The nucleotide sequence of formylmethionine tRNA from *Mycoplasma mycoides* sp. capriR.T. Walker* and U.L. RajBhandary[†]*Chemistry Department, Birmingham University, Birmingham B15 2TT, UK, and [†]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Received 2 November 1977

ABSTRACT

The nucleotide sequence of *Mycoplasma mycoides* sp. capri PG3 formylmethionine tRNA has been determined, using in vitro labeling techniques, to be pC-G-C-G-G-G-G-s⁴U-A-G-A-G-C-A-G-U-D(U)-G-G-D-A-G-C-U-C-G-C-C-G-G-G-C-U-C-A-U-A-A-C-C-C-G-G-A-G-G-C-C-G-C-A-G-G-U-Ψ-C-G-A-G-U-C-C-U-G-C-C-C-C-G-C-A-A-C-C-A_{OH}. This tRNA contains only three modified nucleosides s⁴U, D and Ψ, all of which are derived from uridine. Both in the structural features which distinguish eukaryotic from prokaryotic initiator RNAs and in the overall sequence, this tRNA resembles a typical prokaryotic initiator tRNA. A comparison of the sequence of this tRNA with those of other prokaryotic initiator tRNAs suggests that taxonomically the *Mycoplasma* may be less related to the Cyanophyta (*Anacystis nidulans*) than to the bacteria and less related to the Enterobacteriaceae (*Escherichia coli*) than to the Bacillaceae (*Bacillus subtilis*).

INTRODUCTION

In recent years it has become apparent that initiator tRNAs possess unique features in their sequence which distinguish them from most non-initiator tRNAs¹. In addition, there is a basic difference in these unique features between initiator tRNAs from prokaryotic and those from eukaryotic cytoplasm². The unique structural feature common to all prokaryotic initiator tRNAs is the lack of a Watson-Crick base pair between the 5'-terminal nucleotide and the fifth nucleotide from the 3'-end³. Eukaryotic initiator tRNAs do not share this feature of prokaryotic initiator tRNAs. Instead, they possess the sequence A-U(or Ψ)-C-G in place of T-Ψ-C-G(or A)- present in loop IV of prokaryotic initiator tRNAs and virtually every tRNA which functions in protein synthesis with the single exception of tRNA^{Ala} of *Bombyx mori*^{4,5}. Although recent work on the se-

quence of N. crassa mitochondrial initiator tRNA² may provide an exception to this clear-cut distinction between the unique features present in the initiator tRNAs which are used as formylmethionyl-tRNA from those used as methionyl-tRNAs, it was of interest to sequence a *Mycoplasma* initiator tRNA in order (a) to confirm that these organisms were indeed prokaryotic in origin and (b) to compare the nucleotide sequence with those of other prokaryotic initiator tRNAs for evaluating any possible evolutionary relationship.

We have previously reported⁶ that, similar to prokaryotic initiator tRNAs, the initiator tRNA of M. capri does not possess the Watson-Crick base pair at the end of the amino acid acceptor stem and that it possesses the sequence U-Ψ-C, which is the precursor of T-Ψ-C in loop IV. Using a variety of in vitro labeling techniques, we now report the total sequence of this initiator tRNA and provide a comparison of its sequence with those of other known prokaryotic initiator tRNAs.

MATERIALS AND METHODS

General. Sources of tRNA, enzymes, radioactive materials, chemicals and techniques were all as described previously⁶, M. mycoides sp. capri PG3 was obtained in 1961 from D.G. ff. Edward (Wellcome Research Laboratories) and has subsequently been checked for identity and purity by R.H. Leach (Central Public Health Laboratories). The cells were grown and harvested and the tRNA isolated and fractionated to give tRNA_f^{Met} (previously referred to as tRNA_{fII}^{Met}) at least 90% pure as previously described⁶, except that the final RPC-5 column fractionation was performed at 42° rather than at room temperature as this was found to give an increased resolution².

The methods used for the modified nucleoside composition analysis⁷ of the purified tRNA and for the complete T₁-RNase and pancreatic RNase digestion of the tRNA, followed by the labeling and sequence analysis of the resulting oligonucleotides have been described before⁷⁻⁹.

The 5'-end group labeling of the intact tRNA_f^{Met} with [³²P] was as described before⁶. Use of this 5'-[³²P] labeled tRNA for 5'-terminal sequence analysis by partial digestion with

nuclease P_1 , followed by analysis of the partial digestion products by two-dimensional homochromatography, was as described elsewhere^{7,10}. Partial digestion of the 5'-[³²P] labeled tRNA with T_1 -, T_2 -, U_2 - and B. cereus RNase followed by gel electrophoretic analysis of the partial digestion products is described in an accompanying paper¹¹.

Partial pancreatic RNase digestion of *M. capri* tRNA_f^{Met} and sequence analysis of the digestion products (based on unpublished work of deWeerd and RajBhandary). *M. capri* tRNA_f^{Met} (0.25 A₂₆₀ unit) was incubated in 5 μ l Tris, pH 8.0, buffer (50 mM) at 4° with 0.01 μ g pancreatic RNase for 1 h. The digestion was terminated by the addition of phenol saturated with Tris (pH 8.0) buffer (10 μ l), water (15 μ l) was added, and the reaction mixture was extracted four times with phenol. The combined phenol layers were back-extracted with water (20 μ l), and the combined aqueous layers were extracted with diethylether in the usual way¹⁰. The oligonucleotides present in the partial digest were precipitated by the addition of ethanol (2 vol), treated with bacterial alkaline phosphatase and then labeled at their 5'-ends with [³²P] using polynucleotide kinase and γ -[³²P]-ATP⁶. The 5'-[³²P] oligonucleotides were then separated by two-dimensional polyacrylamide gel electrophoresis, located by autoradiography, and those to be used for sequence analysis were recovered by electrophoretic elution and subjected to partial digestion with nuclease P_1 ¹⁰.

RESULTS

1. Modified Nucleotide Composition of *M. capri* tRNA_f^{Met}

The tRNA was digested with T_2 -RNase, and the nucleoside 3'-phosphates thus produced were converted to [³²P]Np by incubation with polynucleotide kinase in the presence of γ -[³²P]-ATP⁷. Analysis of the products by two-dimensional thin-layer chromatography showed that besides the [³²P]Np spots corresponding to the four major nucleotides, the only other spots detected were those corresponding to pDp and p Ψ p. Although previous studies based on ultraviolet absorption of this tRNA at 338 nm showed that this tRNA contained s⁴U, we were unable to identify the spot corresponding to ps⁴Up. This is most

likely due to the fact that s^4U is quite labile and is easily converted to U; furthermore, s^4Up is phosphorylated by polynucleotide kinase only with difficulty.

2. Sequence Analysis of Oligonucleotides Present in Complete T_1 -RNase and Pancreatic RNase Digests.

The tRNA was digested with T_1 -RNase or pancreatic RNase and the oligonucleotides produced were labeled at their 5'-ends with [^{32}P], as described before^{8,12}. The 5'-[^{32}P] oligonucleotides were separated by two-dimensional electrophoresis^{7,13} (Fig. 1A and B) and their sequences were established by partial digestion with snake venom phosphodiesterase followed by analysis of partial digestion products by electrophoresis on DEAE-cellulose paper and/or two-dimension homochromatography^{6,7,9,14}. Table I lists the sequence and molar yields of these oligonucleotides.

3. Derivation of Longer Oligonucleotide Sequences and the Total Sequence of M. capri tRNA^{Met}_F.

a. Partial digestion of 5'-[^{32}P] tRNA with nuclease P_1 -- 5'-[^{32}P] tRNA was partially digested with nuclease P_1 and the products were subjected to two-dimensional homochromatography (electrophoresis in first dimension, homochromatography in the second dimension)^{7,10}. Homochromatography at pH 4.6 resulted in all oligonucleotides longer than 8 (i.e., those containing nucleotides beyond the s^4U_8 residue) migrating as two components, probably due to partial conversion of s^4U to U and the subsequent separation in the second dimension of oligonucleotides containing s^4U from the corresponding ones containing U (data not shown; see Fig. 4 below for a similar phenomenon). However, if the homochromatography solvent was adjusted to pH 7, only one spot was observed for each oligonucleotide of a given length. The sequence which could be deduced from the 5'-[^{32}P] end using this method was [^{32}P]C-G-C-G-G-G-G- s^4U -A-G-A-G-C-A-G-U (fragment a of Fig. 5).

b. Analysis of oligonucleotides present in partial pancreatic RNase digests -- Figure 2 shows the results of an experiment in which oligonucleotides present in partial pancreatic RNase digests were labeled with [^{32}P] at their 5'-end and then

Figure 1. Autoradiograms of fingerprints of T_1 -RNase (A) and pancreatic RNase (B) digests of *M. capri* tRNA^{Met}. Circled B, xylene cyanole blue tracking dye.

TABLE I. Sequence and molar yield of oligonucleotides present in fingerprints of RNase digests of *M. capri* tRNA^{Met}_f.

T ₁ RNase digest			Pancreatic RNase digest		
Spot Number	Sequence	Molar Ratio*	Spot Number	Sequence	Molar Ratio*
1	CG	1.91	1	pCp	-
2	CCG+CAACCA	1.98	2	AAC	1.8
3	CAG	1.97	3	GC	4.2
4	AG	3.12	4	AGC	1.02
5	CCCCCG	0.83	5	AU	0.7
6	CUCG	0.97	6	AGAGC	1.09
7	DAG ⁺	0.27 [‡]	7	AGU	0.92
8	UAG	0.40	8	ATP	-
9	pGp	-	9	GGGC	0.76
10	CUCAUAACCCG	0.19 ^Δ	10	AGGU+GAGU	2.04
11	UCCUG	1.09	11	GGU*	0.81
12	UΨCG ⁺	0.96	12	GGAGGC	0.81
13	UD(U)G ⁺	0.77	13	GGGGU	0.52
14	UCCUGp [○]	-	14	pUp	-

⁺See Figure 1.

*Based on Cerenkov counting of fingerprint spots.

[‡]Low molar ratio is due to the presence of modified nucleotides at the 5'-ends of these fragments that are often phosphorylated at a slower rate by polynucleotide kinase.

^ΔYield of this fragment is always low, we have no obvious explanation for this observation.

[○]Several minor slow-moving oligonucleotides in the second dimension have been shown to be 2',3'-cyclic phosphates of identified oligonucleotides.

⁺Modified nucleotides present at the 5'-end of these oligonucleotides are identified during routine 5'-end group analysis. The identification of Ψ in the UΨCG sequence has been described before⁶. Identification of UD(U)G in spot t13 is based on the fact that this oligonucleotide, although a trinucleotide, yields four spots upon DEAE-cellulose paper electrophoresis of partial snake venom phosphodiesterase digests. These spots are: a mononucleotide, [³²P]U which comigrates with a non-radioactive marker of pU, two dinucleotides which migrate close to each other and with mobilities identical to those expected for UD, and UU, and the original trinucleotide UD(U)G.

separated by two-dimensional polyacrylamide gel electrophoresis. Radioactive spots 1, 2, 3, 4 and 5 were eluted and their 5'-terminal sequences were derived by partial digestion with nuclease P₁. Figure 3 shows a typical two-dimensional homochromatographic pattern of such a digest on spot 3 of Figure 2, and the sequences thus derived for the oligonucleotides are

listed below:

<u>Spot Number</u>	<u>Sequence</u>
1	GCCCCGCAAC -----
2	AACCCGGAGGC -----
3	AGCUCGCCGGGC -----
4	AGCUCGCCGGGCU ----
5	CGCAGGU ^Ψ CGAGU ----

c. Analysis of partial T_1 -, T_2 -, U_2 - and *B. cereus* RNase digests of 5'-[32 P] tRNA by polyacrylamide gel electrophoresis -- Further evidence for the sequence of the tRNA beyond that obtained above was derived by partial digestion of 5'-[32 P] tRNA with enzymes of different base specificity followed by electrophoretic separation of the products in adjacent lanes of a polyacrylamide slab gel. This procedure, described recently by Donis-Keller *et al.*¹⁵ using yeast 5.8S RNA, has been modified and extended for application to 5'-[32 P] tRNA and mRNA^{2,11}. Figure 4 shows the results obtained when 5'-[32 P] tRNA was partially digested with T_1 -RNase (G specific), T_2 -RNase (non-specific), U_2 -RNase (A specific)

Figure 2. Autoradiographs of 5'-[32 P] labeled oligonucleotides present in a partial pancreatic RNase digests of *M. capri* tRNA^{Met}. The oligonucleotides were labeled at their 5'-ends with [32 P] and then separated by two-dimensional polyacrylamide gel electrophoresis¹¹. Circled B and XC indicate location of bromophenol blue and xylene cyanole tracking dye, respectively.

Figure 3. Autoradiogram of a partial nuclease P_1 digest on spot 3 of Figure 2. First dimension, electrophoresis on cellulose acetate at pH 3.5; second dimension, homochromatography on DEAE-cellulose thin layer plates. B circled is xylene cyanole blue tracking dye.

and the pyrimidine specific RNase (Y) from *B. cereus*¹¹. The sequence near the 5'-end of *M. capri* tRNA_f^{Met} which can be deduced from this experiment is AGAGYAGYGGYAGYY. As mentioned above in the analysis of partial nuclease P_1 digests by two-dimensional homochromatography, radioactive bands beyond Y* (s^4U) are split into two sets of bands (indicated by heavy bars), presumably due to the separation during electrophoresis of oligonucleotides of the same length and sequence which differ only in that one contains s^4U and the other U at a specific location. Based on the sequence of oligonucleotides in Table I and the 5'-terminal sequence derived using nuclease P_1 , most of the pyrimidine (Y) residues in the above sequence could be identified to yield the sequence A-G-A-G-C-A-G-U-D(U)-

Figure 4. Autoradiogram of partial enzymatic digests of 5'-[³²P] M. capri tRNA^{Met}. From left to right, -enzyme, incubated in the absence of enzyme; T₁, T₂, U₂ and Y, incubation of RNA at two different levels of these enzymes. Separation was by electrophoresis on a 20% polyacrylamide gel¹¹.

G-G-D-A-G-Y-Y. The results of Figure 4 also allow us to deduce the sequence A-G-G-Y-Y-G-Y-A-G-G-Y-Y-Y-G-A-G further on from the 5'-end (Fragment n of Fig. 5).

d. Total sequence of *M. capri* tRNA_f^{Met}. -- Figure 5 shows the linear sequence of *M. capri* tRNA_f^{Met} which could be derived using the above results, along with a list of oligonucleotides necessary for the establishment of the sequence. Figure 6 shows the sequence of *M. capri* tRNA_f^{Met} written in the clover-leaf form.

Figure 5. Sequence of *M. capri* tRNA^{Met} in linear form (top) and oligonucleotides necessary for establishment of the sequence. Dashed lines indicate that the fragment concerned contained additional nucleotides, the sequence of which was not completely deduced in the particular experiment described. U* is a mixture of D and U.

Figure 6. Sequence of M. capri tRNA^{Met} written in cloverleaf form. U* is a mixture of U and D.

DISCUSSION

The nucleotide sequence of initiator tRNAs has been conserved markedly during evolution^{9,16}. The sequences of all vertebrate initiator tRNAs are identical⁹, and even between prokaryotic and eukaryotic initiator tRNAs some 65-70% of the sequence is conserved². However, there is considerably more sequence homology between prokaryotic initiators as a group (86-97%) and eukaryotic initiators as another group (73-87%) than between the two groups. The extent of homology of the M. capri tRNA_f^{Met} to prokaryotic initiator tRNAs is 84-91%, whereas that to eukaryotic initiator tRNAs is 66-68%. Thus, both in the structural features which distinguish prokaryotic from eukaryotic initiator tRNAs⁶ and in the overall nucleotide sequence the M. capri tRNA_f^{Met} is distinctly prokaryotic. It is probably worth noting that this is the first unequivocal demonstration of the prokaryotic origin of a member of the Mycoplasma species. Although it has long been assumed that they are prokaryotes, due to their normal close association with and dependence upon eukaryotes, it might have been possible that

the Mycoplasma contained genetic material originating from such organisms. This is apparently not the case.

Table II compares the extent of homology of the *M. capri* tRNA^{Met}_f to the various other prokaryotic initiator tRNAs. The *M. capri* tRNA has the highest extent of homology to the corresponding tRNA from *B. subtilis* than from *E. coli* or *A. nidulans*. The results are generally in agreement with the proposal of Woese and his colleagues from studies on conserved sequences in 16S ribosomal RNA¹⁷, that the blue-green algae (Cyanophyta) diverged first in evolution, to be followed by the divergence of the *Bacillaceae* from the *Enterobacteriaceae*. If this is correct, *M. capri* appears to have diverged from the *Bacillaceae* and, thus, could not be regarded as an example of an evolutionarily primitive organism. Obviously, more data from other Mycoplasma species and from other prokaryotic organisms is required before any more definite conclusions can be reached, although comparison of *M. capricolum* tRNA^{Phe}¹⁸ with other prokaryotic tRNA^{Phe} sequences¹⁹, also suggests the same relationship.

The only nucleoside to be modified in the initiator tRNA of *M. capri* is uridine. This has been modified to 4-thiouridine, pseudouridine and dihydrouridine, but the other common uridine modification (to ribothymidine) has not occurred nor are there any other alkylated or hypermodified nucleosides present. Presumably, these latter modifications are thus not essential for the correct functioning of this tRNA in initiation of protein synthesis. Where they occur in tRNAs from other sources, they may only be present because of the use of tRNAs in other, more elaborate control functions. The func-

TABLE II. Percent Nucleotide Sequence Homology Between *M. capri* and Various Prokaryotic Initiator tRNAs.

	<i>E. coli</i>	<i>T. thermophilus</i>	<i>B. subtilis</i>	<i>A. nidulans</i>	<i>M. capri</i>
<i>E. coli</i>		97	92	90	84
<i>T. thermophilus</i>			94	90	86
<i>B. subtilis</i>				88	91
<i>A. nidulans</i>					84

tion of the modified nucleosides which do occur in M. capri tRNA is, however, not obvious, but due to the limited amount of genetic material in these organisms²⁰, it is tempting to assume that some overwhelming benefit is gained by the organisms from the presence of the uridine modification in its tRNAs.

Despite the high A-T content of the DNA from M. capri (75%)²¹, the initiator tRNA contains only 32% A+U residues and has among the highest G+C content of any tRNA yet sequenced, and is in direct contrast with the composition of the N. crassa mitochondrial initiator tRNA², which has an A+U content of 60%. And yet, both these tRNAs can be aminoacylated by the E. coli methionyl-tRNA synthetase and subsequently formylated by the E. coli met-tRNA formylase.

Finally, the tRNA_f^{Met} from M. capri whose sequence has been determined here is the same tRNA which was previously referred to as tRNA_{fII}^{Met}⁶. We have now also purified the species previously designated as tRNA_{fI}^{Met}, and have shown in further studies (unpublished results) that the purified tRNA is not formylated by E. coli met-tRNA formylase; moreover, this tRNA has a 5'-terminal pG and is, therefore, most likely to be the tRNA_m^{Met} species.

ACKNOWLEDGEMENTS

We are grateful to Mr. E.T.J. Chelton for growing the organisms and to Miss C. Aspey for technical assistance. We also thank M. Silberklang and J. Heckman for providing us with enzymes and various chromatography materials. This work was supported by grants from NATO, Grant No. 1043, and the Medical Research Council, and by the National Institute of General Medical Sciences, Grant No. GM 17151, of N.I.H. and Grant No. NP-114 from the American Cancer Society.

REFERENCES

1. Rich, A. and RajBhandary, U.L. (1976) Ann. Rev. Biochem. **45**, 805-860.
2. Heckman, J.E., Hecker, L., Schwartzbach, S., Barnett, W.E., Baumstark, B. and RajBhandary, U.L. (1977) Cell, in press.
3. Dube, S.K., Marker, K.A., Clark, B.F.C. and Cory, S. (1968) Nature **218**, 232-234.

4. Simsek, M. and RajBhandary, U.L. (1972) Biochem. Biophys. Res. Commun. **49**, 508-517.
5. Sprague, K.U., Hagenbüchle, O. and Zuniga, M.C. (1977) Cell **11**, 561-570.
6. Walker, R.T. and RajBhandary, U.L. (1975) Nucleic Acids Res. **2**, 61-78
7. Silberklang, M., Prochiantz, A., Haenni, A.-L. and RajBhandary, U.L. (1977) Eur. J. Biochem. **72**, 465-478.
8. Simsek, M., Ziegenmeyer, J., Heckman, J. and RajBhandary, U.L. (1973) Proc. Nat. Acad. Sci. USA **70**, 1041-1045.
9. Gillum, A.M., Urquhart, N., Smith, M. and RajBhandary, U. L. (1975) Cell **6**, 395-405.
10. Silberklang, M., Gillum, A.M. and RajBhandary, U.L. (1977) Nucleic Acids Res., in press.
11. Lockard, R.E., Alzher-DeWeerd, B., Heckman, J.E., Tabor, M.W., MacGee, J. and RajBhandary, U.L. (1977) Nucleic Acids Res., submitted.
12. Simsek, M., Petrissant, G., and RajBhandary, U.L. (1973) Proc. Nat. Acad. Sci. USA **70**, 2600-2604.
13. Sanger, F., Brownlee, G.G. and Barrell, B.G. (1965) J. Mol. Biol. **13**, 373-398.
14. Sanger, F., Donelson, J.E., Carlson, A.R., Kössel, H. and Fischer, D. (1973) Proc. Nat. Acad. Sci. USA **70**, 1209-1213.
15. Donis-Keller, H., Maxam, A. and Gilbert, W. (1977) Nucleic Acids Res. **4**, 2527-2538.
16. Ecarot-Charrier, B. and Cedergren R. (1976) FEBS Letters **63**, 287-290.
17. Balch, W.E., Magrum, L.J., Fox, G.E., Wolfe, R.S. and Woese, C.R. (1977) J. Mol. Evol., **9**, 305-311.
18. Kimball, M.E., Szeto, K.S. and Söll, D. (1974) Nucleic Acids Res. **1**, 1721-1732.
19. Chang, S.H., Brum, C.K., Silberklang, M. and RajBhandary, U.L., Hecker, L.I. and Barnett, W.E. (1976) Cell **9**, 717-723.
20. Morowitz, H.J. and Wallace, D.C. (1973) Ann. N.Y. Acad. Sci. **225**, 62-73.
21. Jones, A.S. and Walker, R.T. (1963) Nature **198**, 588-589.