all eucaryotic 5 S rRNAs. In 5 S of Euryarchaeota, however, A-A, A-G, G-A and U-A are also observed for this pairing. A-A is the most common opposition at the equivalent positions of eucaryotic and archaean sarcin/ricin stems, while A-C is the most common in bacterial and mitochondrial sarcin/ricin stems. The variations at this position in sarcin/ricin loops are summarized in Table 1. The variations observed in 5 S loop E and sarcin/ricin loops fall into three groups: (1) R-R (A-A, A-G, G-A, but never G-G), (2) R-Y and Y-R (A-Y and Y-A, but rarely G-Y or Y-G), and (3) Y-Y (U-U, U-C). All observed R-R pairings are compatible with \textit{trans}-Hoogsteen-Hoogsteen, as seen in the NMR structure for A-A. The other two groups are both compatible with \textit{trans} Watson–Crick/Watson–Crick pairing. Next to the symmetrical pair, another non-canonical pairing is found in both 5 S loop E and the sarcin/ricin stem. In some phylogenetic groups only Y-Y pairings are found here, while in others, wobble pairs (including A-C and C-A) and Watson–Crick pairs also occur. The next base-pair is usually Watson–Crick, but in some groups Y-Y and wobble pairs are among the variations. A Watson–Crick base-pair, which in the sarcin/ricin site is a conserved C-G in all phylogenetic groups, always follows the sheared A-G pair. Thus, Watson–Crick pairs flank both sides of the motif.

\begin{table}[h]
\centering
\caption{Sequence variation at the parallel base-pairs (positions 2654 and 2666 in the \textit{E. coli} sequence) in the loop E motif of sarcin/ricin loop of 23 S rRNA in three phylogenetic groups and mitochondria.}
\begin{tabular}{|c|c|c|c|c|}
\hline
Res. & A & C & G & U \\
\hline
\textbf{Table 1. Sequence variation at the parallel base-pairs (positions 2654 and 2666 in the \textit{E. coli} sequence) in the loop E motif of sarcin/ricin loop of 23 S rRNA in three phylogenetic groups and mitochondria.}
\end{tabular}
\begin{tabular}{|c|c|c|c|c|}
\hline
Base & 2654 & 2666 \\
\hline
\textbf{Res.} & A & C & G & U \\
\hline
A & 0 & 57 & 166 & 2 & 10 & 1 \\
& 4 & 23 & 108 & 1 & 10 & 0 \\
C & 3 & 1 & & & & \\
G & & 2 & 1 & & & \\
U & 1 & 16 & 0 & 2 & & \\
\hline
\end{tabular}
\begin{tabular}{l}
\textbf{Note the predominant pairing is A-C in bacteria and mitochondria but A-A in eucaryotes and Archaea and that most variants are related to the dominant pair within a given phylogenetic group by a single mutation. (Number of sequences: 180, bacterial; 64, eucaryotes; 24, archaean; and 144, mitochondrial.)}
\end{tabular}
\end{table}

To summarize, the observed sequence variation agrees with stereochemical expectations based on the 3D structures. The U-A is universally conserved, owing to the unique \textit{trans}-Hoogsteen pairing. The bulged base is not restricted to G. Any pairing compatible with a \textit{trans}-parallel orientation of the bases is allowed immediately 5\' to the bulged base. No insertions are observed within the region bracketed by the parallel pair and the sheared pair. In addition, it is observed that the first Watson–Crick pairing occurs at a variable distance from the parallel base-pair; base insertions or deletions may also occur here. The data are insufficient yet for understanding the possible structural constraints underlying these sequence variations, although they do point to a role akin to a flexible joint (see below).
hairpin ribozyme, the first Watson–Crick pairing is several non-canonical pairings removed from the loop E motif.

**Loop E motifs in internal and hairpin loops of 23 S rRNA**

Three hairpin or internal loops in addition to the sarcin/ricin hairpin stem were identified in 23 S rRNA as loop E sites which meet these criteria. These occur in the highly conserved helices 11 and 13 in domain I and in the internal loop between helices 31a and 32 in domain II, in the 5’-most part of the 23 S molecule. (The positions of the proposed loop E sites discussed in this paper within the conventionally drawn secondary structures of 16 S and 23 S rRNA are shown in Figure 3). The possible existence of loop E motifs in helices 11, 13, and in the core of 23 S (see below) was previously suggested, based on consensus sequence comparisons (Wimberly et al., 1993; Wimberly, 1994). The E. coli versions of these are shown in Figure 1B, on which sequence conservation and variations observed in all three phylogenetic groups are superimposed. Very few deviations from the criteria defined above are observed and these are restricted to specific phylogenetic groups. These examples show that the motif may exist in either orientation closing a hairpin loop, which gives additional support for its autonomous existence. Footprinting data are also consistent with existence of the motif at these positions, as summarized in Table 2 (Egebjerg et al., 1987, 1990).

**A loop-E motif in the core of 16 S rRNA**

Universally conserved nucleotides 889-892/907-909 comprise a loop E motif in a hairpin stem (helix 27) in the core of 16 S rRNA (rightmost panel, Figure 1B). The sheared pair is A-A, with a small number of A-G in eucaryotes. The U-A is invariant in all groups. The bulged G shows some variation (A, C, or U) and the symmetrical A-A, a very small number of AG, GA and AU variants. Watson–Crick or wobble pairs separate the motif from a conserved GCAA hairpin loop; the pair adjacent to the sheared A-R pair is a conserved Y-R pairing (C-G in all Archaea, U-G in eucaryotes, C-G, U-G or C-A, in bacteria). Pairing of positions 912–910 with 885–887 was recently proposed on the basis of site-directed mutagenesis, completing helix 27 (Lodmell et al., 1995). This leaves base 888 unpaired, but does not hinder formation of the motif. Helix 27 is part of a complex junction, which includes two more helical elements and a pseudo-knot to organize the 5’ and central domains of 16 S RNA (Masquida et al., 1997). The nucleotides of the loop E motif in helix 27 are protected from chemical probes both in the naked 16 S rRNA and in the 30 S subunits (Moazed et al., 1986). During the first stages of *in vitro* assembly of the RNA with r-proteins, residues G886-G890, on the 5’ side of the Loop E motif become more reactive to chemical probes, indicating that a protein-mediated conformational change probably occurs (Stern et al., 1988b). It was recently proposed that a conformational switch occurs during translation in which nucleotides 912–910 alternately pair with 885-887, as in Figure 1B, or with 888-890 (Lodmell & Dahlberg, 1997). In the latter conformation the loop E motif would be disrupted by pairing of the bulged G890 with C910 and of the reversed A889 with U911. However, in isolated wild-type ribosomes, the predominant conformation is that allowing formation of the loop E motif (Lodmell & Dahlberg, 1997). Further, mutations favouring that conformation are always less severe than those stabilizing the alternative conformation. It can be added that helix H27 can be drawn with a loop E motif in 16 S rRNA from mitochondria as well. Interestingly, a conformational switch between a structure involving a GNRA tetraloop and one with a loop E motif occurs during the pathway of the PSTV viroid processing (Baumstark et al., 1997).

**Three-way junctions organized by loop E motifs**

Three-way junctions occur widely in stable RNA including ribosomal RNAs (5 S, 16 S and 23 S), ribozymes, and viral RNAs. RNA three-way junctions typically have a number of unpaired or non-canonically paired bases in one or more strands linking and organizing the helices in space, as exemplified in the structure of the hammerhead ribozyme (Pley et al., 1994). One of several three-way junctions which are found in the 3’ major domain of 16 S RNA binds S7 (Chiaruttini et al., 1989; Dragon & Brakier Gingras, 1993) and forms part of the P-site for tRNA binding (Moazed & Noller, 1986). This junction is usually drawn in secondary structure maps with 19 unpaired nucleotides, many of which are universally conserved or show very restricted variation (Figure 3). In fact, nucleotides 1345–1349 and 1373–1376 form a loop E motif bridging helices 29 and 43, which together with helix 28 comprise the junction (Figure 4A). The only variable position within the motif is the parallel pairing between positions 1346 and 1375. This is a highly conserved A-A in bacteria and in Archae (no exceptions), while mainly A-U in eucaryotes with variants A-C, A-A, C-U, G-U, U-A, C-A, and U-C, consistent with maintenance of the *trans-trans* pairing at this position. A conserved U-U (Y-U in eucaryotes) can form between positions 1345 and 1376. Usually U1345 is shown Watson–Crick paired with conserved A938 in secondary structures of 16 S. Watson–Crick bp 1350–1372 and 1344–939 flank the motif. The exact position of the three-way junction branch point depends on how U1345 is paired, but in either case the branching does not disrupt the motif. Chemical probing of the Watson–Crick sites of nucleotides in naked 16S rRNA showed that A1346, A1349, A1374, A1375 within the motif are reactive,
Figure 3. Conventional secondary structures of 16S rRNA and domains I and II of 23S rRNA showing locations of proposed loop E sites. 16S rRNA sites: helix 27 in the core region (see also Figure 1B) and the three-way junction comprising helices 28, 29, and 43 in the 3'-major domain (see also Figure 4A). 23S rRNA sites: internal loops at helices 11, 13, 31a and 32 (see also Figure 1B); three-way junction formed by helices 3, 4, and 23 (see also Figure 4B); multi-helix loops adjacent to helices 21 and 36 (see also Figure 5B). Not shown in this figure: the loop E motifs located in the sarcin/ricin stem in domain VI (see Figure 1A) and in the core region of 23S rRNA (see Figure 5A).
whereas bulged G1347, U1348 and G1373 are not (Moazed et al., 1986), in complete agreement with the probing data of the corresponding positions of 5 S loop E discussed above (see Table 2). In the 30 S particle, these bases become protected from reactive probes, due to the binding of protein S7 (see below).

**A three-way junction in domain I of 23 S RNA**

The three-way junction formed by helices 3, 4, and 23 in domain I of 23 S rRNA also contains a loop E motif. As in the 16 S junction, the motif bridges two of the helical arms (helices 3 and 23) creating potential continuous stacking between them (Figure 4B). The branch point of the junction occurs at the equivalent position: two pairings from the bulged base or one after the A-A pair. However, the two junctions differ in the direction of branching. In the 23 S three-way junction, branching occurs in the strand containing the unpaired G, while it was the opposite in the 16 S junction, stressing again the autonomous existence of the motif. Very little sequence variation is observed across three phylogenetic groups. Chemical probing of naked 23 S rRNA and 50 S subunits supports the existence of the junction (Table 2) (Egebjerg et al., 1990). A stretch of five other nominally unpaired bases within this junction loop is also universally conserved (G446-G450). The loop E motif in this junction may, as in the 16 S three-way junction, organize highly conserved single-stranded residues for specific interactions with other RNA sequences.

### Table 2. Chemical probing data of proven and proposed loop E motifs

<table>
<thead>
<tr>
<th>Base-pair</th>
<th>Base-atom</th>
<th>A-N1</th>
<th>A-N7</th>
<th>G-N1</th>
<th>G-N7</th>
<th>U-A trans-Hoogsteen</th>
<th>Bulged G</th>
<th>Parallel</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 S rRNA</td>
<td>R</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>R</td>
<td>U</td>
<td>R</td>
</tr>
<tr>
<td>23 S S/R</td>
<td>r</td>
<td>U</td>
<td>U</td>
<td>r</td>
<td>U</td>
<td>R</td>
<td>U</td>
<td>r</td>
</tr>
<tr>
<td>23 S helix 11</td>
<td>r</td>
<td>U</td>
<td>U</td>
<td>r</td>
<td>U</td>
<td>R</td>
<td>U</td>
<td>r</td>
</tr>
<tr>
<td>23 S helix 13</td>
<td>r</td>
<td>U</td>
<td>U</td>
<td>r</td>
<td>U</td>
<td>R</td>
<td>U</td>
<td>r</td>
</tr>
<tr>
<td>23 S loop 31a/32</td>
<td>r</td>
<td>U</td>
<td>U</td>
<td>r</td>
<td>U</td>
<td>R</td>
<td>U</td>
<td>r</td>
</tr>
<tr>
<td>16 S 3WJ 28/29/43</td>
<td>R</td>
<td>U</td>
<td>U</td>
<td>R</td>
<td>r</td>
<td>R</td>
<td>U</td>
<td>r</td>
</tr>
<tr>
<td>23 S 3WJ 3/4/23</td>
<td>r</td>
<td>U</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>R</td>
<td>U</td>
<td>r</td>
</tr>
<tr>
<td>23 S core</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>U</td>
<td>r</td>
<td>U</td>
<td>r</td>
</tr>
<tr>
<td>23 S MWJ 21</td>
<td>R</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>R</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>23 S MWJ 36</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>r</td>
<td>U</td>
<td>U</td>
<td>U</td>
<td>r</td>
</tr>
</tbody>
</table>

R, very reactive; r, weakly reactive; U, unreactive. Blanks indicate undetermined. 3WJ stands for three-way junction and MWJ for multi-way junction. The numbers indicate helices adjacent to the motifs identified (see also Figures 3 to 5).

*a* The bulged base is A in the probed sequence.

*b* The pair is G-A in the probed sequence. Probing data for 5 S rRNA are from Westhof et al. (1989), for 16 S from Moazed et al. (1986) and Baudin et al. (1987), and for 23 S from Egebjerg et al. (1990).

**A loop E motif in the core of 23 S rRNA**

The core of 23 S rRNA is closed by a helix formed by the 5′ and 3′ termini of the molecule; it is usually drawn in a circular format with the six domains radiating outward (Airport Terminal model, see Figure 5A). Covariation analysis detected Watson–Crick interactions between positions 1262–2017, 1269–2011, and 1270–2010 in 23 S rRNA (Gutell & Woese, 1990), suggesting that an irregular helix forms between nucleotides 1262 and 1270, which link domains II and III, and residues 2010 and 2017, which link domains IV and V (Haselman et al., 1989). In fact these nucleotides comprise a loop E motif, highly conserved across all three phylogenetic groups and also found in the highly reduced mitochondrial sequences (Figure 5A). The only sequence variations occur at the symmetrical A-A parallel pairing which is conserved in bacterial and archaeal sequences but displays three C-A and one A-G variants in eucaryotes. The folding of the motif creates a stretch of continuous base-pairing between domains II and IV. Disruption of either Watson–Crick basepair 1262–2017 or 1269–2011, flanking the motif was shown by site-directed mutagenesis to impair the ability of the 23 S to engage in protein synthesis (Aagaard & Douthwaite, 1994). All adenosine residues in the motif were found to be slightly reactive with DMS in the naked 23 S rRNA. Double-helix specific CV nuclease cut strongly at the Watson–Crick sites flanking the motif and weakly within it, giving further proof for its existence. The motif is completely protected from chemical and enzymatic probes in the 50 S ribosome, consistent with multiple RNA-RNA or RNA-protein interactions in the folded RNA.

**Loop E motifs in multi-helix loops in domain I and II of 23 S**

A loop E motif conserved in all three phylogenetic groups is found at positions 1262–1270.
821 in *E. coli* 23S at the base of helix 36, which connects the seven-way junction loop of domain II to the rest of the molecule. The motif creates potential stacking between helix 36 and either helix 37 or helix 45. The branch point occurs next to the parallel base-pairs and does not interrupt the motif. The parallel pair is predominantly A-A in eucaryotic 23S, G-A in bacteria, and mixed in Archaea. Another loop E motif is found in the variable junction in domain I adjacent to helix 21. Sequence variations for both motifs are superimposed on the *E. coli* sequence (Figure 5B) and chemical probing data are summarized in Table 2 (Egebjerg et al., 1990).

**Loop E motifs in catalytic RNAs**

Besides ribosomal RNAs and the other known examples, PSTV and the hairpin ribozyme, the loop E motif probably exists in other structured or catalytic RNAs. The loop E motif has thus been identified in a subfamily of type B RNase P sequences, in helix P10.1, where it replaces a regular Watson–Crick helix (Massire et al., 1998). The loop E motif is also apparent in the P7.1–P7.2 subdomain of subgroup IA2 of group I introns (Michel & Westhof, 1990) and in domain 3 of one class of group II introns (Michel et al., 1989). These motifs are illustrated in Figure 6.
Discussion

Evolutionary implications

Why are there so many loop E sites in the large RNAs, and why are these so highly conserved? The fact that they are conserved in all phylogenetic groups indicates that loop E motifs were present in the Ur-ribosome. Their occurrence at sites involved in binding primary proteins and RNA sequences which define the domain structure of 16S and 23S rRNA suggests that they evolved within internal and junction loops due to their special ability to participate in the organization and structuring of multiple junctions as well as in RNA-RNA and RNA-protein interactions. The fact that the Watson–Crick sites of all the adenosine residues in the motif are available for interaction is noteworthy and parallels the situation found in the GNRA loop motifs and their receptors (Jaeger et al., 1994; Cate et al., 1996b).

Protein binding by loop E motifs

It is well known that loop E of eucaryal 5S rRNA forms part of the binding site for ribosomal protein L5 (Allison et al., 1991) and transcription factor TFIIIA (Romaniuk, 1989), while elongation factors and cytotoxins interact with the sarcin/ricin loop of 23S rRNA. The bulged G must be present.
for specific recognition and modification of the RNA motif by \( \alpha \)-sarcin (Wool et al., 1992). The interaction of elongation factors EF-G and EF-Tu with the sarcin/ricin loop of 23 S rRNA was demonstrated using chemical footprinting (Moazed et al., 1988). Protection was observed upon binding of EF-Tu or EF-G within the motif at the bulged G2655 and at A2665, which is \( \text{trans} \)-Hoogsteen paired to U2656. In the absence of factors, the Watson–Crick faces of A2665 and G2655, but not A2654 (which probably interacts with C2666 in the \( \text{E. coli} \) sequence) are reactive.

S7 is the sole primary binding protein in the 3' domain of 16 S. Foot-printing and cross-linking experiments show that nucleotides comprising the loop E motif of the 16 S three-way junction in Figure 4A form part of the binding site of this protein (Wower & Brimacombe, 1983; Powers et al., 1988). Concise footprints obtained using solvent-generated hydroxyl radicals show that S7 protects nucleotides on the 3' side of the loop E motif of the three-way junction, and in helices 29 and 43, which are bridged by the motif (Powers & Noller, 1995). The 5' staggering of six to seven nucleotides observed in the footprint is consistent with axial binding to the deep groove of the RNA. The remaining unpaired bases of the three-way junction are lightly protected from hydroxyl radicals by S7. However, some nucleotides become more accessible upon protein binding. Most notably, U1381, adjacent to the loop E motif becomes reactive to kethoxal upon binding by S7 and is strongly reactive in 30 S subunits. It is protected by tRNA binding to ribosomes (Moazed & Noller, 1986; Moazed et al., 1986; Powers et al., 1988). These data suggest that the loop E motif of this three-way junction may play two roles: to provide a binding site for S7, which helps to fine-tune the local RNA structure, and to interact with other nucleotides in the junction or in the nearby four-way junction to help form the binding site for the tRNA anti-codon loop.

The structure of S7, recently solved using X-ray crystallography (Hosaka et al., 1997; Wimberly et al., 1997), indicates extensive protein surface for interaction with double-helical RNA.

The majority of loop E motifs found in 23 S rRNA occur in the 5' half of the molecule. In fact, four occur in domain I, the first to be transcribed and the first to fold (Tumminia et al., 1994). Chemical and enzymatic probing of 23 S rRNA, isolated as well bound to r-proteins in the ribosome, indicates that domain I is more highly structured and compact than the other domains of the molecule (Egebjerg et al., 1987, 1990; Döring et al., 1991). Addition of a single, relatively small ribosomal protein, L24, the only primary binding protein which binds to domain I, completes the folding of the domain and initiates assembly of the 50 S subunit (Cabezon et al., 1977; Spillman & Nierhaus, 1978; Tumminia et al., 1994). L24 has been footprinted to two sites (Egebjerg et al., 1987), one of which is adjacent in the primary and secondary structure to the three-way junction which contains the loop E motif, and which may therefore play a role in helping to structure the neighbouring RNA for interaction with L24.

Loop E and RNA-RNA interactions

In the hairpin ribozyme, it has been shown that the internal loop containing the Loop E motif must interact with the loop containing the cleavable phosphate for catalysis to occur (Burke, 1996; Burke et al., 1996). A recent model with side-by-side interaction of the two internal loops has been proposed (Earnshaw et al., 1997).

Protection by ribosomal proteins S5 and S12 of bases in the multi-branch loop at the junction of the three major domains of 16 S rRNA, including those belonging to the loop E motif that we have identified, was observed using base-specific probes (Stern et al., 1988a), but not using hydroxyl radicals generated with the \( \text{Fe}^{2+}/\text{EDTA} \) reagent (Powers & Noller, 1995). Hydroxyl radicals react primarily with the ribose moieties of the backbone, suggesting that the base modifications were due to conformational changes induced by binding of these proteins to other regions of the RNA (notably the nearby 530 stem-loop region). Thus, it may be that the loop E motif in the core of the 16 S rRNA...
is involved in RNA-RNA rather than RNA-protein interactions. In support of this, two distinct RNA-RNA crosslinks connecting G894 (in the duplex just above the sheared A-A bp of the motif) with bases U244 in the early-folding 5′ domain and A1468 in the 3′-terminal domain indicate the central role of the core in organizing the domain structure of 16S rRNA (Wilms et al., 1997). Crosslinking is also observed between U1345, at the base of the loop E motif in the 16S three-way junction, and unpaired C934, in the same junction (Wilms et al., 1997). Unpaired bases in the junction, U1376 and C1378 have been crosslinked directly to ribosome-bound tRNA (Döring et al., 1994). Taken together with footprinting of U1381 by bound tRNA, discussed above, the evidence is strong for a role for the loop E motif in structuring this three-way junction for RNA-RNA interaction with tRNA.

Conclusions

Comparative sequence analysis and chemical probing data show that the loop E motif is a recurrent, autonomous structural element that occurs not only within RNA “internal” but also around “multi-helix” loops, which points to crucial and multiple roles in structuring and organizing the junctions. Although loop E motifs are often footprinted by proteins, their involvement as RNA-RNA anchors cannot be excluded, as shown by the example of the hairpin ribozyme (Butcher & Burke, 1994b; Earnshaw et al., 1997). The specificity and the structure of the binding motif, of either protein or RNA are still unknown.

Methods

Sequence analysis was carried using the COSEQ program (Massire & Westhof, unpublished). We made extensive use of the secondary structures of the 16S, 23S and group I introns posted on the web (http://pundit.colorado.edu:8080/RNA) (Gutell et al., 1993; Damberger & Gutell, 1994; Gutell, 1994). 5S rRNA sequences were obtained from the Berlin database. Nucleotides are numbered consecutively according to the Xenopus laevis sequence (Specht et al., 1997). Large (23S-like) and small (16S-like) ribosomal subunit RNA sequences were obtained from the Antwerp database and the Ribosomal Database Project (Maidak et al., 1997; van de Peer et al., 1997a,b). Sequences belonging to the three main phylogenetic groups (Woese et al., 1990), eucaryotes (1581 16S rRNA sequences), Archaea (379 16S rRNA sequences) and bacteria (1488 16S and 180 23S rRNA sequences) were downloaded and analysed separately. At positions harbouring likely loop E motifs, the alignments were checked for insertions or deletions as an initial screen for the existence of the motif at a given site. When the absence of insertions (or deletions) was confirmed, covariations between the pairings were analysed using COSEQ. Numbering of nucleotides and helices for 16S and 23S rRNAs are given according to the Escherichia coli sequences (Brimacombe, 1991).

Acknowledgements

We are grateful to Christian Massire for help with the use of the COSEQ program program. N.L. was a Fogarty Senior International Fellow, supported by NIH grants (1-F06-TW02251-01) and 1R15-GM/ODS5898-01 (NBL), ACS-FRF grant no. 31427-B4 and by a ‘Poste Rouge’ from the CNRS. E.W. acknowledges support from the Institut Universitaire de France.

References


tems S4, S5, S7, S8, S11, S12 and S18 to domains 1 and 2 of 16 S rRNA in the *Escherichia coli* 30 S particle. *Biochimie*, 71, 839–852.


