DR H. G. KHORANA

Biochem. J. (1968) **109**, 709
Printed in Great Britain

Synthesis in the Study of Nucleic Acids

THE FOURTH JUBILEE LECTURE

By H. G. KHORANA*

Delivered on 17 April 1968 in the Beveridge Hall, Senate House, University of London, Malet Street, London, W.C. 1, and on 22 April 1968 in the Main Lecture Theatre, Department of Chemistry, University of Glasgow

Historical and introduction

I wish first of all to express my deep appreciation for the honour that The Biochemical Society has done me by inviting me to give this Lecture. It is at once a matter of great pride and pleasure for me to be visiting England on such an occasion, for it is in this country that I spent some of my formative years.

I would like to begin by giving a brief historical account of some of the developments in the field of molecular biology. Although it is always difficult to determine or clearly define the starting point in any area of science, the idea that genes make proteins was a distinct landmark and, in particular, the specific one-gene–one-enzyme hypothesis of Beadle & Tatum (1941) got the field of biochemical genetics going. The next step was taken when it was established that genes are nucleic acids. The transformation experiment of Avery, Macleod & McCarty (1944) and the later experiment of Hershey & Chase (1952) with the bacteriophage established this for DNA, and the experiments with tobacco-mosaic-virus RNA (Gierer & Schramm, 1956; Fraenkel-Conrat, Singer & Williams, 1957) in the middle fifties directly proved the same role for RNA. Chemistry and biochemistry of the nucleic acids and study of the interrelations between nucleic acids and proteins were now rapidly gaining momentum, and some of the landmarks that laid firm foundation for the modern structure of molecular biology may now be mentioned.

The structural chemistry of nucleic acids, which developed over a period of some 60 years in many countries, progressed step by step, from the chemistry of purines and pyrimidines, to work on nucleosides and then on to nucleotides. This work reached a climax in 1952 in Professor Todd's Laboratory in Cambridge with the elucidation of the internucleotidic linkage in nucleic acids (Brown & Todd, 1955). Strikingly enough, the next major landmark also came from Cambridge during the following year, with the Watson–Crick structure for DNA (Watson & Crick, 1953). Kornberg's discovery and characterization of the enzyme DNA polymerase (Lehman, Bessman, Simms & Kornberg, 1958) was a major triumph of modern enzymology, and the methods developed distinctly aided the discovery, a few years later, of DNA-dependent RNA polymerase (Weiss, 1960; Stevens, 1961; Burma, Kroger, Ochoa, Warner & Weill, 1961; Hurwitz, Furth, Anders & Evans, 1962; Chamberlin & Berg, 1962). Interest in the biosynthesis of peptide bonds is also of long standing, but critical progress began to be made in the early fifties, and it is particularly satisfying to note that the pioneer in this work was P. Zamecnik, who is a predecessor of mine in the present Lecture series (Zamecnik, 1962). The pace and intensity of work in the field of protein biosynthesis continued to increase during the late fifties, and by 1961 a relatively well-defined cell-free protein-synthesizing system was available. I have given this account to emphasize that our own work, especially during the past 8 years, has been very much inspired and influenced by these discoveries.

With the knowledge of the chemical structures of nucleic acids, the task before chemists in the early fifties was clearly that of synthesis and sequential analysis of nucleic acids. We devoted ourselves to the problems of synthesis, namely learning to put together nucleosides and nucleotides to form short chains of polynucleotides. There was ample justification for developing organic synthesis in this field, especially if one looked at that time at the enormous amount of organic chemical effort in other fields of natural products, such as complex alkaloids, steroids and terpenes, and in carbohydrates and polypeptides, two other classes of naturally occurring macromolecules. So the preoccupation for several years in my Laboratory was to try to develop methods of chemical synthesis.

* Address: Institute for Enzyme Research, University of Wisconsin, Madison, Wis. 53706, U.S.A.

This forms the first phase of our work, although it continues to demand a great deal of our attention even at present.

Concurrently, we pursued enzymic studies, the main purpose here in early studies being to understand the mode of action of different types of nucleases and phosphodiesterases by using oligo- and poly-nucleotides of completely defined structure, so that precise tools for structural analysis of the nucleic acids would be available (Khorana, 1961). In the early sixties, enzymic studies with defined polynucleotides were extended to DNA polymerase and RNA polymerase. Short polydeoxyribonucleotides could indeed be used as templates, and the results appeared to open up new approaches to the synthesis of high-molecular-weight polynucleotides of interest in precise studies of the genetic code. This formed the second phase of our work. In this Lecture I would like to present a summary review of the above-mentioned two phases and then go on to describe in a little more detail the work that is now in progress in my Laboratory. The immediate aim of this activity is the synthesis of a specific DNA duplex such as the gene for a transfer RNA.

Chemical development of polynucleotide synthesis

At the outset there were the two main divisions, the deoxyribo series and the ribo series. Although many of the problems in the two series were inevitably common, attention was first focussed on the deoxyribo series, although their chemistry at that time was much less developed. The first major problem in the field was of chemical activation of the phosphate group in mononucleotides so as to make a phosphodiester or internucleotide bond. An intensive and prolonged survey has given us satisfactory methods. The condensing agents used today are carbodi-imides and sulphonic acid chlorides.

The next large problem was that of finding suitable protecting groups for the different sites, the 5′-hydroxyl group, the 3′-hydroxyl group, the amino groups on the heterocyclic rings and the phosphate group itself. Organic chemists always face the problems of protecting groups in any synthetic undertaking because reactions performed by the chemist lack the specificity of enzymic reactions. Suffice it to say that satisfactory solutions to all of the problems in the nucleotide field now exist (Khorana, Jacob, Moon, Narang & Ohtsuka, 1965).

In the next phase of this work we were preoccupied with the problems of chemical polymerizations of mononucleotides. It was recognized early that, just as in the protein field the availability of polyamino acids prepared by polymerization methods had contributed greatly to the studies of protein structure, similarly polynucleotides, if they could be prepared by chemical polymerization procedures and contained known sequences, would be useful in the nucleic acid field. It was also clear that analytical methods for separation and characterization of short chains of polynucleotides had to be developed. As a review statement on this aspect of the work, it should suffice to say that we have learned to polymerize every one of the mononucleotides, and preformed di- and tri-nucleotide blocks so that homopolynucleotides and polynucleotides containing repeating but defined sequences can be prepared rather rapidly and, indeed, syntheses of most of the repeating polymers that I am going to discuss later in this Lecture have been done by polymerization methods (Khorana *et al.* 1967).

A desired aim in chemical synthesis is the ability to create step-by-step polynucleotides of defined and specific sequences. Two approaches can be imagined: one in which mononucleotides are added one at a time to a growing polynucleotide chain; and the second approach, which is theoretically more attractive, consisting in the preparation of oligonucleotide blocks and then in a subsequent step bringing these together to form longer chains. Both of these approaches have been investigated systematically over the years (see, e.g., Jacob & Khorana, 1965; Kössel, Büchi & Khorana, 1967*a*; Ohtsuka & Khorana, 1967). The approach using blocks is now making headway and has been used recently for synthesis of chains containing up to 20 nucleotide units. One example of this work is given later in this Lecture.

In the RNA field things are even more complicated. The 2′-hydroxyl group in the ribose ring creates an additional formidable problem. Methods had to be developed for specific linkage of the 3′-hydroxyl group of one nucleoside to the 5′-hydroxyl group of the next. Once again time does not permit any serious review, and it is sufficient to say that satisfactory methods now exist for synthesis of short chains of polyribonucleotides. One major accomplishment of work in this field has been the successful and unambiguous synthesis of all of the possible 64 trinucleotides that can be derived from the four mononucleotides (Lohrmann, Söll, Hayatsu, Ohtsuka & Khorana, 1966). The use of these in work on the genetic code will be mentioned later.

Polynucleotide synthesis and the genetic code

A few words about the experimental development of the coding problem are now appropriate. In the fifties, possible rules governing the genetic code engaged the attention of many theoreticians, Gamow

being the first to speculate on the connexion between DNA structure and polypeptide chains (for a historical account see Crick, 1966). However, until 1961 the only experimental approach was that of direct correlation of the sequence of a nucleic acid with that of its protein. It was hoped to do this either chemically, e.g. by working with the protein coat of a virus and its RNA, or by mutagenic techniques. An ingenious application of the 'frame-shift mutation' idea was, indeed, that of Crick, Barnett, Brenner & Watts-Tobin, (1961), who produced evidence for the three-letter property of the genetic code. These approaches, however, offered little immediate hope of getting directly at the coding problem.

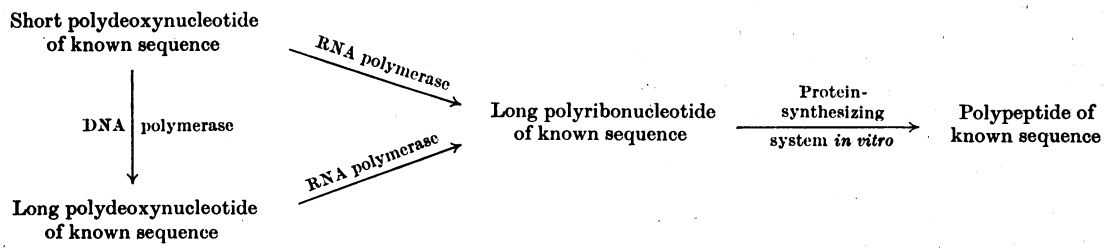
The approach that introduced a new and direct experimental attack on the genetic code was the discovery made by Matthaei & Nirenberg (1961) that polyuridylate directs the synthesis of polyphenylalanine in the bacterial cell-free amino acid-incorporating system. A great deal, in fact, was learned during the years 1961–63, both in the Laboratories of Ochoa and of Nirenberg and their co-workers (Ochoa, 1963; Nirenberg, Matthaei, Jones, Martin & Barondes, 1963), about the overall composition of codons by using polynucleotides made by the agency of polynucleotide phosphorylase, an enzyme discovered some years earlier by Grunberg-Manago and Ochoa. The responsibility for complete elucidation of the genetic code now essentially rested with the chemist. If only a chemist could make a nucleic acid of completely defined structure and analyse the protein specified by it, then one would have a direct correlation of the sequences of the two types of macromolecules. This, then, was a central problem in molecular biology at this time.

Synthesis of polydeoxyribonucleotides with repeating sequences. As mentioned above, we at this time could only make short chains of DNA by chemical methods. On the other hand, both DNA polymerase and RNA polymerase appeared to offer hope of eventually preparing the required high-molecular-weight RNA messengers by the reaction sequence shown in Scheme 1.

The decision to synthesize polydeoxyribonucleotides with repeating nucleotide sequences followed from two important types of considerations: (1) the device of using polymerases for amplification of the short synthetic compounds to high-molecular-weight compounds appeared to be possible only with repeating patterns of sequences; and (2) the cell-free protein-synthesizing system, being essentially a crude extract, contained powerful nucleases, and unequivocal answers could be hoped for only by use of messengers that contained defined but precisely repeating sequences. Finally, the choice of nucleotide combinations in the synthetic polymers was influenced by the important finding that messenger RNA had to be single-stranded. In fact, the only DNA containing more than one type of base and whose sequence was completely known was the high-molecular-weight polymer poly-dAT.* Although RNA polymerase nicely produces from it poly-rAU containing the two bases in strictly alternating sequence, the polynucleotide, because of self-complementarity, has tight double-stranded structure and elicits no response from the ribosomes in the cell-free system. It was therefore clear that those combinations of nucleotides that would lead to overwhelming base-pairing in the polynucleotides should be avoided.

All of the chemical syntheses relevant to the genetic code that have been carried out to date are shown in Table 1. First, we made the two sets of polynucleotides shown on the left, which contained repeating dinucleotide sequences. These are the Watson–Crick complementary sets: one set contains the hexamer of the dinucleotide with alternating thymidylate and guanylate residues and the hexamer of the dinucleotide with alternating

* Abbreviations: in nucleotide sequences, A, C, G, T and U represent residues containing adenine, cytosine, guanine, thymine and uracil respectively, and the prefixes d and r represent deoxyribo and ribo series respectively; t-RNA, transfer RNA.



Scheme 1. Suggested reaction sequence for the preparation of high-molecular-weight RNA messengers for the synthesis of polypeptides of known sequence.

Table 1. *Synthetic polydeoxyribonucleotides with repeating sequences*

Repeating dinucleotide sequences		Repeating trinucleotide sequences				Repeating tetranucleotide sequences	
d[TC]$_6$ / d[AG]$_6$	d[TG]$_6$ / d[AC]$_6$	d[TTC]$_4$ / d[AAG]$_4$	d[CCT]$_{3-5}$ / d[GGA]$_{3-5}$	d[TAC]$_{4-6}$ / d[TAG]$_{4-6}$	d[CCA]$_{3-5}$ / d[GGT]$_{3-5}$	d[TTAC]$_4$ / d[GTAA]$_2$	d[TCTA]$_3$ / d[TAGA]$_2$
		d[TTG]$_{4-6}$ / d[CAA]$_{4-6}$	d[CGA]$_{3-5}$ / d[CGT]$_{3-5}$	d[ATC]$_{3-5}$ / d[ATG]$_{3-5}$			

$$\text{d[TG]}_6 + \text{d[AC]}_6 + \begin{Bmatrix}\text{dTTP}\\ \text{dATP}\\ \text{dCTP}\\ \text{dGTP}\end{Bmatrix} \rightarrow \text{poly-dTG:CA} \qquad (1)$$

$$\text{d[TTC]}_4 + \text{d[AAG]}_3 + \begin{Bmatrix}\text{dTTP}\\ \text{dATP}\\ \text{dCTP}\\ \text{dGTP}\end{Bmatrix} \rightarrow \text{poly-dTTC:GAA} \qquad (2)$$

$$\text{d[TATC]}_3 + \text{d[TAGA]}_2 + \begin{Bmatrix}\text{dTTP}\\ \text{dATP}\\ \text{dCTP}\\ \text{dGTP}\end{Bmatrix} \rightarrow \text{poly-dTATC:GATA} \qquad (3)$$

Scheme 2. Types of reactions catalysed by DNA polymerase. 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.

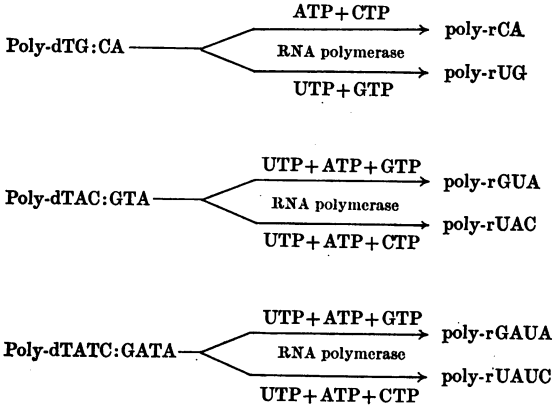
adenylate and cytidylate residues; and the second set consists of the hexamer of alternating thymidylate and cytidylate residues and the hexamer of alternating adenylate and guanylate residues. This work was then extended to polynucleotides with repeating trinucleotide sequences. There is a theoretical maximum of ten such sets that can contain more than one nucleotide base and we have already prepared seven such sets (Jacob & Khorana, 1965; Narang & Khorana, 1965; Narang, Jacob & Khorana, 1965, 1967*a*,*b*). A general point about all of the syntheses is that one has to prepare segments corresponding to both strands of a DNA duplex for the DNA polymerase to catalyse synthesis. Shown also in Table 1 are two sets of polymers with repeating tetranucleotide sequences (Kössel *et al.* 1967*a*; Ohtsuka & Khorana, 1967). In brief, the major considerations for these are that they contain in every fourth place the chain-terminating codons and also that this class of polymers can be used to prove the direction of reading of messenger RNA (Khorana, 1965, 1966; Khorana *et al.* 1966, 1967).

Double-stranded DNA-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 2, a mixture of the two short-chain polynucleotides with repeating dinucleotide sequences directed the extensive synthesis of a double-stranded DNA-like polymer (Scheme 2) containing exactly the sequences present in the short-chain templates (Byrd, Ohtsuka, Moon & Khorana, 1965; Wells, Ohtsuka & Khorana, 1965; Khorana, 1965, 1966; Khorana *et al.* 1966). More recently, the same was true for short chains with repeating trinucleotide sequences, and even more recently also for repeating tetranucleotide sequences (Wells, Jacob, Narang & Khorana, 1967*a*; Wells, Büchi, Kössel, Ohtsuka & Khorana, 1967*b*). Many of the features of these reactions are truly remarkable. Thus: (1) in all these reactions (Scheme 2) the enzyme shows complete fidelity in reproduction of sequences; (2) the synthesis is extensive, 50–200-fold, and the products are of high molecular weight (500000 to over 1000000); (3) the enzyme thus amplifies and multiplies the information created by chemical methods; (4) 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 2 catalogues the different kinds of polymers that have already been prepared in this way and characterized. Thus we have three classes of polymers: two double-stranded polymers with

Table 2. *Synthetic polydeoxyribonucleotides with repeating alternating sequences*

Repeating alternating dinucleotide sequences	Repeating alternating trinucleotide sequences	Repeating alternating tetranucleotide sequences
Poly-dTC:GA	Poly-dTTC:GAA	Poly-dTTAC:GTAA
Poly-dTG:CA	Poly-dTTG:CAA	Poly-dTATC:GATA
	Poly-dTAC:GTA	
	Poly-dATC:GAT	



Scheme 3. Preparation of single-stranded polyribonucleotides from DNA-like polymers containing repeating tri- and tetra-nucleotide sequences.

Table 3. *Synthetic polyribonucleotides 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	

repeating alternating dinucleotide sequences, four polymers with repeating alternating trinucleotide sequences and two polymers with repeating alternating tetranucleotide sequences.

Single-stranded polyribonucleotides with repeating nucleotide sequences. The next step was the transcription of the DNA-like polymers by means of RNA polymerase to form single-stranded polyribonucleotides. The principle used throughout is illustrated in Scheme 3. All of the DNA-like polymers contain two, or a maximum of three, different bases in individual strands. It is therefore possible, by giving only the nucleoside triphosphates required for copying only one strand, to restrict the action of RNA polymerase to that strand. This is the case for all of the polymers, examples of which are shown in Scheme 3. Nearest-neighbour frequency analysis of all of the RNA products again shows that they contain in them strictly repeating nucleotide sequences. The total RNA-like polymers prepared so far are listed in Table 3.

We may now summarize the stage of our work so far. By using a combination of purely chemical methods, which are required to 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 have at our disposal a variety of high-molecular-weight polyribonucleotides of known sequences. Mistake levels, if they occur at all, are insignificant.

Cell-free polypeptide synthesis and the genetic code. Polymers with repeating dinucleotides, $[AB]_n$, contain two triplets, *ABA* and *BAB*, in alternating sequence, and these should direct incorporations of two amino acids in strictly alternating sequence. Repeating trinucleotide polymers, $[ABC]_n$, contain three repeating triplets, depending on 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 repeating tetranucleotide sequences, $[ABCD]_n$, show that four amino acids should go in and repeating tetrapeptide should be the product no matter where the reading starts.

All of these predictions have been fully borne out experimentally without a single exception (Khorana, 1965, 1966; Khorana *et al.* 1966; Kössel, Morgan & Khorana, 1967*b*). Table 4 summarizes the total results obtained with repeating polymers. It consists of three parts: to the left is a set of results with repeating dinucleotides, poly-UC, poly-UG, poly-AC and poly-AG; the next part consists of eight polymers with repeating trinucleotide sequences; and the last part consists of four ribopolynucleotides with repeating tetranucleotide sequences. The following brief comments may be made. (1) Repeating dinucleotide polymers do, in fact, direct incorporation of only two amino acids and the products invariably are the expected co-peptides (Nishimura, Jones & Khorana, 1965*b*; Jones, Nishimura & Khorana, 1966). (2) Repeating trinucleotide polymers have as a rule given three homopolypeptides (Nishimura *et al.* 1965*a*; Morgan,

Table 4. *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	Polypeptides formed	Messenger	Polypeptide formed
Poly-UC	$[Ser\text{-}Leu]_n$	Poly-UUC	$[Phe]_n$, $[Ser]_n$, $[Leu]_n$	Poly-UAUC	$[Tyr\text{-}Leu\text{-}Ser\text{-}Ile]_n$
Poly-AG	$[Arg\text{-}Glu]_n$	Poly-AAG	$[Lys]_n$, $[Glu]_n$, $[Arg]_n$	Poly-UUAC	$[Leu\text{-}Leu\text{-}Thr\text{-}Tyr]_n$
Poly-UG	$[Val\text{-}Cys]_n$	Poly-UUG	$[Cys]_n$, $[Leu]_n$, $[Val]_n$	Poly-GUAA	Di- and tri-peptides
Poly-AC	$[Thr\text{-}His]_n$	Poly-CAA	$[Gln]_n$, $[Thr]_n$, $[Asn]_n$	Poly-AUAG	Di- and tri-peptides
		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')		

Wells & Khorana, 1966), and it should now be emphasized that this was because of the fact that in all of the work with the cell-free system artificially high Mg^{2+} concentration was used and therefore polypeptide chains could start out without a proper initiation signal. We shall return to this point very shortly. With the trinucleotide polymers, therefore, most often three homopeptides are obtained. The exceptions are poly-UAG and poly-AUG. Both of these contain 'nonsense' triplets. UAG is the well-known 'amber' triplet and UGA is now known to be also a 'nonsense' triplet. (3) Finally, repeating tetranucleotide polymers, in fact, direct amino acid incorporations such that the repeating tetrapeptides shown are formed (Kössel *et al.* 1967*b*). This has been proved by analysis for the two products shown. This analysis of the repeating tetrapeptides proves independently that the direction of the reading of the messenger is from the 5′-hydroxyl to the 3′-hydroxyl end. This result is in agreement with those of a number of groups, in particular, from Streisinger and co-workers, Thach and co-workers, Ochoa and co-workers (Crick, 1966) and from Lamfrom, McLaughlin & Sarabhai (1966). The last two polynucleotides shown contain in every fourth place the 'nonsense' triplets, UAG and UAA, and for this reason fail to give any continuous peptides.

What is the significance of these results with regard to the genetic code? We can state the conclusions as follows:

(1) DNA does, in fact, specify the sequence of amino acids in proteins and this information is relayed through an RNA. This is the first time that a direct sequence correlation between DNA and a protein has been accomplished.

(2) All of the results directly prove three-letter non-overlapping properties of the code.

(3) Finally, information on codon assignments can also be derived from these results.

For this large question of codon assignments, however, unless one does a large number of polypeptide syntheses, our individual experiments do not by themselves give unique answers to the codons. For example, of the two triplets, UCU and CUC, which stand for serine and leucine, it is not possible to say which stands for which amino acid. Now the way the code has, in fact, been derived has been by a combination of these results with an elegant technique developed by Nirenberg and Leder. This is the binding technique (Nirenberg & Leder, 1964). The first event in the synthesis of a peptide bond must be the formation of a ternary complex between messenger RNA, the ribosome and aminoacyl-t-RNA. Messenger RNA clearly must select the aminoacyl-t-RNA species from the soluble pool. Several groups of workers between 1962 and 1964 demonstrated the messenger RNA directed binding of aminoacyl-t-RNA to ribosomes. The elegant simplification used by Nirenberg & Leder (1964) was that the messenger may often be as short as a trinucleotide in directing this binding. One therefore looks for the stimulation of the binding of different aminoacyl-t-RNA species to ribosomes in the presence of specific trinucleotides. An example of its use is in Fig. 1 (Nishimura *et al.* 1965*a*).

Of the three sequence isomers AAG, AGA and GAA, the question is: which stands for lysine? One measures the binding of [^{14}C]lysyl-t-RNA to ribosomes in the presence of increasing amounts of these trinucleotides. As shown in Fig. 1, the effect is specific for AAG. The other trinucleotide that gives a similar effect is AAA, which we knew before stands for lysine. AAG and AAA are therefore the codons for lysine. This technique has been used extensively in Nirenberg's Laboratory, and in our own Laboratory we have used all of the 64 synthetic trinucleotides in this type of analysis. Although extremely useful, the technique has not proved to be completely reliable. Often the effects are very

small and there are artifacts. Conversely, there are cases where authentic trinucleotide codons do not give any binding. The way most of the code has actually been worked out is by using this technique in combination with the results from the repeating polymers and often by using evidence from a number of experiments *in vivo*. The structure of the code that has emerged is given in Table 5.

This is by now a familiar method of presentation (Crick, 1966). The first column gives the first letter, the last column gives the third letter and in the middle there is a column for each one of the second letter. Only a few points may be made: (1) the code as shown is for *Escherichia coli* B, but probably will hold essentially for other organisms as well; (2) the code is highly degenerate in a semi-systematic way and there are entries for all of the 64 trinucleotides; (3) in fact, there are two meanings for a few of the codons, e.g. AUG and GUG stand for insertion of amino acids and for chain-initiation signals; (4) there are three 'nonsense' triplets; all of these cause chain termination. It is not known which triplet is used naturally for chain termination. Finally, how much confidence can we have in these assignments? Very probably all of this can be regarded as established. The most encouraging feature is that there is really no conflict with regard to any assignment between the different lines of evidence. Various portions of this code have been derived also from experiments *in vivo* by Yanofsky and co-workers, by Streisinger and co-workers, by Wittmann and Tsugita and by others (Crick, 1966).

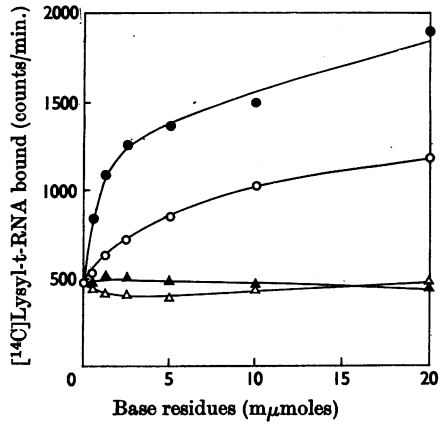


Fig. 1. Stimulation of the binding of [14C]lysyl-t-RNA to ribosomes by trinucleotides: ●, AAA; ○, AAG; ▲, AGA; △, GAA.

Synthetic polynucleotides with known repeating codons have been used in definitive studies of many other aspects of the genetic code and of protein synthesis. I can only list some of the findings. In studies of the initiation signals for protein synthesis, the codons AUG and GUG have been shown to recognize the initiator, formylmethionyl-t-RNA (Ghosh, Söll & Khorana, 1967). A further study of polymethionine synthesis by using poly-AUG has

Table 5. *The genetic code as established for* E. coli

The generally accepted convention is used in showing trinucleotide sequences. Thus in the trinucleotide codons for phenylalanine, UUU and UUC, the first letter is at the left (the 5′-hydroxyl end) and U or C are the third letters at the 3′-hydroxyl end. C.I. stands for chain initiation; abbreviations for amino acids are standard.

		Genetic code				
First letter	Second letter... ...	U	C	A	G	Third letter
U		Phe	Ser	Tyr	Cys	U
U		Phe	Ser	Tyr	Cys	C
U		Leu	Ser	'Nonsense'	'Nonsense'	A
U		Leu	Ser	'Nonsense'	Trp	G
C		Leu	Pro	His	Arg	U
C		Leu	Pro	His	Arg	C
C		Leu	Pro	Gln	Arg	A
C		Leu	Pro	Gln	Arg	G
A		Ile	Thr	Asn	Ser	U
A		Ile	Thr	Asn	Ser	C
A		Ile	Thr	Lys	Arg	A
A		Met. (C.I.)	Thr	Lys	Arg	G
G		Val	Ala	Asp	Gly	U
G		Val	Ala	Asp	Gly	C
G		Val	Ala	Glu	Gly	A
G		Val (C.I.)	Ala	Glu	Gly	G

given direct demonstration of two aminoacyl-t-RNA-binding sites on 70s ribosomes and has shed light on the role of 30 and 50s sub-units in the initiation of protein synthesis (Ghosh & Khorana, 1967). In extended studies of codon–anti-codon recognition (Söll *et al.* 1966*a*,*b*; Söll, Cherayil & Bock, 1967; Söll & RajBhandary, 1967) patterns for multiple recognition of codons by one t-RNA species have been largely elucidated. The results obtained with repeating tetranucleotide polymers have provided further independent evidence on the direction of reading of messengers (Kössel *et al.* 1967*b*). Other aspects of protein synthesis studied have included 'missense' suppression (Gupta & Khorana, 1966), misreading of codons provoked by certain aminoglycoside antibiotics (Davies, Jones & Khorana, 1966) and the direct translation of single-stranded DNA in the presence of certain antibiotics such as neomycin B (Morgan, Wells & Khorana, 1967).

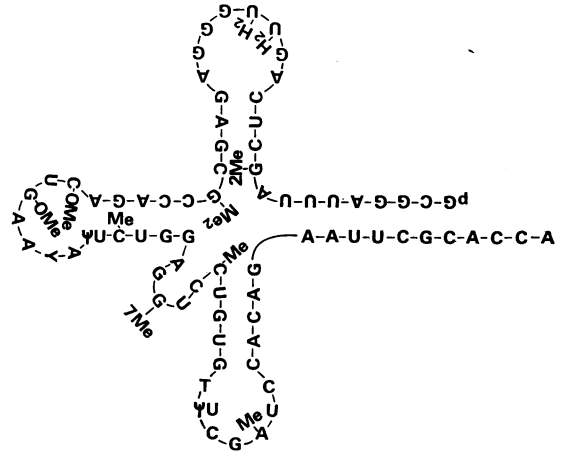


Fig. 2. Cloverleaf model for the secondary structure of t-RNA, as exemplified by the nucleotide sequence of yeast phenylalanine t-RNA. ψU, Pseudouridine; Y, a nucleoside that remains unidentified.

Current work on the synthesis of the gene for a transfer RNA

I would now like to turn to the next part of this Lecture, which reviews our current interest in polynucleotide synthesis.

Although the high-molecular-weight DNA-like polymers with repeating nucleotide sequences have been useful for studies of the genetic code and continue to be used for a variety of physicochemical and enzymic studies of nucleic acids, it has been clear to us for some time that their use is 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 know, for example, what the initiation and termination signals for RNA polymerase are, what kind of sequences are recognized by repressors, by host modification and host restrictive enzymes and by enzymes involved in genetic recombination and so on. For these studies eventually what is required is the ability to synthesize long chains of DNA with specific non-repeating sequences. With this should come the ability to 'manipulate' DNA for different types of studies. We therefore concluded about 3 years ago that, in continuing our interest in polynucleotide synthesis, the next long-range aim must be the development of methods for the total synthesis of biologically specific DNA duplexes. I would at this point interject perhaps one obvious point about the remarkable recent accomplishment of Arthur Kornberg in replication of single-stranded DNA and, similarly, the earlier accomplishment of Spiegelman with viral RNA. Landmarks as these accomplishments are in this field, it is important to remember that in these studies these workers are taking the informational strand and using a purified polymerase to replicate this strand. On the other hand, the objective of our total synthesis would be to put together all of the information by synthesis. It is clear that eventually our ability to manipulate the information content of nucleic acids, which we equate with their sequence, is dependent on our ability to put together the information by chemical synthesis.

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. Thus the general functions of t-RNA are clearly established. These molecules have to be recognized by a rather large number of components of the protein-synthesizing machinery, such as by the pyrophosphorylase that repairs the CCA end, by the aminoacyl-t-RNA synthetases, by ribosomes and by messenger RNA. Further, the t-RNA species are a unique class of molecules, possessing attributes of nucleic acids and proteins. There is a good deal of evidence now to suggest that, in addition to a common secondary structure (cloverleaf model; Fig. 2), these molecules possess a tertiary structure. Also, all t-RNA species abound in minor bases, which are largely found in certain non-hydrogen-bonded regions. It is entirely possible that a good part of the evolution of the genetic code is synonymous with the evolution of t-RNA 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 is clear that chemical synthesis, provided that it could be developed to the point that one can manipulate different parts of the t-RNA structures,

would open up a definitive approach of wide scope. For example, one could have deletions in different parts, one could take an anti-codon loop from a t-RNA specific for one amino acid and replace it with the anti-codon loop from another t-RNA.

It should be further emphasized that our ability to synthesize DNA with known specific sequences will inevitably involve us in the problem of DNA replication—for one would hope that, having once produced a synthetic gene, one would be able to assure its permanent availability by enzymic multiplication. Further, we would clearly have to study the process of transcription of this gene *in vitro* and this would therefore be a precise system for initiation of RNA polymerase action.

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 synthesis 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 have been wanting to exploit is the template principle, i.e. the ability of polynucleotides to form hydrogen-bonded bihelical structures. We want to make short pieces of DNA that would correspond to segments of a double-stranded DNA structure and, when these are properly annealed to form bihelical complexes, to try to join them by either chemical or enzymic methods. It does not matter if we have to make a much larger number of short pieces. This is by far preferable to trying to put together a long but single-stranded structure. Mistakes would accumulate and the currently available methods of separation and purification for polynucleotides, even for chains of 20–50 units long, are hopelessly inadequate.

Chemical synthesis of polydeoxyribonucleotides. In initiating synthetic work, our decision was to set out to synthesize eicosanucleotides with the sequences shown in Fig. 3. These are about the maximum chain lengths that current methods of synthesis, or rather separation, would allow. The points to be made about the two eicosanucleotides shown in Fig. 3 are as follows. (1) They span the length of nucleotide residues 21–50 of yeast alanine t-RNA (Holley *et al.* 1965). Yeast alanine t-RNA was the only t-RNA whose primary sequence was known at the time the present work was undertaken. (2) The principal assumption in defining the DNA sequence complementary to the t-RNA sequence has been made that all the minor bases are produced by modification of the four parent bases and that these modifications occur after transcription of the DNA gene with the four standard bases. Thus inosine is formed by deamination of adenosine and hence it originates from an A–T base pair in DNA. Similarly pseudouridine (5-ribosyluracil) and di-

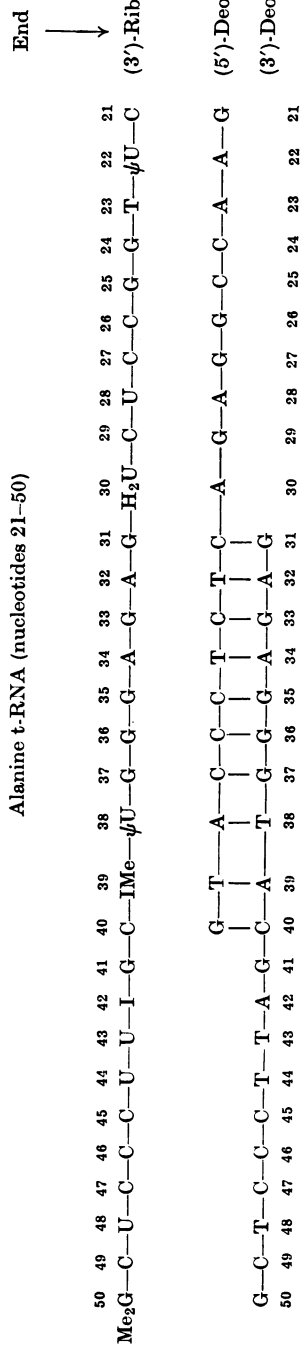
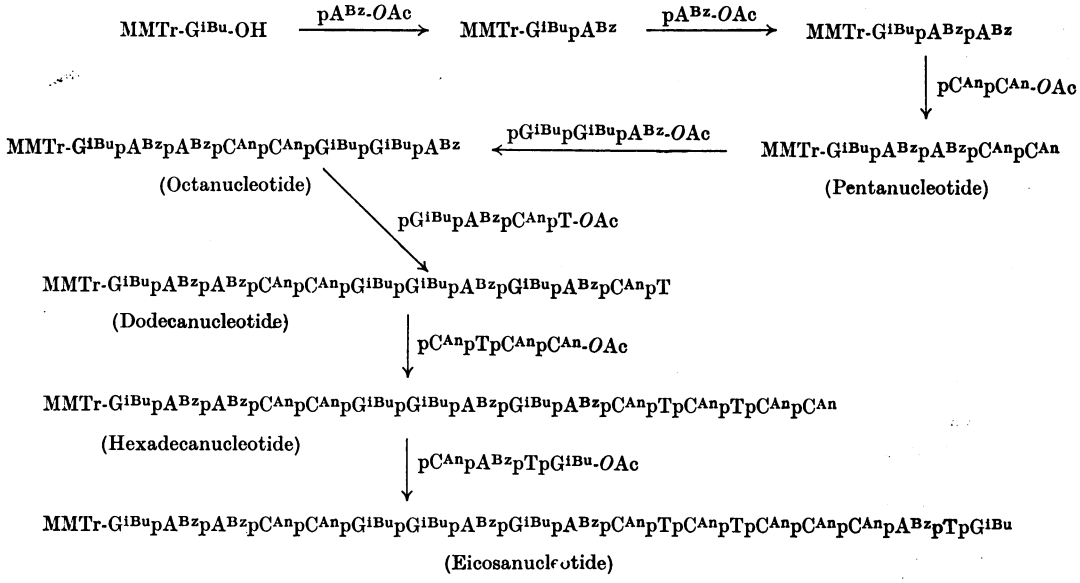


Fig. 3. Polydeoxyribonucleotide sequences complementary to yeast alanine t-RNA. I, Inosine; ψU, pseudouridine.

hydrouridine both originate from uridine and hence correspond to an A–T base pair. (3) The third point about the eicosanucleotides in Fig. 3 is that they are complementary throughout half of their length and are of opposite polarity. The complementary region of ten base pairs was expected to be sufficiently long to allow the eicosanucleotides to align themselves as shown. One possibility considered for enzymic work was that the DNA polymerase of *E. coli* might repair these structures to complete a duplex of 30 nucleotide units (Khorana, 1967). If this failed, then further chemically synthesized polydeoxyribonucleotides with sequences complementary to the single-stranded arms in Fig. 3 may be annealed and the pieces joined end to end. The joining reaction in aqueous solution had been studied in this Laboratory several years ago by P. T. Gilham (Naylor & Gilham, 1966).

As an illustration of the chemical methodology in current use, the stepwise synthesis of the eicosanucleotide complementary in sequence to nucleotides 21–40 (Fig. 3) of alanine t-RNA may be reviewed. The steps used are shown in Scheme 4. At the bottom is the ultimate product of synthesis shown with all the protecting groups that were present throughout the duration of synthesis. The synthesis starts from the 5′-hydroxyl end of the polynucleotide and the first phosphodiester bond synthesis involves a condensation between the protected deoxyguanosine derivative *N*-isobutyryl-5′-monomethyltrityldeoxyguanosine (dMMTr-G^{iBu}) and a protected mononucleotide 3′-*O*-acetyl-*N*-benzoyladenosine 5′-phosphate (dpA^{Bz}-*O*Ac). Then follow successive condensations between blocks of protected di-, tri- and later tetra-nucleotides and the 3′-hydroxyl end of the growing fully protected polynucleotide chain. At each step, the products are separated by prolonged anion-exchange chromatography and then checked for purity by extensive paper chromatography. The yields tend to decline as the chain lengths increase, and larger and larger excess of the incoming blocks must be used and even then the yields of desired products are only moderate in the final stages. These syntheses clearly represent the most demanding and time-consuming part of the total project. Synthesis of every polynucleotide requires careful planning with respect to the choice of blocks. The above example will suffice to illustrate the general principles. The total list of synthetic oligo- and poly-nucleotides that have been made so far is given in Table 6.

Initial studies with DNA-joining enzymes. During the past year several groups of workers (Zimmerman,



Scheme 4. Steps in the blockwise chemical synthesis of an eicosadeoxyribonucleotide. The standard method of presentation of polynucleotide chains is used, with the letter p to the left of a nucleoside symbol indicating a 5′-phosphate and to the right indicating a 3′-phosphate. MMTr indicates a monomethoxytrityl group present at the 5′-hydroxyl end of the terminal nucleoside. The protecting groups on the heterocyclic rings of different deoxyribonucleosides are shown by superscripts on the nucleoside initials: An, anisoyl; Bz, benzoyl; iBu, isobutyryl; *O*Ac at the right-hand end of oligonucleotides stands for 3′-*O*-acetyl.

Little, Oshinsky & Gellert, 1967; Weiss & Richardson, 1967; Olivera & Lehman, 1967; Gefter, Becker & Hurwitz, 1967; Cozzarelli, Melechen, Jovin & Kornberg, 1967) reported the discovery of two distinct enzymes that bring about repair of single-stranded breaks in double-stranded DNA. The overall reaction catalysed by these enzymes is shown in Scheme 5.

The enzymes could clearly be a great boon to our projected DNA studies, and our first preoccupation with these enzymes was therefore to determine the minimum chain length of complementary polynucleotides that these enzymes would require in order to bring about the joining reaction. Two sets of studies were carried out (Gupta, Ohtsuka, Weber, Chang & Khorana, 1968) and the combinations of polynucleotides used are listed in Table 7. The first set of experiments (part A of Table 7) showed that a chain as short as a hexanucleotide, e.g. $[^{32}P]d[TG]_3$, in combination with the complementary long polydeoxyribonucleotide, e.g. poly-dCA, can be joined end to end to form a very long chain (Gupta *et al.* 1968). Even more relevant were the results (part B of Table 7) with the systems in which both complementary chains were short. Thus two molecules of the octanucleotide $[^{32}P]d[TAAG]_2$ could be joined end to end to form the corresponding hexadecanucleotide in the presence of the complementary $d[TTAC]_4$. Further, a direct correlation of the temperature requirement of the reaction and the stability of the bihelical complex between the components also turned up. Thus under the conditions of the enzyme reaction (6mM-Mg^{2+} + 10mM-tris chloride), $[^{32}P]d[TG]_3$ + poly-dCA showed T_m about 35° and the optimum temperature for the joining reaction was about 20°. The T_m of $d[TTAC]_4 + [^{32}P]d[TAAG]_2$ is about 20° and the optimum temperature for the joining reaction was 5–10°. It is therefore concluded that the joining reaction would occur so long as the component polynucleotides are present entirely or largely in bihelical structures.

Joining of polydeoxyribonucleotide segments corresponding to t-RNA gene. We were now in a position to study the joining of the synthetic fragments with

Table 6. *Oligo- and poly-deoxyribonucleotides prepared by chemical synthesis*

Eicosanucleotides and intermediate sizes from 5′-hydroxyl end
G-A-A-C-C-G-G-A-G-A-C-T-C-T-C-C-C-A-T-G (I)
G-C-T-C-C-C-T-T-A-G-C-A-T-G-G-G-A-G-A-G (II)

Dodecanucleotide and intermediate sizes from 5′-hydroxyl end
T-G-G-T-G-G-A-C-G-A-G-T

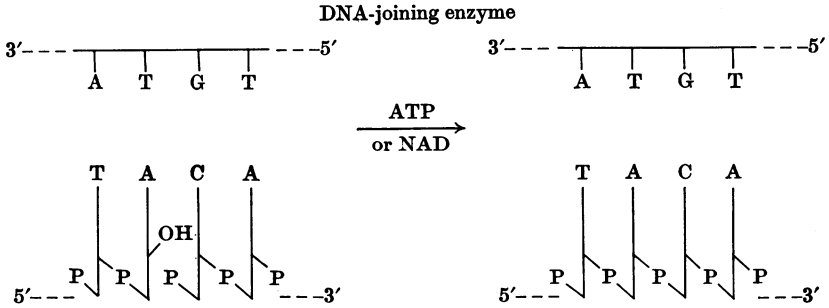
Undecanucleotide and intermediate sizes from 5′-hydroxyl end
C-C-G-G-T-T-C-G-A-T-T

Nonanucleotides and intermediate sizes from 5′-hydroxyl end
T-C-T-C-C-G-G-T-T
C-T-A-A-G-G-G-A-G

Heptanucleotide
A-G-A-G-T-C-T

Table 7. *Combinations of oligo- and poly-deoxyribonucleotides used in the study of DNA-joining enzymes*

	Combinations of one short polynucleotide + one long polynucleotide	
A.	Short	Complementary long
	$d[TG]_{3-5}$	Poly-dCA
	$d[CA]_{3-5}$	Poly-dTG
	$d[T]_{6-10}$	Poly-dA
	$d[A]_{10-15}$	Poly-dT
		Complementary short
B.	Short	but longer
	$d[TAAG]_2$	$d[TTAC]_4$
	$d[AG]_4$	$d[TC]_8$



Scheme 5. Catalysis by DNA-joining enzymes of the formation of a phosphodiester bond at the site of a single-strand break in a duplex DNA molecule.

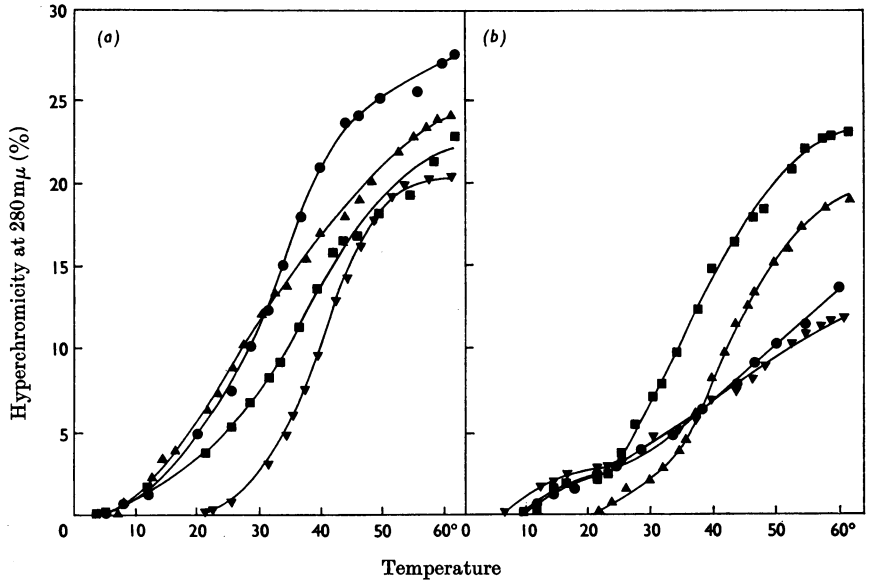


Fig. 4. Temperature–extinction profiles of different combinations of polydeoxyribonucleotides in 10mM-tris–HCl buffer, pH7·6, and 6mM-$MgCl_2$: (*a*) ■, eicosanucleotide I+5′-phosphorylnonanucleotide; ▲, eicosanucleotide I+5′-phosphorylheptanucleotide; ●, decanucleotide+heptanucleotide; ▼, decanucleotide+nonanucleotide; (*b*) ■, eicosanucleotide I+eicosanucleotide II+5′-phosphorylheptanucleotide; ▲, eicosanucleotide I+eicosanucleotide II+5′-phosphorylnonanucleotide; ●, eicosanucleotide I+eicosanucleotide II; ▼, eicosanucleotide I [where eicosanucleotide I and eicosanucleotide II are as described in Table 6, decanucleotide is the sequence GAACCGGAGA, nonanucleotide is the sequence TCTCCGGTT and heptanucleotide is the sequence TCTCCGG, shown in Fig. 5].

specific sequences. Because of the specificity in sequences, in contrast with the simpler systems with repeating sequences discussed above, it was necessary first to make certain that the appropriate single-stranded pieces were annealed to form a bihelical complex under the conditions of the enzymic reaction. No truly relevant study of rates of annealing of very short chains with specific sequences or a correlation of the T_m values with chain length was available, and, further, our primary concern was to form bihelical structures under the conditions of the enzymic reaction. (Stabilization by the customary addition of a salt such as sodium chloride was out of the question because of the observed strong inhibition of the enzymes in the presence of even low concentrations of salt.) Results of the study of temperature–extinction profiles are shown in Fig. 4. Fortunately, the stability of the short complementary structures was high, higher than one might have expected. Thus a combination of the decanucleotide and complementary nonanucleotide showed T_m about 40°. Further, it is interesting that the two eicosanucleotides alone were apparently reluctant to form a double-helical structure, and this is presumably due to the destabilization effect of the single-stranded arms. However, when a nona- or a hepta-nucleotide was added to fill in one of the arms, the formation of a stable complex 'melting' at about 42° was observed.

The combinations of polynucleotides so far tested successfully are shown in Fig. 5. Two types of experiments have been carried out. In the first type, the joining of one short piece, e.g. [^{32}P]-CTAAGGG, is tested in the presence of the two eicosanucleotides. The ^{32}P label is present only at the 5′-hydroxyl end of this component and the assay involves conversion of this label into a phosphatase-insensitive form. The reaction is analysed by chromatography on DEAE-cellulose paper strips. An example is shown in Fig. 6. In the control (bottom part), from which the ligase was omitted, all the label is sensitive to phosphatase and moves as inorganic phosphate. After the ligase reaction, most of the label now stays at the origin as part of a polynucleotide. In the second type of experiments, short pieces corresponding to both of the arms of the complex between the two eicosanucleotides (Fig. 5) can be present simultaneously or they may be added sequentially, after it is made certain by phosphatase assay that the first piece has been quantitatively joined.

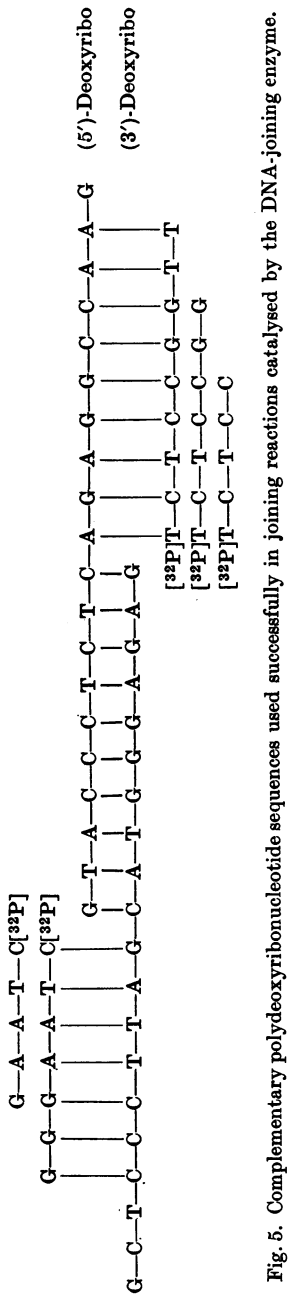


Fig. 5. Complementary polydeoxyribonucleotide sequences used successfully in joining reactions catalysed by the DNA-joining enzyme.

Our results to date show that pieces as short as the pentanucleotides shown in Fig. 6 can be joined, the reaction being carried out at 5° over a period of some hours. A comparison of the *E. coli* and T4-phage DNA-joining enzymes showed that the latter is more effective with the shorter chains. A comparative experiment is shown in Fig. 7. Although the reaction occurred quite well with both enzymes when a nonanucleotide component was tested, with the heptanucleotide very little reaction occurred with the *E. coli* enzyme. All our more recent work has been carried out with the T4-phage-induced enzyme.

It is important to emphasize at this stage that the joining reaction, as measured initially by the phosphatase-insensitive assay, is then followed by degradation of the ligase product to 3′-nucleotides and determination of the ^{32}P in the resulting nucleotides. We know where the ^{32}P is to begin with and where it should be after degradation of the joined product to 3′-nucleotides. All of our assays have been supported fully by this type of analysis.

Repair reactions with DNA polymerase. Shown in Fig. 8 are three of the short DNA molecules that are the products of the joining reactions described above. The nucleotide residues shown within 'boxes' are actually lacking in the joining-enzyme products. Our next concern in the present work has been to use DNA polymerase under repair conditions such that perfect completion of the duplexes shown in Fig. 8 may be achieved, i.e. the nucleotide units shown within the 'boxes' may specifically be introduced. In this way DNA polymerase could serve as an important tool for further characterization of the products. As the duplex chains are elongated step by step, there will always be a protruding ('sticky') end and in every alternate step the strand with the 3′-hydroxyl end will be the shorter one. At the latter steps, the products could be further characterized by testing for the incorporations of mononucleotides by DNA polymerase under conditions of repair reaction. Thus there would be three lines of evidence to show that the enzymic joining of, say, decanucleotide blocks to a growing duplex was under control at every step: (1) the incorporation of terminal [^{32}P]-phosphomonoester group into a phosphatase-insensitive form; (2) the nearest-neighbour analysis of the product formed; and (3) limited but specific incorporation of expected nucleotides at the ends catalysed by DNA polymerase under carefully controlled repair conditions. The results now to be reviewed show that DNA polymerase of *E. coli* can be used in a reliable way for the indicated purpose.

The conditions used for the present DNA polymerase work are those recently used by R. Wu & A. D. Kaiser (personal communication) in their

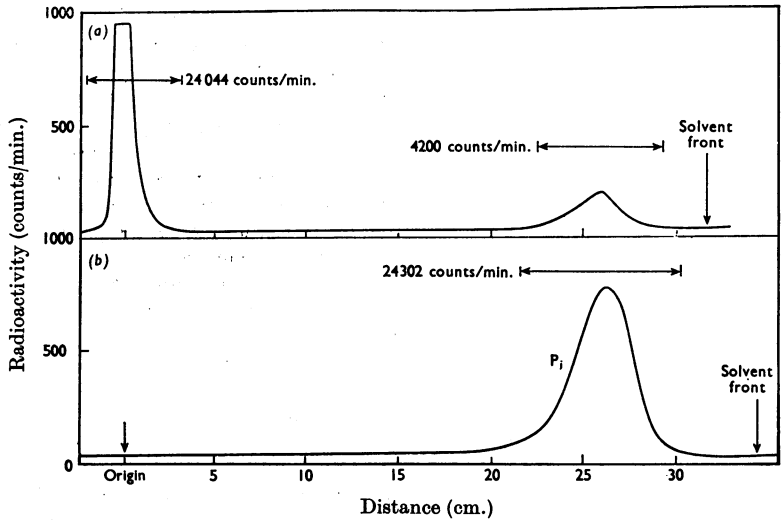


Fig. 6. DEAE-cellulose paper-strip assay of the reaction catalysed by the DNA-joining enzyme: (*a*) eicosanucleotide I+eicosanucleotide II+[^{32}P]CTAAGGG+T4-phage-induced enzyme; (*b*) [^{32}P]CTAAGGG+phosphatase. Details are given in the text.

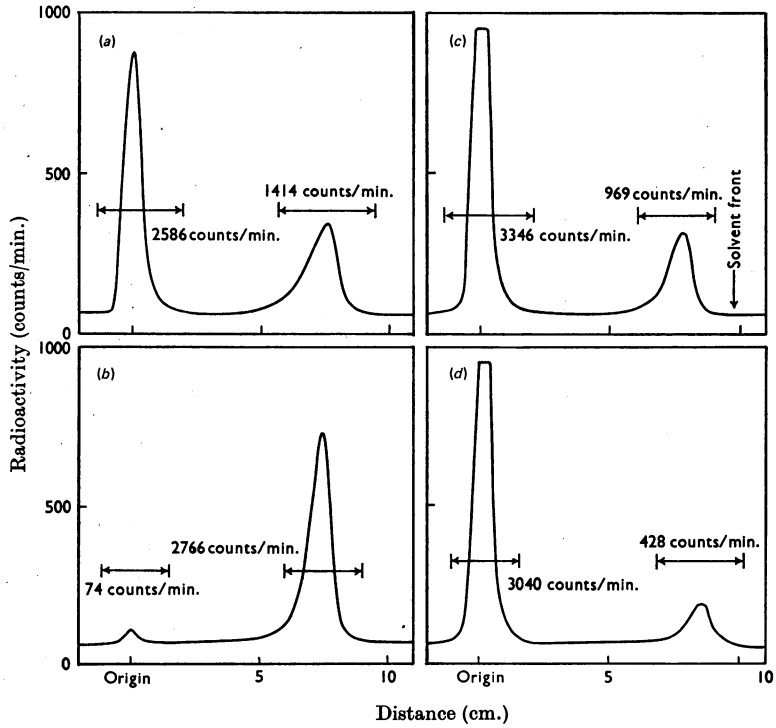


Fig. 7. DEAE-cellulose paper-strip assays of reactions catalysed by the DNA-joining enzymes: comparison of *E. coli* and T4-phage-induced enzymes: (*a*) eicosanucleotide I+eicosanucleotide II+[^{32}P]nonanucleotide (TCTCCGGTT)+*E. coli* enzyme; (*b*) eicosanucleotide I+eicosanucleotide II+[^{32}P]heptanucleotide(TCTCCGG)+*E. coli* enzyme; (*c*) as for (*a*), but with T4-phage-induced enzyme; (*d*) as for (*b*), but with T4-phage-induced enzyme.

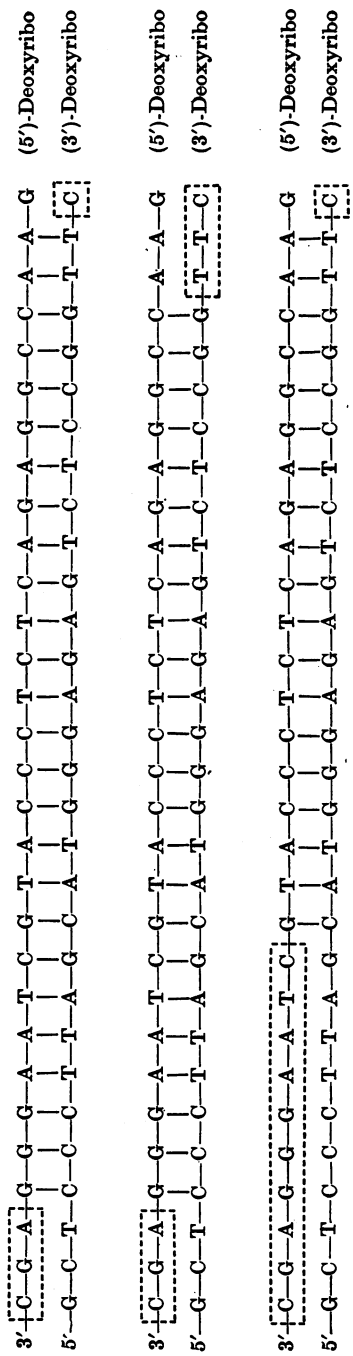


Fig. 8. DNA duplexes prepared by enzymic joining of chemically synthesized polydeoxyribonucleotides (at 5° at pH 6·9).

repair work with the 'sticky' ends of λDNA. The DNA shown on top of Fig. 8 lacks just one deoxycytidylate residue at one terminus and the deoxyadenylate, deoxyguanylate and deoxycytidylate residues at the opposite terminus. [^{3}H]dCTP when used singly was, in fact, incorporated, and, when this labelled nucleotide was used in the presence of unlabelled dATP and dGTP, approximately twice the incorporation of [^{3}H]dCTP was observed. Upon degradation with micrococcal nuclease + spleen phosphodiesterase, all of the radioactivity was released in the form of deoxycytidine and none was found in the corresponding nucleotide. Similarly, [^{3}H]dATP could be incorporated and subsequent degradation released all of the radioactivity as deoxyadenosine. When [^{3}H]dATP was used together with unlabelled dCTP and dGTP and the product now degraded to 3′-nucleotides, most of the radioactivity was now released as deoxyadenosine 3′-phosphate, showing that the incorporation of dATP was followed by that of other nucleotides.

Only two of the several experiments carried out with the DNA species shown in the middle and bottom of Fig. 8 may be mentioned. When [α-^{32}P]dTTP + unlabelled dATP + dGTP + dCTP were used to repair the DNA in the middle of Fig. 8 and the product degraded to 3′-nucleotides, approximately half of the radioactivity was in thymidine 3′-phosphate and the other half was in deoxyguanosine 3′-phosphate; none was found in the other two nucleotides. A similar experiment ([α-^{32}P]dTTP + three unlabelled triphosphates) with the DNA at the bottom of Fig. 8 was carried out and the product again degraded to 3′-nucleotides. All of the radioactivity was found to be in deoxycytidine 3′-phosphate. This experiment is the most definitive in showing: (1) that DNA polymerase catalysed strictly repair reaction under the conditions used; and (2) that exonuclease II activity of the polymerase was completely non-functional under these conditions. If the latter activity were functioning, the first nucleotide to be lost from the bottom right terminus would have been thymidylate. Reinsertion of [α-^{32}P]dTTP at this site by DNA polymerase would have given a product that on subsequent degradation would have released radioactivity in thymidine 3′-phosphate. No radioactivity was, however, detected in this nucleotide.

Projections into the future. The progress to date provides an encouraging start on the complex problem of total synthesis of long DNA duplexes. Thus short chemically synthesized polydeoxyribonucleotides may be added in a stepwise fashion to a developing duplex by using the device of always providing a 'sticky' end. Thus, in some of the work immediately in hand, syntheses of a decanucleotide

complementary to the alanine t-RNA sequence 46–55 and of another decanucleotide (complement to the complement of t-RNA sequence) covering the sequence 51–60 nucleotides are in progress and the total job of synthesis of this gene can now be precisely planned.

Having accomplished the total synthesis of the DNA duplex, replication with a view to multiplication of the synthetic gene would clearly be the next step. Indeed, we would be involved in further studies of the replication, *in vitro*, of bihelical DNA. Once this is done, the process of transcription can be studied systematically and intensively. It is also clear that both ends of the synthetic gene may be extended by the same general principle to include possible initiation and termination signals for the RNA polymerase. Similarly, studies of repression *in vitro* could probably also be brought within the scope of the present studies.

For the first time, concrete plans for addition of the synthetic gene to a genome such as λDNA and the subsequent introduction of the synthetic gene-bearing λDNA into bacterial cells can be made. Similarly the way is clear, in principle, for preparation of synthetic circular (covalently closed) DNA. Although all these possibilities belong to the future, the present results nevertheless would appear to give an encouraging start.

A large number of devoted chemists and biochemists have participated in the studies that have formed the content of this Lecture. I would like to mention in particular the names of the colleagues who are engaged in the current work on the chemical and enzymic methods for the synthesis of the gene for alanine t-RNA: H. Büchi, M. H. Caruthers, N. Gupta, A. Kumar, E. Ohtsuka, V. Sgaramella and H. Weber. I cannot adequately express my indebtedness to these colleagues. In the fifties the work was generously supported by the National Research Council of Canada. In more recent years the work has been and continues to be supported by the National Science Foundation, Washington, by the National Cancer Institute of the National Institutes of Health, U.S. Public Health Service, by the Life Insurance Medical Research Fund and by the Graduate School of the University of Wisconsin.

REFERENCES

Avery, O. T., Macleod, C. M. & McCarty, M. (1944). *J. exp. Med.* **79**, 137.

Beadle, G. W. & Tatum, E. L. (1941). *Proc. nat. Acad. Sci., Wash.*, **27**, 499.

Brown, D. M. & Todd, A. R. (1955). In *Nucleic Acids*, vol. 1, p. 409. Ed. by Chargaff, E. New York: Academic Press Inc.

Burma, D. P., Kroger, H., Ochoa, S., Warner, R. C. & Weill, J. D. (1961). *Proc. nat. Acad. Sci., Wash.*, **47**, 749.

Byrd, C., Ohtsuka, E., Moon, M. W. & Khorana, H. G. (1965). *Proc. nat. Acad. Sci., Wash.*, **53**, 79.

Chamberlin, M. & Berg, P. (1962). *Proc. nat. Acad. Sci., Wash.*, **48**, 81.

Cozzarelli, N. R., Melechen, N. E., Jovin, T. M. & Kornberg, A. (1967). *Biochem. biophys. Res. Commun.* **28**, 578.

Crick, F. H. C. (1966). *Cold Spr. Harb. Symp. quant. Biol.* **31**, 3.

Crick, F. H. C., Barnett, L., Brenner, S. & Watts-Tobin, R. J. (1961). *Nature, Lond.*, **192**, 1227.

Davies, J., Jones, D. S. & Khorana, H. G. (1966). *J. molec. Biol.* **18**, 48.

Fraenkel-Conrat, H., Singer, B. & Williams, R. C. (1957). *Biochim. biophys. Acta*, **25**, 87.

Gefter, M. L., Becker, A. & Hurwitz, J. (1967). *Proc. nat. Acad. Sci., Wash.*, **58**, 240.

Ghosh, H. P. & Khorana, H. G. (1967). *Proc. nat. Acad. Sci., Wash.*, **58**, 2455.

Ghosh, H. P., Söll, D. & Khorana, H. G. (1967). *J. molec. Biol.* **25**, 275.

Gierer, A. & Schramm, G. (1956). *Nature, Lond.*, **177**, 702.

Gupta, N. K. & Khorana, H. G. (1966). *Proc. nat. Acad. Sci., Wash.*, **56**, 772.

Gupta, N. K., Ohtsuka, E., Weber, H., Chang, S. H. & Khorana, H. G. (1968). *Proc. nat. Acad. Sci., Wash.*, **60**, 285.

Hershey, A. D. & Chase, M. (1952). *J. gen. Physiol.* **36**, 39.

Holley, R. W., Apgar, J., Everett, G. A., Madison, J. T., Marquisee, M., Merrill, S. H., Penswick, J. R. & Zamir, A. (1965). *Science*, **147**, 1462.

Hurwitz, J., Furth, J. J., Anders, M. & Evans, A. (1962). *J. biol. Chem.* **237**, 3752.

Jacob, T. M. & Khorana, H. G. (1965). *J. Amer. chem. Soc.* **87**, 2971.

Jones, D. S., Nishimura, S. & Khorana, H. G. (1966). *J. molec. Biol.* **16**, 454.

Khorana, H. G. (1961). In *The Enzymes*, 2nd ed., vol. 5, p. 79. Ed. by Boyer, P. D., Lardy, H. & Myrbäck, K. New York: Academic Press Inc.

Khorana, H. G. (1965). *Fed. Proc.* **24**, 1473.

Khorana, H. G. (1966). In *Symp. Fed. Europ. biochem. Soc., Warsaw: Genetic Elements: Properties and Function*, p. 209. Ed. by Shugar, D. London: Academic Press (Inc.) Ltd.

Khorana, H. G. (1967). *Proc. 7th int. Congr. Biochem., Tokyo*, p. 17.

Khorana, H. G., Büchi, H., Ghosh, H., Gupta, N., Jacob, T. M., Kössel, H., Morgan, A. R., Narang, S. A., Ohtsuka, E. & Wells, R. D. (1966). *Cold Spr. Harb. Symp. quant. Biol.* **31**, 39.

Khorana, H. G., Büchi, H., Jacob, T. M., Kössel, H., Narang, S. A. & Ohtsuka, E. (1967). *J. Amer. chem. Soc.* **89**, 2154.

Khorana, H. G., Jacob, T. M., Moon, M. W., Narang, S. A. & Ohtsuka, E. (1965). *J. Amer. chem. Soc.* **87**, 2954.

Kössel, H., Büchi, H. & Khorana, H. G. (1967*a*). *J. Amer. chem. Soc.* **89**, 2185.

Kössel, H., Morgan, A. R. & Khorana, H. G. (1967*b*). *J. molec. Biol.* **26**, 449.

Lamfrom, H., McLaughlin, C. S. & Sarabhai, A. (1966). *J. molec. Biol.* **22**, 359.

Lehman, I. R., Bessman, M. J., Simms, E. S. & Kornberg, A. (1958). *J. biol. Chem.* **233**, 163.

Lohrmann, R., Söll, D., Hayatsu, H., Ohtsuka, E. & Khorana, H. G. (1966). *J. Amer. chem. Soc.* **88**, 819.

Matthaei, J. H. & Nirenberg, M. W. (1961). *Proc. nat. Acad. Sci., Wash.*, **47**, 1580.

Morgan, A. R., Wells, R. D. & Khorana, H. G. (1966). *Proc. nat. Acad. Sci., Wash.*, **56**, 1899.
Morgan, A. R., Wells, R. D. & Khorana, H. G. (1967). *J. molec. Biol.* **26**, 477.
Narang, S. A., Jacob, T. M. & Khorana, H. G. (1965). *J. Amer. chem. Soc.* **87**, 2988.
Narang, S. A., Jacob, T. M. & Khorana, H. G. (1967*a*). *J. Amer. chem. Soc.* **89**, 2158.
Narang, S. A., Jacob, T. M. & Khorana, H. G. (1967*b*). *J. Amer. chem. Soc.* **89**, 2167.
Narang, S. A. & Khorana, H. G. (1965). *J. Amer. chem. Soc.* **87**, 2981.
Naylor, R. & Gilham, P. T. (1966). *Biochemistry*, **5**, 2722.
Nirenberg, M. W. & Leder, P. (1964). *Science*, **145**, 1399.
Nirenberg, M. W., Matthaei, J. H., Jones, O. W., Martin, R. G. & Barondes, S. H. (1963). *Fed. Proc.* **22**, 55.
Nishimura, S., Jones, D. S. & Khorana, H. G. (1965*b*). *J. molec. Biol.* **13**, 302.
Nishimura, S., Jones, D. S., Ohtsuka, E., Hayatsu, H., Jacob, T. M. & Khorana, H. G. (1965*a*). *J. molec. Biol.* **13**, 283.
Ochoa, S. (1963). *Fed. Proc.* **22**, 62.
Ohtsuka, E. & Khorana, H. G. (1967). *J. Amer. chem. Soc.* **89**, 2195.
Olivera, B. M. & Lehman, I. R. (1967). *Proc. nat. Acad. Sci., Wash.*, **57**, 1426.
Söll, D., Cherayil, J. D. & Bock, R. M. (1967). *J. molec. Biol.* **29**, 97.
Söll, D., Cherayil, J., Jones, D. S., Faulkner, R. D., Hampel, A., Bock, R. M. & Khorana, H. G. (1966*a*). *Cold Spr. Harb. Symp. quant. Biol.* **31**, 51.
Söll, D., Jones, D. S., Ohtsuka, E., Faulkner, R. D., Lohrmann, R., Hayatsu, H., Cherayil, J. D., Hampel, A., Bock, R. M. & Khorana, H. G. (1966*b*). *J. molec. Biol.* **19**, 556.
Söll, D. & RajBhandary, U. L. (1967). *J. molec. Biol.* **29**, 113.
Stevens, A. (1961). *J. biol. Chem.* **236**, PC43.
Watson, J. D. & Crick, F. H. C. (1953). *Nature, Lond.*, **171**, 737.
Weiss, S. B. (1960). *Proc. nat. Acad. Sci., Wash.*, **46**, 1020.
Weiss, B. & Richardson, C. C. (1967). *Proc. nat. Acad. Sci., Wash.*, **57**, 1021.
Wells, R. D., Büchi, H., Kössel, H., Ohtsuka, E. & Khorana, H. G. (1967*b*). *J. molec. Biol.* **27**, 237.
Wells, R. D., Jacob, T. M., Narang, S. A. & Khorana, H. G. (1967*a*). *J. molec. Biol.* **27**, 227.
Wells, R. D., Ohtsuka, E. & Khorana, H. G. (1965). *J. molec. Biol.* **14**, 221.
Zamecnik, P. (1962). *Biochem. J.* **85**, 257.
Zimmerman, S. B., Little, J. W., Oshinsky, C. K. & Gellert, M. (1967). *Proc. nat. Acad. Sci., Wash.*, **57**, 1841.