NUCLEIC ACID SYNTHESIS

H. G. KHORANA

Institute for Enzyme Research, University of Wisconsin, Madison, Wisconsin, U.S.A.

INTRODUCTION

Synthesis in the field of nucleic acids has been a major and continuing interest in my laboratory ever since we knew the structures of the nucleic acids. Clearly the foundations of synthetic work in the nucleotide field were laid by the prolonged and intensive efforts of Professor Todd's laboratory (Cambridge, England) and it was my good fortune to be associated with this laboratory before our own work in the field got going.

With the elucidation of the internucleotidic linkage in the nucleic acids, polynucleotide chemistry entered a new area of complexity in which the major problems to be posed were those of sequence and synthesis of polynucleotides. In principle, organic chemists could attempt to put together nucleosides and nucleotides to form short chains of polynucleotides which would be identical in every way with the naturally-occurring nucleic acids. It was very reminiscent of the developments of protein chemistry at the turn of the century and into the twenties and thirties which followed the proposal of the peptide theory of protein structure. There was a strong and clear case for sustained effort in developing synthesis in the nucleic acid field, especially when one looked around at that time at the level of sophistication that many other areas of natural product chemistry had attained. However, in addition to the intrinsic interest in developing the organic chemistry of nucleic acids, there was the hope that the availability of synthetic polynucleotides of known size and structure would provide many opportunities for chemical, physico-chemical, and enzymatic studies of nucleic acids.

So we began to undertake work on the organo-chemical synthesis of polynucleotides in the middle fifties. While our early work proceeded hand-in-hand with a broader study of the problems of phosphate chemistry, e.g., methods of phosphorylation of nucleosides and of carbohydrates, properties of phosphate esters, specific activation of mononucleotides to form nucleotide coenzymes, polynucleotide synthesis became the most serious commitment in due course of time. Looking back over the years, three phases of our work in this field can be discerned. Up till the early sixties, the main preoccupation was with the organo-chemical synthesis and chemical and enzymatic characterization of polynucleotides. While even at present, organo-chemical methods demand further investigation and refinement, nevertheless, synthesis of short chains of deoxypolynucleotides with known and controlled sequences became possible in the early sixties. A second phase of our work then started and a central question which was posed at this time was "Can chemical synthesis make a contribution to the study of the fundamental process of biological information flow, DNA → RNA → protein?".

In retrospect this exciting phase of our work lasted until about 1967. The third phase, which began some three years ago and developed recently into a major and full interest, has to do with the eventual synthesis of biologically specific DNA's, in other words, genes.

Some of you might agree that it is not an exaggeration to say that I could devote the whole hour to a discussion of either the organo-chemical problems of polynucleotide synthesis, or the genetic code, or our current activity which is concerned with the synthesis of a bihelical DNA corresponding in sequence to the gene for a short ribonucleic acid. Instead I have chosen to give you a summary review of all the three phases of our work, desiccated though this account might appear to be.

CHEMICAL METHODOLOGY FOR POLYNUCLEOTIDE SYNTHESIS

DEOXYRIBOPOLYNUCLEOTIDES

Internucleotide bond synthesis and the protecting groups

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 focussed on the deoxy series. We began by learning to make thymidylyl-thymidine (TpT)(*Figure 1*),

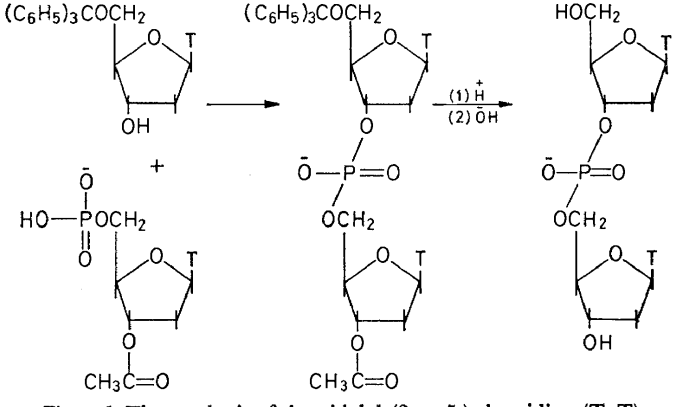


Figure 1. The synthesis of thymidylyl-(3 → 5)-thymidine (TpT).

the simplest dinucleoside phosphate. Thymidine is, relatively speaking, the easiest to work with, for it requires no protection on the ring and presents no special solubility problems. In the synthetic procedure illustrated in *Figure 1* three concepts are worthy of note. (i) One of the components in the reaction is a nucleoside with a free 3′-hydroxyl group, the 5′-hydroxyl group being blocked by the classical bulky trityl group, an acid-sensitive group. (ii) The second component in the condensation is a mononucleotide which has a 3′-hydroxyl group blocked by a simple alkali-labile group. (iii) At the outset of our work on the condensations of the above-mentioned two components, the monoesterified phosphate group, that is, a nucleotide,

was used directly in the condensation reaction. This, in retrospect, represented a very critical advance as will be mentioned again later. Until that time, all of the previous workers had used protected phosphate esters such as benzyl esters. As regards the condensation of the two components shown in *Figure 1* to form the internucleotidic linkage, dicyclohexylcarbodiimide (DCC) was the reagent first used, but an intensive search for activating agents has been maintained in our work. We have investigated all of the different types of activating agents which have been introduced by various groups from time-to-time (see especially ref. 1–3). The condensing agents used today are dicyclohexylcabodiimide and aromatic sulphonyl chlorides, in particular, mesitylenesulphonyl chloride[1] and 2,4,6-triisopropylbenzenesulphonyl chloride.[2] These reagents are also the ones uniformly used in work in this field appearing from other laboratories.

By using stoichiometric amounts of the two components shown in *Figure 1* we can, under mild conditions, obtain 90–95 per cent yield of the internucleotide bond, a situation which is satisfactory by contemporary standards of organic synthesis. The mechanism of activation of the phosphate group is complex. We know some things about it, and at least in the case of DCC, we have demonstrated that the initial phosphorylating agent is a trimetaphosphate[4]. We also know that with the sulphonyl chlorides, the phosphorylating agent is different[1], but clearly as the condensation gets under way, the situation becomes complex because of the fact that pyrophosphate exchange reactions occur at a rate much higher than the phosphorylation of the secondary alcoholic function. The situation must progressively become more complex as the number of diester bonds contained in the oligonucleotidic components used in the condensation reaction increases.

Returning to the problem of protecting groups, we have already noted in *Figure 1* the two fundamental types of protecting groups used for the hydroxyl functions of the deoxyribose moiety. As long as there are only the pyrimidine nucleosides in the oligonucleotide chains to be built, the use of the parent trityl group is satisfactory. However, we soon found that if we wanted to include purine deoxynucleosides in the chains, then the glycosyl bonds in the latter were too sensitive for the prolonged acetic acid treatment required for the removal of the trityl group. Clearly, the need therefore was of a group specific for a primary hydroxyl function, such as a trityl group, but something more labile to acid[5]. This need was fulfilled very satisfactorily by the use of *p*-methoxy substituted trityl groups[6, 7]. The introduction of each methoxy group in the *p*-position of one or more of the phenyl moieties of the trityl group increased the lability to acid by a factor of about 10 and we have used mono- and dimethoxytrityl groups for protection of the 5′-hydroxyl groups in our work.

As soon as we began to work with deoxynucleosides other than thymidine, the problem of protection of the amino functions in the heterocyclic rings became evident. For example, when we attempted a condensation between protected thymidine 5′-phosphate with the then known 5′-*O*-trityl-deoxycytidine, the activated phosphate group reacted mainly with the amino group in the cytosine ring[7]. The product was the phosphoroamidate. In similar experiments evidence was also obtained for the reactivity of the

amino groups in the adenine and the guanine rings, although the reactivity of the amino functions varied a great deal. It was also found that, by and large, the nucleosides and nucleotides lacking any protecting groups on the amino functions were highly insoluble compounds under anhydrous conditions. Therefore, a new class of protecting groups for the amino functions was required, primarily to overcome the reactivity of the amino groups, but also, perhaps, to aid in the solubilization of the compounds.

The reaction sequence shown in *Figure 2* is an example of a general method now available for preparing deoxynucleosides containing the amino groups

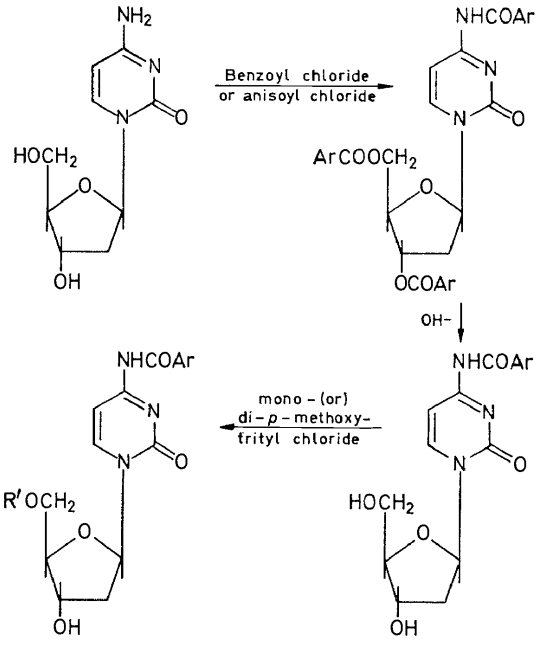


Figure 2. Preparation of protected deoxycytidine derivatives.

as well as the 5′-hydroxyl groups blocked. In the first step, acylation of the unprotected deoxynucleoside, exemplified in *Figure 2* by deoxycytidine, is carried out with an excess of the acylating agent. The fully protected derivative thus obtained is plunged into alkali at room temperature under carefully controlled conditions and the desired *N*-protected derivative is obtained in quantitative yield. The selective de-esterification is due to the fact that an ionization of the amide group in the aromatic system occurs at alkaline pH which extends the resonating system and thus stabilizes the acyl group on the amino function. The rate of loss of acyl groups from the hydroxyl functions, on the other hand, is simply proportional to the hydroxyl ion concentration. In this way, all of the deoxynucleosides and -nucleotides

can be obtained in the *N*-protected form. The succeeding step to protect the 5′-hydroxyl function (*Figure 2*) is a standard one. The derivatives thus prepared can be used in a standard condensation reaction such as has been described above in *Figure 1*. The important point to note is that after the condensation step, all of the protecting groups on the amino functions can be effectively removed by treatment with concentrated ammonia, presumably by the direct nucleophilic attack on the carbonyl carbon.

The total protected derivatives used currently for different deoxynucleosides and deoxynucleotides are shown in *Figure 3*. By a combination of these very acid-sensitive and ammonia-sensitive protecting groups all of the possible dinucleotides can be prepared in excellent yield.

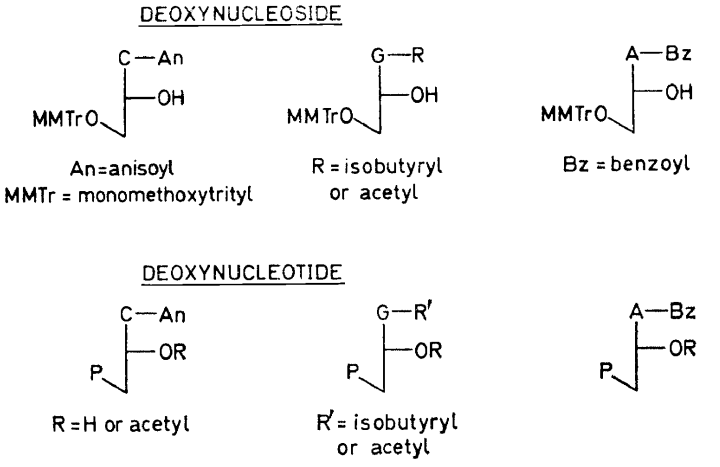


Figure 3. Protected derivatives of deoxyribonucleosides and deoxyribonucleotides.

One other protecting group which has not been introduced so far is that which is used for protecting the phosphomonoester group. There has been, and continues to be, great interest in the synthesis of fully protected oligonucleotides containing 5′-phosphate end groups. The general method in use for the synthesis of dinucleotides of this type is shown in *Figure 4*. The component which is eventually to carry the 5′-phosphate group is converted to a monocyanoethyl derivative (I→ II, *Figure 4*) and this is now brought into condensation with the second component (III) containing the 3′-hydroxyl and the amino groups blocked. After condensation, the cyanoethyl group on the 5′-phosphate group and the 3′-*O*-acetyl group may both be removed by a mildly alkaline treatment, the cyanoethyl group suffering ready elimination in the form of acrylonitrile (IV→ V, *Figure 4*). More recently an alternative group, the trichloroethyl group, originally introduced by Woodward and his coworkers[8], has been investigated by Eckstein[9] for work in the nucleotide field. Our own comparative study shows the cyanoethyl group to be preferable, at least in conjunction with the total strategy that we have available for our work, and we have therefore continued to use this group.

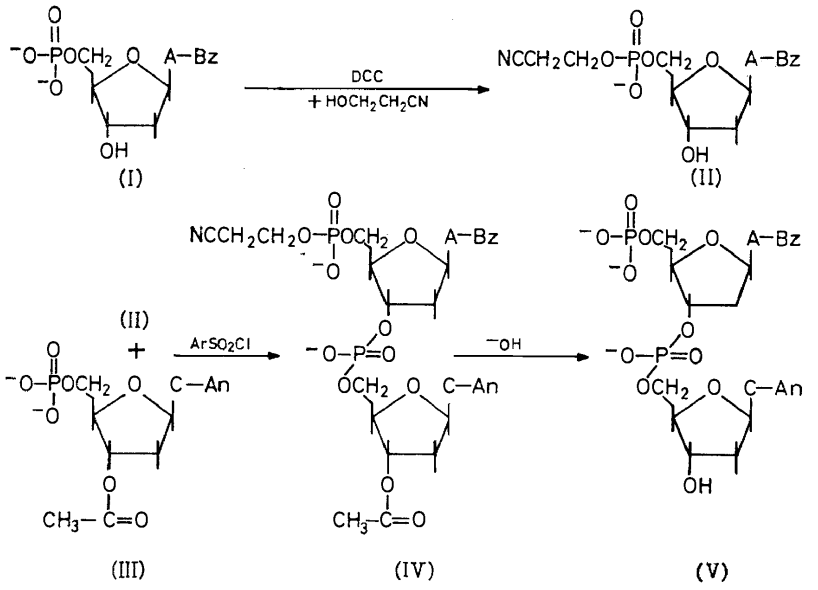


Figure 4. The synthesis of an *N*-protected dinucleotide d-pA^{Bz}pC^{An} carrying a 5′-phosphate end group.

For the synthesis of a trinucleotide with a 5′-phosphate group, the dinucleotide as prepared above, may be quantitatively converted to the cyanoethyl derivative again and this may then be condensed with the mononucleotide. An example of this is given below in *Figure 5.*

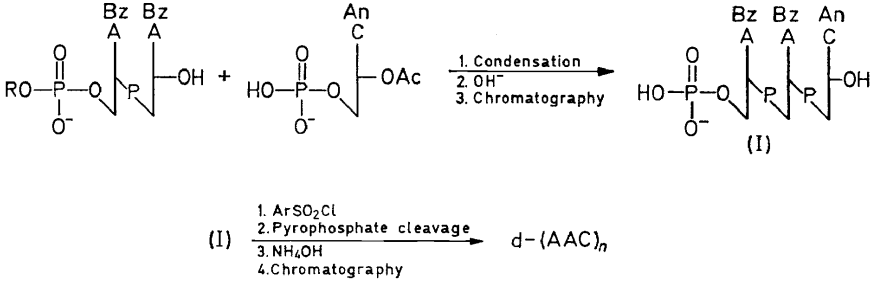


Figure 5. Chemical synthesis of polynucleotides containing a repeating trinucleotide sequence.

It will be noted that the protecting groups, which we have developed in this field, are perhaps the simplest and, by modern standards of organic chemistry, rather conventional in concept; thus, mild acid and mild base, the classical instruments in organic chemistry, are mostly used for their removal. One could, indeed, consider a number of other fundamental

types of procedures. One could consider, for example, some oxidative or reductive process, use something sensitive to heat or photochemical exposure, or one could use some metal-assisted elimination process, and so on. However, so far it has been difficult to coordinate the use of any such ideas with the total strategy that we already have, and it is, of course, to be borne in mind that the sensitivity, structural complexity, and, above all, the physical properties of nucleotides and oligonucleotides simply do not allow the use of many elegant ideas which can be applied in more standard type of organo-chemical operations. Nevertheless, one, of course, hopes that with passage of time further simplification in the total methodology and perhaps more elegance would be introduced. On the other hand, it is also true that the present protecting groups, which we have had for some eight years now, have stood the test of time: as the syntheses have become more and more demanding, especially in the duration of the total operations, their use has been completely satisfactory. Certainly, the problem of protecting groups has not been a progress-determining factor in recent years.

Polymerization of protected mono-, di- and tri-nucleotides

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 were to contain 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, preformed di- and trinucleotide 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 have been used in studies of the genetic code (see later) have been made possible by polymerization methods[10, 11]. An illustration of the steps involved in the preparation of polynucleotides containing repeating trinucleotide sequences is given in *Figure 5*.

Stepwise synthesis of longer chains of deoxyribopolynucleotides with specific sequences

A logical goal of 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-by-one to a growing polynucleotide chain and the second approach, which is theoretically more attractive, consists in the preparation of oligonucleotide blocks and their subsequent condensation to form successively longer chains. Both of these appraches have been investigated systematically over the years.

The approach used first with greater success was that which would involve the stepwise addition of protected mononucleotide units, one at a time, to the 3′-hydroxyl end of a growing oligonucleotide chain[12, 13]. The first synthetic steps are illustrated in *Figure 6* for the synthesis of polynucleotides containing

the repeating T–T–C sequence. Thus, the synthesis started with the condensation of d–pT–*O*Ac and 5′-*O*-trityl-thymidine (Tr–T) and the product on an alkaline treatment gave d–Tr–TpT (I in *Figure 6*). Condensation of the latter with the protected nucleotide, d–pC^{An}–*O*Ac, followed by an alkaline treatment gave d–Tr–TpTpC^{An}. Successive steps involved condensations with either of the two protected nucleotides, d–pT–*O*Ac or d–pC^{An}–*O*Ac. Although this approach entails the maximum number of synthetic steps in putting together a polynucleotide chain of a given size,

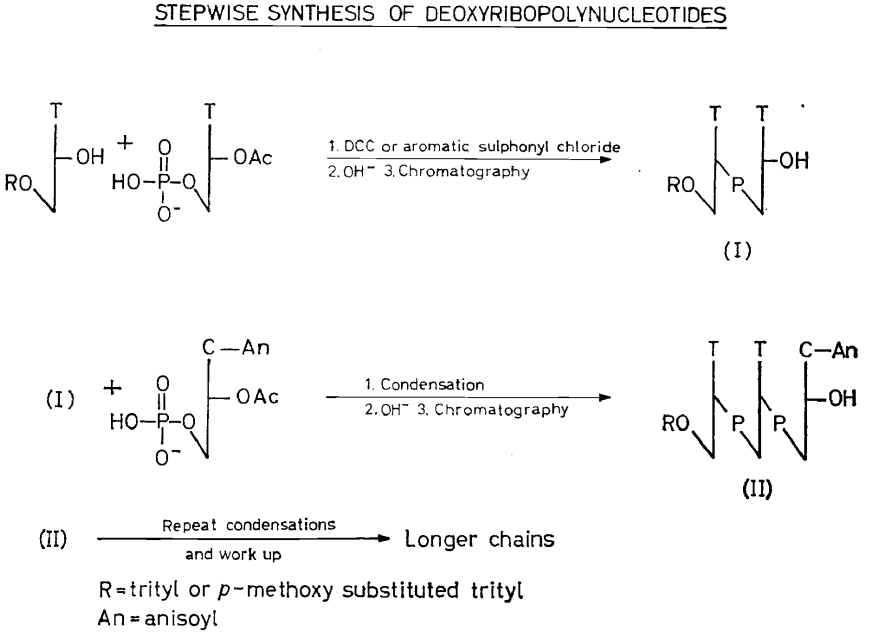


Figure 6. Sequence of chemical reactions for the stepwise synthesis of deoxyribopolynucleotides containing the repeating TpTpC sequence.

it embodies a highly important compensating feature. The growing oligonucleotidic component in successive condensations becomes more and more valuable and consistently high yields with respect to this component may be sustained by providing an increasing excess of the mononucleotidic component. The latter, of course, is much the more readily available component. In *Figure 7* are shown the molar equivalents of the protected mononucleotides, relative to the protected polynucleotides, used during the later condensations and the yields of the desired products. It is seen that the syntheses were satisfactory throughout with respect to the polynucleotidic component.

Before proceeding further, it may be noted that the use of protecting groups for the heterocylic rings necessitated the development of new techniques for the separation and isolation of the protected oligo- and poly-nucleotides (such as those listed in *Figure 7*). It was necessary to ensure that the protected intermediates would be completely stable throughout the prolonged

periods (many months) of chemical synthesis. Column chromatographic techniques which were developed for the isolation of the products after each condensation step involved use of DEAE–cellulose anion exchangers and of buffered eluting agents (triethylammonium acetate or triethylammonium bicarbonate at 2–4°) in mixed aqueous-organic solvents.

Product	*Molar equivalent of mono-nucleotide*	*Yield* (%)	$\frac{270\ m\mu}{302\ m\mu}$
d–TrTpTpC^{An}pTpTpC^{An}pTpT	40	73	1·9
d–TrTpTpC^{An}pTpTpC^{An}pTpTpC^{An}	85	75	1·5
d–TrTpTpC^{An}pTpTpC^{An}pTpTpC^{An}pT	120	70	1·65
d–TrTpTpC^{An}pTpTpC^{An}pTpTpC^{An}pTpT	185	70	1·8
d–TrTpTpC^{An}pTpTpC^{An}pTpTpC^{An} pTpTpC^{An}	220	56	1·5

Figure 7. Conditions for the stepwise synthesis of deoxyribopolynucleotides containing the repeating TpTpC sequence.

The approach using preformed oligonucleotide blocks, as mentioned above is of course potentially more useful and has recently been reinvestigated[3, 14]. It has been used in particular for the synthesis of several series of polynucleotides with repeating tetranucleotide sequences[15, 16]. The oligonucleotide blocks used in condensations have been di-, tri- and tetra-nucleotides. An example is the synthesis of the hexadecanucleotide containing the repeating tetranucleotide sequence, T–T–A–C, steps for which are illustrated in *Figure 8.*

The same approach continues to be used in our current work (see last section of this lecture) and we may further illustrate its use in the stepwise synthesis of the icosanucleotide complementary in sequence to nucleotide 21 through 40 of ala–tRNA (see later). The steps used are shown in *Figure 9.* The bottom line shows the ultimate product of synthesis written in a characterless form using one and two-letter symbols for nucleosides and protecting groups. It is a product of about 8,000 mol. weight and shows all the protecting groups which had to survive the duration, a period of about two years, of synthesis. The first phosphodiester bond synthesis involves a condensation between the protected deoxyguanosine derivative d–MMTr-G^{iBu} and a protected mononucleotide d–pA^{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 and, in addition, by DEAE–cellulose–urea chromatography after removal of the protecting groups. 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.

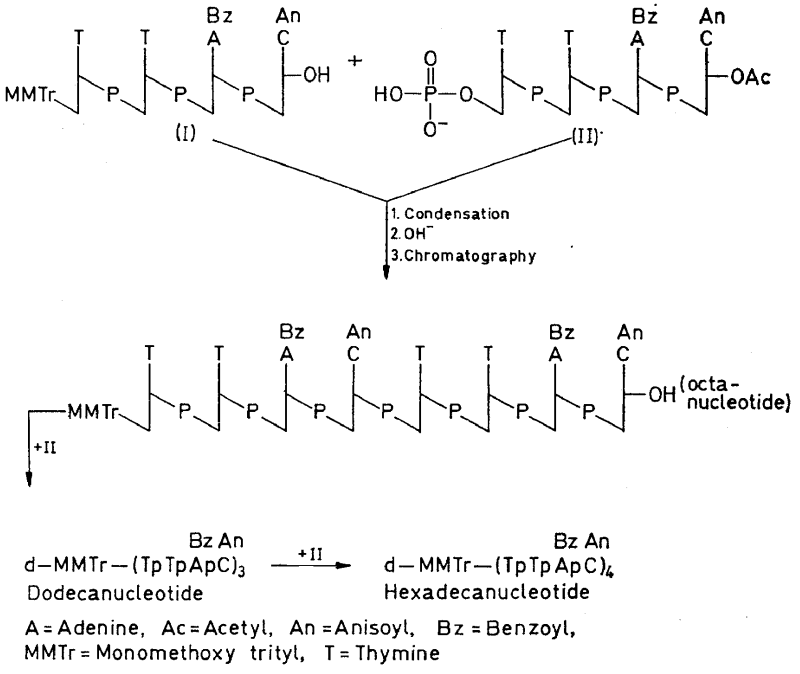


Figure 8. The synthesis of a hexadecanucleotide with repeating tetranucleotide sequence using preformed tetranucleotide blocks.

The synthesis of such long polynucleotide chains requires careful planning with respect to the choice of blocks for a variety of reasons. Thus, in *Figure 9*, it may be noted that the strategy involves the use of larger preformed oligonucleotide blocks as the chain length increases. Apart from the fact that this reduces the number of synthetic steps, the important consideration is the more appreciable total negative charge difference between the starting and the product polynucleotide—the charge difference on the polynucleotides being a main factor, but not the only one, in the separation on ion exchange columns. That separation was reasonably satisfactory up to the last step of the synthesis in *Figure 9* is shown in *Figure 10* for the separation of the hexadecanucleotide from the icosanucleotide product.

Before concluding this section, we may add some overall and general comments on the problem of the stepwise synthesis of specific polynucleotide chains. It has already been mentioned that as the chain length increases, larger excesses of the preformed oligonucleotide blocks must evidently be used and the yields certainly decrease. Further, the problem, which makes synthesis in the polynucleotide field uniquely difficult, in contrast, for example, with the synthesis in the protein field, is the presence of phosphoryl dissociations on the internucleotide bonds. During the activation process intended for the activation of the terminal phosphomonoester group, the phosphodiester bonds are also capable of being further activated and this further activation promotes side reactions which can lead to the cleavage of the internucleotide bonds[17, 18]. With an increase in the number of

Blockwise synthesis of a deoxyriboeicosanucleotide

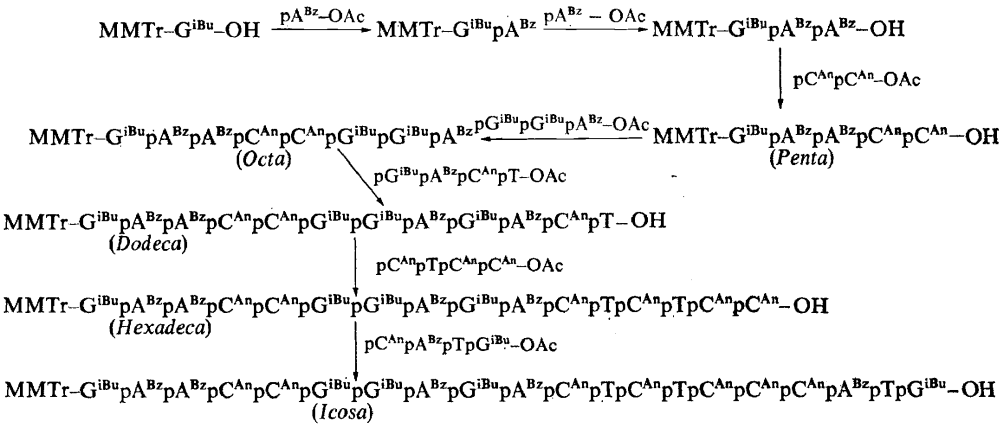


Figure 9. Steps in the chemical synthesis of an icosanucleotide. The standard method of presentation of polynucleotide chains is used. MMTr is the abbreviation for monomethoxytrityl group and is present at the 5′-OH of the terminal nucleoside. The protecting groups on the heterocyclic rings of different deoxynucleosides are shown by superscripts on the nucleoside initials; Ac stands for acetyl, Bz for benzoyl and iBu for isobutyryl groups. *O*Ac at the right-hand end of oligonucleotides stands for 3 -*O*-acetyl.

phosphodiester bonds in the polynucleotidic components to be condensed, the side reactions become, of course, cumulative. Can alternative approaches be devised which would minimize, abolish or obviate these problems? One fundamentally different approach would be to use fully protected

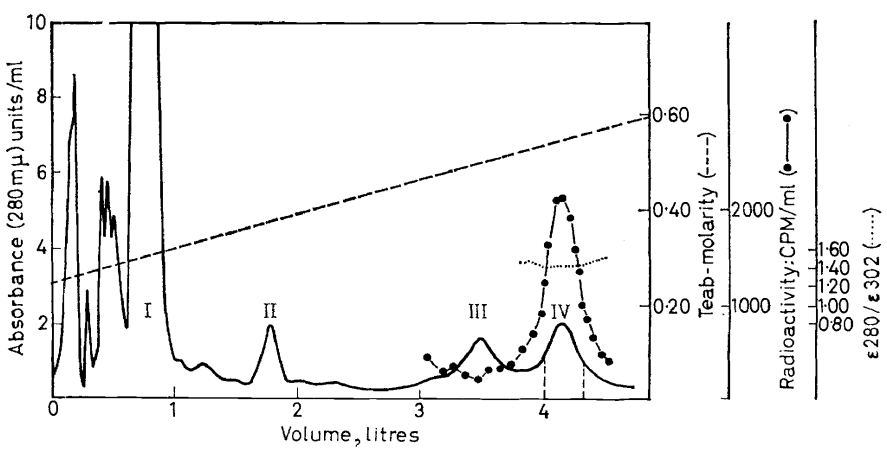


Figure 10. Condensation of the protected hexadecanucleotide with d–pC^AnpA^BzpTpG^iBu–*O*Ac. Separation of the reaction products on a DEAE–cellulose (bicarbonate) column (1·8 × 60 cm) preequilibrated at 4° with 0·05 M triethylammonium bicarbonate (Teab) (pH 7·5) in 40 per cent ethanol. The bicarbonate concentration gradient used for elution was as shown by the dotted line, the solvent being 40 per cent ethanol. Peak I was the excess tetranucleotide, peak II the corresponding pyrophosphate, peak III the unreacted hexadecanucleotide and peak IV the icosanucleotide.

phosphate esters, such that instead of a phosphodiester linkage, a neutral triester is created at every step and only at the end are the triesters converted to the regular phosphodiester bonds. While several laboratories, our own included, have investigated and continue to be interested in this approach, it is difficult to see at this time how a significant overall advantage could be achieved over the current methods. For this approach to be significantly superior, the yields at every condensation step would have to be very high indeed and this is even less likely at the level of triester bond synthesis than in our present reactions where diester bonds are formed directly. If the yields are not going to be 100 per cent or close to it, then there is the eternal problem of separation of the products: entirely new and powerful methods would have to be developed for this purpose, presumably based on partition principle. We would then actually be throwing away the most important device which everyone has hitherto exploited for separation in the polynucleotide field, namely, the polyelectrolyte character of the polynucleotides. In order to retain this last feature and still use the triester approach in condensation reactions, we have investigated the possibility of converting the phosphodiester bonds to the neutral cyanoethyl esters just prior to the condensation reactions and to remove the highly labile cyanoethyl groups after every condensation and prior to separation on columns. All in all this approach, although workable, appears to offer no great superiority over the current methods. Another alternative would be the possible preparation of nucleoside phosphites and build up chains in which the individual nucleosides would be linked to each other through phosphite bridges. Only at the end the di-substituted phosphite linkages could be oxidized to phosphate ester linkages. Again for this approach to be attractive, the yields at every step would have to be quantitative, otherwise formidable separation problems of all the components, neutral in character, would arise.

Polynucleotide synthesis on polymer supports

There is much current interest in carrying out organic synthesis on polymer supports and evidently a great deal of success has been achieved in developing rapid polypeptide synthesis, especially by Merrifield and his coworkers. Naturally the question arises whether similar concepts can be developed in the polynucleotide field. A number of groups of workers[19–22] has been investigating various approaches. It is most important to again stress the obvious, that after all what one is hoping to avoid is the separation step after every condensation. So the whole success of polymer support synthesis depends on realizing *exactly* 100 per cent reaction at every step. Otherwise, clearly, one soon reaches a point of diminishing returns. No one in the polynucleotide field has attained yields on insoluble polymer supports anything approaching 100 per cent, and, in fact, the yields reported have often been less than those that we are accustomed to in our work using the conventional methods.

We, ourselves, have been investigating a number of approaches using both insoluble and organic solvent-soluble polymers[20]. The idea of using solvent-soluble polystyrene-type polymers was first expressed by Professor

Shemyakin and his coworkers[23]. This idea is very attractive for many reasons in that the condensations could be carried out in completely homogeneous medium using the knowledge acquired in our previous extensive syntheses. At the end of the reactions one would simply precipitate the polymer carrying the oligonucleotide chains from a predominantly aqueous medium. In our approach, we start with a commercially available polystyrene of high molecular weight and convert a few percent of the phenyl groups to methoxytrityl chloride groups by standard organic reactions (*Figure 11*). We thus are in a position to attach a deoxynucleoside through its 5′-hydroxyl function to the polymer-supported methoxytrityl group. In this way, we have prepared all of the four 5′-*O*-protected deoxynucleosides supported on a polystyrene backbone (*Figure 12*). The subsequent principles for building on oligonucleotide chains can all be standard and in this way step-by-step extension of the oligonucleotide chains may be realized.

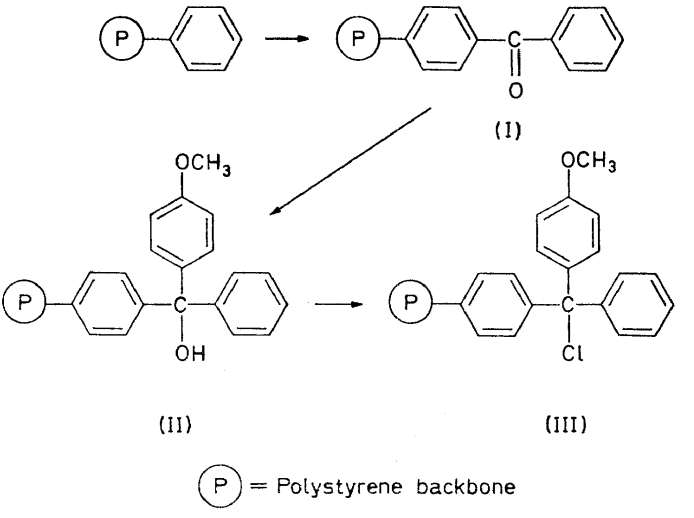


Figure 11. The preparation of *p*-methoxytrityl chloride supported on a polystyrene backbone.

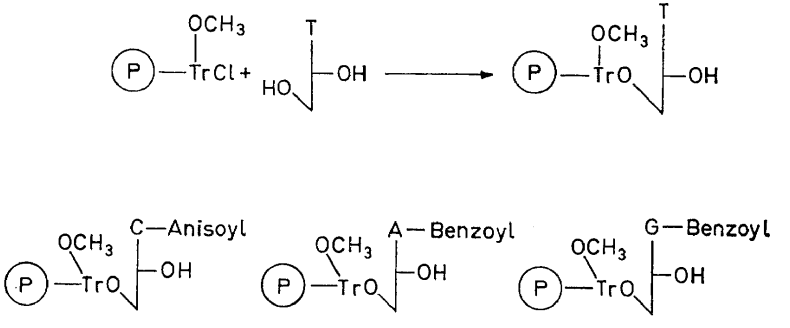


Figure 12. The preparation of 5′-methoxytrityl-*N*-protected deoxyribonucleosides supported on a polystyrene backbone.

While there are still problems to be solved, this approach appears to us to be the most promising and we have already reported the synthesis of a few oligonucleotides in high yield[20]. In a preliminary report Cramer and coworkers[22] have independently described essentially the same approach.

SPECIFIC SYNTHESIS OF RIBO-OLIGONUCLEOTIDES CONTAINING 3′→5′ INTERNUCLEOTIDE BONDS

In the RNA field things are even more complicated. 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′-hydroxylic group of one nucleoside to the 5′-hydroxylic group of the next. Assuming that, as in the above-discussed syntheses, ribomononucleotides are to be one of the component starting materials in ribo-oligonucleotide synthesis, the two types of approaches which can be considered are shown in *Figure 13*. In the approach on the left a fully-protected nucleoside 3′-

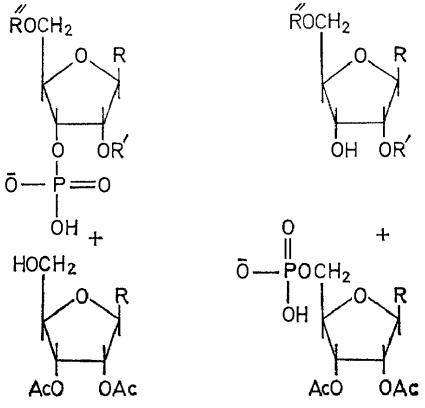


Figure 13. Two alternative approaches to the synthesis of C_3 –C_5 inter-ribonucleotidic linkage.

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) which would have only the 3′-hydroxyl group free and all the other functions suitably protected. A systematic investigation of both approaches in our laboratory (see e.g. refs. 24, 25) led to a clear preference for the former approach discussed above. It should be mentioned that in Dr. C. B. Reese's laboratory in Cambridge, the latter approach continues to be investigated at present.

Two discoveries played very important roles in the practical development of our work in this area. The first was the finding that ribonucleoside-3′ phosphates, which, incidentally, can nowadays be prepared free from the 2′-isomers in large amounts, could be directly reacted with trityl or methoxytrityl halides to form the 5′-*O*-trityl ribonucleoside 3′-phosphates (II) (*Figure 14*). The second crucial finding was that 2′-OH group in these products could be quantitatively acetylated provided carefully defined

conditions were used. In the fifties, the lesson that the nucleotide chemists learned from the RNA chemistry was the ready migration of the phosphate group between the 2′ and 3′-positions in ribonucleotides. At that time 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 acylating agent, would not lead to cyclic phosphate formation and consequent randomization of the phosphate group. Presently, it was discovered, fortunately, that when the acetylation reaction is performed under rigidly anhydrous conditions in the presence of an excess of acetate ions, provided in the form of tetraethylammonium acetate, then only the acetylation of the 2′-OH group occurs and this esterification is quantitative[26].

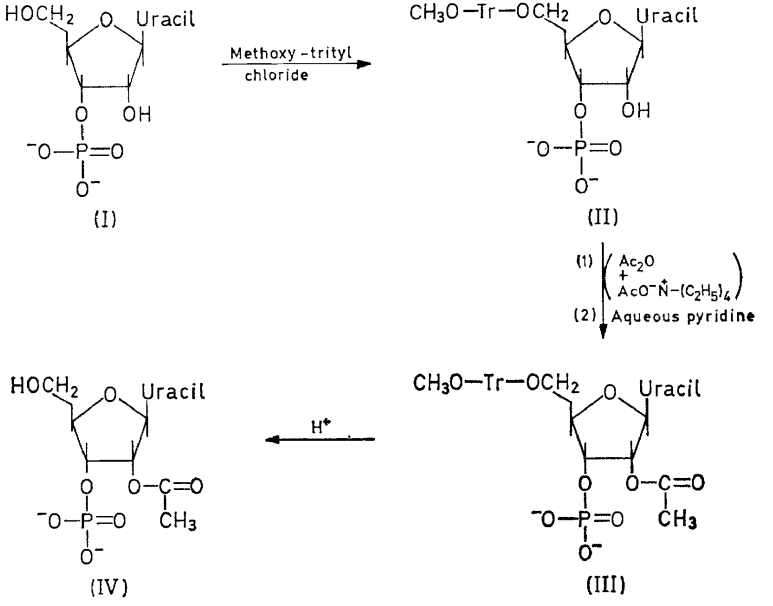


Figure 14. Currently used method for the preparation of protected derivatives of uridine 3′-phosphate.

This reaction thus leads to the preparation of the key intermediates, 5′-*O*-trityl-2′-*O*-acyl-ribonucleoside 3′-phosphates (III in *Figure 14*) (*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 (*Figure 14*) which are suitable starting materials for polymerization reactions. Alternatively, III may be condensed with protected ribonucleosides carrying free 5′-OH groups (*Figure 15*) 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.

It is interesting to observe that the overall strategy for work in the ribonucleotide field turns out to be 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.

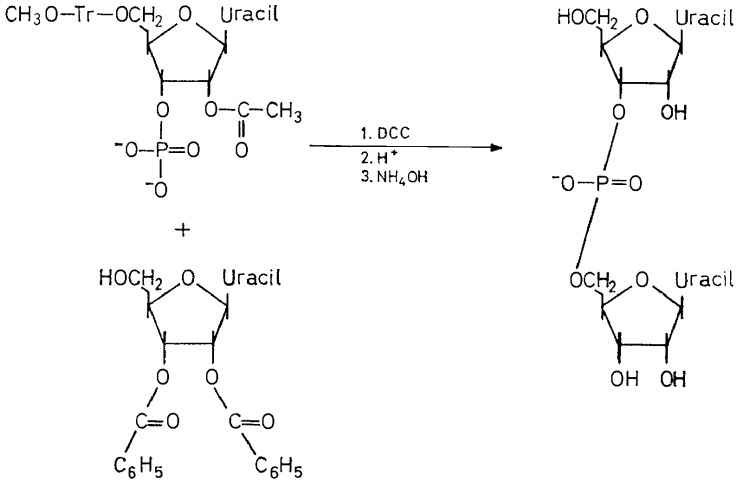
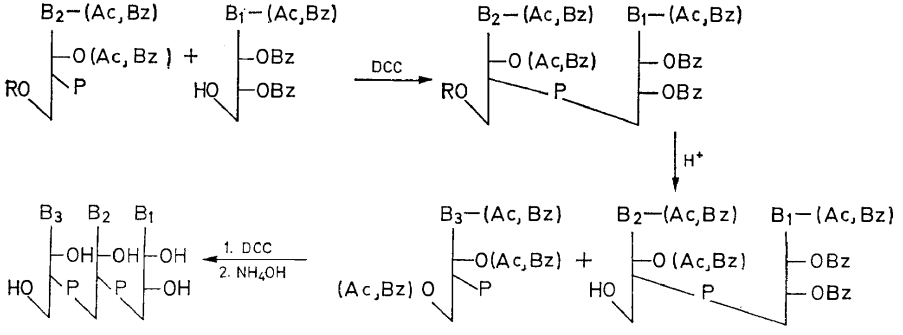


Figure 15. The synthesis of uridylyl-(3′→ 5′)-uridine.

The above approach has recently been 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 *Figure 16.* It is readily seen that the synthesis of the 64 possible trinucleotides



B_1, B_2, B_3 = Purines or pyrimidines
R = Trityl, methoxytrityl, Ac = Acetyl, Bz = Benzoyl

Figure 16. Generalized scheme for the synthesis of ribotrinucleotides.

requires three sets of protected nucleoside or nucleotide intermediates: (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) ribonucleoside 3′-phosphates. Satisfactory methods for the preparation of all of the three sets of compounds were developed. Using these protected derivatives, unambiguous syntheses of all of the possible ribotrinucleotides were realized[27].

POLYNUCLEOTIDE SYNTHESIS AND THE GENETIC CODE

I should now like to turn to the second part of my lecture and here I need to review at least some of the landmarks in the field of Molecular Biology if I am going to attempt to see our own work in relation to the total development in this field.

The inference that genes make proteins goes back to more than fifty years. However, what put this concept into sharp focus was the one gene—one enzyme hypothesis of Beadle and Tatum proposed in the early forties[28] and, I believe, it is this that got the field of biochemical genetics going. The next step was taken when it was established that genes are nucleic acids. The transformation experiments of Avery and coworkers in 1944[29] followed by the bacteriophage experiment of Hershey and Chase in 1952[30] established this for DNA and the work with TMV –RNA a few years later[31, 32] established the same for RNA. By early 1950's 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—the genes—and, 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. There soon came the now classical paper by Brown and Todd[33] in 1952 to which I have already referred. In 1953 came the Watson–Crick structure[34] which elucidated the macromolecular organization of DNA, and focussed attention, in particular, 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 polypeptides was born. A few years later enzymology of DNA got into stride with the discovery in 1956 by Kornberg and coworkers of the enzyme DNA polymerase[35] which showed that DNA could be replicated to produce more DNA in a well-characterized enzymatic reaction. In 1960, several groups of workers almost simultaneously discovered the enzyme, DNA-dependent RNA polymerase which clarified the manner in which information from DNA may be transcribed to a ribonucleic acid and this product we now equate with messenger RNA after Jacob and Monod. The next important biochemical step was the development of a cell-free amino acid incorporating system. Work on this really began with efforts to understand the biosynthesis of peptide bond. The subject has a long history, but one thinks in particular of the work of Zamecnik and Hoagland[36], of Berg, of Lipmann, of Tissieres and Watson at Harvard and then the remarkable observation with the cell-free system made by Nirenberg and Matthaei[37] in 1961. The observation was that a simple polynucleotide, polyuridylate, directs the synthesis of polyphenylalanine in the bacterial cell-free amino acid incorporating system. In

retrospect, what this development did was to bypass the question of sequence determination of nucleic acids for the determination of the code. Instead, one began to prepare polynucleotides of, as far as possible, defined composition and to feed these into the system. The responsibility for complete elucidation of the genetic code now essentially rested with the chemist. Only if a chemist could make a nucleic acid of completely defined structure and analyze the protein 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 *Figure 17.* It was also clear that one would have to study the action of DNA polymerase and of RNA polymerase on the short synthetic deoxypolynucleotides to make ribopolynucleotide messengers of known structure. Results with these enzymes obtained in the early sixties had, in fact, given encouraging results in this direction[38, 39].

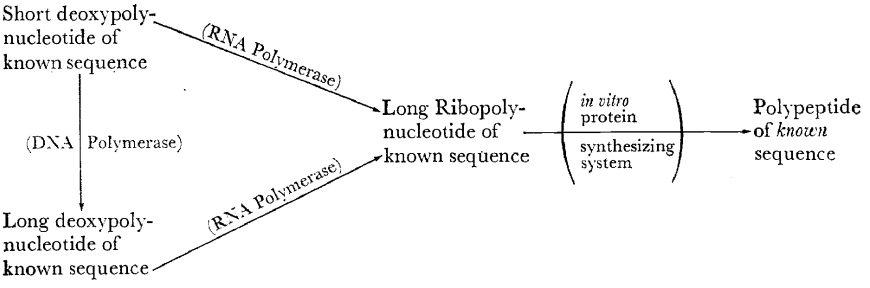


Figure 17. Sequence of enzymatic reactions for the synthesis of specific polypeptides using chemically synthesized specific deoxyribopolynucleotides.

The first question about synthetic work was "What kind of sequences should one start to put together in the deoxy series?" From a number of conisderations, 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.

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

Table 1. Synthetic deoxyribopolynucleotides with repeating sequences

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

A = adenine; C = cytosine; U = uracil; G = guanine; T = thymine.

polynucleotides shown on the left which contained repeating dinucleotide sequences[18]. These are the Watson–Crick complementary sets; one, of the hexamer of alternating T and G and of the hexamer of alternating A and C; and the second set consists of alternating T and C and that of alternating A and G polymers. This work was then extended to polynucleotides with repeating trinucleotide sequences. There are a theoretical maximum of 10 such sets which can contain more than one nucleotide base and we have already prepared 7 such sets[10–13]. 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 catalyze synthesis. Shown also on this Table are 2 sets of polymers with repeating tetranucleotide sequences[15, 16]. In brief, the consideration for these are that they contain in every 4th place the chain-terminating codons and also that this class of polymers also can be used to prove the direction of reading of messenger RNA[40].

A critical advance in our work came when we found, as is seen in *Table 2*, that in presence of DNA-polymerase a mixture of the two short-chain polynucleotides with repeating dinucleotides directs the extensive synthesis of a double-stranded DNA-like polymer (*Table 2*) containing exactly the sequences present in the short-chain polynucleotides[41, 42]. More recently, the same was true for short chains with repeating trinucleotide sequences and even more recently, also for repeating tetranucleotide sequences[43, 44]. Many of the features of these reactions are truly remarkable. (i) Thus: in all these reactions (*Table 2*) the enzyme shows complete fidelity

Table 2. Types of reactions catalysed by DNA-polymerase†

(1) $d(TG)_6 + d(AC)_6 +$	{ dTTP, dATP, dCTP, dGTP } → Poly d-TG:CA
(2) $d(TTC)_4\, d(AAG)_3 +$	{ dTTP, dATP, dCTP, dGTP } → Poly d-TTC:GAA
(3) $d(TATC)_3 + d(TAGA)_2 +$	{ dTTP, dATP, dCTP, dGTP } → Poly d-TATC:GATA

† All 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. d = deoxy

in the reproduction of sequences. (ii) The synthesis is extensive, 50–200 fold and the products are high molecular weight (500 000 to over a million). (iii) The enzyme thus amplifies and multiplies the information created by chemical methods. (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 simply catalogues the different kinds of polymers which have already been prepared in this way and characterized. Thus we have three

classes of polymers; two double-stranded alternating dinucleotide polymers, four polymers with repeating trinucleotide sequences and two polymers with repeating tetranucleotide 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	

For designation of DNA-like polymers see *Table 2.*

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, and these directed the synthesis of a variety of polypeptidic products in the cell-free protein synthesizing system. At the same time 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, initiation and termination of protein synthesis, could all be precisely studied. I do not wish to devote any space here to any of these aspects, especially because these studies have formed the subject of recent reviews at other places[45–48]. However, what I would like to emphasize here is that having prepared DNA's of completely known structures it was imperative that we rigorously characterize the polypeptidic products specified by the polynucleotides. All the results gave abundant proof for the correctness of the basic properties of the genetic code. First of all, it showed the most important thing that the sequence of amino acids in peptide chains is under the direct control of the nucleotide sequence in DNA, this control being exerted via RNA and secondly, all of the results were consistent with the three-letter, non-overlapping properties of the code.

Finally, I would only like to take this opportunity to present the list of total three-letter codon assignments for different amino acids. *Table 4* presents the catalogue which is now universally accepted. The most encouraging feature is that there is really no conflict with regard to any assignment between the different lines of experimental evidence. Various portions of the code have been independently derived from *in vivo* experiments by a large number of different workers[49]. To me, as an organic chemist, it is a matter of great personal satisfaction to record that my confidence in most of these assignments is comparable to that in the organic chemists' assignment of structure to any of the usual complex natural products.

With the solution to the problem of the genetic code, a central problem of our time may be claimed to have been solved. It may be hoped that this knowledge will form a theoretical framework for further work in many areas of molecular biology.

Table 4. The Genetic code.

1st letter	*2nd letter*				*3rd letter*
	U	C	A	G	
U	PHE	SER	TYR	CYS	U
	PHE	SER	TYR	CYS	C
	LEU	SER	C.T.	C.T.	A
	LEU	SER	C.T.	TRY	G
C	LEU	PRO	HIS	ARG	U
	LEU	PRO	HIS	ARG	C
	LEU	PRO	GLN	ARG	A
	LEU	PRO	GLN	ARG	G
A	ILEU	THR	ASN	SER	U
	ILEU	THR	ASN	SER	C
	ILEU	THR	LYS	ARG	A
	MET (C.I.)	THR	LYS	ARG	G
G	VAL	ALA	ASP	GLY	U
	VAL	ALA	ASP	GLY	C
	VAL	ALA	GLU	GLY	A
	VAL (C.I.)	ALA	GLU	GLY	G

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 in this Table follows the conventional way of writing of trinucleotides: thus, the first letter (base) of the trinucleotide is on the left and the third letter (base) 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 U–U–U and U–U–C; codons for the amino acid ALA are GCU, GCC, GCA and GCG.

PROGRESS IN THE TOTAL SYNTHESIS OF THE GENE FOR A TRANSFER-RIBONUCLEIC ACID

Introductory

Some three years ago, when the work on the problems of the genetic code was reaching a satisfactory conclusion, we began to define the orientation of further work in the synthetic field and this is what I would like to turn to now.

While the high molecular-weight DNA-like polymers with repeating nucleotide sequences continue to be used for a variety of physico-chemical and enzymatic studies of nucleic acids, it has been clear to us for sometime 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 turns genes on and off and, as parts of this question, what the initiation and termination signals for RNA polymerase are and what kind of sequences are recognized by repressors. What determines the specificity of the host modification and host restrictive enzymes and of enzymes involved in genetic recombination and so on. For these studies eventually what is required is the ability to synthesize long chains of bihelical deoxyribonucleic

acids with specific non-repeating 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 must 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.

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[50, 51]. The choice of the gene for a transfer RNA followed from a variety of considerations. First of all we know the primary structures and, secondly, we know that these 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, by the pyrophosphorylase which repairs the CCA end, by the aminoacyl-tRNA synthetases, by ribosomes and by messenger RNA. Furthermore, transfer RNA's are a unique

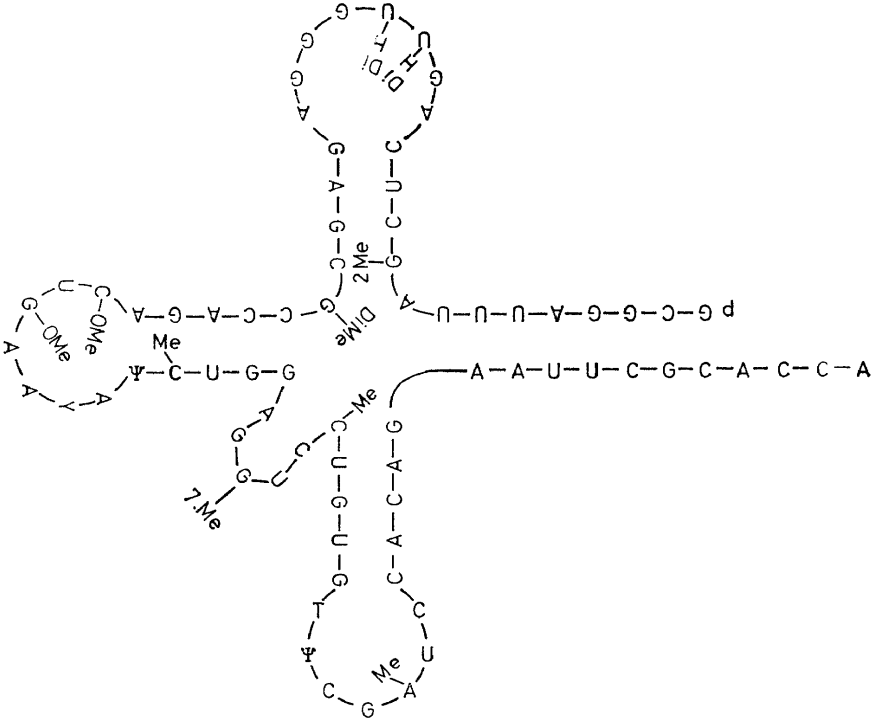


Figure 18. Clover-leaf model for the secondary structure of yeast phenylalanine tRNA[52]. Abbreviations for different nucleosides are standard: Thus, Ψ stands for pseudouridine. Y stands for a nucleoside which is as yet unidentified.

Despite variations in base composition and sequence, all of the tRNA's, whose primary nucleotide sequences are known, can adopt a clover-leaf secondary structure remarkably similar to that shown here for phe-tRNA. The clover-leaf model was one of the models included in the original publication of Holley and coworkers on yeast ala-tRNA.[53]

class of molecules, possessing attributes of both *nucleic acids* and *proteins*. There is a good deal of evidence now to suggest that in addition to a common secondary structure (e.g., *Figure 18*, ref. 52) these molecules possess a tertiary structure. Then, all tRNA's abound in minor bases which are largely found in the looped-out 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 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 is 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. For example, one could have deletions in different parts, one could take an anti-codon loop from a tRNA specific for one amino acid and replace it with the anti-codon loop from another tRNA.

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 have been wanting to exploit is the template principle, that is, the ability of polynucleotides to form hydrogen-bonded bihelical structures. We want to make short pieces of DNA which 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 either by chemical or enzymatic 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. In the latter case, mistakes would accumulate and the currently available methods of separation and purification for polynucleotides, even for chains of 20 to 50 units long, are hopelessly inadequate.

Chemical synthesis of deoxyribopolynucleotide segments

In initiating synthetic work, our decision was to set out to synthesize icosanucleotides with the sequences shown in *Figure 19*. These are about the

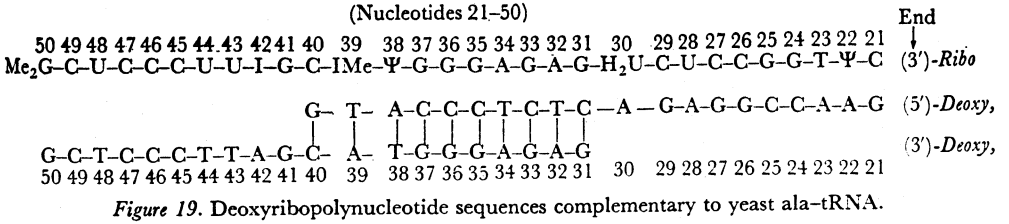


Figure 19. Deoxyribopolynucleotide sequences complementary to yeast ala-tRNA.

maximum chain lengths that current methods of synthesis and, in particular, separation would reasonably allow. The points to be made about the two icosanucleotides shown in *Figure 19* are: (i) They span the length 21 through

50 nucleotide residues of yeast alanine tRNA[53]. Yeast alanine tRNA was the only tRNA whose primary sequence was known at the time the present work was undertaken. (ii) The principal assumption in defining the DNA sequence complementary to the tRNA 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 Ψ and dihydro–U both originate from U and hence correspond to an A–T base-pair. (iii) The third point about the icosanucleotides in *Figure 19* is that they are complementary through half of their length and are of opposite polarity. The complementary region of 10 base-pairs was expected to be sufficiently long to allow the icosanucleotides to align themselves as shown. One possibility considered for enzymatic work was that the DNA-polymerase of *E. coli* might repair these structures to complete a duplex of 30 nucleotide units[50]. If this failed, then further chemically synthesized deoxyribopolynucleotides with sequences complementary to the single-stranded arms in *Figure 19* 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[54].

The chemical synthesis of one of the two icosanucleotides of *Figure 19* has already been reviewed in an earlier section. The total list of synthetic oligo- and poly-nucleotides which have been prepared so far is given in *Table 5*.

Table 5. Deoxyribo-, oligo- and polynucleotides prepared by chemical synthesis

Icosanucleotides and intermediate sizes from 5′-end
(5′) G–A–A–C–C–G–G–A–G–A–C–T–C–T–C–C–C–A–T–G (3′)
(5′) G–C–T–C–C–C–T–T–A–G–C–A–T–G–G–G–A–G–A–G (3′)

Dodecanucleotide and intermediate sizes from 5′-end
(5′) T–G–G–T–G–G–A–C–G–A–G–T (3′)

Undecanucleotide and intermediate sizes from 5′-end
(5′) C–C–G–G–T–T–C–G–A–T–T (3′)

Decanucleotide and intermediate sizes from 5′-end
(5′) T–C–G–G–T–A–G–C–G–C (3′)

Nonanucleotides and intermediate sizes from 5 -end
(5′) T–C–T–C–C–G–G–T–T (3′)
(5′) C–T–A–A–G–G–G–A–G (3′)

Octanucleotide and intermediates from 5 -end
(5′) C–C–G–G–A–A–T–C (3′)

Heptanucleotide
(5′) A–G–A–G–T–C–T (3′)

Joining of short deoxyribo-oligonucleotides by DNA-joining enzymes

During the past year, while the syntheses of the icosanucleotides of *Figure 19* were nearing completion, several groups of workers[55–59] reported the discovery of two distinct enzymes which bring about repair of single-stranded breaks in double-stranded DNA, the reaction being illustrated by using the hypothetical sequence of a DNA segment in *Figure 20*. Thus, wherever a break in one of the strands occurs so as to form a 3′-OH and a 5′-phosphate group, the enzymes, in the presence of an energy source such as ATP or, curiously, diphosphopyridine nucleotide, restore the regular phosphodiester linkage.

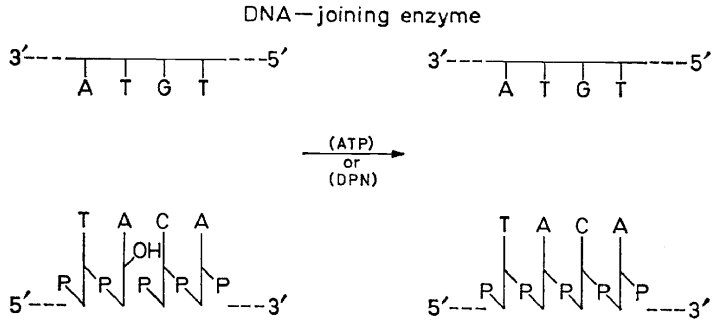


Figure 20. Polynucleotide joining enzyme catalyzes the formation of a phosphodiester bond at the site of a single-strand break in a duplex DNA molecule.

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 which these enzymes would require in order to bring about the joining reaction. Two sets of studies were carried out[60] and the combinations of polynucleotides used are listed in *Table 6*. The first set of experiments (part A of *Table 6*) showed that a chain as short as a hexanucleotide, e.g., 5′-P$(TG)_3$, in combination with the complementary long deoxyribopolynucleotide, e.g., poly d-CA, can be joined end-to-end to form a very long chain. Even more relevant were the results (B in *Table 6*) with the systems in which both complementary chains were short. Thus two molecules of the octanucleotide 5′-$P^{32}(TAAG)_2$ could be joined end-to-end to form the corresponding hexadecanucleotide in the presence of the complementary $d(TTAC)_4$. Finally, as one might have expected, a direct correlation of the temperature

Table 6. Deoxyribo-, oligo- and polynucleotide combinations used in study of DNA-joining enzymes

A.	*Short*	*Complementary Long*
	$d(TG)_{3-5}$	poly d-CA
	$d(CA)_{3-5}$	poly d-TG
	d–T_6–T_{10}	poly dA
	d–A_{10-15}	poly dT
B.	*Short*	*Complementary short but longer*
	$d(TAAG)_2$	$d(TTAC)_4$
	$d(AG)_4$	$d(TC)_8$

requirement of the joining reaction with the stability of the bihelical complex between the components was evident. Thus under the conditions of the enzyme reaction (0·006 M Mg^{2+} + 0·01 M Tris chloride), d–P$(TG)_3$ + poly d–CA showed a T_m (midpoint of helix-random coil transition) of about 35° (*Figure 21*) and the optical temperature for the joining reaction was around 20°. The T_m of $d(TTAC)_4$ + d–P$(TAAG)_2$ was around 20° (*Figure 22*) and the optimal temperature for the joining reaction was 5–10°. It was therefore concluded that the joining reaction would occur so long as the component polynucleotides were able to form spontaneously bihelical structures under the conditions of the reaction.

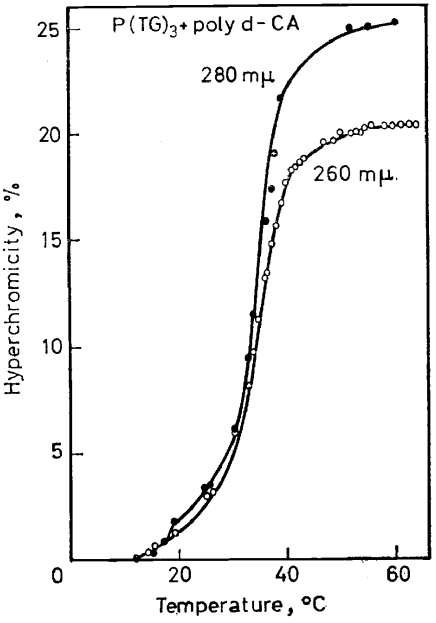


Figure 21. Temperature-absorbance profile of d–$P^{32}(TG)_3$ + poly d–CA in 0·01 M Tris–HCl buffer, pH 7·6, and 0·006 M $MgCl_2$.

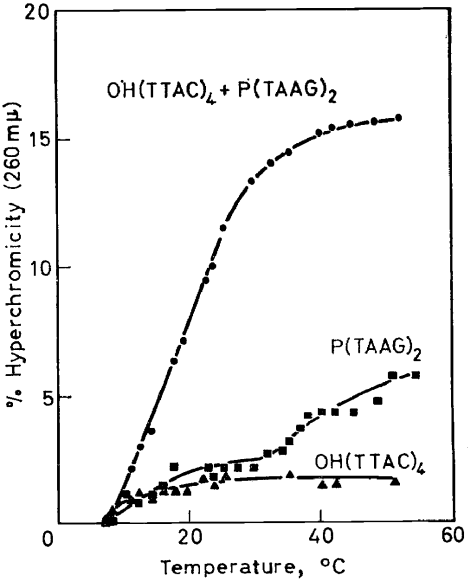


Figure 22. Temperature-absorbance profile of $d(TTAC)_4$ and d–$P^{32}(TAAG)_2$ in 0·01 M Tris–HCl buffer, pH 7·6, and 0·006 M $MgCl_2$.

Joining of deoxypolynucleotide segments corresponding to tRNA gene

We were now in a position to study the joining of the synthetic segments with specific sequences. The complementary segments whose joining has been studied are shown in *Figure 23*. Because of the specificity in sequences

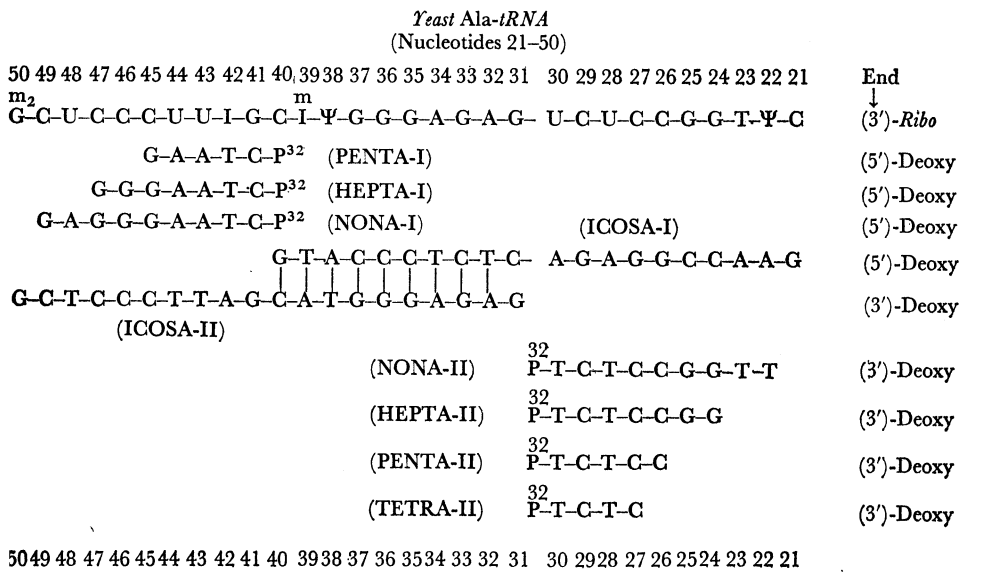


Figure 23. Chemically synthesized deoxyribopolynucleotides corresponding to sequences 21–50 of yeast ala-tRNA. The icosanucleotide-I (Icosa-I) represents a sequence complementary to nucleotide 21–40 of the tRNA and has polarity opposite to that of the tRNA; the nona-, hepta-, and penta-nucleotides (designated Nona-I, Hepta-I, etc.) similarly contain sequences complementary to nucleotides 41–49 or less and again have polarity opposite to that of the tRNA. The deoxyribopolynucleotides, Icosa-II, Nona-II, Hepta-II, Penta-II and Tetra-II, are segments, complements of the complement and therefore contain the same sequences and polarity as the tRNA itself. P^{32} represents 5′-phosphate end group wherever shown. In the synthetic work, the assumption was made that the rare bases present in the tRNA arise by subsequent modification of the four standard bases used by the transcribing enzyme. Thus inosine is formed by deanimation of adenosine and so comes from an A–T base-pair in DNA.

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's with chain length was available and furthermore our primary concern was to test for the formation of bihelical structures under the conditions of the enzymic reaction.

The results given in *Figure 24* showed that stabilities of very short duplexes were surprisingly high. Thus, a decanucleotide + the complementary nonanucleotide showed a T_m of slightly higher than 40°. A combination of the decanucleotide and the complementary heptanucleotide had a T_m of around 33°. There are many features of general interest which are apparent in the data of *Figure 24* but attention may be focussed on the stabilities of

polynucleotide mixtures of interest for the joining reactions. Thus a mixture of the two icosanucleotides and nona-nucleotide-II gave an ordered structure with T_m of about 43°. A combination of the two icosanucleotides and the heptanucleotide (Hepta-II) also gave a rather stable complex ($T_m \sim 36°$). From these results, it was concluded that temperatures of 20° or below would be appropriate for testing of the enzyme-catalyzed joining reactions. It is, further, of interest that the bihelical complexes described formed very rapidly. Although, initially, prolonged incubations were given at temperatures varying between 0° and 20°, this was unnecessary and annealing appeared to require no more than some minutes at the most.

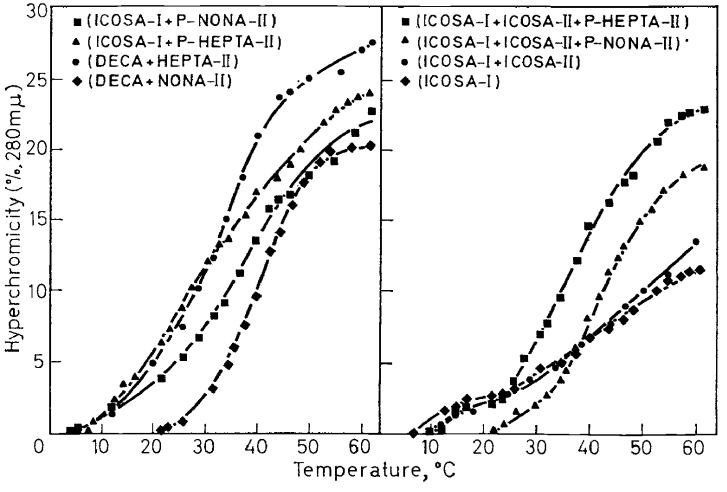


Figure 24. Temperature-absorbance profiles of different combinations of complementary deoxyribopolynucleotides. The medium used was 0·01 M Tris–HCl buffer, pH 7·6, and 0·006 M $MgCl_2$. The structures of the different polynucleotides are shown in *Figure 23*. Deca is G–A–A–C–C–G–G–A–G–A, which is complementary to the Nona-II.

Without going into details of the biochemical procedures used to follow the joining reactions, I simply wish to conclude that the joining of all the short oligonucleotides (shown in *Figure 23*) as short as the tetranucleotide p–TCTC, to the 3′–OH end of the appropriate icosanucleotide has been shown to go to completion[51, 61].

General strategy for DNA synthesis

The most fundamental general concept which emerges from results described above for future work is the following: For the step-by-step and ordered synthesis of long duplexed DNA, there should always be, after every joining reaction, a protruding single-stranded end (designated the sticky end). The latter will guide in the next single-stranded segment belonging to the complementary strand of the duplex. The length of the individual pieces to be chemically synthesized need be no more than 10 nucleotide units or may, in certain cases, be even less. About five nucleotide units may be used to complement with the sticky end and the other five would be left over to serve in turn as the new sticky end. (These principles

are illustrated in *Figure 25* for further immediate work on the gene for ala-tRNA.) The total approach offers an incredibly simple solution to what has so far appeared to be an extremely complex problem, namely, the unambiguous and accurate synthesis of very long deoxypolynucleotide chains with strictly defined sequences. The demands on organo-chemical synthesis are *relatively* light in that only decanucleotides will, in general, be required. In this size range, the purification and characterization of the products of organic synthesis can be rigorously satisfactory.

Chemical syntheses of the segment will be followed by the end-to-end joining by DNA-joining enzymes as reviewed above. By the enzymatic procedures described elsewhere[51, 61], the joining reactions can be carefully scrutinized to be specific and quantitative. Further, as discussed above, as the duplexed chains are elongated step-by-step, there will always be a single-stranded end and, in every alternate joining step, the strand bearing the 3′-OH group will be the shorter one. Another method has been developed for further characterization of the developing duplex at these stages[62]. The important tool is the *E. coli* DNA polymerase used under conditions of strict repair action[63, 64]. Shown in *Figure 26* are three of the short DNA's which are the products of the enzymatic joining reactions and represent a part of the gene for ala-tRNA (see above). The nucleotide residues shown within dash-line boxes are actually lacking in the joined products. Detailed studies on nucleotide incorporations, using one or more appropriate isotopically labelled deoxynucleoside 5′-triphosphate substrates, have demonstrated that the pattern of nucleotide incorporations are exactly as expected for additions at the shorter 3′-OH ends: the reaction simply leads to the completion of the duplexes. Conversely, the DNA polymerase may be used as a precise tool for monitoring the step-by-step joining by the DNA-joining enzymes, for the pattern of nucleotide incorporations could be predicted accurately for the elongating DNA duplex.

With the above principles and methodology, the more immediate job of the total synthesis of ala–tRNA gene can be precisely planned.

Prospects for the future

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. The principle of 'the sticky end' as the main device in DNA synthesis opens up additional possibilities in a number of directions. For example, totally synthetic genes with appropriate sticky ends may be covalently joined to the ends of genomes such as λ DNA by using a sticky end of the latter. Circularization of the synthetic DNA to form covalently closed duplexed circle is another possibility. While all these possibilities belong to the future, the present results nevertheless would appear to provide an encouraging start.

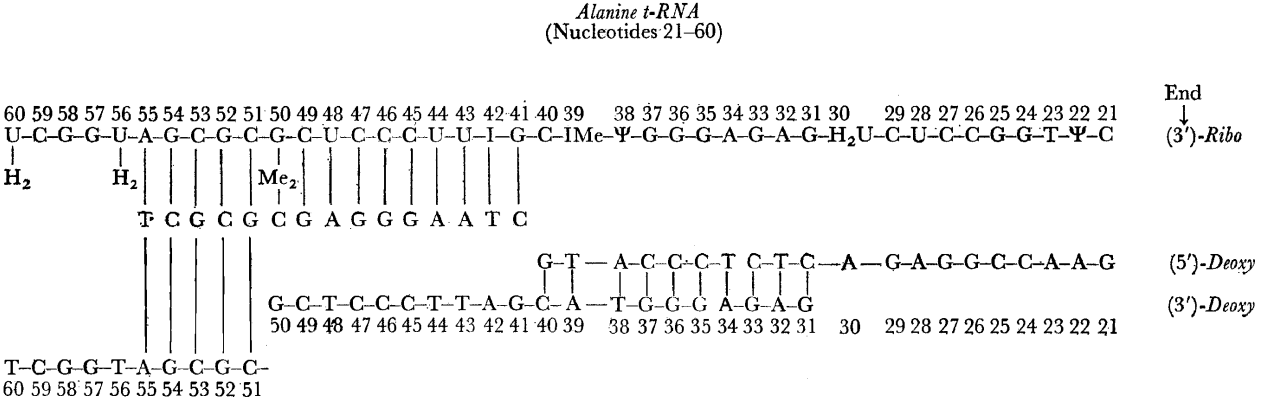


Figure 25. Principles for further work on the synthesis of the gene for ala-tRNA (for explanation see text).

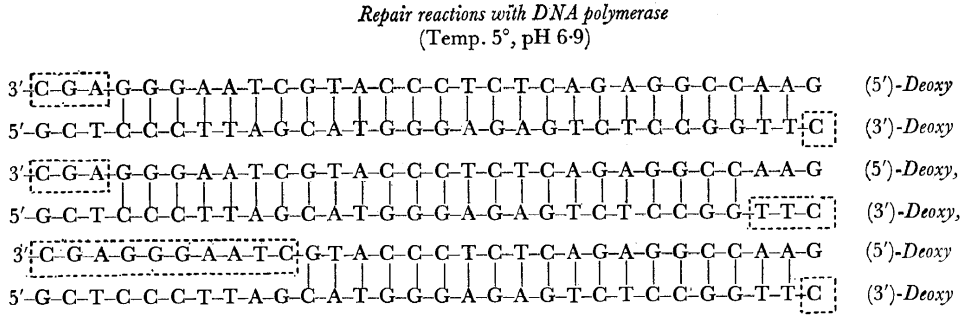


Figure 26. DNA-duplexes prepared by enzymatic joining of chemically synthesized deoxyribonucleotides. The nucleotide sequences within the dash-line boxes are lacking in the joinase products.

SUMMARY AND CONCLUSION

I have attempted a swift review of a concentrated and sustained effort on the part of a rather large group of people working in my laboratory during the past fifteen years or so. All of this work has had behind it the experimental and theoretical philosophy of organic chemistry. In retrospect, three phases of this work can be discerned. In the first phase, the major preoccupation was with the organo-chemical methodology of polynucleotide synthesis. This still occupies us today: there is room for elegance and further simplification especially when one looks ahead to the tremendous future that the whole area of molecular genetics promises to the organic chemist. In the second phase of work, which began around 1962, we were able to use methods of organic chemistry and enzymology in conjunction. An exciting phase of work ensued with the preparation in this way of high molecular-weight double-stranded DNA's of limited but specified information. These made possible a variety of studies on DNA-directed protein synthesis and, consequently, the elucidation of the fundamental properties of the biological code. A start has now been made in what appears to be the third phase of our work in the nucleic acid field: the ultimate objective of the present work must be the total synthesis of genes. As in the work on the genetic code, there is again the very fortunate circumstance that short chemically-synthesized oligonucleotides can be joined end-to-end by specific enzymes to form DNA duplexes. The central concept in this work is that of using a single-stranded stretch (the sticky end) to guide the specific joining reactions. As expected, the same concept clearly must be the central device used by *Nature* in DNA manipulation and recombination. When the synthesis of bihelical long DNA is mastered in the laboratory, it may be hoped that many of the outstanding problems in molecular biology can be studied with chemical precision.

I wish to conclude by hazarding the following rather long-range predictions. In the years ahead, genes are going to be synthesized. The next steps would be to learn to manipulate the information content of genes and to learn to insert them into and delete them from the genetic systems. When, in the distant future, all this comes to pass, the temptation to change our biology will be very strong.

ACKNOWLEDGEMENTS

A large number of devoted chemists and biochemists have participated in the studies which 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 enzymatic methods for the synthesis of the gene for ala–tRNA: H. Buchi, 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, the National Cancer Institute of the National Institutes of Health, U.S. Public Health Service, and the Life Insurance Medical Research Fund, and by the Graduate School of the University of Wisconsin.

References

[1] T. M. Jacob and H. G. Khorana. *J. Am. Chem. Soc.* **86**, 1630 (1964).
[2] R. Lohrmann and H. G. Khorana. *J. Am. Chem. Soc.* **88**, 829 (1966).
[3] H. R. Kossel, M. W. Moon and H. G. Khorana. *J. Am. Chem. Soc.* **89**, 2148 (1967).
[4] G. Weimann and H. G. Khorana. *J. Am. Chem. Soc.* **84**, 4329 (1962).
[5] P. T. Gilham and H. G. Khorana. *J. Am. Chem. Soc.* **81**. 4647 (1959).
[6] M. Smith, D. H. Rammler, I. H. Goldberg and H. G. Khorana. *J. Am. Chem. Soc.* **84**, 430 (1962).
[7] H. Schaller, G. Weimann, B. Lerch and H. G. Khorana. *J. Am. Chem. Soc.* **85**, 3821 (1963).
[8] R. B. Woodward, K. Heusler, J. Gosteli, P. Naegeli, W. Oppolzer, R. Ramage, S. Ranganathan, and H. Vorbruggen. *J. Am. Chem. Soc.* **88**, 852 (1966).
[9] F. Eckstein. *Angew. Chem. Intertnat. Edit.* **5**, 671 (1966).
[10] S. A. Narang, T. M. Jacob and H. G. Khorana. *J. Am. Chem. Soc.* **89**, 2158 (1967).
[11] S. A. Narang, T. M. Jacob and H. G. Khorana. *J. Am. Chem. Soc.* **89**, 2167 (1967).
[12] T. M. Jacob and H. G. Khorana. *J. Am. Chem. Soc.* **87**, 2971 (1965).
[13] S. A. Narang and H. G. Khorana. *J. Am. Chem. Soc.* **87**, 2981 (1965).
[14] J. Hachmann and H. G. Khorana. *J. Am. Chem. Soc.*, in press.
[15] H. Kossel, H. Buchi and H. G. Khorana. *J. Am. Chem. Soc.* **89**, 2185 (1967).
[16] E. Ohtsuka and H. G. Khorana. *J. Am. Chem. Soc.* **89**, 2195 (1967).
[17] T. M. Jacob and H. G. Khorana. *J. Am. Chem. Soc.* **87**, 368 (1965).
[18] E. Ohtsuka, M. W. Moon and H. G. Khorana. *J. Am. Chem. Soc.* **87**, 2956 (1956).
[19] R. L. Letsinger and V. Mahadevan. *J. Am. Chem. Soc.* **87**, 3526 (1965); **88**, 5319 (1966); R. L. Letsinger, M. H. Caruthers, P. S. Miller and K. K. Ogilvie, *J. Am. Chem. Soc.* **89** 7146 (1967); R. L. Letsinger, M. H. Caruthers and D. M. Jerina. *Biochemistry* **6**, 1379 (1967).
[20] H. Hayatsu and H. G. Khorana. *J. Am. Chem. Soc.* **89**, 3880 (1967).
[21] L. R. Melby and D. R. Strobach. *J. Am. Chem. Soc.* **89**, 450 (1967).
[22] F. Cramer, R. Helbig, H. Hettler, H. K. Scheit and H. Seliger. *Angew. Chem.* **78**, 640 (1966).
[23] M. M. Shemyakin, Y. A. Ovchinnikov, A. A. Kinyushkin, and I. V. Kozhevnikova. *Tetrehedron Letters* 2323 (1965).
[24] D. H. Rammler and H. G. Khorana. *J. Am. Chem. Soc.* **84**, 3112 (1962).
[25] Y. Lapidot and H. G. Khorana. *J. Am. Chem. Soc.* **85**, 3852 (1963).
[26] D. H. Rammler, Y. Lapidot and H. G. Khorana. *J. Am. Chem. Soc.* **85**, 1989 (1963).
[27] R. Lohrmann, D. Soll, H. Hayatsu, E. Ohtsuka and H. G. Khorana. *J. Am. Chem. Soc.* **88**, 819 (1966).
[28] G. W. Beadle and E. L. Tatum. *Proc. Natl. Acad. Sci. U.S.* **27**, 499 (1941).
[29] O. T. Avery, C. M. Macleod and M. McCarty. *J. Exptl. Med.* **79**, 137 (1944).
[30] A. D. Hershey and M. Chase. *J. Gen. Physiol.* **36**, 39 (1952).
[31] A. Gierer and G. Schramm. *Nature* **177**, 702 (1956).
[32] H. Fraenkel-Conrat, B. Singer and R. C. Williams. *Biochim. Biophys. Acta* **25**, 87 (1957).
[33] D. M. Brown and A. R. Todd, in *Nucleic Acids*, E. Chargaff and J. N. Davidson, (eds.), Academic Press, New York, 1955, Vol. I, p. 409.
[34] J. D. Watson and F. H. C. Crick. *Nature* **171**, 737 (1953).
[35] I. R. Lehman, M. J. Bessman, E. S. Simms and A. Kornberg. *J. Biol. Chem.* **233**, 163 (1958).
[36] P. C. Zamecnik. *Harvey Lectures, LIV*, 256 (1958–1959).
[37] J. H. Matthaei and M. W. Nirenberg. *Proc. Natl. Acad. Sci. U.S.* **47**, 1580 (1961).
[38] A. Falaschi, J. Adler and H. G. Khorana. *J. Biol. Chem.* **238**, 3080 (1963).
[39] A. Kornberg, L. L. Bertsch, J. F. Jackson and H. G. Khorana. *Proc. Natl. Acad. Sci. U.S.* **51**, 315 (1964).
[40] H. G. Khorana, H. Buchi, T. M. Jacob, H. Kossel, S. A. Narang and E. Ohtsuka. *J. Am. Chem. Soc.* **89**, 2154 (1967).
[41] C. Byrd, E. Ohtsuka, M. W. Moon and H G. Khorana. *Proc. Natl. Acad. Sci. U.S.* **53**, 79 (1965).
[42] R. D. Wells, E. Ohtsuka and H. G. Khorana. *J. Mol. Biol.* **14**, 221 (1965).
[43] R. D. Wells, T. M. Jacob, S. A. Narang and H. G. Khorana. *J. Mol. Biol.* **27**, 237 (1967).
[44] R. D. Wells, H. Buchi, H. Kossel, E. Ohtsuka and H. G. Khorana. *J. Mol. Biol.* **27**, 265 (1967).
[45] H. G. Khorana. *Fed. Proc.* **24**, 1473 (1965).
[46] H. G. Khorana, in *Genetic Elements—Properties and Functions*, edited by D. Shugar, Academic Press, 1967, p. 209.
[47] H. G. Khorana, H. Buchi, H. Ghosh, N. Gupta, T. M. Jacob, H. Kossel, A. R. Morgan, S. A. Narang, E. Ohtsuka and R. D. Wells. *Cold Spring Harbor Symp. Quant. Biol.* **31**, 39 (1966).
[48] H. G. Khorana, in *The Harvey Lectures*, Series 62, p. 79, Academic Press, N.Y., 1968.

[49] *Cold Spring Harbor Symposium* Vol. **31**, 1966.
[50] H. G. Khorana, *Proceedings of the Seventh International Congress of Biochemistry*, Tokyo, August 19–25, 1967, p. 17 (1968).
[51] H. G. Khorana, H. Buchi, M. H. Caruthers, S. H. Chang, N. K. Gupta, A. Kumar, E. Ohtsuka, V. Sgaramella, and H. Weber. *Cold Spring Harbor Symp. Quant. Biol.*, 1968, in press.
[52] U. L. RajBhandry, S. H. Chang, A. Stuart, R. D. Faulkner, R. M. Hoskinson and H. G. Khorana. *Proc. Natl. Acad. Sci. U.S.* **57**, 751 (1967).
[53] 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).
[54] R. Naylor and P. T. Gilham. *Biochemistry* **5**, 2722 (1966).
[55] B. S. Zimmerman, J. W. Little, C. K. Oshinsky and M. Gellert. *Proc. Natl. Acad. Sci. U.S.* **57**, 1841 (1967).
[56] B. Weiss and C. C. Richardson. *Proc. Natl. Acad. Sci. U.S.* **57**, 1021 (1967).
[57] B. M. Olivera and I. R. Lehman. *Proc. Natl. Acad. Sci. U.S.* **57**, 1426 (1967).
[58] M. L. Gefter, A. Becker and J. Hurwitz. *Proc. Natl. Acad. Sci., U.S.* **58**, 240 (1967).
[59] N. R. Cozzarelli, N. E. Melechen, T. M. Jovin and A. Kornberg. *Biochem. Biophys. Res. Comm.* **28**, 578 (1967).
[60] N. K. Gupta, E. Ohtsuka, H. Weber, S. H. Chang and H. G. Khorana. *Proc. Natl. Acad. Sci. U.S.* **60**, 285 (1968).
[61] N. K. Gupta, E. Ohtsuka, V. Sgaramella, H. Buchi, A. Kumar, H. Weber and H. G. Khorana. *Proc. Natl. Acad. Sci., U.S.*, **60**, 1338 (1968).
[62] N. K. Gupta and H. G. Khorana. *Proc. Natl. Acad. Sci. U.S.*, **61**, 215 (1968.)
[63] C. C. Richardson, C. L. Schildkraut, H. V. Aposhian and A. Kornberg. *J. Mol. Biol.* **9**, 46 (1964).
[64] R. Wu and D. Kaiser. *J. Mol. Biol.* in press and private communication.