Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3309–3312, June 1983
Biochemistry

Putative introns in tRNA genes of prokaryotes

(intervening sequence/archaebacteria/*Sulfolobus solfataricus*)

BRIAN P. KAINE, RAMESH GUPTA, AND CARL R. WOESE*

Department of Genetics and Development, University of Illinois, 515 Morrill Hall, 505 South Goodwin Avenue, Urbana, Illinois 61801

Communicated by Ralph S. Wolfe, March 7, 1983

ABSTRACT **Sequences of two putative tRNA genes, for serine and leucine, from the archaebacterium *Sulfolobus solfataricus* contain intervening sequences in the anticodon region. Furthermore, the genes lack encoded CCA 3′ termini and are flanked by A+T-rich DNA segments. The introns can both form the same secondary structure, which is a double-helical extension of the anticodon stalk. The resulting structure contains two symmetrically placed 3-base bulge loops, in which are located cleavage sites for the introns. In the one case tested, the gene occurs as a single copy in the genome.**

It is generally believed that intervening sequences (introns) are characteristic of eukaryotic genes (1, 2). Attempts to find them in prokaryotic genes have had uniformly negative results. However, prokaryotes comprise two phylogenetically distinct groups, the eubacteria and the archaebacteria, neither apparently any closer to the other than to the eukaryote (3, 4). The search for introns in prokaryotes to date has been confined exclusively to the eubacteria. We have cloned a number of tRNA genes and gene clusters from several (distantly related) species of archaebacteria to look for introns in that group and have found them. We report here several examples thereof.

MATERIALS AND METHODS

A *Sulfolobus solfataricus* (5) culture was kindly provided by T. A. Langworthy (University of South Dakota). DNA from this organism was a gift of Wolfram Zillig (Max-Planck-Institut für Biochimie, Munich). The organism was grown in a medium containing, per liter, 1.3 g of $(NH_4)_2SO_4$, 0.3 g of KH_2PO_4, 0.3 g of $MgSO_4 \cdot 7H_2O$, 0.1 g of $CaCl_2 \cdot 2H_2O$, and 1.0 g of yeast extract (Difco), pH being adjusted to 3.0 with H_2SO_4. Growth temperature was routinely 70°C.

Restriction enzymes were purchased from New England BioLabs or Bethesda Research Laboratories. The conditions used for digestion were those suggested by the manufacturers.

An *Eco*RI digest of genomic DNA of *S. solfataricus* was randomly cloned into plasmid pBR322 (6). The resulting clones were selected for tRNA structural genes by colony hybridization (7) by using *in vitro* labeled *S. solfataricus* total tRNA (8) as probe. Subfragments of this clone were produced by restriction endonuclease digestion, resolved by electrophoresis on 1% agarose gels, and transferred to nitrocellulose filters (9), and tRNA genes were localized by hybridization with ^{32}P-labeled total tRNA. The fragment to be subjected to sequence analysis was isolated from low-melting temperature agarose gels (10) and ^{32}P-end-labeled at internal restriction sites with T4 polynucleotide kinase (11). This (double-stranded) DNA was cleaved secondarily to produce fragments labeled in one of the strands only, these being reisolated from 5% polyacrylamide gels. Sequence analysis of DNA was done by the chemical method (11, 12).

DNA·DNA hybridization to Southern blots of restriction enzyme digests of genomic DNA was done by published procedures (11).

RESULTS AND DISCUSSION

Fig. 1 is a restriction map of *S. solfataricus* DNA clone of about 7,000 base pairs length. Hybridization of various restriction fragments with labeled *S. solfataricus* total tRNA reveals the clone to contain a number of presumptive tRNA genes. A portion of this clone produced by restriction endonuclease *Hin*dIII was subcloned and more extensively mapped, as shown in Fig. 1. Fig. 2 presents the sequence of the region so indicated in Fig. 1, which appears to contain two tRNA genes.

As judged by their anticodons the two genes encode serine and leucine tRNAs. Both genes contain intervening sequences in their anticodon loops. This can be seen best in Fig. 3, in which each gene is folded into the typical tRNA cloverleaf conformation. Note also that neither gene encodes the tRNA 3′ CCA terminus. Both of these properties were previously thought to be characteristic only of eukaryotic tRNA genes (13–19).

The 103-base-pair fragment that is produced from the serine tRNA gene by restriction endonuclease *Hpa* II (from tRNA positions 3–68, by the numbering convention of ref. 20) was used to probe Southern blots of *S. solfataricus* DNA digested by endonucleases *Eco*RI, *Hin*dIII, *Hin*fI, or *Pst* I. As Fig. 4 shows, each gives a unique hybridization band, indicating that the sequence in question exists as a single copy in the genome.

A direct demonstration that the two genes in question encode functional tRNAs has not been given. Nevertheless, the circumstantial case here is very strong. (*i*) The serine tRNA gene (at least) exists in one copy in the genome. (*ii*) ^{32}P-Labeled total tRNA hybridizes with the *Hin*dIII fragment (see Fig. 1) that contains the two genes in question. (*iii*) Both tRNA genes will fold into perfectly normal tRNA cloverleaf configurations (except for the lack of a CCA 3′ terminus and the additional intervening sequence). (*iv*) Only leucine and serine (and in eubacteria, tyrosine) tRNAs are so-called class II tRNAs—i.e., they possess a large "extra arm"; this is the case for the two genes in question (20). (*v*) Except for the CCA terminus these tRNA genes contain all of the "invariant" and "semi-invariant" residues characteristic of normal tRNAs (21, 22)—i.e., T(U)-8, C(Y)-11, A-14, G(R)-15, G-18, G-19, A-21, G(R)-24, C(Y)-32, T(U)-33, G(R)-37, C(Y)-48, G-53, T-54, T(Ψ)-55, C-56, A(R)-57, A-58, T(Y)-60, and C-61; see ref. 20 for numbering (Y, pyrimidine; R, purine). (*vi*) The serine tRNA gene is specifically related (74% sequence homology) to a *bona fide* serine tRNA (same anti-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed.

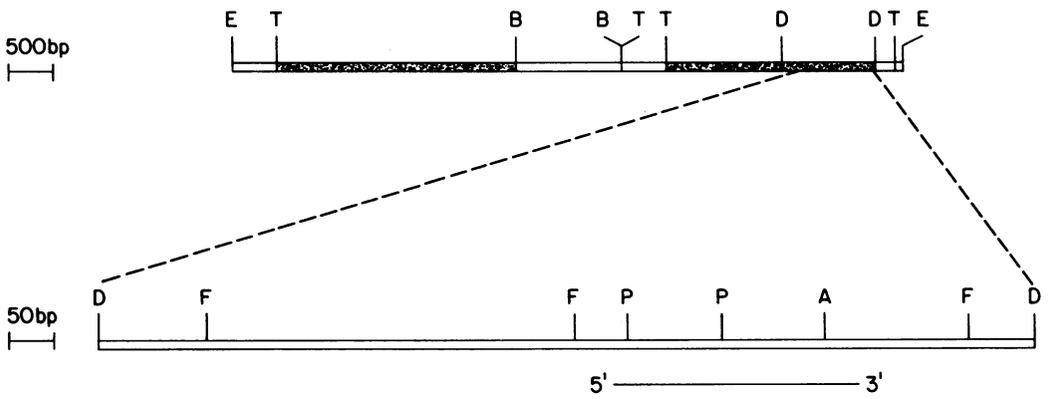

FIG. 1. Restriction map of a *S. solfataricus* clone containing tRNA genes. The upper portion is the complete *Eco*RI-produced fragment and the lower one, the subfragment produced by *Hin*dIII, which contains the tRNA genes in question. Size scales, in terms of base pairs (bp), are given at the left. In the upper portion the shaded areas are those that hybridize with ^{32}P-labeled total tRNA from *S. solfataricus*. Points cut by various restriction endonucleases are designated as follows: A, *Hae* III; B, *Bgl* II; D, *Hin*dIII; E, *Eco*RI; F, *Hin*fI; P, *Hpa* II; and T, *Pst* I. The fragment whose sequence is given in Fig. 2 is underlined in the lower portion.

codon) from a distant archaebacterial relative *Halobacterium volcanii* (20, 23). Also, both of the tRNA genes are flanked by A+T-rich regions, a common characteristic of known tRNA genes (13–19). Given all of these characteristics it would seem highly likely that these are true tRNA genes and not pseudogenes.

The two introns shown in Fig. 3 are not of the same length—that in the serine tRNA gene comprises 25 residues and that in the leucine gene, 15. However, both occur at exactly the same place in the gene, and both can be folded into the same interesting secondary structure in the anticodon arm region. Specifically, the double-stranded stalk of the anticodon arm can be extended (presumably coaxially) by six additional base pairs if the two 3-base bulge loops shown in Fig. 3 are created. (In that the two examples of this structure are not quite identical in sequence, they provide some comparative support for the proposed structure.) In both cases the two cleavage points that would remove the intron are symmetrically located in the two small bulge loops, suggesting perhaps a simultaneous cleavage by a symmetrical (dimeric) processing enzyme. (However, the exact cleavage points cannot be determined because of sequence identities in the loops.) The structure appears to have sequence symmetry as well; note the terminal double C·G pairs defining the ends of the 3-base bulges.

The known examples of intron-containing eukaryotic tRNA genes seem not to form the exact same secondary structure as found here. In most of these cases all three bases of the anticodon can form pairs (13–16), whereas in the present archaebacterial examples the 3′ base of the anticodon is *not* paired (see Fig. 3).

As the molecular organization of archaebacteria becomes increasingly understood, their relationship to eukaryotes becomes increasingly apparent. For example, the two groups share specific features in their translation elongation factors (24) and tRNA sequences (23). Repetitive elements are characteristic of both genome types (25). The tRNA gene characteristics reported here—introns and lack of encoded CCA 3′ termini—are another important addition to this list.

The authors are grateful to Prof. Norman R. Pace for a number of valuable suggestions concerning this work, Prof. W. Zillig for the gift of *S. solfataricus* DNA, Prof. T. Langworthy for a culture of this organism, and W. Weisburg for help in preparing figures. B.P.K. is supported by Damon Runyon–Walter Winchell Cancer Fund Fellowship DRG-543. The work is otherwise supported by National Aeronautics and Space Administration Grant NSG-7044 (to C.R.W.).

```
          I
5' GGTAAATAGTGCCGGGGTGCCCGAGCGGACCAAGGGGGTAGGCTCGAGACCATTTCCAGCGTCAAGGCGGGCGACCTACTGCCTCTCCGA

                                                    TCCGAGCTCTGGTAAAGGTCGCAGTTCCGCCCGCTGGATGACGGAGAGGCT

                                                                                    II
GGCACGCGGGTTCAAATCCCGCCCCCGGCGTTAGTTTAAAATTCTCCAATTTATATCATATATTGATTGCGGGGGTGCCCGAGCAAGGTC

CCGTGCGCCCAAGTTTAGGGCGGGGGCCGCAATCAAATTTTAAGAGGTTAAATATAGTATATAACTAACGCCCCCACGGGCTCGTTCCAG

AAAGGGGTCGGGCTCAGGACCATATGGTGAGACCCCCGATGGTGTAGGCCTGCGTGGGTTCAAATCCCACCCCCCGCAATTTTAGATATG 3'

TTTCCCCAGCCCGAGTCCTGGTATACCACTCTGGGGGCTA
```

FIG. 2. *S. solfataricus* DNA sequence containing tRNA genes. Gene sequences are denoted by overlining and anticodon positions by additional underlining. Both strands of the DNA were subjected to sequence analysis over most of the length. In regions where sequence is only from a single strand, it has been determined independently a number of times. I, putative serine tRNA gene (UCG codon); II, putative leucine tRNA gene (CUG codon).

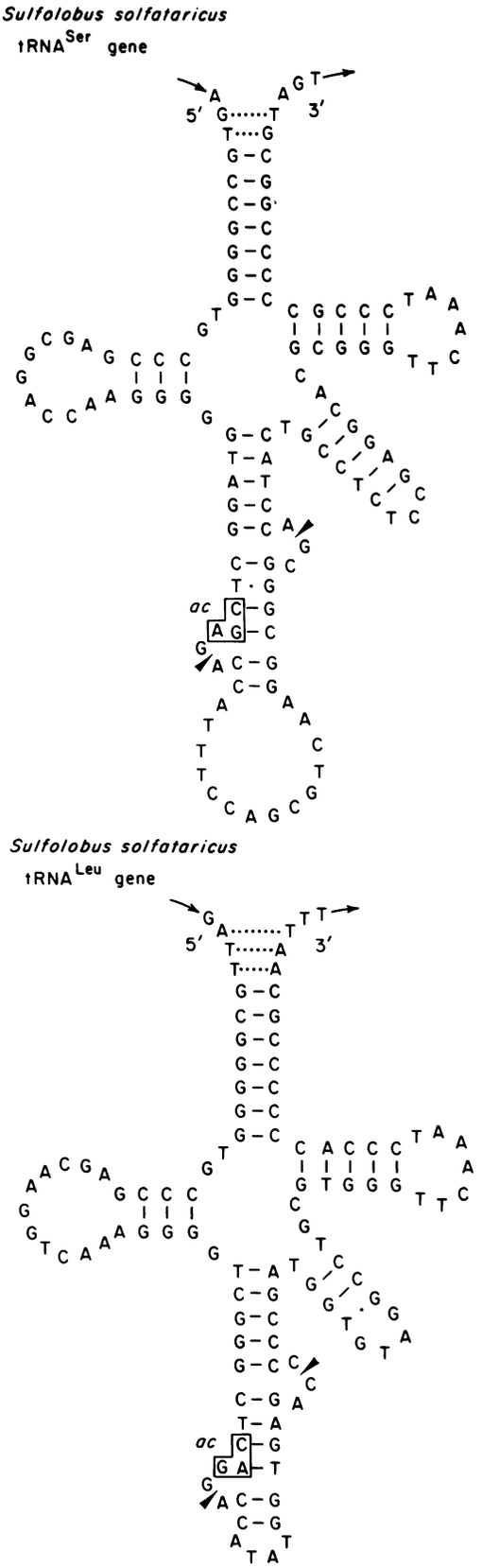

FIG. 3. Serine and leucine tRNA genes folded into the typical "cloverleaf" secondary structure. Anticodon sequences (ac) are boxed. Possible paired extensions of the acceptor stems are indicated by dotted pairings. Arrows indicate possible (symmetric) sites for cleavage of introns and splicing.

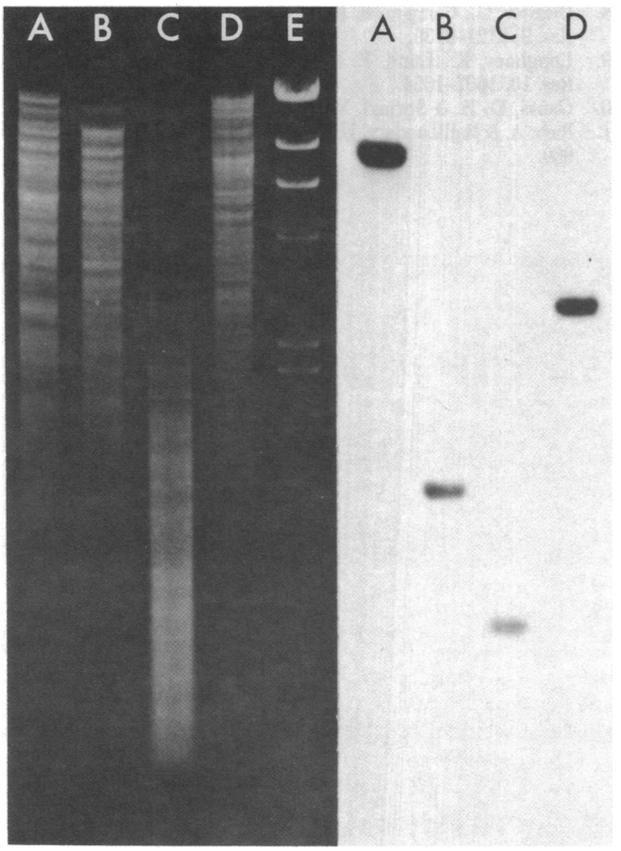

FIG. 4. Electrophoretic pattern (*Left*) and autoradiograph (*Right*) of a Southern blot of *S. solfataricus* genomic DNA cut with various restriction endonucleases and hybridized with the *Hpa* II fragment containing the serine tRNA gene. Lanes: A, *Eco*RI; B, *Hin*dIII; C, *Hin*fI; D, *Pst* I; and E, phage λ DNA digested with *Hin*dIII, as a size standard.

1. Abelson, J. (1979) *Annu. Rev. Biochem.* **48**, 1035–1069.
2. Breathnach, R. & Chambon, P. (1981) *Annu. Rev. Biochem.* **50**, 349–383.
3. Woese, C. R. & Fox, G. E. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5088–5090.
4. Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N. & Woese, C. R. (1980) *Science* **209**, 457–463.
5. Prangishvilli, D., Zillig, W., Gierl, A., Biesert, L. & Holz, I. (1982) *Eur. J. Biochem.* **122**, 471–477.
6. Bolivar, F., Rodriguez, R. L., Greene, P. J., Betlach, M. C., Heyneker, H. L. & Boyer, H. W. (1977) *Gene* **2**, 95–113.
7. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 3961–3965.
8. England, T. E., Bruce, A. G. & Uhlenbeck, O. C. (1980) *Methods Enzymol.* **65**, 65–74.
9. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
10. Wieslander, L. (1979) *Anal. Biochem.* **98**, 305–309.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), pp. 122–123; 312–315; 387–389; 475–477.
12. Maxam, A. M. & Gilbert, W. G. (1980) *Methods Enzymol.* **65**, 499–560.
13. Valenzuela, P., Venegas, A., Weinberg, F., Bishop, R. & Rutter, W. J. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 190–194.
14. Venegas, A., Quiroga, M., Zaldivar, J., Rutter, W. J. & Valenzuela, P. (1979) *J. Biol. Chem.* **254**, 12306–12309.
15. Selker, E. & Yanofsky, C. (1980) *Nucleic Acids Res.* **8**, 1033–1042.
16. Robinson, R. R. & Davidson, N. (1981) *Cell* **23**, 251–259.
17. Nakajima, N., Ozeki, H. & Shimura, Y. (1981) *Cell* **23**, 239–249.

18. Duester, G., Campen, R. K. & Holmes, W. M. (1981) *Nucleic Acids Res.* **9**, 2121–2139.
19. Loughney, K., Lund, E. & Dahlberg, J. E. (1982) *Nucleic Acids Res.* **10**, 1607–1624.
20. Gauss, D. H. & Sprinzl, M. (1983) *Nucleic Acids Res.* **11**, rl–r53.
21. Rich, A. & RajBhandary, U. L. (1976) *Annu. Rev. Biochem.* **45**, 805–860.
22. Kim, S.-H. (1978) in *Transfer RNA*, ed. Altman, S. (MIT Press, Cambridge, MA), pp. 248–293.
23. Gupta, R. (1981) Dissertation (Univ. of Illinois, Urbana).
24. Kessel, M. & Klink, F. (1980) *Nature (London)* **287**, 250–251.
25. Sapienza, C. & Doolittle, W. F. (1982) *Zentralbl. Bakteriol. Mikrobiol. Hyg. I. Abt. Orig. C* **3**, 123–127.