Nucleic Acids Research, Vol. 20, No. 17 ***4397–4400***

Conformity of RNAs that interact with tetranucleotide loop binding proteins

Christian Zwieb
Department of Molecular Biology, The University of Texas Health Science Center at Tyler, PO Box 2003, Tyler, TX 75710, USA

Received May 14, 1992; Revised and Accepted August 1, 1992

ABSTRACT

A group of RNA binding proteins, termed tetraloop binding proteins, includes ribosomal protein S15 and protein SRP19 of signal recognition particle. They are primary RNA binding proteins, recognize RNA tetranucleotide loops with a GNAR consensus motif, and require a helical region located adjacent to the tetraloop. Closely related RNA structures that fit these criteria appear in helix 6 of SRP RNA, in helices 22 and 23A of 16 S ribosomal RNA, and, as a pseudoknot, in the regulatory region of the *rpsO* gene.

Tetraloops and protein binding

RNA molecules commonly associate with proteins to form multi-component complexes dedicated to a wide variety of essential cellular functions (1). With the exception of a few tRNAs, little is known about the precise three-dimensional structure of RNA, making it difficult to determine molecular details of protein RNA-interactions. Evolutionary conserved motifs are useful guides to structural relationships, and these components help define structure and function of protein-RNA complexes. This paper describes the RNA binding sites of a group of proteins that recognize RNA tetranucleotide loops (tetraloops) with the consensus motif GNAR.

Tetraloops occur frequently in RNA molecules (1–6). They tend to exhibit either an UNCG- or GNRA-motif where N is any nucleotide and R is either G or A (7). Both types of tetraloops were studied by NMR, showing that they differ significantly in their overall three-dimensional structure, but contain an unusual base pair for increased thermodynamic stability (8, 9). Because UNCG and GNRA exchange in evolution, the major role of tetraloops appears to be the formation of stable RNA helices (7,10). However, it was shown recently that a protein in signal recognition particle (SRP19) recognizes the helix 6-tetraloop of SRP RNA in a sequence specific manner, demonstrating that tetraloops have more specific functions (11).

Comparative sequence analysis of more than thirty SRP RNA sequences identifies the helix 6-tetraloop motif as GN*AR*, which is slightly different from the GN*RA*-motif commonly found in the ribosomal RNAs (6). Moreover, in SRP RNA, the GNAR does not alternate in evolution with the dissimilar UNCG-motif. The variable base at the second tetraloop position is freely accessible but it is not used for SRP19 binding, whereas the conserved adenosine at the third tetraloop-position is essential for the interaction with SRP19. Numerous base pairs in the stem of helix 6, where SRP19 prefers one RNA-strand, are required (11).

The GNAR-tetraloop located at position 727 in 16 S rRNA (*E.coli* numbering) is unique in that it is part of the binding site of ribosomal protein S15 (Table 1). The protein protects nucleotides G727, A729 and G730, but not A728, from chemical modification (12), indicating that S15, like SRP19, interacts not with the nucleotide at the second, but rather with nucleotides at the first, third, and forth tetraloop position.

Primary sequence comparison of SRP19 with ribosomal protein S15 and its homologs, reveals a weak, but significant, similarity (Figure 1). The similarity covers nearly all of ribosomal protein S15 and excludes about twenty-five amino acids from each terminus of SRP19 (11). It is in agreement with this alignment that the twenty-four amino acids from the C-terminus of SRP19 are not required for binding to SRP RNA (13,14). Although the proteins are relatively small (*E.coli* ribosomal protein S15: Mr 10, 139; human SRP19: Mr 16,157), they seem to bind to more than one helical turn of RNA, which appears to involve a large portion of the region that is shared between SRP19 and S15 (11). This is probably why it is difficult to discern a short protein binding motif. However, hydropathy plots of the tetraloop binding proteins reveal a common hydrophobic region between amino acids at position 51 and 61 (human SRP19 numbering) followed by a hydrophilic region of the amino acids between positions 63 and 79 (not shown). The conserved amino acids found in this domain may play an important role in protein folding or in the interaction with the RNA (Figure 1).

S15 and SRP19 are related also on the functional level. Both are primary RNA binding proteins, in the sense that they do not require other proteins for their interaction with RNA. SRP19 promotes binding of SRP54 to SRP RNA (15), whereas S15 enables ribosomal proteins S6 and S18 to interact with 16 S rRNA (16). In both cases, binding appears to be mediated by conformational changes in the RNA and not by protein-protein interactions (12,17–20).

A mutation in *E.coli* ribosomal protein S15 (*secC*) suppresses mutations in genes that code for components of the protein export machinery (21). Although the *secC* mutation may affect protein export indirectly by altering the rate of protein synthesis (22,23), it is interesting to find that ribosomal protein S15 is similar to protein SRP19, since SRP19 is part of the machinery that targets

secretory proteins. The intriguing possibility that the ribosome plays an active role in the early stages of protein targeting was considered previously (24).

SRP19 and S15 recognize RNAs with similar structures

The structural and functional similarities between S15 and SRP19 suggest that both proteins recognize RNA sites with related structures. The RNA binding site of SRP19 consists of a tetraloop with the consensus motif GNAR, and several adjacent base pairs in the corresponding stem (Figure 2). Site-directed mutagenesis of helix 6 in SRP RNA showed that the second tetraloop base is dispensable for the interaction with SRP19, and that one RNA strand is preferred in the distal part of the helix. A second minor binding site is located in helix 8 of SRP RNA, supporting the idea that SRP19 is sandwiched between helices 6 and 8 (11).

The binding site of ribosomal protein S15 was characterized by protection of the 16 S rRNA toward chemical and enzymatic probes (12,25). The degree of dimethyl sulfate modification of A728 is unchanged, confirming that the second tetraloop base is not part of the S15 binding site, whereas the bases at the first, third, and fourth tetraloop positions become less accessible (12). Since S15 protects only one of the RNA strands in helix 22, the protein appears to lie parallel to helix 22. This configuration is similar to that of SRP19 as it interacts with the SRP RNA. The structure of the RNA in the binding site of S15 would be virtually identical to the binding site of SRP19, if, in the presence of S15, helix 23A stacks on helix 22. The appearance of a new S15-dependent cut by the double-strand specific RNase V1 between G671 and U672 (26) provides evidence for this structure (Figure 2).

A second minor S15 binding site in rRNA is not yet identified, but is suspected by the enhanced reactivity toward chemical and enzymatic probes found in the distal portion of helix 23. An interaction may also occur at the internal loop of helix 20 that shows some sequence similarity with the minor binding site in SRP RNA. S15 would then be sandwiched between helices 23 and 20, which is compatible with the position of the protein in the three-dimensional model of the 30 S ribosomal subunit (27). A third possibility is that the minor binding site of protein S15 is located on 4.5 S RNA in the region that is homologous to SRP RNA (6,28).

Table 1. GNAR-tetraloops in SRP RNA and bacterial 16 S rRNA.

	Position[a]	Loop sequence[b]		Main alternatives	Protein[c]
SRP RNAs	147	GGAG	49%	GAAA 13%, GUAG 9%	SRP19
	197	GGAA	82%	GCUUCA 12%[d]	—
16S rRNAs	159	GAAA	100%		—
	380	GAAA	64%	GCAA 29%, GGAA 5%	—
	727	GAAG	86%	GAAA 12%	S15
	898	GCAA	100%		—
	1450	GCAA	33%	UUCG 15%, GUAA 11%	—
	1516	GGAA	95%		—

[a]SRP RNAs numbering is according to the human sequence; rRNAs are numbered in reference to the 16 S rRNA from *E.coli*. [b]Frequency of tetraloop sequences was calculated using the alignment in the SRP database (28); data for the bacterial 16 S rRNA were taken from reference 7. [c]SRP19 and S15 (11,12) are tetraloop binding proteins; no proteins were found to interact as indicated by the dashes (30). [d]The main alternative sequence forms a hexanucleotide loop that is characteristic for the plant SRP RNAs (28).

Figure 1. Alignment of tetraloop binding protein sequences related to SRP19 and ribosomal protein S15. Sequences of human protein SRP19 (Hom. sap. SRP19 (31)), the SRP19 homolog from *Saccharomyces cerevisiae* (Sac. cer. SEC65 (32)), rat ribosomal protein S13 (Rat. nor. S13 (33)), ribosomal protein S11 from *Halobacterium marismortui* (Hal. mar. S11 (34)), and the homologous ribosomal proteins of *Escherichia coli* (Esc. col. S15 (35,36,37)), *Bacillus stearothermophilus* (Bac. ste. S15 (34)), *Saccharomyces cerevisiae* mitochondria (Sac. cer. mt. S15 (38)), *Nicotiana tabacum* chloroplast (Nic. tab. ch. S15 (39)), *Marchantia polymorpha* chloroplast (Mar. pol. ch. S15 (40), *Zea mays* chloroplast (Zea. may. ch. S15 (41)), *Oryza sativa* chloroplast (Ory. sat. ch. S15 (42)) and *Secale cereale* chloroplast (Sec. cer. ch. S15 (43)). Shaded areas indicate amino acids that are conserved in at least nine of the twelve sequences.

The biosynthesis of ribosomal protein S15 is regulated by a negative feed-back mechanism that involves binding of S15 to its own mRNA. An essential part of this scheme is a structural similarity between the S15 binding sites in the S15-mRNA (*rpsO*) and the 16 S rRNA. Chemical and enzymatic probing of the regulatory region of the *rpsO* mRNA supports two separate helices with two large loops (Figure 2D). This RNA folding pattern is distinctly different from the one found in rRNA or SRP RNA. However, binding of S15 causes an unusual high degree of protection in the mRNA that is compatible with the formation of a pseudoknot (29). In this alternative structure, the Shine-Dalgarno sequence and parts of the initiating AUG are unavailable for translation (Figure 2C). The common features of the S15 and SRP19 binding sites, proposed here, support this scheme, because the pseudoknot generates a single continuously stacked helix. Moreover, a GNAR-loop motif is revealed by including twenty-three nucleotides of the 5'-translated region. Although the naked RNA contains no tetraloop, such a conformation may be induced by binding to protein S15 for increased stability of the complex.

In summary, the binding sites of protein SRP19 on the SRP RNA and of ribosomal protein S15 on rRNA and mRNA are similar. Because of the wide distribution of RNA tetraloops, corresponding similarities between the tetraloop binding proteins may have developed independently by convergent evolution. Alternatively, these proteins could be phylogenetically related and may be the first recognized members of a single family of proteins. Detailed comparative analysis of RNA tetraloop binding protein complexes will resolve this question in the future. This will establish the role of tetraloop binding proteins in ribonucleoprotein particle assembly and translation regulation, and increase the knowledge of the molecular interactions that take place between protein and RNA.

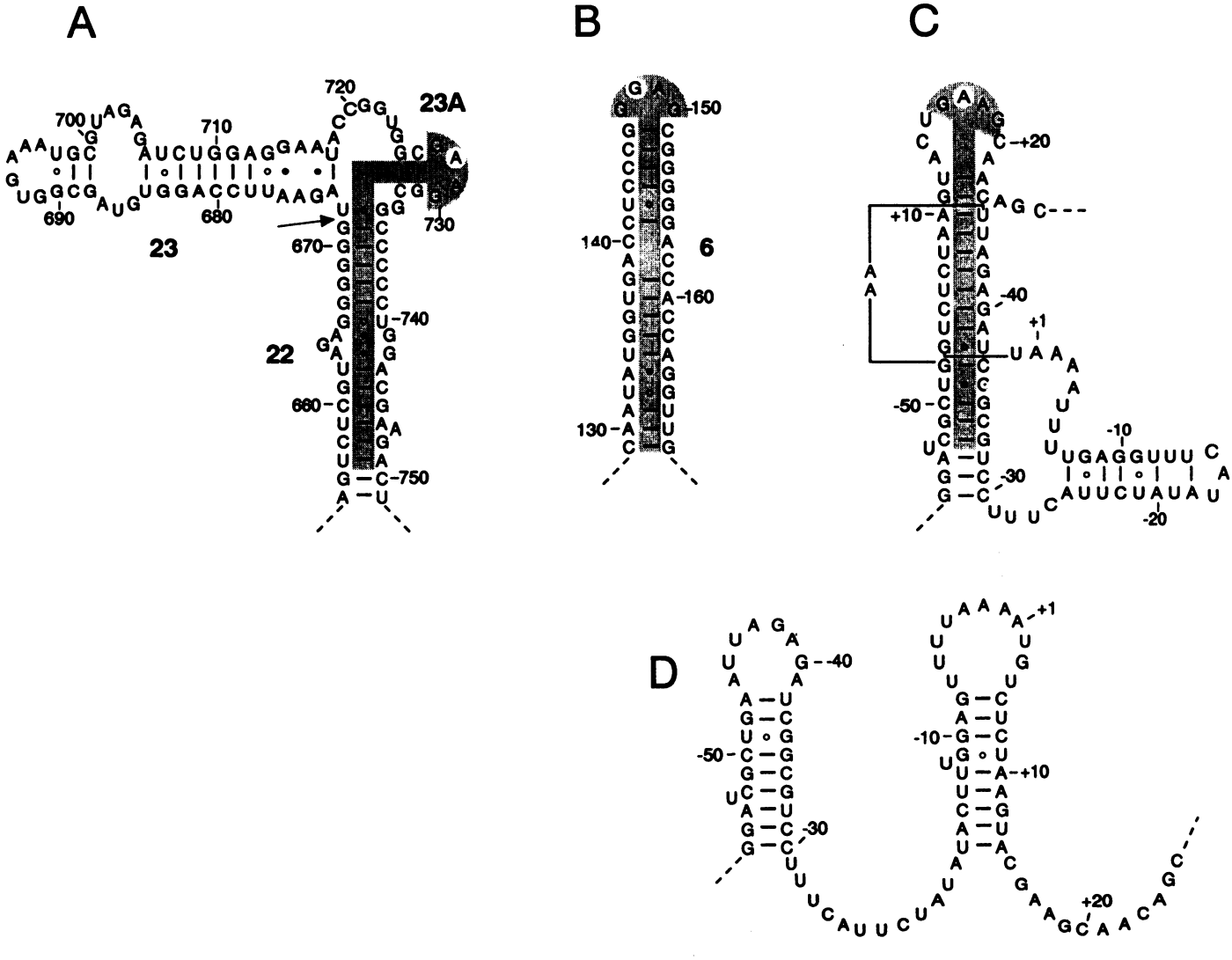

Figure 2. Secondary structures of RNAs that interact with protein SRP19 and ribosomal protein S15. A: Secondary structure of helices 23, 23A and 22 in *E.coli* 16 S rRNA near the S15 binding site; B: Secondary structure of helix 6 of human SRP RNA with the SRP19 binding site; C: Postulated pseudoknot structure of the regulatory region of the *rpsO* mRNA that interacts with S15; D: Open structure of the regulatory region of the *rpsO* mRNA. Helices are numbered according to the nomenclature of Brimacombe et. al., 1988 (rRNA) and Larsen and Zwieb, 1991 (SRP RNA). Watson-Crick pairs are shown by lines, G-U pairs by open circles and A-G pairs by filled circles. Experimentally determined and proposed binding sites of SRP19 (B) and ribosomal protein S15 (A and C) are shown in gray; the second base of the tetraloop-motif GNAR is dispensable for the interaction with protein. The position of an S15-dependent cut by RNase V1 in 16 S ribosomal RNA (26) is marked by the arrow.

ACKNOWLEDGMENTS

I thank Niels Larsen for helpful discussions and Alice Johnson for critical reading of the manuscript. This work was supported, by Grant-in-Aid 91G-556 from the American Heart Association, Texas Affiliate.

REFERENCES

1. Frankel, A.D., Mattaj, I.W., and Rio, D.C. (1991) *Cell* 67, 1041–1046
2. Waring, R.B., and Davies, R.W. (1984) *Gene* 28, 277–291
3. Jacquier, A., and Michel, F. (1987) *Cell* 50, 17–29
4. Pace, N.R., Smith, D. K., Olsen, G.J., and James, B.D. (1989) *Gene* 82, 65–75
5. Adhin ,M.R., Albas, J., and Van Duin J. (1990) *Biochim. Biophys. Acta* 1050, 110–118
6. Larsen, N., and Zwieb, C. (1991) *Nucleic Acids Res.* 19, 209–215
7. Woese, C. R., Winker, S., and Gutell, R. R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 8467–8471
8. Cheong, C., Varani, G., and Tinoco, Jr. I. (1990) *Nature (London)* 346, 680–682
9. Heus, H. A., and Pardi, A. (1991) *Science* 253, 191–194
10. Uhlenbeck, O. C. (1990) *Nature (London)* 346, 613–614
11. Zwieb, C. (1992) *J. Biol. Chem.* in the press
12. Svensson, P., Changchien L., Craven, G. R., and Noller H. F. (1988) *J. Mol. Biol.* 200, 301–308
13. Zwieb, C. (1991) *Biochem. Cell. Biol.* 69, 649–654
14. Rümisch, K., Webb, J., Lingelbach, K., Gausepohl, H., and Dobberstein, B. (1990) *J. Cell Biol.* 111, 1793–1802
15. Walter, P., and Blobel, G. (1983) *Cell* 34, 525–533 (1983)
16. Mizoshima, S., and Nomura, M. (1970) *Nature (London)* 226, 1214–1218
17. Held, W. A., Ballou, B., Mizushima, S., and Nomura, M. (1974) *J. Biol. Chem.* 249, 3103–3111
18. Zwieb, C., and Ullu, E. (1986) *Nucleic Acids Res.* 14, 4639–4657
19. Zwieb, C. (1989) *Prog. Nucleic Acids Res. Mol. Biol.* 37, 207–234
20. Andreazzoli, M., and Gerbi, S. A. (1991) *EMBO J.* 10, 767–777
21. Ferro-Novic, S., Honma, M., and Beckwith, J. (1984) *Cell* 38, 211–217
22. Watanabe, T., Hayashi, S., and Wu, H. C. (1988) *J. Bact.* 170, 4001–4007
23. Schatz, P., and Beckwith, J. (1990) 24, 215–248
24. Ibrahimi, I., and Gentz, R. (1987) *J. Biol. Chem.* 262, 10189–10194
25. Mougel M., Philippe C. , Ebel J.-P., Ehresmann B. and Ehresmann C. (1988) *Nucleic Acids Res.* 16, 2825–2839
26. Mougel M., Eyermann F., Westhof E., Romby P., Expert-Bezanìon A., Ebel J.P., Ehresmann B. and Ehresmann C. (1987) *J. Mol. Biol.* 198, 91–107
27. Brimacombe, R., Atmadja, J., Stiege, W., and Schüler, D. (1988) *J. Mol. Biol.* 199, 115–136
28. Zwieb, C., and Larsen, N.(1992) *Nucleic Acids Res.* Supplement 20, 2207
29. Philippe, C., Portier, C. , Mougel, M., Grunberg-Manago, M., Ebel, J.-P., Ehresmann, B., and Ehresmann C. (1990) *J. Mol. Biol.* 211, 415–426
30. Stern, S., Powers, T., Changchien, L., and Noller, H.F. (1989) *Science* 244, 783–790
31. Lingelbach, K., Zwieb, C., Webb, J.R., Marshallsay, C., Hoben, P.J., Walter, P., and Dobberstein, B. (1988) *Nucleic Acids Res.* 16, 9431–9442
32. Stirling, C.J., and Hewitt E.W. (1992) *Nature* 356, 534–537
33. Suzuki, K., Olvera J., and Wool, I.G. (1990) *Biochem. Biophys. Res. Comm.* 171, 519–524
34. Arndt E., Breithaupt G., and Kimura M. (1986) *FEBS Letters* 194, 227–234
35. Morinaga T., Funatsu G., Funatsu M., and Wittmann H.G. (1976) *FEBS Letters* 64, 307–309
36. Takata, R., Mukai, T., Aoyagi, M., and Hori, K. (1984) *Mol Gen. Genet.* 197, 225–229
37. Regnier P., Grunberg-Manago M., and Portier C. (1987) *J. Biol. Chem.* 262, 63–68
38. Dang H., and Ellis S.R. (1990) *Nucleic Acids Res.* 18, 6895–6901
39. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., and Sugiura, M.(1986) *EMBO J.* 5, 2043–2049
40. Ohyama,K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S.I., Inokuchi, H., and Ozeki, H. (1986) *Nature* 322, 572–574
41. Fitzky, B., and Subramanian, A.R. (1990) *Nucleic Acids Res.* 18, 3407
42. Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.R., Meng, B.Y., Li, Y.Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K., and Sugiura, M. (1989) *Mol. Gen. Genet.* 217, 185–194
43. Prombona, A., and Subramanian, A.R. (1989) *J. Biol. Chem.* 264, 19060–19065