BIOORGANIC CHEMISTRY 7, 351–393 (1978)

REVIEW ARTICLE

Studies on Nucleic Acids: Total Synthesis of a Biologically Functional Gene[1,2]

H. GOBIND KHORANA

Alfred P. Sloan Professor of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

INTRODUCTION

The most exciting periods in the development of different fields of science are often characterized by seminal interactions and fusion of ideas originating from seemingly unrelated disciplines. The ideas may have been evolving simultaneously, albeit haphazardly, among different scientific communities, but their convergence, recognition, and synthesis at the opportune time signify great moments. There follow exponential phases of growth and visualization of new and fertile frontiers. The recent history and accomplishments of molecular biology are examples *par excelence* of what is meant. The momentum and vigor, very probably unparalleled in the history of science, which this newly christened field gained in the fifties and sixties, has not shown signs of abating. It is my interpretation that the rise of the tide in the fifties was a result of ideas, concepts, and experimental techniques nurtured by chemists, biochemists, physicists, geneticists, cytologists, virologists, and bacteriologists. It was my good fortune to have been totally consumed by this tide just when the excitement over DNA structure and the birth of molecular biology were on the horizon. With a small group of enthusiastic colleagues, especially Gordon Tener, John Moffatt, Bob Chambers, Michael Smith, Bill Razzell, and Peter Gilham, every one of whom has since made his own mark on science, work in the field of organic phosphates and nucleotides was started. In offering a tribute to Bill Johnson, I thought that instead of focusing in detail on a single problem I would attempt to present a review of some of the main topics which have formed the subject of our research during the past 25 years. It is hoped that such an account would reveal continuity and perspective and be of greater interest to readers of *Bioorganic Chemistry*.

SOME HISTORICAL DEVELOPMENTS OF CHEMICAL INTEREST

(*A*) *Phosphate Esters of Biological Interest*

Ever since the original demonstration in 1905 by Harden and Young (*1*) of the dependence of alcoholic fermentation on the presence of inorganic phosphate,

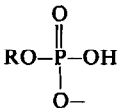
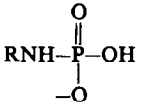
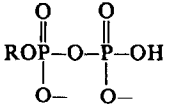
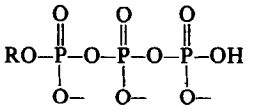
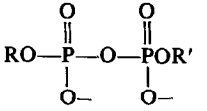
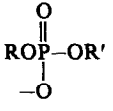
[1] To Bill Johnson, a great scientist, with deep affection and great admiration.

[2] Work in the author's laboratory has been generously supported since 1956 by grants from the National Cancer Institute of the U.S. Public Health Service and since 1960 by grants from the National Science Foundation in Washington, D.C.

0045-2068/78/0073-0351 $02.00/0

TABLE 1

GENERAL CLASSIFICATION OF PHOSPHATE ESTERS OF BIOLOGICAL INTEREST

	Customary organic designation	General structures	Nature of organic substituents	Examples
I.	Monoesters of phosphoric acids		R = sugars, nucleosides, glycerol and glycerol lipids, proteins, unsaturated compounds	Sugar phosphates, nucleotides, phosphatidic acids, phosphoenol pyrurate
II.	Phosphoramidates		R = creatinine, arginine	
III.	Monoesters of pyrophosphoric acid		R = nucleoside–5′-	Ribo- and deoxyribonucleoside 5′-diphosphates, thiamine pyrophosphate
IV.	Monoalkyl triphosphates		R = nucleoside–5′-	Ribo- and deoxyribonucleoside-5′ triphosphates (ATP, dATP, etc.)
V.	Diesters of pyrophosphoric acid		R = nucleoside–5′- R′ = nicotinamide, sugars, choline	Nucleotide coenzymes and group transfer, intermediates in biosynthesis
VI.	Diesters of phosphoric acid		R, R′ = nucleosides, sugars, diglyceride and choline, ethanolamine	RNA and DNA, phospholipids in biomembranes, cell-wall polymers, e.g., teichoic acids

biochemical research has uncovered a bewildering variety of the manifestations and functions of organic phosphate esters in living processes. Indeed, there is hardly anything that goes on in a living cell in which phosphate esters, in one form or another, are not involved at some stage or another. Some of the early pioneers in discoveries of phosphate esters (nucleotide coenzymes, ADP, ATP, Coenzyme A) were Warburg, Meyerhof, Lohmann, Emden, Lipmann, and many others. The progress in the understanding of biochemical and intermediary pathways, biosynthesis of macromolecules, carbohydrates, membrane phospholipids, proteins, and nucleic acids became inseparable from the involvement and discoveries of nucleotides and derivatives. Van Potter correctly described the phenomenon of the early fifties, when the number of known nucleotides rose from under 10 to well over 100 (*2*), as the

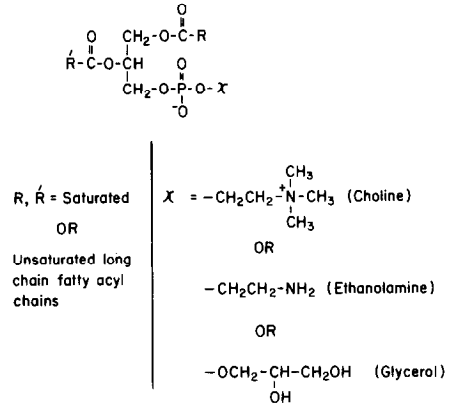


FIG. 1. Examples of phospholipids occurring in biological membranes. 1,2-Di-fatty-acyl-*sn*-glycero-3-phosphoryl choline, ethanolamine, or glycerol.

"nucleotide explosion." An epoch-making discovery made in the late fifties was that of adenosine 3′,5′-cyclic phosphate by Sutherland and co-workers and subsequently of the corresponding analogs of other nucleotides, especially of guanosine cyclic nucleotide.

Although the organic structures associated with the phosphate esters cover an extremely wide spectrum in type and complexity, the categories considered on the basis of substitution on the phosphoric or pyrophosphoric acid molecules are relatively few. An attempt is made here to introduce the large family of phosphate esters of biological interest by classifying them, as shown in Table 1.

An introduction to each one of the above classes was given in an earlier work (*3*). Here, only a few structures are selected for illustration. Figure 1 shows the general structures for a few of the naturally occurring phospholipids which are universal constituents of biological membranes. The general structures for DNA are shown below.

(*B*) *Chemistry of the Nucleic Acids*

After the discovery of the nucleic acids by F. Miescher in 1869, their structural chemistry developed over a period of about 80 years in many countries. This development progressed step by step from the chemistry of the constituent purines and pyrimidines, the sugars, ribose and deoxyribose, to the chemistry of the glycosides

(ribo- and deoxyribonucleosides) of the heterocyclic bases. Some of the early pioneers were Piccard, Kossel, E. Fischer, O. Hammerston, Levene, Jones, and Gulland. By the late forties, the structures of the nucleosides had been largely defined, and the next focus was on the structures of the nucleotides and the internucleotide bonds in nucleic acids.

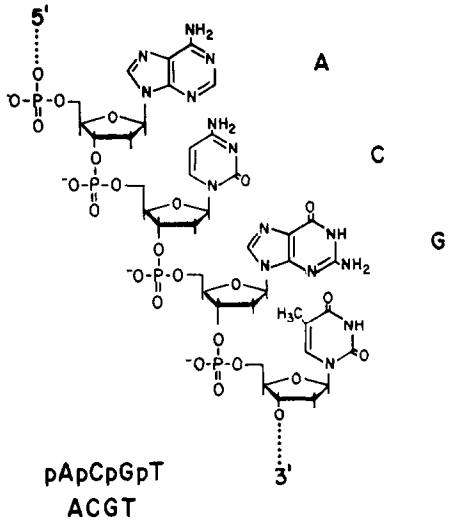


FIG. 2. A hypothetical tetranucleotide sequence in DNA illustrating details of covalent bonds. At bottom left are shown the conventional abbreviations: the letter p at left stands for 5'-phosphate group while the same letter between the nucleoside initials stands for 3' → 5'-linked phosphodiester bond. Further abbreviations may be effected by omitting the letter p altogether. In this case, too, the polarity of the oligonucleotide chain is 5' → 3' from left to right.

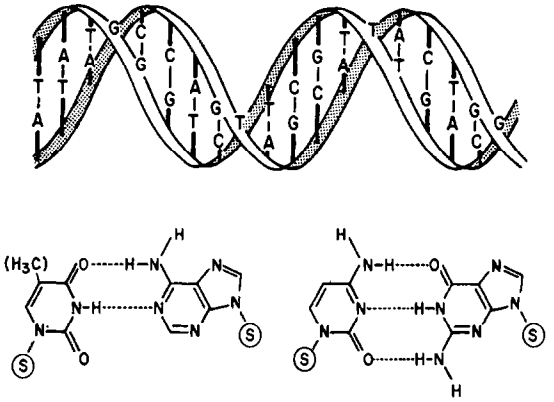


FIG. 3. Watson–Crick structure for DNA. The two base pairs, adenine : thymine and cytosine : guanine, are shown in detail to the left.

Levene and colleagues had already proposed the correct linkage in DNA. However, the RNA structure was complicated by the behavior of the internucleotide bond carrying a neighboring (2') hydroxyl group. A distinct climax was reached in 1952 with the elucidation of the chemical bonds in both RNA and DNA, notably by the efforts of Brown and Todd and their colleagues. A superb and classic account is given in the monograph by Chargaff and Davidson (*4*). The structural details present in DNA are

illustrated with respect to the hypothetical tetranucleotide segment illustrated in Fig. 2. (Note also the inserted conventional abbreviations.)

The Watson–Crick bihelical structure for DNA, which was proposed only about a year later, is shown in Fig. 3.

CHEMICAL SYNTHESIS OF POLYNUCLEOTIDES

(*A*) *Deoxyribopolynucleotides*

At the outset there were the two main divisions in the field: the deoxyribo series and the ribo series. While many of the problems in the two series were inevitably common, attention was first focused on the deoxy series. We began by learning to make thymidylyl–thymidine (TpT) (Fig. 4), the simplest dinucleotide. Thymidine is, relatively

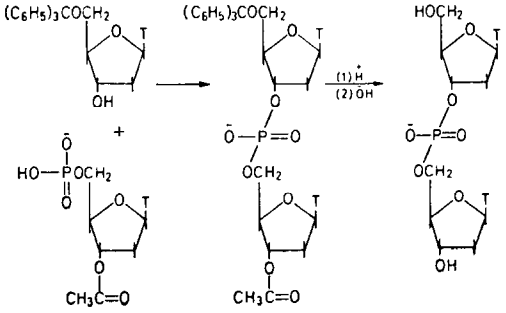


FIG. 4. Condensation of two protected components to form a dinucleotide. T stands for thymine. On top left is the 5′-protected thymidine and in bottom left is the 3′-protected thymidine 5′-phosphate. Condensation to form the internucleotide bond (middle) is effected by activation of the phosphate group. The protecting groups are removed by mild acid and alkali to give the unprotected dinucleotide (right). Mild alkaline treatment alone would free the 3′-OH end group for further chain elongation by condensation with protected mono- or oligonucleotides carrying 5′-phosphate groups.

speaking, the easiest to work with, for it requires no protection on the ring and presents no special solubility problems. In the synthetic procedures illustrated in Fig. 4, three concepts are worthy of note. (i) One of the starting components is a nucleoside with a free 3′-hydroxyl group, the 5′-hydroxyl group being blocked by the bulky trityl group, an acid-sensitive group. (ii) The second component in the condensation is a mononucleotide which has a 3′-hydroxy group blocked by a simple alkali-labile group. (iii) The 5′-phosphomonoester group of a nucleotide is directly activated by a reagent so as to effect condensation with the 3′-OH group of the nucleoside to form an internucleotide bond. This example focuses on the problems of suitable protecting groups in dinucleotide synthesis as well as in the stepwise synthesis of higher oligonucleotides and, second, on the question of activating agents (*5*). Figure 5 gives a summary of the protecting groups, all readily and safely removable, and of the condensing agents which are currently used. All the groups shown in Fig. 5 can be removed with very mildly acidic conditions (methoxytrityl), ammonia (acyl-protecting groups for amino functions), or fluoride ion (silyl groups). A practically important feature is the selective unblocking of the terminal 3′-OH group in order to elongate the

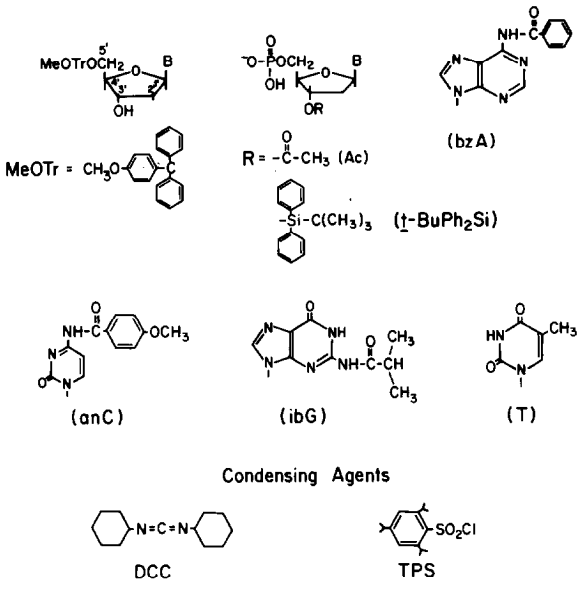


FIG. 5. Groups used for the protection of deoxyribonucleosides and deoxyribonucleotides and the condensing agents used in the chemical synthesis of polynucleotides. The abbreviations used for the different groups are also shown.

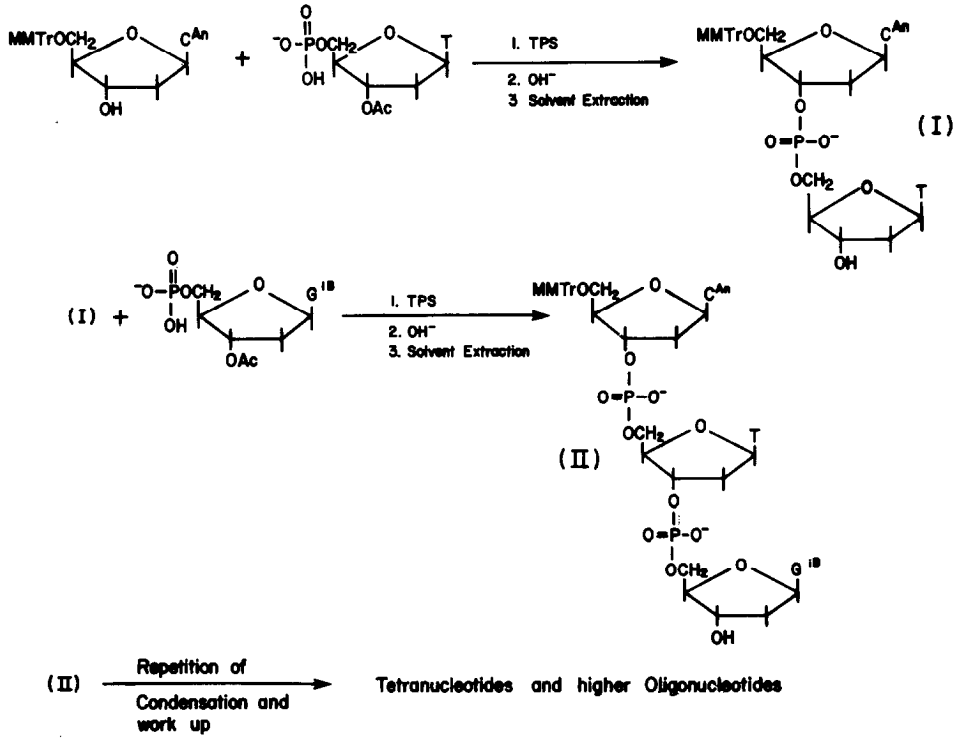


FIG. 6. Principles in the stepwise synthesis of tri- and higher oligonucleotides. See also the legend to Fig. 4.

chain by condensation with a 5′-mononuclotide or olignonucleotide. Thus, the example shown in Fig. 6 illustrates the principle for stepwise synthesis of tri-nucleotides and higher oligonucleotides. In elongation, the incoming units may consist of protected mononucleotide or performed di-, tri-, and tetranucleotide blocks carrying 5′-phosphate groups. An illustration of the steps used in the synthesis of a deoxydodecanucleotide is given in Fig. 7. The accepted shorthand abbreviations are used for the nucleosides and the protecting groups. Unambiguous synthesis requires the separation and characterization of the required products at every step of the synthesis. Currently used procedures, which are much more rapid than the older anion exchange

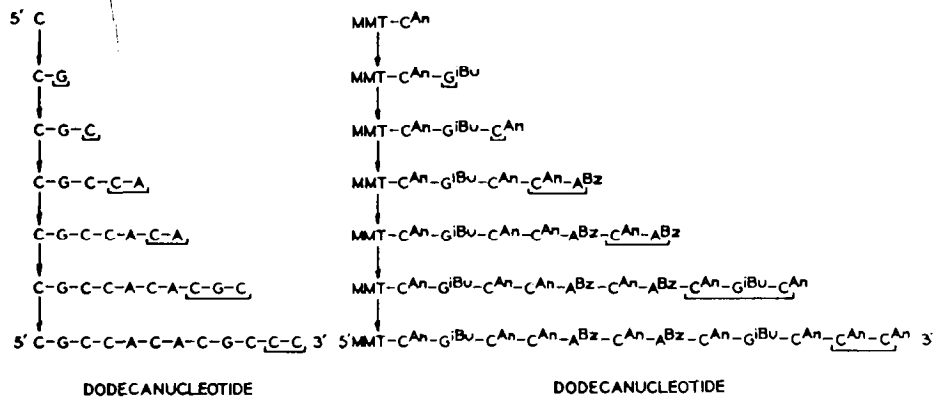


FIG. 7. Steps in the chemical synthesis of a dodecanucleotide. The incoming mono- and oligonucleotides are shown by horizontal brackets under the oligonucleotide blocks. The single letters C, G, and A stand for deoxyribonucleosides. The protecting groups used to protect the amino groups in different bases are shown above the nucleoside initials. The phosphate groups connecting the individual nucleosides are shown simply by hyphens between the nucleoside initials. The 5′-OH end group is at the left and the chain grows in the 3′-direction. The designation of the polarity of the oligonucleotide chains is as described above in the legend to Fig. 2.

chromatography, include extraction methods and high-pressure liquid chromatography (*6*, *7*). Syntheses of the oligonucleotide segments used in the following sections were all performed using the above principles.

(*B*) *Chemical Synthesis of* C_3-, C_5-*Linked Ribooligonucleotides*

The 2′-hydroxyl group in the ribose ring creates an additional formidable problem. Consequently, methods had to be developed for the specific linkage of the 3′-hydroxyl group of one nucleoside to the 5′-hydroxyl group of the next. A brief review of the methods eventually developed is desirable because the synthesis of all the 64 possible trinucleotides derivable from the four mononucleotides greatly aided the elucidation of the genetic code, as described later.

Assuming that, as in the above-discussed syntheses, ribomononucleotides are to be one of the component starting materials in ribooligonucleotide synthesis, the two types of approaches which can be considered are shown in Fig. 8. In the approach on the left a fully protected nucleoside 3′-phosphate is condensed with another protected ribonucleoside carrying a free 5′-hydroxyl group. In the second approach, a protected nucleoside 5′-phosphate is to be condensed with a ribonucleoside (or a nucleotide)

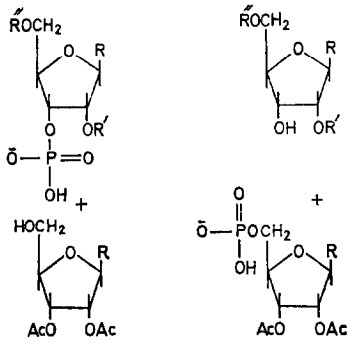


FIG. 8. Two alternative approaches to the synthesis of C_3–C_5 inter-ribonucleotidic linkage (details in text).

which would have only the 3′-hydroxyl group free and all the other functions suitably protected. A systematic investigation of both approaches [see, e.g., Refs. (*8*, *9*)] led to a clear preference for the former approach discussed above.

Two discoveries proved to be decisive in the practical development of our work in this area. The first was the finding that ribonucleoside 3′-phosphates, which, incidentally, can be prepared free from the 2′-isomers in large amounts, could be directly treated with trityl or methoxytrityl halides to form the 5′-*O*-trityl ribonucleoside 3′-phosphates (II) (Fig. 9). The second crucial finding was that the 2′-OH group in these products could be quantitatively acetylated, provided carefully defined conditions were used. In the fifties, it would have been difficult to believe that treatment of a ribonucleoside 3′- (or 2′-) phosphate carrying a *cis*-OH group free on the adjacent position with an acetylating agent would not lead to cyclic phosphate formation and consequent randomization of the phosphate group. It was discovered, fortunately, that when the acetylation reaction is performed under rigidly anhydrous conditions in the presence of an excess of acetate

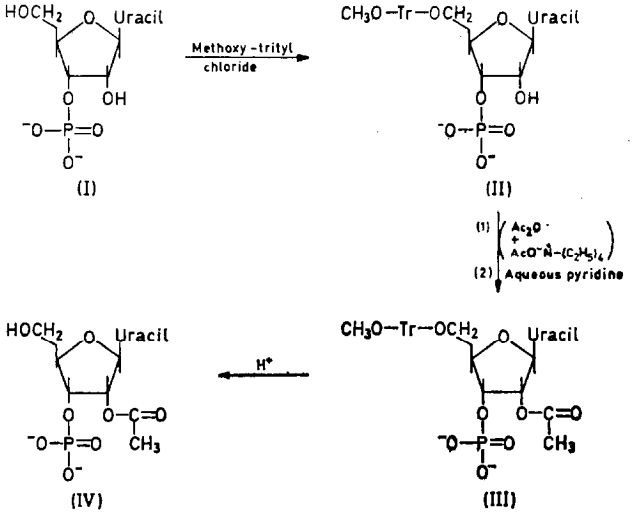


FIG. 9. A direct method for the preparation of 5′- and 2′-protected derivatives of uridine 3′-phosphate from the unprotected ribonucleoside 3′-phosphates.

ions, provided they are in the form of tetraethylammonium acetate, then only the acetylation of the 2′-OH group occurs; this esterification is quantitative (*10*). This reaction thus leads to the preparation of the key intermediates, 5′-*O*-trityl-2′-*O*-acylribonucleoside 3′-phosphates (III in Fig. 9) (*N*-acylation would also occur at this stage, if the amino groups have not already been protected). These protected derivatives on treatment with mild acid give compounds of the type IV (Fig. 9) which are suitable starting materials for polymerization reactions. Alternatively, III may be condensed with protected ribonucleosides carrying free 5′-OH groups (Fig. 10), and at the end the total protecting groups may be removed by successive mildly acidic and ammoniacal treatments, which are safe for the internucleotidic linkage.

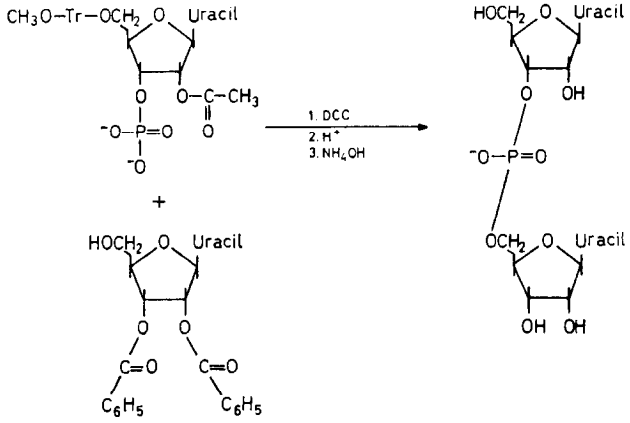


FIG. 10. The synthesis of uridylyl-(3′ → 5′)-uridine. The components used were the protected uridine 3′-phosphate as prepared in Fig. 9 and 2′,3′-di-*O*-benzoyluridine.

It is interesting to observe that the overall strategy for work in the ribonucleotide field became just the opposite of that developed in the deoxy series. Thus, we use the ribonucleoside 3′-phosphates instead of the 5′-phosphates in the deoxy series and therefore the direction of chain elongation is the opposite in the two series. Then, just before a repeat of the condensation step, a 5′-hydroxyl group is uncovered in the ribose series by a mildly acidic treatment, while in the corresponding case of the deoxy series the repetitive steps involve mildly alkaline treatment to expose the 3′-hydroxyl group. All this is exactly as is allowed, respectively, by the RNA and DNA chemistry.

The above approach was successfully used in the unambiguous synthesis of all of the 64 ribotrinucleotides derivable from the four common mononucleotides. These were required for the work on the genetic code, specifically, the assignment of different trinucleotide codons to different amino acids. The synthetic approach in its generalized form is illustrated in Fig. 11. It is readily seen that the synthesis of the 64 possible trinucleotides requires three sets of protected compounds: (i) a set of the four protected ribonucleosides with 5′-OH group free; (ii) a set of the four protected (5′-*O*-methoxytrityl, *N*,2′-*O*-acyl) ribonucleoside-3′-phosphates; and (iii) a set of the four protected (*N*,2′,5′-O-acyl) ribonucleosides 3′-phosphates. Satisfactory methods for the preparation of all of the three sets of compounds were developed. Unambiguous syntheses of all of the possible ribotrinucleotides were realized using these unprotected derivatives. (*11*).

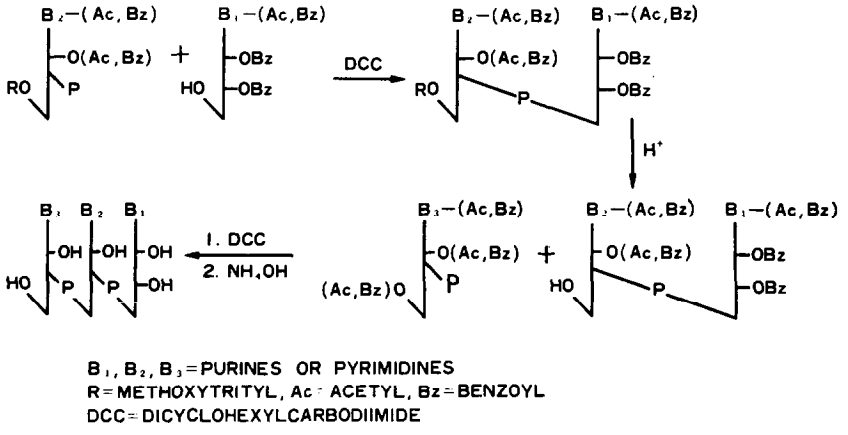


FIG. 11. Generalized method for the synthesis of ribotrinucleotides (details in text).

THE MODE OF ACTION OF PHOSPHODIESTERASES AND NUCLEASES

At the stage in the development of biochemistry of the nucleic acids when the above synthetic methods were being developed, the precise mode of action of the enzymes which degrade nucleic acids was largely unknown. Indeed, many such enzymes were still to be discovered. The availability of a variety of oligonucleotides of defined size, end groups, and internucleotide linkages provided new opportunities for elucidation of the different modes of cleavages of nucleic acids. While a review was presented elsewhere (*12*), two aesthetically pleasing examples are chosen for inclusion here. The snake venom phosphodiesterase was shown to be an *exonuclease*, which began by degrading a polynucleotide chain from the terminus bearing a 3′-OH group. It released successively nucleoside-5′-phosphate units and at every step again produced a new 3′-OH group. A second enzyme, spleen phosphodiesterase, was also shown to be an exonuclease. Its mode of action was just the reverse of the enzyme described above. Thus, this enzyme began at the end bearing a 5′-OH group and successively released nucleoside-3′-phosphate units. The modes described here are illustrated in Fig. 12 with respect to the simplest case of a trinucleoside containing only the two phosphodiester bonds.

The inclusion of a segment of the work on phosphodiesterases and nucleases, which was carried out in the late fifties and early sixties, is made with several thoughts in mind. First, work with these enzymes was our first exposure to enzymology and, in a sense, this meant preparation for the later work with RNA and DNA polymerases, amino acid-activating enzymes, and a number of other crude or purified enzymatic systems. Second, the availability of enzymes with defined modes of action on polynucleotides was to provide highly useful reagents for the characterization of synthetic oligo- and polynucleotides. Classical methods of organic analysis and even the newly introduced physical methods were simply not applicable to the characterization of polynucleotides. Third, the above-described enzymes, especially the venom phosphodiesterase, began to serve as a universal reagent for partial degradation in nucleotide sequence work (*13*). The finger-printing methods, giving maps from which the sequence can be directly read off, require the formation of successively lower homologs, and the venom phosphodiesterase has proved essentially indispensable.

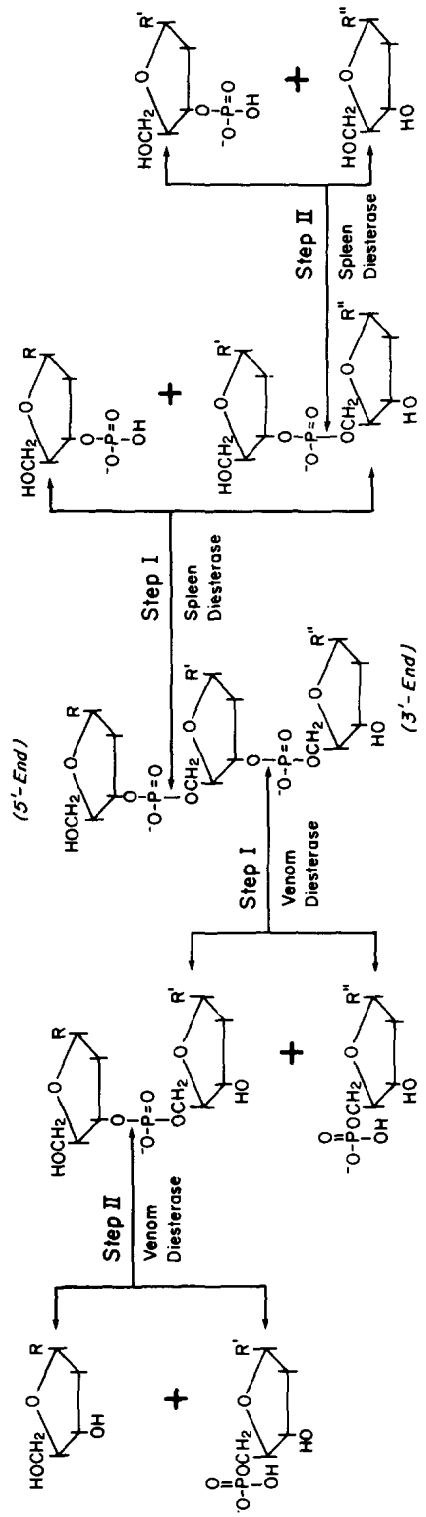


FIG. 12. Illustration of the stepwise enzymatic degradation of an oligonucleotide chain (a trinucleotide) possessing the required 3'-*OH* and 5'-*OH* groups at the termini. Snake venom phosphodiesterase is an exonuclease which attacks a chain at the end bearing a 3'-hydroxyl end and successively removes ribo- or deoxyribonucleoside 5'-phosphate units from that end. Spleen phosphodiesterase is also an exonuclease but its mode of action is complementary to that described above, in that it begins at the end of the chain bearing the 5'-OH end and it successively releases ribo- or deoxyribonucleoside 3'-phosphate units.

POLYNUCLEOTIDE SYNTHESIS AND THE GENETIC CODE

(*A*) *Introduction*

A brief review of some of the major developments, which had origins in different fields, is desirable in tracing the formulation and eventual solution of the problem of the genetic code. The inference that genes make proteins is more than 60 years old, and this was clearly an important idea. However, what put this concept into sharp focus was the one gene–one enzyme hypothesis which Beadle and Tatum proposed in the early forties (*14*) which, I believe, got the field of biochemical genetics going. The next step was taken when it was established that genes are nucleic acids. What indeed is the biological role of nucleic acids? Are they or the proteins the vehicles of heredity? This discussion has a long and troubled history. An excellent and profound analysis of the early and varied evidence supporting the genetic role of nucleic acids was provided by Hotchkiss in 1955 (*15*). Developments which decisively turned the tide were the transformation experiments of Avery and co-workers in 1944 (*16*), followed by the bacteriophage experiment of Hershey and Chase in 1952 (*17*). These experiments established the genetic role for DNA and the work with TMV-RNA a few years later (*18*, *19*) established the same for RNA. By the early fifties it was therefore clear that genes are nucleic acids and that nucleic acids direct protein synthesis.

Clearly the first task was to know more about the chemistry of nucleic acids and the genes; indeed the accelerated pace of development in molecular biology that soon followed was a result of work on the chemistry and biochemistry of nucleic acids. The parallel but completely independent chemical developments mentioned above in the chemistry of nucleic acids [Brown and Todd in Ref. (*4*)] were most remarkable. The latter were followed, again miraculously, by the proposal of the Watson–Crick structure for DNA (*20*). The double-helical organization of DNA focused attention particularly on the biological meaning of its physical structure. It is also about this time that the hypothesis that a linear sequence of nucleotides in DNA specifies the linear sequence of amino acids in proteins was born. A few years later, the enzymology of DNA got into stride with the work of Kornberg and his co-workers (*21*): their discovery and characterization of the enzyme DNA polymerase, now DNA polymerase I, was a major triumph of modern enzymology, and the methods developed distinctly aided a few years later the characterization of DNA-dependent RNA polymerase (*22–26*). The discovery of this enzyme clarified the manner by which information in DNA is transcribed into RNA. The last biochemical landmark to be introduced was the development of a cell-free amino acid-incorporating system. Work on this really began with efforts to understand the biosynthesis of the peptide bond. The subject has a long history, but critical progress began to be made in the early fifties. One thinks, in particular, of the pioneering work of Zamecnik and Hoagland (*27*), of Lipmann (*28*), of Berg (*29*), and in regard to the bacterial system of Watson's laboratory (*30*), and of the important refinement made in 1961 by Matthaei and Nirenberg (*31*). The latter workers made the most revolutionary observation that a simple polynucleotide, polyuridylate, directs the synthesis of polyphenylalanine in the bacterial cell-free amino acid-incorporating system. What this development did in effect was to bypass the question of sequence determination of nucleic acids for the determination of the genetic code.[3]

[3] Fifteen years later, sequencing in the nucleic acid field also experienced a revolution, but in the early sixties sequence analysis appeared a most formidable and refractory problem.

Instead, one began to prepare polynucleotides of, as far as possible, defined composition and to feed these into the protein-synthesizing system. The responsibility for complete elucidation of the genetic code now essentially rested with the chemist. If only a chemist could make nucleic acids of completely defined structure and analyze the proteins specified by it, then one would have a direct correlation of the sequences of the two types of macromolecules. This, then, was the central goal of molecular biology at this time.

As discussed above, we, at this time, could only make short chains of DNA and therefore the scheme we devised to study the coding problem is the one shown in Fig. 13. It was also clear that one would have to study the action of DNA polymerase and of RNA polymerase on the short synthetic oligonucleotides. Early studies with both RNA polymerase and DNA polymerase were very encouraging in that short polynucleotides

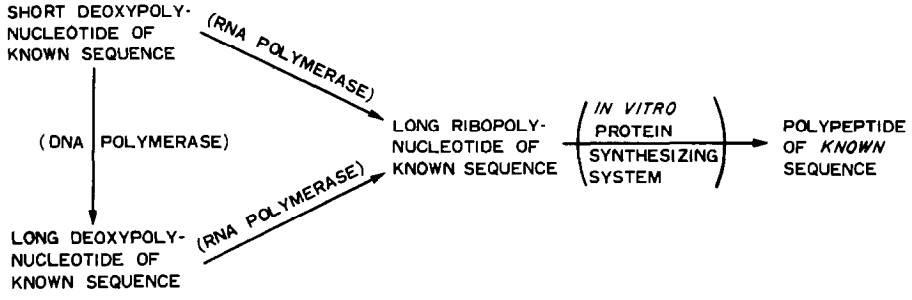


FIG. 13. Sequence of enzymatic reactions for the synthesis of specific polypeptides using chemically synthesized specific deoxyribopolynucleotides. Details are in text.

with repeating sequences did indeed serve as templates for both polymerases and the products contained the expected repeating sequences; but fortunately they had also been greatly amplified in size.

(*B*) *Short Chains of Synthetic Deoxyribopolynucleotides Containing Repeating Nucleotide Sequences*

The above enzymatic results were of great practical value in deciding about the kind of sequences one should start to put together in the deoxy series. There were a number of considerations, for example, the single-stranded requirement for messenger function and the fact that the amino acid-incorporating system contained nucleases. Syntheses were undertaken of polynucleotides with repeating nucleotide sequences because they seemed to offer clear-cut results.

All of the chemical syntheses relevant to the genetic code that were carried out are shown in Table 2. First, we made the two sets of polynucleotides shown on the left, which contained repeating dinucleotide sequences: one set contains the hexamer of the dinucleotide with alternating thymidylate and guanylate residues and the hexamer of the dinucleotide with alternating adenylate and cytidylate residues; the second set consists of the hexamer of alternating thymidylate and guanylate residues and the hexamer of alternating adenylate and guanylate residues (*32*). This work was then extended to polynucleotides with repeating trinucleotide sequences. There is a theoretical maximum of 10 such sets that can contain more than one nucleotide base, and we prepared seven such sets (*33–37*). Shown also in Table 2 are two sets of polymers with repeating

tetranucleotide sequences (*38*, *39*). Two additional considerations for the selection of the nucleotide sequences in them are: (i) they contain the chain-terminating codons in every fourth place and (ii) this class of polymers can be used to prove the direction of reading of the messenger RNA (*40–42*). Two general points about all the synthetic polynucleotides shown in Table 2 may be noted. The first point is that every set comprises two polynucleotides which are complementary in the *antiparallel* Watson–Crick base-pairing sense. A set of repeating trinucleotide polymers, which was complementary in the *parallel* sense, was found to be unacceptable to the DNA polymerase (*43*). The second point is that it *was* necessary to synthesize segments corresponding to both strands of the DNA polymer eventually desired (see below). DNA polymerase failed to bring about polymerization reactions when given only one of the segments of a set as a template.

TABLE 2

SYNTHETIC DEOXYRIBOPOLYNUCLEOTIDES WITH REPEATING SEQUENCES

Repeating dinucleotides		Repeating trinucleotides		
$\begin{bmatrix}(TC)_6\\(AG)_6\end{bmatrix}$	$\begin{bmatrix}(TG)_6\\(AC)_6\end{bmatrix}$	$\begin{bmatrix}(TTC)_4\\(AAG)_4\end{bmatrix}$	$\begin{bmatrix}(CCT)_{3-5}\\(GGA)_{3-5}\end{bmatrix}$	$\begin{bmatrix}(TAC)_{4-6}\\(TAG)_{4-6}\end{bmatrix}$
Repeating tetranucleotides		$\begin{bmatrix}(TTG)_{4-6}\\(CAA)_{4-6}\end{bmatrix}$	$\begin{bmatrix}(CGA)_{4-6}\\(CGT)_{3-5}\end{bmatrix}$	$\begin{bmatrix}(ATC)_{3-5}\\(ATG)_{3-5}\end{bmatrix}$
$\begin{bmatrix}d[TTAC]_4\\d[GTAA]_2\end{bmatrix}$	$\begin{bmatrix}d[TCTA]_3\\d[TAGA]_2\end{bmatrix}$			$\begin{bmatrix}(CCA)_{3-5}\\(GGT)_{3-5}\end{bmatrix}$

(C) *Double-Stranded DNA-like and Single-Stranded RNA-like Polymers with Repeating Nucleotide Sequences*

As expected, the next stage, namely, the work with DNA polymerase, provided a critical advance. As shown in Scheme 1, a mixture of the two short-chain

$$dT_{11} + dA_7 + \begin{bmatrix}dTTP\\dATP\end{bmatrix} \rightarrow \text{poly(dA : dT)} \quad (1)$$

$$d[TG]_6 + d[AC]_6 + \begin{bmatrix}dTTP\\dATP\\dCTP\\dGTP\end{bmatrix} \rightarrow \text{poly(dTG : CA)} \quad (2)$$

$$d[TTC]_4 + d[AAG]_3 + \begin{bmatrix}dTTP\\dATP\\dCTP\\dGTP\end{bmatrix} \rightarrow \text{poly(dTTC : GAA)} \quad (3)$$

$$d[TATC]_3 + d[TAGA]_2 + \begin{bmatrix}dTTP\\dATP\\dCTP\\dGTP\end{bmatrix} \rightarrow \text{poly(dTATC : GATA)} \quad (4)$$

Scheme 1. Types of reactions catalyzed by DNA polymerase.[4] All of the DNA-like polymers are written so that the colon separates the two complementary strands. The complementary sequences in the individual strands are written so that antiparallel base pairing is evident.

polynucleotides with repeating dinucleotide sequences directed the extensive synthesis of a double-stranded DNA-like polymer (Scheme 1) containing exactly the sequences present in the short-chain templates (*44–48*). The same was true for short chains with repeating trinucleotide sequences and later also for repeating tetranucleotide sequences (*49*, *50*). Many of the features of these reactions are truly remarkable. Thus, (i) in all these reactions (Scheme 1) the enzyme showed complete fidelity in reproduction of sequences; (ii) the synthesis was extensive, 50 to 200-fold, and the products were of high molecular weight (500,000 to over 1,000,000); (iii) the enzyme thus amplified and multiplied the information created by chemical methods;[4] (iv) finally, from the standpoint of an organic chemist, the most satisfying aspect is that the DNA polymers thus made can be used for further synthesis. We never have to go back to time-consuming chemical synthesis for obtaining that particular sequence again. DNA polymerase assures the continuity of these sequences.

TABLE 3

DNA-LIKE POLYMERS WITH REPEATING NUCLEOTIDE SEQUENCES

Repeating dinucleotide sequences	Repeating trinucleotide sequences	Repeating tetranucleotide sequences
Poly(d-TC : GA)	Poly(d-TTC : GAA)	Poly(d-TTAC : GTAA)
Poly(d-TG : CA)	Poly(d-TTG : CAA)	Poly(d-TATC : GATA)
	Poly(d-TAC : GTA)	
	Poly(d-ATC : GAT)	

Table 3 catalogs the different kinds of polymers that were prepared in this way and characterized. Thus there were three classes of polymers: two double-stranded polymers with repeating dinucleotide sequences and four polymers with repeating tetranucleotide sequences.

The availability of a variety of DNA-like polymers with completely defined nucleotide sequences made possible an attack on many aspects of the studies of the biological reaction sequence DNA → RNA → protein. There ensued a highly fruitful and exciting phase of our research. Thus, by relatively straightforward biochemical methods and minor innovations, we were able to prepare a large variety of ribopolynucleotide messengers by using the transcribing enzyme, RNA polymerase (*51*). Thus, because all the DNA-like polymers contain two, or a maximum of three, different bases in individual strands, it was possible, by giving the nucleoside triphosphates required for copying only one strand, to restrict the action of RNA polymerase to that strand. In this way, each one of the DNA-like polymers described yielded two RNA messengers, again with the expected repeating sequences. The total RNA-like messengers thus prepared are shown in Table 4.

In summary, by using a combination of purely chemical methods, required to

[4] In retrospect, it is remarkable how well the DNA polymerase I worked in the amplification and multiplication of the synthetic polymers. The enzyme preparations then made must have had the "right" amounts of nucleases in them. As purification was taken further by Kornberg and co-workers, no net synthesis but only repair synthesis was observed.

TABLE 4

SYNTHETIC RIBOPOLYNUCLEOTIDES WITH REPEATING NUCLEOTIDE SEQUENCES

Repeating dinucleotide sequences	Repeating trinucleotide sequences	Repeating tetranucleotide sequences
Poly(rUG)	Poly(rUAC)	Poly(rUAAG)
Poly(rAC)	Poly(rGUA)	Poly(rUAGA)
Poly(rUC)	Poly(rAUC)	Poly(rUCUA)
Poly(rAG)	Poly(rGAU)	Poly(rUUAC)
	Poly(rUUG)	
	Poly(rCAA)	
	Poly(rUUC)	
	Poly(rGAA)	

produce new and specified information, and then following through with the two enzymes, DNA polymerase and RNA polymerase, which are beautifully precise copying machines, we had at our disposal a variety of high molecular weight ribopolynucleotides of known sequences.

(D) Polypeptide Synthesis In Vitro and the Genetic Code: Cell-Free Polypeptide Synthesis Using Polynucleotides with Repeating Sequences

Although far from being proven, the genetic code had been suspected in the early sixties to be nonoverlapping with a trinucleotide sequence standing for each amino acid. If this were to be so, polymers with repeating dinucleotide sequences (AB) which contain two triplets, ABA and BAB, in alternating sequence should direct incorporations of two amino acids in strictly alternating sequence. Repeating trinucleotide polymers $(ABC)_n$ contain three repeating triplets depending upon the starting point. These are ABC, BCA, and CAB. Here one would predict that one amino acid should be incorporated at a time to form a homopolypeptide chain, and a maximum of three such chains should result. Similar considerations for polynucleotides with repeating tetranucleotide sequences $(ABCD)_n$ show that *in vitro* polypeptide synthesis should give

TABLE 5

SYNTHESES OF POLYPEPTIDES BY USING THE CELL-FREE *E. Coli B* SYSTEM, AND SYNTHETIC MESSENGERS CONTAINING REPEATING NUCLEOTIDE SEQUENCES

With repeating dinucleotide as messenger		With repeating trinucleotide as messenger		With repeating trinucleotide as messenger	
Messenger	Polypeptide formed	Messenger	Polypeptide formed	Messenger	Polypeptide formed
Poly(UC)	$[Ser–Leu]_n$	Poly(UUC)	$[Phe]_n,[Ser]_n,[Leu]_n$	Poly(UAUC)	$[Tyr–Leu–Ser–Ile]_n$
Poly(AG)	$[Arg–Glu]_n$	Poly(AAG)	$[Lys]_n,[Glu]_n,[Arg]_n$	Poly(UUAC)	$[Leu–Leu–Thr–Tyr]_n$
Poly(UG)	$[Val–Cys]_n$	Poly(UUG)	$[Cys]_n,[Leu]_n,[Val]_n$	poly(GUAA)	Di- and tripeptides
Poly(AC)	$[Thr–His]_n$	Poly(CAA)	$[Gln]_n,[Thr]_n,[Asn]_n$	Poly(AUAG)	Di- and tripeptides
		Poly(GUA)	$[Val]_n,[Ser]_n$,("nonsense")		
		Poly(UAC)	$[Tyr]_n,[Thr]_n,[Leu]_n$		
		Poly(AUC)	$[Ile]_n,[Ser]_n,[His]_n$		
		Poly(GAU)	$[Met]_n,[Asp]_n$,("nonsense")		

products containing repeating tetrapeptide sequences, irrespective of the starting point in the reading of the messengers. All these predictions were fully borne out experimentally without a single exception. The results obtained with the three classes of polynucleotide messengers are listed in Table 5. These results led to the following general conclusions: (i) DNA does, in fact, specify the sequence of amino acids in proteins, and this information is relayed through an RNA. (This was the first time that a direct

TABLE 6

THE GENETIC CODE[a]

1st↓ \ 2nd →	U	C	A	G	↓3rd
U	Phe	Ser	Tyr	Cys	U
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Term.	Term.	A
	Leu	Ser	Term.	Trp	G
C	Leu	Pro	His	Arg	U
	Leu	Pro	His	Arg	C
	Leu	Pro	Glum	Arg	A
	Leu	Pro	Glum	Arg	G
A	Ileu	Thr	Aspn	Ser	U
	Ileu	Thr	Aspn	Ser	C
	Ileu	Thr	Lys	Arg	A
	Met (Init.)	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	U
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (Init.)	Ala	Glu	Gly	G

[a] The abbreviations for amino acids are standard. C.T. stands for chain termination, i.e., the trinucleotide sequence does not stand for any amino acid but probably signals the end of protein chain formation. C.I. stands as a signal for chain initiation in protein synthesis. The method of presentation used follows the conventional way of writing of trinucleotides; thus, the first letter (base) of the trinucleotide is on the left (the 5′-end) and the third letter (the 3′-end) is to the right of the middle (second) base. The use of the table for derivation of codons for different amino acids is exemplified as follows: codons for the amino acid, Phe, are UUU and UUC; codons for the amino acid, Ala, are GCU, GCC, GCA, and GCG.

sequence correlation between DNA and a protein had been established.) (ii) All the results prove the three-letter and nonoverlapping properties of the code. (iii) Finally, information on codon assignments could also be derived from these results.

In further work, by a combination of the above approach and the ribosome–tRNA–trinucleotide binding technique developed by Nirenberg and Leder (*52*) as well as the extensive genetic mutation work (*53*), it proved possible to deduce the total structure of the genetic code as shown in Table 6. Only a few general observations may be made. (i) The code as shown is for the microorganism *Escherichia coli* B, but probably will hold essentially for other organisms as well, although detailed and systematic checking in

other systems (plant and animal) remains to be carried out. (ii) There are entries for all of the 64 trinucleotides (there is no absolute nonsense). The code is highly degenerate in a semisystematic way. Most of the degeneracy pertains to the third letter, where all of the four bases may stand for the same amino acid, or where the two purine bases may stand for one amino acid and the two pyrimidines may stand for another amino acid. An exception is the box with the first letter A and the second letter U. Here, AUU, AUC, and AUA represent isoleucine whereas the fourth codon, AUG, stands for methionine. Three amino acids show additional degeneracy in positions other than the third letter: thus, leucine and arginine are degenerate in the first letter, whereas serine is unique in changing its position with regard to both the first and second letters. (iii) While the code is now generally accepted to be essentially universal, it should not be inferred that all organisms use the same codons for protein synthesis. What the universality means is that a trinucleotide codon does not change its meaning from one organism to the next. After all, there is very great divergence in the DNA compositions of diverse organisms and they therefore probably use different codons for the same amino acid to varying extents. (iv) The codons AUG and GUG, which stand, respectively, for methionine and valine, are also used as signals for initiation of polypeptide chain synthesis (see also a later section for initiation of protein synthesis). (v) There are three trinucleotides, UAA, UAG, and UGA, which cause termination of polypeptide chain growth. (vi) Finally, all aspects, including assignments, initiation, and termination codons have since been confirmed in more recent years directly by total sequence determination of genes and inter-genic spaces.

In addition to the studies bearing directly on the genetic code, defined synthetic polynucleotides provided new opportunities for defined work on many related problems of protein synthesis such as codon–anticodon interactions, nature of genetic suppression, misreading of the genetic code as induced by various antibiotics, direct translation of single-stranded DNA, and initiation and termination of protein synthesis (*54*). Finally, there was the large complementary problem of the "adapter" molecules, the transfer RNAs (see below), which contained the anticodons corresponding to each amino acid. The first problem was one of nucleotide sequence determination, which was solved first for the major yeast alanine tRNA by Holley and co-workers (*55*). The synthesis of the gene specifying this tRNA was soon to become the next major interest of our research.

TOTAL SYNTHESIS OF TRANSFER RNA GENES

(*A*) *Introduction*

In 1966–1967, when the work on the problems of the genetic code was reaching a satisfactory conclusion, the question was that of further orientation of synthetic work in the nucleic acid field. While the high molecular weight DNA-like polymers with repeating nucleotide sequences continued to be used for a variety of physicochemical and enzymatic studies of nucleic acids, it was clear that their use was restricted for many of the outstanding questions in molecular biology. These questions, broadly speaking, are concerned with the problem of DNA recognition and DNA punctuation. We would like to known, for example, what turns genes on and off and, as part of this

question, what the initiation and termination signals for RNA polymerase[5] are and what kind of sequences are recognized by repressors, what determines the specificity of the host modification and host restriction enzymes and of enzymes involved in genetic recombination, and so on. Eventually what would be required for these studies is the ability to synthesize long chains of bihelical deoxyribonucleic acids with specific nonrepeating sequences. With this should come the ability to "manipulate" DNA for different types of studies. We therefore concluded that, in continuing our interest in polynucleotide synthesis, the next long-range aim should be the development of methods for the total synthesis of biologically specific DNA duplexes. It is emphasized that the objective of a total synthesis would be to put together all of the information, i.e., both strands, by chemical synthesis. Eventually our ability to manipulate the information content of nucleic acids depends upon this. It is also of personal interest to me to recall that the "distant" hope of the synthesis of biologically functional nucleic acids was expressed as early as the late fifties (*56*).

As a specific objective, the decision we made was to start work on the total synthesis of the double-stranded DNA corresponding in sequence to the entire length of a transfer RNA. The choice of the gene for a transfer RNA followed from a variety of considerations. First, one must know the primary structure, and, second, we knew that the tRNAs are specified by the DNA genome. The general functions of tRNA are clearly established. These molecules have to be recognized by a rather large number of components of the protein-synthesizing machinery such as the pyrophosphorylase which repairs the CCA end, the aminoacyl-tRNA synthetases, ribosomes, and messenger RNA. Indeed, biological specificity in protein synthesis is determined at the step of loading of the amino acids onto the tRNAs. Furthermore, tRNAs are unique in possessing attributes of both proteins and nucleic acids. Indeed, X-ray analyses of single crystals have since been performed and three-dimensional structures which have been defined (*57*) are most remarkable. Then, all tRNAs abound in minor bases which are largely found in the looped-out nonhydrogen-bonded regions. It is entirely possible that a good part of the evolution of the genetic code is synonymous with the evolution of tRNA molecules. The total area of the structure–function relationships in these molecules is an open field despite the great current research activity in this field. It was clear that chemical synthesis, provided it could be developed to the point where one can manipulate different parts of the tRNA structures, would open up a definitive approach of wide scope.

(*B*) *Strategy for the Controlled Synthesis of Bihelical DNA*

How is the job of synthesizing a long DNA to be approached? However efficient organic synthesis might become, it is difficult to imagine that nucleic acid syntheses of the future would be done entirely by chemical methods alone, as in fact was proved in our work on the genetic code. It is clear that new concepts would have to be introduced. The central idea that we decided to exploit is the template principle, that is, the ability of polynucleotides to form hydrogen-bonded bihelical structures. We would make short overlapping pieces of DNA which would correspond to segments of a double-stranded

[5] Remarkable progress has been made during the last decade and continues to be made in these areas (see following text).

DNA structure and, when these are properly annealed to form bihelical complexes, their end-to-end joining would be attempted either by chemical or enzymatic methods.

The discoveries of the enzymes polynucleotide ligase (*58*) as well as that of polynucleotide kinase (*59*) in the late sixties proved of crucial value in these studies, and fortunately early experiments on the stabilities of short synthetic double-stranded complexes with overlapping sequences and on the utilization of such duplexes by polynucleotide ligase for joining reactions were encouraging in that deoxyoligonucleotide segments within the practical range of organic synthesis were adequate (*60*). These results led to the three-phase strategy shown in Fig. 14 for the total synthesis of a bihelical DNA. Thus, the first phase requires the chemical synthesis of short polynucleotides of chain lengths in the range of 10 to 12 nucleotides; these would

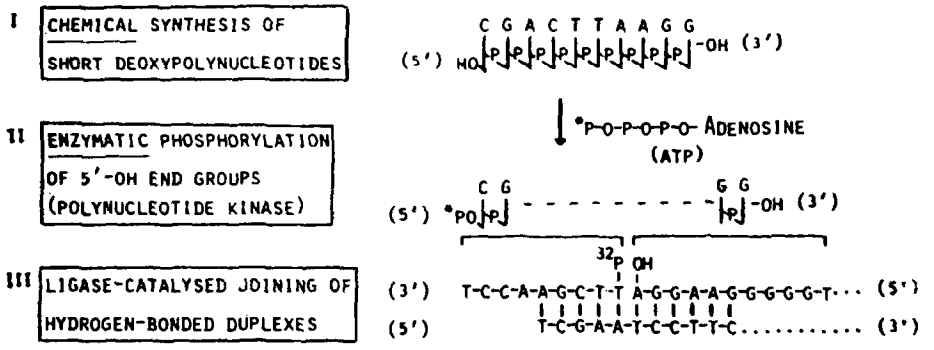


FIG. 14. Three-stop strategy for the total synthesis of double-stranded DNA. Details are in text.

correspond to the entire two strands. A hypothetical sequence showing the 3'- and 5'-OH ends is shown at top right in the figure using the Fischer projection method. In the second step, the 5'-OH ends of the segments, except those which are to be at the 5'-termini of the duplexes, are phosphorylated by using γ-^{32}P-labeled ATP and the polynucleotide kinase. This enzyme is admirably suited for the purpose because of its absolute specificity for phosphorylating the 5'-OH end group of a short or a long polynucleotide chain without any regard to the sequence.

Since the main task in the next phase (phase III) is the accurate and specific end-to-end joining of all of the synthetic segments, the presence of radioactive phosphate groups at 5'-ends of the synthetic segments, as introduced above, is crucial. Three, four, or as many as six or seven segments with overlapping complementary sequences are brought together under suitable conditions of ionic strength and temperature and in the presence of divalent ions, and the polynucleotide ligase is then used to bring about covalent joinings to form covalently linked duplexes. The joinings can be monitored by analyses of the joined products in a variety of ways. For example, the double-stranded product and the individual strands after separation may be sized quite precisely by electrophoresis in polyacrylamide gels. Furthermore, they may be degraded to 5'- or 3'-mononucleotides, and the distribution of radioactivity in different mononucleotides will immediately show the accuracy in joining. Characterization is further aided by manipulation of the specific activity of [γ-^{32}P]ATP used in phosphorylation reactions. Thus, specific activity at the internal linkages in the short duplexes formed first (see

below) may be much weaker than that at the termini which are to connect the short duplexes to form the large ultimate DNA. Different joinings may also be further distinguished by concomitant use of ^{32}P and ^{33}P isotopes. Indeed, both ^{32}P and ^{33}P-labeled phosphoryl groups may be used at two levels of specific activity at different sites, allowing completely error-free characterization.

(C) Total Synthesis of the Structural Gene for the Yeast Alanine Transfer RNA

This tRNA, whose primary structure is shown in Fig. 15, was the first one whose sequence was determined (*61*). Knowledge of the primary sequence obviously was decisive in the synthetic work. Early ligase and methodology experiments were, in fact,

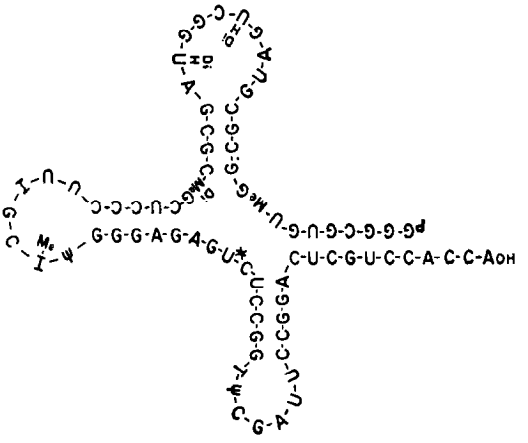


FIG. 15. Structure of the major yeast alanine transfer RNA as deduced by Holley and his co-workers. The structure is shown in the familiar cloverleaf model for the secondary structure of the tRNA. The abbreviations U, C, A, and G stand, respectively, for the nucleotides of uracil, cytosine, adenine, and guanine. ψ stands for pseudouridine, T for ribothymidine, I for inosine, MeI for methylinosine, and DiH-U for 5,6-dihydrouridine.

all performed using synthetic segments corresponding to parts of the gene for this tRNA (*60*). The plan for the total synthesis which finally emerged is shown in Fig. 16. [For detailed validation of the considerations leading to this plan, see Reference (*62*).] Thus, the gene was divided into three parts shown as A, B, and C (or C′), and each part was to consist of several chemically synthesized fragments. The segments are indicated in Fig. 16 by brackets, the serial number being inserted into the bracket. Detailed procedures for the joining reactions and for the construction and characterization of the total duplex have been given elsewhere (*62*) and will not be reviewed here, because related joinings are to be described later in this article.

The above task was completed by the summer of 1970. At the chemical level, there was considerable satisfaction and encouragement that the synthesis of the genetic material and specific bihelical DNAs was well on the way. The strategy for overall synthesis in the macromolecular size range, the techniques and procedures for isolation and characterization of short and medium size duplexes, and, finally, the total duplexes had been satisfactorily worked out. Chemical synthesis had found new frontiers in tackling biological problems. However, for the immediate future of taking the alanine

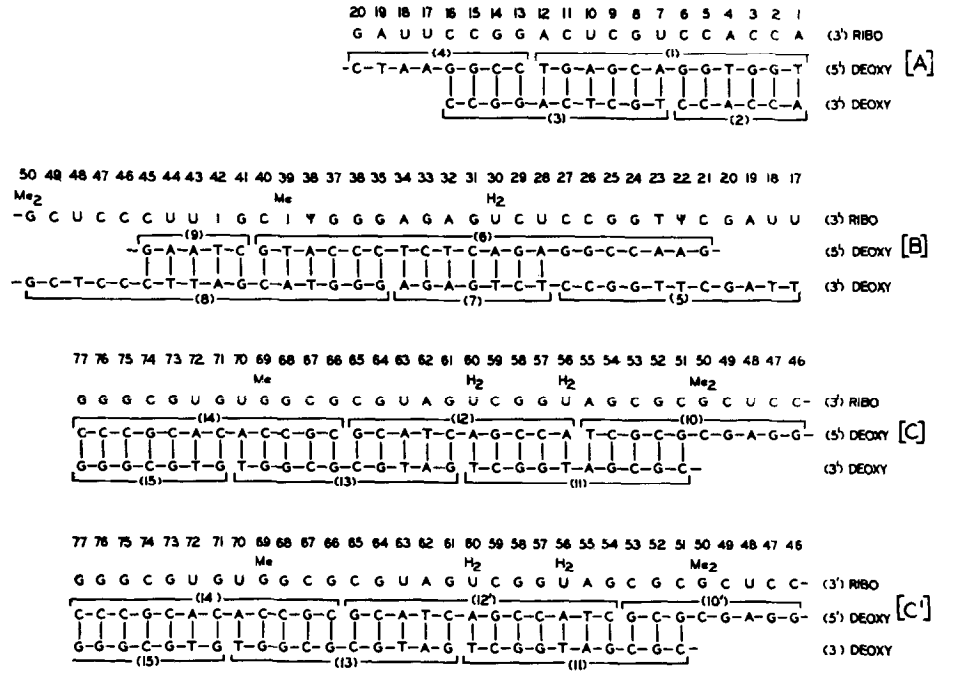


FIG. 16. Total plan for the synthesis of the yeast alanine tRNA gene. The chemically synthesized segments are shown by horizontal brackets, the serial number of the segment being in parentheses inserted into brackets. A total of 17 segments (including 10′ and 12′) varying in chain length from penta- to icosanucleotides was synthesized.

tRNA gene further for biochemical work, the yeast system presented formidable hurdles. Indeed, concurrently with the above synthesis, the bacterial tRNA gene, that for tyrosine tRNA suppressor, was selected as the next target.

(*D*) *Choice of the* E. coli *Tyrosine Suppressor tRNA Gene*

The main arguments, which fortunately became even more compelling as we proceeded with this project, for this choice were as follows: (i) firm knowledge of the primary nucleotide sequence coming from at least two groups of workers (*63*); (ii) the marked advantages which this tRNA offered in studying the biochemistry of the amino acid-charging reaction; and (iii) the dramatic progress in the biochemistry of *in vitro* protein synthesis in the cell-free *E. coli* system and in study of *in vitro* suppression of amber mutation using suppressor tRNAs; (iv) the insertion of the gene into the transducing bacteriophage ϕ80 by Brenner and Smith and co-workers (*64*) and the convenience of working with $\phi 80psu^{+}_{III}$ and related derivatives for probing the nucleotide sequences adjoining the tyrosine tRNA structural gene; (v) extensive work on the structure–function relationships in the same gene by the same group (*65*) using the genetic approach, and, finally, (vi) the discovery of a precursor for this tRNA (*66*) (Fig. 17). This structure containing a 5′-triphosphate end group unambiguously defined the site of initiation of transcription, and therefore attention could be focused precisely on the promoter (preinitiation) region and on the process of transcription of this gene. In fact, one could envisage a linear arrangement for the different regions and control

elements of the gene, as shown in Fig. 18. The region of the structural gene, 126 base-pairs long, that corresponds to the precursor RNA is shown (Fig. 17). The primary RNA product is then cleaved at the point shown to give the tRNA length. The starting point of transcription must be at the point shown by the arrow, and the DNA that precedes it must by definition be the promoter region, the region which is recognized by the RNA polymerase. Transcription begins at the indicated point and continues, presumably, until there is a termination signal. Until recently the view was held that the

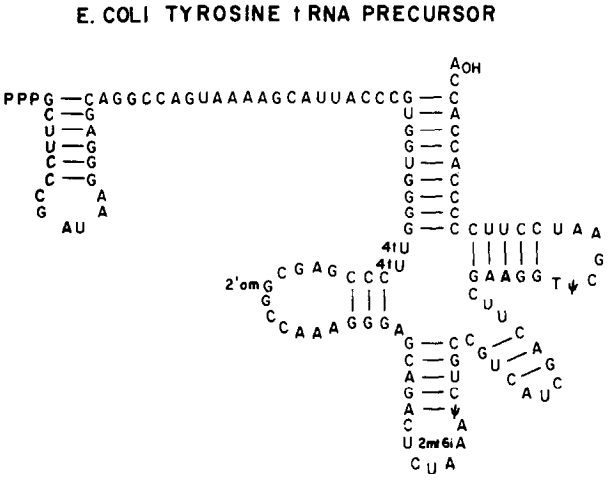


FIG. 17. The primary nucleotide sequence of an *E. coli* tyrosine tRNA precursor. The sequence is written using the standard cloverleaf structure for the tRNA portion and the possible hairpin at the 5′-end.

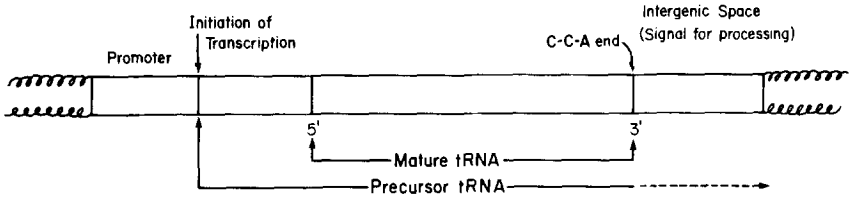


FIG. 18. Linear arrangement of the different parts of the tyrosine suppressor tRNA gene. The gene, which is one of the two duplicate genes in tandem, is located at 25 min on the *E. coli* genetic map. The suppressor gene arises by a single nucleotide change in the anticodon of one of the two tyrosine genes in tandem. Details are in the text.

signal for termination in the tyrosine tRNA gene was probably soon after the C-C-A end, but it has now been shown that the termination of transcription in fact occurs about 225 nucleotides downstream from the structural gene (*67*). As shown later, a cleavage of the primary transcript soon after the CCA sequence occurs as a first event in the processing and maturation of the tRNA.

(*E*) *Total Synthesis of the 126 Base-pair Long DNA Corresponding to the Tyrosine tRNA Precursor* (*Fig. 17*)

The plan adopted for the synthesis of this DNA consisted of a total of 26 chemically synthesized deoxypolynucleotide segments (Fig. 19). How was this plan deduced?

Obviously there are an enormously large number of possibilities for dividing a DNA of this size. The main rules that were followed arise from the demands of chemical and enzymatic work. Thus, there is, of course, the basic requirement of the repetitive overlap of about five base pairs between the adjacent segments. Second, it is important from the standpoint of enzymatic work that partial or total complementarity (self-structure) within the single-stranded segments as well as self-complementarity at 5′-protruding ends be avoided as far as possible (*68*). Another important consideration is economy in the laborious chemical work. Multiple use of synthetic oligonucleotide blocks, short or

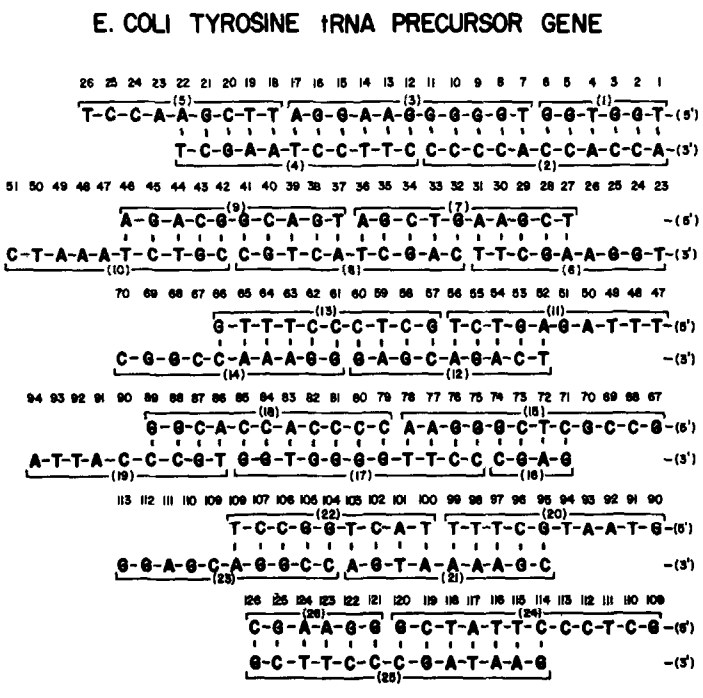


FIG. 19. Plan for the total synthesis of the structural gene for the tyrosine tRNA. Considerations leading to this particular mode of segmentation of the total 126 base-pair-long DNA are given in text.

long, should clearly be maximized. Thus, in the plan adopted (Fig. 19) a systematic search (computer program) revealed, for example, that the nonanucleotide sequence C-C-C-C-A-C-C-A-C occurs twice (segments 2 and 18), the octanucleotide sequence, G-C-T-C-C-C-T-T, occurs twice (segments 13 and 24), and the hexanucleotide sequence, T-T-C-G-A-A (self-complementary) occurs four times. Multiple occurrences of this kind can be found for parts of a large number of other segments. Finally, at least some synthetic intermediates or complete segments were available from previous work (*62*).

When syntheses of all of the required segments had been accomplished, subgrouping with a view to optimal and error-free enzymatic joinings became the next objective. Inevitably, there would be a large number of alternatives, but the formation of perfectly ordered double-helical complexes would be expected to be favored, resulting in quantitative joinings to form covalently linked duplexes. Paradoxically, the extent of

joinings in different systems has varied a great deal, the yields being usually less or much less than quantitative. The protocols for optimal joinings, the rates of reactions, and the yields of the required duplexes from single-stranded oligonucleotides vary considerably (*62*, *69*). Therefore, ultimate choices in grouping of segments continue to require a large amount of empirical work. A further desirable feature to be borne in mind while grouping the single-stranded segments is that, as far as possible, the termini bearing 5′-OH groups should be protruding. Phosphorylations of the termini by the polynucleotide kinase for the subsequent purpose of joining the duplexes to form larger duplexes go to completion extremely rapidly and with complete certainty when this is the

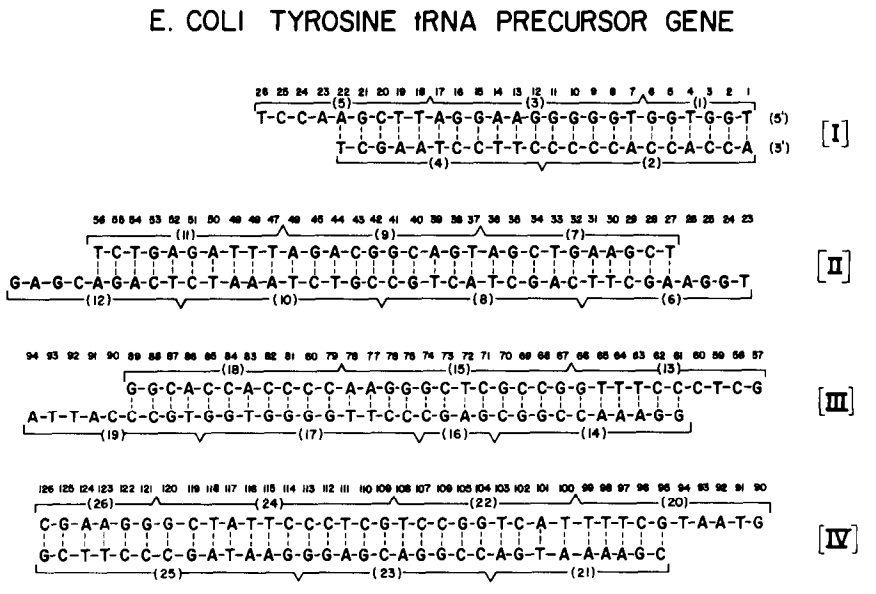


FIG. 20. Plan for the enzymatic joining of the segments in the total synthesis of the gene for the precursor of tyrosine tRNA. Grouping of the total chemically synthesized segments into four duplexes for the purpose of enzymatic joining is shown.

case. The four groups which were finally arrived at in the present work are shown in Fig. 20, duplex [I] to duplex [IV]. Thus, each duplex consisted of five to seven chemically synthesized segments. While the enzymatic joinings to form each one of these duplexes had unique characteristics, only one example, that of the synthesis of duplex [I], will suffice to illustrate the procedures (Fig. 21; further details in Figure legend).

In the last phase, the duplexes representing different parts of the total sequence are joined through their protruding single-stranded ends to form the total duplex. Fortunately joinings at this stage are very efficient and rapid, for which two illustrations may be given. One (Fig. 22) is the joining of duplex [I] to duplex [II]. Both duplexes carried distinctive radioactive labeling patterns to facilitate analysis. The second is the last step joining of the two halves, i.e., joining duplex [I + II] to the duplex [III + IV] (Fig. 23). Finally, Fig. 24 shows the total synthetic duplex corresponding to the tRNA precursor and depicts the synthetic, chemical, and enzymatic steps used.

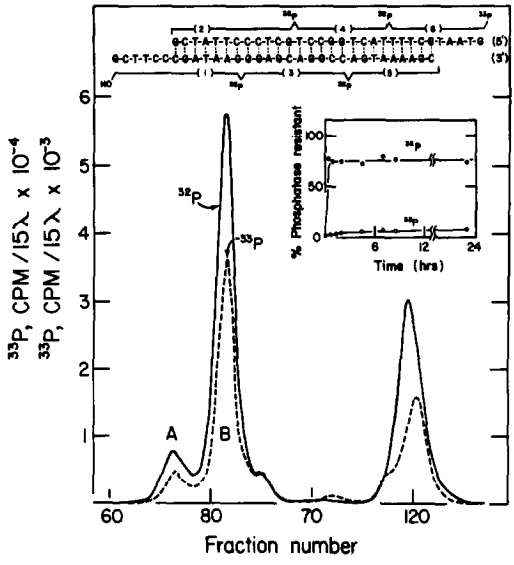


FIG. 21. The synthesis and purification of duplex [I]. The radioactive (^{32}P or ^{33}P) labeling pattern of the 5′-ends in different segments is shown in the inset, as are the kinetics of joining. The latter were followed by the formation of phosphatase-resistant radioactivity. Separation of the products was accomplished on a Bio-Gel (0.5-m) column. Peak B contained the pure duplex [I]. Peak A contained a dimer, the structure of which has been discussed elsewhere (*81*). Unreacted segments appeared last.

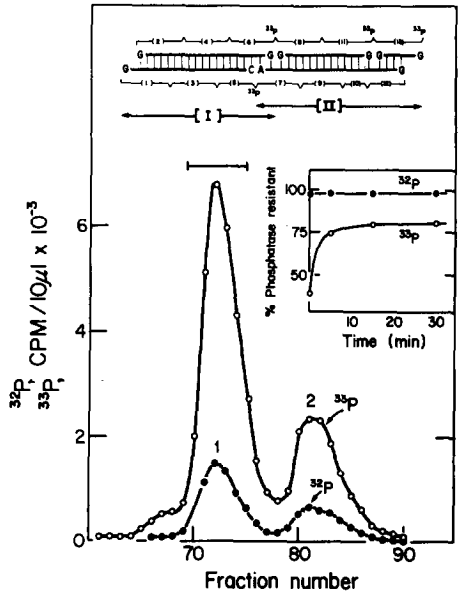


FIG. 22. Joining of duplex [I] to duplex [II]. The ^{33}P radioactive labeling sites are shown in the duplexes. At the remaining sites of enzymatic joining, there was weak ^{32}P radioactivity. The kinetics of joining are shown in the inset. Separation was on a Bio-Gel (0.5-m) column. Peak I contained the joined product. Unjoined products were in Peak 2. Further details are in Ref. (*82*).

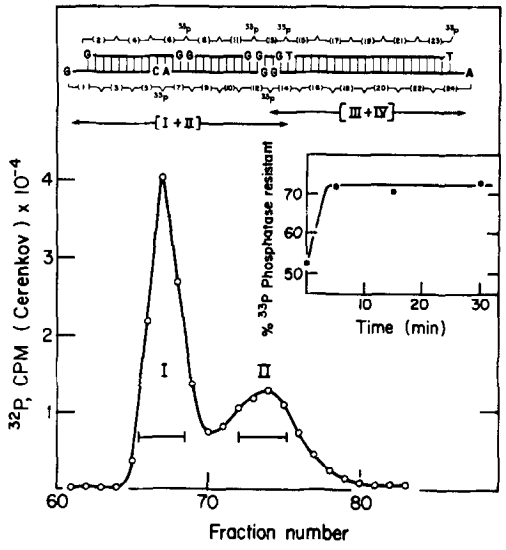


FIG. 23. The joining of duplex [I + II] to duplex [III + IV] to form the total duplex [I + II + III + IV]. The kinetics of joining are shown in the inset. Separation by flow through an Agarose column is shown. For further details, see Ref. (*82*).

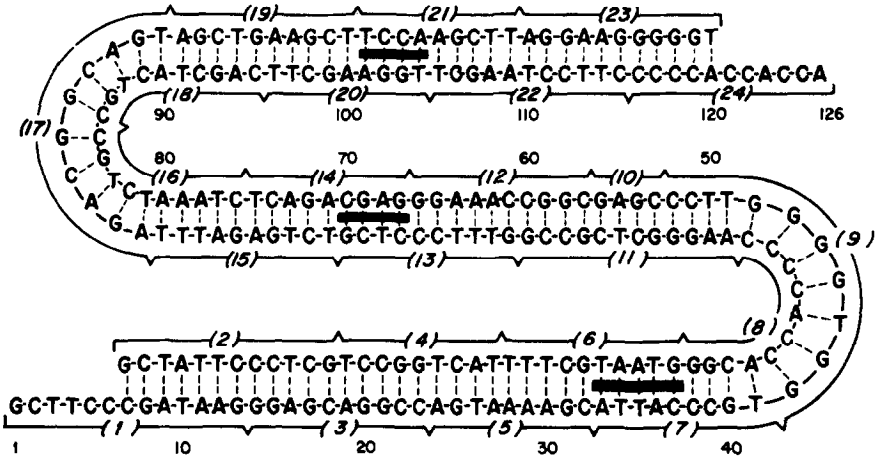


FIG. 24. The totally synthetic DNA duplex corresponding to the entire sequence (126 nucleotides) of the tRNA precursor shown in Fig. 17. The numbers and distances between the carets show the oligonucleotides which were synthesized chemically. The carets indicate the sites where joining was accomplished by the use of polynucleotide ligase. The three thick strips placed between the strands of the duplex at different positions indicate the sites where the preformed duplexes ([I]–[IV]) were joined to each other. The sequence of their joinings was as follows: duplex [I] was joined to duplex [II] (Fig. 22). Separately, duplex [III] was joined to duplex [IV]. Duplex [I + II] was finally joined to duplex [III + IV] (Fig. 23).

(F) Determination of Nucleotide Sequences in the Promoter Region and in the Region Adjoining the CAA End of Tyrosine tRNA Suppressor Gene

As shown in the diagram of Fig. 18, the promoter region lies to the bottom left of the DNA just shown (Fig. 24) and the signal for the processing of the primary gene transcript is at the distal end. As mentioned above, the *E. coli* tyrosine suppressor tRNA gene can be integrated into the temperate bacteriophage $\phi 80$, and the derivative ($\phi 80 psu^{+}_{III}$) provides a much more convenient starting material for sequence work on the suppressor gene. The bacteriophage DNA still is enormously large (single strand, MW $\sim 15 \times 10^6$). Clearly, the aim should be to focus selectively on the very short regions of interest which adjoin the DNA shown in Fig. 18. Because of the fortunate circumstance that, from the synthetic work described above, short or long deoxyribopolynucleotides corresponding to both strands of the structural gene (Fig. 24) are available, the following general approach to the desired sequences becomes possible. The two strands of the bacteriophage $\phi 80 psu^{+}_{III}$ DNA containing the above gene may be separated, and suitable synthetic deoxyribopolynucleotides may be specifically hybridized to the separated strands at sites close to the regions whose sequences are to be determined. Primer–template relationships are thus established with proper polarities. The 3′-ends of the primers may then be extended into the two regions of unknown sequence using DNA polymerases. The latter would bring about nucleotide incorporations according to the nucleotide sequence in the template strand. The nucleotide sequence of the unknown region can then be deduced from the pattern of nucleotide incorporation (*70*, *71*).

Only one set of experiments (Fig. 25) concerned with the determination of the sequence in the promoter region will be outlined to show the concepts; no details of the actual methods used for sequence determination will be given, since these are no longer novel. The primers, DNA I and, preferably, DNA II, are hybridized to the appropriate (l-strand) of $\phi 80 psu^{+}_{III}$ DNA. In the figure are shown the starting point and direction of transcription (to the right) and the start of the promoter region (to the left). DNA II, when extended by five nucleotides, will be at the start of the promoter region. This limited extension will further verify that hybridization of the primer occurred at the correct site. As illustrations of the methods used to keep the newly growing chains within manageable size, the following points may be noted. In the first step (A in Fig. 25), three deoxynucleoside triphosphates were provided in the polymerase reaction mixture. The absence of dTTP limited the chain growth to the sequence shown in the dotted box in A. The chain thus elongated was used again as a primer with a different set of three triphosphates (B in Fig. 25). A third procedure used for selective fragmentation of the newly grown chain was the inclusion of rCTP in place of dCTP, which conferred alkaline lability at specific sites in the chain. Finally, in Expt C of Fig. 25, another principle was used, namely, to add three nucleoside triphosphates at standard concentrations but to add the fourth triphosphate (in this case, dTTP) in a rate-limiting amount. In this way successively longer chains useful in sequencing were obtained. This work was carried out several years ago, and it is of interest to note that these basic concepts are widely used in the current methodology for sequencing of DNA of much greater lengths.

The total promoter sequence as far as determined is shown in Fig. 26A. The arrow at

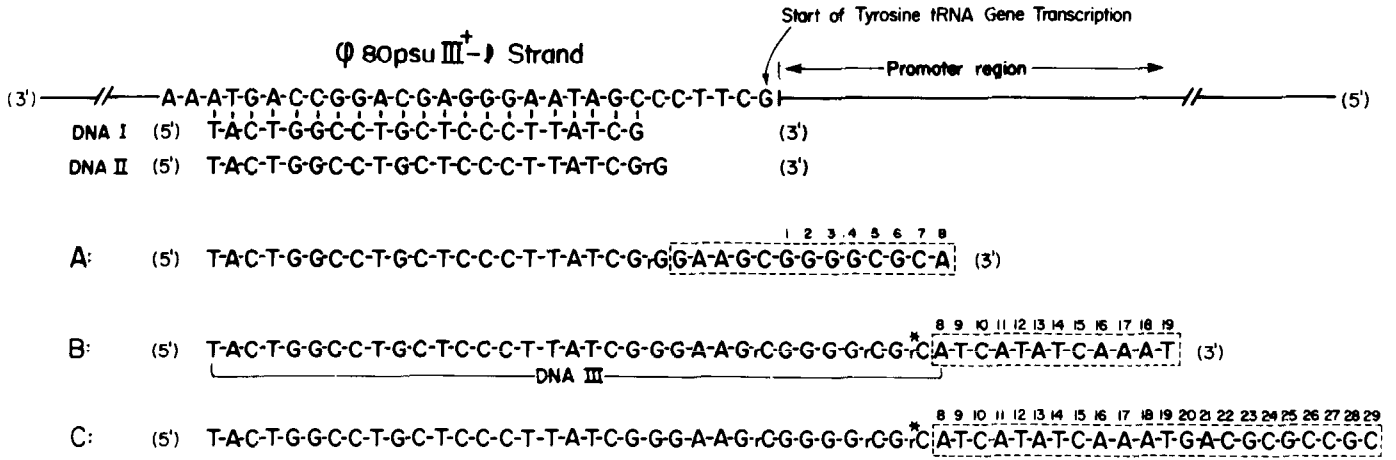


FIG. 25. Experimental design for sequencing and the partial nucleotide sequence determined in the promoter region of the tyrosine tRNA gene. The primer–template complexes prepared at different stages were obtained by hybridizing DNA I, DNA II, or DNA III to the 1-strand of the $\phi 80psu^{+}_{III}$ DNA. DNA polymerase-catalyzed elongations were carried out using three nucleoside triphosphates at a time (further details in text) and substituting ribo-CTP for dCTP. The new nucleotide sequence discovered after each elongation, subsequent alkaline cleavage, and analysis of the new fragment is shown in the appropriate dashed box.

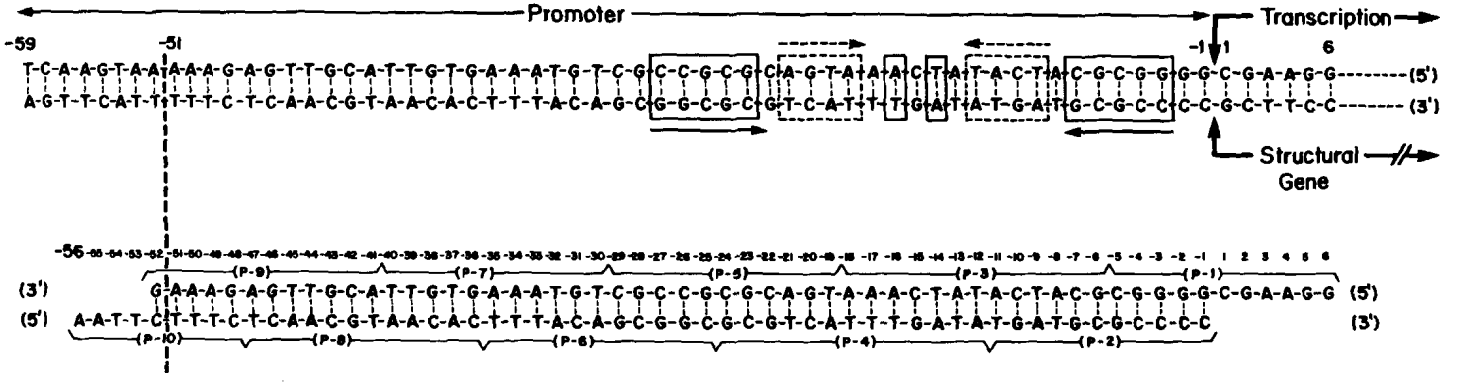


FIG. 26. (A) The nucleotide sequence in the promoter region of the tyrosine suppressor tRNA gene, the point of initiation of transcription, and its direction into the structural gene being at the right. The elements of twofold symmetry in the sequence are shown in the boxes, their correspondence being indicated by arrows. (B) Plan for the total synthesis of the promoter region of the tyrosine suppressor tRNA gene. Included are the 51 nucleotide base pairs in the promoter region plus one C:G base pair plus the single-stranded A-A-T-T sequence at the 5'-end. Together the A-A-T-T-C- sequence is the recognition sequence for the $EcoR_1$ restriction enzyme. The 10 segments (P-1 to P-10) to be synthesized are indicated by horizontal brackets.

the top right shows the direction of transcription, nucleotide No. 1 indicating the starting point. The promoter sequence as far as determined is from nucleotide −1 to −59. The sequence possesses interesting elements of palindromic symmetry that are indicated by dashed and solid boxes and their relationship to each other is shown by the direction of the arrows. Unfortunately, the significance of the remarkable symmetry in the sequence remains unclear, especially since its presence or extent is variable in the different promoters that have been sequenced. How could one tell that the sequence determined is adequate to constitute a functional promoter? Work from our own laboratory (*72*) and cumulative evidence forthcoming from different laboratories did strengthen the notion that this was so. The minimal functional length was estimated to be between 40 and 50 nucleotides.

FIG. 27. Palindromic symmetry in the nucleotide sequence found in the tyrosine tRNA gene beyond the 3'-end of the tRNA. The sequence is shown in the double-stranded configuration.

Turning to the part of the gene beyond the C-C-A, the amino acid acceptor end of the tRNA, we ask the question, How much sequencing will have to be done to cover the signal for the processing of the gene transcript to the tRNA? [It may be noted that the processing to the CCA terminus evidently had occurred *in vivo* in the tRNA precursor isolated (Fig. 17)]. An arbitrarily short sequence of 23 nucleotides adjoining the amino acid acceptor end was determined and the corresponding duplex was synthesized. The sequence is shown in Fig. 27. The palindromic sequence turned out to be of special significance. Thus, by virtue of this symmetry, the RNA counterpart could assume a hairpin structure which, as seen later, is evidently recognized by a specific endonuclease. The signal having been located within the sequence of Fig. 27, it was assumed that it was unneccessary to pursue the sequence work further.

(*G*) *Total Synthesis of the Tyrosine tRNA Suppressor Gene*

In sum, the work reviewed above required the synthesis and joining of a total of eight duplexes shown in Fig. 28. Of these duplexes the synthesis of duplex [I] to [IV] and their joining to form the "structural gene" have already been outlined (Fig. 24). The duplexes P_{1-3} and P_{4-10} at the bottom mostly correspond to the promoter sequence described above (Fig. 26A); in synthetic work, however, modification was introduced at the terminus distal to the structural gene (Fig. 26B). Thus, the natural sequence was retained up to nucleotide −51, and the remaining terminal sequence was changed so as to include the restriction enzyme, $EcoR_1$ recognition sequence, which would enable the insertion of the synthetic gene into a suitable vector. The total duplex corresponding to the promoter region involved the chemical synthesis of 10 segments shown in Figure 26B. The enzymatic joinings required special strategy because of the common sequences within segments forming parts of the palindromes—hence, the synthesis of the promoter in two parts, duplexes P_{1-3} and P_{4-10}. Analogous to the above considerations for modification of the promoter terminus, the duplex at the distal end, presented

and discussed above (Figs 27 and 28, duplex [Va]), was also modified to the duplex [Vb] (Fig. 28).

Thus, the naturally occurring sequence adjoining the CCA end was shortened to only 16 nucleotides. It still completely retained the palindrome, and evidently, because of

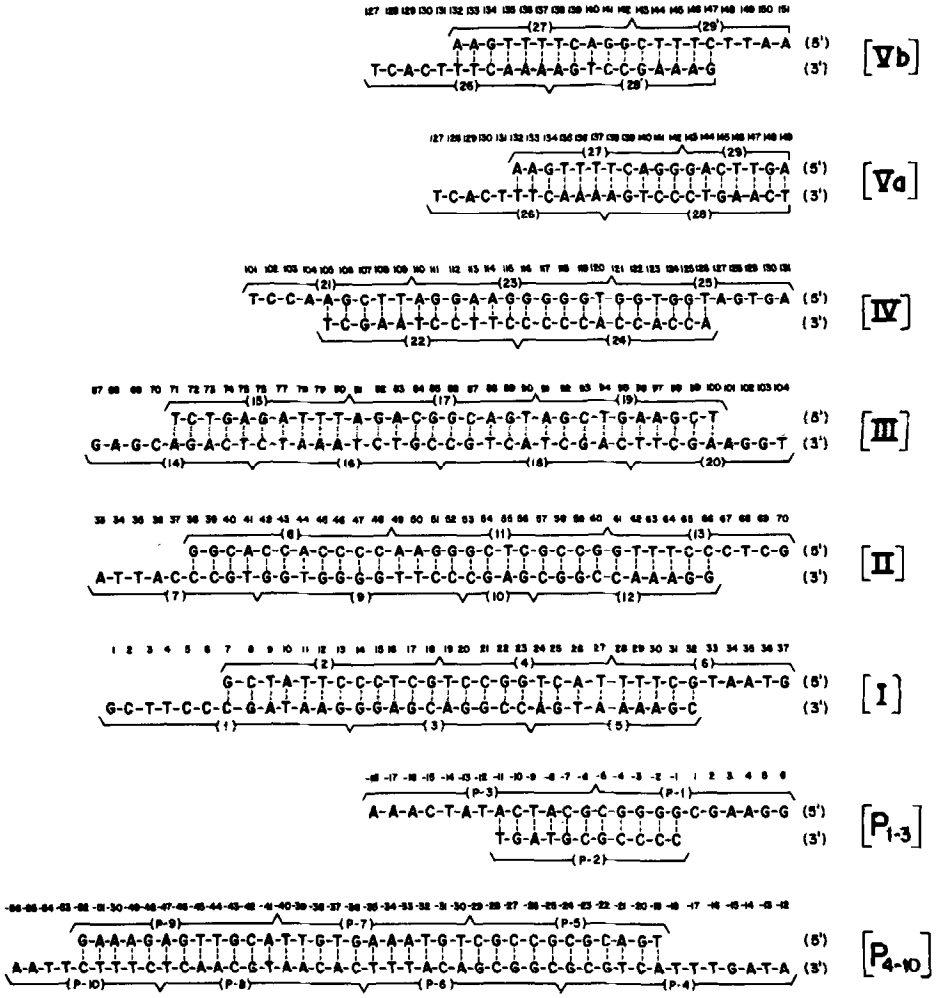


FIG. 28. Plan for the total synthesis of the *E. coli* tyrosine suppressor tRNA gene, including the promoter region and the region containing the processing signal adjoining the C-C-A end. Duplexes [I], [II], [III], and [IV] contain a total of 25 chemically synthesized segments and represent the entire tRNA precursor sequence already shown in Fig. 17. The 5′-pppG terminus of the precursor sequence corresponds to nucleotide 1 (segment 1) of duplex [I]. (The numbers of segments are in parentheses.) The 3′-end of the precursor sequence is contained in segment 25 (duplex [IV]) (nucleotide 126). The grouping of the total segments to the four duplexes shown was arrived at following a large amount of systematic enzymatic work. For details regarding duplexes [P_{1-3}], [P_{4-10}], and [Vb], see the text.

this, there was no deleterious effect on processing of the transcript by the endonuclease. Then, there were a short artificial sequence, which was also used at the promoter end, and the $EcoR_1$ restriction enzyme sequence at the terminus. Therefore both termini of

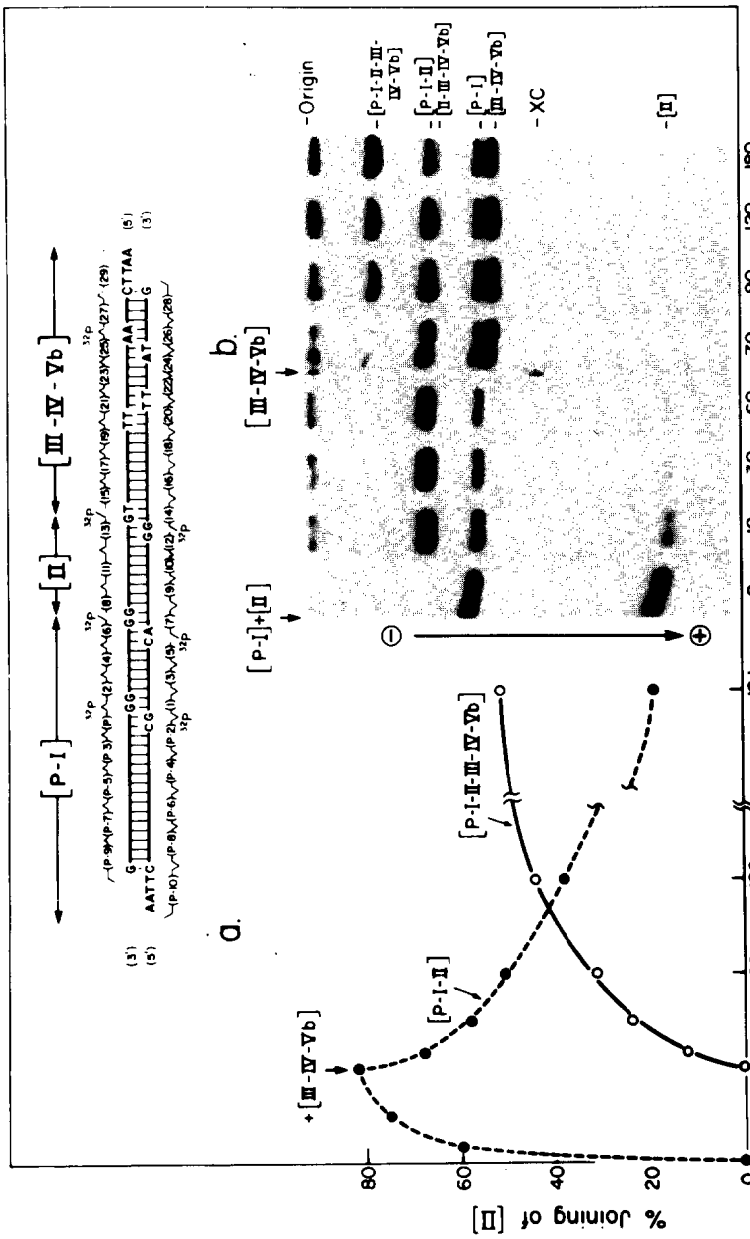


FIG. 29. Stepwise synthesis of the synthetic gene [P-I-II III-IV-Vb]: an example of the enzymatic joining of the duplexes. The promoter duplex was first joined to duplex [I] to form [P + I]. To the latter was added duplex [II], the joining to form [P-I-II] being rapid as seen in the first curve. Preformed [III-IV-Vb] was then added and the kinetics of formation of the total gene was followed by electrophoresis on a polyacrylamide gel. The total joinings performed in stepwise fashion can all be monitored conveniently by shifts in mobilities on gel electrophoresis.

the synthetic gene would carry the protruding sequence for insertion into suitable vectors.

The joining of the eight duplexes in Fig. 28 could be approached in many alternative ways. One final-step experiment leading to the totally synthetic gene is shown in Fig. 29. As mentioned above, the joining of preformed duplexes through their complementary protruding oligonucleotide sequence proceeds rapidly and in high yields. The totally

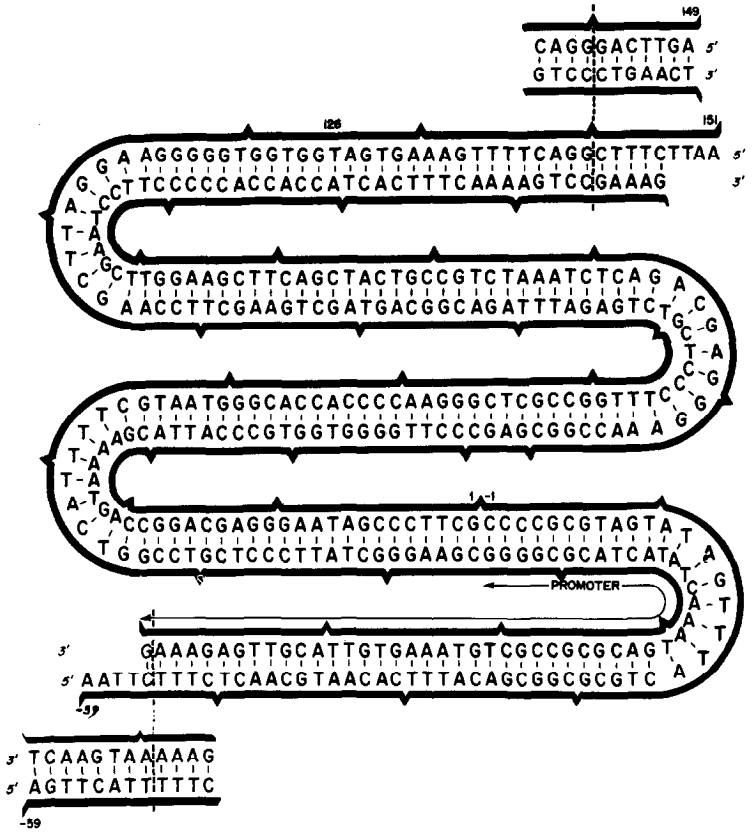


FIG. 30. Synthetic tyrosine suppressor tRNA gene. At bottom left is the terminal $EcoR_1$ endonuclease-specific sequence. Then follow a 51-nucleotide-long promoter region, a 126-nucleotide-long DNA corresponding to the precursor RNA and, finally, a 25-nucleotide-long region. Of the latter, 16 nucleotides belong to the natural sequence including the $EcoR_1$ endonuclease-specific sequence. Additional duplex segments shown on top right and bottom left are the sequences as they naturally are found in these regions and continue in the two directions.

synthetic gene, isolated and characterized in a variety of ways, is shown in Fig. 30. The additional sequences shown at top right and at bottom left are the natural sequences as they would continue in the two directions in the *E. coli* genome.

TRANSCRIPTION *IN VITRO* OF THE SYNTHETIC GENE

Two of the first aims following total synthesis were (i) the *in vitro* transcription followed by processing of the primary transcript and base modifications to form the

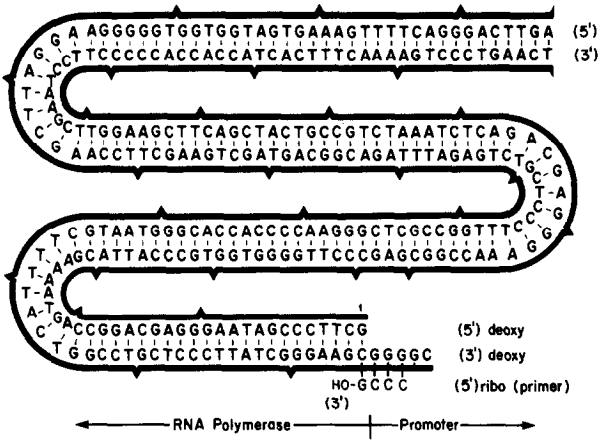


FIG. 31. Primer-dependent transcription of the synthetic gene lacking the promoter but containing an additional 23-nucleotide-long DNA beyond the C-C-A end. Only one synthetic oligonucleotide segment at the distal end has five nucleotides protruding into the promoter region (bottom right). The direction of transcription by RNA polymerase is shown. Transcription was specific to the required DNA strand, complementary to the primer sequence.

mature tyrosine tRNA and (ii) the demonstration, *in vivo*, of suppression of the amber mutation following insertion of the gene into a suitable vector and transfection into a suitable strain of *E. coli*. Transcription was first studied in detail using the synthetic DNA shown in Fig. 31. The structure of the promoter was largely unknown at this time, and only a segment containing a single-stranded sequence corresponding to the first five nucleotides of the promoter region was linked to the structural gene. The DNA shown in Fig. 31 also contained the duplex [Va] (Fig. 28) at the distal (C-C-A) end. After earlier studies of primer-dependent transcription of single- and double-stranded

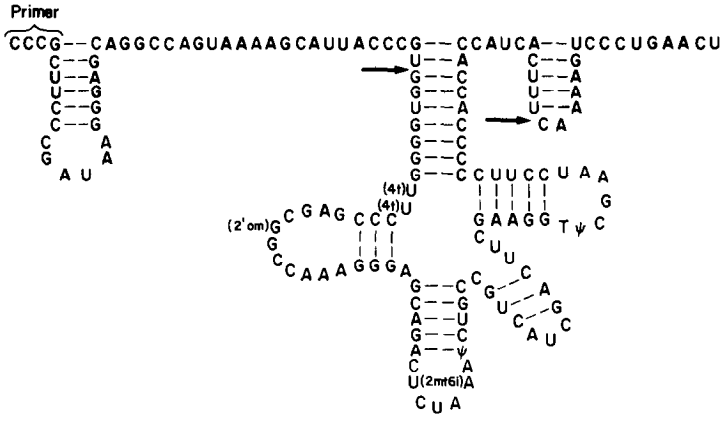


FIG. 32. Structure of the main primary transcript using primer-dependent transcription of synthetic DNA, as shown in Fig. 31. Completion of the different modifications of bases as shown is not certain Probably many modifications were incomplete. The two crucial endonucleolytic cleavages (arrows) for processing to functional tRNA are shown. The arrow on the left is the previously determined site of cleavage by endonuclease P (*74*). The arrow on the right is the site now determined as the first step in processing at the 3'-end of the transcript.

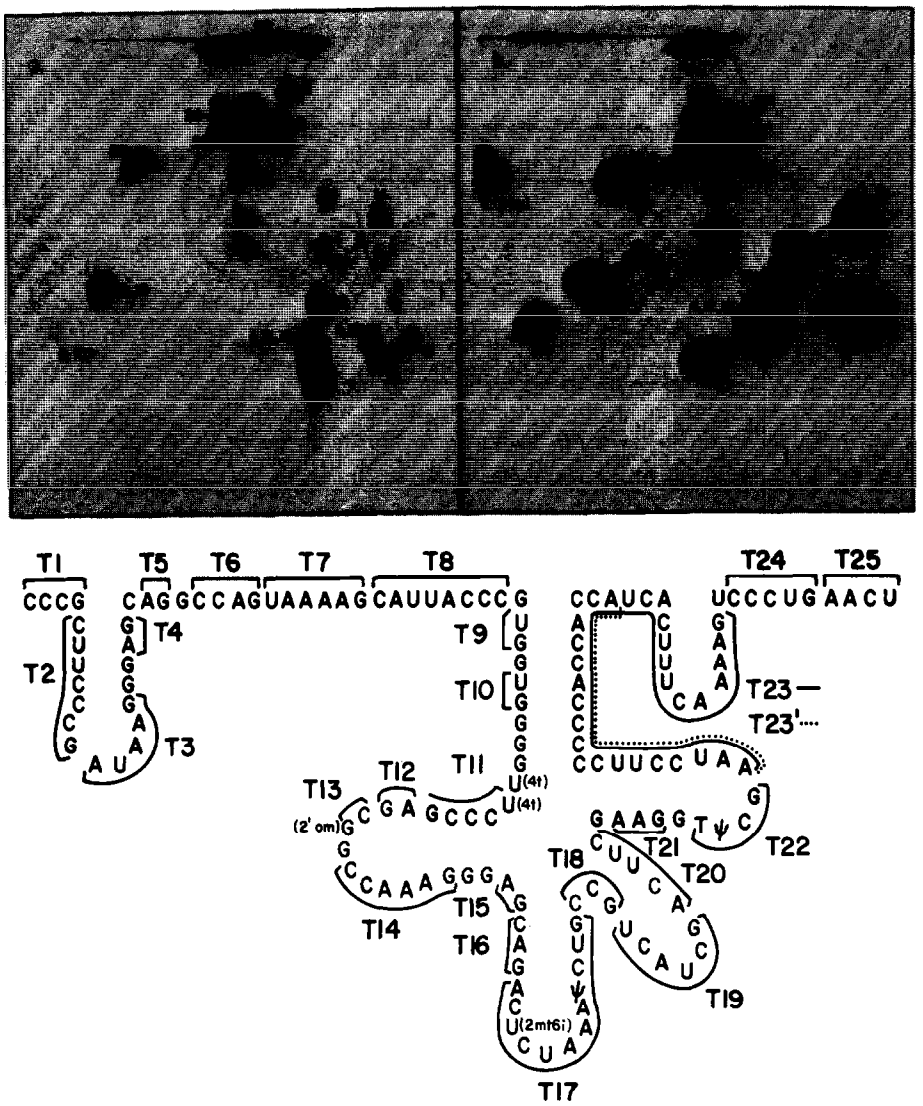


FIG. 33. Fingerprints (a) and (b), respectively, of the products obtained on T_1-RNase digestion of the primary transcript (transcription experiment performed as in Fig. 31) and of the tRNA obtained by processing with *E. coli* extracts. The oligonucleotides to be expected in the primary transcript and in the processed transcript are shown systematically in (c). In the sequence, modified nucleosides are present in mature $tRNA_1^{Tyr}$ are shown. In the primary transcript, the modifications are, of course, absent and the parent nucleosides are present.

synthetic deoxyribopolynucleotides (*73*), the ribotetranucleotide (5′)-C-C-C-G-(3′) complementary to the 3′-terminus sequence was used as the primer (Fig. 31). Under suitable conditions, transcription was essentially restricted to that of the desired strand, and the primary transcript was shown to consist largely of the RNA with the structure shown in Fig. 32. When the latter was exposed to an *E. coli* extract, 30,000 *g* supernatant, processing and maturation occurred, as expected. Thus, the first two endonucleolytic cleavages occurred at the sites indicated by the two arrows. The

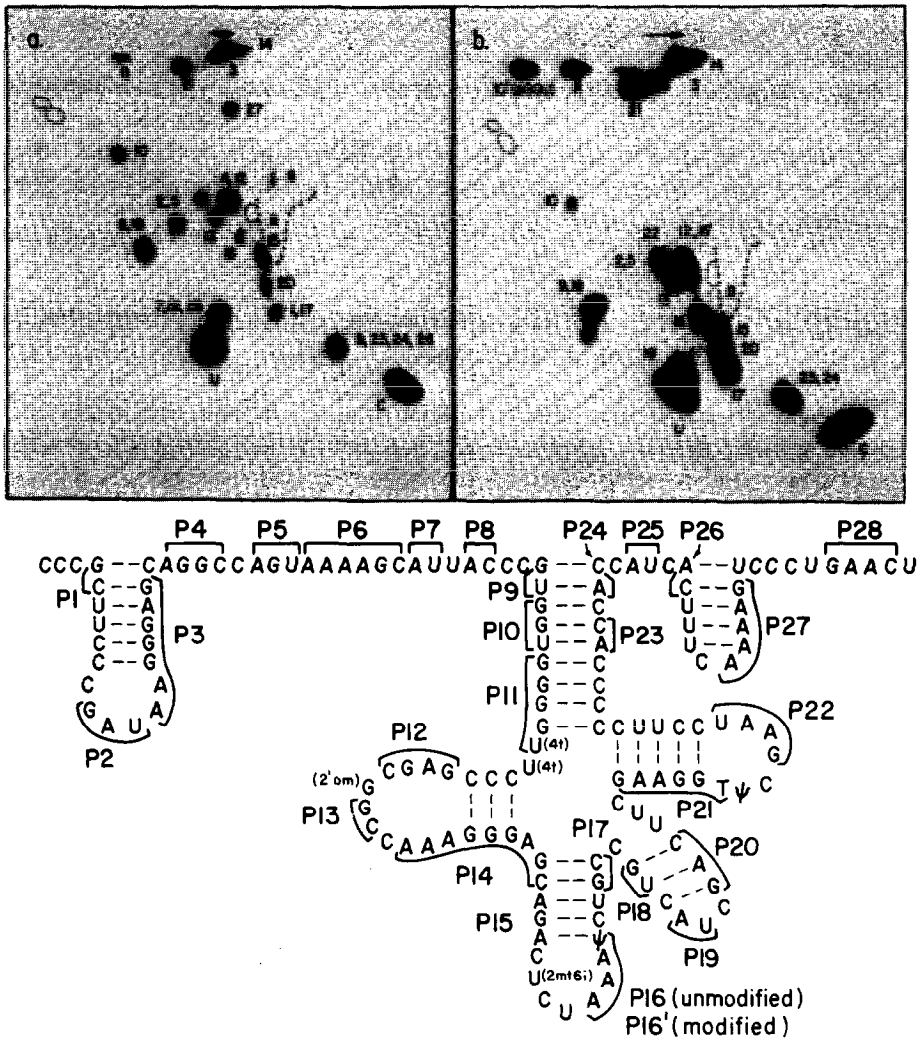


FIG. 34. Fingerprints (a) and (b), respectively, of the products obtained by pancreatic RNase digestion of the primary transcript (transcription experiment performed as in Fig. 31) and of the tRNA obtained by processing with *E. coli* extracts. The oligonucleotides to be expected are shown schematically in (c). As in Fig. 32, the sequence contains modified nucleosides as present in mature $tRNA_1^{Tyr}$. The modifications are absent in the primary transcript.

nuclease generating the 5′-terminus of the tRNA had been described before (*66*, *74*), but the nuclease acting beyond the C-C-A sequence was new and it evidently recognized the hairpin shown. The seven nucleotides remaining at the 3′-end were demonstrated to be removed stepwise by another exonuclease (*75*). Base modifications, at least partial, were shown to occur in the resulting tRNA. The latter as well as the primary transcript were thoroughly characterized by fingerprinting and by comprehensive nearest-neighbor analyses of the oligonucleotides produced by digestion with T_1 and pancreatic RNases (Figs. 33 and 34). All of the data were completely consistent with those expected for the total sequence. Full analyses of the transcripts provided

sensitive means of confirming the accuracy of the synthetic work, at least in the transcribed part.

Identical transcription experiments were next performed to study transcription of the synthetic gene containing the promoter region (Fig. 30). Transcription was indeed found to be determined by the promoter, in that it was initiated with pppG as the →5′-end nucleotide and was followed by the sequence which is found naturally at the 5′-end. Furthermore, the transcription was strand specific.

BIOLOGICAL ACTIVITY OF THE SYNTHETIC GENE (SUPPRESSION OF AMBER MUTATION *IN VIVO*)

Suppression of amber (nonsense) mutation was a decisive experiment to be performed with the synthetic gene. Indeed, the demonstration of suppression and the clear-cut demonstration that this involved a single nucleotide change in the anticodon of the tyrosine tRNA (change in anticodon from 5′---G*UA---3′ to 5′---CUA---3′) to enable recognition of the amber codon, UAG, were strong arguments for the choice of tyrosine suppressor tRNA as the synthetic target.

A number of experiments performed with the synthetic gene demonstrated its ability to suppress amber mutations in intact bacteria and in bacteriophages (*76*). The protocol

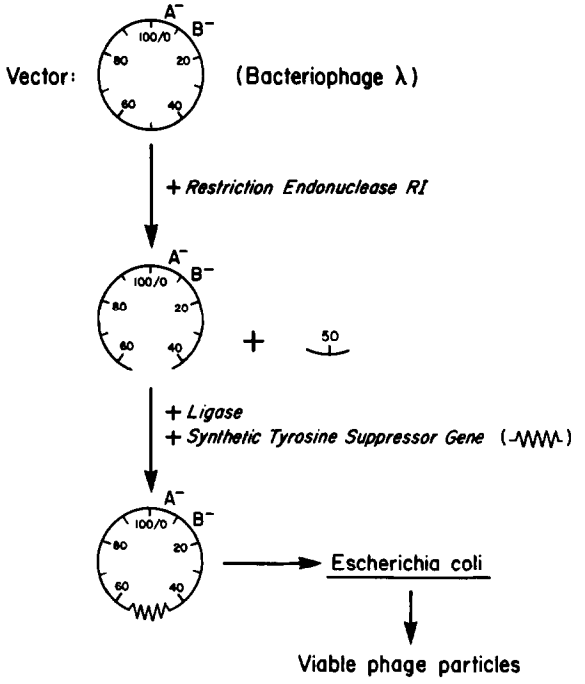


FIG. 35. Cloning of the synthetic gene for the tyrosine suppressor tRNA. The vector used was a derivative of bacteriophage λ with two amber mutations (A^-, B^-). Digestion with restriction endonuclease R_1 excised a nonessential piece (at 50 min of genetic map) out of the total genome. Subsequent addition of the synthetic gene (*76*) and ligase gave the circular phage containing phage particles. The results are shown in Fig. 36.

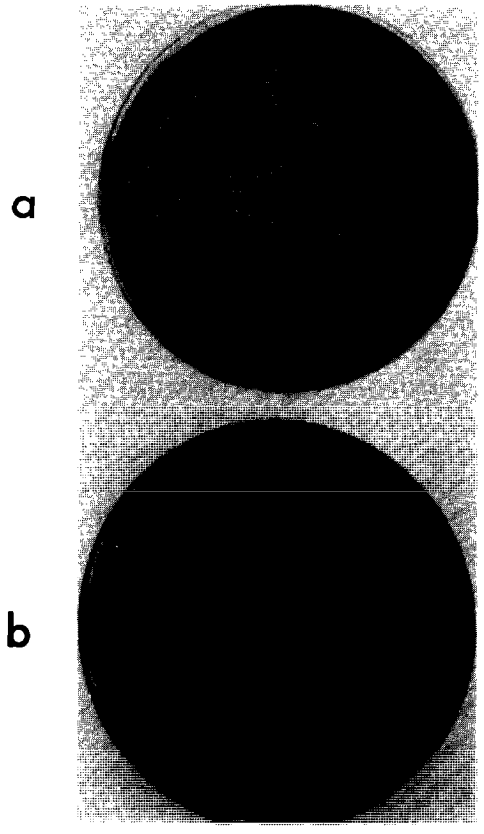


FIG. 36. Experiment showing phage growth in *E. coli* following insertion of the synthetic amber suppressor gene. The two culture dishes show (a) bacteriophage λ with amber mutation used as the vector for the synthetic gene and (b) the same vector with inserted suppressor tRNA gene, prepared as that shown in Fig. 35. The phage plaques were present only in (b).

for a typical experiment is shown in Fig. 35. Bacteriophage λ DNA carrying *A*am 32 and *B*am1 mutations (*77*) was digested with $EcoR_1$ endonuclease. A nonessential segment was thus excised (Fig. 35). The synthetic gene was then covalently inserted into the linear bacteriophage DNA by using T_4 polynucleotide ligase (the insertion must have occurred randomly in regard to direction). The resulting circular molecules (Fig. 35) were used to transform *E. coli* bacterial cultures. As seen in Fig. 36, suppression of the two amber mutations in the phage DNA and consequent formation of phages plaques were observed only with the phage DNA carrying the synthetic suppressor gene.

CONCLUDING REMARKS

In more recent years, the methodology for the step-by-step synthesis of bihelical DNA of any given sequence has been developed. There are complete control and unambiguity at all stages. Chemically synthesized segments of specific sequences can be rigorously purified and analyzed. Similarly, enzymatic joinings can be monitored stringently. In principle, there is no limit on the size or structure of a DNA that can be

constructed by the organobiochemical methods. In practice, organic synthesis of the segments still takes a lot of time and repetitive work, and there is great room for ingenuity and simplification of the procedures, although the use of new protecting groups as well as of high-pressure liquid chromatography represent very significant advances. Furthermore, it remains a complete mystery and a chemical paradox why the enzymatic joinings of single-stranded segments to form DNA duplexes do not go to completion. Clearly, disclosure of the reasons for variations in yields and consequent improvements would be of marked significance in the laboratory synthesis of DNA.

Total synthesis of a DNA containing specific nucleotide sequences differs conceptually from the *in vivo* or *in vitro* DNA replication catalyzed by the nucleic acids polymerizing enzymes. The latter require a preformed template strand to bring about polynucleotide synthesis and, therefore, the important biological function served is the faithful replication of information rather than the creation of new information. In total synthesis, as illustrated by the work herein described, both strands are constructed in response to a predetermined sequence. Since the sequence is generated by chemical synthesis, there is full choice on the subsequent manipulation of the sequence information by designed chemical change. This ability is the essence of the chemical approach to the study of biological specificity in DNA and RNA.

Completion of the total synthesis of the gene for a transfer RNA serves as but one example of the enormous opportunities for the studies of structure–function relationships in different genes. In the present work, the large questions which can all be studies by the "chemical" approach are (i) mechanism of the action of the promoter which "turns on" the transcription of the tRNA gene; (ii) structure–function studies in the mature tRNA itself; (iii) detailed mechanisms of specificity of the host of enzymes which bring about processing and base modifications (maturation) of the tRNA molecule. While comments have been made above in regard to the intriguing questions of the specificity of the tRNA molecules in protein synthesis, the following brief comment may be made regarding the promoter regions which control transcription. For example, in *Escherichia coli*, all of the promoters and those of the infecting viruses are evidently recognized by one RNA polymerase, and, despite the fact that about 25 promoters have been sequenced, the structural features important in (a) the recognition of the promoter region by the RNA polymerase, (b) the formation of the stable complex, and (c) the point of initiation of transcription remain poorly understood. Systematic synthetic modifications in one promoter may prove to be the only approach toward dissecting the role of different sections comprising the promoter region.

In the general field of work involving recombinant DNA, the present methodology may complement the genetic approaches. For example, it may not be necessary to synthesize full lengths of the genes with their own control elements, as was done in the present work. Instead, shorter, more easily manageable segments corresponding to the necessary gene products, e.g., peptide hormones, may be synthesized and inserted into the plasmids or vectors.

An area in molecular biology which will receive increasing focus is the large problem of nucleic acid–protein interactions; the promoter–RNA polymerase interactions mentioned above are an example of this type. Until the recent dramatic progress in the determination of DNA sequences, the degrees of freedom in applying synthesis to such problems were too great. However, now with the framework in protein–nucleic acid

interactions being narrowed down, synthesis can provide systematic approaches to the precise studies of the control regions in different systems. The synthetic approach is now being used increasingly in such studies (*78–80*) and may often be the only one that enables a deeper understanding of biological controls at the molecular level. Therefore, in an overall assessment of the prospects, the predictions may be made that organobiochemistry of the type illustrated in this article will interact increasingly and more effectively with studies related to the mechanisms of control and regulation of the expression of genetic information.

REFERENCES

1. A. HARDEN AND W. J. YOUNG, *Proc. Roy. Soc.* (*London*), **B77** (1906).
2. V. R. POTTER, "Nucleic Acid Outlines," Vol. 1. Burgess, Minneapolis, Minn., 1960.
3. H. G. KHORANA, "Some Recent Developments in the Chemistry of Phosphate Esters of Biological Interest," Wiley, New York, 1961.
4. D. M. BROWN AND A. R. TODD, "Nucleic Acids" (E. Chargaff and J. N. Davidson, Eds.), Vol. 1, pp. 409–445. Academic Press, New York, 1955.
5. H. G. KHORANA, *Pure Appl. Chem.* **17**, 349 (1968).
6. R. A. JONES, H.-J. FRITZ, AND H. G. KHORANA, *Biochemistry* **17**, 1268 (1978).
7. H.-J. FRITZ, R. BELAGAJE, E. L. BROWN, R. H. FRITZ, R. A. JONES, R. G. LEES, AND H. G. KHORANA, *Biochemistry* **17**, 1257 (1978).
8. D. H. RAMMLER AND H. G. KHORANA, *J. Amer. Chem. Soc.* **84**, 3112 (1962).
9. Y. LAPIDOT AND H. G. KHORANA, *J. Amer. Chem. Soc.* **85**, 3852 (1963).
10. D. H. RAMMLER, Y. LAPIDOT, AND H. G. KHORANA, *J. Amer. Chem. Soc.* **85**, 1989 (1963).
11. R. LOHRMANN, D. SOLL, H. HAYATSU, E. OHTSUKA, AND H. G. KHORANA, *J. Amer. Chem. Soc.* **88**, 819 (1966).
12. H. G. KHORANA, "The Enzymes" (P. D. Boyer and H. A. Lardy, Eds.), 2nd ed., Vol. 5, pp. 79–94. Academic Press, New York, 1961.
13. F. SANGER, J. E. DONELSON, A. R. COULSON, H. KÖSSEL, AND D. FISHER, *Proc. Nat. Acad. Sci. USA* **70**, 1209 (1973).
14. G. W. BEADLE AND E. L. TATUM, *Proc. Nat. Acad. Sci. USA* **27**, 499 (1941).
15. R. D. HOTCHKISS, "Nucleic Acids" (E. Chargaff and J. N. Davidson, Eds.), Vol. 2, p. 435. Academic Press, New York, 1955.
16. O. T. AVERY, C. M. MACLEOD, AND M. MCCARTY, *J. Exp. Med.* **79**, 137 (1944).
17. A. D. HERSHEY AND M. CHASE, *J. Gen Physiol.* **36**, 39 (1952).
18. A. GIERER AND G. SCHRAMM, *Nature* (*London*) **177**, 702 (1956).
19. H. FRAENKEL-CONRAT, B. SINGER, AND R. C. WILLIAMS, *Biochem. Biophys. Acta* **25**, 87, (1957).
20. J. D. WATSON AND F. H. C. CRICK, *Nature* (*London*) **171**, 737 (1953).
21. I. R. LEHMAN, M. J. BESSMAN, E. S. SIMMS, AND A. KORNBERG, *J. Biol. Chem.* **233**, 163 (1958).
22. S. B. WEISS, *Proc. Nat. Acad. Sci. USA* **46**, 1020 (1960).
23. A. J. STEVENS, *J. Biol. Chem.* **236**, PC43 (1961).
24. J. HURWITZ, J. J. FURTH, M. ANDERS, AND A. J. EVANS, *J. Biol. Chem.* **237**, 3752 (1962).
25. M. CHAMBERLIN AND P. BERG, *Proc. Nat. Acad. Sci. USA* **48**, 81 (1962).
26. D. P. BURMA, H. KROGER, S. OCHOA, R. C. WARNER, AND J. D. WEILL, *Proc. Nat. Acad. Sci. USA* **47**, 749 (1961).
27. P. C. ZAMECNIK, "The Harvey Lectures," Series 54, p. 256.Academic Press, New York, 1958–1959.
28. F. LIPMANN, W. C. HULSMANN, H. G. HARTMANN, H. G. BOMAN, AND G. ACS, *J. Cellular Comp. Physiol., Suppl.* 1, **54**, 75 (1969).
29. P. BERG, *Annu. Rev. Biochem.* **30**, 293 (1961).
30. A. TISSIERES, D. SCHLESSINGER, AND F. GROS, *Proc. Nat. Acad. Sci. USA* **46**, 1450 (1960).
31. J. H. MATTHAEI AND M. W. NIRENBERG, *Proc. Nat. Acad. Sci. USA* **47**, 1580 (1961).
32. E. OHTSUKA, M. W. MOON, AND H. G. KHORANA, *J. Amer. Chem. Soc.* **87**, 2956 (1965).

33. T. M. JACOB AND H. G. KHORANA, *J. Amer. Chem. Soc.* **87**, 2971 (1965).
34. S. A. NARANG AND H. G. KHORANA, *J. Amer. Chem. Soc.* **87**, 2981 (1965).
35. S. A. NARANG, T. M. JACOB, AND H. G. KHORANA, *J. Amer. Chem. Soc.* **87**, 2988 (1965).
36. S. A. NARANG, T. M. JACOB, AND H. G. KHORANA, *J. Amer. Chem. Soc.* **89**, 2158 (1967).
37. S. A. NARANG, T. M. JACOB, AND H. G. KHORANA, *J. Amer. Chem. Soc.* **89**, 2167 (1967).
38. H. KÖSSEL, H. BÜCHI, AND H. G. KHORANA, *J. Amer. Chem. Soc.* **89**, 2185 (1967).
39. E. OHTSUKA AND H. G. KHORANA, *J. Amer. Chem. Soc.* **89**, 2195 (1967).
40. H. G. KHORANA, H. BÜCHI, T. M. JACOB, H. KÖSSEL, S. A. NARANG, AND E. OHTSUKA, *J. Amer. Chem. Soc.* **89**, 2154 (1967).
41. H. KÖSSEL, A. R. MORGAN, AND H. G. KHORANA, *J. Mol. Biol.* **26**, 449 (1967).
42. H. KÖSSEL, *Biochim. Biophys. Acta* **157**, 91 (1968).
43. R. D. WELLS, T. M. JACOB, S. A. NARANG, AND H. G. KHORANA, *J. Mol. Biol.* **27**, 237 (1967).
44. C. BYRD, E. OHTSUKA, M. W. MOON, AND H. G. KHORANA, *Proc. Nat. Acad. Sci. USA* **53**, 79 (1965).
45. R. D. WELLS, E. OHTSUKA, AND H. G. KHORANA, *J. Mol. Biol.* **14**, 221 (1965).
46. H. G. KHORANA, *Fed. Proc.* **24**, 1473 (1965).
47. H. G. KHORANA, "Genetic Elements: Properties and Function" (D. Shugar, Ed.), p. 209. Federation of European Biochemical Societies, Academic Press, London/New York, 1966.
48. H. G. KHORANA, H. BÜCHI, H. GHOSH, N. GUPTA, T. M. JACOB, H. KÖSSEL, A. R. MORGAN, S. A. NARANG, E. OHTSUKA, AND R. D. WELLS, "Symposia on Quantitative Biology," Vol. 31, p. 39. Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring Harbor, N.Y., 1966.
49. R. D. WELLS, T. M. JACOB, S. A. NARANG, AND H. G. KHORANA, *J. Mol. Biol.* **27**, 227 (1967).
50. R. D. WELLS, H. BÜCHI, H. KÖSSEL, E. OHTSUKA, AND H. G. KHORANA, *J. Mol. Biol.* **27**, 237 (1967).
51. S. NISHIMURA, D. S. JONES, AND H. G. KHORANA, *J. Mol. Biol.* **13**, 302 (1965); H. G. KHORANA, "*Genetic Elements: Properties and Functions*" (D. Shugar, Ed.), p. 209). Federation of European Biochemical Societies, Academic Press, New York/London, 1966.
52. M. W. NIRENBERG AND P. LEDER, *Science* **145**, 1399 (1964).
53. "Symposia on Quantitative Biology," Vol. 31. Cold Spring Harbor Laboratory of Quantitative Biology, Cold Spring Harbor. N.Y. 1966.
54. H. G. KHORANA, "The Harvey Lectures," Series 62, p. 79. Academic Press, New York, 1968.
55. R. W. HOLLEY, J. APGAR, G. A. EVERETT, J. T. MADISON, M. MARQUISEE, S. H. MERRILL, J. R. PENSWICK, AND A. ZAMIR, *Science* **147**, 1462 (1965).
56. H. G. KHORANA, *Fed. Proc.* **19**, 931 (1960).
57. *Acc. Chem. Res.* **10**, No. 11 (1977).
58. M. GELLERT, *Proc. Nat. Acad. Sci. USA* **57**, 148 (1967); S. B. ZIMMERMAN, J. W. LITTLE, C. K. OSHINSKY, AND M. GELLERT, *Proc. Nat. Acad. Sci. USA* **57**, 1841 (1967); B. WEISS AND C. C. RICHARDSON, *Proc. Nat. Acad. Sci. USA* **57**, 1021 (1967); B. M. OLIVERA AND I. R. LEHMAN, *Proc. Nat. Acad. Sci. USA* **57**, 1426 (1967); M. L. GEFTER, A. BECKER, AND J. HURWITZ, *Proc. Nat. Acad. Sci. USA* **58**, 240 (1967).
59. C. C. RICHARDSON, *Proc. Nat. Acad. Sci. USA* **54**, 158 (1965).
60. N. K. GUPTA, E. OHTSUKA, H. WEBER, S. H. CHANG, AND H. G. KHORANA, *Proc. Nat. Acad. Sci. USA* **60**, 285 (1968); N. K. GUPTA, E. OHTSUKA, V. SGARAMELLA, H. BÜCHI, A. KUMAR, H. WEBER, AND H. G. KHORANA, *Proc. Nat. Acad. Sci. USA* **60**, 1338 (1968).
61. R. W. HOLLEY, J. APGAR, G. A. EVERETT, J. T. MADISON, M. MARQUISEE, S. H. MERRILL, J. R. PENSWICK, AND A. ZAMIR, *Science* **147**, 1462 (1965).
62. K. L. AGARWAL, H. BÜCHI, M. H. CARUTHERS, N. GUPTA, H. G. KHORANA, K. KLEPPE, A. KUMAR, E. OHTSUKA, U. L. RAJBHANDARY, J. H. VAN DE SANDE, V. SGARAMELLA, H. WEBER, AND T. YAMADA, *Nature (London)* **227**, 27 (1970); H. G. KHORANA, K. L. AGARWAL, H. BÜCHI, M. H. CARUTHERS, N. GUPTA, K. KLEPPE, A. KUMAR, E. OHTSUKA, U. L. RAJBHANDARY, J. H. VAN DE SANDE, V. SGARAMELLA, T. TERAO, H. WEBER, AND T. YAMADA, *J. Mol. Biol.* **72**, 209 (1972), and accompanying papers.
63. H. M. GOODMAN, J. ABELSON, A. LANDY, S. BRENNER, AND J. D. SMITH, *Nature (London)* **217**, 1019 (1968); U. L. RAJBHANDARY, S. H. CHANG, H. J. GROSS, F. HARADA, F. KIMURA, AND S. NISHIMURA, *Fed. Proc.* **28**, 409 (1969).
64. J. D. SMITH, J. N. ABELSON, B. F. CLARK, H. M. GOODMAN, AND S. BRENNER, "Symposia on Quantitative Biology," Vol. 31, p. 479. Cold Spring Harbor Laboratory of Quantitative Biology, Cold

Spring Harbor, N.Y. 1966.; R. L. RUSSELL, J. N. ABELSON, A. LANDY, M. L. GEFTER, S. BRENNER, AND J. D. SMITH, *J. Mol. Biol.* **47**, 1 (1970).
65. J. D. SMITH, *Brit. Med. Bull.* **92**, 220 (1973).
66. S. ALTMAN, *Nature New Biol.* **229**, 19 (1971); S. ALTMAN AND J. D. SMITH, *Nature New Biol.* **233**, 35 (1971).
67. KÜPPER, T. SEKIYA, M. ROSENBERG, J. EGAN, AND A. LANDY, *Nature* (*London*), **270**, 423 (1978).
68. V. SGARAMELLA AND H. G. KHORANA, *J. Mol. Biol.* **72**, 427 (1972); J. H. VAN DE SANDE, M. H. CARUTHERS, V. SGARAMELLA, T. YAMADA, AND H. G. KHORANA, *J. Mol. Biol.* **72**, 457 (1972); P. C. LOEWEN, R. C. MILLER, A. PANET, T. SEKIYA, AND H. G. KHORANA, *J. Biol. Chem.* **251**, 642.
69. 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, *J. Biol. Chem.* **251**, 565 (1976), and accompanying papers.
70. P. C. LOEWEN AND H. G. KHORANA, *J. Biol. Chem.* **248**, 3489 (1973); P. C. LOEWEN, T. SEKIYA, AND H. G. KHORANA, *J. Biol. Chem.* **249**, 217 (1974).
71. T. SEKIYA AND H. G. KHORANA, *Proc. Nat. Acad. Sci. USA* **71**, 2978 (1974); T. SEKIYA, M. J. GAIT, K. NORRIS, B. RAMAMOORTHY, AND H. G. KHORANA, *J. Biol. Chem.* **251**, 4481 (1976); T. SEKIYA, R. CONTRERAS, H. KÜPPER, A. LANDY, AND H. G. KHORANA, *J. Biol. Chem.* **251**, 5124 (1976).
72. T. SEKIYA, T. TAKEYA, R. CONTRERAS, H. KÜPPER, H. G. KHORANA, AND A. LANDY, "RNA Polymerase" (R. Losick and M. Chamberlin, Eds.), pp. 455–472. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. 1976.
73. T. TERAO, J. E. DAHLBERG, AND H. G. KHORANA, *J. Biol. Chem.* **247**, 6157 (1972); and unpublished work of Dr. R. Contreras.
74. H. D. ROBERTSON, S. ALTMAN, AND J. D. SMITH, *J. Biol. Chem.* **247**, 5243 (1972).
75. E. K. BIKOFF AND M. L. GEFTER, *J. Biol. Chem.* **251**, 6240 (1975).
76. M. J. RYAN, R. BELAGAJE, E. L. BROWN, H.-J. FRITZ, M. J. GAIT, H. KÜPPER, R. G. LEES, NORRIS, T. SEKIYA, AND T. TAKEYA, *Fed. Proc.* **36**, 732, Abstr. No. 3, (1977).
77. F. R. BLATTNER, B. G. WILLIAMS, A. E. BLECHL, K. DENNISTON-THOMPSON, H. E. FABER, L.-A. FURLONG, D. J. GRUNWALD, D. O. KIETER, D. D. MOORE, J. W. SCHUMM, E. L. SHELDON, AND O. SMITHIES, *Science* **196**, 161 (1977).
78. E. L. BROWN, R. BELAGAJE, H.-J. FRITZ, K. NORRIS, AND H. G. KHORANA, *Fed. Proc.* **36**, 732 (1977).
79. D. V. GOEDDEL, D. G. YANSURA, AND M. H. CARUTHERS, *Proc. Nat. Acad. Sci. USA* **74**, 3292 (1977) and references cited therein.
80. C. P. BAHL, R. WU, J. STOWINSKY, AND S. A. NARANG, *Proc. Nat. Acad. Sci. USA* **74**, 966 (1977).
81. M. H. CARUTHERS, R. KLEPPE, K. KLEPPE, AND H. G. KHORANA, *J. Biol. Chem.* **251**, 568 (1976).
82. R. KLEPPE, T. SEKIYA, P. C. LOEWEN, K. KLEPPE, K. L. AGARWAL, P. BESMER, H. BÜCHI, M. H. CARUTHERS, P. J. CASHION, M. FRIDKIN, E. JAY, A. KUMAR, R. C. MILLER, K. MINAMOTO, A. PANET, U. L. RAJBHANDARY, B. RAMAMOORTHY, N. SIDOROVA, T. TAKEYA, J. H. VAN DE SANDE, AND H. G. KHORANA, *J. Biol. Chem.* **251**, 667 (1976).