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

1. GENERAL INTRODUCTION*

H. Gobind Khorana, Kan L. Agarwal,† Peter Besmer,§ Henry Büchi,¶ Marvin H. Caruthers,∥ Peter J. Cashion,* Matt Fridkin,* Ernest Jay,* Kjell Kleppe,* Ruth Kleppe,* Ashok Kumar,* Peter C. Loewen,~ Robert C. Miller,~ Katsumaro Minamoto,~ Amos Panet,* Uttam L. Rajbhandary, Belagaje Ramamurthi, Takao Sekiya, Tatsuo Takeya, and J. Hans van de Sande*

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 deoxyribonucleotide 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.
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 tyrosine suppressor tRNA 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 (8-17). 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 coworkers (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'Tyr and E. coli tRNA'Met 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'Tyr 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_Met (B).
Synthesis of Tyrosine Suppressor tRNA Precursor Gene

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 ribonucleotide triphosphates are used. A large body of evidence, which has since accumulated, abundantly supports this conclusion. Thus, \( \psi \) 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 of their 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 di[1-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 di(G-C-A-G-C-T), which was segment 7 in the alanine gene synthesis, was now included in segment 11 and the heptanucleotide di(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, di(G-C-
T-C-C-C-T-T-A), also derived from the above heptanucleotide and belonging to segment 8 in the previous work, now was used in segment 24. Further, a systematic search (for occurrences occurring more than once) in the present DNA* showed that the nonanucleotide sequence d(C-C-C-A-C-C-A-C) occurs in segment 9 as well as in segments 15-18. Similarly, the hexanucleotide sequence, d(T-T-C-G-A-A), occurs twice (nucleotides 18-23 and 26-31) in the same strand and because it is self-complementary, it is also present twice in the complementary strand. Shorter sequences can be found to repeat with increasing frequency, but their use may not always be allowed by other considerations.

As mentioned above, the hexanucleotide sequence d(T-T-C-G-A-A) occurs four times. The best use of this common sequence would be to place it at the 5'-end of four segments and the same protected hexanucleotide intermediate could then be rapidly elongated to give the four segments. How- ever, it is not possible to do so because of the previous experience with a situation of this type (35). Thus, a duplex such as the one shown in Fig. 44 would rather undergo dimerization to give B in Fig. 4 than to add the required terminal 5'-OH groups at the protruding single-stranded ends. 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 tetradecanucleotide 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

**ENZYMIC 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 duplexes* showed, however, that the rates and extent of phosphorylation of 5'-OH groups at the termini of DNA duplexes by the polynucleo- tide 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

---

*Present work and unpublished work from the laboratory of Dr. K. Kleppe, Bergen, Norway.
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 DNA.
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 to the tRNA has also been determined (33). Studies are continuing on the sequence work, as well as on synthesis as the sequence becomes known. Concurrently, work is in progress on the mechanism of action of the DNA-dependent RNA polymerase. In addition, a number of other promoters which are recognized by the E. coli polymerase are under intensive study. Work is also being done on synthesis and its successful application in the determination of the nucleotide sequences in the control regions in a variety of genetic systems, it seems certain that synthesis as the general method of binding of the enzyme, selection of the initiation site and related aspects of the mechanism of transcription will be gained in the near future. Synthetic work could further aid in more precisely defining the chemistry of the various steps in the overall process. Consequently, controlled transcription of the synthetic gene for the precursor to the tRNA should be possible.

With the recent dramatic progress in methodology for DNA sequencing and its successful application in the determination of the nucleotide sequences in the control regions in a variety of genetic systems, it seems certain that synthesis as exemplified in the present series of papers will play an important role in understanding the mechanisms of the DNA-protein interactions and the expression of genetic information in general.

REFERENCES

Total synthesis of the structural gene for the precursor of a tyrosine suppressor transfer RNA from Escherichia coli. 1. General introduction.


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