Sequence analysis of a cluster of twenty-one tRNA genes in Bacillus subtilis

Christopher J.Green and Barbara S.Vold

Biomedical Research Laboratory, SRI International, 333 Ravenswood Avenue, Menlo Park, CA 94025, USA

Received 25 April 1983; Revised and Accepted 25 July 1983

ABSTRACT

The DNA sequence of a cluster of twenty-one tRNA genes distal to a rRNA gene set in <u>B. subtilis</u> was determined. None of the tRNA genes are repeated in the sequence. The only classes of tRNAs that are not represented are those for cysteine, glutamine, tryptophan, and tyrosine. Three of the tRNA genes in this cluster do not have the 3'-CCA sequence encoded in the gene. There is no RNA polymerase terminator sequence in the region between the 5S gene and the first tRNA gene or within the tRNA gene. This rRNA and tRNA gene cluster probably represents one transcriptional unit. However, there may be an RNA polymerase promoter site within this sequence, which raises some interesting questions concerning the regulation of transcription for these tRNA genes.

INTRODUCTION

The transcriptional and translational apparatus of Bacillus subtilis differs from that of Escherichia coli in several respects, and the studies presented here along with those of other investigators indicate that these two organisms also differ in the organization of tRNA genes. Hybridization studies performed using density-transfer experiments for the localization of DNA regions first suggested that B. subtilis might have a type of tRNA gene organization different from that of \underline{E} . coli in that the majority of tRNA genes of B. subtilis tended to be clustered in two regions of the genome, one near the origin of replication and one further away (1,2). Both regions also hybridized with the rRNAs for 5, 16, and 23S species, however, the majority of rRNA genes were clustered near the origin of replication at a position of about 3 to 12 (on a genetic map from 0-100). Early hybridization studies estimated the numbers of tRNA genes to be between 35 and 40 in B. subtilis (1,2). RNA-DNA hybridization studies by Jeng and Doi (3) suggest that there may be more tRNA genes than this. The maximum estimated number of tRNA genes using tRNA from vegetative cells was 44 and using tRNA from cells at Stage III of sporulation was 69. Jeng and Doi suggest that

new tRNA species exist in sporulating cells (3). However, Henner and Steinberg (4) analyzed tRNAs from cells at various growth stages using twodimensional gel electrophoresis and concluded that there were minimal changes in the species of tRNAs. Also RPC-5 column chromatography comparisons of tRNAs from vegetative cells and spores indicated a few possibilities for the existence of unique tRNAs at different growth stages (5,6). However, those tRNA species have not yet been sequenced.

Hybridization of tRNAs to Eco Rl restriction fragments has been carried out by Wawrousek and Hansen (7). Although the map positions of these fragments are not known, six fragments containing tRNA genes associated with rRNA genes were observed. This supports earlier evidence (1,2) that tRNA genes are clustered in a few regions of the genome.

Up to now, DNA sequence analysis of regions containing tRNA genes in <u>B</u>. <u>subtilis</u> have revealed that tRNA genes can exist either as spacers between the 16 and 23S rRNAs (8,9), as a clustered group between two ribosomal RNA gene sets (7,10), or as a clustered group apparently not associated with a ribosomal RNA gene set (11).

MATERIALS AND METHODS

Strains

The Eco RI fragment described in this paper was obtained from the collection of Ken Bott. It was designated as rrnB by Bott and it has been mapped (G. Stewart, M. Lampe, and K. Bott, unpublished observations) to be in the area between argA (position 72.5 on a map from 0-100) and thr (position 81). This 4.5 kb insert of <u>B. subtilis</u> DNA was cloned into a pDH5109 plasmid and designated pBC204 by Ken Bott. tRNA hybridization studies performed in our laboratory indicated that this piece contained several tRNA genes. <u>E. coli</u> K12 JM103 and M13 mp7 phage were purchased from New England Biolabs.

Enzymes and Reagents

Enzymes and reagents were obtained from the following sources: restriction enzymes and polynucleotide kinase, Bethesda Research Labs; T4 DNA ligase and M13 pentadecamer primer DNA, New England Biolabs; the Klenow fragment of DNA polymerase, Boehringer Mannheim. $[\alpha^{-32}P]dCTP$ (specific activity 600-800 Ci/mmol), New England Nuclear; all deoxy- and dideoxy-nucleotide triphosphates, PL Biochemicals.



scale: = 100 bases

Figure 1. Physical map of the rRNA and tRNA gene cluster. The amino acid designations denote the presumed isoaccepting class of each tRNA gene. \underline{T} and \underline{H} show the Taq I and Hpa II restriction sites, respectively. The arrows indicate the origin and direction of each M13 clone sequence analysis.

DNA Preparation

The pBC204 plasmid was isolated by the procedure of Norgard (12). The 4.5 kb Eco Rl insert of <u>B. subtilis</u> DNA was cut out with Eco RI and purified by RPC-5 chromatography (13). This fragment was further reduced by cleavage with Sma I at a single site to yield a 1.5 kb piece that hybridized to 23S RNA but not tRNA and a 3.0 kb piece that hybridized strongly to tRNA by Southern analysis (14). The 3.0 kb DNA fragment was purified by RPC-5 and used in this sequence analysis.

DNA Sequencing

The M13 single-strand phage dideoxynucleotide chain termination technique was used to perform the sequence analysis (15,16). M13 mp7 doublestranded replicating factor was cut with Acc I to yield ends that could hybridize to either Hpa II or Taq I cut DNA fragments. The 3.0 kb Sma I-Eco RI <u>B. subtilis</u> DNA fragment from pBC204 was digested completely with Hpa II or partially with Taq I and ligated into the M13 cloning vector. Since most tRNA genes have a Taq I site (TCGA) in their GT $^{\rm WC}$ loop region, a partial digest with this enzyme resulted in a set of overlapping M13 clone inserts. The final sequence was constructed as shown in Figure 1. Approximately 200 bases at either end of the 3.0 kb fragment were not sequenced.

RESULTS AND DISCUSSION

The sequence analysis of this piece of <u>B. subtilis</u> DNA shows the presence of 21 tRNA genes immediately following an rRNA gene set (Figure 2).

TCGTAGGTTCCAGGTCCTACAGGTCCACCTATACGGAGGAATACCCCAAGTCTGGCTGAAGGGATCGGTCTTGAAAACCGACAGGGTGTCAAAGCCCGC 1500 	0
GGGGGTTCGAATCCTCTCCTCCGCCATACATATTCCTAATCATCGCGGGGTGGAGCAGTTCGGTAGCTCGGGGGCTCATAACCCGAAGGTCGCAGGT 1800	0
TCAAATCCTGCCCCGCAACCAAATTTTAAAATGGTCCGGTAGTTCAGTTGGTTAGAATGCCTGCC	0
CGGACCGCCATTTAAATACTTAGGCTCGGTAGCTCGGTTGGTAGAGCAACGGACTGAAAATCCGTGTGGGGGGGG	•
TATCAATATGCTTTGGCGGTTGTGGCGAAGTGGTTAACGCACCAGATTGTGGCTCTGGGCATTCGTGGGTTCGATTCCCATCAATCGCCCCCAAATAAAAAT 1900	•
TGCGGGTGTAGTTTAGTGGTAAAACCTCAGCCTTCCAAGCTGATGTCGTGGGTTCGATTCCCATCACCCGGCTCCATTTCTATATCGTCATGGGCCTGTAG 2000 	0
CTCAGCTGGTTAGAGCGCCGGCTGATAAGCGTGAGGTCGATGGTTCGAGTCCATTCAGGCCCACCATGACTTTTGTTCCACAGTAGCTCAGTGGTAGAG 210	8
CTATCGGCTGTTAACCGATCGGTCGGAGGTTCGAATCCTGCCTG	8
AGGTCGTGTAAGCGGCGGAGGGTTCAAATCCCTCCTCCCGCCATATGATTACAGATATCATAATTATCGGCCCGTTGGTCAAGCGGTTAAGACACCG 230	8
CCCTTTCACGGGCGGTAACACGGGTTCGAATCCCGTACGGGTCATCCCGGAAGGCTTGCCTGCAAGGTTTTTTGTTTTTGTTTTATAAATCATGTATATG 240 tRNa(Glu, UUC)	8
TCTTAGATTTTGTTCTTTATTTTAAAAAACAGACTACAAAAATCTCCATATATTTCGTTTTTCTTCAGAAAATGAAGTTAATTGTCTATAAGTATAAGCCG 250	8
TTTCAGGGAAAGGGGTTTTTTTTTTTTTTCTTCGA	
Figure 2. The nucleotide sequence of the cluster of 21 tRNA genes distal to an rRNA operon. The rRNA, tRNA, promoter and the terminator sequences are underlined.	

There are no repeated tRNA genes; when isoaccepting species are found they have different anticodons. The only classes of tRNAs that are not represented are those for cysteine, glutamine, tryptophan, and tyrosine. The cloverleaf representation of the tRNA genes is shown in Figure 3. Of the 21 tRNA genes in this cluster, 9 are known from the published tRNA sequence: tRNAAla, tRNAArg, tRNAGLY, tRNALYS, tRNAMEt, tRNAFMEt, tRNAPhe, tRNAThr, and tRNAVAC (17). All 9 of these known tRNAs are present in both vegetative cells and spores.

Several other clusters of <u>B.</u> <u>subtilis</u> tRNA genes have recently been sequenced: tRNA^{Ile}_{GAU} and tRNA^{Ala}_{UGC} (8); tRNA^{Asn}_{GUU}, tRNA^{Thr}_{GGU}, tRNA^{Arg}_{ACG}, tRNA^{Pro}_{UGG}, and tRNA^{Ala}_{UGC} (7); and tRNA^{Glu}_{UUC}, tRNA^{Lys}_{UUU}, tRNA^{Asp}_{GUC}, and tRNA^{Phe}_{GGA} (11). The sequences of the tRNA genes in these clusters match ours with the following exceptions: 1) a base pair in the stem of the GT^VC loop of tRNA^{Ile}_{GAU}; 2) two base pairs in tRNA^{Asn}_{GUU}, one in the stem of the GT^VC loop and one in the aminoacyl stem; and 3) the tRNA^{Thr}_{GGU} which has a significantly different structure from the tRNA^{Thr}_{GUU} found in our cluster.

Three of the tRNA genes found in this cluster appear to code for the The tRNA^{fMet} gene codes for an RNA identical to methionine anticodon CAU. that of the known tRNA^{fMet} (18). The tenth tRNA gene in the cluster differs from the reported sequence of tRNA^{Met} of B. subtilis (19) by an extra T in the dihydrouridine loop. This extra nucleotide increases the homology of this predicted B. subtilis tRNA^{Met} with the known E. coli tRNA^{Met} (20) and gives a definite band on our sequencing gels. The eleventh tRNA gene in this cluster also has a CAU anticodon but RPC-5 studies of B. subtilis tRNAs indicate the presence of only two species that can be aminoacylated with methionine (5). No other tRNAs for Met or fMet were observed by Yamada and Ishikura during the purification for sequence analysis of the methionine tRNAs from Bacillus subtilis cells harvested in late logarithmic stage (Yamada and Ishikura, personal communication). The presence of an A-U base pair in the third position of the aminoacyl stem of this predicted tRNA may rule out its being recognized by a methionyl-tRNA synthetase (21). Phage T4, E. coli, and spinach chloroplasts have an unknown modified nucleoside in the wobble position of the anticodon of a tRNANI that may have originated from a C in the original transcript and may allow it to recognize AUA codons (22-25). By analogy with these other examples and because of the homology with the $tRNA_{CAU}^{IIe}$ gene, we conclude that the eleventh tRNA gene in this cluster is actually a gene for a tRNAILe and not a methionine isoacceptor.

5768

All known tRNA^{His} sequences have an extra nucleotide at the 5' end of the molecule, usually a G (17). In <u>Drosophila melanogaster</u> and <u>Schizosaccharomyces pombe</u> this nucleotide is added to the tRNA^{His} posttranscriptionally and is not encoded in the gene (26). No other prokaryotic tRNA^{His} gene sequences have been published. The <u>B. subtilis</u> tRNA^{His} gene that we have sequenced contains an extra G at the 5' end. None of the other 20 tRNA genes in this cluster or the 14 <u>B. subtilis</u> tRNA genes in other clusters (7,8,9,11) have a G before the normal coding sequence. Therefore, we believe that this is evidence that <u>B. subtilis</u> tRNA^{His} has an additional G residue at the 5' end and that this residue is encoded in its gene and is not a posttranscriptional modification.

The tRNA^{Leu} gene reported here is unusual in that the sequence contains an A at the 5' end of the anticodon loop. This was confirmed by sequencing both strands of the DNA fragment containing this gene. All prokaryotic tRNAs and tRNA gene sequences so far reported contain only a pyrimidine at this site. A purine has been found in this position only in tRNAs of chloroplast and mitochondrial tRNAs (17).

Three out of the 21 tRNA genes in this cluster, $tRNA_{UGC}^{Ala}$, $tRNA_{UUC}^{Glu}$, and $tRNA_{UAA}^{Leu}$, did not have the 3' terminal CCA. This appears to be a relatively common feature in <u>B</u>. <u>subtilis</u> as the $tRNA_{GUU}^{Asn}$, $tRNA_{GUU}^{Thr}$ (7), $tRNA_{UUC}^{Glu}$, and $tRNA_{GAA}^{Phe}$ genes (11) in other clusters also are missing the CCA end. It is interesting to note that the same tRNA genes in different clusters are not always without their CCA end.

The last 233 bases of the 23S rRNA gene were sequenced. Only the last 25 bases of this gene have been previously reported (7). The 55 bases between the 23S and 5S gene differs at only one position (position 253 in our sequence) from that reported by Wawrousek and Hansen (7) for a different rRNA gene set. The spacer region between the 5S and tRNA genes has no homology with the previously reported rRNA-tRNA gene set (7). Beginning at position 406 of the sequence, one base down from the 5S gene, there begins a seven-base paired hairpin loop (Figure 2). This loop does not have the stretch of T residues that are characteristic of terminator hairpins found in <u>E. coli</u> (27). Hairpin loops have been shown to occur the spacer regions in <u>E. coli</u> (28) and <u>B. subtilis</u> tRNA gene clusters (11) and may play a role in the processing of this primary transcript (29).

The first tRNA gene, tRNA^{Val}, begins 21 bases downstream from the end of the 5S rRNA gene. The spacing between the 21 tRNA genes varies from 2 to





Figure 3. The secondary structures of the tRNA genes in this gene cluster. The genes for tRNAUGC, tRNAUUC, and tRNAUAA do not have CCA ends. The gene for tRNAMI has a methionine anticodon (CAU) but it has a 78% homology with the gene for tRNAGAU and the C in the wobble position of its anticodon may be posttranscriptionally modified to recognize AUA codons.

37 bases. The spacer regions on either side of the structural genes for both tRNA^{Thr} and tRNA^{Lys} have six-base sequences that could pair causing the CCA ends of these tRNAs to loop out. This type of tRNA precursor structure has been found in <u>E. coli</u> and may create recognition sites for processing

enzymes (28,30). One common characteristic that the spacer regions share is that they are AT rich. Inside the largest spacer region, between the second (tRNA^{Thr}) and third (tRNA^{Lys}) genes, a perfect consensus sequence (TATAAT) for a Pribnow box (-10 region) for an RNA polymerase promoter was found (Figure 2). Upstream of this site, separated by 17 bases, is a corresponding -35 region (TTGCCG) that conserves the most important part of the -35 consensus sequence (TTGNNN) (27). It is interesting to note that this possible -35 region is inside the $tRNA_{IIGII}^{Thr}$ gene and that the first two T residues in the -35 sequence results in the tRNA having two suboptimal U-G pairs at this junction of the aminoacyl and $GT\Psi C$ stems (Figure 3). The transcription start site for this predicted promoter is just upstream of the tRNALUS gene. This promoter would be recognized by the major, σ^{55} containing, RNA polymerase found in both vegetative and sporulating B. subtilis cells (31,32). B. subtilis contains at least four different holoenzymes for RNA polymerase activity during vegetative growth and at least one new holoenzyme during sporulation (33). The multiple forms are characterized by different sigma factors, and recognize different base sequence in the -35and -10 regions of the promoter. However, relatively few promoters for the sporulating and minor RNA polymerases have been analyzed so that sequence information on their consensus recognition sites is incomplete. This tRNA gene cluster does not contain any other presently known -10 and -35 sequence combinations for these other RNA polymerases (33). However, there may be some sites that approximate these known minor promoter sequences. As more polymerase recognition sequences are found, other promoter sites may be discovered on this sequence.

Immediately following the last tRNA gene, $tRNA_{UUC}^{elu}$, there is a 12 basepair hairpin loop with a long stretch of Ts at the 3' end, characteristic of an RNA polymerase terminator site (27) (Figure 2). Such terminator structures have also been found in the other <u>B. subtilis</u> gene clusters (7,11). The sequencing was continued for approximately 150 bases beyond this termination site without encountering any further tRNA genes or obvious RNA polymerase recognition sites.

There are no transcription termination sites between the end of the 23S rRNA gene and the last tRNA gene in this sequence. Thus it appears that this rRNA-tRNA gene cluster can be transcribed as a single unit. The putative RNA polymerase promoter, between the second and third tRNA genes, for the major vegetative RNA polymerase containing σ^{55} , may serve to increase the relative rates of transcription of the 19 downstream tRNA genes compared

to the tRNA $_{\rm UAC}^{\rm Val}$, tRNA $_{\rm UGU}^{\rm Thr}$, and rRNA genes lying upstream from this promoter. In addition, it may serve to relieve the last 19 tRNA genes from the same type of regulation controlling the rRNA promoter.

The extensive clustering of tRNA genes in B. subtilis demonstrated here is on a significantly greater scale than that found in E. coli. Approximately 50 tRNA genes representing 35 individual species of tRNAs have been mapped on the E. coli chromosome (34). In E. coli, tRNA genes occur in all seven rRNA operons and also occur as separate transcriptional units. The largest tRNA cluster known in E. coli contains seven tRNA genes (28). This E. coli tRNA gene cluster exists as a separate operon with its own promoter and terminator and is apparently not linked to rRNA operons. B. subtilis also contains at least one cluster of four tRNA genes that appears to be a separate transcriptional unit (11). However, the bulk of the tRNA genes may be associated with rRNA gene sets in B. subtilis (7). Hybridization studies have shown that the great majority of tRNA genes are on only six Eco RI fragments of the B. subtilis genome and that these fragments also hybridize to rRNA. Three of these hybridize to both 16S and 23S rRNA and probably represent tRNA clusters located between tandemly arranged rRNA gene sets (7). Another Eco RI fragment is derived from the spacer region between 16S and 23S rRNA genes that contain tRNA^{Ile} and tRNA^{Ala} genes (8). The remaining two Eco RI fragments strongly hybridize to tRNA and 23S rRNA but not 16S rRNA. We believe that the larger Eco RI fragment, designated trrnE by Wawrousek and Hansen (7) is the 4.5 kb piece cloned by Bott (35) and designated rrnB that contains the sequence presented in this paper. The presence of a 4.2 kb Eco RI fragment that also hybridizes to 23S rRNA and tRNA may indicate the existence of another tRNA gene cluster similar to this one.

The clustering of tRNA genes on the <u>B</u>. <u>subtilis</u> genome should allow the complete characterization of the organism's tRNA gene organization in a relatively short period of time. The possible discovery of other RNA polymerase recognition sites controlling the expression of these genes may answer the questions of the importance of any specific tRNA species to the process of sporulation.

ACKNOWLEDGEMENTS

We are indebted to Ken Bott for providing us with the pBC204 plasmid. We would also like to thank Y. Yamada, H. Ishikura, H. Yoshikawa, L. Schulman, and B. Dudock for helpful discussions.

This work was supported by Public Health Services Research Grant GM 29231 from the National Institutes of Health.

REFERENCES

- Oishi, M., Oishi, A., and Sueoka, N. (1966) Proc. Natl. Acad. Sci. U.S.A. 1. 55, 1095-1103.
- Smith, I., Dubnau, D., Morell, P., and Marmur, J. (1968) J. Mol. Biol. 33, 2. 123-140.
- 3.
- 4.
- Jeng,Y.-H. and Doi,R.H. (1975) J. Bacteriol. <u>121</u>, 950-958. Henner,D.J. and Steinberg,W. (1979) J. Bacteriol. <u>140</u>, 555-566. Vold,B.S. and Minatogawa,S. (1972) in Spores V. Halvorson,H.O., Hanson,R., and Campbell,L.L. Eds., American Society for Microbiology, 5. Washington, D.C. pp. 254-263.
- Vold, B.S. (1973) J. Bacteriol. 113, 825-833. 6.
- Wawrousek.E.F. and Hansen, J.N. (1983) J. Biol. Chem. 258, 291-298. 7.
- 8. Loughney, K., Lund, E., and Dahlberg, J.E. (1982) Nucleic Acids Res. 10. 1607-1624.
- Ogasawara, N., Seiki, M., and Yoshikawa, H. (1983) J. Bacteriol., in 9. press.
- Zuber, P. (1982) Ph.D. Thesis, Microbiol. Dept. Univ. Virginia. 10.
- Yamada, Y., Ohki, M., and Ishikura, H. (1983) Nucleic Acids Res., in 11. press.
- Norgard, M.V. (1981) Anal. Biochem. 113, 34-42. 12.
- Hardies, S., Horn, G., Klein, B., Larson, J., Neuendorf, S., 13. Wells,R. Panayotatos, N., Patient, R., and Selsing, E. (1980) in Methods in Enzymol, Nucleic Acids, Pt. I, Grossman, L. and Moldave, K. Eds., Vol. 65, pp. 327-347, Academic Press, New York.
- Southern, E.M. (1975) J. Mol. Biol. 98, 503-517. 14.
- Sanger, F., Nicklen, S., and Coulsen, A. (1977) Proc. Natl. Acad. Sci. USA 15. 74, 5463-5467.
- 16. Messing, J., Crea, R., and Seeburg, P.H. (1980) Nucleic Acids Res. 9, 309-321.
- Gauss, D. and Sprinzl, M. (1983) Nucleic Acids Res. 11, r1-r105. 17.
- Yamada,Y., Kuchino,Y., and Ishikura,H. (1980) J. Biochem. 87, 18. 1261-1269.
- 19. Yamada, Y. and Ishikura, H. (1980) Nucleic Acids Res. 8, 4517-4520.
- 20. Cory, S. and Marker, K. (1970) Eur. J. Biochem. 12, 177-194.
- 21. Schulman,L. (1979) in Transfer RNA: Structure, Properties and Recognition, Schimmel, P.R., Söll, D., and Abelson,J.N. Eds., pp. 311-324, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Fukada, K. and Abelson, J. (1980) J. Mol. Biol. 139, 377-391. 22.
- 23. Kuchino, Y., Watanabe, S., Harada,F., and Nishimura,S. (1980) Biochemistry 19, 2085-2089.
- Kashdan, M. and Dudock, B. (1982) J. Biol. Chem. 257, 11191-11194. 24.
- 25. Francis, M.A. and Dudock, B. (1982) J. Biol. Chem. 257, 11195-11198.
- Cooley, L., Appel, B., and Söll, D. (1982) Proc. Natl. Acad. Sci. USA 79, 26. 6475-6479.
- Rosenberg, M. and Court, D. (1979) Ann. Rev. Genet. 13, 319-353. 27.
- 28.
- Nakajima, N., Ozeki, H., and Shimura, Y. (1981) Cell 23, 239-249. Young, R. and Steitz, J. (1978) Proc. Natl. Acad. Sci. USA 75, 3593-3597. 29.
- 30. Rossi, J.J. and Landy, A. (1979) Cell 16, 523-534.
- Losick, R. and Pero, J. (1981) Cell 25, 582-584. 31.
- 32. Doi, R. (1982) Arch. Biochem. Biophys. 214, 772-781.
- 33. Johnson, W. C., Moran, C. P., and Losick, R. (1983) Nature 302, 800-804.
- Ozeki,H. (1980) in Genetics and Evolution of RNA Polymerase, tRNA 34. and Ribosomes, Osawa, S., Ozeki, H., Uchida, H., and Yura, T. Eds., pp. 173-183, Elsevier/North-Holland Biomedical Press, New York.
- Stewart, G., Wilson, F., and Bott, K. (1982) Gene 19, 153-162. 35.
- 36. Hasegawa, T. and Ishikura, H. (1978) Nucleic Acids Res. 5, 537-548.