Total Synthesis of the Structural Gene for the Precursor of a Tyrosine Suppressor Transfer RNA from *Escherichia coli*

1. GENERAL INTRODUCTION*


From the Departments of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 and, the Institute for Enzyme Research of the University of Wisconsin, Madison, Wisconsin 53706

With the ultimate objective of the total synthesis of a tRNA gene including its transcriptional signals, an *Escherichia coli* tyrosine suppressor tRNA gene was chosen. The arguments in favor of this choice are presented. A plan for the total synthesis of the 126-nucleotide-long DNA duplex corresponding to a precursor (Altman S., and Smith, J. D. (1971) *Nature New Biol.* **233**, 35) to the above tRNA is formulated. The plan involves: (a) the chemical synthesis of 26 deoxyribooligonucleotide segments, (b) polynucleotide ligase-catalyzed joining of several segments at a time to form a total of four DNA duplexes with appropriate complementary single-stranded ends, and (c) the joining of the duplexes to form the entire DNA duplex. Ten accompanying papers describe the experimental realization of this objective.

Methods have been developed in recent years for the synthesis of oligonucleotides of defined nucleotide sequences. These involve: (a) the chemical synthesis of short deoxyribooligo-

* This work has been supported by grants from the National Cancer Institute of the National Institutes of Health, United States Public Health Service (CA05178 and CA11981), the National Science Foundation, Washington, D.C. (GB-7434X, GB-21053X, GB-36881X, and BMST3-06757), the American Cancer Society (NP-140) and by funds made available to the Massachusetts Institute of Technology by the Sloan Foundation. This is Paper CXXXI in the series “Studies on Polynucleotides.” The preceding paper is Ref. 1.

† Present address, Department of Biochemistry, University of Chicago, Chicago, Illinois 60637.
§ Present address, Department of Biology, Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.
¶ Present address, Department of Zoology, University of Colorado, Boulder, Colorado 80302.
|| Present address, Biology Department, University of New Brunswick, Fredericton, New Brunswick, Canada.
* Present address, Department of Organic Chemistry, The Weizmann Institute of Science, Rehovot, Israel.
* Present address, Department of Biochemistry and Molecular Biology, Cornell University, Ithaca, New York 14850.
~ Present address, Department of Biochemistry, University of Bergen, 5000 Bergen, Norway.
' Present address, Department of Biochemistry, All India Institute of Medical Sciences, New Delhi, India.
Present address, Department of Biochemistry, University of Manitoba, Winnipeg, Manitoba, Canada.

nucleotide segments corresponding to the entire two strands of the intended DNA; (b) phosphorylation of the 5'-hydroxyl end groups in the synthetic oligonucleotides using polynucleotide kinase, and (c) the head to tail joining of the appropriate segments when they are aligned to form bihelical complexes using the T<sub>4</sub> polynucleotide ligase. This methodology has been successfully applied to the total synthesis of the 77-nucleotide-long DNA corresponding to the major yeast alanine tRNA (2). While the accomplishment of this synthesis established confidence in the general methodology for DNA synthesis, and the availability of several relatively short DNA duplexes of defined nucleotide sequences made it possible to study aspects of transcription (3, 4) and of DNA enzymology (5-7), the synthetic DNA corresponding to the yeast alanine tRNA proved, at least for some time, unsuitable for studies of certain problems of central biochemical interest. For example, it had been hoped that the availability of synthetic DNAs would permit further studies of the following two problems: (a) the mechanism of initiation and termination of transcription and (b) precise structure-function relationship

---

* Present address, Department of Microbiology, University of British Columbia, Vancouver, British Columbia, Canada.
† Present address, Department of Applied Organic Chemistry, Nagoya University, Nagoya, Japan.
‡ Present address, Department of Virology, Hebrew University, The Hadassah Medical School, Jerusalem, Israel.
§ Present address, Division of Medical Biochemistry, Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada.
Synthesis of Tyrosine Suppressor tRNA Precursor Gene

in tRNA. With the continued hope of being able to apply the synthetic approach to these and related problems, the total synthesis of the DNA corresponding to an *Escherichia coli* transfer RNA gene was undertaken. We now wish to report the total synthesis of a DNA corresponding to the entire length (126 nucleotides) of the precursor to an *E. coli* tyrosine suppressor tRNA. The present paper gives the main arguments for the choice of this RNA and introduces the synthetic plan, while ten accompanying papers document the experimental realization of the objective (6-17). Brief reports on portions of this work have appeared during the last 4 years (18-21).

The first requirement for undertaking synthesis of a DNA is the specification of its sequence. For RNAs whose sequences are known, the sequences of the genes can be deduced directly. Further, among RNAs, the choice was made in favor of tRNA genes because of a variety of reasons. Of the various classes of gene products, the tRNAs are easily the most intriguing in regard to structure and function. These molecules have to be recognized by a rather large number of components of the protein-synthesizing machinery, such as by the aminoacyl-tRNA synthetases, by the nucleotidyltransferase which repairs the C-C-A end, by the ribosomes and by several proteins involved in protein chain initiation, elongation, and termination, and finally by messenger RNA. Also, tRNA molecules abound in modified bases and the nascent tRNA molecules have to be recognized by several modifying enzymes. Indeed, the tRNAs are a unique class of molecules, which evidently possess common secondary structure characteristic of nucleic acids but they also undergo folding to adopt tertiary structures. This has been amply demonstrated by the establishment of tertiary structures in a number of cases and the elucidation of the structure by x-ray diffraction methods (22, 23). Despite this recent progress, understanding of the structure-function relationships is largely lacking. It is hoped that chemical synthesis could, in principle, offer a definitive approach of wide scope. Different parts of the tRNA structure could be systematically modified at the gene level. The modifications could involve additions, deletions, or substitutions of single or a few bases, or could be more extensive, such as the replacements of loops and stems by those present in different tRNAs.

**CHOICE OF *ESCHERICHIA COLI* TYROSINE tRNA SUPPRESSOR GENE**

The first major consideration in favor of an *E. coli* tRNA gene was the fact that biochemical work in the tRNA field is much more advanced with *E. coli* than with other organisms. Thus, the cell-free protein-synthesizing system, its characterization, the biochemistry of the ribosomes, and the understanding of the various factors required for initiation, elongation, and termination of polypeptide chains are all much better understood than with other systems. Specifically in the case of tyrosine tRNA, the aminoacyl-tRNA synthetase had been purified and characterized by Calendar and Berg (24).

A second consideration was the accuracy of the nucleotide sequence of the tRNA chosen. An assurance on this account at the start of the synthetic work was obviously desirable. The sequence of the *E. coli* tyrosine tRNAs (including the amber suppressor tRNA) was first determined by Goodman et al. (25). Fortunately, the same sequence was determined independently for this tRNA by RajBhandary, Nishimura, and their co-workers (26) by using a separate set of methods.

From the standpoint of studies on structure-function relationships, among many other general considerations, two specific lines of reasoning in favor of tyrosine tRNA were as follows. Firstly, a comparison of the *E. coli* tRNA<sup>Tyr</sup> and *E. coli* tRNA<sup>Met</sup> sequences (Fig. 1) (27) showed remarkable similarities in parts of the cloverleaf structures but a striking difference was in the size of the loop III in the two tRNAs. It seemed reasonable to investigate the minimal changes in tRNA<sup>Tyr</sup> which would be required to elicit an initiator function in protein synthesis (see also Kleppe et al. (17)). Secondly, the

![Fig. 1](http://www.jbc.org/)

**Fig. 1.** Cloverleaf models of the primary nucleotide sequences of an *Escherichia coli* tyrosine suppressor tRNA (A) and the *E. coli* tRNA<sup>Met</sup> (B).
Synthesis of TyrosineSuppressor tRNA Precursor Gene

The choice of tyrosine suppressor tRNA gene proved to be uniquely fortunate from the standpoint of the initiation of transcription of this gene. During the progress of the synthetic work, Altman and Smith (29) discovered that the tyrosine tRNA is actually formed in vitro as a precursor, which is then processed to give the functional tRNA. The structure of the tRNA precursor is shown in Fig. 2. A unique feature in this structure, which had been absent in all the other tRNAs studied, is that it carries a triphosphate group at the 5'-terminus. This identifies beyond any doubt the initiation point for the transcription of the tRNA precursor. The knowledge of the precise initiation point for transcription permits a focus on the binding site of the RNA polymerase and on the mechanism of initiation of transcription. Further, the availability of the φ80psu111 DNA made it convenient to undertake work on the nucleotide sequences in the regions adjoining the two ends of the precursor RNA (Fig. 2). These regions presumably correspond to the promoter and the terminator regions. The two strands of φ80psu111 DNA could be separated at the start of the sequence work, and it was readily determined which strand of the DNA hybridizes with the tRNA or its deoxyribonucleotide analogs (30). By specific hybridization of adequately long synthetic deoxyribonucleotide primers corresponding to the 3'-end (31, 32) and to the 5'-end (33) with the appropriate DNA strands, template-primer systems could be set up. DNA polymerase-catalyzed incorporation of nucleotides would lead to the extension of the primer according to the sequence in the template strand. The nucleotide sequence of the unknown region could then be deduced from the pattern of nucleotide incorporation.

The mechanisms of initiation and of termination of transcription would hopefully be clarified by a combination of sequence work and synthesis. After the attachment of these DNA regions to the synthetic structural gene, it would naturally be expected that it would be possible to insert the gene into an amber mutant of E. coli. The tyrosine suppressor tRNA gene appeared to be eminently suitable in this regard for demonstrating biological activity in vitro.

 PLAN FOR TOTAL SYNTHESIS

Chemical Synthesis of Deoxyribonucleotide Segments—
In deriving the nucleotide sequence of the DNA from precursor RNA (Fig. 2), the only assumption made was that base modifications all occur after transcription in which only the four standard ribonucleoside triphosphates are used. A large body of evidence, which has since accumulated, abundantly supports this conclusion. Thus, ψ is derived from uridine and all the other modifications (methylation, sulfurylation, isopentenylation) involve the parent recognizable bases in the tRNA backbone (34).

Given only the basic requirement of providing overlaps of 4 to 6 base-pairs between the adjoining chemically synthesized segments, the possibilities for division into segments of the two strands representing the precursor length are enormous. In deriving a plan, three types of considerations were borne in mind. (a) The plan should aim at maximum economy in chemical synthesis, compatible with the two considerations given below. Chemical synthesis continues to be the most time-consuming, demanding, and, therefore, the major progress-determining step in DNA synthesis, and multiple use of oligonucleotide blocks and of total or parts of segments is of great practical importance. (b) For reasons which are not as yet fully understood, the enzymatic joinings do not go to completion. Indeed, the yields vary widely in different systems, as shown in the accompanying papers. Therefore, the synthetic plan should, in principle, optimize the yields in joining reactions. (c) Self-complementary structures, if present at the 5'-end (see below) would lead to undesired dimer formation (35). These and the presence of even short complementary sequences in the same or in different segments to be used in one joining reaction should be avoided. Presumably, the formation of short, partially correct base-paired structures can compete with the desired perfectly ordered Watson-Crick duplexes.

A systematic weighing of various factors in the synthesis of a long DNA is largely precluded at this time because of the absence of adequate data. In the present work, the following practical considerations aided the derivation of the total plan shown in Fig. 3. First, there were several segments or parts which were available from the previous synthetic work on the DNA corresponding to the alanine tRNA and their use was given a primary consideration.

Thus, the hexanucleotide d([C-G-G-T-G-G]), which formed a part of segment 1 in the previous work, now became segment 1 in the present plan. The heptanucleotide d([A-G A G T C T]), which was segment 7 in the alanine gene synthesis, was now included in segment 11 and the heptanucleotide d([G-C-T-C-C-T]), which earlier formed a part of segment 8, now was used as a part of segment 13. Similarly, the nonanucleotide, d(G-C-G-
reason being that initially the aim was to synthesize only the 85-nucleotide unit-long DNA corresponding to the tyrosine tRNA, the synthetic plan in this region was reconsidered. The plan proposed aimed at the chemical synthesis of the pentadecamer containing the undecamerule of segment 17 and the tetranucleotide shown as segment 16. Unfortunately, the synthesis of segment 17 alone proved to be rather overwhelming and an extension of this synthesis to include segment 16 was not practical. It was therefore hoped that conditions might be found for the subsequent enzymatic reactions such that the tetranucleotide (segment 16) would join to the neighboring segments 14 and 17. The enzymatic joinings were indeed carried out successfully although the yields left a great deal to be desired (15). An improvement in the enzymatic joinings in this part is still under investigation by undertaking the synthesis of the tetradecamer in which combines the present segment 16 with segment 14. The results of this study will be reported upon at a later date.

Finally, it may be noted that in the plan shown in Fig. 3, single-stranded runs (hexanucleotide sequences) are available at the ends of the double-stranded DNA for extension to the regions which would, presumably, form the promoter and terminator regions for the transcription of the gene.¹

**ENZYMATIC JOINING OF CHEMICALLY SYNTHESIZED SEGMENTS**

As mentioned above, the results of joining experiments are frequently unpredictable. In experimental systems containing three or four segments, the yields vary very widely. Therefore, a large amount of empirical work is necessary to determine the combination of segments which would give optimal yields in the overall joining reactions. Following extensive experimentation, the 26 chemically synthesized segments (Fig. 3) were divided into four groups shown in Fig. 5. While the detailed arguments are presented in the individual papers dealing with different sections (13-16), it may simply be mentioned here that duplex [I] could only go as far as segment 5. In duplex [II], as many as eight segments could be used in a one-step joining reaction without any ambiguity. Duplex [III] required particularly detailed investigation for reasons mentioned above. The yield was rather low. Duplex [IV] consisted of six to seven segments (segments 19 to 25). A great surprise was the failure to join segment 26 to the remainder of duplex [IV].

Having prepared the four duplexes shown in Fig. 5, the next step was to quantitatively phosphorylate the terminal 5'-OH groups in the duplexes in preparation for the ligase-catalyzed joining to complete the synthesis of the total duplex. Work with a number of defined duplexes² showed, however, that the rates and extent of phosphorylation of 5'-OH groups at the termini of DNA duplexes by the polynucleotide kinase are influenced very much by the duplex structures around the 5'-OH groups. To ensure facile and complete phosphorylation, it seemed clearly desirable to have the terminal 5'-OH groups at the protruding single-stranded ends of the duplexes. The grouping was therefore amended in regard to duplexes [II], [III], and [IV], as shown in Fig. 6. The modified grouping largely met the above requirement and no difficulty was experienced in the phosphorylation reactions with the duplexes and, therefore, in the completion of the total synthesis (17).

Five accompanying papers (8-12) describe in a condensed

---

¹ Present work and unpublished work from the laboratory of Dr. K. Kleppe, Bergen, Norway.

---

---
Synthesis of Tyrosine Suppressor tRNA Precursor Gene

**Fig. 4.** A plan for the enzymatic joining of segments involving a self-complementary single-stranded protruding end. The 5'-end of the self-complementary region carries a 5'-32P-phosphate group. Treatment with polynucleotide ligase leads to the undesired symmetrical dimer.

**Fig. 5.** Escherichia coli tyrosine tRNA-precursor gene. Grouping of the 26 chemically synthesized segments for the purpose of polynucleotide ligase-catalyzed reactions. The four duplexes [I] to [IV] were thus prepared.

**Fig. 6.** Escherichia coli tyrosine tRNA precursor gene. Modification of the plan shown in Fig. 5 for the enzymatic joining of the chemically synthesized segments. The four duplexes now shown contained mostly 5'-OH groups at protruding single-stranded ends. Phosphorylation by polynucleotide kinase followed by the ligase-catalyzed joining led to the successful total synthesis of the continuous bihelical DNA.

form the large body of work comprising the syntheses of the 26 segments. Four following papers describe the next phase of the synthetic effort, namely, the polynucleotide ligase-catalyzed joining of the segments to form duplexes [I] to [IV] (13-16). Finally, the last paper (17) successfully accomplishes the total synthesis of the double-stranded DNA.

**CONCLUDING REMARKS**

The next immediate goal is the controlled transcription of the totally synthetic gene and subsequent maturation of the transcript to the functional tyrosine tRNA. While means could be found to carry out the transcription at the present stage of synthesis, a fundamentally more interesting approach (see above) would be to understand the biological signals for transcription. The latter require (a) the elucidation of the nucleotide sequences of the regions adjoining the two ends of the DNA corresponding to the precursor for the tRNA; and (b) elucidation of the mechanism of initiation and termination of transcription and precise determination of the lengths of the DNA regions involved in these processes. Then should follow the synthesis of the required DNA duplexes and their attachment to the appropriate ends of the synthetic structural
gene. Towards these objectives, the sequence of 23 nucleotides in the region adjoining the C-C-A end has been determined (32) and the corresponding DNA duplex has already been synthesized (36).

Similarly, the sequence of 29 nucleotides immediately adjacent to the initiation point of transcription of the precursor gene. Towards these objectives, the sequence of 23 nucleotides determination is rapid. It seems very likely that insights into a number of laboratories and progress in sequence are recognized by the E. coli polymerase are under intensive polymerase. In addition, a number of other promoters which the mechanism of action of the DNA-dependent RNA sequence becomes known. Concurrently, work is in progress on continuing on the sequence work, as well as on synthesis as the to the tRNA has also been determined (33). Studies are to the tRNA has also been determined (33). Studies are

**REFERENCES**


Downloaded from http://www.jbc.org/ by guest on September 22, 2017
Total synthesis of the structural gene for the precursor of a tyrosine suppressor
transfer RNA from Escherichia coli. 1. General introduction.
H G Khorana, K L Agarwal, P Besmer, H Büchi, M H Caruthers, P J Cashion, M
Fridkin, E Jay, K Kleppe, R Kleppe, A Kumar, P C Loewen, R C Miller, K Minamoto,
A Panet, U L RajBhandary, B Ramamoorthy, T Sekiya, T Takeya and J H van de Sande

Access the most updated version of this article at http://www.jbc.org/content/251/3/565

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/251/3/565.full.html#ref-list-1