Derivation of the Three-dimensional Architecture of Bacterial Ribonuclease P RNAs from Comparative Sequence Analysis

Christian Massire, Luc Jaeger and Eric Westhof*

Institut de Biologie Moléculaire et Cellulaire du CNRS
UPR 9002, 15 rue Descartes
67084 Strasbourg, France

The secondary structure of bacterial RNase P RNA, a ribozyme responsible for the maturation of the 5' end of tRNAs, is well established on the basis of sequence comparison analysis. RNase P RNA secondary structures fall into two types, A and B, which share a common core formed by the assembly of two main folding domains, but differ in their peripheral elements.

A revised alignment of 137 available sequences reveals new covariations allowing for the refinement of both types of secondary structures. Phylogenetic evidence is thus provided for the extension of stems P11, P14, P19, P10.1 and P15.1 through further canonical base-pairs or GA...GA mismatches. These refinements led in turn to a new organization of the catalytic core, with coaxial stackings of helices P2 and P19 as well as P1 and P4. New inter-domain tertiary interactions involve loop L9 and helix P1 and loop L8 with helix P4. These features were incorporated into atomic-scale 3D models of RNase P RNA for representatives of each structural type, namely Escherichia coli and Bacillus subtilis. In each model, the juxtaposition of the core helices creates a cradle onto which the pre-tRNA substrate binds with most evolutionarily conserved residues converging towards the cleavage site. The inner cores of both types are stabilized similarly, albeit by different peripheral elements, emphasizing the modular and hierarchical organisation of the architecture of RNase P RNAs. Similarities are thus apparent between the type A modules, P16/P17/P6 and P13/P14, and their type B analogs, P5.1/P15.1 and P10.1/P10.1a, respectively. Other noteworthy features of these models include compactness and good agreement with published crosslinking data.

© 1998 Academic Press Limited

Keywords: ribonuclease P RNA; ribozyme; RNA folding; RNA-RNA interactions; RNA modelling

*Corresponding author

Introduction

Ribonuclease P (RNase P) is the endoribonuclease responsible for removing the leader sequence from the tRNA precursors during the maturation of the 5' end of tRNAs (for reviews, see Altman et al., 1993; Pace & Brown, 1995). RNase P is a ribonucleoprotein whose catalytic function, at least in bacteria, is carried by its RNA component (RNase P RNA) rather than by the protein (Guerrier-Takada et al., 1983). Another remarkable feature of RNase P RNA is its ability to recognize the tertiary structure of its pre-tRNA substrates (Kahle et al., 1990). The secondary structure of bacterial RNase P RNA was inferred from a first comparative analysis of sequences (James et al., 1988) and refined as further sequences became available (Haas et al., 1991; Brown et al., 1993; Haas et al., 1994). An unexpected result was the recognition of two main structural types, namely type A and type B, which nevertheless share a common core (Haas et al., 1996b). Type A is the ancestral form (see the Escherichia coli sequence in Figure 1(a)), whereas type B has later emerged within the Gram-positive lineage (see the Bacillus subtilis sequence in Figure 1(b))). Previous work on group I introns has shown that large RNAs are likely to be organized in self-folded structural domains (Murphy & Cech, 1993; Douorna & Cech, 1995). Such domains are also visible in RNase P RNA, which is primarily organized in two distinct domains (Pan, 1995;
Modelling of the 3D Structure of RNase P RNA

Figure 1 (legend opposite)
Modelling of the 3D Structure of RNase P RNA

Loria & Pan, 1996). Domain I, involved in the recognition of the T-loop of pre-tRNAs (Pan et al., 1995; Loria & Pan, 1997), encompasses the "upper part" of RNase P RNA, i.e. stems P7 to P12, the large internal loop L11/12, as well as stems P13/ P14 (in type A) or P10.1 (in type B; see Figure 1). Domain II includes the "bottom" half of the molecule, and comprises most nucleotides thought to be involved in the formation of the catalytic site, i.e. stems P1 to P5 and their flanking strands (Frank et al., 1996; Pan & Jakacka, 1996; see Figure 1). Domain II includes as well stems P15, P19, and extensions P16/P17/P6 and P18 (in type A) or P5.1, P15.1/P15.2 (in type B). In spite of the early recognition of both structural types, most attention has been focused on the type A sequence from E. coli, for which two structural models for the RNase P RNA-pre-tRNA complex were independently proposed (Harris et al., 1994; Westhof & Altman, 1994). The Harris & Pace model was primarily designed as a frame for inter- and intramolecular crosslinking data (Burgin & Pace, 1990; Nolan et al., 1993; Oh & Pace, 1994), whereas the Westhof & Altman model was based on alternative cross-linking data (Guerrier-Takada et al., 1989) and chemical modification data (Lumelsky & Altman, 1988; Shiraishi & Shimura, 1988; Kirsebom & Altman, 1989; Knap et al., 1990; Talbot & Altman, 1994b; Westhof et al., 1996b). They share nonetheless some common details, such as the stacking of P1, P2 and P3, as well as an interaction between the L15/16 internal loop and the RCCA sequence at the 3′ end of pre-tRNA (Kirsebom & Svård, 1994; LaGrandeur et al., 1994). However, these working models remain organised in a radically different way, reflecting incompatibilities between some of the experimental data. The number of available sequences has dramatically increased within the last two years, bringing phylogenetic evidence for long-range interaction between L14 and P8, L18 and P8 (Brown et al., 1996), as well as between L9 and P1 (Massire et al., 1997). None of these contacts was actually predicted by the 3D working models, but the Harris & Pace model has been recently refined to accommodate them (Harris et al., 1997). Our previous model, however, could not incorporate these interactions without breaking the proposed catalytic site. This fact led us to reconsider the entire modelling process on the primary basis of an exhaustive phylogenetic analysis of the 137 bacterial sequences available from the ribonuclease P database (Brown, 1997) and from genome sequencing projects (see Methodology). In the present study, we propose 3D models for the two main structural types of RNase P RNA. Although they display different peripheral elements, they share a similar core centred around two inter-domain tertiary interactions involving loops L8 and L9. The proposed models should be viewed as "consensus" 3D structures that aim at emphasizing the folding architecture and relationships between the folding elements rather than precise atomic details.

Results

Type A and type B bacterial RNase P RNAs have both been shown to be formed of two independent folding domains that can self-assemble into a final active structure (Loria & Pan, 1996). The assembly of RNase P RNAs structural elements is thus fully compatible with a hierarchical folding process where prefolded RNA modules are used to create larger domains capable, once assembled, to associate into the final structure. The strategy we have used for modelling the three-dimensional architectures of type A and type B RNase P ribozymes reflects this modular view of RNA structures. In a first step, we began by modelling domain I and domain II separately. Once built, it became possible to orient and adjust the structure of the two domains with respect to their interdomain long-range interactions and tertiary contacts with the pre-tRNA substrate. The resulting three-dimensional models for the architecture of type A and type B RNase P ribozymes were finally refined with respect to the available published cross-linking data.

Modelling of domain I

The P10/P11 stack

All bacterial RNase P RNA sequences present a wide internal loop L11/12 linking stems P11, P12, P13 and P14 (James et al., 1988; and see Figure 1). This loop contains two conserved regions, 130-ACAGRNA-136 and 177-GUGNAA-182 (Chen & Pace, 1997) that are found also in RNase P RNA sequences from eukarya (Pagán-Ramos et al., 1996).

Figure 1. Classical representation of the 2D structure of (a) the E. coli RNase P RNA and of (b) the B. subtilis RNase P RNA. (Haas et al., 1994). Strands involved in the formation of the P4 and P6 pseudoknots are connected by square brackets. Extensions of stems P11, P10.1a, P15.1 and P19 are highlighted by continuous lines (see the text). Alternative representations for the 2D structures of (c) the E. coli RNase P RNA and of (d) the B. subtilis RNase P RNA. The alternative representations constitute an attempt at rendering the respective spatial arrangement of the helical stems and stacks in the 3D models. Thick lines link nucleotides adjacent in the sequence with arrows indicating the 5′ to 3′ polarity. Nucleotides invariant among bacteria are bold. Nucleotides involved in tertiary interactions are enclosed in boxes and linked by thin, dotted lines. Known non-canonical pairings are indicated by open circles. For each 2D diagram, the thick dotted line between P5 and P7 indicates the separation between the two main structural domains (Loria & Pan, 1996). Stems are colour-coded according to the helical stacks proposed in the present 3D models (see Figure 2).
and archaea (Haas et al., 1996a). In spite of this homology, the pairing that closes loop L11/12 varies between kingdoms. In bacteria, L11/12 is closed by a composite helical structure formed of the 2-bp stems P10 and P11 (Tallsjö et al., 1993; Haas et al., 1994) and the lone base-pair 128:230 (see Table 1 and Mattson et al., 1994). In eukarya, stems P10 and P11 are merged within a single stem of at least 6 bp (Tranguch & Engelke, 1993), whereas a continuous 8 or 9 bp stem P10/P11 may form in the archaea Methanococcus jannaschii and Archaeoglobus fulgidus. This suggests that, in bacterial RNase P RNA, P10, P11, and the 128:230 interaction could also be part of a long irregular helix. Nucleotides adjacent to the 128:230 interaction have been proposed to form additional base-pairs (Mattson et al., 1994; Westhof & Altman, 1994). Indeed, the base-pair 129:229 shows transversions from C:G to A:G and A:U (but no C:U is observed, see Table 2). The putative base-pair 127-G . . . 231-YAA motif is much more conserved, however. When the two adjacent bulging nucleotides A232 and A233 are present (in all bacteria but bacteriodes), this base-pair exhibits a strong bias towards G:U and G:C pairs (then mostly in thermophilic species). Interestingly, in the 113 RNase P RNA sequences displaying the bulge adenine bases A173 and A174 at the base of stem P12, a similar pattern of variation is observed for the adjacent 144-G . . . 172-YAA motif. We suggest, therefore, that the 127-G . . . 231-YAA motif should be viewed as a part of the 7 bp stem P10/P11, in the same way the 144-G . . . 172-YAA motif is incorporated into stem P12 (see Figure 1(c)). A233 has been shown to be involved in the formation of a contact with the hydroxyl O2' of residue 62 in the pre-tRNA, which suggests that this internal loop motif, G . . . YAA, is particularly well suited for tertiary interactions. In fact, in our A type model, 144-G . . . 172-YAA is proposed to interact with loop L13 (see below).

Table 1. Covariations between bases 128 and 230 (136 sequences)

<table>
<thead>
<tr>
<th>Base 230</th>
<th>Base 128</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>126</td>
</tr>
<tr>
<td>U</td>
<td></td>
</tr>
</tbody>
</table>

Data are collected with the program COSEQ (unpublished).

Table 2. Covariations between bases 129 and 229 (135 sequences, both bases are deleted in the H. mobilis sequence)

<table>
<thead>
<tr>
<th>Base 229</th>
<th>Base 129</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>28</td>
</tr>
<tr>
<td>U</td>
<td>13</td>
</tr>
</tbody>
</table>

The P13/P14 stack

Type B and most β-purple bacteria sequences do not possess hairpins P13 and P14, which, otherwise, always appear simultaneously in loop L11/12 (see Figure 1(a) and (b)). This suggests that, in type A RNase P RNA, these stems are coaxially stacked, so that their 5' and 3' dangling strands are brought close to each other (see Figure 1(c)). Interestingly, the secondary structure of P14 can be refined in order to leave an identical number of unpaired nucleotides in the L11/12 loop of type A and B RNase P RNAs (see Figure 1(c) and (d)), in agreement with the high level of sequence conservation of L11/12. Indeed, in all type A RNase P RNAs, stem P14 can be extended by joining nucleotide(s) in J13/14 to adjacent nucleotide(s) in J14/11, leaving in most sequences the conserved G225 bulging (see Tables 3 and 4). Phylogenetic evidence for this extension is mainly provided by flavobacteria sequences, where the bulge in P14 is absent and where all Watson-Crick combinations are encountered for the base-pairs extending P14. Thus, the length of P14 is more constant than it was previously proposed (9 or 10 bp long, instead of 7 to 9 bp) in agreement with the tertiary interaction involving L14 and P8 (Brown et al., 1996).

Table 3. Covariations between bases 203 and 226 (110 sequences possessing P13/P14)

<table>
<thead>
<tr>
<th>Base 226</th>
<th>Base 203</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>101</td>
</tr>
<tr>
<td>U</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 4. Covariations between bases 203.0 (5' to base 213) and 227 in the 51 sequences having base 203.0

<table>
<thead>
<tr>
<th>Base 227</th>
<th>Base 203.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
</tr>
<tr>
<td>U</td>
<td>44</td>
</tr>
</tbody>
</table>
similar to the P5b-P6a interaction seen in the crystallographic structure of the P4-P6 domain of the Tetrahymena group I intron (Cate et al., 1996a). Although the Clostridium innocuum sequence contains a GAAA loop L12, a particular loop receptor rich in G:U pairs has been suggested for this sequence in lieu of the expected 11 nt receptor motif (Haas et al., 1996b). This receptor can be rearranged as a variant of the 11 nt motif (see Figure 3). Indeed, the resulting motif CCUG-G…UAUGG, with a GU platform instead of the canonical AA platform, has actually been found by in vitro selection to interact specifically with GAAA tetraloops (see sequence C7.10 in Figure 5B of Costa & Michel, 1997). The L12-P10.1a interaction is also feasible in the Brevibacillus brevis sequence, this time through a GYRA loop-helix interaction (Michel & Westhof, 1990; Jaeger et al., 1994).

Most non-mycoplasma sequences possess an internal loop L10.1 of consensus Bsu 139-RAAA…166-RAGUA and can, therefore, be aligned with confidence (see Figure 3). Strikingly, this internal loop motif is similar to the loop E of eukaryotic 5 S RNA (Shen et al., 1995), which is characterized by a well-defined structure involving the stacking of non-canonical base-pairs. By analogy, we propose the formation of the following non-canonical pairs within L10.1 (see also Figure 3): R139:A170 (sheared), A140:U169 (Hoogsteen), A141:A167 (Hoogsteen) and A142:R166 (sheared). As a result, the P10.1 extension should appear as a straight, irregular helix that keeps constant the distance between the base of P10.1 and the 11 nt GAAA receptor (see Figures 1(d), 2(b) and (d)). This is emphasized by the observation than the loop E motif L10.1 can be slide one base-pair, since it does not affect the overall length of P10.1/P10.1a. Even the Bacillus megaterium sequence which possesses a longer AAAAACC…CGAGUA internal loop (extra nucleotides are underlined) but lacks two base-pairs in P10.1a can potentially have a similar overall length for its P10.1/P10.1a extension. The formation of an AC platform, followed by a C.C pair can indeed substitute for the two missing base-pairs (Figure 3; and see Cate et al., 1996b). This latter motif has actually been recognized as an alternative for the AA platform and the G:U pair within the GAAA receptor (see sequence C7.16 in Figure 5B of Costa & Michel, 1997). The overall stacking of the P10.1 extension is moreover confirmed by the remaining type B sequences: the Acholeplasma laidlawii sequence lacks the L10.1 internal loop and possesses instead a continuous P10.1 stem and, in the C. innocuum sequence, the three purine:purine pairs that constitute L10.1 are likely to stack on each other. Interestingly, when the L12-P10.1a interaction is present, the overall length of P10.1/P10.1a seems to correlate with the length of P12: indeed, the three mycoplasma sequences M. capricolum, M. pneumoniae and M. genitalium display a simultaneous shortening of stem P12 and of the L10.1 internal loop, while retaining the L12-P10.1a interaction.

Analogy between P13/P14 and P10.1a/P10.1

It was previously suggested that P10.1a/P10.1 could be a structural equivalent for the P13/P14 stems (Loria & Pan, 1996). The fact that the RNase P protein component from E. coli can bind either to a type A or to a type B RNase P RNA could plead for a high level of similarity between tertiary structures of both types of RNase P RNA (Guerrier-Takada et al., 1983). In type A sequences, the loop-helix interaction L14-P8 brings P14 close to the 3 nt bulge between P10 and P11, i.e. close to the insertion site of P10.1 in type B sequences (see Figures 1 and 2(c) and (d)). The type B stem P10.1a/P10.1 could therefore take the place of the type A stems P13/P14 if one assumes a tetraloop conformation for the GAAA fragment that links P10.1 and P11 in most type B sequences (see Figures 1(c) and (d)). By considering the analogy between P13/P14 and P10.1a/P10.1, one could expect that a tertiary interaction, specific to type A sequences, replaces the type B interaction L13-P10.1a. Actually, P13, which is always 6 bp long and is capped by a loop of consensus GUAAGAC, could be involved in a tertiary contact with the 144-G…172-YAA motif at the base of stem P12. Indeed, Brown et al. (1996) have already noticed that the stem–loop P13/L13 and the 144-G…172-YAA motif always appear simultaneously, both features having been deleted in type B and β-purine bacteria sequences. Interestingly, the mitochondrial RNase P RNA from Reclinomonas americana lacks both the L13 loop and the P12 bulge (Lang et al., 1997). Since mitochondrial sequences have been derived from the z lineage of purple bacteria, the Recl. americana sequence provides another instance of the simultaneous deletion of the P12 bulge and the L13 loop, providing further evidence for their implication in a common interaction. However, due to the lack of experimental evidence for specific contacts between L13 and P12, we did not attempt to model this hypothetical interaction in detail.

Conformation of the P7/P8/P9/P10 cruciform junction

The juxtaposition of stems P7 to P10 creates a cruciform in the RNase P RNA structure (Haas et al., 1994). In such a four-way junction, stems are expected to be stacked by pairs, leaving only two theoretically feasible conformations: a P7/P8 and a P9/P10 stack or, alternatively, a P7/P10 and a P8/P9 stack. We constructed the cruciform junction with both conformations. Only the latter was found compatible with the geometry of the L14-P8 interaction and with the simultaneous extension of stems P10/P11 and P14, which leaves at most two nucleotides between P14 and P11 (see Figure 1(a) and (c)). This conformation is moreover the only one that is in agreement with the cross-link of pre-
Figure 2. Ribbon representations of the complete 3-D models of RNase P RNA. Front view of (a) the *E. coli* model and of (b) the *B. subtilis* model. The colour code is the same as in Figure 1. Stems of domain I are red, green and deep blue; stems of domain II are orange, purple, light blue and yellow. The invariant nucleotides of the catalytic site are represented with large white spheres, while the conserved nucleotides from the T-loop recognition site are represented with smaller spheres. Back view of (c) the *E. coli* model and of (d) the *B. subtilis* model. A rotation of 180° along the vertical axis has been applied to each model between front and back views. The loop-helix interactions L9-P1, L8-P4, L14-P8, L18-P18 in the *E. coli* model, and L12-P10.1 and L15.1-P5.1 in the *B. subtilis* model are indicated by wide arrowheads. The ribbon drawings were made with the DRAWNA software (Massire *et al*., 1994).
tRNA G64 with both A118 in L9 and G100 in L8 (Nolan et al., 1993).

In the present models, domain I is therefore made of the tight assembly of the nearly coaxial stacks P9/P8, P11/P10/P7, and alternatively P13/P14 or P10.1a/P10.1 (see Figure 2). The position of stem P12, capping the P11/P10/P7 stack, is uncertain due to the lack of evidence for specific contacts in the large loop L11/12, but is nevertheless partially constrained by the known geometry of the L12-P10.1 interaction in the B. subtilis model.

Modelling of domain II

The stem P19

Stem P19 appears sporadically between P2 and the well-conserved fragment J19/4 of consensus ACARAA. Fragment J19/4 constitutes with the 3' strand of P4 the last universally conserved region (Chen & Pace, 1997) and is unchanged regardless of the occurrence of P19. Stem P19 is present in most branches of the bacterial tree, except in ψ-purine bacteria (other taxa provide evidence for a late deletion of P19). We noticed that in nearly all proposed secondary structures of RNase P RNA with P19, the distance between P2 and P19 is the same as that between P19 and the ACARAA consensus fragment (from zero to three nucleotides). However, the secondary structure of stem P19 can be refined in order to leave no unpaired nucleotide between P2 and P19 as well as between P19 and the conserved part of J11/12. Table 5 summarizes the different types of pairing that are found within the six proximal base-pairs of P19 (see also Figure 1(b) and (d)). Although pairs Bsu 329:368

---

Table 5. Comparative analysis of the pairing types within the six proximal base-pairs of P19 (B. subtilis numbering)

<table>
<thead>
<tr>
<th>Base-pair</th>
<th>No. of sequences</th>
<th>Regular</th>
<th>Pairing type</th>
<th>Mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>329:368</td>
<td>88</td>
<td>57</td>
<td>21</td>
<td>7</td>
</tr>
<tr>
<td>330:367</td>
<td>86</td>
<td>65</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>331:366</td>
<td>86</td>
<td>75</td>
<td>4</td>
<td>7</td>
</tr>
<tr>
<td>332:365</td>
<td>83</td>
<td>75</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>333:364</td>
<td>83</td>
<td>72</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>334:363</td>
<td>79</td>
<td>69</td>
<td>4</td>
<td>6</td>
</tr>
</tbody>
</table>

and 330:367, located at the base of P19, tend to form more non-classical base-pairs than 331:366 to 334:363, they nevertheless favour regular Watson-Crick base-pairs, or alternative A:A, G:A, A:G or A:C pairs rather than other mismatches. Purine:purine pairs have been observed in crystallographic structures extending regular helical stems, e.g. base-pair G26:A44 in tRNA\textsuperscript{Asp} (Westhof \textit{et al}, 1985) and in the P4-P6 domain of group I introns (Cate \textit{et al}, 1985). Similarly, A:C pairs are observed at the edge of RNA helices; for example, an A:C pair is found at the edge of the anticodon stem of tRNA\textsuperscript{Asp} (Westhof \textit{et al}, 1985). Therefore, the presence of purine:purine or A:C pairs at the base of P19 is fully compatible with the extension of the helical stack. As a consequence, stem P19 appears to be adjacent to P2 in bacteria, as already noticed in eukarya (Tranguch & Engelke, 1993).

**Main helical stacks in domain II**

Since the major driving force for RNA folding is base stacking, a single stem is likely to be inserted in a given structure at the top of an already existing helical stack. Thus, the fact that the single stem P19 is adjacent to P2 suggests a P19/P2 stack. The sequences that lack P19 possess instead a 4 nt linker (R19) between P2 and the ACARAA fragment J19/4. Fragment J19/4 is highly conserved and should therefore adopt the same conformation in all bacterial RNase P RNA sequences. Consequently, the 3' ends of P19 and R19 should also adopt a similar conformation in both type A and type B models. This is achieved if R19 adopts a loop conformation, mimicking the right-handed stack between P19 and P2 (see Figure 4). Thus, in both 3D models, the conserved J19/4 fragment lies on the edge of the deep groove of P2 and along the shallow groove of P4. At the other end of P2, a single, invariant nucleotide (G19) is inserted between P2 and P3 (see Figure 1). In the absence of any other feasible alternative stacking, we conserved in our models the P2/P3 stack already present in previous models (Harris \textit{et al}, 1994; Westhof \& Altman, 1994). Therefore, the assembly of P19, P2 and P3 forms the first continuous helical stack in domain II (see Figure 1 and 2). The choice of the P19/P2 stack in our models prevents the once proposed stacking of P1 on P2. Instead, we propose to stack P1 on P4, since a single conserved nucleotide (A361) is always inserted between the 3' strands of P4 and P1 (Haas \textit{et al}, 1996b). P4 is in turn likely to stack on P5, because their 5' strands are always contiguous (Harris \textit{et al}, 1994). Finally, P5 is adjacent to both P7 and P5.1 in type B sequences. However, P5 and P7 belong to different folding domains and are not adjacent in some cyanobacteria sequences, whereas P5 and P5.1 are always contiguous. Moreover, the fact that the single stem P5.1 is present in type B sequences argues for its stacking on a conserved stem, i.e. P5. Thus, P1, P4, P5 and P5.1 constitute the second continuous helical stack in domain II (see Figures 1(d) and 5(b)).

**The P15/P16/P17/P6 extension in type A RNase P RNA**

In our models, stem P15 is the only conserved stem from the domain II that is not part of the main helical stacks. P15 is poorly constrained in type B sequences because of the lack of evidence of specific contacts within the conserved dangling strands J5/15 and J15/18. In type A sequences, however, P15 constitutes the base of the P15/P16/P17/P6 extension, which is strongly constrained by the formation of pseudoknot P6. The orientation of P15 may therefore be inferred from the modelling of the whole P15/P16/P17/P6 extension from type A RNase P RNA. P6 is usually 4 bp long, although its size can reach 7 bp in \textit{z}-purple bacteria and cyanobacteria (Vioque, 1997). One to four nucleotides link P5 and P6, whereas a single adenosine base joins P6 to P7 in most cases. Therefore, P6 was stacked below P7 in the \textit{E. coli} model (see Figures 1(c) and 5(a)). The other end of P6 is always adjacent to the 3' strand of P17, justifying the stacking of P6 and P17 within a type III pseudoknot (see Figure 5(a) and Westhof \& Jaeger, 1992). Since the P15/P16/P17/P6 extension is constrained at both ends, it should contain a sharp turn either between P16 and P17, or between P15 and P16 (see Figures 1 and 2). The internal loop between P16 and P17 is extremely variable in composition or in length, and is often interrupted by the insertion of an extra stem, either between P16 and P17 (e.g. in planctomycete or \textit{z}-purple bacteria sequences, see Brown, 1997) or between P17 and P16 (e.g. in Chlamydiae, see Herrmann \textit{et al}, 1996). The fact that these extra stems are likely to stack
on either P16 or P17 strongly suggests the presence of a kink between P16 and P17 (see Figures 1 and 2). The L15/16 internal loop, or its homologue in type B sequences, L15, is involved in intermolecular interactions with the pre-tRNA 3'-RCCA sequence (LaGrandeur et al., 1994) through at least two canonical base-pairs (Kirsebom & Svard, 1994). An NMR structure of a 31-mer RNA comprising P15, P16 and the L15/16 internal loop (Glemarec et al., 1996) shows that the bases within the internal loop L15/16 stack upon each other, leaving the Watson-Crick side of G292 and G293 available for additional pairings. More recently, the L15/16 loop and its homologue in B. subtilis, L15, were modelled in detail on the basis of a phylogenetic analysis (Easterwood & Harvey, 1997). Our model of this region, which does not differ significantly from the latter work, presents a nearly coaxial stacking of P15 and P16, allowing the pre-tRNA 3'-RCCA sequence to interact in the minor groove of L15/16.

**Putative interaction between L5.1 and L15.1 in type B RNase P RNA**

Type B RNase P RNA sequences lack the P16/P7/P6 extension that contributes to the stabilisation of type A RNase P RNA structures via the formation of the long-range interaction P6 (Haas et al., 1991). Nevertheless, they are characterized by the additional stems P5.1 and P15.1 that could potentially form an interaction in order to stabilize domain II (Haas et al., 1996b). The recent availability of further type B sequences allowed us to reconsider the previous alignment of P5.1 and P15.1 (see Figure 6). Stem P15.1, previously considered to have 6 or 7 bp, can be extended up to 11 canonical base-pairs in each type B sequence (see Figure 6). This extension is made possible in some species by replacing two contiguous base-pairs by a GA...GA motif or one of its variants, or by bulging a nucleotide on the 3' strand. The L15.1 loop

![Diagram](image)

Figure 5. Close-up stereo view of the junction (a) between P5, P6/P17 and P7 in the *E. coli* model and (b) between P5, P5.1 and P7 in the *B. subtilis* model. Unpaired nucleotides A79, U80, A81, A86 and G279 in the *E. coli* model and A81 to U85 in the *B. subtilis* model are drawn as sticks.

![Diagram](image)

Figure 6. Refined alignment of stems P5.1 and P15.1 for the 18 type B bacterial sequences. Canonical Watson-Crick pairs are green, wobble pairs are blue, sheared pairs are red, nucleotides characteristic to the lone Acholeplasma laidlawii sequence are purple. Unpaired nucleotides are black. Nucleotides that are invariant within type B sequences (but in *Acp. laidlawii*) are bold. Square brackets stand for fragments of sequence not shown.
is then reduced to the consensus RAAN_{5,5}AA. It is striking that the presence of such a loop in L15.1 appears to correlate with the presence of a highly conserved stem P5.1 of 6 bp long and capped by an UGNRAU loop (see Figure 6). Indeed, in the Acp. laidlawii sequence, where L15.1 is reduced to the well-known GCAA tetraloop (Heus & Pardi, 1991), stem P5.1 contains an extra C.G base-pair and is not capped by a UGNRAU loop. Furthermore, at a finer scale, the sequences of M. genitalium and M. pneumoniae may also be distinguished from one in both the L5.1 and L15.1 loops (see Figure 6). In these sequences, the L5.1 loop follows the consensus UGGAAAY instead of UGGGAY (significant differences are underlined, see also Figure 6). This variation correlates with the presence in L15.1 of a loop following the consensus RAAN_{5,5}AUAA observed in all L8 decaloops (see below), instead of consensus RAAN_{5,5}AA (see Figure 6). These correlated changes suggest that L5.1 and L15.1 are indeed interacting. The structure of the putative L5.1-L15.1 interaction was not modelled in detail due to the lack of evidence for any specific contact. Nevertheless, the fact that a GCAA tetraloop in L15.1 correlates with the presence of two G.C base-pairs in P5.1 pleads for a classical GCAA loop-helix interaction between L15.1 and P5.1 in the RNase P RNA from Acp. laidlawii, in place of the more complex interaction in the other type B sequences. Since P5.1 and P15.1 are expected to be located at the same place within all type B structures regardless of the actual nature of the P5.1-P15.1 interaction, the permutation of P5.1-P15.1 modules between different type B structures should preserve the active conformation. Therefore, in the B. subtilis model we have replaced the native P5.1-P15.1 module by its counterpart from the Acp. laidlawii sequence, where the relative orientation of P5.1 and P15.1 is constrained by the known geometry of a GNRA-loop interaction (Jaeger et al., 1994; Pley et al., 1994). The imposed stacking of P5.1 below P5 (see Figure 2(d)) thus constrains the position of P15.1.

**Homology between P18 and P15.1/P15.2**

Stems P15.1 and P15.2 are always contiguous in type B sequences (Haas et al., 1996b). Their simultaneous insertion S to the fragment Bsu 314-AGANNAU-321 (nucleotides 328 to 335 in E. coli; see Figure 1) occurs in replacement of one or two nucleotides (Haas et al., 1994, 1996a). Therefore, P15.1 and P15.2 are likely to stack on each other within an independently folded subdomain P15.1/P15.2. Most type A sequences possess instead a single stem P18 (see Figure 1(c) and (d)). In our models, P18 and P15.2 are placed similarly and can even be considered as two variants of the same helical stem (see Figure 2(c) and (d)). We keep, however, the different names to emphasize their structural differences: P18 is always 8 bp long and capped by a GNRA loop, whereas P15.2 does not exhibit significant conservation in either sequence or length. P18 is constrained in the E. coli model by the L18-P8 interaction (Brown et al., 1996), whereas P15.1/P15.2 in the B. subtilis model is instead constrained by the L15.1-L5.1 interaction. It is noteworthy that these interactions place similarly the P18 and P15.1-P15.2 extensions behind P4, so that the neighbouring Eco 328-AGA-330 fragment can adopt the same conformation in both models, which is expected, since this fragment is conserved in all bacterial sequences.

**Interdomain association**

The fact that the two independent folding domains, when synthesized separately, can self-assemble into a catalytically active complex (Loria & Pan, 1996) suggests the formation of multiple tertiary contacts between the two pre-folded domains. The first identified interdomain interaction occurs between L18 and P8 (Brown et al., 1996). However, RNase P RNA mutants with a deletion of P18 retain a weak catalytic activity, indicating that the interdomain association may occur in the absence of the L18-P8 interaction (Haas et al., 1994). We recently reported phylogenetic evidence for another interdomain tertiary interaction between L9 and P1 (Massire et al., 1997). The conservation of a GNRA loop L9 on top of the 5 bp stem P9 is correlated with the conservation of a G:C pair in the distal end of P1. Alternatively, L9 and the 3' end of P1 are complementary in at least two Mycoplasma sequences, where they have therefore the ability to be involved in the formation of a pseudoknot in lieu of the loop-helix interaction. Nonetheless, the L9-P1 interaction is absent in one-third of the sequences, indicating that additional interdomain contacts may still remain to be identified.

**Putative interaction between L8 and P4**

L8 is a good candidate for such an additional interaction. In type A RNAse P RNA, P8 is indeed the anchoring site for both L14 and L18, suggesting that P8 itself should be tightly bound to the core. Moreover, in all bacterial RNAse P sequences, there is a striking conservation of two adenine bases within L8 (Eco A98 and A99 or Bsu A105 and A106, see Figures 1 and 7(a)), pointing to their potential role in folding or/and catalytic function of bacterial RNAse P RNA. Interestingly, we have noticed that all 11 to 12 bp separate these conserved adenine bases from loop L9, which was suggested to interact with P1 in the majority of RNAse P RNA sequences (Figure 1(c); see Massire et al., 1997). Since P8 and P9 are stacked, loops L8 and L9 are separated by a complete helical turn and point in the same direction. The L1-L9 interaction could thus act as a ruler for positioning the conserved adenine bases of L8 in front of a specific site of interaction. Considering that P1 and P4 are stacked in our current RNAse P RNA models, the best candidate for this kind of interaction is P4.
served G:C pairs within core element P4 are, indeed, located at a distance of one helical turn from the two base-pairs supposed to interact with L9 (see Figure 1(c)). Therefore, whereas L9 interacts with the shallow groove side of 3:371 and 4:370 base-pairs of P1 in our type-A model, the two adenine bases of L8 interact with the shallow groove side of the conserved Eco\textsubscript{C}71:G356 and C70:G357 base-pairs of P4 (see Figure 1(c) and (d)). Although this putative tertiary interaction is not corroborated by a direct covariation, its formation is fully compatible with both type A and type B models, and can help in understanding the subtle variations observed for the length of the stems P5, P7 and P8, which are the structural elements separating P4 from L8. Most type A and type B sequences are characterized by a different number of nucleotides separating the base of stem P8 and the two conserved adenine bases in L8 (see Figure 7(a)). Interestingly, this length variation within stem–loop P8/L8 correlates with the length variations of stems P5 and P7 (see Figure 7(a)). If one assumes an antiparallel orientation between P5 and P7 as well as between P7 and P8, the anchoring point of P5 to P4 is then located at a conserved distance from the two conserved adenine bases in L8, in both type A and type B RNase P structures (see Figure 7(b)). Therefore, the positioning of the conserved adenine bases of L8 in front of the fourth and fifth C:G base-pairs of P4 is possible in both
type A and type B three-dimensional models (see Figure 1(c) and (d)) and can justify the concerted changes of length observed among RNase P sequences within P5, P7 and P8. Several experimental data support the existence of a tertiary contact between L8 and P4. Indeed, protections towards the Fe(II)-EDTA hydroxyl radical of both L8 and P4 are lost when domains I and II are constructed and folded separately (Loria & Pan, 1996). Moreover, recent cross-linking data have shown the proximity of L8 and P4 in the catalytically active RNase P RNA (Harris et al., 1997). With this interaction, the P8/P9 stack is anchored to the catalytic domain at both ends through loop-helix interactions (see the green arrows in Figure 2(c)), providing a rigid frame for the L14-P8 and L18-P8 interactions (see the blue arrows in Figure 2(c)).

**The overall three-dimensional architecture**

According to our three-dimensional models, the architecture of bacterial RNase P RNAs can be seen as the assembly of helical stacks packed together through the formation of long-range tertiary interactions (see Figures 1 and 2). Several intradomain long-range interactions participate in the independent folding and stabilization of domains I and II. So far, no less than three long-range interactions have been found to constrain the folding of domain I: interactions L14-P8 and L13-P12 in type A RNase P RNA, and L12-P10.1a in type B RNase P RNA. Domain II from type A RNase P RNA is stabilized by the long-range interaction P6, whereas loop-loop interaction L5.1-L15.1 represents the P6 natural counterpart in type B RNase P RNA. Interdomain tertiary interactions like L18-P8, L9-P1 and L8-P4 are then favouring the association of domains I and II into the final active structure. The core from both models presents a side-by-side assembly of the four main helical stacks: P11/P10/P7, P9/P8, P1/P4/P5 and P2/P3 (see Figures 1 and 2). This assembly is smoothly bent, with an internal and an external side (see Figure 8(a)). The internal side is concave and acts as a cradle for the pre-tRNA substrate. It contains also most of the universally conserved nucleotides from P4, J5/15, J18/2, J19/4 that are thought to form the catalytic site (see Figure 2). Contacts with the pre-tRNA substrate are clustered in three regions. (i) The 3’-terminal RCCA sequence binds to the internal loop L15/16 in the E. coli model or to the L15 loop in the B. subtilis model (see Figure 8(b)). (ii) The catalytic site, in the neighbourhood of the 5’ termini of the tRNA, comprises J5/15, the 5’ strand of P4, J18/2 as well as the conserved A351 and 352 from J19/4. (iii) The P11 stem and the A118 from P9 are part of the recognition site for the T-loop of the substrate (Pan et al., 1995; Loria & Pan, 1997). The external side of the three-dimensional architecture possesses, in contrast, the most variable extensions that maintain the active conformation of the catalytic site. On this external side can be found the anchoring sites for the L14 and L18 loops (see Figure 2(b) and (d)) and most nucleotides involved in the binding of the RNase P protein component (Talbot & Altman, 1994a).

**Validation of the models**

The structure of the E. coli RNase P RNA has been investigated through crosslinking experiments (Burgin & Pace, 1990; Nolan et al., 1993; Harris et al., 1994) using a 9 Å long crosslinking agent (azidophenacyl bromide) specifically attached to the 5’ end of either RNase P RNA or mature tRNA. The main crosslinks obtained in that way are summarized in Table 6. For each attachment site, the crosslinked nucleotides are specified,
with the mean inter-nucleotide distance (measured between phosphorus atoms) indicated in parentheses. We consider that a given crosslink is compatible with our model if this distance is less than 30 Å, since the localization of these crosslinks is not known at an atomic scale. Moreover, since most attachment sites are obtained by circular permutations of the RNA, crosslinking agents are attached to new 5' ends, for which the backbone itself is expected to add some extra flexibility. The majority of intra- and inter-molecular crosslinks are compatible with our models (see Table 6). Two distinct sets of incompatible crosslinks are visible. The first set involves the G179 crosslinks, i.e. within the region of L11/12, which is poorly conserved. The second set contains specific contacts between its bases, which we did not attempt to model. The second set of unsatisfied constraints occurs in the region of P15. Once again, the actual 3D structure of this region is likely to be more compact than in our models and remains undetermined.

**Discussion**

**General organization of the core of RNase P RNAs**

Type A and type B RNase P secondary structures share a common conserved core that encompasses parts of structural domains I and II: roughly, stems P1 to P5 with P15 from domain II and stems P7 to P11 from domain I (see Figure 9). These core elements may be conserved because: (i) they are part of the catalytic site (P4, J5/15, J19/4 and J18/2); (ii) they are involved in the correct recognition of the substrate (P11, L15); and (iii) they form part of the frame that stabilizes the catalytic core or constitutes the interdomain interface (P8, P9). Therefore, deletion of some of the normally conserved core elements may affect or abolish the specificity for pre-tRNA substrates, while still allowing cleavage of alternative substrates (Guerrier-Takada & Altman, 1992; Pan & Jakacka, 1996). Figure 9(c) illustrates the decomposition of the core into two main domains, each formed by the coaxial stacking of several helices. We do not wish to convey the impression that Figure 9(c) represents the 2D structure of an Ur-RNase P, precursor of the two bacterial types, although we cannot exclude the fact that it constitutes a possibility. The common interdomain contacts involve P1 and L9 as well as P4 and L8. However, type A and type B RNase P RNAs present their own set of peripheral, tertiary elements that contribute to the stabilization of the catalytic core through the formation of specific tertiary interactions (see The overall three-dimensional architecture in Results and Figure 9(a) and (b)). Insertion sites of less conserved, peripheral elements occur on the core periphery. Type A and type B RNase P RNAs are best distinguished by the elements inserted in the single-stranded regions between P5 and P7 and L15 (see Figure 9). Several long-range interactions characteristic of each type of RNase P occur between peripheral elements (e.g. L5.1-L15.1 in type B or P6 in type A) as well as between the periphery and the core (e.g. L14-P8 and L18-P8 in type A). Despite the fact that these peripheral elements are non-homologous, they nevertheless have similar, functional purposes by contributing to the folding and stabilization of a common catalytic core.

**Comparison with other models**

The present models are somewhat similar, in their general domain organisation, to the last model described by Harris and Pace (later referenced as the HP model, see Harris et al., 1997). Common features include a roughly similar orientation of the pre-RNA substrate, an identical flattened conformation of the four-way junction, a network of tertiary interactions L14-P8, L18-P8 and L9-P1, and a C-shaped extension P15/P16/P17/P6. Significant divergences are, however, clearly visible, in particular in domain II. Whereas the P4/P5 and P1/P2/P3 stacks form a T in the HP model, the stacks P1/P4/P5 and (P19)/P2/P3 are nearly parallel in our models.

<table>
<thead>
<tr>
<th>Attachment</th>
<th>Crosslinked nucleotides (inter-phosphate distance in Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G 63</td>
<td>335 (17.8) 351 (25.5) 352 (26.4)</td>
</tr>
<tr>
<td>G 135</td>
<td>112 (15.0) 118 (18.0)</td>
</tr>
<tr>
<td>G 179</td>
<td>112 (29.0) 118 (30.4)</td>
</tr>
<tr>
<td>G 229</td>
<td>127 (17.9) 129 (18.4)</td>
</tr>
<tr>
<td>G 292</td>
<td>247 (24.9) 248 (23.9)</td>
</tr>
<tr>
<td>G 304</td>
<td>247 (28.0) 248 (24.6)</td>
</tr>
<tr>
<td>G 332</td>
<td>247 (15.4) 248 (13.7)</td>
</tr>
<tr>
<td>G 350</td>
<td>247 (20.0) 248 (19.4)</td>
</tr>
<tr>
<td>G 1</td>
<td>246 (10.1) 331 (20.9)</td>
</tr>
<tr>
<td>G 53</td>
<td>111 (21.6) 114 (24.4)</td>
</tr>
<tr>
<td>G 64</td>
<td>67 (25.5) 68 (21.2)</td>
</tr>
</tbody>
</table>

The attachment site of 5'-azidophenacyl crosslinking agents is indicated in the first column. For each attachment site, the crosslinked nucleotides are indicated, with the mean internucleotide distance measured between P atoms. Nucleotides within the pre-tRNA substrate are indicated with italicized numbers. Distances greater than the chosen threshold value (30 Å) are bold.
Figure 9. Simplified 2D representations of (a) a typical type A sequence and of (b) a type B sequence with their characteristic tertiary interactions. Stems are colour-coded as in Figures 1 and 2 with conserved regions represented with thick lines. Main insertion sites are indicated by arrowheads. (c) A 2D representation of the minimal core. Insertion sites common to both types (P12, P18 and P19) are indicated by small arrowheads. The insertion sites of peripheral elements that are characteristic of a given type are indicated by large arrowheads.
More importantly, the incorporation of P19 in our models leads to a different topology of the central pseudoknot with an alternative crossing of the junctions J19/4 and J3/4 (see Figure 4). In Results, we proposed that the conserved J19/4 lies on the edge of the deep groove of P2 and along the shallow groove of P4, if one assumes a right-handed stack between P19 and P2. The crossing between J19/4 and the 5' strand of the P1 and P2 helices is thus reversed compared to that of the HP model. In order to avoid the passage of J19/4 between the 5' strand of P2 and J3/4, the crossing between the J19/4 and J3/4 strands also must be reversed. As a consequence, J19/4 lies on the substrate side in our models, instead of being on the protein side as in the HP model (compare the boxed 2D sketches in Figure 10(a) and (b)).

This different placement of J19/4 has important consequences for the sequence of folding of the central pseudoknot. P2 was among the first helices to be firmly identified by phylogenetic sequence comparison (James et al., 1988) and has, therefore, been naturally thought to be part of the secondary structure. Evidence for four out of the eight base-pairs of P4 was then also provided. Because of its assumed shorter length, P4 has been implicitly considered to fold after the formation of P2. This hierarchical folding is present in the HP model: the unfolding of its pseudoknot starts with the disruption of P4 (see Figure 10(a), right) and not by the opening of P2, which would lead to an entanglement (see Figure 10(a), left). As new sequences provided new instances of covariations in the distal base-pairs, P4 has been extended to its present length of 8 bp plus a bulge (Haas et al., 1991). However, the folding of the central pseudoknot has not been re-analysed. The following facts argue for an alternative folding of the pseudoknot:

(i) P4 is indeed longer than P2 (8 bp and a conserved bulge versus 7 bp); (ii) P2 displays greater variation in length than P4, and is even absent in some mitochondrial sequences where P4 lacks only one base-pair; (iii) protections towards the Fe(II)-EDTA hydroxyl radical show that P2 is almost completely exposed to solvent, whereas P4 is deeply buried in the structure (Loria & Pan, 1996); and (iv) experimental evidence for a slower formation of P2 has been reported (Zarrinkar et al., 1996). These arguments would indicate that P2 is indeed a tertiary interaction with P4 forming the 2D structure. This is noticeable in our models (see Figure 10(b), right), where the 3' strand of P2 and its flanking junctions, J18/2 and J19/4, make a broad loop that, upon folding, creates the pairing of P2. One can therefore expect that the unfolding of the pseudoknot starts with the disruption of P2 and not of P4 (see Figure 10(b), left).

In domain I, the main difference between the HP model and our models concerns the positioning of P13 and P14. In our models, the fact that P13 and P14 are stacked implies a rather different orientation of P13 and P12. The latter stem is partially constrained by the maintenance of the L12-P10.1 interaction. But, since P10.1 is itself constrained only in its lower part, small variations in the structure of the P10/P10.1/P11 junction would result in greater changes in the position of the distal end of P10.1. This would dramatically alter the orientation of P12. This is the reason why the modelling of P12 and L11/12 will probably need refinement.

**Long-range interactions within RNase P RNAs**

**RNA-RNA long-range interactions**

The long-range interactions within RNase P RNAs belong essentially to two broad categories: base-pairing between two loops and the family of GNRA loop-receptor interactions (Westhof et al., 1996a; and see Figure 1). Interestingly, both kinds of contacts are alternatively found between L9 and P1, since this interaction may be realized either through a GNRA-helix interaction (mostly in type A sequences) or through the formation of a pseudoknot P21 in some mycoplasmas (Massire et al., 1997). Nevertheless, GNRA loop-receptor interactions are found at four distinct locations and are thus the most widespread long-range interactions among RNase P RNAs.

In addition, several tertiary interactions proposed in our models involve adenine-rich terminal loops (L8, L13, L5.1 and L15.1) distinct from the classical GNRA tetraloop, as well as adenine-rich internal loops, such as those found within P10/P11 and P12. Recently, Abramowitz & Pyle (1997) reported that GNRA tetraloops may be considered as members of a larger family of GN..RA motifs that all share the ability to form tertiary interactions through the recognition of the minor groove of a helix. GN..RA motifs may thus be enlarged to hexa- or heptaloops, or even internal loops if the insertion sequence has the ability to fold itself into a regular hairpin stem-loop structure. Such a motif is indeed found in the Deinococcus radiodurans sequence, which displays an unusually long P9 stem interrupted by a GA...GAA internal loop after 5 bp, i.e. at the exact place of the L9 tetraloop found in most sequences. Accordingly, the G:C pairs that constitute the putative loop receptor are seen in P1 as expected. Other probable members of the GN..RA motif family are the L8 decalooop and the L15.1 loop from type B RNase P RNAs, of respective consensus GAAN,AUAA and GAAN,AA (see Figure 7). The conservation of both ends suggests that these loops are closed by at least one sheared G:A or A:A base-pair, thus extending the helical stacks P8 and P15.1. The Watson-Crick side of at least the 3' adenine base is therefore available for further long-range interaction, as is the case with a canonical GNRA tetraloop. Thus, whereas adenine bases of the L8 decalooop are proposed to interact with the shallow groove of the G:C base-pairs within P4, adenine bases of L15.1 could well be involved in a new type of loop-loop interaction with the hexaloop UGNRAU, L5.1. The higher
degree of conservation within the decaloop L8 is moreover striking, and suggests specific pairings between 5' and 3' ends of the loop. Internal loops of similar sequence GAAN...AUAAG are found also within stem P9 of purple bacteria sequences (Brown et al., 1996) in place of the GNRA loop L9, providing additional evidence for the implication of such motifs in tertiary interactions.

In type A RNase P RNAs, the natural counterpart of the L8 decaloop is a pentaloop of consensus UAAAYR, which is also proposed to interact with the G-C base-pairs in P4. This loop has no obvious similarity with the classical GNRA loop and seems to be much less widespread.

The G...YAA internal loop motif is another interesting motif found at two different locations among RNase P RNAs (see Figure 1(c)). This motif within P10/P11 was shown to recognise specifically the pre-tRNA backbone (Pan et al., 1995; Loria & Pan, 1997). As proposed in our 3D models, the same motif within P12 is potentially involved in the recognition of the conserved heptaloop L13 of consensus GUAAAG. It remains to be seen whether the motif G...YAA is found in other large RNA structures.

Possible compensation of a missing RNA-RNA interaction by the protein component

The conservation of the L8-P4 interaction is of key importance in our models, since it maintains the association of the two main folding domains in a cradle-like conformation. This interaction is also expected to occur in archaea, since the archaenal
consensus core is almost identical with the bacterial core (Haas et al., 1996a). However, the whole stem-loop P8/L8 is absent in both archaeal Methanococcus jannaschii and Archaeoglobus fulgidus. These sequences are, moreover, the only archaeal ones where the whole P16/P17/P6 extension has also been deleted, raising the question of a possible link between the deletion of these structural elements. One can nevertheless speculate that, in both these organisms, the protein component of RNase P has co-evolved in order to compensate for the lack of P8 and/or P6 so as to maintain the active conformation of the RNA. Unfortunately, no RNase P protein from archaea has been fully identified to date (Nieuwlandt et al., 1991; Darr et al., 1992), and no analogue of bacterial (Brown & Pace, 1992) or eukaryal (Kim et al., 1997) proteins is found in complete archaeal genomes (Bult et al., 1996; Klenk et al., 1997; Smith et al., 1997). The absence of a bacterial-like protein in archaeal RNase P is particularly striking when considering the very strong structural homology between archaeal and bacterial RNase P RNA. In eukarya, the secondary structure of RNase P RNA is not yet firmly established (Tranguch & Engelke, 1993), but resembles the bacterial or archaenal consensus in many points (Chen & Pace, 1997). Animal sequences nevertheless seem to lack stems P5 and P15, which may be linked with the emergence of a stem-loop-stem structure within P3. Yeast sequences may be distinguished by their unique structure of the cruciform junction, illustrating once again the modular use of alternative RNA structural motifs and their possible stabilization through either RNA-RNA or protein-RNA interactions.

Conclusion

The modelling of the three-dimensional architecture of group I introns emphasized the hierarchical folding of large RNA molecules (Michel & Westhof, 1990; Jaeger et al., 1996; Westhof et al., 1996a). First, the pairings of the secondary structure join proximate regions in sequence, followed by end-to-end stacking of contiguous helices. Those preformed domains then associate to constitute the compact tertiary structure via interactions between tertiary architectural motifs (for a review, see Brion & Westhof, 1997). The first recurrent 3D motif discovered (Michel & Westhof, 1990) consists of a GNRA tetraloop interacting with two consecutive base-pairs in the shallow groove of an RNA helix. The primordial role played by this motif and its variants in the hierarchical self-assembly process of large RNAs has since been established (Jaeger et al., 1994; Murphy & Cech, 1994; Pley et al., 1994; Costa & Michel, 1995, 1997; Cate et al., 1996a; Costa et al., 1997). The modelling of four complete self-splicing group I introns belonging to four subgroups illustrated the widespread use of two types of long-range RNA-RNA anchors between peripheral domains: GNRA-helix receptor and loop-loop interactions (e.g. see Lehner et al., 1996). The picture that emerged from that study was that non-homologous peripheral domains, characteristic of each subgroup, engage in either one of those two RNA-RNA long-range anchoring interactions and, thereby, contribute to either the stabilization of the helical stems of the catalytic core or the correct positioning of the helical substrate.

The present work extends the view of a modular and hierarchical RNA self-assembly, as seen in the architecture of group I introns, to the bacterial RNase P ribozymes. Again, coaxial stacking of helical stems (one of which contains a pseudoknot) is largely prevalent in the formation of subdomains that create a cradle into which the substrate, the amino acid arm of pre-tRNA, is bound. And, again, the maintenance and assembly of the subdomains is performed by a recurrent and systematic use of GNRA-helix and loop-loop anchoring contacts. Finally, again, different and non-homologous peripheral elements promote different and mutually exclusive long-range anchoring contacts with identical functional purposes.

Only a fraction of group I introns self-assemble and most of them require the help of protein cofactors for proper folding. A well-described system is the Neurospora crassa mitochondrial tyrosyl tRNA synthetase (CYT-18 protein), which binds to the P4-P6 domain of introns belonging to subgroup IA2, like the N. crassa NcLSU intron (Caprara et al., 1996b). Those introns lack the P5abc extension, which forms a self-folding domain (Murphy & Cech, 1993; Cate et al., 1996a) and is situated at the back of the substrate binding site. It has been shown that CYT-18 binds at the same place as P5abc (Caprara et al., 1996a). Similarly, in the present models, the P RNA forms a layer of helices sandwiched between the protein cofactor and the tRNA substrate. One can speculate that a fully autonomous RNase P functioning without a protein cofactor would require the presence of additional peripheral domains for adequate stabilization of the helical core at the back side of the substrate recognition surface.

Methodology

Phylogenetic analysis

The initial alignment of 87 bacterial sequences from the sixth edition of the book of P was first downloaded from the ribonuclease P database (Brown, 1997). New sequences were added and manually aligned using SeqPup (Don Gilbert, Indiana University) as they were directly brought to publication (J. W. Brown, personal communication; and Haas et al., 1996b; Herrmann et al., 1996; Vioque, 1997) or derived from genome sequencing projects (Fleischmann et al., 1995; Fraser et al., 1995; Himmelreich et al., 1996; Tomb et al., 1997). Although the sequencing of the complete gen...
ome of Neisseria gonorrhoeae, Streptococcus pyogenes, (Roe et al., 1997, http://dna1.chem.uoknor.edu. Neisseria meningitidis, Vibrio cholerae (TIGR, 1997, http://www.tigr.org/tdb/mdb/mdb.html) and Treponema pallidum (Weinstock & Norris, 1997, http://utmmg.med.uth.tmc.edu/treponema/t pall.html) are still under progress, their unique RNase P RNA gene (rnpB) has been identified by BLAST searches (Altschul et al., 1997). These sequences contain the whole set of conserved nucleotides that are characteristic of bacterial RNase P RNA. They display therefore a secondary structure and have been added to the alignment. Sequences were then classified by phylogeny before being sought for finer alignment within the most variable stems (P3, P6, P9, P12, P16, P17, P19, P10.1) and within single-stranded regions (L11/12, L15/16, L8, L15.1). For clarity, references and access numbers of sequences of particular interest are grouped in Table 7. Throughout the text, any base-pair is denoted by a colon. Residues involved in internal loops are separated by three dots. Other standard abbreviations are: N, any nucleotide; R, A or G; Y, U or C; H is A, C or U. Nucleotides in bold letters represent invariant nucleotides.

Three-dimensional modelling

Molecular modelling was performed as described (Michel & Westhof, 1990; Westhof, 1993) using the program MANIP (unpublished) implemented on an SGI workstation. The 3D structures were geometrically refined with the restrained least-squares program NUCLIN-NUCLSQ (Westhof et al., 1985). Drawings were produced by DRAWNA (Massire et al., 1994) on an SGI workstation. Coordinates are available by anonymous ftp (130.79.17.244).

Acknowledgements

We are grateful to Dr J. W. Brown for communicating RNase P RNA sequences from green non-sulphur bacteria prior to publication, and more generally for his involvement in the creation and the updating of the RNase P database (Brown, 1997). We thank Dr S. Norris for the identification of the rmpB gene in the Treponema pallidum genome. C.M. acknowledges support from the Ministere de l’education nationale, de l’enseignement superieur, de l’/recherche et de l’insertion professionnelle and the Human Frontier Science Program (CRG0291/1997-M to E.W.). E.W. thanks the “Institut Universitaire de France” for support.

References


Siegel, R. W., Banta, A. B., Haas, E. S., Brown, J. W. & Pace, N. R. (1996). Mycoplasma fermentans simplifies...
our view of the catalytic core of ribonuclease P RNA. RNA, 2, 452–462.


(Received 26 November 1997; received in revised form 23 March 1998; accepted 26 March 1998)