The nucleotide sequence of threonine transfer RNA coded by bacteriophage T4

Christine Guthrie*, Clara Ann Scholla**, Harry Yesian** and John Abelson**

*Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94143 and ** Department of Chemistry, University of California, San Diego, La Jolla, CA 92036, USA

Received 13 March 1978

ABSTRACT

The nucleotide sequence of a low molecular weight RNA coded by bacteriophage T4 (and previously identified as species a) has been determined. The molecule is of particular biological interest for its associated biosynthetic properties. This RNA is 76 nucleotides in length, contains eight modified bases, and can be arranged in a cloverleaf configuration common to tRNAs. The anticodon sequence is UGU, which corresponds to the threonine-specific codons ACU.

The nucleotide sequence was determined primarily by nearest-neighbor analysis of RNA synthesized in vitro using [α-32P]nucleoside triphosphates. Using the single-strand specific nuclease SI, two in vivo labeled half-molecules were generated and analysed. This information together with restrictions imposed by nearest-neighbor data, provided a unique linear sequence of nucleotides with the features of secondary structure common to tRNA molecules.

INTRODUCTION

Bacteriophage T4 codes for eight unique species of transfer RNA (1). Acceptor activities for arginine, glycine, isoleucine, leucine and proline have been previously identified (2,3). Nucleotide sequences of tRNAs specific for serine (4), proline (5), glycine (6,7), glutamine (8), leucine (9) and arginine (10), have now been determined. A primary biological interest in this system derives from the existence of biosynthetic intermediates which can be easily detected (11). To date, two dimeric precursor species have been extensively characterized: the precursor to proline and serine tRNAs (5) and that to glutamine and leucine tRNAs (12). A third dimeric precursor has been shown to contain the sequences of tRNA species designated a and δ (11). This precursor is of particular interest because the cognate mature tRNAs are found in distinctly unequal amounts in the infected cell (1; Guthrie, manuscript in preparation). Knowledge of the nucleotide sequences are prerequisite to an evaluation of the contribution to this phenomenon by the tRNA and precursor structure.
We report here the nucleotide sequence of the low molecular weight RNA component previously designated α (1). Initial characterization of molecules labeled uniformly in vivo was performed using the standard RNA sequencing techniques of Sanger and colleagues (13). The sequence determination was based on nearest-neighbor analysis of RNA synthesized in vitro using [α-32P]nucleoside triphosphates. Using these data and the canonical features of tRNA secondary structure, a consistent linear sequence could be established. The composition of half-molecules, generated by the single-strand specific nuclease SI, places such constraints on the sequence overlaps that this information, together with nearest neighbor data, allow the prediction of a unique nucleotide sequence.

The tRNA contains 76 nucleotides. Arranged in a cloverleaf structure, the anticodon corresponds to threonine-specific codons; the amino acid acceptance has not been tested.

MATERIALS AND METHODS

Synthesis and Isolation of 32P-Labeled RNA.

In vivo RNA was obtained from wild type T4D infection of Escherichia coli strain B/5 in the presence of [32P] as described previously (14). The in vitro synthesis of tRNA\textsuperscript{Thr} using T4 DNA as a template in a coupled system containing purified RNA polymerase and a 0.2 M NH\textsubscript{4}Cl ribosomal wash from T4-infected E. coli strain A19 is detailed elsewhere (15). [α-32P]nucleoside triphosphates were purchased from New England Nuclear Corp. In both cases the labeled RNA was resolved by two steps of polyacrylamide gel electrophoresis and eluted from the gel as described earlier (14).

Sequence Analysis

The techniques, methods and materials used were those described in detail by Barrell (13). Analysis of modified nucleotides following complete digestion by RNase T\textsubscript{2} was as described previously (14).

Nuclease SI Digestion of tRNA

SI nuclease was purified essentially according to the method of Vogt (16); and is described elsewhere (17). The digestion mixture (240 μl) contained 0.3 M NaCl, 0.05 M Na acetate, pH 4.5, 3 x 10^{-3} M ZnCl\textsubscript{2}, =5 x 10^{6} cpm of [32P]labeled in vivo tRNA, 20 μg of carrier tRNA and 10 units of SI. The reaction was carried out for 30 min at 15°C and terminated by addition of EDTA (1 mM), 25 μg polyU and 2.5 volumes of ethanol. The precipitated RNA was resuspended in H\textsubscript{2}O and fractionated on a 20% polyacrylamide gel (17).
RESULTS

1. Sequence analysis of in vivo RNA

[^32]P-labeled RNA synthesized in vivo was subjected to RNase T₁ digestion and fractionated as shown in Figure 1A. Each oligonucleotide was quantitated and its composition determined by alkaline hydrolysis and by digestion with RNase T₂ to allow identification of minor nucleotides (Figure 2). Oligonucleotides were further characterized by analysis of the products of digestion with pancreatic RNase A, as described in Table 1. These operations allowed determination of the sequences of the small oligonucleotides T₁, T₂, T₃, T₅, T₆, T₇ and T₉.

Oligonucleotide T₄ is the only product containing a 3' hydroxyl group. Alkaline hydrolysis gave Cp(3):Ap(1), while complete digestion with snake venom phosphodiesterase yielded pC(1):pA(1), indicating the sequence C(C₄,A)A₀H⁹. Digestion with RNase A gave 2C:1AC, indicating the sequence C(AC,C)A₀H⁹. Digestion with RNase U₂ to give the products CA and CCA₀H⁹ established the sequence as CACCA₀H⁹.

Similar analyses were performed on oligonucleotides obtained from complete digestion of the RNA with RNase A (Figure 3). In this case the products were further characterized by digestion with RNase T₁. As shown in Table 2, these procedures allowed the determination of the sequences of all the oligonucleotides except T₁₃ and T₁₄.

2. Sequence analysis of in vitro RNA

Comparison of Figures 1A and C reveals that the RNase T₁ oligonucleotide pattern of tRNA synthesized in vitro with [α-[^32]P]GTP differs from that of in vivo RNA in several respects. Sequence analysis of the oligonucleotides (Table 1) reveals that these differences can be accounted for by the absence of most nucleotide modifications from RNA synthesized in vitro. Thus the in vivo sequence DAG (T₆) is replaced by UAG (T₈) in in vitro RNA; the molar yield is unchanged. Oligonucleotide T₁₂ has the partial sequence (AU, me GU, C)G. This product is not found in in vitro RNA which contains the oligonucleotide (T₉) AUG(U) in full molar yield. In addition, T₇ is now found in two molar yield and is shown to contain the sequence isomers CUG and UCG. These data suggest that the sequence of T₁₂ is AUme⁷GU₇G. (Since the terminal G of T₇ is labeled by both GTP and ATP it is not possible to determine the nearest neighbor of T₁₂ from this information). Thus in vivo there is nearly full modification (AUG is found in only 0.2 molar yield), while no me⁷G is synthesized in vitro. The final difference between the in vivo and in vitro oligonucleotide
Figure 1. Products of RNase T1 digestion of threonine tRNA were fractionated on cellulose acetate at pH 3.5 in the first dimension and on DEAE paper with 7% formic acid in the second dimension. Fig. 1A shows digestion products of threonine tRNA labeled in vivo with \(^{32}\)P; Fig. 1C shows products of this tRNA synthesized in vitro with \([\alpha-^{32}\text{P}]\)GTP; Fig. 1B is a composite linedrawing, indicating products found in both in vivo and in vitro RNA (•), in in vivo RNA only (○) and in in vitro RNA only (©). The sequences of the numbered products are given in Table 1. The long exposure of the autoradiograph in panel A reveals several oligonucleotides near the origin which are not labelled; they correspond to sequences from a tRNA species which has a similar electrophoretic mobility and thus contaminates the tRNA\(^{\text{Thr}}\) preparation.
Figure 2. Composite line drawing of component nucleotides of in vivo $^{32}P$-labeled tRNA Thr digested with RNase T2 and fractionated by two-dimensional thin layer chromatography. Up is found in several positions of the chromatogram. On 540 paper at pH 3.5, Up has a mobility slightly faster than G ($R_G = 1.12$) and $\hat{A}p$ slightly slower than G ($R_G = 0.90$). Abbreviations and other details are given in the legend to Table 1.

patterns is the absence of T16 and T16' in Figure 1C and the concomitant appearance of T15. T16 and T16' contain two minor bases, $\Psi$ and a modified A residue. They both have the partial sequence (AAΨ, U)G and appear to differ only in the state of modification of A, which causes the oligonucleotides to migrate to different positions in the fingerprint. T15 is shown (Table 1) to have the sequence UAAΨG. (It is interesting to note that the $\Psi$ modification is synthesized in vitro.) We conclude that the sequence of T16 and T16' is UAAΨG; the placement of $\hat{A}$ is based on the location of this oligonucleotide in the cloverleaf configuration (Figure 6): a hypermodified A residue is commonly found adjacent to the 3' end of the anticodon.

As shown in Tables 1 and 2, all of the oligonucleotide sequences can be directly deduced from the nearest neighbor data with the exception of T18. Determination of this sequence is described in Table 3.

3. Determination of Sequence Overlaps.

Data obtained from nearest neighbor analyses (Tables 1, 2) provide substantial information on the order of oligonucleotides in the linear sequence. Analogous information from in vivo analysis is primarily
Oligonucleotides as shown in Figure 1 were quantitated by scintillation counting and molar yields calculated by normalizing to T10. The molar yields reported here are the average of three RNA preparations. Suggested molar yields were determined from the sequence shown in Figure 5. The composition of each oligonucleotide was determined by alkaline hydrolysis; each nucleotide was quantitated relative to the yield of Gp. These data are not given here but were consistent in each case with the composition of digestion products of pancreatic RNase A. These products were identified by electrophoretic mobility and by their alkaline hydrolysis products; the yields of RNase A products are given in moles of oligonucleotide normalized to pyrimidine content. The molar yields of in vitro labeled oligonucleotides are normalized to the nearest integer. Nearest neighbors are given in brackets. N.L. indicates a product was not labeled by a given input triphosphate. VP'ase is snake venom phosphodiesterase. Other sequence methods are given in the legend of Table 3. Modified bases were identified as in Figure 2. Modifications are abbreviated as follows: D, dihydrouridine; T, ribothymidine; $\Psi$, pseudouridine; me$^7$G, 7-methylguanosine. A and U are unknown modifications of A and U, respectively, as determined from the unmodified sequences found in in vitro RNA.

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Molar Yield in vivo</th>
<th>Pancreatic RNase A Digestion Products</th>
<th>Labeled Products of Pancreatic RNase A Digestion from $\alpha$-32P Input label</th>
<th>Deduced Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td>4.2</td>
<td>F.L.</td>
<td>G.A + G.C + G.U</td>
<td>CACACACCCAC</td>
</tr>
<tr>
<td>T2</td>
<td>0.9</td>
<td>C(1.1), G</td>
<td>G(1), C(2)</td>
<td>TTTCCAAAGCC</td>
</tr>
<tr>
<td>T3</td>
<td>1.8</td>
<td>2 AG</td>
<td>U(1), C(1)</td>
<td>UUCCCAAGCC</td>
</tr>
<tr>
<td>T4</td>
<td>0.7</td>
<td>1 AC, C</td>
<td>C(1)</td>
<td>CACACACCCAC</td>
</tr>
<tr>
<td>T5</td>
<td>0.7</td>
<td>1 Gp</td>
<td>G(1)</td>
<td>CACACACCCAC</td>
</tr>
<tr>
<td>T6</td>
<td>1.0</td>
<td>C(0.9), G</td>
<td>G(1), C(1)</td>
<td>U(1), C(1)</td>
</tr>
<tr>
<td>T7</td>
<td>0.2</td>
<td>U of AG(1)</td>
<td>G(1), A or AG(1)</td>
<td>AGU(1)</td>
</tr>
<tr>
<td>T8</td>
<td>0.9</td>
<td>1 C(1.9), U(0.9), AG</td>
<td>A of AG(1)</td>
<td>C(1), U(1)</td>
</tr>
<tr>
<td>T9</td>
<td>0.9</td>
<td>1 T(1.1), C(1.1), G</td>
<td>C(1)</td>
<td>CACACACCCAC</td>
</tr>
<tr>
<td>T10</td>
<td>0.8</td>
<td>1 A(1.8), C(2.1), C(1.1), G(1)</td>
<td>A of AG(1)</td>
<td>A of AG(1)</td>
</tr>
<tr>
<td>T11</td>
<td>0.8</td>
<td>1 U(0.7), C(2.1), C(1.1), C(1)</td>
<td>A of AG(1)</td>
<td>A of AG(1)</td>
</tr>
<tr>
<td>T12</td>
<td>0.8</td>
<td>1 A(1.1), C(2.1), C(1.1), C(1)</td>
<td>A of AG(1)</td>
<td>A of AG(1)</td>
</tr>
<tr>
<td>T13</td>
<td>0.8</td>
<td>1 A(1.1), C(2.1), C(1.1), C(1)</td>
<td>A of AG(1)</td>
<td>A of AG(1)</td>
</tr>
<tr>
<td>T14</td>
<td>0.8</td>
<td>1 C(1.8), C(2.1), C(1.1), C(1)</td>
<td>A of AG(1)</td>
<td>A of AG(1)</td>
</tr>
<tr>
<td>T15, T16</td>
<td>0.9</td>
<td>1 A(1.1), U(0.9), G</td>
<td>U of AG(1)</td>
<td>U of AG(1)</td>
</tr>
<tr>
<td>T17</td>
<td>0.9</td>
<td>1 A(1.1), U(0.9), G</td>
<td>U of AG(1)</td>
<td>U of AG(1)</td>
</tr>
<tr>
<td>T18</td>
<td>0.8</td>
<td>1 A(1.1), C(2.1), C(1.1), C(1)</td>
<td>C(2.1)</td>
<td>A of AG(1)</td>
</tr>
</tbody>
</table>

Footnotes:
1. See Figure 1
2. RNA synthesized with CTP was heavily contaminated where indicated by brackets, thus only qualitative conclusions can be drawn from these data.
3. G modification not found in in vitro RNA; in vivo RNA is close to fully modified.
4. Modification not found in in vitro RNA; see text for explanation of deduced sequence.
5. See Table 3 for explanation of deduced sequence.
dependent on the presence of modified nucleotides in RNase A and Tl
digestion products and is therefore of limited usefulness. Guided by
the features of the cloverleaf pattern governing tRNA secondary structure,
it is possible to predict a consistent linear sequence from the sum of
these data. It is desirable, however, to obtain independent information
establishing a unique nucleotide sequence. Examination of the predicted
sequence revealed that overlap ambiguities could be resolved if it were
possible to examine the two halves of the molecule independently. This
was achieved by digestion of the tRNA with the single strand specific
nuclease Sl, previously shown by Harada and Dahlberg (18) to split tRNAs
by an endonucleolytic cleavage in the anticodon loop. The half-molecules
thus generated were fractionated by polyacrylamide gel electrophoresis
and analysed following complete digestion with RNase Tl. As shown in
Figure 4, two distinctly unique patterns are obtained, although there is
some cross-contamination (e.g., Tl1, Tl2, Tl3, Tl7) at the level of 0.1-
0.2 molar yield. It can be seen that T3, present in 2 molar yield in
the intact molecule, segregates 1:1 between the two halves. The most
### Table 2

Sequences and Molar Yields of Pancreatic RHase A Oligonucleotides

<table>
<thead>
<tr>
<th>Oligo-</th>
<th>PI</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>P8</th>
<th>P9</th>
<th>P10</th>
<th>P11</th>
<th>P12</th>
<th>P13</th>
<th>P14</th>
<th>P15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide</td>
<td>8.0</td>
<td>12.5</td>
<td>1.3</td>
<td>1.1</td>
<td>1.2</td>
<td>1.0</td>
<td>0.9</td>
<td>3.2</td>
<td>1.1</td>
<td>1.1</td>
<td>1.1</td>
<td>1.0</td>
<td>0.9</td>
<td>0.8</td>
<td>0.7</td>
<td>1.0</td>
</tr>
<tr>
<td>CEIL.</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Molar Yield</td>
<td>8.9</td>
<td>10.0</td>
<td>1.3</td>
<td>2.6</td>
<td>1.1</td>
<td>1.0</td>
<td>0.7</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

**Table 3**

Sequences of Substrate and Digestion Products

<table>
<thead>
<tr>
<th>Nucleotide</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
<th>SI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Digestion</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
</tr>
<tr>
<td>Products</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
<td>U</td>
<td>T</td>
<td>C</td>
<td>G</td>
</tr>
</tbody>
</table>

**Footnote 1:** The assignment of the modified bases in this oligonucleotide is based on the arguments presented in the text and in the derivation of overlap sequences in Figure 5.

Products of digestion with RNase A, RNase U2 and CMCT (Ncyclohexyl-N'-β-morpholinyl-(4)-ethyl)-carbo diimide-methyl-P-toluene sulfonate) were identified by their electrophoretic mobility and by their alkaline hydrolysis products. Figures in parentheses indicate relative molar yield of each product. XTP signified one of the four [α-32P] nucleoside triphosphates used in the in vitro synthesis.
Figure 4. Products of RNase T₁ digestion of 5'(A) and 3'(B) half-molecules of tRNA^{Thr} labeled with ³²P in vivo generated by nuclease SI cleavage were fractionated as in Fig. 1. Oligonucleotides were quantitated and characterized by digestion with RNase A and by alkaline hydrolysis. Products present in molar yield are indicated as in Fig. 1 and Table 1. Product T₁₆ is present in only partial yield and T₁₈ is completely missing, consistent with their location in the anticodon (see text and Fig. 6). The arrows indicate new oligonucleotides generated by SI cleavage: (1) arises from T₁₈ and contains at least the first 6 nucleotides of that sequence; (2) is generated from T₁₆/T₁₆' and is tentatively identified as P₅A₅Ψ₉G (since SI is known to be associated with a 3' exonuclease activity (16-18), we presume this product to be more indicative than intact 5₁₆ of the initial site of SI cleavage); (3) results from removal by SI of the terminal adenosine of T₄.

Figure 5. The nucleotide sequence of tRNA^{Thr} showing overlap sequences, complete RNase A and T₁ digestion products, as well as the two SI partial digestion products (see Figure 4). Unambiguous overlaps can be constructed by using the SI data and the nearest neighbor data given in Tables 1 and 2. Five such overlapping sequences are shown, but many other blocks can be derived. To construct these sequences, start with a single product, for example P₁₁: AGAGC(A). There is only one T₁ product with ...AG(A); this is T₆/T₈(DAG(A)), giving the sequence (G)DAGAGC(A). There are two RNase A products ending ...GD(A). These are P₁₀' (AGD(A)) and P₁₃ (AGGD(A)). P₁₀' is uniquely linked to T₁₀, so P₁₃ must be used here, giving (pyrimidine) AGGDAGG(C(A)). T₆/T₈ plus P₁₀' and T₁₀ dictate the sequence CUCAGDAGGAGGC(A)). The SI data show that T₁₀ is in the 5' half of the molecule. The only T₁ product starting with CA... except T₄, which is in the 3' half of the molecule, is T₁₈. This finally dictates the sequence CUCAGDAGGAGGCACCUAC(U). By similar lines of reasoning, the other overlapping products shown can be derived.
striking observation is that oligonucleotides T16 and T18 are not found in either fragment, strongly predicting that these comprise the anticodon region of the tRNA (see Figure 6). Concomitantly, several unique oligonucleotides can be seen in the digested fragments. Sequence analysis identifies them as products of S1 cleavage within the missing oligonucleotides (see legend to Figure 4).

Using the constraints imposed by these data, it was then possible to derive oligonucleotide overlaps which determine a unique linear sequence. The critical overlaps are indicated in Figure 5. This sequence is identical with that predicted from secondary structure arguments and is shown in Figure 6.

DISCUSSION

The nucleotide sequence is shown in cloverleaf form in Figure 6. The tRNA is 76 nucleotides in length. The anticodon sequence UGU corresponds to the threonine specific codons ACg3 if the modified U can recognize both A and G in the wobble position. Codon-directed binding and amino acid acceptance of this tRNA have not been tested.
The data amply demonstrate the power of nearest neighbor analysis in direct sequencing of oligonucleotide products and in ordering these products. Proof of a unique linear sequence has been traditionally demonstrated by the analysis of limited ribonuclease digestion products to obtain overlapping fragments. We have been able to circumvent this process by taking advantage of the sequence constraints derived from consideration of the 5' and 3' halves of the molecule separately. This was enabled by the single-stranded specificity of nuclease S1, which cleaves tRNAs endonucleolytically in the anticodon loop (17,18). These restrictions, together with the nearest neighbor data, allowed the determination of a unique linear sequence identical to that predicted from the features of secondary structure common to tRNA molecules.

There are no apparent unusual features of this tRNA sequence. Identification of structural anomalies were of particular interest in view of the biosynthesis of this tRNA. It arises on a dimeric precursor molecule which also contains the 6 tRNA sequence (11), yet the ratio of the cognate mature tRNAs is distinctly asymmetric: there are 3-4 copies of each threonine tRNA for every 6 tRNA in Th infections (l; Guthrie, manuscript in preparation). The possibility of multiple gene copies now appears extremely unlikely (K. Fukada, pers. comm.), nor have we found here any evidence for sequence heterogeneity. The determination of the nucleotide sequence of the precursor should aid in elucidating the mechanism of this differential gene expression.

ACKNOWLEDGMENTS

This work was supported by U.S. Public Health Service grants to C.G. (GM 21119) and to J.A. (CA 10984) and by the Cancer Research Coordinating Committee from the University of California to C.G.

+Present address: Department of Physiological Chemistry, Brown University, Providence, RI 02912, USA

REFERENCES

17. Manale, A., Colby, D., and Guthrie, C. manuscript submitted for publication.