Identification and location of nine T5 bacteriophage tRNA genes by DNA sequence analysis

Suresh M. Desai, Judith Vaughan and Samuel B. Weiss

Franklin McLean Memorial Research Institute, Ben May Laboratory for Cancer Research, and Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, IL 60637, USA

Received 31 March 1986; Accepted 25 April 1986

ABSTRACT

Sequence analysis of two DNA fragments generated from bacteriophage T5 DNA by restriction with Hpa I and Hae III has resulted in the detection and localization of nine tRNA genes (His, two Ser genes, Leu, Val, Lys, fMet, Pro, and Ile). The genes which code for tRNAs His and Leu are partials, whereas the remaining genes are complete. A majority of the tRNA genes are located in close proximity to one another. A unique feature of the Pro and Ile genes is that their DNA sequence overlap.

INTRODUCTION

Earlier reports from this and other laboratories (1-8) have shown that the genome of bacteriophage T5 contains tRNA genes that are expressed in T5 phage infected Escherichia coli. Our studies indicate that T5 DNA codes for tRNAs that are recognized by the 20 different amino acids including several tRNA isoacceptors (2-5). Figure 1A depicts the entire T5 genome (approximately 80 x 10^6 daltons), with four single strand interruptions that divide the light strand into five unequal segments designated A, B, C, D, and E (9,10). The T5 tRNAs are coded by the heavy continuous strand (4), and all of the tRNA genes map in the C segment (4,5) except for Arg, which is located at the left end of the D segment (11). These genes span a DNA length of about 13.8 kb that represents 11.2% of the T5 chromosome.

The mapping of tRNA genes were located in six cluster groups (5). Several of these cluster groups contain 5 to 8 tRNA genes, but their arrangement within each cluster has not been determined. The structure for a number of T5 tRNA species has been elucidated by Kryukov et al. (12). This same group has reported the T5 gene order for tRNA^Gln^,His^,Asp^ and several other stable non-tRNA species (13); they have also defined the location and structure of two tRNA^Ser^ genes by using DNA sequence analysis (14).

In this report, we describe the results of DNA sequence analysis for two T5 DNA fragments; this analysis has led to the identification and loca-
ization of nine tRNA genes (His, two Ser genes, Leu, Val, Lys, fMet, Pro, and Ile). The DNA sequence and the location of the two Ser genes are in general agreement with the findings by Kryukov et al. (14). Two of the tRNA genes identified here, Pro and Ile, show overlapping sequences.

MATERIALS AND METHODS

Materials

DNA restriction endonucleases, T4 polynucleotide kinase, T4 DNA ligase, E. coli polymerase I, and Bam HI linkers were obtained from P-L Biochemicals. Calf intestine alkaline phosphatase was purchased from Boehringer Mannheim, and (γ-32P)ATP from ICN. Hind III-restricted 32P-labeled φX174 DNA (RF) was used as a marker for sizing of DNA fragments separated by electrophoresis on acrylamide gels.

Preparation of T5 DNA and T5 32P-tRNAs

T5 phage DNA was isolated from T5+ phage which was grown and purified as described earlier (4). T5 32P-tRNAs were prepared from E. coli F cells infected with T5+ phage, or with one of the T5 deletion mutants st(14) and b3, and pulsed with carrier-free 32Pi under conditions reported previously (5). Cells were collected, tRNAs isolated, and individual tRNA species purified by two-dimensional gel electrophoresis (5).

Hybridization analysis of restricted DNA fragments

T5 DNA fragments, generated by reaction with restriction endonucleases under conditions described by the supplier, were separated by electrophoresis on 1% agarose and transferred to nitrocellulose as reported by Southern (15). Annealing of individual T5 32P-tRNA species with the DNA-impregnated nitrocellulose strips was conducted under conditions reported elsewhere (11). Autoradiography was used for detection of tRNA-DNA hybrids following extensive washing and drying of the nitrocellulose strips.

DNA sequence analysis

The Maxam-Gilbert technique was employed for DNA sequence analysis (16). DNA fragments were end-labeled with (γ-32P)ATP by the action of T4 polynucleotide kinase under conditions described by the supplier. The labeled DNA was restricted with an appropriate endonuclease that made a single cleavage, and the two subfragments generated were separated by acrylamide gel electrophoresis. Alternatively, the end-labeled DNA was heat denatured in 90% formamide, and the two single strands were separated by electrophoresis on acrylamide gels containing 8 M urea. The radioactive DNAs were located by autoradiography, electroeluted, purified on DEAE-cellulose, and used directly.
for sequence analysis after precipitation with ethanol in the presence of carrier RNA.

Cloning of T5 DNA fragments

The general procedure for the cloning of restriction fragments was essentially that described by Maniatis et al. (17). Blunt-ended T5 DNA fragments were ligated with Bam HI linkers and restricted with Bam HI endonuclease, and the purified reconstructed fragment was inserted into the unique Bam HI site of pBR322 plasmid DNA by reaction with T4 DNA ligase. E. coli was transformed as described by Cohen et al. (18). Transformants were selected from colonies grown on L plates containing ampicillin (50 μg/ml). Plasmid DNAs isolated from amplified colonies were screened for DNA inserts by their slower mobility on gel electrophoresis relative to that of nonmodified pBR322 DNA.

RESULTS

Maps of T5 DNA fragments generated by different restriction endonucleases have been described previously (19,20). Figure 1B shows the restriction map of T5 DNA when it is cleaved with Hpa I (20). Our hybridization analysis of the Hpa I fragments indicated that tRNA genes were located in fragments 5, 6, and 8, with fragment 6 containing the majority of these genes (Table 1). Restriction of Hpa I fragment 6 with Hae III generates 9 subfragments whose order we established as shown in Figure 1C. Hybridization

![Figure 1. The T5 genome: restriction with Hpa I and Hae III endonucleases.](image)

(A) The physical map of T5 DNA shows the heavy continuous and the light DNA strands designated H and L, respectively. The letters A, B, C, D, and E indicate the DNA segments produced by the major nicks in the L strand. The numbers below each DNA segment are the approximate molecular weights (x 10^-6) of the segments. The region of T5 DNA where tRNA genes are located is indicated above the H strand. (B) Hpa I restriction map of T5 DNA (20). Only those DNA segments containing tRNA genes are numbered. (C) Hae III restriction map of Hpa I fragment 6. The numbers above each fragment represent their approximate base pair content.
Table 1. Hybridization of T5 tRNA Species with Hpa I Restricted T5 DNA Fragments

<table>
<thead>
<tr>
<th>DNA Fragment Number</th>
<th>Hybridization Analysis&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Gly, Met</td>
</tr>
<tr>
<td>6</td>
<td>His&lt;sub&gt;1&lt;/sub&gt;, His&lt;sub&gt;2&lt;/sub&gt;, Ser&lt;sub&gt;2&lt;/sub&gt;, Ile&lt;sub&gt;1&lt;/sub&gt;, Leu, fMet, Pro, Gln, Thr, Asn, Phe</td>
</tr>
<tr>
<td>8</td>
<td>Ile&lt;sub&gt;2&lt;/sub&gt;, Ser&lt;sub&gt;1&lt;/sub&gt;, Arg</td>
</tr>
</tbody>
</table>

<sup>a</sup>Purified <sup>32</sup>P-tRNA species were used for hybridization analyses as described in Materials and Methods.

Analysis indicated the presence of multiple tRNA genes in Hae III fragments B and C; hence, we attempted to clone these two fragments. We were successful in cloning Hae III fragment B in pBR322 plasmid DNA, but we were unable to clone the complete Hae III fragment C using this same vector. The inability to clone certain T5 DNA fragments in various vectors has also been reported by others (13,21).

Subfragments from Hae III fragments B and C were generated for DNA sequence analysis by cleavage with additional restriction enzymes, as shown in Figure 2. The sequence of nucleotides derived by the method of Maxam and Gilbert (16) for Hae III fragment B is shown in Figure 3 for the noncoding (RNA-like) DNA light strand. In this DNA segment, we identified two separate

![Hae III Fragment B](image1)

![Hae III Fragment C](image2)

Figure 2. Restriction maps of T5 DNA Hae III fragments B and C. The strategy used to determine the DNA sequence is shown beneath each fragment. The dashed arrows indicate the origin and direction of the DNA regions sequenced.
Figure 3. Nucleotide sequence of Hae III fragment B (RNA-like strand). Sequence direction is 3' to 5', left to right. The sequences for the partial genes tRNA<sub>His</sub> and tRNA<sub>Leu</sub>, and the two tRNA<sub>Ser</sub> genes are indicated by arrows. The DNA sequence between the Hind III and Bgl II restriction sites were reported by Kryukov et al. (14). Differences in nucleotide sequence from those reported here are indicated by an asterisk (*) designating a nucleotide that is either different, missing, or extra, and by an inverted “V” over two bases which designates an inversion of sequence.

Figure 4. Nucleotide sequence of Hae III fragment C (RNA-like strand). Sequence direction is the same as for fragment B (Fig. 3). Arrows above sequences indicate the location of the tRNA genes for Val, Lys, fMet, Pro, and Ile<sub>3</sub>. 

4201
genes, both coding for tRNA$^{Ser}$, at residues 610-699 and 707-799. The genes for these two tRNAs were first located and sequenced by Kryukov and coworkers (14) in a Hind III/Bgl II T5 DNA fragment whose restriction sites are indicated in Figure 3. Hae III fragment B also contains the partial genes for tRNA$^{His}$ and tRNA$^{Leu}$ at residues 1-44 and 1388-1442, respectively. These partial genes occur at either end of fragment B and were identified from their reported RNA sequences (12).

In Hae III fragment C, the tRNA genes for Val, Lys, fMet, and Pro were identified at residues 402-475, 483-560, 749-828, and 835-911, respectively (Fig. 4). In addition to the above genes, the nucleotide residues from 902-977 can be folded into a tRNA structure with an anticodon (GAU) for the recognition of isoleucine. The 3' terminus of this tRNA$_{3}^{Ile}$ gene shows no CCA sequence and overlaps the 5' terminus of the tRNA$_{3}^{Pro}$ gene by 10 nucleotides. The folded T5 tRNA structures for Val, Lys, fMet, and Ile$_{3}$ are illustrated in Figure 5; the structures for the other tRNAs described here have been reported elsewhere (12,14).
DISCUSSION

In this study, sequence analysis of two T5 DNA fragments allowed us to detect nine tRNA genes. Based on previous studies, the location of several of these genes were somewhat unexpected. We described two tRNA$^{\text{Ser}}$ genes in an earlier report (5). One gene (Ser$_2$) was mapped in a tRNA cluster (Region I) located near the left end of T5 DNA segment C, and the other gene (Ser$_1$) was mapped in a cluster (Region IVa) approximately 7.5 kb downstream from Region I, at the right end of DNA segment C (Fig. 1A). In the present study, the tRNA$_{\text{Ser}}$ genes in Hae III fragment B were found to be relatively close to the tRNA$^{\text{His}}$ gene, which was previously mapped in the Region I cluster (5).

We assume, therefore, that neither of the two tRNA$_{\text{Ser}}$ genes in fragment B represents the Ser$_1$ gene of Region IVa, but that one gene is Ser$_2$ (Region I) and the other is a third Ser gene which we had not detected in our earlier studies. It is also possible that we had previously mismapped the Ser$_1$ tRNA gene.

We had previously identified the presence of two different tRNA$^{\text{Ile}}$ genes, Ile$_1$ and Ile$_2$, which were mapped in Region I and Region IIIb cluster groups, respectively (4,5). These two clusters are separated by about 5.7 kb. The Region II cluster, which lies between Regions I and IIIb, was reported to contain the tRNA genes for Leu, Ala, Pro, Val, Lys, and fMet, but not Ile (4,5). In the present study, the tRNA genes identified by DNA sequence analysis of Hae III fragment C (Val, Lys, fMet and Pro) are Region II genes. Therefore, the tRNA Ile gene in fragment C appears not to be either Ile$_1$ or Ile$_2$, but a different Ile gene (Ile$_3$) not previously identified. Positive hybridization of Hae III fragment C with $^{32P}$-tRNA$^{\text{Ile}}$, but not with $^{32P}$-tRNA$^{\text{Ile}}$, suggests that the Ile$_3$ gene is Ile$_1$-like and may be a partial or duplicate gene copy of Ile$_1$. However, as with the Ser$_1$ gene, we cannot exclude the possibility that the Ile$_1$ gene was previously mismapped.

A number of the tRNA genes in the Hae III B and C fragments are close to one another. This raises the strong possibility that gene expression may involve primary transcripts which are polycistronic, similar to T4 tRNA transcripts (22). The Ser-Ser and the Val-Lys genes are each separated by seven nucleotides and the fMet-Pro genes are separated by six nucleotides. The spacer nucleotides between these closely positioned genes are rich in T and A residues, as are the nucleotides proximal to the 3' and 5' ends of the gene pairs, e.g., 3'-ATAATTTT[fMET]TAACAA[Pro]TTTTAATA-5'. Transcription of such spacer and flanking sequences could provide endonucleolytic sites for
processing of multimeric transcripts into smaller tRNA precursors by RNase PC and P (23-25).

All of the tRNA genes identified in this study, except for the Ile<sub>3</sub> gene, have an encoded CCA sequence at their 3' termini. If the overlapping Pro-Ile<sub>3</sub> genes are transcribed in a single transcription unit (possibly including the fMet gene), maturation of tRNA<sub>3</sub><sup>lle</sup> would necessitate the cleavage of tRNA<sub>Pro</sub> sequences and the action of nucleotidyl transferase for the addition of CCA sequences to the 3' terminus of tRNA<sub>3</sub><sup>lle</sup>. However, this mechanism for tRNA<sub>3</sub><sup>lle</sup> maturation would preempt the maturation of complete tRNA<sub>Pro</sub> species. At present, we have no evidence concerning the expression of the Ile<sub>3</sub> gene in T5 infected cells.

Although we are unaware of overlapping tRNA genes in other organisms, the expression of overlapping reading frames that result in the synthesis of two different proteins was first recognized for bacteriophage φX174 (26) and has since been reported for a number of animal viruses (27-31).

ACKNOWLEDGEMENTS

This work was supported by the National Institutes of Health, Grant AI-12479.

*Present address: Abbott Laboratories, Department of Molecular Biology-L3, North Chicago, IL 60064, USA

*To whom reprint requests should be sent

REFERENCES