Studies on Polynucleotides. XLII.[1] The Synthesis of Deoxyribopolynucleotides Containing Repeating Nucleotide Sequences.[2] Introduction and General Considerations[3]

H. G. Khorana, T. M. Jacob, M. W. Moon, S. A. Narang, and E. Ohtsuka

Contribution from the Institute for Enzyme Research of the University of Wisconsin, Madison, Wisconsin. *Received March 22, 1965*

Deoxyribopolynucleotides of repeating di- and trinucleotide sequences are of interest in the study of the amino acid code. The total considerations for chemical and enzymatic work with this general class of compounds are presented. Accomplishments of enzymatic work already made possible by their availability are reviewed.

Until recently a major concern of synthetic work in the deoxyribopolynucleotide field has been the development of total methodology for the assembly of the different mononucleotide units to form C_3'–C_5'-linked polynucleotides. Thus, in the work reported previously in this series of papers, we have devoted attention to (1) the development of satisfactory methods for the activation of the phosphomonoester group in a mononucleotide so as to cause the phosphorylation of the hydroxyl group of another nucleoside or nucleotide,[4-6] (2) the design of suitable protecting groups for the various functional groups (primary and secondary hydroxyl groups, amino groups, phosphoryl dissociation in phosphomonoester groups),[7] (3) the development of methods for the polymerization of mononucleotides and of methods for the separation and characterization of the resulting polynucleotides,[5,8] and (4) the evaluation of approaches to the stepwise synthesis of deoxyribopolynucleotides containing specific sequences.[7,9] While all these aspects of the synthetic problem continue to be under study, nevertheless, reasonably satisfactory approaches have already emerged for the synthesis of short chains (chain length, between ten and twenty nucleotide units) of deoxyribopolynucleotides containing defined nucleotide sequences. Therefore, recently we have sought to define specific objectives for further synthetic work in this field. The present introductory paper outlines the general objectives for undertaking the synthesis of deoxyribopolynucleotides containing repeating di- and trinucleotide sequences. Accomplishments of biochemical studies made possible by the availability of these two classes of synthetic deoxyribopolynucleotides are also reviewed.

A central thesis in modern biology has been that DNA in all living organisms directs the synthesis of all proteins and that this control is exerted through the intermediate formation of an RNA which contains the information originally present in DNA. The continuity of DNA itself, the genetic material, is secured by the manner of replication originally propounded by Watson and Crick.[10] Three biochemical discoveries in recent years have highlighted the progress which has been made in uncovering the molecular mechanism of the vital biological reactions mentioned. The first is the discovery by Kornberg and co-workers of an enzyme which brings about the synthesis of DNA in the presence of the four deoxyribonucleoside 5′-triphosphates and a DNA template.[11] The second is the discovery of the enzyme DNA-dependent RNA-polymerase which brings about the synthesis of ribonucleic acid from the four ribonucleoside 5′-triphosphates in the presence of DNA[12]; the latter serves as the template in this reaction in that it determines the composition and the sequence of nucleotide units in the synthesized RNA according to the Watson–Crick base-pairing principle. The discovery of this enzyme makes it clear how a linear deoxyribopolynucleotide sequence is "transcribed" into a linear ribopolynucleotide sequence. The third important development in recent years has been the preparation of an *in vitro* amino acid incorporation system[13] which does in fact bring about the formation of polypeptide material in response to certain varieties of RNA.[14] By using in this system enzymatically prepared ribopolynucleotides which, however, contain random sequences of nucleotide units, and analyzing for the extent of incorporation of different amino acids into polypeptides, a great deal has been learned about the "translation"[15] of the information contained in the 4-letter language of the

(1) Paper XLI: R. M. Hoskinson and H. G. Khorana, *J. Biol. Chem.*, in press.

(2) This is an extension of the previous series of papers dealing with the synthesis of specific deoxyribopolynucleotides. Previous paper which deals directly with this topic: T. M. Jacob and H. G. Khorana, *J. Am. Chem. Soc.*, 87, 368 (1965); references to earlier papers are given in this paper.

(3) This work has been supported by grants from the National Science Foundation, (Grant No. GB-976), the National Cancer Institute of the National Institutes of Health (Grant No. CA-05178), and the Life Insurance Medical Research Fund (Grant No. G-62-54).

(4) P. T. Gilham and H. G. Khorana, *J. Am. Chem. Soc.*, **80**, 612 (1958).

(5) H. G. Khorana, "Some Recent Developments in the Chemistry of Phosphate Esters of Biological Interest," John Wiley and Sons, Inc., New York, N. Y., 1961, Chapter 5.

(6) T. M. Jacob and H. G. Khorana, *J. Am. Chem. Soc.*, **86**, 1630 (1964).

(7) For leading references to earlier papers see H. Schaller and H. G. Khorana, *ibid.*, **85**, 3841 (1963).

(8) See, *e.g.*, G. M. Tener, H. G. Khorana, R. Markham, and E. H. Pol, *ibid.*, **80**, 6224 (1958); H. G. Khorana and J. P. Vizolyi, *ibid.*, **83**, 675 (1961); R. K. Ralph, W. J. Connors, H. Schaller, and H. G. Khorana, *ibid.*, **85**, 1983 (1963).

(9) For leading references to earlier papers on this subject see T. M. Jacob and H. G. Khorana, *ibid.*, **87**, 368 (1965).

(10) J. D. Watson and F. H. C. Crick, *Nature*, 171, 737 (1953); *Cold Spring Harbor Symposia Quant. Biol.*, **18**, 123 (1953).

(11) I. R. Lehman, J. J. Bessman, E. S. Simms, and A. Kornberg, *J. Biol. Chem.*, **233**, 163 (1958).

(12) S. B. Weiss and T. Nakamoto, *Proc. Natl. Acad. Sci. U. S.*, **47**, 1400 (1961); A. Stevens, *J. Biol. Chem.*, **236**, PC 43 (1961); J. J. Furth, J. Hurwitz, and M. Anders, *ibid.*, **237**, 2611 (1962); D. P. Burma, H. Kroger, S. Ochoa, R. C. Warner, and J. D. Weill, *Proc. Natl. Acad. Sci. U. S.*, **47**, 749 (1961).

(13) A. Tissieres, D. Schlessinger, and F. Gros, *ibid.*, **46**, 1450 (1960).

(14) J. H. Matthaei and M. W. Nirenberg, *ibid.*, **47**, 1580 (1961); M. W. Nirenberg and J. H. Matthaei, *ibid.*, **47**, 1588 (1961).

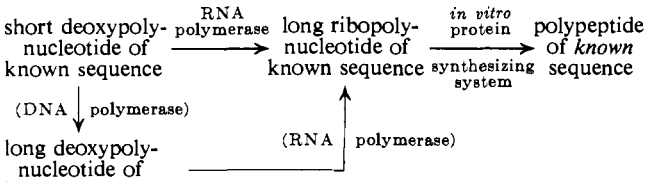
(15) See, for example, F. A. Lipmann in "Progress in Nucleic Acid Research," Vol. 1, J. N. Davidson and W. E. Cohn, Ed., Academic Press Inc., New York, N. Y., 1963, p. 135.

polynucleotide material into the 20-letter language of the proteins (amino acid code).[16,17]

Certain aspects of enzymatic work carried out using synthetic deoxyribopolynucleotides of defined size and structure deserve mention. In studies with DNA polymerase of *Escherichia coli*, it was demonstrated that short-chain synthetic deoxyribopolynucleotides containing alternating deoxyadenylate and thymidylate units serve as templates and bring about the synthesis of the previously characterized high molecular weight DNA-like polymer containing again deoxyadenylate and thymidylate units in alternating sequence.[18] In experiments with DNA-dependent RNA polymerase, short-chain thymidine polynucleotides were found to serve as templates and to bring about the specific synthesis of ribopolyadenylate from adenosine 5′-triphosphate. As with the reaction mentioned above with DNA polymerase, an important and practically advantageous feature of this reaction was that the size of the ribopolyadenylate formed was much larger than that of the deoxypolynucleotide template used in the reaction.[19] Furthermore, of interest in the study of the amino acid code was the experimental verification of the expectation that the ribopolyadenylate product formed under the influence of deoxypolythymidylate should bring about the synthesis of polylysine in the *in vitro* protein synthesizing system.[20]

The above results with the enzymatic systems were available to us at the outset of the present series of investigations. These results encouraged the hope that short chains of deoxypolynucleotides *containing repeating but specific* patterns of nucleotide sequences would in general serve as templates for the nucleic acid polymerases and, consequently, large ribopolynucleotides[21] of known sequences would become available for precise studies of the coding problem.[16] The total envisaged scheme for the *in vitro* synthesis of defined peptides using short-chain deoxypolynucleotides of defined nucleotide sequence is shown in Chart I.

Chart I. Proposed Sequence of Enzymatic Reactions for the Synthesis of Specific Polypeptides[a]



[a] Using chemically synthesized specific deoxyribopolynucleotides.

It was also hoped that the availability of a variety of DNA-like polymers with known nucleotide sequences would open up new opportunities for further studies of the chemistry and enzymology of DNA.

Ribopolynucleotides containing repeating di- and trinucleotide sequences would be expected to give clear-cut answers to some of the primary questions regarding the amino acid code.[16] Thus, if (1) it indeed is a "three-letter" code, (2) it is nonoverlapping, and (3) the reading of the contiguous "triplets" occurs sequentially without omission of any nucleotides, then homopeptides should be produced using ribopolynucleotides with repeating trinucleotide sequence. As applied to a polynucleotide with repeating dinucleotide sequence, using the same assumptions, a copeptide containing two amino acids in alternating sequence should result. Furthermore, information should also be forthcoming from these results regarding the sequence of the nucleotides in the different coding units. A further important consideration in favor of the synthesis of polynucleotides with patterns of repeating nucleotide sequences was the fact that the *in vitro* protein-synthesizing system in current use contains powerful nucleases[22,23] (exo- and endonucleases), and a large proportion of the polynucleotide added as "messenger RNA" suffers extensive degradation during the period of incubation used in the amino acid incorporation experiments.[24] The use of polynucleotides containing, for example, repeating di- and trinucleotide sequences circumvents the ambiguity that the presence of nucleases would otherwise create. Despite exonucleolytic and endonucleolytic cleavages, a polynucleotide with a repeating dinucleotide sequence should invariably stimulate the formation of a copeptide with two amino acids in strictly alternating sequence. The use of polynucleotides with repeating trinucleotide sequences should always give homopeptides, although three (or even more if there is ambiguity[16]) types of homopeptides, each composed of a different amino acid, may be formed. Thus, for example, depending upon where the reading starts, a polynucleotide with a repeating trinucleotide sequence, ABCABC, may give homopeptides of the amino acids whose coding triplets are ABC, BCA, and CAB.

The actual choice of mononucleotide combinations in the projected syntheses was primarily determined by the observations reported in literature on the influence of the secondary structure of messenger RNA on its capacity to stimulate amino acid incorporation in the *in vitro* amino acid incorporation system. To be able to form the necessary complex with ribosomes, messenger RNA must not have a pronounced secondary structure of its own.[25] Therefore, in synthetic work

(16) F. H. C. Crick in "Progress in Nucleic Acid Research," Vol. 1, J. N. Davidson and W. E. Cohn, Ed., Academic Press Inc., New York, N. Y., 1963, p. 164.

(17) J. C. Bennet and W. J. Dreyer, *Ann. Rev. Biochem.*, **33**, 205 (1964).

(18) A. Kornberg, L. R. Bertsch, J. Jackson, and H. G. Khorana, *Proc. Natl. Acad. Sci. U. S.*, **51**, 315 (1964).

(19) A. Falaschi, J. Adler, and H. G. Khorana, *J. Biol. Chem.*, **237**, 3752 (1962).

(20) P. Leder, B. F. C. Clark, W. S. Sly, S. Pestka, and M. W. Nirenberg, *Proc. Natl. Acad. Sci. U. S.*, **50**, 1135 (1963).

(21) For full discussion of the reiterative mechanism which permits the synthesis of products longer than the templates see ref. 22 and B. D. Mehrotra and H. G. Khorana, *J. Biol. Chem.*, **240**, 1750 (1965).

(22) I. R. Lehman in "Progress in Nucleic Acid Research," Vol. 2, J. N. Davidson and W. E. Cohn, Ed., Academic Press Inc., New York, N. Y., 1963, p. 83.

(23) P. F. Spahr, *J. Biol. Chem.*, **239**, 3716 (1964); M. F. Singer and G. Tolbert, *Science*, **145**, 593 (1964).

(24) G. J. Spyrides and F. A. Lipmann, *Proc. Natl. Acad. Sci. U. S.*, **48**, 1977 (1962); S. H. Barondes and M. W. Nirenberg, *Science*, **138**, 810 (1962).

(25) Thus in a series of copolymers of uridylate and guanylate units, as the proportion of the latter nucleotide increased the extent of secondary structure in the polymers increased [guanine-containing homopolymers have strong tendency for the formation of multistranded structures; see, *e.g.*, R. K. Ralph, W. J. Connors, and H. G. Khorana, *J. Am. Chem. Soc.*, **84**, 2265 (1962)]. The progressive increase in secondary structure, in turn, decreased the capacity to stimulate amino acid incorporation [M. F. Singer, O. W. Jones, and M. W. Nirenberg, *Proc. Natl. Acad. Sci. U. S.*, **49**, 392 (1963)]. Similarly, ribopolynucleotides containing alternating units of adenylate and uridylate, which are obtained by the use of the DNA-like polymer containing alternating

those nucleotide combinations which result in Watson–Crick types of hydrogen bonds were excluded. Thus, in repeating dinucleotide structure, syntheses were undertaken of the homologous series of deoxypolynucleotides containing the following dinucleotides in repeating sequence: thymidylyldeoxycytidine (two pyrimidines), deoxyadenylyldeoxyguanosine (two purines), thymidylyldeoxyguanosine (two keto bases), and deoxycytidylyldeoxyadenosine (two amino bases). In the series of deoxypolynucleotides with repeating trinucleotide sequences, the syntheses of the repeating sequences thymidylylthymidylyldeoxycytidine and thymidylylthymidylyldeoxyinosine were undertaken first. Subsequent to these, hexanucleotides containing the complementary triplets deoxyguanylyldeoxyadenylyldeoxyadenosine and deoxycytidylyldeoxyadenylyldeoxyadenosine were synthesized.

Four accompanying papers[26–29] describe the synthetic work on the above series of deoxyribopolynucleotides. So far as the enzymatic studies with the resulting polynucleotides are concerned, all of the major expectations have already been realized. Thus, the DNA polymerase utilizes a mixture of the two decanucleotides, containing the repeating sequences thymidylyldeoxycytidine and deoxyguanylyldeoxyadenosine, as the templates, and, in the presence of the 5′-triphosphates of the four deoxynucleosides brings about the extensive synthesis of a high molecular weight DNA-like polymer containing alternating deoxycytidylate and thymidylate units in one strand and deoxyguanylate and deoxyadenylate units in the complementary strand.[30] Analogously, the synthesis of another two-stranded DNA-like polymer, containing deoxyguanylate and thymidylate units in alternating sequence and deoxycytidylate and deoxyadenylate units in alternating sequence in the two complementary strands, is brought about in response to the two decanucleotides containing the above repeating dinucleotide sequences.[31] High molecular weight ribopolynucleotide of repeating di- and trinucleotide sequences have been prepared by using the short-chain synthetic deoxyribopolynucleotides as templates for DNA-dependent RNA polymerase.[32] Ribopolynucleotides containing repeating dinucleotide sequences have also been prepared by the use of DNA-like polymers mentioned above as templates for the RNA polymerase. Finally, the synthesis of homopeptides and copeptides as directed by ribopolynucleotides containing, respectively, repeating trinucleotide and dinucleotide sequences has also been demonstrated.[33–35]

deoxyadenylate and thymidylate units as template for RNA polymerase, fails to stimulate amino acid incorporation (unpublished work of Professor S. Ochoa and co-workers and of Dr. S. Nishimura in this laboratory).

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

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

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

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

(30) C. Byrd, E. Ohtsuka, M. W. Moon, and H. G. Khorana, *Proc. Natl. Acad. Sci. U. S.*, **53**, 79 (1965).

(31) R. D. Wells, E. Ohtsuka, and H. G. Khorana, unpublished work.

(32) S. Nishimura, T. M. Jacob, and H. G. Khorana, *Proc. Natl. Acad. Sci. U. S.*, **52**, 1494 (1964), and unpublished work of Dr. S. Nishimura.

(33) S. Nishimura, D. S. Jones, R. D. Wells, T. M. Jacob, and H. G. Khorana, *Federation Proc.*, **24**, 409 (1965).

(34) S. Nishimura, D. S. Jones, E. Ohtsuka, H. Hayatsu, T. M. Jacob, and H. G. Khorana, *J. Mol. Biol.*, in press.

(35) S. Nishimura, D. S. Jones, and H. G. Khorana, *ibid.*, in press

Studies on Polynucleotides. XLIII.[1] The Synthesis of Deoxyribopolynucleotides Containing Repeating Dinucleotide Sequences[2,3]

E. Ohtsuka, M. W. Moon, and H. G. Khorana

Contribution from the Institute for Enzyme Research of the University of Wisconsin, Madison, Wisconsin. *Received March 22, 1965*

The synthesis and characterization of four series of homologous deoxyribopolynucleotides, up to the dodecanucleotides, containing the following dinucleotides in repeating sequences is described: thymidylyldeoxycytidylate, thymidylyldeoxyguanylate, deoxyadenylyldeoxyguanylate, deoxycytidylyldeoxyadenylate. The general method of synthesis involved the polymerization of suitably protected dinucleotides by reaction with dicyclohexylcarbodiimide followed by removal of the protecting groups and separation by a combination of anion exchange and paper chromatography. The nature of side products which were encountered is discussed.

The synthesis of deoxyribopolynucleotides containing repeating di- and trinucleotide sequences is of interest in the study of the amino acid code, and a review of the major objectives of synthetic work has been given in the preceding paper.[1] The present paper records the synthesis and characterization of several series of homologous short-chain deoxyribopolynucleotides containing the repeating dinucleotide sequences thymidylyldeoxycytidine, thymidylyldeoxyguanosine, deoxyadenylyldeoxyguanosine, and deoxycytidylyldeoxyadeno-

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

(2) For leading references to previous papers which deal directly with the synthesis of deoxyribopolynucleotides containing specific sequences, see ref. 1.

(3) This work has been supported by grants from the National Science Foundation (Grant No. GB-976), the National Cancer Institute of the National Institutes of Health (Grant No. CA-05178), and the Life Insurance Medical Research Fund (Grant No. G-62-54).