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 ORMONDT,† 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:

(3') ---GGGGGCGCATCATATCAAA TGACGCGC CGC- (5')
(5') ---CCC CGCGTAGTATAGTTTACTGCGCGGCG--- (3')

The general approach used for the sequence determination involved the DNA polymerase I-catalyzed elongation of suitable deoxyribopolynucleotide primers when hybridized to the 3'-strand of φ80psu11 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.

Methodology now exists for the synthesis of bireehal 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 closely defined the initiation and termination sites for the transcription of this gene by the isolation and characterization of a precursor to the tRNA (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, 9–9) 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 φ80psu11 DNA carrying the gene for the tyrosine suppressor tRNA, the hybridization of appropriate deoxyribopolynucleotide primers at the tyrosine tRNA gene terminal r- or l-strand of the above DNA, and the controlled DNA polymerase-catalyzed

* This work has been supported by National Institutes of Health, United States Public Health Service Grant CA05178, and the National Science Foundation, Washington, D.C. Grant GB-21053X. The authors also wish to acknowledge the receipt of financial support from the European Molecular Biology Organization. This is Paper CXXVII in the series on studies on polynucleotides. The preceding paper in this series is Ref. 1.

† Permanent address, Laboratory for Physiological Chemistry, the University of Leiden, Leiden, Holland.
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 l-strand of the 880 base-pair 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 deoxynucleotide 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.

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 T4 polynucleotide ligase-catalyzed joining of the chemically synthesized oligonucleotides 5'-d-G-C-T-C-C-C-T-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'-d-G-C-T-C-C-C-T-T-A-T-C-G was elongated 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 two 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 dodecanucleotide 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.

3 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 ml of water, and to the solution were added 15 ml of 60% glycerol containing xylene cyanol and bromphenol blue as dye markers.

The solution was placed on a 15% polyacrylamide gel slab (30 × 20 × 0.2 cm) containing 7 mM urea and electrophoresis was performed at 600 mA Tris-base, 90 mA boric acid, and 4 mA EDTA for 16 hours at 10 mA. The radioactive bands were eluted from the gel with 2 M TEAB. The eluate was evaporated to dryness, and the residue was dissolved in 1 ml of H2O and dealted 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°C. After neutralization with acetic acid, the products were separated by homomixograph 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 M TEAB. The eluate was evaporated in vacuo to dryness. The radioactive oligonucleotides thus obtained were subjected to enzymatic degradation to 3'- or 5'-mononucleotides and to the depurination reaction. For partial digestion with phosphodiesterases, the oligonucleotides were treated with 0.5 M KOH for 18 hours at 37°C to degrade the RNA fragments which came from the homomixture. Hydrolyzates were neutralized with acetic acid and passed through Sephadex G-50 to remove the ribonucleotides and salt.

**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 DE51 paper directly and the products were separated by electrophoresis in 75% formic acid. The papers were autoradiographed and the radioactive spots were cut out and eluted with 2 M TEAB for further analysis.

Partial digestion with snake venom phosphodiesterase was carried out in 10 mM glycine buffer (pH 9.2) at 20°C 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°C for 30 min in 10 mM Pipes buffer (pH 6.9) with 10 mM MgCl2 and 1 to 30 units per ml of the enzyme. After the reaction, cold water was added to the digestion mixture (about 100°C) and the mixture was applied to 20% polyacrylamide gels containing 0.5 M KOH for 18 hours at 37°C to degrade the RNA fragments. The materials (56 pmol) from the first peak in Fig. 4 were treated with 0.3 M KOH at 37°C for 18 hours. The above product was isolated enzymatically by dT-A-C-T-G-G-C-C-T and two complementary polynucleotides and isolation of the joined product was performed by gel filtration through a Sephadex G-75 column (1 × 100 cm).

**Isolation of Elongated Primers**

After incubation, the mixtures were heated at 100°C for 2 min and, to separate the elongated primers and 32P-labeled DNA 1-strand, immediately loaded onto an agarose 1.5 cm column (0.9 × 24 cm) pre-equilibrated with 50 mM TEAB. The elution peak corresponding to the elongated primer was evaporated to dryness, the

---

3 The abbreviations used are: TEAB, triethylammonium bicarbonate; Pipes, piperazine-N,N'-bis(2-ethanesulfonic acid).

**RESULTS**

**Hybridization of DNA I to l-strand of 32P-labeled 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 1-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 is d-G-G-G-G-C-G-C-A (Fig. 1)—Elongation of DNA I1 was carried out using DATP, dGTP, and dCTP. The kinetics of incorporation were shown in the inset in Fig. 5. The elongated product was isolated as in the figure (Peak I). It was further purified by polyacrylamide gel electrophoresis (Fig. 6a) when...
shown in Figs. 7 and 8. The longest sequence (H5), a trideca-

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 [α-32P]dATP, [α-32P]dCTP, and [α-32P]dGTP were used as the labeled triphosphates. For isolation, the reaction mixture was dialyzed, then heated at 100°C for 2 min, and loaded on an Agarose 1.5m column (0.9 X 24 cm). Peak I contained the l-strand which had incorporated some radioactivity and Peak II was the required elongated DNA II, whereas Peak III contained the excess of nucleoside triphosphates.

| Experiment | l-strand | [32P]/DNA l | Mol. ratio DNA l/l-strand | [32P]/DNA l hybridized |
|------------|----------|-------------|--------------------------|-----------------------|
| 1          | 0.361    | 0.665       | 1.9                      | 0.77                  |
| 2          | 0.443    | 1.75        | 3.8                      | 0.88                  |

Hybridization was performed as described under “Materials and Methods.” Analysis of the primer hybridized to the l-strand was carried out by filtration through an Agarose 1.5m column as described previously (8).

The bands, Hz to H5, were sequenced by the fingerprinting protocol of the nucleotide incorporations is given in the text. As seen in the inset, the three nucleoside triphosphates, dATP, dGTP, and dCTP were present in each reaction mixture; in one experiment, all three were labeled with 32P in the α position (O—O), in the second, [α-32P]dATP was the labeled nucleotide (O—O), in the third, [α-32P]dGTP, (Δ—Δ), and in the fourth [α-32P]dCTP (Δ—Δ) were used as the labeled triphosphates. For isolation, the reaction mixture was dialyzed, then heated at 100°C for 2 min, and loaded on an Agarose 1.5m column (0.9 X 24 cm). Peak I contained the l-strand which had incorporated some radioactivity and Peak II was the required elongated DNA II, whereas Peak III contained the excess of nucleoside triphosphates.

three bands, A1 to A3, were observed. After alkaline treatments to cleave at the rG site, the fragments containing the new sequences were separated by one-dimensional homochromatography. As seen in Fig. 6, b to d, several products were obtained from each of the bands shown in Fig. 6a. A1 gave more of the shorter fragments whereas, as expected, the longer product, A3, gave longer oligonucleotides. Although the oligonucleotides shown in Fig. 6, b to d were obtained when 1 μM each of [α-32P]-dATP, [α-32P]-dGTP, and [α-32P]-dCTP were used in the extension reaction, the results obtained with one [α-32P]-dNTP (1 μM) + two unlabeled dNTPs (5 μM) were similar; however, the main bands in these cases corresponded to H1 to H5.

The bands, H5 to H7, were sequenced by the fingerprinting methods after partial enzymatic degradations. The fingerprints of partial digests with venom and spleen phosphodiesterases are shown in Figs. 7 and 8. The longest sequence (H5), a trideca-

nucleotide, corresponded to the structure shown in the dashed box in Fig. 1 (Experiment A). 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 [α-32P]-dATP was used, radioactivity was found in dAp, dGp, and dCp in the ratio of 1:1:1. When [α-32P]-dGTP was used, radioactivity was found in dAp, dGp, and dCp in the ratio of 1:3:2. Finally, when [α-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 oligonucleotides were separated by electrophoresis on DES1 paper and the results are summarized in Table III. The bands H5, H6, H4, and H3 gave P1 and pdCP in the expected molar ratio. The compounds, H4, and H3 gave pdC in addition to P1 or pdCP or both as would be expected from their 3′-terminal sequences.

Preparation of Primer, DNA III, for Further Sequencing—The sequence derived 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 concentr-
FIG. 7. Fingerprints obtained on partial venom phosphodiesterase digests of the products designated $H_2$ to $H_7$. 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, $32P$-[G-G-A-A-G], the decanucleotide, $32P$-[G-G-A-G-C-A-G-G-C-C], and the dodecanucleotide, $32P$-[G-C-T-C-C-C-T-T-A-T-C-G].

The elongation reactions. In the first reaction, $[\alpha-32P]$dTTT, $[\alpha-32P]$dCTT, and $[\alpha-32P]$dGTT were used as the triphosphates. Only a few nucleotides were added as shown by the mobility of the extended primer on a polyacrylamide gel (Fig. 10, Channel 6, Ad). After alkaline cleavage of $\Lambda_4$, separation by homochromatography (Fig. 11a) gave an oligonucleotide ($H_2$) which was identified as d-A-T-C. Thus, degradation to 3’- and 5’-nucleotides (Table IV) gave the results expected for this sequence. Because the dA unit in this sequence is the 3’-terminal nucleotide (eighth nucleotide in the promoter) in DNA III, the sequence of the next two nucleotides is d-T-C. This conclusion was also confirmed by the following experiment.

DNA III was next elongated using dATP, dTTP, and dCTP. With this combination of the triphosphates, only the component carrying the complete octanucleotide sequence in DNA III (Band 1 in Fig. 10a) would undergo elongation. The second component (Band 2 in Fig. 10a) lacking rC-A at the 3’ terminus would be left out because of the absence of dATP.

DNA III was isolated from excess nucleoside triphosphates and the template DNA by gel filtration through an Agarose 1.5m column. The product was analyzed by electrophoresis on a 15% polyacrylamide gel (Fig. 10, Channel a). The DNA III preparation contained two compounds, designated Band 1 and 2. From their mobilities on the gel and the results described below, it was concluded that Band 1 corresponded to DNA III containing the full eight nucleotides in the promoter region, whereas Band 2 lacked the dinucleotide sequence rC-A at the 3’ end.

Sequence of Next 11 Nucleotides is d-T-C-A-T-A-T-C-A-A-T (Fig. 1)—DNA III prepared as above was used in two elongation reactions. In the first reaction, $[\alpha-32P]$dTTT, $[\alpha-32P]$dCTT, and $[\alpha-32P]$dGTT were used as the triphosphates. Only a few nucleotides were added as shown by the mobility of the extended primer on a polyacrylamide gel (Fig. 10, Channel 6, Ad). After alkaline cleavage of $\Lambda_4$, separation by homochromatography (Fig. 11a) gave an oligonucleotide ($H_2$) which was identified as d-A-T-C. Thus, degradation to 3’- and 5’-nucleotides (Table IV) gave the results expected for this sequence. Because the dA unit in this sequence is the 3’-terminal nucleotide (eighth nucleotide in the promoter) in DNA III, the sequence of the next two nucleotides is d-T-C. This conclusion was also confirmed by the following experiment.

DNA III was next elongated using dATP, dTTP, and dCTP. With this combination of the triphosphates, only the component carrying the complete octanucleotide sequence in DNA III (Band 1 in Fig. 10a) would undergo elongation. The second component (Band 2 in Fig. 10a) lacking rC-A at the 3’ terminus would be left out because of the absence of dATP.

$^4$ Analysis of the composition of the products $H_1$ to $H_7$ shows that the incorporations of dGTP and dCTP were particularly rate-limiting, whereas dA incorporation evidently went to completion. Therefore, in this experiment, higher concentrations of dGTP and CTP were used.

$^5$ With this combination of the triphosphates, only the component carrying the complete octanucleotide sequence in DNA III (Band 1 in Fig. 10a) would undergo elongation. The second component (Band 2 in Fig. 10a) lacking rC-A at the 3’ terminus would be left out because of the absence of dATP.
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. 10a. 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]-dCTP 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 H10 to 3'-nucleotides gave radioactive dGp and dTp in the ratio of 1:2, whereas degradation of H4 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 H8 and H10 by the Burton method gave the results described in Table V. As can be seen, both products contain 2 pdTpdCp residues followed by 1 or more dA residues (formation of radioactive 3') and also a pdTp residue again followed by 1 or more dA residues. When [α-32P]dCTP was used, both H8 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 H8 and H10. When [α-32P]dATP was used, radioactive pdCp was found only in the case of H10. All of the above results are in accord with the sequences in Fig. 9 for H8 and H10.

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

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.

| Oligonucleotides | dNTP    | Radioactivity Distribution (cpm) | dAp | dGp | dTp | dCp |
|------------------|---------|----------------------------------|-----|-----|-----|-----|
| H₁               | [α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP | 693 (2.0) | 578 (1.7) | 19   | 46  |
| H₂               | [α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP | 959 (2.0) | 804 (1.7) | 26   | 500 (1.0) |
| H₃               | [α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP | 1465 (2.0) | 2000 (2.7) | 34   | 815 (1.1) |
| H₄               | [α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP | 1144 (2.0) | 2187 (3.8) | 57   | 654 (1.2) |
| H₅               | [α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP | 1017 (2.0) | 3149 (6.2) | 20   | 614 (1.2) |
| H₆               | [α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP | 966 (2.0) | 2549 (5.3) | 87   | 999 (2.1) |
| H₇               | [α⁻³²P]dATP+[α⁻³²P]dGTP+[α⁻³²P]dCTP | 1118 (2.0) | 3535 (6.1) | 90   | 1748 (3.1) |

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 P₁ and pdCp + 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.

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

(Fig. 1)—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, A₈, 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
Section of the text:

Fig. 10. Polyacrylamide gel electrophoresis of the different elongation products using DNA III as the primer. Preparation of the primer, DNA III, elongation of DNA III, and isolation of elongation products were carried out as described under "Materials and Methods." Channel a contains the newly formed oligonucleotide (II) obtained from Band A4 in Fig. 10c. Channel b contains the oligonucleotides (H11 and H12) obtained from Band A3 in Fig. 10c. Channel c contains the oligonucleotides (H11 and H12) obtained from Band A1 in Fig. 10d. Homochromatography as shown in Channel a was carried out using Homomix III. Homochromatography shown in Channels b and c was carried out using Homomix I and I, respectively. The markers, Penta, Deca, and Dodeca, are as described in Fig. 6. The marker, DNA I, indicates the position taken by [32P]DNA I.

Table VI. Thus, when H11 was prepared using [α-32P]dATP as the labeled nucleotide, radioactivity was found in dAp, dGp, dTp, and dCp in the ratio of 2 : 1 : 1 : 2. A similar preparation of H12 gave the same results except that dCp was present in the molar ratio of 3. When [α-32P]dCTP was used in the elongation reaction, both H11 and H12 gave radioactivity in dTtp and dCp in the identical ratio of 1 : 3. When [α-32P]dTTP was used as the labeled nucleotide, again both H11 and H12 gave dAp as the sole labeled nucleotide. Finally, when [α-32P]dGTP was used in the preparation of H11, radioactivity was found in dAp, dGp, dTp, and dCp in the ratio of 1 : 3 : 2 : 1. All of these results support the conclusions drawn from the fingerprints.

The results of Burton degradation on the products H11 and H12 are shown in Table VII. When [α-32P]dTTP was used, H11 gave pdTpdCp in the ratio of 2 : 2. H11 prepared by using [α-32P]dCTP gave pdCp, pdTp, pdCpdCp, and pdTpdcCp in the ratio of 2 : 1 : 1 : 2. These results indicate that H11 contains in its sequence 2 pdCp, 1 pdCpdCp, 2 pdTp, and 2 pdTpdCp as pyrimidine tracts and pdC as the 3' end nucleotide. When [α-32P]dATP was used, pdTp and pdTpdCp were produced in the ratio of 1 : 2 from H11. This suggests that both of the pdTpdCp sequences and 1 of the 2 pdTp residues are followed by a dA residue. H11 prepared from [α-32P]dGTP gave pdCp, pdCpdCp, and pdTp in the ratio of 2 : 1 : 1, and the result indicated that pdCpdCp, both the pdCp and 1 of the 2 pdTp are followed by a dG residue. These results completely agree with the above described sequence for H11.

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 reaction.
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

| Oligonucleotides | dNTP                  | Radioactivity Distribution (cpm) |
|------------------|-----------------------|-----------------------------------|
|                  | pdA                   | pdG                               | pdT                             | pdC                             |
| H₈               | [α-¹³P]dTTP+[α-¹³P]dGTP+[α-¹³P]dCTP | 28                                | 71                              | 708(1.0)                        | 1110(1.6)                       |
| H₉               | [α-¹³P]dATP+[α-¹³P]dTTP+[α-¹³P]dCTP | 966(6.0)                          | 28                              | 646(4.0)                        | 370(2.3)                        |
| H₁₀              | [α-¹³P]dATP+[α-¹³P]dTTP+[α-¹³P]dCTP | 1214(6.0)                         | 21                              | 705(3.5)                        | 671(3.3)                        |

(b) 3′ Nucleotide analysis

| Oligonucleotides | dNTP                  | Radioactivity Distribution (cpm) |
|------------------|-----------------------|-----------------------------------|
|                  | dAp                   | dGp                               | dTt                             | dCp                             |
| H₈               | [α-¹³P]dTTP+[α-¹³P]dGTP+[α-¹³P]dCTP | 883(1.0)                          | 21                              | 1062(1.2)                       | 28                             |
| H₉               | [α-¹³P]dATP+[α-¹³P]dTTP+[α-¹³P]dCTP | 973(6.0)                          | 16                              | 567(3.5)                        | 372(2.3)                        |
|                  | [α-¹³P]dATP+[α-¹³P]dTTP+[α-¹³P]dCTP | 3439(2.0)                         | 200                             | 1894(1.1)                       | 2620(2.1)                       |
|                  | dATP+                 | dTTP+                            | dCTP                           |                                  |
| H₁₀              | [α-¹³P]dATP+[α-¹³P]dTTP+[α-¹³P]dCTP | 2543(4.0)                         | 32                              | 32                             | 55                             |
|                  | [α-¹³P]dATP+          | dTTP+                            | dCTP                           |                                  |
|                  | [α-¹³P]dATP+          | dTTP+                            | dCTP                           |                                  |
|                  | [α-¹³P]dATP+          | dTTP+                            | dCTP                           |                                  |
|                  | [α-¹³P]dATP+          | dTTP+                            | dCTP                           |                                  |
|                  | [α-¹³P]dATP+          | dTTP+                            | dCTP                           |                                  |
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.

| Oligonucleotides | dNTP | Radioactivity Distribution (cpm) |
|------------------|------|----------------------------------|
|                  |      | PI | pdTp | pdCp | pdTpdCp | pdT |
| H9               |      |    |      |      |         |     |
| [α-32P]dATP+     | dTTp+| 4070(1.9) | 2140(1.0) | -     | 3662(1.7) | -   |
| dATP+[α-32P]dTTp+| dTtp+| 42  | 776(1.0) | -     | 1436(2.0) | 694(1.0) |
| dATP+            | dTTp+[α-32P]dTCP | 12  | 300  | -     | 6587(2.0) | -   |
| H10              |      |    |      |      |         |     |
| [α-32P]dATP+     | dTTp+| 1956(2.6) | 747(1.0) | 644(0.9) | 1501(2.0) | 729(0.9) |
| dATP+[α-32P]dTTp+| dTtp+| 61  | 888(1.1) | -     | 1646(2.0) | -   |

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.

| Oligonucleotides | dNTP | Radioactivity Distribution (cpm) |
|------------------|------|----------------------------------|
|                  |      | dAp | dGp | dCp | dTp |
| H11              |      |     |     |     |     |
| [α-32P]dATP+     | dGTP+| dTTp+| dTCP | dCp | dT | 655(2.0) | 414(1.3) | 367(1.1) | 664(2.0) |
| dATP+[α-32P]dGTP+| dTTp+| dTCP | dCTP | 177 | 234 | 729(1.3) | 1650(3.0) |
| dATP+            | dGTP+[α-32P]dTTp+| dTCP | 1431(4.0) | 80 | 143 | 100 |
| H12              |      |     |     |     |     |
| [α-32P]dATP+     | dGTP+| dTTp+| dTCP | 1614(1.5) | 2863(2.7) | 2114(2.0) | 1095(1.0) |
| dATP+[α-32P]dGTP+| dTTp+| dTCP | 1032(2.5) | 602(1.5) | 586(1.4) | 1227(3.0) |
| dATP+            | dGTP+[α-32P]dTTp+| dTCP | 313  | 363 | 929(1.3) | 2144(3.0) |
| dATP+            | dGTP+| dTTp+| dTCP | 1270(4.0) | 37  | 25  | 59 |

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.

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 ter 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 DNA1 and the leftward promoter

1 H. Schaller and colleagues, personal communication.
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**

The numbers in parentheses are the experimental molar ratios.

| Oligonucleotides | dNTP | Radioactivity Distribution (cpm) |
|------------------|------|----------------------------------|
|                  | Pi   | pTdpC  | pCpCpC  | pTdpCpC  | pTdpCpC |
| H11              |      |        |        |        |        |
| [~12p]dATP+[~32p]dGTP+ [~12p]dCTP | 2339(3.0) | 4544(2.0) | 837(0.8) | 3699(1.1) | 2349(2.1) |
| [~32p]dGTP+ [~12p]dTTP+ [~12p]dGTP+ [~12p]dCTP + [~12p]dTTP+ [~12p]dGTP+ [~12p]dCTP | 2073(3.0) | 1277(2.2) | 64(1.0) | 357(1.0) | 570(2.6) |
| [~12p]dGTP+ [~12p]dTTP+ [~12p]dCTP | 213 | 1277(2.2) | 49 | 574(1.0) | 704(2.2) | 80 |
| [~12p]dGTP+ [~12p]dTTP+ [~12p]dCTP | 101 | 2300(2.0) | 1277(1.0) | 2059(1.9) |

| H12              |      |        |        |        |        |
| [~12p]dATP+[~32p]dGTP+ [~12p]dCTP | 401(3.4) | 7572(2.4) | 984(0.8) | 8600(1.0) | 2774(1.0) |
| [~32p]dGTP+ [~12p]dTTP+ [~12p]dCTP | 1595(3.6) | 426(1.0) | 99 | 3936(1.2) | 878(0.2) |
| [~12p]dATP+ [~12p]dGTP+ [~12p]dTTP+ [~12p]dCTP | 3377 | 2752(2.9) | 82 | 774(1.0) |
| dATP+ dGTP+ dCTP | 77 | 12 | 2 | 11 | 604(1.7) | 701(1.0) |

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

The numbers in parentheses are the experimental molar ratios.

---

recognized by the *E. coli* polymerase (32, 33). Finally, the sequence of the promoter region in the lactose operon is also known.8 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 λ 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-

---

* W. M. Barnes, J. Abelson, F. Blattner, B. Dickson, B. Reznikoff, and K. Thornton, personal communication.
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. DNA's 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
1. Loewen, P. C., Sekiya, T., and Khorana, H. G. (1974) J. Biol. Chem. 249, 217–226
2. Khorana, H. G., Agarwal, K. L., Buchi, H., Caruthers, M. H., Gupta, N. K., Kleppe, K., Kumar, A., Ohtsuka, E., RajBhandary, U. L., van de Sande, J. H., Sgaramella, V., Terao, T., Weber, H., and Yamada, T. (1970) J. Mol. Biol. 72, 209 and accompanying papers
3. Kleppe, R., and Khorana, H. G. (1972) J. Biol. Chem. 247, 6149
4. Terao, T., Dahlberg, J. E., and Khorana, H. G. (1972) J. Biol. Chem. 247, 6157
5. Russell, R. L., Abelson, J. N., Landy, A., Gelffer, M. L., Brenner, S., and Smith, J. D. (1970) J. Mol. Biol. 47, 1
6. Altman, S., and Smith, J. D. (1971) Nature New Biol. 233, 35
7. Miller, R. C., Jr., Besmer, P., Khorana, H. G., Flanzy, M., and Szilaske, W. (1971) J. Mol. Biol. 66, 363
8. Besmer, P., Miller, R. C., Jr., Caruthers, M. H., Kumar, A., Minamoto, K., van de Sande, J. H., Siderova, N., and Khorana, H. G. (1972) J. Mol. Biol. 72, 503
9. Loewen, P. C., and Khorana, H. G. (1973) J. Biol. Chem. 248, 3489
10. Sekiya, T., and Khorana, H. G. (1974) Fed. Proc. 33, 1424
11. Sekiya, T., and Khorana, H. G. (1974) Proc. Nat. Acad. Sci. U. S. A. 71, 2978–2982
12. Kössel, H., and Roychaudhury, R. (1971) Eur. J. Biochem. 22, 271
13. Brownlee, G. G., and Sanger, F. (1969) Eur. J. Biochem. 11, 365
14. Kleppe, K., Ohtsuka, E., Kleppe, R., Molineux, I., and Khorana, H. G. (1971) J. Mol. Biol. 66, 341
15. Sgaramella, V., and Khorana, H. G. (1972) J. Mol. Biol. 72, 427
16. Burton, K. (1967) Methods Enzymol. 12, 222
17. Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H., and Fisher, D. (1973) Proc. Nat. Acad. Sci. U. S. A. 70, 1209
18. Berg, P., Fancher, H., and Chamberlin, M. (1963) in Symposium on Informational Macromolecules, p. 467, Academic Press, New York
19. van de Sande, J. H., Loewen, P. C., and Khorana, H. G. (1972) J. Biol. Chem. 247, 6140
20. Wu, R., Donelson, J., Padmanabhan, R., and Hamilton, R. (1972) Bull. Inst. Pasteur 70, 203
21. Hedgpeth, J., Goodman, H. M., Boyer, H. W. (1972) Proc. Nat. Acad. Sci. U. S. A. 69, 5418
22. Boyer, H. W., Chow, L. T., Dugaiczky, A., Hedgepeth, J., and Goodman, H. M. (1973) Nature New Biol. 244, 40
23. Kelly, T. J., Jr., and Smith, H. O. (1970) J. Mol. Biol. 51, 393
24. Roy, P. H., and Smith, H. O. (1973) J. Mol. Biol. 81, 445
25. Weigel, P. H., Englund, P. T., Murray, K., and Old, R. W. (1975) Proc. Nat. Acad. Sci. U. S. A. 70, 1151
26. Gilbert, W., and Maxam, A. (1973) Proc. Nat. Acad. Sci. U. S. A. 70, 3581
27. Kleid, D., Agarwal, K. L., and Khorana, H. G. (1975) J. Biol. Chem., in press
28. Maniatis, T., Ptashne, M., Barrell, B. G., and Donelson, J. (1974) Nature 250, 394–397
29. Wilson, D. A., and Thomas, C. A., Jr. (1974) J. Mol. Biol. 84, 348
30. Locker, J., Rabinowitz, M., and Getz, G. S. (1974) Proc. Nat. Acad. Sci. U. S. A. 71, 1366
31. Bautz, E. K. F. (1973) Fed. Eur. Biochem. Soc. Lett. 36, 123
32. Zain, B. S., Weissman, S. M., Dhar, R., and Pan, J. (1974) Nucleic Acids Res., in press
33. Dhar, R., Weissman, S. M., Zain, B. S., Pan, J., and Lewis, A. M., Jr. (1974) Nucleic Acids Res., in press
34. Gilbert, W. (1966) Nature 212, 1480
35. Crick, F. (1971) Nature 234, 25
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

J. Biol. Chem. 1975, 250:1087-1098.

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