VI, and VIII for the preparation of the tri-, tetra-, penta-, hexa-, and octanucleotides, respectively. After completion of the reaction an equal volume of water was added to the reaction solution, which was set aside for 18 hr. Concentrated ammonia solution (an equal volume) was then added for 1 hr, after which time the solution was evaporated to a gum, dissolved in aqueous pyridine, and chromatographed on a DEAE-cellulose column under the conditions described in the text for the individual reactions studied.

**Synthesis of the Tetranucleotide Tr-TpTpTpT, Using Picryl Chloride.** Pyridinium Tr-TpT (22 mg, 20.8 μmoles) and triethylammonium pTpT-OAc (25 mg, 29.6 μmoles) were dissolved in anhydrous pyridine (2 ml) and dried by repeated evaporation of pyridine. The residual gum was dissolved in anhydrous pyridine (1 ml), and picryl chloride (28 mg, 0.11 mmole) was added. The picryl chloride was only partly soluble at the start of reaction but had dissolved after 1 day. Samples (about 0.2 ml) were removed after 6, 24, and 48 hr. Each sample was diluted with water (0.5 ml) and extracted with ether (two 2.5-ml portions). A 2 *N* sodium hydroxide solution (0.3 ml) was added to the aqueous layer and after 15 min at room temperature an excess of pyridinium Dowex 50 ion-exchange resin was added. An aliquot of the sample was analyzed following paper chromatography in solvent A. The yield of Tr-TpTpTpT estimated in this way was: 6 hr, 24%; 24 hr, 44%; and 48 hr, 51%.

**Synthesis of the Pentanucleotide Tr-TpTpTpTpT Using Picryl Chloride.** Triethylammonium Tr-TpTpT (375 $OD_{267}$ units) and triethylammonium pTpT-OAc (45 mg, 53 μmoles) were dissolved in anhydrous pyridine and rendered anhydrous by evaporation of pyridine. The dry gum was dissolved in anhydrous pyridine (1 ml) and picryl chloride (36 mg, 0.14 mmole) was added. The sealed reaction mixture was shaken at room temperature for 5 days. Water (3 ml) was added and the mixture was set aside at room temperature overnight. The aqueous layer was extracted with ether (three 10-ml portions) and treated with concentrated ammonium hydroxide solution (2 ml) for 1 hr. Ammonia was evaporated, and the product was dissolved in aqueous pyridine and applied to the top of a DEAE-cellulose column (45 × 1 cm, bicarbonate form) which was eluted with a linear salt gradient (3 l. of 0.02 *M* ammonium bicarbonate in the mixing vessel and 3 l. of 0.3 *M* salt in the reservoir); 15-ml fractions were collected every 20 min. The pure pentanucleotide was eluted in fractions 165–190 and amounted to 312 OD units (50% of theoretical).

**Table XI.** Characterization of Products by Enzymic Degradation

| Compd | Enzyme | $OD_{267}$ units degraded | Nucleoside: nucleotide | Theor ratio |
|---|---|---|---|---|
| TpTpTpT | Venom | 4.7 | 1:3.5 | 1:3 |
| TpTpTpT | Spleen | 3.75 | 1:3.1 | 1:3 |
| TpTpTpTpTpT | Venom | 11.5 | 1:5.3 | 1:5 |
| TpTpTpTpTpT | Spleen | 5.5 | 1:4.5 | 1:5 |
| TpTpTpTpTpTpTpT | Venom | 14.5 | 1:6.8 | 1:7 |

**Characterization of Products.** The trityl-containing products were characterized as pure by chromatography in at least two solvent systems (Table X) and by electrophoresis. Their chromatographic behavior was also compared with authentic samples previously prepared[4a] at pH 7.1 (Table X). Following detritylation they were again shown to be pure on paper chromatography (Table X). The tetra-, hexa-, and octanucleotides obtained by detritylation were completely characterized by degradation with venom phosphodiesterase and spleen phosphodiesterase (Table XI).

# Studies on Polynucleotides. LXI.[1] Polynucleotide Synthesis in Relation to the Genetic Code. General Introduction[2]

**H. G. Khorana, H. Büchi, T. M. Jacob, H. Kössel,[3] S. A. Narang, and E. Ohtsuka**

*Contribution from the Institute for Enzyme Research, University of Wisconsin, Madison, Wisconsin 53706.* *Received December 12, 1966*

**Abstract:** An approach to the study of the genetic code has previously been developed which involves the use of ribopolynucleotides containing known repeating nucleotide sequences as messengers in the cell-free protein-synthesizing system. With the aim of a rather systematic extension of this approach for further studies of the genetic code, chemical synthesis of a variety of deoxyribopolynucleotides containing repeating tri- and tetranucleotide sequences has been undertaken. This introductory paper examines the general considerations underlying the synthetic work which is described in a series of accompanying papers. Biochemical accomplishments and experiments made possible by the availability of DNA-like and RNA-like polymers containing known sequences are briefly reviewed.

The approaches to the study of the genetic code which have been developed in this laboratory[4–7] have comprised investigations along the following lines: (1) chemical synthesis of several series of short-chain deoxyribopolynucleotides containing repeating di- and trinucleotide sequences;[8–11] (2) use of chemically synthesized deoxyribonucleotides with repeating nucleotide sequences as templates for DNA-dependent RNA polymerase and characterization of the products as high molecular weight ribopolynucleotides containing repeating nucleotide sequences;[5,12,13] (3) use of chem-

(1) Paper LX: H. Kössel, M. W. Moon and H. G. Khorana, *J. Am. Chem. Soc.*, **89**, 2148 (1967).

(2) This work has been supported by grants from the National Science Foundation (Grant No. GB-3342), the National Cancer Institute of the National Institutes of Health, U. S. Public Health Service (Grant No. CA-05178), and the Life Insurance Medical Research Fund (Grant No. G-62-54).

(3) H. K. wishes to acknowledge the receipt of a postdoctoral fellowship (1964–1966) from the Deutscher Akademischer Austauschdienst, Bad Godesberg, Germany.

(4) H. G. Khorana, T. M. Jacob, M. W. Moon, S. A. Narang, and E. Ohtsuka, *J. Am. Chem. Soc.*, 87, 2954 (1965).

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

(6) H. G. Khorana, Proceedings of the Third Meeting of the Federation of European Biochemical Societies, Warsaw, April 1966, in press.

(7) H. G. Khorana, H. Büchi, H. Ghosh, N. Gupta, T. M. Jacob, H. Kössel, A. R. Morgan, S. A. Narang, E. Ohtsuka, and R. D. Wells, *Cold Spring Harbor Symp. Quant. Biol. (June 1966)*, in press.

(8) E. Ohtsuka, M. W. Moon, and H. G. Khorana, *J. Am. Chem. Soc.*, 87, 2956 (1965).

(9) T. M. Jacob and H. G. Khorana, *ibid.*, 87, 2971 (1965).

(10) S. A. Narang and H. G. Khorana, *ibid.*, 87, 2981 (1965).

(11) S. A. Narang, T. M. Jacob, and H. G. Khorana, *ibid.*, 87, 2988 (1965).

(12) A. Falaschi, J. Adler, and H. G. Khorana, *J. Biol. Chem.*, **238**, 3080 (1963); B. D. Mehrotra and H. G. Khorana, *ibid.*, **240**, 1750 (1965).

(13) S. Nishimura, T. M. Jacob, and H. G. Khorana, *Proc. Natl. Acad. Sci. U. S.*, **52**, 1494 (1964).

ically synthesized deoxyribopolynucleotides containing repeating mono- and dinucleotide sequences as templates for DNA polymerase and characterization of the products of the enzymatic reaction as high molecular weight double-stranded DNA-like polymers containing the same repeating nucleotide sequences as were present in the short-chain templates;[14,15] (4) use, in turn, of the high molecular weight DNA-like polymers as templates for DNA-dependent RNA polymerase and the preparation, in this way, of long single-stranded ribopolynucleotides containing repeating nucleotide sequences;[16] (5) use of the above-prepared ribopolynucleotides with defined repeating di- and trinucleotide sequences as messengers in the cell-free protein-synthesizing system and complete characterization of the resulting polypeptidic products;[16-18] (6) chemical synthesis of all of the 64 possible ribotrinucleotides derivable from the four major ribomononucleotides[19] and extensive testing of all of the trinucleotides according to the technique of Nirenberg and Leder[20] with a view to the assignment of codon sequences for the 20 amino acids.[21]

The principal types of results obtained in the above studies may be summarized. First, the extensive correlations of the nucleotide sequences of the deoxyribopolynucleotides and of the ribopolynucleotides prepared from them with the amino acid sequences of the polypeptidic products formed in the cell-free protein-synthesizing system have unambiguously proven the following basic features of the genetic code:[5-7] (1) a linear sequence of deoxyribonucleotides in DNA specifies the linear sequence of amino acids in a polypeptide chain, and this information is indeed expressed through the intermediate formation of an RNA (messenger) under the direction of DNA; (2) a group of three nucleotide units in a linear sequence specifies the incorporation of a particular amino acid (three-letter code); (3) each nucleotide unit in a polynucleotide chain is used only once in forming groups of three nucleotides (nonoverlapping property of the code); finally (4) the reading of a polynucleotide chain occurs in units of (codons) three letters without omission of a single nucleotide unit. Second, a major application of the above general approach has been in the derivation of nucleotide sequences within the codons corresponding to different amino acids.[5-7] Third, the availability of precise transcription–translation (DNA → RNA → protein) systems has made possible studies at the biochemical level of a number of biologically important questions such as genetic suppression,[22,23] the misreading of the genetic code in the presence of antibiotics,[24] and polypeptide chain initiation[25] (see below).

It is therefore clear that the general approach involving preparation of DNA-like and RNA-like polymers offers a wide scope in further studies of the genetic code and of the chemistry and enzymology of nucleic acids themselves. Consequently, a sytematic extension of the previous work dealing with the preparation of polynucleotides with known repeating sequences has been undertaken. The present paper, which serves as a general introduction to a series of accompanying papers,[26-30] reviews some of the considerations for the synthesis of deoxyribopolynucleotides containing repeating tri- and tetranucleotide sequences.

**The Preparation of DNA-like Polymers.** The starting point in the synthesis of ribopolynucleotides with repeating nucleotide sequences is the chemical synthesis of short-chain deoxyribopolynucleotides. These products may either be used directly as templates for DNA-dependent RNA polymerase or, alternatively, they may first be used as templates for DNA polymerase and the resulting DNA-like polymers may subsequently be used for transcription by RNA polymerase. Studies with both polymerases have shown the approach involving the prior synthesis of DNA-like polymers by the use of DNA polymerase to be by far the most satisfactory and desirable. Briefly, the important advantages of DNA polymerase catalyzed synthesis are: (1) extensive synthesis; (2) the fact that the products are of high molecular weight containing faithfully replicated sequences of the short-chain templates; (3) the possibility of reutilization of the DNA-like polymers as templates for DNA polymerase for synthesis of more of the same polymers, thus ensuring their permanent availability.

Further, the availability of a variety of DNA-like polymers with completely defined sequences would enable systematic physicochemical studies of DNA structure and, finally, at the enzymatic level, the DNA-like polymers would offer well-defined systems for studies of the mechanisms of action of DNA-dependent RNA polymerase and of DNA polymerase itself.

Thus the case for the "amplification multiplication" of the chemically synthesized deoxyribopolynucleotides to DNA-like polymers with DNA polymerase is overwhelmingly strong. However, an unfavorable feature of the action of this enzyme has been that short chains of deoxyribopolynucleotides corresponding to one strand alone have so far failed to prime the synthesis of the DNA-like polymers. In other words, the enzyme-catalyzed reactions have proceeded only when deoxyribopolynucleotide templates corresponding to both of the complementary strands of the required DNA-like product are provided. Consequently all of the syntheses described in the accompanying papers consist of sets such that every set contains two series of deoxyribopolynucleotides which are complementary with one another in the antiparallel Watson–Crick base-pairing sense.

(14) C. Byrd, E. Ohtsuka, M. W. Moon, and H. G. Khorana, *ibid.*, **53**, 79 (1965).
(15) R. D. Wells, E. Ohtsuka, and H. G. Khorana, *J. Mol. Biol.*, 14, 221 (1965).
(16) S. Nishimura, D. S. Jones, and H. G. Khorana, *ibid.*, **13**, 302 (1965).
(17) S. Nishimura, D. S. Jones, E. Ohtsuka, H. Hayatsu, T. M. Jacob, and H. G. Khorana, *ibid.*, **13**, 283 (1965).
(18) D. S. Jones, S. Nishimura, and H. G. Khorana, *ibid.*, **16**, 454 (1966).
(19) R. Lohrmann, D. Söll, H. Hayatsu, E. Ohtsuka, and H. G. Khorana, *J. Am. Chem. Soc.*, **88**, 819 (1966).
(20) M. Nirenberg and P. Leder, *Science*, **145**, 1399, (1964).
(21) D. Söll, E. Ohtsuka, D. S. Jones, R. Lohrmann, H. Hayatsu, S. Nishimura, and H. G. Khorana, *Proc. Nat. Acad. Sci. U. S.*, 54, 1378 (1965).
(22) N. K. Gupta and H. G. Khorana, *ibid.*, **56**, 772 (1966).
(23) J. Carbon, P. Berg, and C. Yanofsky, *ibid.*, **56**, 764 (1966).
(24) J. Davies, D. S. Jones, and H. G. Khorana, *J. Mol. Biol.*, **18**, 48 (1966).
(25) H. P. Ghosh, D. Söll, and H. G. Khorana, *ibid.*, in press.
(26) S. A. Narang, T. M. Jacob, and H. G. Khorana, *J. Am. Chem. Soc.*, **89**, 2158 (1967).
(27) S. A. Narang, T. M. Jacob, and H. G. Khorana, *ibid.*, **89**, 2167 (1967).
(28) T. M. Jacob, S. A. Narang, and H. G. Khorana, *ibid.*, **89**, 2177 (1967).
(29) H. Kössel, H. Büchi, and H. G. Khorana, *ibid.*, **89**, 2185 (1967).
(30) E. Ohtsuka and H. G. Khorana, *ibid.*, **89**, 2195 (1967).

Finally, for the preparation of the single-stranded ribopolynucleotide messengers from the double-stranded DNA-like polymers, DNA-dependent RNA polymerase is used. The copying by Watson–Crick base-pairing principle may be restricted to either of the two strands of the DNA-like polymers by providing as substrates only those ribonucleoside triphosphates which are necessary for the copying of a single strand. This principle was used previously in the preparation of the four ribopolynucleotides with repeating dinucleotide sequences from the two double-stranded DNA-like polymers, and the same method is applicable whenever the DNA-like polymers with repeating sequences contain at the most three different bases in each strand.[4–7] This condition is met by all of the deoxyribopolynucleotides whose syntheses are described in the accompanying series of papers. Examples of the transcription of individual strands of DNA-like polymers with repeating tetranucleotide sequences are given later.

**Polynucleotides with Repeating Trinucleotide Sequences.** A primary concern of more recent work on the genetic code has been to determine unambiguously the nucleotide sequences within the trinucleotide codons which stand for different amino acids. An important advance in this field was made by Nirenberg and Leder[20] by their discovery that a ribotrinucleotide, which represents the codon for an amino acid, can specifically direct the binding of the corresponding aminoacyl-sRNA to ribosomes. Extensive use of this approach permitted many assignments rather conclusively;[21,31] however, the technique proved not to be completely reliable. Often the effects observed were too small and, in certain cases, ambiguous binding by one trinucleotide of more than one aminoacyl-sRNA was observed. In addition, certain trinucleotides which independently had been shown to be authentic codons failed to give any detectable effect in the binding test.[21,31] On the other hand, the above approach, in which amino acid incorporations directed by ribopolynucleotides containing known nucleotide sequences are determined, has given no evidence of any artifacts. Therefore, it was decided to try to *prove* as many of the codon assignments as possible by preparing ribopolynucleotide messengers containing repeating nucleotide sequences.

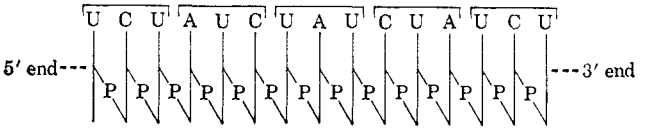
Of the ribopolynucleotides containing repeating dinucleotide sequences, there are a maximum of only four types which are of interest as messengers (for a detailed discussion see ref 4). All of these have previously been prepared and successfully used in amino acid incorporation experiments. Of the class of polymers containing repeating trinucleotide sequences, containing more than one type of nucleotide, the maximum number derivable from the four common mononucleotides is 20 (*cf.* the following paper). Of these, only one member containing the repeating trinucleotide sequence adenylyladenylylguanylyl has been previously prepared and used in amino acid incorporation experiments.[13,17] Further work on the genetic code clearly required the preparation of a wide variety of ribopolynucleotides containing repeating trinucleotide sequences. Consequently a major aim of the present work has been the synthesis of deoxyribopolynucleotides containing repeating trinucleotide sequences, which would serve as templates for the DNA-polymerase reactions.

Additional considerations for the synthesis of DNA-like polymers with repeating trinucleotide sequences are given later.

**Ribopolynucleotides with Repeating Tetranucleotide Sequences.** The three-letter, nonoverlapping properties of the genetic code dictate that a ribopolynucleotide containing a repeating tetranucleotide sequence direct the synthesis of a polypeptide containing four amino acids (Chart I). Although the amino acids at

**Chart I.** The Formation of a Polypeptide with Four Amino Acids in Repeating Sequence as Directed by a Ribopolynucleotide Containing a Repeating Tetranucleotide Sequence

COOH end
– – – serine–isoleucine–tyrosine–leucine–serine – – – $H_2N$ end
or
$N_2H$ end
– – – serine–isoleucine–tyrosine–leucine–serine – – –
COOH end




the ends will be determined by where the reading started and ended, the polypeptide product will have the repeating tetrapeptide sequence no matter where the reading of the messenger started and ended (Chart I). First of all, actual demonstration that, with these polymers, only four amino acids are in fact incorporated and that the polypeptidic products contain the amino acids in a uniquely repeating sequence would further confirm the conclusions already drawn about the basic properties of the genetic code. Second, the nucleotide sequences in the repeating tetranucleotide polymers could be so designed that they contain codons about whose assignment uncertainty still prevailed. The work with the repeating tetranucleotide polymers would thus augment the work on codon assignments using polymers with repeating di- and trinucleotide sequences. Further considerations for the synthesis of repeating tetranucleotide polymers were as follows.

**Direction of Reading of Messenger RNA.** Direction of polypeptide chain growth in protein biosynthesis is established to be from the $\alpha$-amino end to the carboxyl end. The question of the direction of the reading of the messenger RNA for the synthesis of a polypeptide chain has recently been studied by a number of approaches.[32–37] At the time that the present work was undertaken, the two sets of results had led to opposite conclusions.[32–35] It was considered that the structural characterization of the polypeptidic product formed under the direction of a polynucleotide containing a repeating tetranucleotide sequence such as shown in Chart I should give an independent and

(31) R. Brimacombe, J. Trupin, M. Nirenberg, P. Leder, M. Bernfield, and T. Jaouni, *Proc. Natl. Acad. Sci. U. S.*, 54, 954 (1965), and earlier papers in this series.

(32) M. Salas, M. A. Smith, W. M. Stanley, A. J. Wahba, and S. Ochoa, *J. Biol. Chem.*, 240, 3988 (1965).

(33) R. E. Thach, M. A. Cecere, T. A. Sundarajan, and P. Doty, *Proc. Nat. Acad. Sci. U. S.*, 54, 1167 (1965).

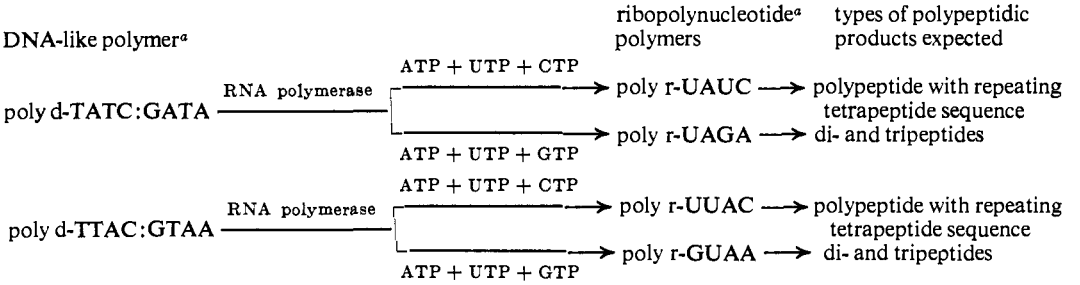
(34) E. F. Eikenberry and A. Rich, *ibid.*, 53, 668 (1965).

(35) A. R. Williamson and R. Schweet, *Nature*, 206, 29 (1965).

(36) E. Terzaghi, Y. Okada, G. Streisinger, J. Emrich, M. Inouye, and A. Tsugita, *Proc. Natl. Acad. Sci. U. S.*, 56, 500 (1966).

(37) J. R. Guest and C. Yanofsky, *Nature*, 210, 799 (1966).

**Chart II.** Reaction Sequence Starting with DNA-like Polymers Containing Repeating Tetranucleotide Sequences for the Study of Polypeptide Chain Termination




[a] 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. The repeating tetranucleotide sequences in the DNA-like and RNA-like polymers are shown simply by the one-letter abbreviations for the four bases.

unequivocal answer to this important question. If the reading of the messenger RNA proceeded from the 5′ end (left to right in Chart I), then the polypeptide chain formed will contain isoleucine residues at the amino end of tyrosine units and leucine residues at the carboxyl end of tyrosine units. If the reading of the messenger RNA proceeded from the 3′ end of the polynucleotide chain, then the polarity of the peptidic linkages would be reversed and, therefore, leucine residues would occur at the amino ends of tyrosine and isoleucine residues at the carboxyl ends of the tyrosine units. The requirement would therefore be one of determining the amino acid sequence within the repeating tetrapeptide units of the polypeptidic products.

**Polypeptide Chain Termination.** It has recently become apparent from genetic work that two trinucleotide sequences, UpApA and UpApG, cause polypeptide chain termination in *Escherichia coli* B.[38,39] The repeating tetranucleotide sequences of the deoxyribopolynucleotides to be synthesized were so chosen that ribopolynucleotides prepared from them would contain, in one case, the trinucleotide UpApA and, in the second case, the trinucleotide UpApG, as one of the four repeating trinucleotide units. Shown in Chart II are the double-stranded DNA-like polymers which were to be synthesized, the ribopolynucleotide messengers which these polymers would afford, and the types of polypeptides which, in turn, would be expected in the cell-free protein-synthesizing system. It is seen that the ribopolynucleotide polymers poly r-UAUC (see also above, Chart I) and poly r-UUAC would be expected to direct the synthesis of "continuous" polypeptides with repeating tetrapeptide sequence whereas poly r-UAGA and poly r-GUAA, which contain the trinucleotides UAG and UAA in every fourth place, would, at most, give di- and tripeptides. The results should provide direct confirmation or otherwise of the conclusions reached about UAG and UAA from genetic evidence. Furthermore, if the di- and tripeptides could be shown to be formed, then simple and well-defined systems would be available for study of the mechanism of polypeptide chain termination.

**Genetic Suppression.** Another biologically important process is that of genetic suppression. This process, which works at the translation (RNA → protein) step, can often correct the damage due to a harmful mutation involving a single base change. The mutation may give rise to a nonsense or chain-terminating codon such that premature polypeptide chain termination would occur. Alternatively, a single base change may lead to a change in the original codon such that a different amino acid is inserted in the polypeptide chain (missense mutation). Suppressor genes are known which correct the effects of both types of mutation. Thus the premature chain termination caused by the codons UAG and UAA is now known to be overcome by two distinct classes of suppressors.[40-42] It is clear that the polymers poly r-UAGA and poly r-GUAA (Chart II), when used in protein-synthesizing systems prepared from bacterial strains which contain the suppressors, should give rise to polypeptides longer than the di- and tripeptides. If so, then the polymers with repeating tetranucleotide sequences would offer simple systems for *in vitro* studies of nonsense suppression.

Studies of the mechanism of missense suppression using ribopolynucleotides containing repeating dinucleotide sequences have already been reported from two laboratories, and the results obtained have permitted clear-cut conclusions.[22,23] It is clear that studies along these lines can be extended using a variety of ribopolynucleotides containing repeating nucleotide sequences.

**Additional General Considerations.** The availability of ribopolynucleotides containing known sequences has made possible studies of a number of additional biological problems, and the range of such studies continues to grow. Some lines of investigation may be briefly cited.

Like polypeptide chain termination, the initiation of polypeptide chains is governed by certain trinucleotide sequences. Ribopolynucleotides containing repeating dinucleotide sequences (available from previous work) and certain ribopolynucleotides containing repeating trinucleotide sequences (available from work described in accompanying papers) have already given definitive information from cell-free polypeptide synthesis experiments on the nature of the codons which signal polypeptide chain initiation.[25] It is hoped that polymers containing certain repeating tetranucleotide sequences would enable further studies of this process.

Amino glycoside antibiotics such as streptomycin have been shown to cause errors in translation step by

(38) S. Brenner, A. O. W. Stretton, and S. Kaplan, *Nature*, **206**, 994 (1965).
(39) M. G. Weigert and A. Garen, *ibid.*, **206**, 992 (1965).
(40) G. W. Notani, D. L. Engelhardt, W. Konigsberg, and N. D. Zinder, *J. Mol. Biol.*, 12, 439 (1965).
(41) M. G. Weigert and A. Garen, *ibid.*, 12, 448 (1965).
(42) A. O. W. Stretton and S. Brenner, *ibid.*, 12, 456 (1965).

provoking misreading of the genetic message.[24,43] Similar antibiotics such as neomycin B have been shown to elicit messenger response from denatured DNA.[44] Ribopolynucleotides containing repeating dinucleotide sequences and certain single-stranded deoxyribopolynucleotides prepared by separation of the double-stranded DNA-like polymers have already permitted more precise studies of these processes.[45] The availability of a large selection of polynucleotides with repeating tri- and tetranucleotide sequences would permit an extension of these studies.

A series of accompanying papers describe the work on the synthesis of deoxyribopolynucleotides with repeating tri- and tetranucleotide sequences.[26-30] The synthetic deoxyribonucleotides have been used successfully in the synthesis of DNA polymerase catalyzed synthesis of DNA-like polymers containing the appropriate repeating sequences. The synthesis and characterization of the DNA-like polymers and their physicochemical properties are being reported elsewhere.[46-48] The DNA-like polymers have afforded, in turn, the expected sets of single-stranded ribopolynucleotides.[49] The use of the resulting ribopolynucleotides as messengers in the cell-free protein-synthesizing system have given new information, in particular on the codons whose assignments have hitherto been uncertain. As a consequence, essentially all of the genetic code is now established with certainty for the microorganism *E. coli* B.[50] Other biological studies made possible by the availability of the DNA-like polymers will be reported subsequently.

(43) J. Davies, W. Gilbert, and L. Gorini, *Proc. Natl. Acad. Sci. U. S.*, 51, 883 (1964).
(44) B. J. McCarthy and J. J. Holland, *ibid.*, 54, 880 (1965).
(45) A. R. Morgan, R. D. Wells, and H. G. Khorana, *J. Mol. Biol.*, in press.
(46) R. D. Wells, T. M. Jacob, S. A. Narang, and H. G. Khorana, *ibid.*, in press.
(47) R. D. Wells, H. Büchi, H. Kössel, E. Ohtsuka, and H. G. Khorana, *ibid.*, in press.
(48) R. D. Wells and J. Blair, *ibid.*, in press.
(49) H. Kössel A. R. Morgan, and H. G. Khorama, *ibid.*, in press.
(50) A. R. Morgan, R. D. Wells, and H. G. Khorana, *Proc. Natl. Acad. Sci. U. S.*, **56**, 1899 (1966).

# Studies on Polynucleotides. LXII.[1] Deoxyribopolynucleotides Containing Repeating Trinucleotide Sequences (4).[2] Preparation of Suitably Protected Deoxyribotrinucleotides[3]

**S. A. Narang, T. M. Jacob, and H. G. Khorana**

*Contribution from the Institute for Enzyme Research, University of Wisconsin, Madison, Wisconsin 53706.* *Received December 12, 1966*

**Abstract:** Suitably protected deoxyribotrinucleotides were required as starting materials for the preparation of deoxyribopolynucleotides containing repeating trinucleotide sequences. Using the stepwise method illustrated in Chart I, the syntheses of the following 13 protected deoxyribotrinucleotides[4] have been accomplished: d-pTpA$^{Bz}$pC$^{An}$, d-pTpA$^{Bz}$pG$^{Ac}$, d-pA$^{Bz}$pTpC$^{An}$, d-pA$^{Bz}$pTpG$^{Ac}$, d-pC$^{An}$pG$^{Ac}$pA$^{Bz}$, d-pC$^{An}$pG$^{Ac}$pT, d-pTpTpG$^{Ac}$, d-pA$^{Bz}$pA$^{Bz}$pC$^{An}$, d-pA$^{Bz}$pA$^{Bz}$pG$^{Ac}$, d-pC$^{An}$pC$^{An}$pA$^{Bz}$, d-pC$^{An}$pC$^{An}$pT, d-pG$^{Bz}$pG$^{Bz}$pA$^{Bz}$, and d-pG$^{Bz}$pG$^{Bz}$pT. Acetylation in acetic anhydride–pyridine gave the corresponding 3′-O-acetyl derivatives. All the trinucleotides were characterized to be pure with and without the protecting groups.

As discussed in the preceding introductory paper,[1] the synthesis of DNA-like polymers containing repeating trinucleotide sequences is of interest for further studies of the genetic code. It is also clear that such DNA-like polymers would be useful for further physicochemical studies of DNA. The maximum number of double-stranded DNA-like polymers with repeating trinucleotide sequences which can be formed from the four common deoxyribomononucleotides and which contain more than one base in each strand are ten. These are listed in Table I. Copying of the individual strands in these polymers by DNA-dependent RNA

(1) Preceding paper in this series: H. G. Khorana, H. Büchi, T. M. Jacob, H. Kössel, S. A. Narang, and E. Ohtsuka, *J. Am. Chem. Soc.*, **89**, 2154 (1967).
(2) Previous papers dealing with the preparation of deoxyribopolynucleotides with repeating trinucleotide sequences are: (a) T. M. Jacob and H. G. Khorana, *ibid.*, **87**, 2971 (1965); (b) S. A. Narang and H. G. Khorana, *ibid.*, **87**, 2982 (1965); (c) S. A. Narang, T. M. Jacob, and H. G. Khorana, *ibid.*, **87**, 2988 (1965).
(3) This work has been supported by grants from the National Science Foundation (Grant No. GB-3342), the National Cancer Institute of the National Institutes of Health, U. S. Public Health Service (Grant No. CA-05178), and the Life Insurance Medical Research Fund (Grant No. G-62-54).
(4) For convenience and economy of space, abbreviations are used extensively for both the protected and unproctected series of compounds described in this and the accompanying papers. The basic system of abbreviations used for polynucleotides and their protected derivatives is as has been used in previous papers in this series and is in current use in *Biochemistry* and *The Journal of Biological Chemistry*. Thus the single letters A, T, C, and G represent the nucleosides of, respectively, adenine, thymine, cytosine, and guanine. The letter p to the left of the nucleoside initial indicates a 5′-phosphomonoester group and the same letter to the right indicates a 3′-phosphomonoester group. Thus, in going from the left to the right the polynucleotide chain is specified in the 3′→5′ direction. The protecting groups on the purine or pyrimidine rings are designated by two-letter abbreviations added as superscripts after the nucleoside initial: thus d-A$^{Bz}$ for N-benzoyldeoxyadenosine, C$^{An}$ for N-anisoylcytidine, G$^{Ac}$ for N-acetylguanosine. The acetyl group at the 3′-hydroxyl group of a nucleoside is shown by -OAc added after the nucleoside initial. Thus, pT-OAc is the abbreviation of 3′-O-acetylthymidine 5′-phosphate. CE is the abbreviation for β-cyanoethyl: thus, d-CE-pG$^{Ac}$ stands for N-acetyldeoxyguanosine 5′-β-cyanoethyl phosphate. Using this system, the abbreviation d-pC$^{An}$pG$^{Ac}$pA$^{Bz}$-OAc stands for the protected trinucleotide, 5′-O-phosphoryl-N-anisoyldeoxycytidylyl-(3′→5′)-N-acetyldeoxyguanylyl-(3′→5′)-3′-O-acetyl-N-benzoyldeoxyadenosine.