The Nucleotide Sequence in the Promoter Region of the Gene for an Escherichia coli Tyrosine Transfer Ribonucleic Acid*

(Received for publication, June 10, 1974)

TAKAO SEKIYA, HANS VAN ORMOND'T,† AND H. GOBIND KHORANA
From the Departments of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

SUMMARY

The sequence of 29 nucleotides in the region preceding the initiation of the transcription of the Escherichia coli tyrosine tRNA gene has been determined. This is:

(5') ---GCCGACCTACATACAAATTGACGCCGCG--- (3')
(3') ---CCGCGTAGTATAGTTTACTGCCGCC--- (5')

The general approach used for the sequence determination involved the DNA polymerase I-catalyzed elongation of suitable deoxyribopolynucleotide primers when hybridized to the l-strand of φ80psuH'' DNA at the appropriate site. Sequences of the newly grown oligonucleotide chains were determined by a combination of two-dimensional fingerprinting following partial exonucleolytic degradation, nearest neighbor analyses, and determination of pyrimidine tracts. Primer elongations were carried out in a controlled and stepwise manner and the newly grown oligonucleotide chains were kept short by incorporating the following features into the method: (a) the insertion of a ribonucleotide unit at or near the 3' terminus of the primers; (b) the use of a maximum of three nucleotide 5'-triphosphates in the first stage of the elongation reaction, isolation of the elongated primer, and its reuse in a second step together with different sets of deoxyribonucleoside triphosphates; and (c) elongation of the primer using all of the four nucleoside triphosphates with one of the triphosphates being supplied in a limiting concentration.

The methodology now exists for the synthesis of biblical DNA with defined nucleotide sequence and the total synthesis of the DNA corresponding to a yeast alanine tRNA has been accomplished (2). For many biochemical studies, controlled transcription of the synthetic DNAs is necessary. Model studies carried out with short single- and double-stranded DNAs as templates for the Escherichia coli DNA-dependent RNA polymerase showed (3, 4) that there was a lack of specificity in both the initiation and termination of transcription. Consequently, the RNA products were heterogeneous and, furthermore, extensive synthesis of the transcripts was not realized. Therefore, with the aim of realizing properly initiated and terminated synthesis of an RNA, a system was chosen which would allow the investigation of the promoter and terminator regions as well. Thus, work was initiated on the synthesis of the DNA corresponding to the E. coli tyrosine tRNA. The gene for this tRNA can be inserted into the transducing bacteriophage φ80 (5) and biochemical studies using the transducing bacteriophage have clearly defined the initiation and termination sites for the transcription of this gene by the isolation and characterization of a precursor to the tRNAi (6). Further, the transducing bacteriophage containing the above tRNA gene provides a convenient starting material for structural work in the promoter and terminator regions by the approach described previously (1, 9) and below.

Therefore, concurrently with the synthetic work on the DNA corresponding to the precursor for the above tRNA, investigations of the nucleotide sequences in the promoter and terminator regions of this gene was undertaken. In previous papers, the sequence of 23 nucleotides beyond the C-C-A end of the above tRNA gene, the terminator region, has been reported (1). We now wish to report on the sequence of 29 nucleotides in the promoter region of this gene, the region preceding the starting point of the initiation of transcription. Preliminary accounts of this work have already appeared (10, 11).

The general approach used in the sequencing work involves the separation of the strands of the bacteriophage φ80psuH DNA carrying the gene for the tyrosine suppressor tRNA, the hybridization of appropriate deoxyribopolynucleotide primers at the tyrosine tRNA gene terminal in the r- or l-strand of the above DNA, and the controlled DNA polymerase-catalyzed synthesis of an RNA, a system which would allow the investigation of the promoter and terminator regions as well.
Fig. 1. Experimental plan for sequencing and the nucleotide sequences determined in the promoter region of the tyrosine tRNA gene. The primer-template complexes were initially obtained by hybridizing DNA I through DNA III to the 1-strand of the &delta;80psu DNA. DNA polymerase-catalyzed elongations were carried out using the triphosphates shown. The new nucleotide sequence discovered after each elongation and subsequent alkaline cleavage and analysis is shown in the appropriate dashed box. The asterisk after C, the 33rd nucleotide in the elongated primers, indicates that this position was occupied partly by dC and partly by rC. This was because the DNA III used contained two components; one contained eight nucleotides into the promoter region, whereas the second was shorter by two nucleotides (rC-A) at the 3' end. Elongation of this mixture with the deoxyribonucleotide triphosphate mixtures shown, therefore, gave dC in addition to rC at the seventh nucleotide in the promoter region.

Fig. 2. The nucleotide sequence in the promoter region of the tyrosine tRNA gene. Elements of 2-fold symmetry in the sequence are shown in the boxes with matching arrows pointing to the axis of symmetry.

Fig. 3. Fingerprint of the product obtained from the terminal nucleotidyltransferase-catalyzed addition of rG units to the dodecanucleotide, d-G-C-T-C-C-C-T-A-T-A-T-C-C. Conditions for the enzymatic reaction and isolation of the product have been given in the text. Partial degradation was with venom phosphodiesterase. Homochromatography in the second dimension was performed using Homomix II.

EXPERIMENTAL PROCEDURE

Materials and Methods

Except for the following, these were as described in a previous paper (1). DNA I—This was prepared by the T<sub>4</sub> polynucleotide ligase-catalyzed joining of the chemically synthesized oligonucleotides 5'-G-P-d-G-C-T-C-C-C-T-A-T-C-G and the unphosphorylated d-T-A-C-T-G-G-C-C-T in the presence of the complementary oligonucleotides. The details will be described elsewhere.

DNA II—The chemically synthesized and 5'-phosphorylated oligonucleotide, 5'-G-P-d-G-C-T-C-C-C-T-A-T-C-G was elongated primers at their 3' ends and determination of the nucleotide sequence of the newly added nucleotides to the primers. The plan of the present experiments for the determination of the promoter sequence and the results obtained are illustrated in Fig. 1. The primer, DNA I, was extended by a guanine ribonucleotide unit at the 3' end (DNA II). Controlled chain elongation of the latter was performed by using dATP, dGTP, and dCTP. The new nucleotide chain thus formed was isolated by alkaline cleavage at the rG site. Its sequence was shown to be as in the dashed box in A (Fig. 1). Next, the primer, DNA I, was first elongated using the three triphosphates mentioned above except that rCTP replaced dCTP. The product, designated DNA III (Fig. 1), was elongated further in the presence of dATP, dTTP, and dCTP. The addition of 11 new nucleotides was now observed in the major product and their sequence was determined after cleavage at the rC sites (dashed box in B, Fig. 1). In a third experiment, DNA III was extended using dATP, dCTP, dGTP, and a very low concentration of dTTP. One of the products formed contained the new decanucleotide sequence shown in the dashed box in C (Fig. 1). The total sequence thus obtained is shown in the double-stranded form in Fig. 2.

We are extremely grateful to Dr. R. Roychoudhury for his assistance and for suggesting the use of cobalt ions in this reaction.
residue was dissolved in 15 μl of water, and to the solution were added 15 μl of 60% glycerol containing xylene cyanol and bromphenol blue as dye markers.

The solution was placed on a 15% polyacrylamide gel slab (20 × 20 × 0.2 cm) containing 7 M urea and electrophoresis was performed in 90 mM Tris-base, 90 mM boric acid, and 4 mM EDTA for 18 hours at 10 mA. The radioactive bands were eluted from the gel with 2 μl TEAB. The elute was evaporated in vacuo to dryness, and the residue was dissolved in 1 ml of H2O and desalted by filtration through a Sephadex G-50 column (0.9 × 24 cm).

Isolation of Oligonucleotides Containing New Nucleotide Sequences

Elongated primer DNAs, purified as above, were subjected to alkaline hydrolysis in 0.5 M KOH for 18 hours at 37°. After neutralization with acetic acid, the products were separated by homochromatography on DEAE-cellulose thin layer plates (13). The radioactive bands were scraped and the powder was collected in small plastic conical tips (Eppendorf). The cellulose powder was first washed with 5 ml of ethyl alcohol to remove urea and then with 500 μl of 2 μl TEAB. The eluate was evaporated in vacuo to dryness.

Nucleotide Sequence Determination

Complete degradation of the radioactive oligonucleotides to 3'- or 5'-mononucleotides was performed by the previously described procedures (14, 15). Degradation to isolate the pyrimidine tracts and radioactive inorganic phosphate was performed by the method of Burton (16). The reaction mixture was applied to DE81 paper directly and the products were separated by electrophoresis in 7% formic acid. The papers were autoradiographed and the radioactive spots were cut out and eluted with 2 μl TEAB for further analysis.

Partial digestion with snake venom phosphodiesterase was carried out in 10 μl glycine buffer (pH 9.2) at 20° for 20 min with 50 to 1200 μg per ml of the commercial enzyme preparation. The partial digestion with spleen phosphodiesterase was carried out at 20° for 30 min in 10 μl Pipes buffer (pH 6.9) with 10 μM MgCl2 and 1 to 10 units per ml of the enzyme. After the reaction, cold ethanolic precipitation of the oligonucleotides was carried out. The precipitation residue was dissolved in 500 μl of 2 M TEAB. The eluate was evaporated in vacuo to dryness, and the residue was dissolved in 1 ml of H2O. The radioactive bands were scraped and the powder was collected in small plastic conical tips (Eppendorf). The cellulose powder was first washed with 5 ml of ethyl alcohol to remove urea and then with 500 μl of 2 μl TEAB. The eluate was evaporated in vacuo to dryness.

RESULTS

Hybridization of DNA 1 to l-strand of ϕ8psω11 DNA—The results of two experiments are shown in Table I. Thus, DNA I, and therefore DNA II and DNA III, can hybridize to the l-strand at only one specific site. Elongation of the primer while it forms a complex with the template should give a unique sequence and the 5 nucleotide units first incorporated should be the ones known to be at the 5' terminus of the RNA precursor and to be lacking in the primers.

Sequence of First Eight Nucleotides Preceding Start of Transcription

The primers (DNA I to DNA III) (3 to 3.6 pmol per ml) were annealed to the 1-strand of ϕ8psω11 DNA (3 pmol per ml) by heating the mixture at 100° C for 2 min and, after cooling to 5°C, MgCl2 was added to 10 mM, dithiothreitol to 10 mM, Tris-Cl (pH 8.0), 0.5 mM pdT, and 2 μl of 1 mg per ml of bacterial alkaline phosphatase. After incubation of the mixture at 65°C for 30 min, DNA II was isolated by Sephadex G-50 (0.9 × 24 cm) filtration using 50 mM TEAB as the eluate.

DNA II was further characterized by degradation to 3'-mononucleotides. Radioactivity was found, as expected, in dGp and dTp, the ratio of the radioactivity being 0.86: 1.0.

Some reactions were carried out in the presence of CTP instead of dCTP and contained 300 μM CTP and 1 mM MnCl2 in place of dMgCl2. The total volume of the reaction mixtures varied between 0.2 to 2.0 ml. Incubation for nucleotide incorporation was at 5°C for 6 hours. The kinetics of incorporation were followed as described previously (9).

Isolation of Elongated Primers

After incubation, the mixtures were heated at 105°C for 2 min and, to separate the elongated primers and ϕ8psω11 DNA 1-strand, immediately loaded onto an agarose 1.5 cm column (0.9 × 24 cm) pre-equilibrated with 50 μl TEAB. The elution peak corresponding to the elongated primer was evaporated to dryness, the

The abbreviations used are: TEAB, triethylammonium bicarbonate; Pipes, piperazine-N,N'-bis(2-ethanesulfonic acid).
shown in Figs. 7 and 8. The longest sequence (H₅), a trideca-
dodec of partial digests with venom and spleen phosphodiesterases are
methods after partial enzymatic degradations. The fingerprints
shown in Fig. 6, b to d were obtained when 1 pM each of [a-V]-
gave longer oligonucleotides. Although the oligonucleotides
shorter fragments whereas, as expected, the longer product, AS,
gave more of the
one-dimensional homochromatography. As seen in Fig. 6, b to d, several products were obtained
the case of increasing size, A₁, A₂, and A₃, were thus obtained. After elution, each
with KOH as in text and then separated by
radioactivity was found in dAp, dGp, and dCp in the ratio of 1:3:2.
[a-3²P]dGTP was used, radioactivity was in dAp, dGp, and
dCp in the ratio of 1:1:1. When [a-3²P]dGTP was used, radio-
activity was found in dAp, dGp, and dCp in the ratio of 1:3:2.
Finally, when [a-3²P]dCTP was used, radioactivity was in dGp
only. The results are all consistent with the sequence derived
above. Similar analyses of the shorter fragments were also
completely consistent with the sequences listed in Fig. 9.
The products formed on Burton degradation of the oligonucleo-
tides were separated by electrophoresis on DE81 paper and the
results are summarized in Table III. The sequences of all of the
oligonucleotides investigated are shown in Fig. 9. The sequences
of the above products were confirmed by extensive nearest
neighbor analyses and by Burton degradation. The results are
shown in Tables II and III. As is seen in Table II for H₅, when
[a-3²P]dATP was used, radioactivity was found in dAp, dGp,
and dCp in the ratio of 1:1:1. When [a-3²P]dGTP was used, radio-
activity was found in dAp, dGp, and dCp in the ratio of 1:3:2. Finally,
when [a-3²P]dCTP was used, radioactivity was in dGp
only. The results are all consistent with the sequence derived
above. Similar analyses of the shorter fragments were also
completely consistent with the sequences listed in Fig. 9.

Preparation of Primer, DNA II, for Further Sequencing—
The sequence deduced above, contains dC units at the initiation
site and at positions 5 and 7 in the promoter region (Figs. 1 and 9).
Because CTP can substitute for dCTP in the DNA polymerase
reaction (18, 19), the polymerase reaction described in the
preceding section was repeated except that CTP replaced dCTP.
The expected product (DNA III) would be an attractive primer
for obtaining a manageable new sequence. In view of the
incomplete elongation encountered above using 1 µM concentra-
FIG. 7. Fingerprints obtained on partial venom phosphodiesterase digests of the products designated H₂ to H₇. The products isolated as in Fig. 6 and under "Materials and Methods" were treated with venom phosphodiesterase and the digests were subjected to the two-dimensional fingerprinting procedure as that described under "Materials and Methods." The homomix used in A, B, C, and D was III, whereas that in E and F was II.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.

The sequences are derived from the patterns as shown in the reproductions shown on the left of each fingerprint. The markers, Penta, Deca, and Dodeca, indicate the positions taken by the pentanucleotide, 3²P-d-G-G-A-A-rG, the decanucleotide, 3²P-d-G-G-A-G-C-A-G-G-C-C, and the dodecanucleotide, 3²P-d-G-G-T-C-C-C-T-T-A-T-C-G.
After the usual separation, the elongated product was subjected to electrophoresis on a polyacrylamide gel. As seen in Channel c in Fig. 10, one main product, designated A, was obtained with the estimated size of 45 nucleotides. The products formed after alkaline hydrolysis are shown in Fig. 11b. Two products, called H4 and H10, were obtained which were sequenced by partial enzymatic degradations followed by fingerprinting (Fig. 12). The sequences derived are also shown in Fig. 9.

Degradation to 3'- and 5'-nucleotides gave results which confirmed the above sequences (Table IV). Thus, when [α-32P]-dTTTP was used and degraded to 3'-nucleotides, the radioactivity was found only in dAp in both H4 and H10. When [α-32P]dATP was used, the radioactivity in H4 was found in dAp, dTp, and dCp in the ratio of 2:1:2; however, the ratio of radioactivity in H10 was 2:1:3. The latter results are consistent with the fact that one of the components in DNA III lacked the terminal rC-A sequence. Elongation of this primer would lead first to the repair of this sequence and, therefore, when [α-32P]dATP is used, an extra mole of radioactive dCp would be found relative to H4 after degradation.

When [α-32P]dCTP was used, degradation of H4 to 3'-nucleotides gave radioactive dGp and dTp in the ratio of 1:2, whereas degradation of H10 gave radioactivity only in dTp. These results are again consistent with the structures shown in Fig. 9 and the fact that H10 must have arisen from the component in DNA III which lacked rC-A.

Depurination of H4 and H10 by the Burton method gave the results described in Table V. As can be seen, both products contain 2 pdTpCp residues followed by 1 or more dA residues (formation of radioactive P1) and also a pdTp residue again followed by 1 or more dA residues. When [α-32P]dTTTP was used, both H4 and H10 gave radioactivity in pdT. This result showed that the 3'-terminal nucleotide in both cases was dT. Further, as seen in Table V, one difference was observed between H4 and H10. When [α-32P]dATP was used, radioactive pdGp was found only in the case of H10. All of the above results are in accord with the sequences in Figs. 9 for H4 and H10.

Sequence of Next 10 Nucleotides is d-G-A-C-G-C-G-C-G-C-G-C.

![Image of a diagram showing sequences and products](http://www.jbc.org/)

---

**Fig. 9.** Sequences of the oligonucleotides belonging to the promoter region and isolated and characterized after DNA polymerase-catalyzed elongation of the primers (see also Fig. 1).
TABLE II

Nearest neighbor analysis for oligonucleotides (H₁ to H₇) obtained by elongation of DNA II with radioactively labeled deoxynucleoside triphosphates followed by alkaline treatment

The details of the preparation are in the legend to Fig. 6. The numbers in parentheses are the experimental molar ratios.

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>dNTP</th>
<th>Radioactivity Distribution (cpm) dAP dGP dTP dCP</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₁</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>693 (2.0) 578 (1.7) 19 46</td>
</tr>
<tr>
<td>H₂</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>959 (2.0) 804 (1.7) 26 500 (1.0)</td>
</tr>
<tr>
<td>H₃</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>1465 (2.0) 2000 (2.7) 34 815 (1.1)</td>
</tr>
<tr>
<td>H₄</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>1144 (2.0) 2187 (3.8) 57 654 (1.2)</td>
</tr>
<tr>
<td>H₅</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>1017 (2.0) 3149 (6.2) 20 614 (1.2)</td>
</tr>
<tr>
<td>H₆</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>966 (2.0) 2549 (5.3) 87 999 (2.1)</td>
</tr>
<tr>
<td>H₇</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>1118 (2.0) 3535 (6.1) 90 1748 (3.1)</td>
</tr>
</tbody>
</table>

TABLE III

Results of Burton degradation of oligonucleotides (H₁ to H₇) obtained by elongation of DNA II followed by alkaline cleavage

The products were separated into Pₗ and pdC + pdC by electrophoresis on DE81 paper. pdCp and pdC were separated by cellulose thin layer chromatography using Solvent III described previously (12). The numbers in parentheses are the observed molar ratios.

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>dNTP</th>
<th>Radioactivity Distribution (cpm) P₁ pdCp pdC</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₁</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>1067 (3.0) - 290 (0.8)</td>
</tr>
<tr>
<td>H₂</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>1118 (3.0) 675 (0.9) -</td>
</tr>
<tr>
<td>H₃</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>3171 (4.0) 1415 (0.9) -</td>
</tr>
<tr>
<td>H₄</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>2167 (5.0) 813 (0.9) -</td>
</tr>
<tr>
<td>H₅</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>4011 (6.0) 1200 (0.8) 392 (0.6)</td>
</tr>
<tr>
<td>H₆</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>3025 (6.0) 1700 (1.7) -</td>
</tr>
<tr>
<td>H₇</td>
<td>[α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP</td>
<td>3655 (6.0) 2854 (2.3) -</td>
</tr>
</tbody>
</table>

(Fig. 4)—During the above primer elongation experiments using a mixture of dATP, dGTP, and dCTP, the formation of products longer than those expected was sometimes noticed. Appearance of the new products varied with the preparations of the triphosphates used and seemed to depend on the presence of a small amount of dTTP as an impurity. Therefore, further elongation reactions were performed by the deliberate addition of 0.015 μM dTTP to the mixture of the remaining three triphosphates. In this way, it was hoped to obtain an elongated primer of the desired size (55 to 60 nucleotides) as a major product. When DNA III was used as the primer and the extended product examined by electrophoresis, the pattern shown in Fig. 10d was obtained. The main band, A₄, was subjected to alkaline hydrolysis and the polynucleotide with the new sequence was further purified (Fig. 11c). The main product now obtained (H₁₁) was sequenced by fingerprinting of its partial enzymic digest (Fig. 13). Two sets of spots were seen, the major set corresponding to the sequence shown in Fig. 9 for H₁₁. The faint spots, which also corresponded to a set, belonged to the contaminant, H₁₂ (Fig. 9), present in H₁₁. The contaminant, which was purified in a small amount by careful elution from the thin layer plate (Fig. 11), was examined for nearest neighbor analysis and for pyrimidine tracts. All of the evidence and its lower mobility pointed to its having the additional d-G-C sequence at the 5' end. These results were expected because the DNA III contained a component which lacked the rC-A sequence at the 3' end and terminated in the preceding rC-G sequence. Elongation of this sequence would, therefore, start with the d-C-A sequence and, after alkaline cleavage, the sequence at the 5' end in the new product would be d-G-C-A. Results of the nearest neighbor analyses performed on different preparations of the products, H₁₁ and H₁₂, are compared in
tains two components, Band 1 and Band 2. Channel b contains the DNA III elongated with a labeled nucleotide. Finally, when [\(\alpha-^{32}\)P]dCTP was used in the elongation reaction, both H\(_1\) and H\(_2\) gave radioactivity in dAp, dGp, and dCp in the ratio of 2:1:1. A similar preparation of the elongated primer, again both H\(_1\) and H\(_2\) gave dAp as the sole labeled nucleotide. Finally, when [\(\alpha-^{32}\)P]dGTP was used in the preparation of H\(_1\), radioactivity was found in dAp, dGp, dCp, and dCp in the ratio of 1:2:3:1. All of these results support the conclusions drawn from the fingerprints.

The results of Burton degradation on the products H\(_1\) and H\(_2\) are shown in Table VII. When [\(\alpha-^{32}\)P]dATP was used, H\(_1\) gave pdTp and pdTpGp in the ratio of 2:2. H\(_1\) prepared by using [\(\alpha-^{32}\)P]dCTP gave pdCp, pdCpCp, pdCpCpCp, and pdCpCpCpCp in the ratio of 2:1:1. These results indicate that H\(_1\) contains in its sequence 2 pdCp, 1 pdCpCp, 2 pdCp, and 2 pdCpCpCp as pyrimidine tracts and pdC as the 3' end nucleotide. When [\(\alpha-^{32}\)P]dATP was used, pdCp and pdCpCp were produced in the ratio of 1:2 from H\(_1\). This suggests that both of the pdCpCp sequences and 1 of the 2 pdCp residues are followed by a dA residue. H\(_1\) prepared from [\(\alpha-^{32}\)P]dATP gave pdCpCp and pdCpCpCp in the ratio of 2:1:1, and the result indicated that pdCpCpCp, both the pdCp and 1 of the 2 pdCp residues are followed by a dG residue. These results completely agree with the above described sequence for H\(_1\).

Results of the parallel experiments carried out with H\(_1\) (Table VII) were also consistent with the sequence deduced for H\(_2\). There was, however, one deviation from the expected results which was observed when [\(\alpha-^{32}\)P]dCTP was used as the radioactive triphosphate. The ambiguity evidently arose from the lack of separation of H\(_1\) and H\(_2\) in this particular experiment and, because H\(_1\) was the major product, analyses carried out on the product corresponded more closely to H\(_1\) than to H\(_2\). Thus, only 2 mol of pdCp were recovered in place of the expected 3 mol (data not shown).

**DISCUSSION**

The approach used in the present work has involved the DNA polymerase I-catalyzed elongation of a primer hybridized at the appropriate site to the single-stranded DNA template containing the polynucleotide sequence of interest. The same approach has been used previously for the determination of the nucleotide sequence in the terminator region of the same gene (1, 9), and it is also being used extensively in other laboratories (17, 20). The potential and flexibility of the method have now been enhanced by the introduction of three types of procedures: the insertion of ribonucleotide units at strategic places so as to allow specific cleavages and isolation of shorter chains containing the new nucleotide sequences; carrying out controlled and stepwise elongation by using three (or less) nucleoside triphosphates, isolating the extended primer, and repeating the elongation reac-
Fig. 12. Fingerprints of the partial venom and spleen phosphodiesterase degradations of the products H8 and H10. The oligonucleotides, H8 and H10, purified as in Fig. 11 b and under ‘Materials and Methods’ were treated with snake venom phosphodiesterase (A and C) or spleen phosphodiesterase (B and D). The digests were subjected to the two-dimensional fingerprinting procedure as that described under ‘Materials and Methods.’ The homomix used was II. The markers, Penta, Deca, and Dodeca, are as described in Fig. 6. The sequence derivation from the pattern of products is shown at the left in each case.

| TABLE IV |

Nucleotide (5' and 3') analysis of oligonucleotides (H8 to H10) obtained by alkaline treatment of DNA III after elongation with deoxynucleoside triphosphates

The details of the preparation are in the legend to Fig. 11. The numbers in parentheses are the experimental molar ratios.

(a) 5' Nucleotide analysis

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>dNTP</th>
<th>Radioactivity Distribution (cpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pdA pdG pdT pdC</td>
<td></td>
</tr>
<tr>
<td>H8</td>
<td>([a-32P]dGTP+[a-32P]dTTP+[a-32P]dCTP)</td>
<td>28 71 708(1.0) 1110(1.6)</td>
</tr>
<tr>
<td>H9</td>
<td>([a-32P]dATP+[a-32P]dTTP+[a-32P]dCTP)</td>
<td>966(6.0) 28 646(4.0) 370(2.3)</td>
</tr>
<tr>
<td>H10</td>
<td>([a-32P]dATP+[a-32P]dTTP+[a-32P]dCTP)</td>
<td>1214(6.0) 21 705(3.5) 671(3.3)</td>
</tr>
</tbody>
</table>

(b) 3' Nucleotide analysis

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>dNTP</th>
<th>Radioactivity Distribution (cpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>dAp dGp dTp dCp</td>
<td></td>
</tr>
<tr>
<td>H8</td>
<td>([a-32P]dGTP+[a-32P]dTTP+[a-32P]dCTP)</td>
<td>883(1.0) 21 1062(1.2) 28</td>
</tr>
<tr>
<td>H9</td>
<td>([a-32P]dGTP+ [a-32P]dTTP+ [a-32P]dCTP)</td>
<td>973(6.0) 16 567(3.5) 372(2.3)</td>
</tr>
<tr>
<td></td>
<td>([a-32P]dATP+ [a-32P]dTTP+ [a-32P]dCTP)</td>
<td>3439(2.0) 200 1894(1.1) 2620(2.1)</td>
</tr>
<tr>
<td></td>
<td>([a-32P]dATP+ [a-32P]dTTP+ [a-32P]dCTP)</td>
<td>70 50 556(2.0) 55</td>
</tr>
<tr>
<td>H10</td>
<td>([a-32P]dGTP+ [a-32P]dTTP+ [a-32P]dCTP)</td>
<td>1090(6.0) 230(1.3) 677(3.7) 593(3.3)</td>
</tr>
<tr>
<td></td>
<td>([a-32P]dATP+ [a-32P]dTTP+ [a-32P]dCTP)</td>
<td>1390(2.0) 74 734(1.1) 2072(3.0)</td>
</tr>
<tr>
<td></td>
<td>([a-32P]dATP+ [a-32P]dTTP+ [a-32P]dCTP)</td>
<td>2576(4.0) 36 22 20</td>
</tr>
<tr>
<td></td>
<td>([a-32P]dATP+ [a-32P]dTTP+ [a-32P]dCTP)</td>
<td>83 593(0.7) 1643(2.0) 55</td>
</tr>
</tbody>
</table>
TABLE V

Analysis of products formed on Burton degradation of oligonucleotides (H₉ and H₁₀) obtained by elongation of DNA III with dATP, dTTP, and dCTP followed by alkaline treatment

The numbers in parentheses are the experimental molar ratios.

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>dNTP</th>
<th>Radioactivity Distribution (cpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P₁</td>
</tr>
<tr>
<td>H₉</td>
<td></td>
<td>4070</td>
</tr>
<tr>
<td></td>
<td></td>
<td>42</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>H₁₀</td>
<td></td>
<td>1956</td>
</tr>
<tr>
<td></td>
<td></td>
<td>61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>655</td>
</tr>
<tr>
<td></td>
<td></td>
<td>177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1431</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1614</td>
</tr>
</tbody>
</table>

TABLE VI

Nearest neighbor analysis of oligonucleotides (H₁₁ and H₁₂) obtained by elongation of DNA III with dATP, dGTP, dCTP, and a limited amount of dTTP followed by alkaline treatment

The details of the preparation are in the legend to Fig. 11. The numbers in parentheses are the experimental molar ratios.

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>dNTP</th>
<th>Radioactivity Distribution (cpm)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>dAp</td>
</tr>
<tr>
<td>H₁₁</td>
<td></td>
<td>655</td>
</tr>
<tr>
<td></td>
<td></td>
<td>177</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1431</td>
</tr>
<tr>
<td>H₁₂</td>
<td></td>
<td>1614</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1031</td>
</tr>
<tr>
<td></td>
<td></td>
<td>313</td>
</tr>
</tbody>
</table>

FIG. 13. Fingerprints of partial venom phosphodiesterase degradation of the product H₁₁, H₁₂, as contaminant, as obtained in Fig. 11c. The oligonucleotide H₁₁ obtained as in Fig. 11c and under “Materials and Methods” was treated with snake venom phosphodiesterase and the digests were subjected to the two-dimensional fingerprinting procedure. For homochromatography, Homomix II was used. The sequence is derived from the pattern as shown in the reproduction on the left of the fingerprint.

Several methods were used for the determination of the sequences of the purified oligonucleotides. These included two-dimensional fingerprinting of the digests obtained on partial degradations with exonuclease enzymes, the extensive nearest neighbor analysis by the use of different α-³²P-labeled deoxynucleoside triphosphates, and the determination of pyrimidine tracts by Burton's method. Finally, in data not given, confirmation of the sequences thus derived was also obtained by the characterization of the oligonucleotide fragments obtained by pancreatic DNase degradation of the extended primers.

The nucleotide sequence now discovered (Fig. 2) possesses remarkable features. It contains two outer regions, which bear 2-fold symmetry relationship with one another and contain five exclusively G:C base-pairs. The central part of the sequence is A:T rich and also contains regions which have 2-fold symmetry. The total sequence possesses a rotational axis around the 15th base-pair which is a G :C base-pair and which is flanked by A:T base-pairs on both sides. The regular double helix can open up to adopt the secondary structure shown in Fig. 14. Thus, the sequence has not only regions with asymmetrical base composition, but can exist in two alternative structures (Figs. 2 and 14). The structure in Fig. 14 also has the same regions of 2-fold symmetry, but their orientation is reversed relative to that in Fig. 2. These features must have biological significance because their occurrence on a random basis seems most improbable.

Regions with 2-fold symmetry are being found with increasing frequency in DNA. Thus, they have been found around the sites of action of several restriction enzymes and related methylases (21–24), of the enzyme cleaving the covalently closed DNA of bacteriophage λ to generate cohesive ends, of the σ function in bacteriophage λ (25), and in the lac operator (26). The sequences previously described for the terminator region in the tyrosine tRNA gene (1) and the sequences found for the promoter regions of the bacteriophage fd DNA (1) and the leftward promoter

H. Schaller and colleagues, personal communication.
The numbers in parentheses are the experimental molar ratios.

<table>
<thead>
<tr>
<th>Oligonucleotides</th>
<th>dATP</th>
<th>dGTP</th>
<th>dCTP</th>
<th>dTTP</th>
</tr>
</thead>
<tbody>
<tr>
<td>H11</td>
<td>4011(3.4) 5712(2.4) 9841(0.8) 3509(1.0) 2254(1.9) 4725(2.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H12</td>
<td>1595(3.6) 426(1.0) 39 34 508(1.2) 878(2.0)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table VII**

Analysis of products formed on Burton degradation of oligonucleotides (H11 and H12) obtained by elongation of DNA III with dATP, dGTP, dCTP, and a limited amount of dTTP followed by alkaline treatment.

Fig. 14. A secondary structure model for the promoter region of the tyrosine tRNA gene.

**FIG. 14.** A secondary structure model for the promoter region of the bacteriophage X DNA (27, 28) all contain similar features. Similarly, Wilson and Thomas (29) and Locker et al. (30) have discovered the presence of "palindromes" in eukaryotic DNA. Because all of the transcription in E. coli is evidently performed by the same enzyme (31), the latter would appear to recognize all of the promoters on the E. coli chromosome as well as at least some of the promoters on the genomes of the bacteriophages which infect E. coli. It is therefore of immediate interest to compare the sequences of some of the different promoters recognized by the E. coli transcriptase. Progress is just beginning to be reported in this area. The sequence of 36 nucleotides in the leftward promoter of bacteriophage X has been determined in this laboratory (27) and by Ptashne and colleagues (28). The sequence of one of the strong RNA polymerase binding sites present in the Rf of the bacteriophage fd has been determined by Schaller and colleagues. Information has also been forthcoming on the sequence of the promoter in SV40 DNA which is recognized by the E. coli polymerase (32, 33). Finally, the sequence of the promoter region in the lactose operon is also known. The striking fact which emerges is that the above promoters all differ widely in primary sequence. Therefore, the important concept must forthwith be invoked that the polymerase recognizes a structure rather than a linear sequence in the double helix. That DNA may be recognized by proteins by virtue of specific looped-out structures has already been proposed by Gicrer (34, see also Ref. 35). However, it remains for future work to determine the nature of the presumed three-dimensional structure recognized by the RNA polymerase because a comparison at this time of the known promoter sequences does not readily reveal a common pattern such as was the case for the tRNA sequences. Thus, of the five promoters whose sequences are known, elements of symmetry can be detected only in the promoter in the tyrosine tRNA gene, the leftward promoter for the N gene in bacteriophage X and the strong promoter for the RF of the bacteriophage fd, and only the first two promoters could possibly be regarded as being similar in regard to the symmetry elements. The sequences for the lac promoter and for the SV40 DNA promoter for the E. coli polymerase evidently lack any recognizable symmetry patterns. Despite the fact that there are large unknowns at present regarding the RNA polymerase-promoter interaction, it is tempting to see some significance in the similarities between the structures shown in Figs. 2 and 14 for the tRNA gene promoter. It is possible that RNA polymerase at first recognizes all or a part of the linear double helix containing regions of a 2-fold symmetry (structures of Fig. 2) and binding of the enzyme to this site takes place. Then, there ensues a conformational change in the enzyme concomitant with a transition in the DNA structure to that shown in Fig. 14. One of two things may then follow. Either the enzyme binds to one or the other looped-out arms such that the sequence recognition is still maintained. By this new mode of binding of the enzyme, selection of the strand as well as the initiation site could both be accomplished. Alternatively, the enzyme after the conformation change may be able to recognize the symmetrical regions in both arms of the structure in Fig. 14. In this case also, strand selection and initiation site would be accomplished simply by the asymmetrical configuration of the multisubunit enzyme. Studies are now under way to gain insights into these aspects of the mechanism of transcription by using the approach described below.

Although it seems highly likely that the interesting structures shown in Figs. 2 and 14 are important in the initiation of tran-
scription, it is not clear at this time how much of the promoter region is actually represented in the sequence now known. Further sequence work would certainly be desirable. However, definitive answers can come only by actual in vitro studies of the transcriptional process, including specificity of initiation. These require, in turn, DNA segments which contain varying lengths of sequences into the preinitiation region and also an adequate length of the DNA into the post initiation region. DNAs of this kind can be obtained by the synthetic methodology which has been developed in this laboratory. However, for initial studies, the desired DNAs can also be prepared from the primer-template complexes used in the present work. After controlled elongation, the primer-template complexes may be digested with an endonuclease specific for single-stranded DNA. In this way, double-stranded DNAs corresponding in length exactly to the elongated primers may be isolated. Experiments along these lines are in progress.

Studies are also in progress on the mechanism of the termination of transcription in the tyrosine tRNA gene, by using DNA segments corresponding to the elongated primers described previously (1). It is hoped that with the elucidation of the exact chain lengths of the promoter and terminator regions, it should prove possible to reconstruct, by total synthesis, a gene containing its control elements. Finally, the availability of such a totally synthetic gene should enable systematic studies of the structure-function relationships in the above tRNA by predetermined alterations in the structural gene.

Acknowledgments—Initial experiments on primer-dependent nucleotide incorporations in the promoter region were carried out by Drs. Peter Loewen and Marvin H. Caruthers. Their assistance in this work is gratefully acknowledged.

REFERENCES

The nucleotide sequence in the promoter region of the fene for an Escherichia coli tyrosine transfer ribonucleic acid.
T Sekiua, H V Ormondt and H G Khorana


Access the most updated version of this article at http://www.jbc.org/content/250/3/1087

Alerts:
• When this article is cited
  • When a correction for this article is posted

Click here to choose from all of JBC’s e-mail alerts

This article cites 0 references, 0 of which can be accessed free at http://www.jbc.org/content/250/3/1087.full.html#ref-list-1