Proc. Nat. Acad. Sci. USA
Vol. 71, No. 8, pp. 2978–2982, August 1974

Nucleotide Sequence in the Promoter Region of the *Escherichia coli* Tyrosine tRNA Gene*

(tRNA precursor/initiation of transcription/primer elongation/DNA polymerase-I/2-fold symmetry)

TAKAO SEKIYA AND H. GOBIND KHORANA

Departments of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge, Mass. 02139

Contributed by H. Gobind Khorana, May 6, 1974

ABSTRACT **The sequence of 29 nucleotides immediately preceding the starting point of transcription of the *E. coli* tyrosine tRNA gene has been determined. This is:**

```
     1          10         20        29
(5') --GGGGCGCATCATATCAAATGACGCGCCGC (3')
(3') --CCCCGCGTAGTATAGTTTACTGCGCGGCG (5')
```

The sequence contains regions of 2-fold symmetry. Transcription of the gene begins at the first nucleotide to the left of nucleotide 1 and is leftward.

The nucleotide sequences in the DNA regions that, presumably, determine the initiation and termination of transcription have, until recently, remained unknown. In connection with our interest in the total synthesis of the *Escherichia coli* tyrosine tRNA gene, including its transcriptional signals, we have been investigating the nucleotide sequences adjoining the two ends of the DNA corresponding to the known precursor for the above tRNA (1). In a recent paper, the sequence of 23 nucleotide units beyond the C-C-A end of this tRNA, the terminator region, was reported (2). We have now determined the sequence of 29 nucleotides at the opposite end, namely, the region preceding the initiation of transcription. The sequence, which is shown in Fig. 1, possesses remarkable features. It contains two outer regions, which bear 2-fold symmetry relationship with one another and contain exclusively G·C base pairs. The central part of the sequence is A·T rich and also contains regions that show 2-fold symmetry. The sequence can loop out from the regular DNA double helix to form the secondary structure shown in Fig. 2. The present paper constitutes a preliminary report of these findings.

EXPERIMENTAL PROCEDURES

The general approach to the sequencing of the promoter region involves: (*1*) separation of the strands of the bacteriophage $\phi 80psu_{III}^{+}$ DNA which carries the gene for the tyrosine suppressor tRNA (3); (*2*) hybridization of appropriate deoxyribopolynucleotide primers at the tyrosine tRNA gene terminus in the l-strand of the above DNA; (*3*) controlled DNA polymerase-catalyzed extension of the primers at their 3′-ends, and determination of the nucleotide sequence of the newly added nucleotides to the primers. The experiments carried out and the results obtained in the present work are illustrated in Fig. 3. 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 and purified. Its sequence was shown to be as in the *dashed box* in *A* (Fig. 3). 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. 3), was elongated further in the presence of dATP, dTTP, and dCTP. The addition of 11 new nucleotides was now observed and their sequence was determined after cleavage at the rC site (*dashed box* in *B*, Fig. 3). 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. 3).

Materials. These were mostly as described earlier (2). The 21 unit long deoxyribopolynucleotide (DNA I, Fig. 3) corresponds to the nucleotide sequence 7-27 of the tyrosine tRNA precursor, numbering from the initiation site. The DNA was prepared by Dr. E. Ohtsuka by the T_4 ligase-catalyzed joining of the appropriate chemically synthesized segments (unpublished work of Drs. K. L. Agarwal, M. H. Caruthers, and J. H. van de Sande). DNA II (Fig. 3) was prepared by the terminal nucleotide transferase-catalyzed addition of rG (4) to the chemically synthesized dodecanucleotide segment, d(G-C-T-C-C-C-T-T-A-T-C-G); subsequent joining of this segment to the nonanucleotide was as for DNA I.

* This is paper 127 in the series "Studies on polynucleotides." Paper 126 in this series is ref. 2.

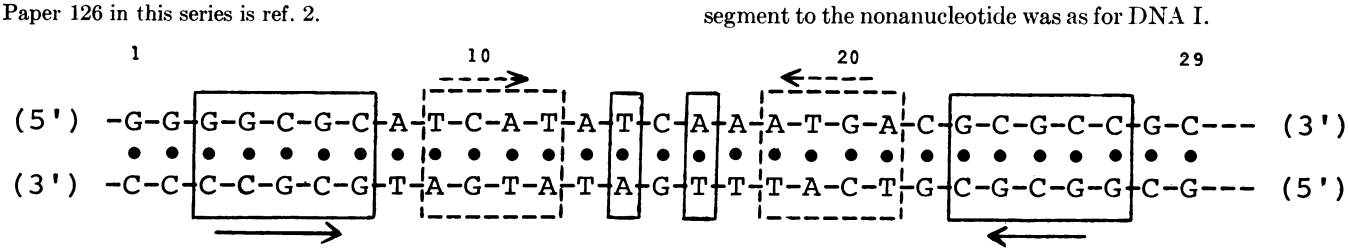

FIG. 1. The nucleotide sequence in the promoter region of the tyrosine tRNA gene. Elements of 2-fold symmetry in the sequence are shown in the *boxes* with matching *arrows* pointing to the axis of symmetry.

Nucleotide Incorporation Reactions. The primers, DNA I to DNA III (3–3.6 pmoles/ml), were hybridized to the l-strand of ϕ80psu$_{III}^+$ DNA (3 pmoles/ml) in 100 mM NaCl + 100 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (Pipes, pH 6.9) buffer as before (2). The conditions for the DNA polymerase reactions have been described (2).

Isolation of the Newly Formed Oligonucleotides. The radioactively labeled extended primers were separated from the l-strand, further purified by polyacrylamide gel (15%) electrophoresis (2) and then hydrolyzed with KOH to cleave the chain at rC or rG sites. The radioactive oligonucleotide chains representing the new sequences were purified by homochromatography (5). The individual radioactive products were used for partial venom and spleen phosphodiesterase degradations followed by a two-dimensional fingerprinting (6), degradation to 3′-nucleotides (2), and diphenylamine–formic acid degradation (2).

RESULTS

The Sequence of the First 8 Nucleotides is G-G-G-G-C-G-C-A (Fig. 3). In an initial experiment, it was established that the primer, DNA I, hybridized to the l-strand of ϕ80psu$_{III}^+$ DNA in a 1:1 molar ratio. The DNA polymerase-catalyzed reaction was carried out using DNA II as the primer and the three deoxynucleoside triphosphates, dATP, dGTP, and dCTP, one or more of the triphosphates being labeled with ^{32}P in the α position. The elongated primer was purified by gel electrophoresis, the bands obtained being shown in Fig. 4*a*. After an alkaline hydrolysis, homochromatography gave several radioactive bands, which are shown in Fig. 4*b*. These multiple products were produced evidently because the repair did not reach the same point in all molecules. The sequences of the bands were determined by fingerprinting, as exemplified in Fig. 5*A*, and were all consistent with the tridecanucleotide sequence G-A-A-G-C-G-G-G-G-C-G-C-A for the longest product. The results of extensive nearest-neighbor analyses and of the pyrimidine tracts were all consistent with this sequence. Therefore, the new sequence beyond the known initiation site is G-G-G-G-C-G-C-A. The fact that the primer, which lacked five known terminal nucleotides in the precursor

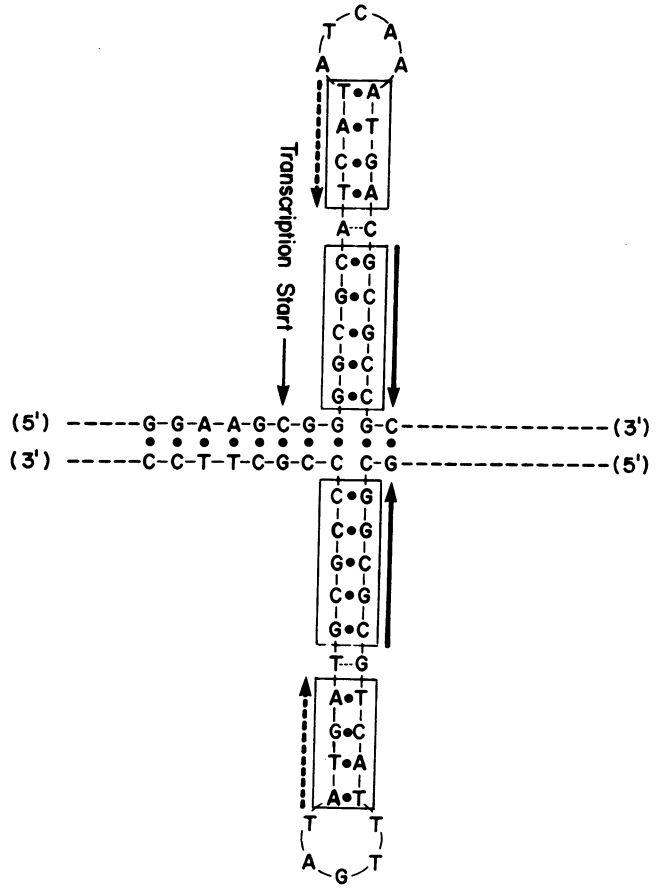

FIG. 2. A secondary structure model for the promoter region of the tyrosine tRNA gene.

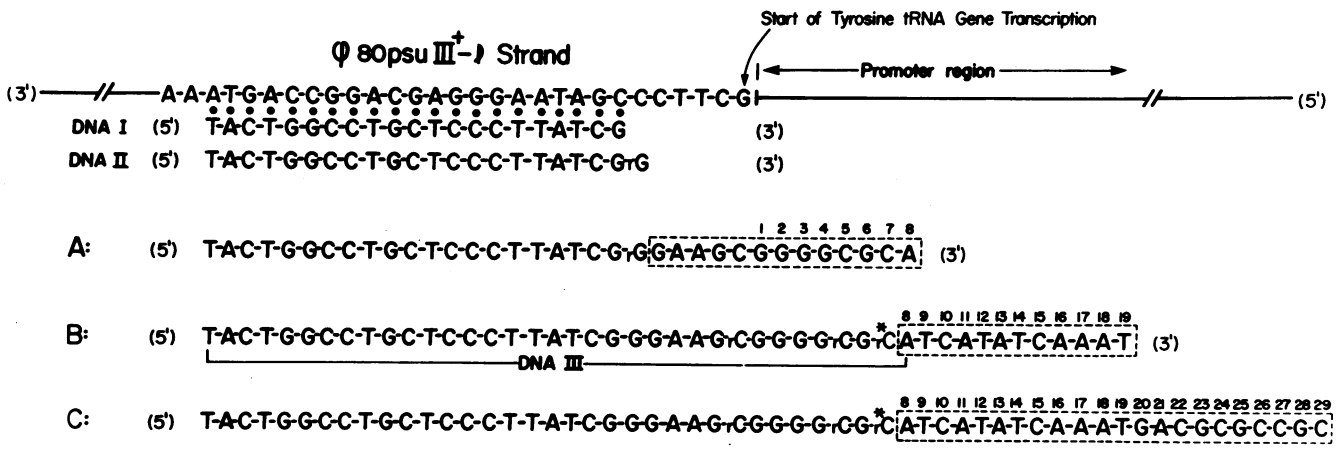

FIG. 3. Experimental plan for sequencing and the nucleotide sequences determined in the promoter region of the tyrosine tRNA gene. The primer-template complexes were initially obtained by hybridizing DNA I–DNA III to the l-strand of the ϕ80psu$_{III}^+$ DNA. DNA polymerase-catalyzed elongations were carried out using the nucleoside 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 34th 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, while the second was shorter by two nucleotides (rC-A) at the 3′-end. Elongation of this mixture with the deoxynucleoside triphosphate mixtures shown, therefore, gave dC in addition to rC at the 7th nucleotide in the promoter region.

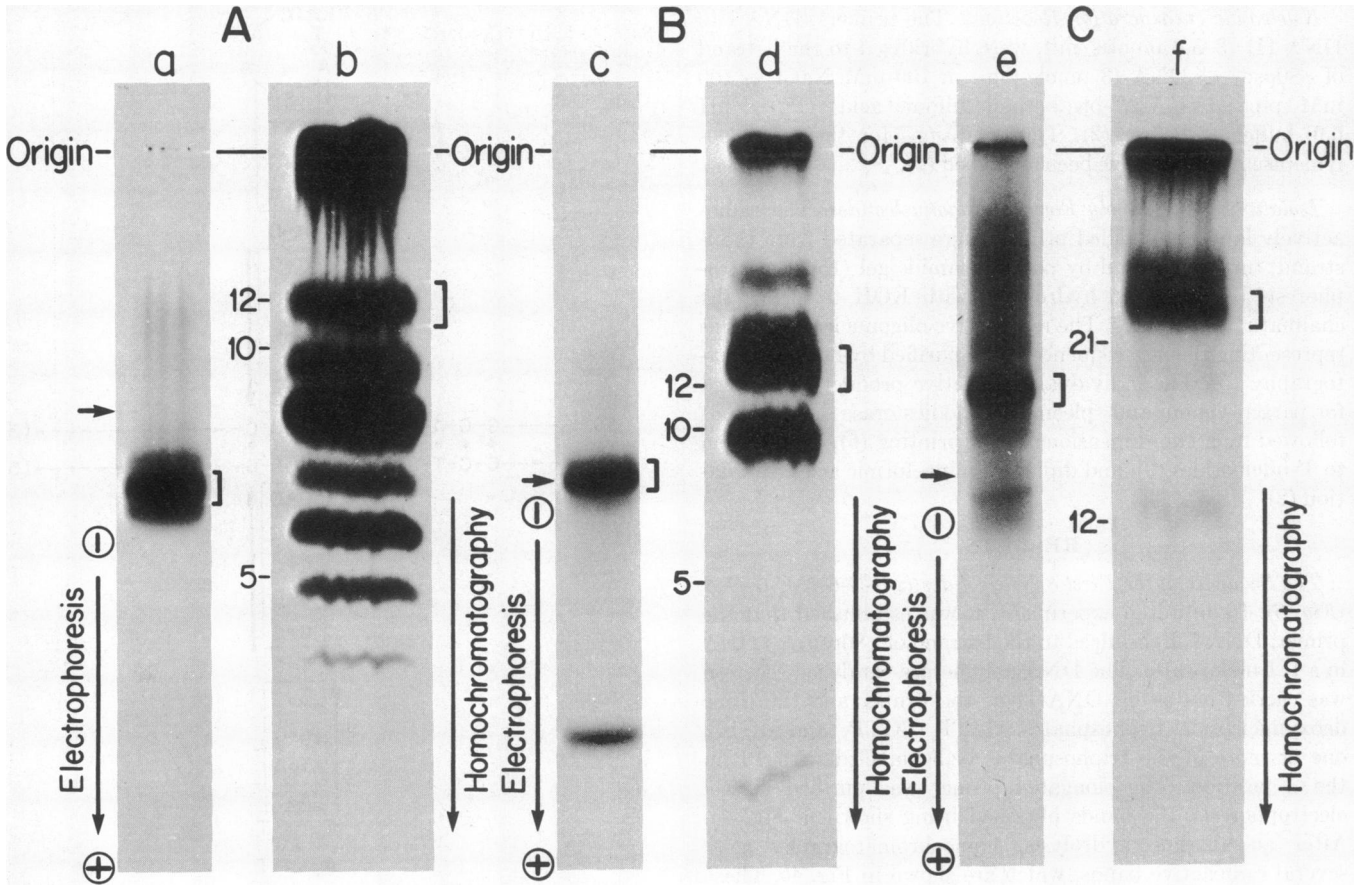

FIG. 4. Purification of the elongated primers and separation of the newly formed oligonucleotides. Elongation of the primers was carried out as described in the *text* and previously published (2). The elongated primers were separated from the excess deoxyribonucleoside triphosphates and the template DNA by gel filtration through an Agarose 1.5 m column (0.9 × 24-cm) and subjected to electrophoresis on a 15% polyacrylamide gel slab (20 × 20 × 0.2-cm) in 7 M urea in 90 mM Tris·borate (pH 8.3), 4 mM ethylenediaminetetraacetate (EDTA) for 18 hr. at 10 mA. Channel *a* contains the product when DNA II was elongated with [α-^{32}P]dATP, [α-^{32}P]dGTP, and [α-^{32}P]-dCTP. Channel *c* contains the product from DNA III elongated with [α-^{32}P]dATP, [α-^{32}P]dTTP, and [α-^{32}P]dCTP. Channel *e* contains the DNA III elongated with [α-^{32}P]dATP, [α-^{32}P]dGTP, and [α-^{32}P]dCTP in the presence of a limited amount of unlabeled dTTP (0.015 μM). The horizontal *arrow* in each of these channels indicates the position of the dye, xylene cyanol, used as a marker. The bands indicated by *brackets* in these channels were eluted and hydrolyzed with 0.5 M KOH at 37° for 18 hr. After neutralization with acetic acid, the hydrolysate was subjected to homochromatography. Channel *b* contains the newly formed oligonucleotides obtained from the indicated band in channel *a*. Channel *d* contains the oligonucleotides obtained from the indicated product in channel *c*. Channel *f* contains the oligonucleotide obtained from the indicated band in channel *e*. Homochromatography as shown in channels *b* and *d* was carried out using the oligonucleotide mixture prepared by hydrolysis of 3% yeast RNA in 7 M urea (pH 7.0) in 20 mM KOH at 65° for 18 hr. Homochromatography shown in channel *f* was carried out using a mixture which was prepared as described above except that 10 mM KOH was used. The *numbers* (5, 10, 12, and 21) indicate the positions taken by the pentanucleotide, [5′-^{32}P]pG-G-A-A-rG, the decanucleotide, [5′-^{32}P]p-G-G-A-G-C-A-G-G-C-C, the dodecanucleotide, [5′-^{32}P]pG-C-T-C-C-C-T-T-A-T-C-G, and [5′-^{32}P]DNA I, respectively. The bands purified by homochromatography and indicated were eluted and subjected to sequence analysis.

sequence, first incorporated these nucleotides in the correct sequence, gave added assurance that the primer elongation occurred at the correct site.

The Sequence of the Next 11 Nucleotides is T-C-A-T-A-T-C-A-A-A-T (*Fig. 3*). DNA I was first extended using dATP, dGTP, and rCTP. The concentrations of the triphosphates were increased relative to those used in the preceding experiment in order to achieve more complete repair. The extended primer (DNA III) was isolated free from the triphosphates and the template. When examined by gel electrophoresis, the product contained two components, one in which the new octanucleotide sequence G-G-G-G-C-G-C-A was present in full, and the second component in which the sequence C-A was missing from the 3′-end. Since no complication was expected in deducing the sequence from the use of this mixture, DNA III was used in the next step without further purification.

Two different sets of triphosphates were used separately for further elongation of DNA III. In one experiment, dGTP, dCTP and dTTP were used. The reaction resulted in the addition of only the T-C sequence to the terminal A. This was shown by alkaline hydrolysis of the extended primer, when the trinucleotide A-T-C was identified as the major product. In the second experiment, DNA III was elongated using dATP, dTTP, and dCTP. The newly added oligonucleotide chain† was isolated by the above procedures (Fig. 4*c* and *d*)

† One product formed concomitantly clearly arose from DNA III that lacked the C-A sequence at the 3′-end.

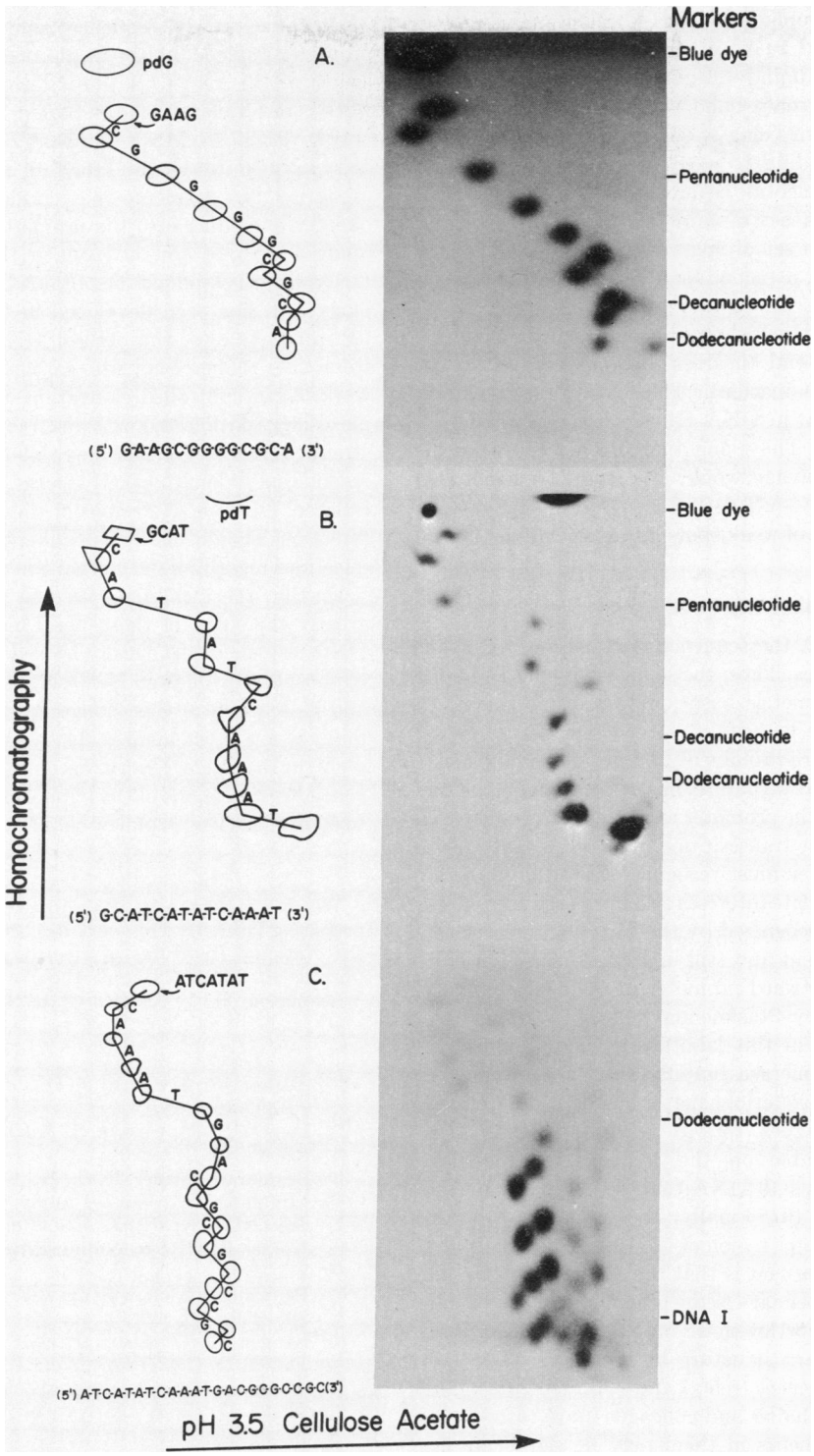

FIG. 5. Fingerprints of the products obtained by partial digestion with venom phosphodiesterase of the newly formed oligonucleotides. The oligonucleotides obtained as described in Fig. 4 were digested partially with snake venom phosphodiesterase. The digests were subjected to the two-dimensional fingerprinting procedure (6). For homochromatography, the RNA hydrolysate with 20 mM KOH was used. The sequences of the full-length oligonucleotides in *A*, *B*, and *C* are shown in Fig. 3 and at the *bottom* in the *frames to the left*. The oligonucleotide fingerprinted in *A* came from Fig. 4*b*. The oligonucleotide in *B* came from Fig. 4*d* and was obtained by elongation of DNA III lacking the rC-A sequence at the 3'-end. In *C*, the product from Fig. 4*f* was used. Two sets of radioactive spots were visible. The explanation for these is given in the *text*.

and sequenced by the fingerprinting method (Fig. 5*B*). The unique sequence for the 11 nucleotide units derived by this method was confirmed by extensive nearest-neighbor analyses using different combinations of the α-^{32}P-radioactive triphosphates and by analyses of the pyrimidine tracts.

Sequence of the Next 10 Nucleotides is G-A-C-G-C-G-C-C-G-C (Fig. 3). Elongation of DNA III was further studied by the addition of low concentration (0.015 μM) of TTP to the mixture (1–10 μM) of dATP, dCTP, and dGTP. Products somewhat longer were formed, which were fractionated by

polyacrylamide gel electrophoresis (Fig. 4*e*). The main product, which was estimated to be about 60 nucleotides long, was hydrolyzed with alkali and was purified by homochromatography (Fig. 4*f*). The main band was then sequenced by fingerprinting, the pattern being shown in Fig. 5*C*. A series of main spots corresponding to a set was found that corresponded to the 22 nucleotide sequence shown in Fig. 5*C* (*bottom left*). However, a set of faint spots showing correspondence with the main set of spots was also present: this set evidently arose from a second product, which contaminated the main product. The contaminant, which could be purified in a small amount, was shown to contain an additional G-C sequence at the 5′-end (nearest-neighbor analysis and pyrimidine tracts determination). This product evidently arose from the component in DNA III which lacked the C-A sequence at the 3′-end. The elongation of this component followed by alkaline cleavage would give the 24 nucleotide long product, G-C-A-T-C-A-T-A-T-C-A-A-A-T-G-A-C-G-C-G-C-C-G-C. The pattern of weak spots in Fig. 5*C* is consistent with this sequence.

DISCUSSION

As seen in Figs. 1 and 2, the sequence now determined possesses remarkable features. Thus, the sequence is composed of regions with very asymmetric base-composition and having 2-fold symmetry. In the looped-out form (Fig. 2), the G·C rich regions are even more prominent and provide stable continuous stems. The total structure has a rotational axis at the fifteenth basepair in the promoter region.

Since, evidently, all the transcription in *E. coli* is performed by one enzyme, the latter must recognize all the promoters. It is, therefore, of immediate interest to compare the sequences of different promoters recognized by the *E. coli* transcriptase. The progress here is recent and still limited. The sequence of 36 nucleotides in the leftward promoter in the bacteriophage λ has been determined by Ptashne and colleagues (7) and by Kleid and Agarwal (8) in this laboratory. The sequence of one of the strong polymerase binding sites present in the replicative form of the bacteriophage fd has been determined by Schaller and colleagues (private communication). Information is also available on the sequence of the promoter in the simian virus 40 DNA that is recognized by the *E. coli* polymerase (9, 10). Finally, the sequence of a few nucleotides in the promoter region of the lactose operon is also known (11). The striking fact is that the above promoters all differ widely in their primary sequence. Therefore, the important concept must forthwith be invoked that the polymerase recognizes a tertiary structure in the DNA. The fact that the promoter sequences (refs. 7 and 8, present work, and unpublished work of Schaller and colleagues) are turning out to have symmetry elements in them may be significant for the attainment of three-dimensional structures. However, it remains for future work to determine the nature of the three-dimensional structure and the lengths of the DNA regions around the initiation site that contribute to it. It is recalled that the concept that DNA may be recognized at certain sites by the attainment of three-dimensional folded structures was propounded by Gierer in 1966 (12).

Does the present sequence comprise all or a part of the promoter sequence? How would one precisely define the total length of this region? While, at this stage, further sequence work is desirable, it is clear that definitive answers can come only by *actual* studies of the transcriptional process, the specificity in the initiation, and the rate of initiation. These require, in turn, DNA segments that contain various lengths of sequences extending into the promoter 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 that has been developed in this laboratory. However, for initial studies, the desired DNA's 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 DNA's corresponding in length exactly to the elongated primers may be isolated. Obvious extensions of these experiments in different directions should make it possible to answer some of the finer questions regarding the mechanism of transcription.

Finally, the techniques used in the present work increase greatly the scope of the general method proposed earlier for DNA sequencing (2, 13). The use of a restricted number of triphosphates followed by isolation of the elongated primer and its reuse in further elongation greatly increases the scope for growing short manageable oligonucleotide chains for unambiguous sequence work. Secondly, the use of a rate-limiting concentration of one of the triphosphates also makes it possible to control the length of new chains at will.

We are grateful to Dr. Ranajit Roychoudhury for assistance in preparing DNA II. This work has been supported by grants from the National Institutes of Health, U.S. Public Health Service (Grant no. CA05178) and the National Science Foundation, Washington, D.C. (Grant no. GB-21053X).

1. Altman, S. & Smith, J. D. (1971) *Nature New Biol.* **233,** 35–39.
2. Loewen, P. C., Sekiya, T. & Khorana, H. G. (1974) *J. Biol. Chem.* **249,** 217–226.
3. Besmer, P., Miller, R. C., Jr., Caruthers, M. H., Kumar, A., Minamoto, K., van de Sande, J. H., Sidarova, N. & Khorana, H. G. (1972) *J. Mol. Biol.* **72,** 503–522.
4. Kössel, H. & Roychoudhury, R. (1971) *Eur. J. Biochem.* **22,** 271–276.
5. Brownlee, G. G. & Sanger, F. (1969) *Eur. J. Biochem.* **11,** 395–399.
6. Sanger, F., Donelson, J. E., Coulson, A. R., Kössel, H. & Fisher, D. (1973) *Proc. Nat. Acad. Sci. USA* **70,** 1209–1213.
7. Maniatis, T. & Ptashne, M. (1974) *Nature,* in press.
8. Kleid, D. G. & Agarwal, K. L. (1974) *J. Biol. Chem.,* in press.
9. Zain, B. S., Weissman, S. M., Dhar, R. & Pan, J. (1974) *Nucleic Acids Res.,* in press.
10. Dhar, R., Weissman, S. M., Zain, B. S., Pan, J. & Lewis, A. M., Jr. (1974) *Nucleic Acids Res.,* in press.
11. Gilbert, W. & Maxam, A. (1973) *Proc. Nat. Acad. Sci. USA* **70,** 3581–3584.
12. Gierer, A. (1966) *Nature* **212,** 1480–1481.
13. Wu, R., Donelson, J., Padmanabhan, R. & Hamilton, R. (1972) *Bull. Inst. Pasteur* **70,** 203.