Proc. Natl. Acad. Sci. USA
Vol. 78, No. 2, pp. 889–892, February 1981
Biochemistry

DNA sequence of the transfer RNA region of bacteriophage T4: Implications for transfer RNA synthesis

GAIL P. MAZZARA*, GUY PLUNKETT, III, AND WILLIAM H. MCCLAIN†

Department of Bacteriology, The University of Wisconsin, Madison, Wisconsin 53706

Communicated by Oliver E. Nelson, Jr., November 3, 1980

ABSTRACT **Sequences encoding eight tRNAs and two stable RNAs of bacteriophage T4 are grouped together on the T4 genome in two clusters, separated by approximately 500 base pairs. The DNA sequence of part of this region was determined. Within each cluster coding sequences are separated by only one or a few base pairs. These findings imply that the RNAs may be processed from a single multimeric transcript, with initial endonucleolytic cleavages generating the previously characterized monomeric and dimeric precursors.**

The tRNAs encoded by bacteriophage T4 have provided a useful system for the study of tRNA biosynthesis. Bacteriophage T4 encodes eight tRNAs and two low molecular weight stable RNAs of unknown function. The nucleotide sequences of small precursor RNAs for most of these have been determined (ref. 1 and references cited therein; unpublished data), and the synthetic steps leading from these intermediates to the mature molecules are known (2). Less is known about the initial steps responsible for formation of the small precursor RNAs. We know that the small precursor RNAs are not primary transcripts, and evidence from *in vitro* transcription studies of the T4 tRNA coding sequences (3), which are grouped together in a small region of the T4 genome (Fig. 1) (4), indicates that these sequences are cotranscribed. However, attempts to isolate this large transcript from infected cells have thus far been unsuccessful. In the present study, restriction endonuclease fragments containing the coding sequences of seven of the stable RNAs were inserted into a plasmid vector and cloned to allow determination of their sequences. The sequence information thus obtained allows us to predict the entire biosynthetic pathway leading from the coding sequences to the mature RNAs.

RESULTS AND DISCUSSION

DNA Sequence Analysis. We determined the nucleotide sequences of T4 DNA encoding $tRNA^{Arg}$, $tRNA^{Ile}$, $tRNA^{Thr}$, $tRNA^{Ser}$, part of $tRNA^{Pro}$, and band C and band D stable RNAs, using the chain-termination method of Sanger *et al.* (5). Prior to nucleotide sequence determination, we cloned T4 restriction endonuclease fragments carrying these genes (Fig. 1) in the plasmid vector pBR322. Cloning procedures are given elsewhere (6). All cloning procedures followed the National Institutes of Health guidelines.

An *Eco*RI fragment carrying the sequences of $tRNA^{Arg}$, band D RNA, and band C RNA was inserted in both orientations into the unique *Eco*RI site of pBR322. Cleavage of these two recombinant plasmids at the unique *Hin*dIII site of pBR322, followed by exonuclease III digestion, generated two molecules in which opposite strands of the cloned DNA fragments were exposed for use as templates in the DNA sequence determination reactions. Similarly, the *Sau*3A fragment carrying T4 sequences encoding $tRNA^{Ile}$, $tRNA^{Thr}$, $tRNA^{Ser}$, and part of $tRNA^{Pro}$ was inserted into the unique *Bam*HI site of pBR322. Template DNA for sequence analysis was then prepared by exonuclease III digestion after cleavage at either the *Hin*dIII or the *Sal* I site of pBR322. For all plasmids, smaller restriction endonuclease fragments from within the cloned T4 DNA or from adjacent regions of the plasmid DNA served as primers in the DNA sequence determination reactions. Figs. 2–5 summarize the sequencing strategy and the sequence data thus obtained.

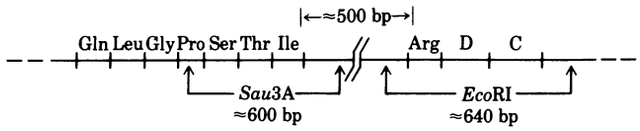

FIG. 1. Region of the T4 genome encoding the stable RNAs. bp, Base pairs.

Implications for Processing. The sequences of the 10 low molecular weight RNAs have been determined‡, and they match the DNA sequences. Knowledge of the DNA sequence allows us to make some predictions as to how the coding sequences are transcribed and processed. As Figs. 4 and 5 show, the known small precursor RNA sequences§ are contiguous in the DNA within each cluster. We surmise from this arrangement that the small precursor RNAs are generated from a common transcript by endonucleolytic cleavages between precursor RNA segments.

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.

* Present address: The Biological Laboratories, Harvard University, Cambridge, MA 02138.
† To whom reprint requests should be addressed.

‡ The sequences of all eight tRNAs and band C RNA have been published (ref. 1 and references cited therein). We recently established the sequence of band D RNA from a combination of RNA and DNA sequence determination data. Band D RNA uniformly labeled with ^{32}P was digested with RNase T1 and the products were subjected to two-dimensional analysis (7). The product that contained a 5′-monophosphate was pA-U-G, and that which lacked a 3′-terminal G was $C\text{-}A\text{-}C\text{-}C\text{-}A_{OH}$. Examination of Fig. 4 shows that C-A-C-C-A occurs only once, ending at residue 319. Potential 5′-terminal A-U-Gs occur starting at residues 200, 238, and 285. We concluded that residues 200–319 code for band D RNA because only this segment would produce an RNA compatible with the RNase T1 products obtained from band D RNA.

§ *E. coli* A49 (which has temperature-sensitive RNase P) incubating at 42°C was infected with wild-type phage T4. Purification and fingerprint analysis of ^{32}P-labeled RNA (7) showed that mature band C RNA was replaced by a precursor RNA that had a 5′ extension of six residues. The sequence of the extension was that expected from the DNA sequence in Fig. 4. Sequences of precursor forms of seven of the tRNAs had been determined previously (ref. 1 and references cited therein; unpublished data). No precursor forms of $tRNA^{Arg}$ or band D RNA have been observed as yet.

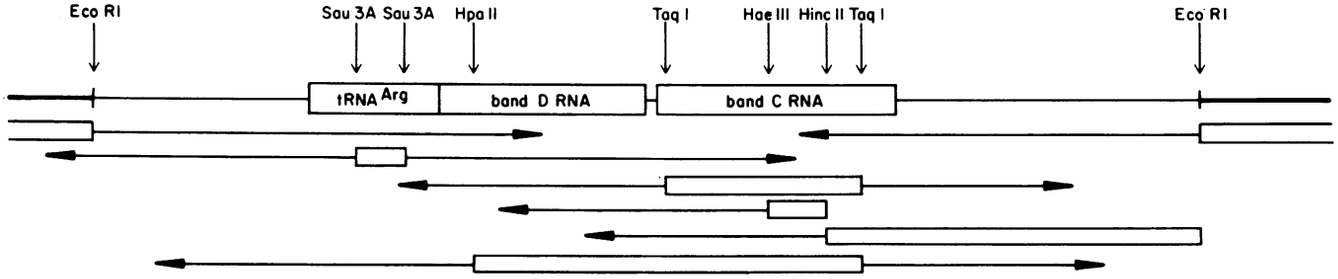

FIG. 2. Schematic representation of the *Eco*RI fragment encoding T4 $tRNA^{Arg}$, band D RNA, and band C RNA. The top line shows the structure of the fragment, including the locations of the T4 RNA coding sequences; the thick lines represent pBR322 DNA. Vertical arrows indicate restriction endonuclease sites used to generate primer fragments. Below this line, boxes represent individual primers used for DNA-sequencing reactions, with arrowed lines indicating the direction of DNA synthesis and the extent of DNA sequence information obtained from each primer. Rightward arrows designate sequence determination of the coding strand; leftward arrows, of the opposite strand.

The predicted common transcript, containing all of the T4 stable RNAs, has not yet been observed in infected cells. However, results from *in vitro* transcription studies indicate that the T4 tRNA coding sequences can be transcribed into a single RNA encompassing all ten coding sequences (3). The initiation site (and presumably the promoter) for this primary transcript is about 1 kilobase upstream from $tRNA^{Gln}$ (to the left in Fig. 1) (4) and hence is not included in the sequences presented here. Examination of the sequence 48 residues downstream from that encoding band C RNA reveals a potential terminator sequence (see Fig. 4). This sequence has the following features in common with known terminators: (*i*) a region of hyphenated dyad symmetry preceding the (presumptive) termination site; (*ii*) a run of U residues at the terminus of the transcript (in this case, five corresponding T residues are seen in the DNA strand that is identical to the RNA transcript); and (*iii*) G+C-rich sequences preceding the stop site (in this sequence, the potential stem and loop formed by the region of dyad symmetry would contain seven G·C base pairs) (8).

The *in vitro* transcription studies also indicated that a ρ-dependent termination of transcription could occur somewhere between the sequences encoding $tRNA^{Ile}$ and $tRNA^{Arg}$ (3). The sequence of this region has not been completely determined, but the available sequence was examined for potential terminators. The best candidate (none were as good as the one described above) is indicated in Fig. 5; surprisingly, this sequence falls *within* the $tRNA^{Ile}$ sequence. If it is functional, the consequences with respect to $tRNA^{Ile}$ synthesis would be of interest—termination of transcription within $tRNA^{Ile}$ could result in yields of $tRNA^{Ile}$ lower than those of $tRNA^{Thr}$. Two experimental observations are relevant in this context. First, the yield of $tRNA^{Ile}$ in T4-infected cells is low relative to $tRNA^{Thr}$. Second, in hosts lacking RNase P activity, a monomeric precursor to $tRNA^{Thr}$ is found in roughly the same (or greater) molar yield as the dimeric $tRNA^{Thr}$-$tRNA^{Ile}$ precursor, whereas no comparable $tRNA^{Ile}$ precursor has been observed (ref. 1; unpublished data). Any relationship between these observations and the proposed terminator structure is, however, purely speculative.

Fig. 4 also shows a potential promoter sequence about 100 base pairs upstream from $tRNA^{Arg}$. The boxed sequence is an excellent match with the strongly conserved "Pribnow sequence," found in the −10 region of prokaryotic promoters, and the 6 residues immediately upstream agree well with the weakly conserved sequence homology found there (8). We cannot say anything about the −35 (or "recognition") region of this potential promoter, because the *Eco*RI cleavage site that generated the restriction fragment was between the −10 region and any potential −35 region. The sequence downstream from the Pribnow sequence also shows some homology with weakly conserved sequences of functional promoters; if this is a functional promoter, transcription would probably start at or near residue 30, resulting in a 93-nucleotide leader prior to the mature $tRNA^{Arg}$ sequence.

The relationship of the promoter and terminators described above to the *in vitro* transcripts described by Goldfarb and Daniel (3) is unknown, and none of them has been shown to function *in vitro* or *in vivo*.

The endonuclease responsible for cleavages of the putative initial transcript (or transcripts) to generate the previously characterized precursor RNAs has not been identified. Ribonuclease III may function to remove most of the transcribed residues to the 5′ side of $tRNA^{Gln}$ (9), but it is not a likely candidate for the other cleavages required. We should consider the possibility that these cleavages need not proceed via a unique enzyme, but can instead be mediated by any of a number of cellular or T4-encoded ribonucleases. Examination of the sequences within the two clusters of RNA coding regions reveals short A+U-rich regions separating the stable RNA sequences. Conceivably these stable RNA sequences could be protected from nucleolytic cleavage by virtue of their tRNA-like conformations; by contrast, the A+U-rich regions would be unprotected and hence susceptible to a variety of ribonuclease activities. It has been shown experimentally that, at least within the small dimeric precursors, the individual tRNA sequences assume tRNA-like conformations (10). By assuming a tRNA-like structure while still part of the primary transcript, the stable RNA sequences may both limit and direct cleavage of that transcript.

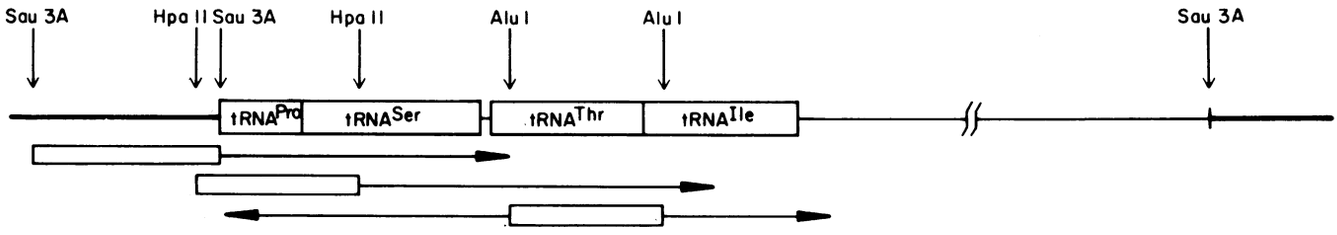

FIG. 3. Schematic representation of the *Sau*3A fragment encoding several T4 tRNAs. All conventions are as in Fig. 2.

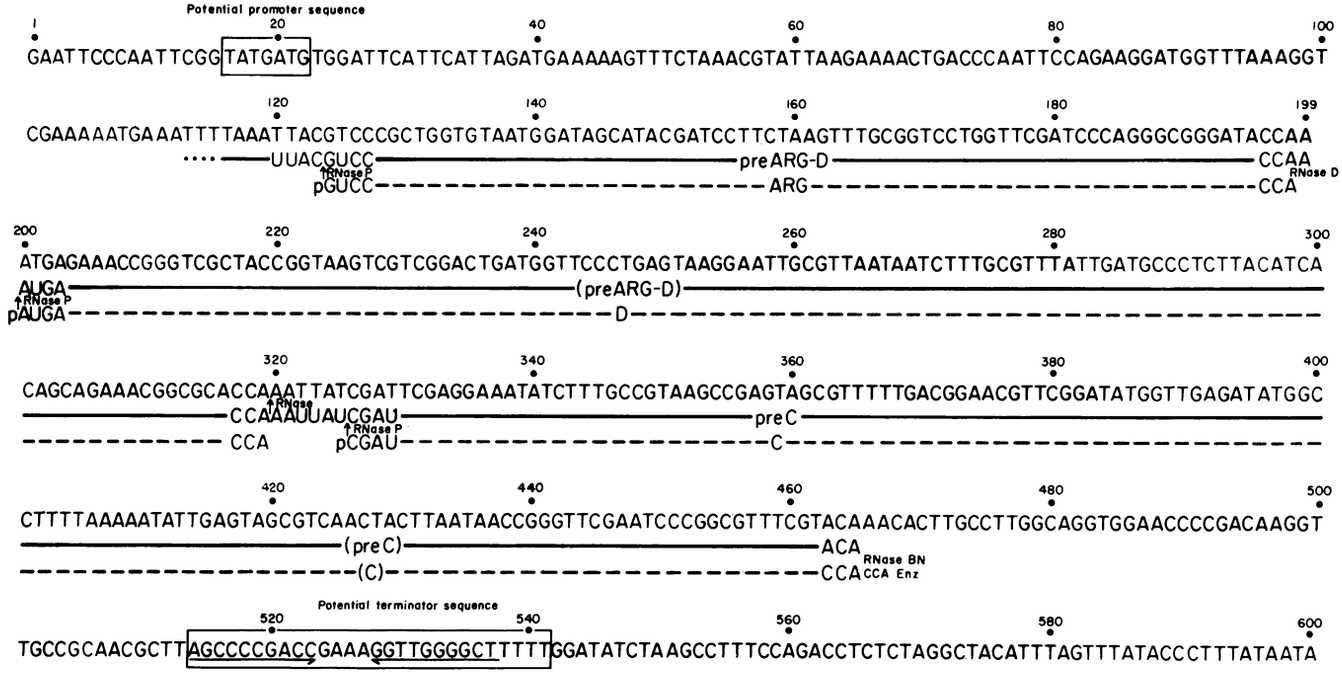

FIG. 4. Nucleotide sequence of the *Eco*RI fragment. The sequence shown is from the strand identical to the RNA transcript; transcription occurs on the complementary strand, whose sequence is not shown. Nucleotide 1 corresponds to the 5′ end of the restriction site defining one end of the fragment. Boxed sequences are potential transcriptional regulatory sequences, with arrowed lines indicating regions of dyad symmetry in the potential terminator. Below the DNA sequence, the corresponding precursor RNAs (solid lines) and mature RNAs (broken lines) are represented, with 5′- and 3′-terminal sequences as indicated. Individual tRNAs are identified by their cognate amino acids. The dimeric precursor to $tRNA^{Arg}$ and band D RNA (preARG-D) is hypothetical, based in part on the absence of mature forms of these molecules in infected cells lacking active RNase P. Vertical arrows indicate known or predicted sites of endoribonucleolytic cleavages during processing. RNase, endonuclease(s) responsible for cleavage of primary transcript(s). RNase BN, the 3′-exonuclease activity missing in *Escherichia coli* BN. CCA Enz, tRNA nucleotidyltransferase.

In Figs. 4 and 5, the activity or activities responsible for these cleavages are indicated simply as RNase (without further identification), and its (their) predicted sites of action are indicated. These cleavages may occur subsequent to transcription or concomitantly with transcription as is the case for rRNA processing in *E. coli* (11).

Once the primary transcript is cleaved to produce the smaller precursors, processing follows the pathways previously described (2, 11). In all cases RNase P generates the 5′ ends of the mature RNA species. Maturation of the 3′ termini occurs via two different pathways. When the complete C-C-A sequence is absent in the precursor, as is the case with $tRNA^{Pro}$, $tRNA^{Ser}$, $tRNA^{Ile}$, and band C RNA, RNase BN acts to exonucleolytically remove extra nucleotides from the 3′ termini of the precursor RNAs. Completion of the 3′ C-C-A sequence is then mediated by tRNA nucleotidyltransferase. In the case of $tRNA^{Thr}$, $tRNA^{Arg}$, and band D RNA the 3′-terminal C-C-A is transcriptionally derived, occurring in the small precursors as C-C-A-N. The extra residues (represented as N) are removed by a 3′-exonuclease other than RNase BN; RNase D is a likely candidate for this activity (12). The processing pathway also includes the modification of specific nucleosides within tRNA sequences. The modified nucleosides, with the exception of 2′-*O*-methylguanosine, are present in the smaller precursors, implying that these modifications occur prior to the cleavage events that generate the monomeric and dimeric precursors. On the other hand, the ribose methylation to produce 2′-*O*-methylguanosine is a terminal step in tRNA maturation (13). The product of at least one T4 gene, provisionally termed M1 or *mb*, may also be involved in the biosynthesis of T4 tRNAs; its role is unknown at present (14).

Recently, Abelson (4) reported the DNA sequence encoding

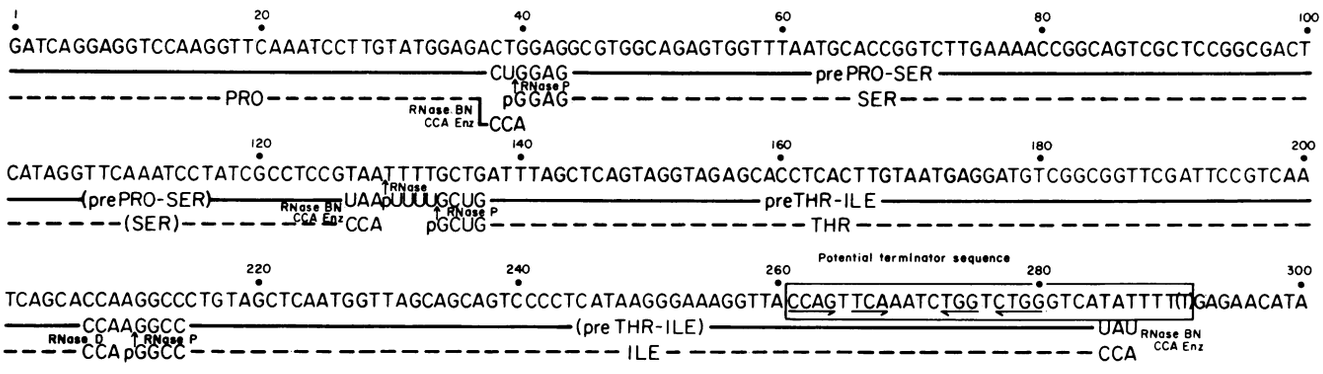

FIG. 5. Nucleotide sequence of the *Sau*3A fragment. All conventions are as in Fig. 4. The monomeric precursor to $tRNA^{Thr}$ (see text) is not indicated; it corresponds to residues 130–209 (with some forms longer at the 3′ ends).

seven T4 tRNAs. The reported DNA sequence agrees with that presented here, and, in addition, shows that the remainder of the precursor RNAs are also contiguous in the DNA sequence in the order shown in Fig. 1. Subsequent to submission of this manuscript, Fukada and Abelson (15) published the data on which the previously reported DNA sequence was based, and they presented a processing model with some similarities to ours. They also presented preliminary sequences for regions encoding the termini of $tRNA^{Arg}$ and band C and band D RNAs; these preliminary sequences agree with the sequence shown in Fig. 4, with one exception: their sequence shows the insertion of a C residue between residues 121 and 122.

We thank F. Sanger, B. G. Barrell, and A. J. H. Smith for instructing us in DNA sequence determination techniques. This work was supported by the U.S. Public Health Service (Grant AI 10257).

1. Guthrie, C. & Scholla, C. A. (1980) *J. Mol. Biol.* **139,** 349–375.
2. McClain, W. H. (1977) *Acc. Chem. Res.* **10,** 418–425.
3. Goldfarb, A. & Daniel, V. (1980) *Nature (London)* **286,** 418–420.
4. Abelson, J. (1979) *Annu. Rev. Biochem.* **48,** 1035–1069.
5. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
6. Mazzara, G. P. (1980) Dissertation (Univ. of Wisconsin, Madison, WI).
7. Barrell, B. G. (1971) in *Procedures in Nucleic Acid Research*, eds. Cantoni, G. L. & Davies, D. R. (Harper & Row, New York), Vol. 2, pp. 751–799.
8. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13,** 319–353.
9. McClain, W. H. (1979) *Biochem. Biophys. Res. Commun.* **86,** 718–724.
10. McClain, W. H. & Seidman, J. G. (1975) *Nature (London)* **257,** 106–110.
11. Mazzara, G. P., Plunkett, G., III & McClain, W. H. (1980) in *Cell Biology, A Comprehensive Treatise*, eds. Goldstein, L. & Prescott, D. M. (Academic, New York), Vol. 3, pp. 439–545.
12. Cudny, H. & Deutscher, M. P. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 837–841.
13. Guthrie, C., Seidman, J. G., Altman, S., Barrell, B. G., Smith, J. D. & McClain, W. H. (1973) *Nature (London) New Biol.* **246,** 6–11.
14. McClain, W. H., Guthrie, C. & Barrell, B. G. (1973) *J. Mol. Biol.* **81,** 157–171.
15. Fukada, K. & Abelson, J. (1980) *J. Mol. Biol.* **139,** 377–391.