RNA (1997), 3:553–556. Cambridge University Press. Printed in the USA.

LETTER TO THE EDITOR

Phylogenetic evidence for a new tertiary interaction in bacterial RNase P RNAs

CHRISTIAN MASSIRE, LUC JAEGER, and ERIC WESTHOF

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

Long-range interactions involving terminal hairpin loops are important for the self-assembly process of large RNA molecules into their active conformation (Lehnert et al., 1996; Westhof et al., 1996a). Interactions between GNRA tetraloops and the shallow groove of RNA helices, which are found in all three classes of large catalytic RNAs, are one of the most widespread long-range structural motifs (Michel & Westhof, 1990; Costa et al., 1997). In ribonuclease P RNA, three of the four well-conserved GNRA loops are involved in such interactions: in most bacterial sequences, GYRA loops L14 and L18 recognize adjacent base pairs in P8 (Brown et al., 1996). In *Bacillus*-like RNase P RNA sequences, the GAAA loop L12 interacts with an 11-nt motif in stem P10.1a (Tanner & Cech, 1995), resulting in a highly specific interaction similar to that found in group I introns (Costa & Michel, 1995; Cate et al., 1996). Here, we report phylogenetic evidence for an interaction involving the last well-conserved GNRA loop, L9, with the terminal helix P1, and its replacement by a pseudoknot in some *Mycoplasma* species.

For more than half of the 41 complete bacterial sequences from the RNAse P database (Brown, 1997), the presence of a consensus stem P9 of 5 bp plus a bulge capped by a GNRA tetraloop appears to be correlated with the presence of a G/C at the eighth base pair of stem P1 (bp 4/370 in *Escherichia coli* numbering, see Fig. 1). By contrast, there is no constraint in the sequence of P1 when the loop L9 or/and the length of stem P9 vary (Table 1). A further indication for this tertiary contact comes from the covariation between the third nucleotide of the GNRA tetraloop L9 and the ninth base pair of stem P1 (bp 3/371, see Table 2). A GNGA loop implies an A/U base pair at positions 3/371, whereas a GYAA loop allows both A/U and G/C pairs. This pattern of covariation has been observed previously for the L14-P8 and L18-P8 tertiary interactions within the bacterial RNAse P RNAs (Brown et al., 1996), as well as for long-range interactions within group I and group II introns (Michel & Westhof, 1990; Jaeger et al., 1994; Costa & Michel, 1995).

Some of the remaining complete sequences display other interesting features. Whereas most sequences within the homogeneous γ subdivision of the purple bacteria possess a L9 of either type GYAA or GYGA, sequences from *E. coli* (Reed et al., 1982) and *Salmonella thyphimurium* (Baer & Altman, 1985) have a GAAA loop on top of a P9 shortened by one base pair, and the closely related sequences from *Klebsiella pneumoniae* and *Erwinia agglomerulans* (Lawrence et al., 1987) have a longer loop UGUCACAG. The occurrence of these loops seems to be related to the presence of A/U and G/U base pairs at positions 3/371 and 4/370, respectively. However, the geometry of these interactions, if they indeed occur, remains unclear. The RNase P RNA from *Deinococcus radiodurans* (Haas et al., 1991) presents a longer stem P9, with an asymmetrical internal loop GA...GAA in lieu of the GNRA loop. Interestingly, this sequence possesses two adjacent G/C pairs at positions 3/371 and 4/370, suggesting that its internal loop could interact in the same manner as GYAA terminal loops. Formation of adjacent, sheared purine/purine pairs, as they appear in the internal loop J4/5 of group I introns (Cate et al., 1996), would indeed leave the adjacent adenosines in a conformation similar to the one found in GYAA loops.

Finally, sequences from *Mycoplasma genitalium* (Fraser et al., 1995) and *Mycoplasma pneumoniae* (Himmelreich et al., 1996) present a single strand J1/20 between the 3′ end of a shorter P1 and the nonconserved stem-loop P20. These sequences are, moreover, the only ones in which L9 is a 7- or 8-nt loop instead of the classical

Reprint requests to: E. Westhof, UPR 9002, Institut de Biologie Moléculaire et Cellulaire du CNRS, 15 rue Descartes, 67084 Strasbourg cedex, France; e-mail: westhof@ibmc.u-strasbg.fr.

```
                    P1          P9          L9          P9'         P1'
GNGA loops
R.rubrum    ----CCAGUCGGCC-G[]C-GGGG------GCGA------CCCUAG[]CAGGCCGUCUGGCGGU
Ag.tumefac  ----CCAGUUGGCC-G[]C-GGGG------GCGA------CCCCAG[]CAGGCCAACUGGCGAA
H.influenz  ----CGAGUUGGUU-A[]G-CGGC------GUGA------GCCGAC[]UCGACCAACUCAAAAC
Ps.fluores  ----AGAGUCGAUU-G[]G-GGGC------GUGA------GCCUAC[]UAGAUCGACUCUCCAC
Chr.vinosm  ----GGAGUCGGCC-A[]G-GGGC------GAGA------GUCCAC[]UCGGCCGACUCCCUUC
Thb.ferrox  ----GGAGUGGGCC-A[]G-GGGC------GUGA------GCCUAC[]UCGGCCCACUCCAAUU
Bac.thetai  UAAAGCAGUUGGCC-G[]CUGUCC------GCGA------GGGUAG[]UAGGCUGACUGCUUUU
A.nidulans  GGGGAAAGGAGGCG-A[]U-GCGG------GUGA------CCGUGA[]U-GUCCUGCUUUCCCU
Anbn.7120   GGAGAGAGUAGGCGUU[]U-GCGA------GCGA------UCGUGA[]C-CACCAACUCUCUCC
Ctx.7601    GGAGAGAGUAGGCGUA[]U-GCGA------GCGA------UCGUGA[]C-UACCAACUCUUUCC
Sncy.6803   ---GAGAGUUAGGG-A[]U-GCGC------GCGA------GCGUGA[]C-GACUAACUCUCUUU
Tt.maritim  ----GGAGAGGAGC-A[]G-AGGG------GUGA------CCCU-C[]U-GCUCCUCUCCCGUG
Tt.neapoli  ----GGAGAGGGGU-A[]G-AGGG------GUGA------CCCU-C[]U-ACCCCUCUCCCGUG
Stm.bikini  ----CGAGCCGGGC-G[]C-CGGG------GUGA------CCCGCG[]CAGCCCGACUCGUCUG
Stm.livida  ----CGAGCCGGGC-G[]C-CGGG------GUGA------CCCGCG[]CAGCCCGACUCGUCUG
M.tubercul  CGGAUGAGUUGGCG-G[]C-CGAG------GUGA------CUCGCG[]CAGGCCAACUCGUCCG
M.leprae    CGGACGAGUUGGCU-G[]C-CGAG------GUGA------CUCGCG[]UAGGCCAACUCGCCCG

GYAA loops
Alc.eutrop  ----AAAGCAGGCC-A[]C-CACG------GCAA------CGUGCG[]UCGGCCUGCUUUGCUU
Ser.marces  ----GGAGUUGACC-A[]G-AGGC------GCAA------GCCUAC[]CCGGUCAACUCCCUC-
Dsv.desulf  ----GGAGUCGGAC-G[]G-GAGG------GCAA------CCUC-C[]CAGUCCGACUCCCGCA
Psanb.6903  -GAAAGAGUAAG-C-G[]U-GCGA------GCAA------UCGUGA[]U-GUCUUACUCUUUCU
T.aquaticu  GCCCCAGGAUAGGG-G[]G-CGGG------GCAA------CCCGAC[]CGCCCUAUCCUGGGGA
T.thmophl   GGCCCGGGACGAGG-G[]G-CGGG------GUAA------CCCGAC[]CGCCUCGUCCCGGAGG

Other putative interacting loops
E.coli      ----GAAGCUGACC-A[]G-GGG-------GAAA-------CCCAC[]UCGGUCAGUUUCACCU
S.typhimur  ----GAAGCUGACC-A[]G-GGG-------GAAA-------CCCAC[]UCGGUCAGUUUCACCU
K.pneumoni  ----GAAGCUGACC-A[]G-GGG-----UGUCACAG-----CCCAC[]UCGGUCAGUUUCACCU
Er.agglome  ----GAAGCUGACC-A[]G-GGG-----UGUCACAG-----CCCAC[]UCGGUCAGUUUCACCU
D.radiodur  GAACGCGGGGAAAC-U[]U-CGGC-GA[42]GAA------GCCGAA[]CCGUUUCCCCGUGCCA

Loops involved in P21
M.genitali  --------GCUGUC-G[]G-GAUG-----UGAUAGC----CAUAAC[]UUGACAGCAUGCUAUC
M.pneumoni  -------CGCUGUC-G[]G-GACG----CUAAUAGA----CGUCAC[]UUGGCAGCGUUCUAUU

Other loops
Bor.burgdo  ----AGCUGGCAGC-U[]G-GGUU-------UUA------AACCUA[]UGGGUGUCAGCUUAAU
Chl.limico  ---AAACCGCAAGU-G[]G-GGCAG-CGGUGCAAACCGUCUGUCAC[]CAGCUU-CGGUUUCAG
B.brevis    -AUGCAGGAAAUGC-G[]G-CAAGUGAGGCGCAAGCCUCGC-UGAC[]UAGCAUUUCCUGCUGG
B.subtilis  -GUUCUUAACGUUC-G[]G-CAGUCU----UUAG----AGGCUGAC[]CAGAACGUUAGACCAC
B.stearoth  -GUUAAUCAUGCUC-G[]G-CAGCCUGGC-UUCG-GCUGGGCUGAC[]UAGAGCAUGAUUAACG
B.megateri  -UGAAAUAACGUUC-A[]G-CAGCUUGGCU-UAUAGCUUAGCUGAC[]ACGAACGUUGUUGAAA
S.pyogenes  -CUGUGCAAUUUUU-G[]G-GAUGUG----CAUA----CACAUUAC[]UAGAAAAUUGCAUAAU
M.capricol  ---GAGAUUUUAUU-G[]G-CAGUAA----UUAA----UUACUG-C[]UAGAUAAAAUCUCAUC
M.fermanta  --------AGCAUU-G[]G-C----------UUA---------GAC[]UAAAUGCU--------
M.floccula  ---------GCAC-AA[]G-C----------UUA---------GAC[]UUAGUGC---------
M.hyopneum  --------UGCAC-AA[]G-C---------UUUA---------GAC[]UUAGUGCAA-------
```

FIGURE 1. Partial alignment of the 41 complete bacterial sequences from the RNase P database (Brown, 1997). Sequences are classified according to the type of loop L9. Names of secondary structure components are indicated on top of the alignment (Pn designates the 5′ branch of a pairing and Pn′, its 3′ branch). Canonical pairings are green within P1 and red within P9. Bulges, gaps, mismatches, and unpaired bases are black. Nucleotides 3 and 4 in P1, 113 and 114 in L9, 370 and 371 in P1′ are blue when involved in the interaction L9-P1. Underlined nucleotides participate in the pseudoknot P21. Squared brackets stand for segments of sequences not included in the alignment.

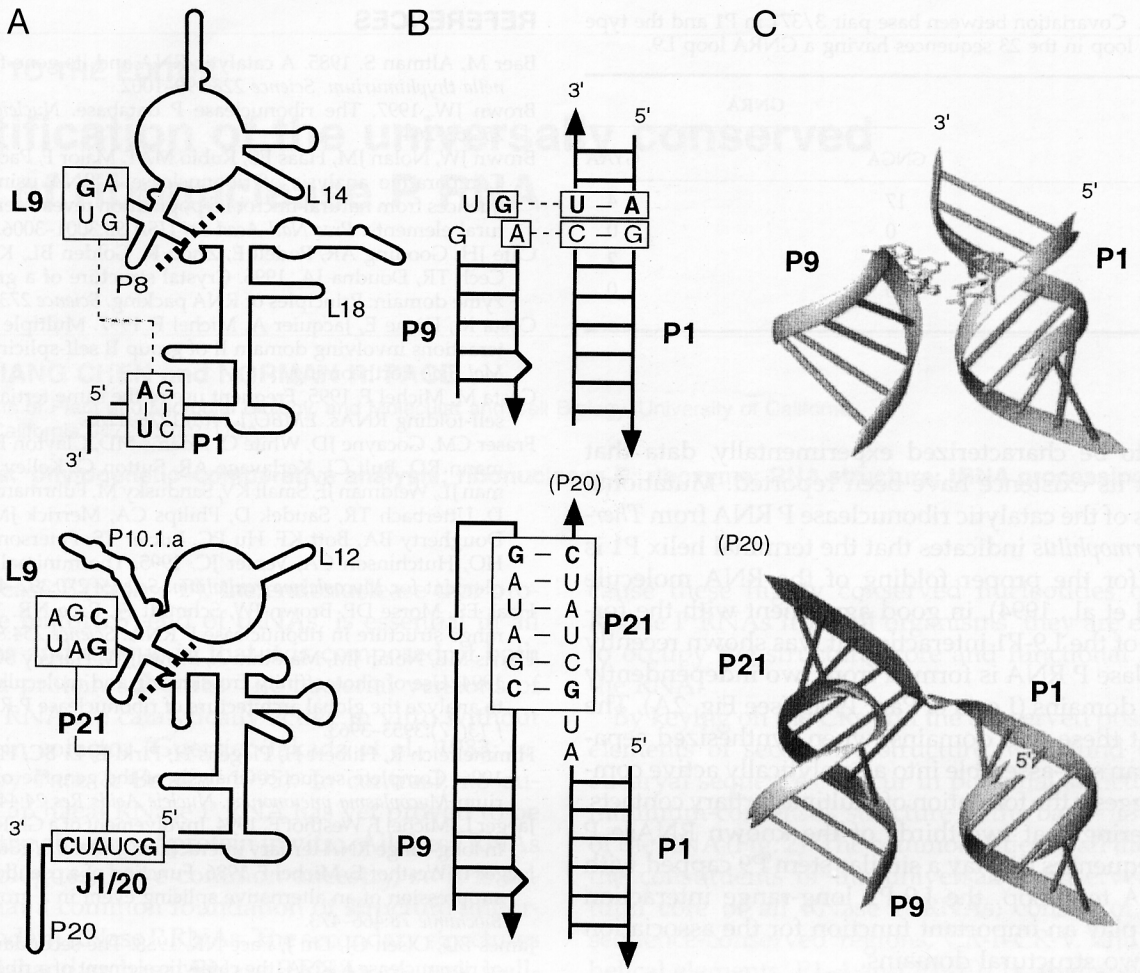

FIGURE 2. **A:** Top: Secondary structure of the RNase P RNA from *Pseudomonas fluorescens* (James et al., 1988). Bottom: Secondary structure of the *M. genitalium* sequence (Fraser et al., 1995). Nucleotides involved in the tertiary interactions are boxed and nucleotides showing covariations are bold. Thin lines indicate tertiary contacts. Thick dotted lines indicate the separation between the two folding domains (Loria & Pan, 1996). **B:** Top: Diagram of the loop–helix interaction. The G-U/A base-triple may be replaced by either an A-C/G or an A-U/A base-triples. Bottom: diagram of the pseudoknot P21 from the *M. genitalium* sequence. Both first and last base pairs are replaced by an A/U base pair in *M. pneumoniae*. **C:** Top: Ribbon model of the loop–helix interaction. The four purines involved in the tertiary interaction are represented as sticks. Bottom: Ribbon model of the pseudoknot P21. The stacking of P9 and P21 is possible while keeping a similar orientation for P1 and P9.

GNRA, although the length of P9 remains the same. Strikingly, in both sequences, L9 and J1/20 are complementary along 6 nt, with covariation on the first and last positions. This new putative pseudoknot, P21, may occur in place of the GNRA–helix interaction encountered in the other sequences (see Fig. 2). Both interactions have been modeled with optimal geometry (Westhof, 1993) and allow similar orientations of stems P1 and P9 with respect to each other (see Fig. 2C).

A similar motif swap has been engineered already in group I introns, where the GNRA–helix interaction between loop L9 and helix P5 could be replaced successfully by a pseudoknot (Jaeger et al., 1994). In fact, the natural counterpart of this pseudoknot was found in a group I intron, the Sd.Cob,1 intron (Jaeger et al., 1996). Although the putative long-range interaction L9-P1 re-

TABLE 1. Correlation between base pair 4/370 in P1 and the type of loop L9 in the 41 complete bacterial sequences from the RNase P database.[a]

	L9	
bp 4/370	GNRA	Other loops
A/U	0	3
C/G	0	3
G/C	23	2
G/U	0	4
U/A	0	2
−/−	0	4

[a]Column GNRA refers to sequences having a GNRA loop on top of a consensus stem P9 of 5-bp plus a bulge. Data were collected using the program COSEQ (C. Massire & E. Westhof, in prep.).

TABLE 2. Covariation between base pair 3/371 in P1 and the type of GNRA loop in the 23 sequences having a GNRA loop L9.

	GNRA	
bp 3/371	GNGA	GYAA
A/U	17	4
C/G	0	0
G/C	0	2
U/A	0	0

mains to be characterized experimentally, data that support its existence have been reported. Mutational analysis of the catalytic ribonuclease P RNA from *Thermus thermophilus* indicates that the terminal helix P1 is crucial for the proper folding of the RNA molecule (Schlegl et al., 1994), in good agreement with the formation of the L9-P1 interaction. It was shown recently that RNase P RNA is formed from two independently folding domains (Loria & Pan, 1996) (see Fig. 2A). The fact that these two domains, when synthesized separately, can self-assemble into a catalytically active complex suggests the formation of multiple tertiary contacts. Considering that two-thirds of the known RNAse P RNA sequences display a similar stem P9 capped with a GNRA tetraloop, the L9-P1 long-range interaction should play an important function for the association of the two structural domains.

This new interaction, if confirmed experimentally, should be important for refining the two existing three-dimensional models of the RNAse P catalytic subunit (Harris et al., 1994; Westhof et al., 1996b). Whereas both models cannot accommodate the new interaction, their reorganization, necessary for including the L9-P1 constraints, will be more extensive in the Westhof and Altman model (C. Massire, L. Jaeger, & E. Westhof, in prep.) than in the Harris and Pace model. More generally, the L9-P1 interaction illustrates a modular view and use of different tertiary interactions as a common feature in the self-assembly of RNA molecules.

ACKNOWLEDGMENTS

We are deeply grateful to Dr. J.W. Brown for making available his compilation of RNase P RNA sequences and secondary structures through his constantly updated World Wide Web site http://jwbrown.mbio.ncsu.edu/RNaseP/. C.M. is supported by a fellowship from the Ministère de l'éducation nationale, de l'enseignement supérieur, de la recherche et de l'insertion professionnelle.

REFERENCES

Baer M, Altman S. 1985. A catalytic RNA and its gene from *Salmonella thyphimurium*. *Science 228*:999–1002.

Brown JW. 1997. The ribonuclease P database. *Nucleic Acids Res 25*:263–264.

Brown JW, Nolan JM, Haas ES, Rubio MAT, Major F, Pace NR. 1996. Comparative analysis of ribonuclease P RNA using gene sequences from natural microbial population reveals tertiary structural elements. *Proc Natl Acad Sci USA 93*:3001–3006.

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science 273*:1678–1685.

Costa M, Dème E, Jacquier A, Michel F. 1997. Multiple tertiary interactions involving domain II of group II self-splicing introns. *J Mol Biol*. Forthcoming.

Costa M, Michel F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J 14*(6):1276–1285.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, Fritchman JL, Weidman JF, Small KV, Sandusky M, Fuhrmann J, Nguyen D, Utterbach TR, Saudek D, Philips CA, Merrick JM, Tomb JF, Dougherty BA, Bott KF, Hu PC, Lucier TS, Peterson SN, Smith HO, Hutchinson CA, Venter JC. 1995. The minimal gene complement for *Mycoplasma genitalium*. *Science 270*:397–403.

Haas ES, Morse DP, Brown JW, Schmidt FJ, Pace NR. 1991. Long-range structure in ribonuclease P RNA. *Science 254*:853–856.

Harris ME, Nolan JM, Malothra A, Brown JW, Harvey SC, Pace NR. 1994. Use of photoaffinity crosslinking and molecular modeling to analyze the global architecture of ribonuclease P RNA. *EMBO J 13*(17):3953–3963.

Himmelreich R, Hilbert H, Plagens H, Pirkl E, Li BC, Herrmann R. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res 24*:4420–4449.

Jaeger L, Michel F, Westhof E. 1994. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J Mol Biol 236*:1271–1276.

Jaeger L, Westhof E, Michel F. 1996. Function of a pseudoknot in the suppression of an alternative splicing event in a group I intron. *Biochimie 78*:466–473.

James BD, Olsen GJ, Liu J, Pace NR. 1988. The secondary structure of ribonuclease P RNA, the catalytic element of a ribonucleoprotein enzyme. *Cell 52*:19–26.

Lawrence NP, Richman A, Amini R, Altman S. 1987. Heterologous enzyme function in *Escherichia coli* and the selection of genes encoding the RNA subunit of RNase P. *Proc Natl Acad Sci USA 84*:6825–6829.

Lehnert V, Jaeger L, Michel F, Westhof E. 1996. New loop–loop tertiary interactions in self-splicing introns of subgroups IC and ID: A complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem & Biol 3*:993–1009.

Loria A, Pan T. 1996. Domain structure of the ribozyme from eubacterial ribonuclease P. *RNA 2*:551–563.

Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of Group I catalytic introns based on comparative sequence analysis. *J Mol Biol 216*:585–610.

Reed RE, Baer MF, Guerrier-Takeda C, Donnis-Keller H, Altman S. 1982. Nucleotide sequence of the gene encoding the RNA subunit (M1 RNA) of ribonuclease P from *Escherichia coli*. *Cell 30*:627–636.

Schlegl J, Hardt WD, Erdmann VA, Hartmann RK. 1994. Contribution of structural elements to *Thermus thermophilus* ribonuclease P RNA function. *EMBO J 13*(20):4863–4869.

Tanner MA, Cech TR. 1995. An important RNA tertiary interaction of group I and group II introns is implicated in Gram-positive RNase P RNAs. *RNA 1*:349–350.

Westhof E. 1993. Modeling the three-dimensional structure of ribonucleic acids. *J Mol Struct 286*:203–210.

Westhof E, Masquida B, Jaeger L. 1996a. RNA tectonics: Towards RNA design. *Folding & Design 1*:R78–R88.

Westhof E, Wesolowski D, Altman S. 1996b. Mapping in three dimensions of regions in a catalytic RNA protected from attack by an Fe(II)-EDTA reagent. *J Mol Biol 258*:600–613.