

H. GOBIND KHORANA

Nucleic acid synthesis in the study of the genetic code

Nobel Lecture, December 12, 1968

1. Introduction

Recent progress in the understanding of the genetic code is the result of the efforts of a large number of workers professing a variety of scientific disciplines. Therefore, I feel it to be appropriate that I attempt a brief review of the main steps in the development of the subject before discussing our own contribution which throughout has been very much a group effort. I should also like to recall that a review of the status of the problem of the genetic code up to 1962 was presented by Crick in his Nobel lecture¹.

While it is always difficult, perhaps impossible, to determine or clearly define the starting point in any area of science, the idea that genes make proteins was an important step and this concept was brought into sharp focus by the specific one gene-one enzyme hypothesis of Beadle and Tatum². The field of biochemical genetics was thus born. The next step was taken when it was established that genes are nucleic acids. The transformation experiments of Avery and coworkers³ followed by the bacteriophage experiments of Hershey and Chase⁴ established this for DNA and the work with TMV-RNA a few years later established the same for RNA^{5,6}. By the early 1950's it was, therefore, clear that genes are nucleic acids and that nucleic acids direct protein synthesis, the direct involvement of RNA in this process being suggested by the early work of Caspersson⁷ and of Brachet⁸. It was important at this stage to know more about the chemistry of the nucleic acids and, indeed, the accelerated pace of discovery that soon followed, was largely because of work at the chemical and biochemical level in the field of nucleic acids.

The structural chemistry of the nucleic acids, which developed over a period of some seventy years in many countries, progressed step-by-step from the chemistry of the constituent purines, pyrimidines and the sugar moieties, to work on the nucleosides and then onto the nucleotides. A distinct climax was reached in 1952 with the elucidation of the internucleotidic linkage in nucleic acids by Brown and Todd and their coworkers^a. (It was my good fortune to be associated with Professor, now Lord, Todd's laboratory before the

start of our own work in the nucleotide field). Shortly thereafter, the Watson - Crick structure¹⁰ for DNA was proposed, which focused 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 proteins was born. A few years later, the enzymology of DNA got into its stride with the work of Kornberg and his coworkers¹¹: their discovery and characterization of the enzyme DNA polymerase was a major triumph of modern enzymology and the methods developed distinctly aided the characterization, a few years later, of DNA-dependent RNA polymerase¹²⁻¹⁶. The discovery of this enzyme clarified the manner by which information in DNA is transcribed into an RNA, which we now equate with messenger RNA¹⁷⁻²⁴. The last biochemical landmark to be introduced in the development of a cell-free amino acid incorporating system. Work on this really began with efforts to understand the biosynthesis of the peptide bond. The subject has a long history but critical progress began to be made in the early fifties. One thinks, in particular, of the pioneering work of Zamecnik and Hoagland²², of Lipmann²³, of Berg²⁴ and in regard to the bacterial system that of Watson's laboratory²⁵, of Berg²⁶, and of the important refinement made in 1961 by Matthaei and Nirenberg²⁷.

With the knowledge of the chemical structures of the nucleic acids, the two major tasks which faced the chemists were those of synthesis and sequential analysis. Chemical synthesis of short-chain oligonucleotides began to be a preoccupation in my laboratory. The types of problems that one faced were: (1) activation of the phosphomonoester group of a mononucleotide so as to phosphorylate the hydroxyl group of another nucleoside or nucleotide; (2) design of suitable protecting groups for the various functional groups (primary and secondary hydroxyl groups in the sugar rings, amino groups in the purine and pyrimidine rings, phosphoryl dissociation in the phosphomonoester group); (3) development of methods for the polymerization of mononucleotides and for the separation and characterization of the resulting polynucleotides, and (4) evaluation of approaches to the stepwise synthesis of polynucleotides of specific sequences.

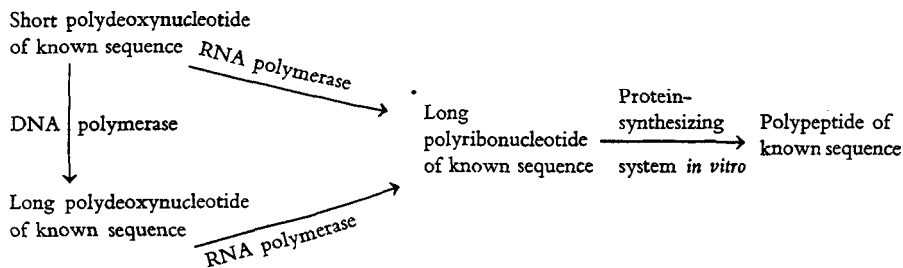
While even at present, organo-chemical methods demand further investigation and refinement, nevertheless, synthesis of short chains of deoxyribo-polynucleotides with predetermined and fully controlled sequences became possible in the early sixties. In addition, unambiguous synthesis of short ribooligonucleotides containing strictly the 3' → 5'-internucleotidic linkages also became feasible. A discussion of the chemical aspects of these problems is out-

side the scope of the present lecture and reviews given elsewhere²⁸⁻³³ should be consulted. The following review will be restricted to that part of the synthetic work which bears on the problem of the genetic code and attention will be focused in the main on the biochemical experiments made possible by the synthetic polynucleotides.

Polynucleotide Synthesis in Relation to the Genetic Code

A few words about the experimental development of the coding problem are now appropriate. In the fifties, possible rules governing the genetic code engaged the attention of many theoreticians, Gamow³⁴ being the first to speculate on the possible relation between DNA and protein structure. However, until 1961 the only experimental approach was that of direct correlation of the sequence of a nucleic acid with that of a protein specified by it. It was hoped to do this either chemically, for example, by working with the coat protein of a virus and its RNA, or by mutagenic techniques. An ingenious application of the <frameshift mutation> idea was, indeed, that of Crick and coworkers, who correctly deduced several of the fundamental properties of the genetic coder. These approaches, however, offered little immediate hope of getting directly at the coding problem.

The discovery which introduced a direct experimental attack on the genetic code was that of Matthaei and Nirenberg²⁷ who observed that polyuridylylate directs the synthesis of polyphenylalanine in the bacterial cell-free amino acid incorporating system. The aim now was to use synthetic polynucleotides of defined composition as messengers in the *in vitro* system. A great deal, in fact, was learned during the years 1961-1963, both in the laboratories of Ochoa³⁵



Scheme 1. Proposed reaction sequence for the preparation of high-molecular-weight RNA messengers and the subsequent *in vitro* synthesis of polypeptides of known amino acid sequences.

and of Nirenberg and their coworkers³⁶, about the overall nucleotide composition of the coding units by using polynucleotides made by the agency of the enzyme, polynucleotide phosphorylase.

*(a) Short chains of deoxyribopolynucleotides containing repeating nucleotide sequences**

The hope in my laboratory was to prepare ribopolynucleotide messengers of completely defined nucleotide sequences. However, chemical methodology permitted at this time the synthesis of oligonucleotides containing but a few ribonucleotide units (see below for trinucleotide syntheses). In the deoxyribonucleotide series, chemical synthesis was a little more advanced and the synthesis of longer, but still short, chains containing ten to fifteen nucleotide units was feasible. Therefore, it was decided to study the RNA-transcribing enzyme with the hope that this enzyme might use short chemically synthesized deoxyribopolynucleotides as templates in the manner that it uses biologically functional DNA.

The initial experiments with the RNA polymerase³⁷ were, in fact, a follow-up of the observation first made by Hurwitz and coworker³⁸ that a mixture of chemically synthesized thymidine oligonucleotides served as a template for the synthesis of ribo-polyadenylate. Our aim was to obtain a ribopolynucleotidic product matching in chain length the deoxypolynucleotide used as the template. However, analysis showed that irrespective of the size of the short deoxypolynucleotide template, the RNA product was always much longer; it contained invariably more than 100 nucleotide units. At first sight, the results appeared to be discouraging in that we were losing control on the exact

*The abbreviations used are as follows. The letters, A, C, G, T and U stand for the nucleosides or the nucleotides of adenine, cytosine, guanine, thymine and uracil respectively; the prefixes d and r represent deoxyribose and ribose series of polynucleotides, respectively. All the polynucleotides containing more than one kind of nucleotide, which are used in the text, in Tables and in Figures, have strictly repeating nucleotide sequences and the repeating unit is shown. For example, poly-rUG and poly-rUAG represent polymers in which the dinucleotide, U-G, and the trinucleotide, U-A-G, sequences repeat. Polythymidylate containing eleven nucleotide units in the chain is abbreviated to dT₁₁ and the hepta-deoxyadenylate is abbreviated to dA₇. tRNA or sRNA stands for transfer RNA; met-tRNA^{met} stands for the non-formylatable species of methionine-specific tRNA which has been charged with the amino acid and fmet-tRNA^{met} stands for the formylatable species of methionine-specific tRNA which has been charged with methionine and the latter residue subsequently formylated.

size of the product, despite our having carefully defined the size of the oligothymidylate template. However, it soon became apparent that this <slipping> or reiterative copying on the part of the enzyme could be a highly useful device to amplify the messages contained in the short chemically-synthesized polynucleotides. In a further study, attention was paid to understand a little better the conditions for the <amplification> to occur³⁹.

Some months later I visited Kornberg's laboratory (this was one of the many pilgrimages that I have made to this great laboratory) and started a few experiments with the DNA polymerase and here, again, very short synthetic oligonucleotides containing alternating A and T units induced the enzyme to bring about extensive synthesis of the previously characterized high molecular weight dAT polymer⁴⁰. These encouraging results led to a generalized scheme (Scheme 1) for the *in vitro* studies of the coding problem. The amplification to produce DNA or RNA products was conceived at this time to be a general behaviour of the polymerases so long as there was a repeating pattern of nucleotide sequences (homopolynucleotides, repeating di- or tri-nucleotides) in the chemically-synthesized deoxypolynucleotide templates. Everything from this point on went remarkably well and the period starting with the spring of 1963 and ending with 1967 was a period, essentially, of uninterrupted success in work devoted to the genetic code.

The decision to synthesize deoxyribopolynucleotides with repeating nucleotide sequences was fortunate for another reason. The cell-free protein-synthesizing system, being a crude bacterial extract, undoubtedly contained powerful nucleases and peptidases. The use of messengers with completely defined but strictly repeating nucleotide sequences could be expected to give unequivocal answers despite (1) the exo- and/or endo-nucleolytic damages to the synthetic messengers and (2) the corresponding activities of the proteolytic enzymes on the polypeptide products.

The actual choice of nucleotide combinations in the deoxyribopolynucleotides to be synthesized was influenced by the then available knowledge that, at least in the cell-free protein-synthesizing system, the messenger RNA appears to be used in the single-stranded configuration. In fact, the only DNA containing more than one type of nucleotide, whose sequence was completely known, was the above-mentioned poly-dAT. Although RNA polymerase nicely produces from it the expected poly-rAU containing the two bases in strictly alternating sequence, this product, because of self-complementarity, has a tight double-stranded structure and elicits no response from the ribosomes in the cell-free system. It was, therefore, clear that those combinations

of nucleotides which would lead to overwhelming base-pairing in the polynucleotides should be avoided³⁰⁻³².

All of the chemical syntheses relevant to the genetic code which were carried out are shown in Table I. First, we made the two sets of polynucleotides shown on the left, which contained repeating dinucleotide sequences: one set contains the hexamer of the dinucleotide with alternating thymidylate and guanylate residues and the hexamer of the dinucleotide with alternating adenylate and cytidylate residues; the second set consists of the hexamer of alternating thymidylate and cytidylate residues and the hexamer of alternating adenylate and guanylate residues⁴¹. This work was then extended to

Table I
Synthetic deoxyribopolynucleotides with repeating nucleotide sequences

Repeating dinucleotide sequences		Repeating trinucleotide sequences				Repeating tetranucleotide sequences							
$\{d[TC]_6\}$	$\{d[AG]_6\}$	$\{d[TTG]_4\}$	$\{d[CAA]_4\}$	$\{d[CCT]_{3-5}\}$	$\{d[GGA]_{3-5}\}$	$\{d[TAC]_{4-6}\}$	$\{d[TAG]_{4-6}\}$	$\{d[CCA]_{3-5}\}$	$\{d[GGT]_{3-5}\}$	$\{d[TTAC]_4\}$	$\{d[GTAA]_2\}$	$\{d[TCTA]_3\}$	$\{d[TAGA]_2\}$
		$\{d[TTG]_{4-6}\}$	$\{d[CAA]_{4-6}\}$	$\{d[CGA]_{3-5}\}$	$\{d[CGT]_{35-}\}$	$\{d[ATC]_{3-5}\}$	$\{d[ATG]_{3-5}\}$						

polynucleotides with repeating trinucleotide sequences. There is a theoretical maximum of ten such sets that can contain more than one nucleotide base: we prepared seven such sets⁴²⁻⁴⁶. Shown also in Table 1 are two sets of polymers with repeating tetranucleotide sequences^{47,48}. Two additional considerations for the selection of the nucleotide sequences in them are: (1) they contain in every fourth place the chain-terminating codons and (2) this class of polymers can be used to prove the direction of reading of the messenger RNA⁴⁹⁻⁵¹. Two general points about all the synthetic polynucleotides shown in Table I may be noted. The first point is that every set comprises two polynucleotides which are complementary in the antiparallel Watson-Crick base-pairing sense. A set of repeating trinucleotide polymers, which was complementary in the *parallel* sense, was found to be unacceptable to the DNA polymerase⁵². The second point is that it *was* necessary to synthesize segments corresponding to both strands of the DNA-polymer eventually desired (see below). DNA polymerase failed to bring about polymerization reactions when given only one of the segments of a set as a template.

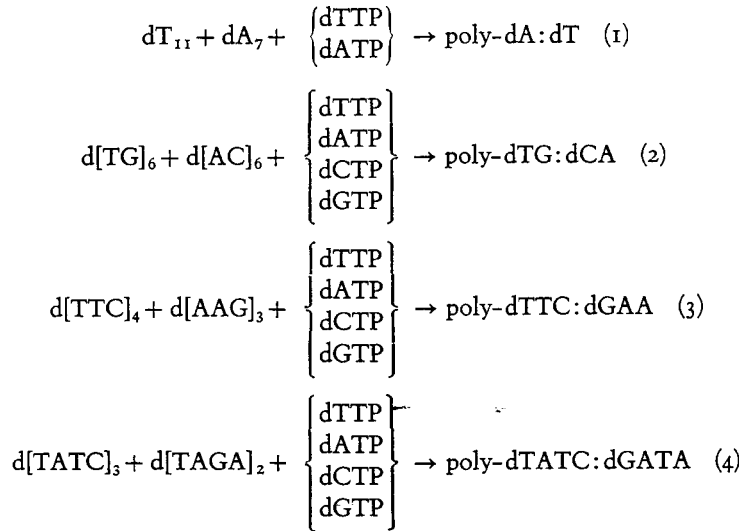
(b) *Double-stranded DNA-like polymers with repeating nucleotide sequences*

In a part of the early work, short-chain deoxyribopolynucleotides with repeating sequences, for example, $(TTC)_3$ and $(TC)_5$, were directly used as templates for the RNA polymerase of *Escherichia coli*. While these experiments were successful^{31,33}, further work soon showed that use of the DNA polymerase as the first <amplification> device was preferable by far. Therefore, a major concern, following the chemical syntheses of the templates, was the study of the DNA polymerase and characterization of the DNA-like products produced by it in response to the short templates.

Scheme 2 lists the four types of reactions which have been elicited from the DNA polymerase, including the use of short homopolynucleotides. As seen, in reaction 1, a mixture of dT_{11} and dA_7 caused the extensive polymerization of dATP and dTTP to give a DNA-like polymer containing polyadenylate and polythymidylate in the two strands. In reaction 2, a mixture of the two short-chain polynucleotides with repeating dinucleotide sequences directed the extensive synthesis of a double-stranded DNA-like polymer containing exactly the sequences present in the short-chain templates^{54,55}. In further work, similar reactions were demonstrated with short-chain templates containing repeating tri- as well as tetra-nucleotide sequences^{52,56}. Characterization of the high molecular weight DNA-like polymers was accomplished by a variety of methods. The techniques used included nearest-neighbor analysis, electron microscopy (in a part of the early work⁵⁵), sedimentation velocity and banding in alkaline cesium chloride density gradients⁵⁷.

Many of the features of the DNA-polymerase catalyzed reactions are truly remarkable. Thus: (1) in all the reactions studied (Scheme 2) the enzyme shows complete fidelity in the reproduction of sequences; (2) the synthesis is extensive, 50-200-fold, and the products are of high molecular weight (300000 to over 1 000000); (3) the enzyme thus amplifies and multiplies the information created by chemical methods; (4) finally, from the standpoint of an organic chemist, the most satisfying aspect is that the DNA polymers thus made can be used repeatedly for further production of the same polymers. It is unnecessary to go back to the time-consuming chemical synthesis for obtaining the templates again. DNA polymerase assures the continuity of these sequences.

Table II catalogues the different kinds of polymers which have so far been prepared and characterized. Thus, we have three classes of polymers: two



Scheme 2. Types of reactions catalyzed by DNA polymerase. All of the DNA-like polymers are written so that the colon separates the two complementary strands. The complementary sequences in the individual strands are written so that antiparallel base-pairing is evident.

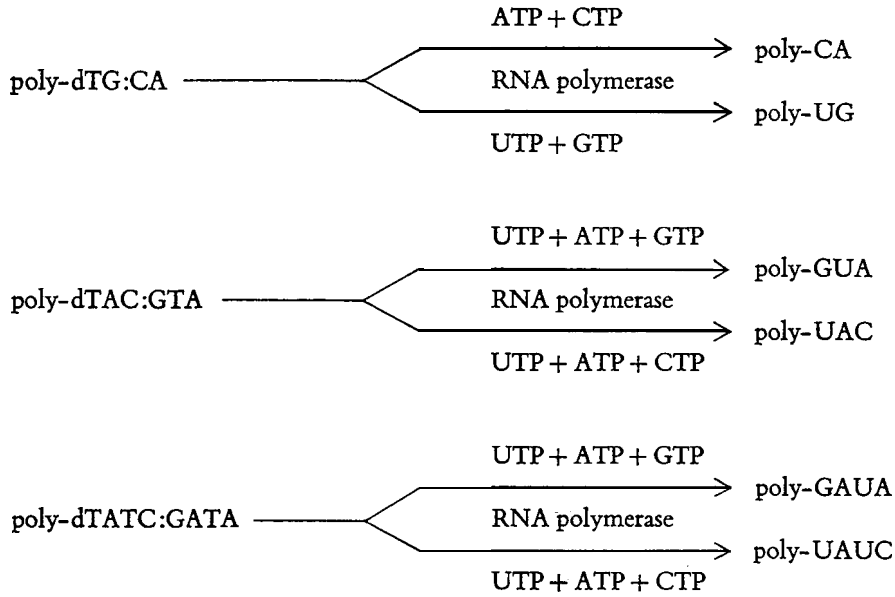
Table II
DNA-like polymers with repeating nucleotide sequences

<i>Repeating dinucleotide sequences</i>	<i>Repeating trinucleotide sequences</i>	<i>Repeating tetranucleotide sequences</i>
Poly-dTC:GA	Poly-dTTC:GGA	Poly-dTTAC:GTAA
Poly-dTG:GA	Poly-dTTG:CAA	Poly-dTATC:GATA
	Poly-dTAC:GTA	
	Poly-dATC:GAT	

double-stranded polymers with repeating dinucleotide sequences, four polymers with repeating trinucleotide sequences and two polymers with repeating tetranucleotide sequences.

(c) *Single-stranded ribo-polynucleotides with repeating nucleotide sequences*

The next step was the transcription of the DNA-like polymers by means of RNA polymerase to form single-stranded ribo-polynucleotides. The prin-



Scheme 3. The preparation of single-stranded ribopolynucleotides from DNA-like polymers containing repeating nucleotide sequences.

ciple used throughout is illustrated in Scheme 3. All of the DNA-like polymers contain two, or a maximum of three, different bases in individual strands. It is therefore possible, by giving the nucleoside triphosphates required for copying only one strand, to restrict the action of RNA polymerase to that strand. This is the case for all of the polymers, examples of which are shown in Scheme 3. Nearest-neighbor frequency analysis of all of the RNA

Table III
Synthetic ribopolynucleotides with repeating nucleotide sequences

<i>Repeating dinucleotide sequences</i>	<i>Repeating trinucleotide sequences</i>	<i>Repeating tetranucleotide sequences</i>
Poly-rUG	Poly-rUAC	Poly-rUAAG
Poly-rAC	Poly-rGUA	Poly-rUAGA
Poly-rUC	Poly-rAUC	Poly-rUCUA
Poly-rAG	Poly-rGAU	Poly-rUUAC
	Poly-rUUG	
	Poly-rCAA	
	Poly-rUUC	
	Poly-rGAA	

products again shows that they contain strictly repeating nucleotide sequences⁵⁸⁻⁶⁰. The total RNA-like polymers prepared so far are listed in Table III.

The work described so far can be summarized as follows. By using a combination of purely chemical methods, which are required to produce new and specified information, and then following through with the two enzymes, DNA polymerase and RNA polymerase, which are beautifully precise copying machines, we have at our disposal a variety of high-molecular-weight ribo-polynucleotides of known sequences. Mistake levels, if they occur at all, are insignificant.

(d) Chemical synthesis of the sixty-four possible ribotrinucleotides

At about the time that the above methods for the synthesis of long ribo-polynucleotides of completely defined nucleotide sequences were developed, the use of ribo-trinucleotides in determining the nucleotide sequences within codons for different amino acids was introduced by Nirenberg and Leder (see below). As mentioned above, chemical methods in the ribonucleotide field developed in this laboratory had previously resulted in general methods for the synthesis of the ribotrinucleotides. In view of the importance of these oligonucleotides in work on the genetic code, all the 64 trinucleotides derivable from the four common ribonucleotides, A, C, U, and G, were unambiguously synthesized and characterized. A separate report⁶¹ should be consulted for the details of the chemical principles used in these syntheses. The use of the trinucleotides in work on the codon assignments for different amino acids is reviewed below.

3. Polypeptide Synthesis in vitro and the Genetic code

(a) Cell-free polypeptide synthesis using polynucleotides with repeating sequences

Polymers with repeating dinucleotide sequences, $(AB)_n$, contain two triplets, ABA and BAB, in alternating sequence. Assuming three-letter, non-overlapping properties of the code, such polymers should direct incorporations of two amino acids in strictly alternating sequence. Repeating trinucleotide polymers, $(ABC)_n$, contain three repeating triplets depending upon the starting point. These are: ABC, BCA, and CAB. Here one would predict that one amino acid should be incorporated at a time to form a homopolypeptide

chain, and a maximum of three such chains should result. Similar considerations for polynucleotides with repeating tetranucleotide sequences, $(ABCD)_n$, show that *in vitro* polypeptide synthesis should give products containing repeating tetrapeptide sequences, irrespective of the starting point in the reading of the messengers. All these predictions have been fully borne out experimentally without a single exception. The results with the three classes of polymers may be reviewed as follows.

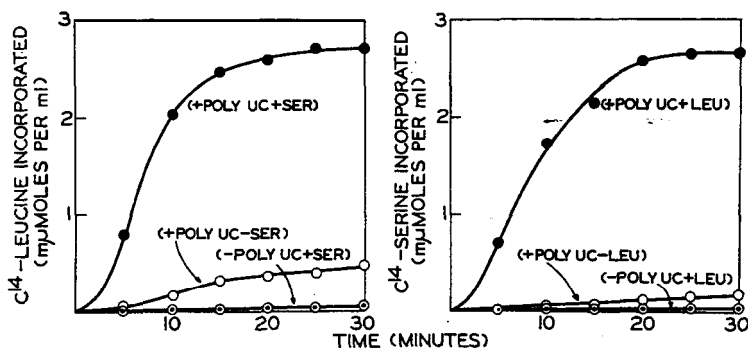


Fig. 1. Characteristics of the incorporation of $[^{14}\text{C}]$ serine and $[^{14}\text{C}]$ leucine into polypeptide in the presence of poly-UC.

Shown in Fig. 1 is an example of the type of results obtained with ribopolynucleotides containing two nucleotides in alternating sequence. Three features of the amino acid incorporations shown in Fig. 1, and which are common for all the messengers of this class, are (1) incorporation of only two amino acids is observed; (2) incorporation of one of these amino acids is dependent on the presence of the second amino acid, and, (3) the incorporations of the two amino acids are equimolar. All these features suggest that in these reactions copolypeptides containing two amino acids in alternating sequence are being produced. This has been demonstrated by extensive analysis of all the four series of polypeptidic products which are listed in Table IV^{58,62}.

Table V shows the results obtained with repeating trinucleotide polymers. Thus, these polymers, have as a rule given three homopolypeptides^{63,64} and it should now be emphasized that this was because in all the work with the cell-free system artificially high Mg^{2+} ions concentration was used and, therefore, polypeptide chains could initiate without a proper signal. Two polymers, poly-rUAG and poly-rAUG, were exceptions in that they stimulated the incorporation of only two amino acids⁶⁴. These polymers contain each a

Table IV

Cell-free copolypeptide syntheses using messengers containing repeating dinucleotide sequences
(System, *Escherichia coli* B)

<i>Polynucleotides</i>	<i>Copolyptides with 2 amino acids in alternating sequence</i>
Poly-UC	(ser-leu) _n
Poly-AG	(arg-glu) _n
Poly-UG	(val-cys) _n
Poly-AC	(thr-his) _n

Table V

Cell-free homopolypeptide syntheses using messengers containing repeating trinucleotide sequences
(System, *Escherichia coli* B)

<i>Polynucleotide</i>	<i>Homopolypeptides of single amino acids</i>
Poly-UUC	phe, ser, leu
Poly-AAG	lys, glu, arg
Poly-UUG	cys, leu, val
Poly-CCA	gln, thr, asn
Poly-GUA	val, ser, (chain-terminator)
Poly-UAC	tyr, thr, leu
Poly-AUC	ileu, ser, his
Poly-GAU	met, asp, (chain-terminator)

chain-terminating triplet; UAG is the well-known amber triplet, and UGA is also now known to be a chain-terminating triplet.

Finally, as seen in Table VI, repeating tetranucleotide polymers, in fact, direct amino acid incorporations such that products containing repeating tetrapeptide sequences are formed except when chain-terminating triplets are present. This has been proved by analysis for the two products shown in Table VI⁵⁰. This analysis of the repeating tetrapeptide sequences proves independently that the direction of reading of the messenger is from the 5'- to the 3'-end. This result is in agreement with that from a number of other laboratories^{65,66}. The last two polynucleotides shown contain in every fourth place the chain-terminating triplets, UAG and UAA; for this reason, they fail to give any continuous peptides, but the formation of tripeptides has been demonstrated⁵¹.

Table VI

Cell-free polypeptide syntheses using messengers containing repeating tetranucleotide sequences
(System, *Escherichia coli* B)

<i>Polynucleotide</i>	<i>Polypeptide</i>
Poly-UAUC	(tyr-leu-ser-ileu) _n
Poly-UUAC	(leu-leu-thr-tyr) _n
Poly-GUAA	di- and tri-peptides
Poly-AUAG	di- and tri-peptides

The results summarized above lead to the following general conclusions: (1) DNA does, in fact, specify the sequence of amino acids in proteins and this information is relayed through an RNA. (This was the first time that a direct sequence correlation between DNA and a protein had been established.) (2) All the results prove the 3-letter and non-overlapping properties of the code. (3) Finally, information on codon assignments can also be derived from these results.

(b) Codon assignments- The structure of the code

For this large question of codon assignments, however, unless one does a large number of polypeptide syntheses, the experiments reviewed above do not individually provide unique assignments to the codons. For example, of the two codons, UCU and CUC, which stand for serine and leucine, it is not possible to say which stands for which amino acid. Now the code has, in fact, been derived by a combination of the results obtained by the use of the binding technique developed by Nirenberg and Leder⁶⁷ and the work with the repeating polymers reviewed above. In Nirenberg's technique, one looks for the stimulation of the binding of different aminoacyl-tRNA's to ribosomes in the presence of specific trinucleotides. An example of its use is shown in Fig. 2, where the question of which of the three sequence isomers, AAG, AGA, GAA, which codes for lysine, is investigated. One measures the binding of [¹⁴C] lysyl-tRNA to ribosomes in the presence of increasing amounts of these trinucleotides. As seen in Fig. 2 the binding is specifically induced by AAG. The other trinucleotide which also promotes a strong binding is AAA and the experiment using this is also included in Fig. 2. The trinucleotides AAG and AAA are, therefore, the codons for lysine. This technique has been used extensively in Nirenberg's laboratory, and my own colleagues have also tested

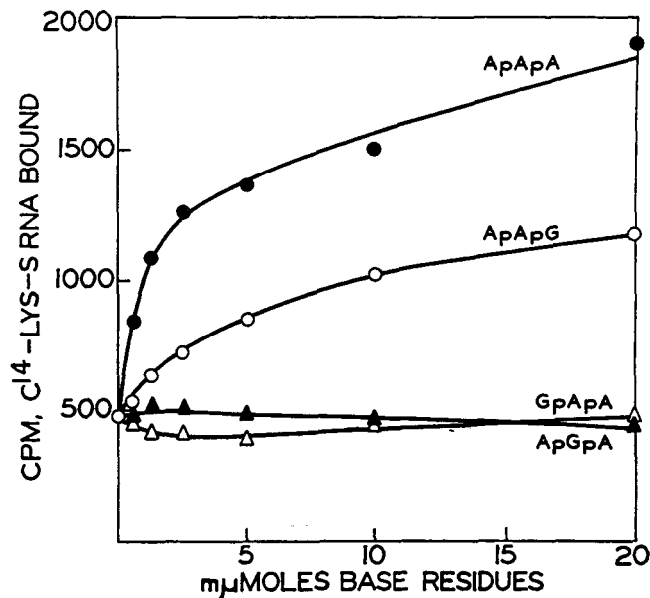


Fig. 2. Stimulation of the binding of [^{14}C] lysyl-tRNA to ribosomes by trinucleotides.

all of the 64 synthetic trinucleotides in this type of analysis. While extremely useful, the technique has not proved to be completely reliable. Often the effects are very small and there are cases where certain trinucleotides stimulate the binding of unexpected tRNA's. Conversely, there are cases where authentic trinucleotide codons do not give any binding. As already mentioned, most of the code actually has been worked out by using this technique in combination with the results from the repeating polymers and often by using evidence from a number of *in vivo* experiments.

The structure of the code which has emerged is shown in Table VII. This is by now a familiar method of presentation⁶⁵. There is a box in the left-hand column for each base as the first letter; within each box in the right-hand column is shown each one of the four bases as the third letter; and in the middle are four columns, one for each base as the second letter. (For use of the Table see the legend). Only a few general observations may be made. (1) The code as shown is for the micro-organism *Escherichia coli* B, but probably will hold essentially for other organisms as well, although detailed and systematic checking in other systems (plants and animals) remains to be carried out. (2) There are entries for all of the sixty-four trinucleotides (there is no absolute nonsense). The code is highly degenerate in a semi-systematic way. Most of the degeneracy pertains to the third letter, where all of the four bases may

Table VII

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 (the 5'-end) and the third letter (the 3'-end) is to the right of the middle (second) base. The use of the Table for derivation of codons for different amino acids is exemplified as follows: codons for the amino acid, PHE, are UUU and UUC; codons for the amino

acid, ALA, are GCU, GCC, GCA and GCG.

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

stand for the same amino acid or where the two purine bases may stand for one amino acid and the two pyrimidines may stand for another amino acid. An exception is the box with the first letter A and the second letter U. Here, AUU, AUC and AUA represent isoleucine while the fourth codon, AUG, stands for methionine. Three amino acids show additional degeneracy in positions other than the third letter: thus, leucine and arginine are degenerate in the first letter while serine is unique in changing its position with regard to, both, the first and the second letters. (3) While the code is now generally accepted to be essentially universal, it should not be inferred that all organisms use the same codons for protein synthesis. What the universality means is that

a trinucleotide codon does not change its meaning from one organism to the next. After all, there is very great divergence in the DNA composition of diverse organisms and they therefore probably use different codons for the same amino acid to varying extents. (4) The codons AUG and GUG, which stand respectively for methionine and valine, are also used as signals for initiation of polypeptide chain synthesis (see also a later section for initiation of protein synthesis). (5) There are three trinucleotides, UAA, UAG and UGA, which cause termination of polypeptide chain growth. It is not clear which ones are used naturally and under what circumstances a particular one is used. More recent work (see the lecture by M. W. Nirenberg, p. 372) indicates that there may be protein factors which have specificity for the different termination codons.

Finally, it should be emphasized that large portions of the code have been derived or confirmed by the prolonged and intensive studies of Yanofsky and coworkers, by the studies of Streisinger and coworkers, by Whitmann and by Tsugita and others (for comprehensive accounts of these studies see ref. 65).

4. *Transfer RNA Structures: The Anticodons and Codon Recognition*

The elucidation of the nucleotide sequence of yeast alanine tRNA by Holley and coworkers⁶⁸ has been followed by similar work on a number of other tRNA's. At present some six yeast tRNA's, four *E. coli* tRNA's, one rat-liver tRNA, and one wheat-germ tRNA, have been sequenced and it is likely that the structures of many more will be known in the near future. Dr. RajBhandary and coworkers⁶⁹ have determined the primary structure of yeast phenylalanine tRNA and this structure is shown in Fig. 3, the usual cloverleaf model being used. In fact, a common feature of all the tRNA's, whose primary structures are known, is that they all can adopt the cloverleaf secondary structure. As discussed in detail by RajBhandary and coworkers⁶⁹ and by others, there is a remarkable overall similarity in regard to many important physical features between the different tRNA's. It is not my intention here to dwell in detail on the broad and exciting subject of tRNA structure and its biological function. The following paragraphs will be confined to those aspects where particularly relevant information has been forthcoming from the work of my own colleagues in Madison.

The first general question that one can ask is: How are the trinucleotide codons recognized by the protein-synthesizing apparatus? The first impor-

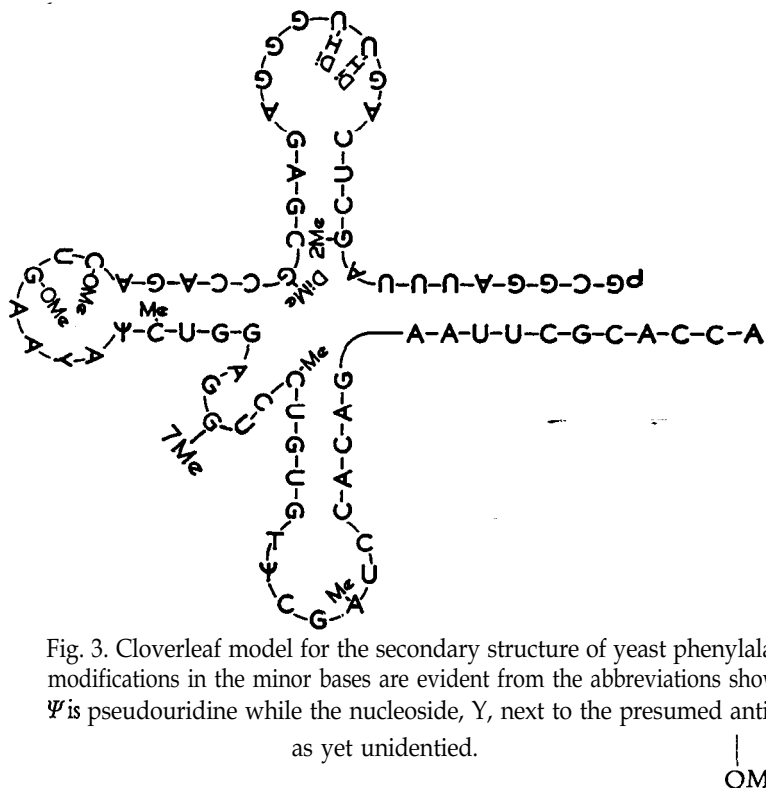


Fig. 3. Cloverleaf model for the secondary structure of yeast phenylalanine tRNA. The modifications in the minor bases are evident from the abbreviations shown against them. Ψ is pseudouridine while the nucleoside, Y, next to the presumed anticodon, G-A-A, is as yet unidentified.

tant advance here was the concept of an adapter molecule, which now is clearly seen to be a tRNA molecule. The experiments of Benzer, Lipmann and their coworkers⁷⁰ brought an elegant confirmation of this hypothesis. The next question is: What is the evidence that the recognition of codons, in fact, involves nucleotide-nucleotide interaction by virtue of base-pairing? If this is so, then one might expect to find in the primary structure of an amino acid specific tRNA three contiguous nucleotide units, <complementary> to the established codons for the particular amino acid. Indeed, the most encouraging common feature of all the tRNA's, whose primary sequence is known, turns out to be that they all contain an identical looped-out region in which such trinucleotide sequences are present. In Table VIII are assembled many of the known anticodons and the established codons for different amino acids. It can be seen that in every case the inferred anticodons show antiparallel Watson-crick base-pairing with the codons. Moreover, the findings (1) that a single nucleotide change in the anticodon of *E. coli* tRNA^{TYR} brings about a change in the coding properties of the tRNA⁷¹ and (2) that an oligonucleotide

Table VIII
Codon-anticodon pairing^a as derived from established codons for certain amino acids and from the primary structures of tRNA's for the corresponding amino acids

Amino acid	ala	tyr	tyr	amber codon	phe	val	ser	met
Codons	→ G C U C A	→ U A U C A	→ U A U C A	→ U A G C A	→ U U U C A	→ G U U C A	→ U C U C A	→ A U G C A
Anticodons ^a	← C G I C A	← A U G C A	← A U G ^b C A	← A U C C A	← A A G-OMe C A	← C A I C A	← A G I C A	← U A C ^c C A
(tRNA source)	← yeast	← yeast	← <i>E. coli</i>	← <i>E. coli</i> (tyr-suppressor tRNA)	← yeast and wheat germ	← yeast	← yeast and rat liver	← <i>E. coli</i>

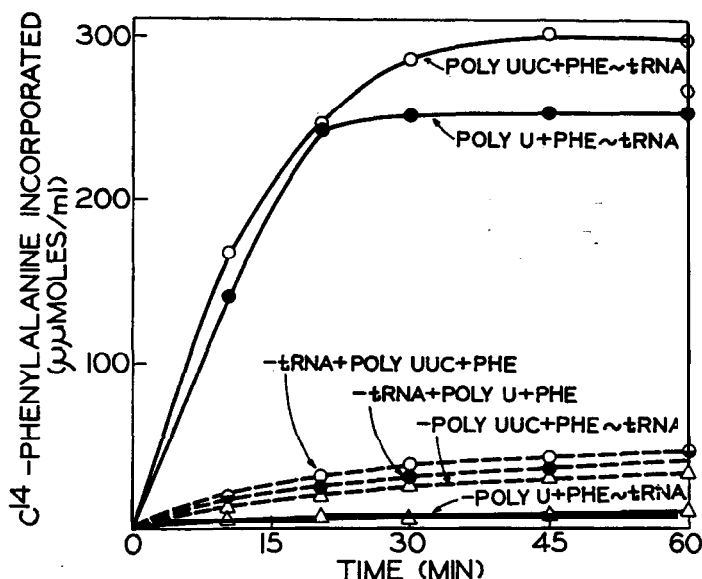
^a Codon-anticodon pairing takes place in the antiparallel direction. Thus, the anticodons, as distinct from the conventional way of writing oligonucleotides, are written in the reverse direction.

^b The structure of this, evidently a derivative of G, is as yet unknown.

^c The structure of this, a derivative of C, is as yet unknown.

fragment of *E. coli* tRNA^{fmet} containing the anticodon sequence binds to ribosomes specifically in response to the codons for formyl-methionine⁷² give us confidence that the concept of an anticodon consisting of three adjacent nucleotides in all tRNA's is correct.

Another important aspect of the biological function of tRNA's deserves comment. Can one tRNA molecule recognize more than one codon? At the present time we believe that this is often the case for the third letter of the codons and this occurs by a certain amount of "wobble" on the tRNA molecule⁷³, thus permitting base-pairs additional to the standard Watson-Crick base-pairs. One such case of multiple recognition of codons by one tRNA has been proven⁷⁴. Thus phenylalanine tRNA, which we have available in a pure state in our laboratory and of which we know the anticodon to be 2'-O-methyl-GAA, can recognize both UUU and UUC which are the established codons for phenylalanine. This has been done by actual polyphenylalanine synthesis using precharged phenylalanyl- tRNA and the two polymers (1) polyuridylylate and (2) poly-UUC which contains a repeating trinucleotide sequence (Fig. 4). There are other possibilities for multiple recognition. For example, it appears that inosine in the first position may recognize U, A and C⁷⁵. Support for this pattern of multiple recognition has also been provided



Polyphenylalanine formation using *Escherichia coli* protein-synthesizing system and purified yeast phe-tRNA. Messengers are poly-U and poly-UUC.

^{75,76}. Possible biological implications of multiple codon recognition by tRNA molecules have been discussed elsewhere⁷⁵.

Transfer RNA's are a unique class of molecules in the biological realm. They clearly have to perform a variety of functions. There is a good deal of evidence to suggest that in addition to a common secondary structure, these molecules possess a tertiary structure^{6,5}. Further, a very plausible and attractive model for the anticodon loop has been put forward⁷⁷. The very recent success in several laboratories⁷⁸⁻⁸³ in obtaining crystals of tRNA's signifies in all probability a new era in the study of tRNA structure and function. The progress here would be very exciting not only for deepening our understanding of the mechanism of protein synthesis but also because of the possibility that a good part of the evolution of the genetic code is synonymous with the evolution of tRNA molecules.

5. Further Aspects of the Code and of Protein Synthesis

(a) Initiation of protein synthesis

As far as the initiation of protein synthesis in *E. coli* is concerned, a surge of activity occurred with the discovery of formylmethionyl-RNA by Marcker and Sanger⁶⁵. It is now generally believed that formylmethionine (fmet) as carried by a particular species of methionine-specific tRNA is the initiation signal in protein synthesis. As mentioned above, in most of the work on the codon assignments using synthetic polynucleotides as messengers, artificially high Mg^{2+} ion concentrations were used. Under those conditions, the need for specific initiation of polypeptide chain synthesis is obviated. However, the requirement for the latter can be introduced by lowering the Mg^{2+} ion concentration to about 4-5 mM (compared with 10-15 mM used in earlier work). It is then found that prompt response in the cell-free system is elicited by only those messengers which contain codons that can recognize fmet-tRNA^{fmet}. Now, the peptide synthesis starts with fmet at the amino terminus. Once again, time does not permit a complete account of the work reported from different laboratories on this subject. My attention will be restricted to those experiments from my own laboratory which (1) permit derivation of codons involved in chain initiation in *E. coli*⁸⁴ and (2) shed a little light on the role of the ribosomal subunits in protein synthesis⁸⁵.

Poly-rAUG, as mentioned above, directs polymethionine synthesis. When this experiment is carried out at 5 mM Mg^{2+} ion concentration, the results

shown in Fig. 5 are obtained. Thus, synthesis proceeds with a lag and poorly, when only $\text{met-tRNA}^{\text{met}}$ is provided. Addition of $\text{fmet-RNA}^{\text{fmet}}$ gives a dramatic acceleration of the rate of polymethionine synthesis. It is therefore concluded that $\text{fmet-tRNA}^{\text{fmet}}$ is required for initiation and $\text{met-tRNA}^{\text{met}}$ is required for chain propagation.

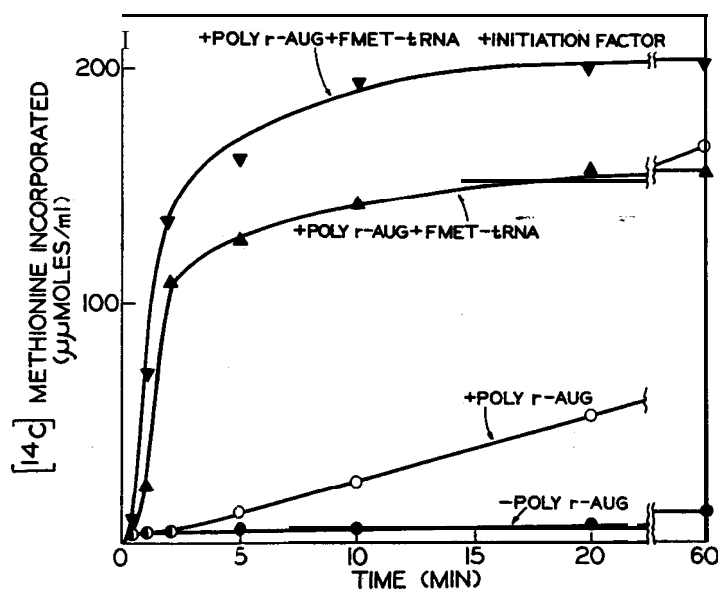


Fig.5. Polymethionine synthesis in the presence of poly-rAUG and $[^{14}\text{C}]\text{met-tRNA}^{\text{met}}$. The effects obtained on supplementing the system with $\text{fmet-tRNA}^{\text{fmet}}$ and initiation factors are shown.

Similarly, at 5 mM Mg^2 ion concentration, poly-rUG-directed synthesis of val-cys copolypeptide (Fig.6) requires the presence of $\text{fmet-tRNA}^{\text{fmet}}$ and there is a striking effect following the addition of protein fractions which have been designated initiation factors^{6,5}. Analysis of the terminal sequence of the polypeptidic product formed showed that fmet was present at the amino end and it was followed by cys and then by val.

From the above results, it is concluded that AUG and GUG are the codons for initiation in *E. coli*⁸⁴. An intriguing point here is degeneracy in the first letter.

That the 70S ribosomal particle from *E. coli* can be split to 30S and 50S subunits was already evident in the late fifties. However, the significance of the two subunits in protein synthesis has remained obscure until recently. Gentle lysis of *E. coli* cells was recently found to yield mainly the 30S and 50S sub-

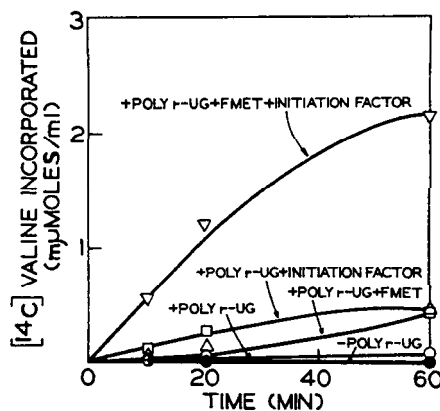


Fig. 6. The synthesis of val-cys copolypeptide as directed by poly-rUG. The synthesis was carried out at 5 mM Mg^{2+} ion concentration. The effects of fmet-tRNA^{fmet} and of the initiation factors are shown.

units and this finding further suggested a role for the 30S-50S couples in protein synthesis⁸⁶. It should be added that prior biochemical investigations had indicated at least two binding sites on the 70S ribosomes. More recently, Nomura and coworkers⁸⁷ showed that in the presence of the viral f2-RNA, fmet-tRNA^{fmet} showed specific binding to the 30S particles, whereas the noninitiator tRNA's showed no binding to the 30S particles.

A further study of the above-described polymethionine synthesis as directed by poly-rAUG gave results⁸⁵ which can be summarized as follows. (1) The 30S particles bind fmet-tRNA^{fmet} in the presence of poly-rAUG at 5 mM Mg^{2+} ion concentration. (2) The noninitiator rRNA, met-tRNA^{met} is bound only after the addition of 50S particles to the 30S particles. (3) It was possible to demonstrate the synthesis of the dipeptide formylmethionyl-methionine (fmet-met) by stepwise formation of the appropriate complex containing all the components. Thus fmet-tRNA^{fmet} was bound to the 30S particles in the presence of poly-rAUG and the <initiation factors>. The complex was isolated by centrifugation, supplemented first with 50S particles and then with met-tRNA^{met}. The resulting complex, containing now 30S particles, poly-rAUG, fmet-tRNA^{fmet}, 50S particles and met-tRNA^{met}, was again isolated by centrifugation. This complex when supplemented with the S-100 supernatant fraction gave the dipeptide fmet-met. Therefore, it is clear that both fmet-tRNA^{fmet} and met-tRNA^{met} were being bound simultaneously to 30S + 50S ribosomal particles. The results provide direct evidence for the presence of two tRNA binding sites on 70S ribosomes. Furthermore, the

picture of the role of the ribosomal subunits which emerges from this work is that the primary event in the initiation of protein synthesis is the binding of the initiator tRNA to 30S ribosome + messenger RNA complex. The resulting initiation complex is then joined by the 50S particles and is now able to accept another aminoacyl-tRNA so that a peptide bond may be formed. Nomura and coworkers⁸⁷ arrived at the same conclusion from their work and the above results support their conclusions. Several other laboratories have subsequently obtained similar results.

As mentioned above, certain protein factors that can be released from the ribosomes are required for the initiation of protein synthesis. These factors, the general chemistry of the ribosomal proteins and the ribosomal subunits themselves are all areas which are currently the subject of investigation in many laboratories. Very recently, striking progress has been made in Nomura's laboratory on the reconstitution of the 30S subunit. These and related studies are rapidly opening up new approaches to a deeper understanding of the mechanism of protein synthesis.

(b) *Missense suppression: tRNA involvement*

Another application of the ribopolynucleotide messengers with repeating nucleotide sequences was in the study of the mechanism of genetic suppression (missense to sense). From the work of Yanofsky and his coworkers it is known that many mutants of *E. coli* can only make a defective protein A of tryptophan synthetase. In one case, mutant A-78, one glycine residue in the A protein is replaced by cysteine. A suppressed mutant (A-78-Su-78) restores, to a small extent, the original glycine in place of cysteine. Using the cell-free protein-synthesizing system from *E. coli* B, it was shown⁸⁸ that this system when supplemented with the tRNA from the strain A-78-Su-78, incorporates [¹⁴C]glycine in the presence of valine under the direction of poly-r-U G (Fig. 7). As reviewed above, the latter polymer normally directs the synthesis of valine-cysteine copolypeptide. Valine-glycine copolypeptide formed *specifically* in the presence of tRNA from A-78-Su-78 strain was thoroughly characterized⁸⁸. Similarly, Carbon, Berg and Yanofsky⁸⁹ showed that another missense suppressor of glycine to arginine mutation in protein A also acted at the level of tRNA. Previously, tRNA had been shown to be responsible for suppression of an *amber* codon in a bacteriophage RNA^{90,91}. As already described above, in one case amber suppression has now been shown to be due to a single nucleotide change in the anticodon of a tRNA⁷¹.

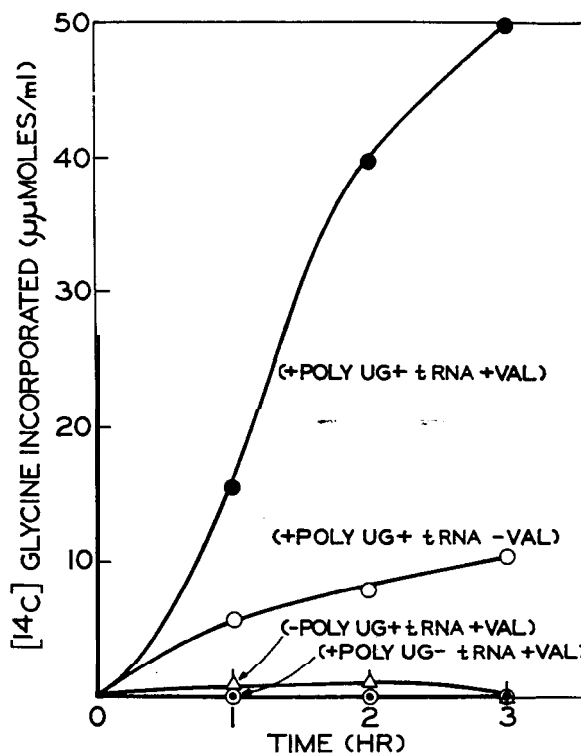


Fig. 7. The incorporation of [^{14}C]glycine into valine-glycine copolypeptide as stimulated by poly-rUG in the presence of tRNA from A-78-Su-78 strain of *Escherichia coli*.

(c) Translation of single-stranded DNA-like polymers

Recently the striking finding was reported⁹² that in the presence of aminoglycoside antibiotics such as neomycin B, denatured DNA stimulated the incorporation of amino acids in the bacterial cell-free protein-synthesizing system. A further study of these observations using single-stranded deoxyribopolynucleotides with defined nucleotide sequences, poly-dTG, poly-dAC and poly-dT, gave most encouraging results⁹³. Thus, the response from the DNA-like polymers was excellent and, surprisingly, the mistakes were very rare and small. For example, poly-dCA directed the synthesis of thr-his copolypeptide and no other amino acid was incorporated. In addition to providing a further opportunity for the study of ribosome function, these results may have important practical applications in future work. As pointed out elsewhere⁹⁴, it is not inconceivable that the laboratory synthesis of specific

proteins will be carried out using nucleic acid templates. For this purpose, protected trinucleotides representing different codons will be made in quantity and on a commercial basis, and these will be used in the synthesis of nucleic acid templates for proteins, the approach offering flexibility and selectivity in amino acid substitutions at the template level.

6. Conclusion

While clarity in some of the detailed aspects of the genetic code is still lacking, it has been a most satisfying experience in the lives of many of us, who have worked on the problems, to see complete agreement reached in regard to its general structure. Evidence coming from a variety of techniques, genetic and biochemical, from *in vivo* and *in vitro* experiments, has furnished the codon assignments reviewed above. It is unlikely that any of the assignments would be revised. However, much remains to be done at chemical and biochemical level to obtain an adequate understanding of the very elaborate protein-synthesizing system. Nevertheless, the problem of the genetic code at least in the restricted one-dimensional sense (the linear correlation of the nucleotide sequence of polynucleotides with that of the amino acid sequence of polypeptides) would appear to have been solved. It may be hoped that this knowledge would serve as a basis for further work in molecular and developmental biology.

Acknowledgements

I wish to emphasize again that the work which has formed the content of this lecture has been so much a collaborative effort. I am deeply indebted to a very large number of devoted colleagues, chemists and biochemists, with whom I had the good fortune to be happily associated.

Work and progress in science becomes more and more interdependent: this certainly has been true in work on the genetic code. Many of the great scientists, who influenced directly or indirectly the work herein reviewed, have been mentioned in the text. I wish to make a personal acknowledgement to one more scientist. Fortunately, I was accepted by Professor V. Prelog of the Eidgenössische Technische Hochschule, Zurich, as a postdoctoral student.

The association with this great scientist and human being influenced immeasurably my thought and philosophy towards science, work and effort.

The work had its start at the British Columbia Research Council, Vancouver, Canada, where it was made possible by the encouragement of Dr. Gordon M. Shrum (now Chancellor of Simon Fraser University, B. C.) and with the financial support from the National Research Council of Canada. In more recent years, generous support has been received from the National Cancer Institute of the National Institutes of Health, U.S. Public Health Service, the National Science Foundation, Washington, the Life Insurance Medical Research Fund and the Graduate School of the University of Wisconsin.

1. F. H. C. Crick, in *On the genetic code, in Nobel Lectures, Physiology or Medicine 1942-1962*, Elsevier, Amsterdam, 1964, p.811.
2. G. W. Beadle and E. L. Tatum, *Proc. Natl. Acad. Sci. (U.S.)*, 27 (1941) 499.
3. O. T. Avery, C. M. Macleod and M. McCarty, *J. Exptl. Med.*, 79 (1944) 137.
4. A. D. Hershey and M. Chase, *J. Gen. Physiol.*, 36 (1952) 39.
5. A. Gierer and G. Schramm, *Nature*, 177 (1956) 702.
6. H. Fraenkel-Conrat, B. Singer and R. C. Williams, *Biochim. Biophys. Acta*, 25 (1957) 87.
7. T. Caspersson, *Naturwissenschaften*, 28 (1941) 33.
8. J. Brachet, *Arch. Biol. (Liège)*, 53 (1942) 207.
9. D. M. Brown and A. R. Todd, in E. Chargaff and J. N. Davidson (Eds.), *Nucleic Acids*, Vol. 1, Academic Press, New York, 1955, p.409.
10. J. D. Watson and F. H. C. Crick, *Nature*, 171 (1953) 737.
11. I. R. Lehman, M. J. Bessman, E. S. Simms and A. Kornberg, *J. Biol. Chem.*, 233 (1958) 163.
12. S. B. Weiss, *Proc. Natl. Acad. Sci. (U.S.)*, 46 (1960) 1020.
13. A. Stevens, *J. Biol. Chem.*, 236 (1961) PC43.
14. J. Hurwitz, J. J. Furth, M. Anders and A. Evans, *J. Biol. Chem.*, 237 (1962) 3752.
15. M. Chamberlin and P. Berg, *Proc. Natl. Acad. Sci. (U.S.)*, 48 (1962) 81.
16. D. P. Burma, H. Kroger, S. Ochoa, R. C. Wamer and J. D. Weill, *Proc. Natl. Acad. Sci. (U.S.)*, 47 (1961) 749.
17. A. D. Hershey, J. Dixon and M. Chase, *J. Gen. Physiol.*, 36 (1953) 777.
18. E. Volkin and L. Astrachan, *Virology*, 2 (1956) 149.
19. F. Jacob and J. Monod, *J. Mol. Biol.*, 3 (1961) 318.
20. S. Brenner, F. Jacob and M. Meselson, *Nature*, 190 (1961) 576.
21. F. Gros, W. Gilbert, H. Hiatt, C. G. Kurland, R. W. Risebrough and J. D. Watson, *Nature*, 190 (1961) 581.
22. P. C. Zamecnik, *Harvey Lectures*, Ser. 54 (1958-1959) 256.
23. F. Lipmann, W. C. Hulsmann, G. Hartmann, Hans G. Boman and George Acs, *J. Cell. Comp. Physiol.*, 54, Suppl. 1 (1959) 75.

24. P.Berg, *Ann.Rev.Biochem.*, 30 (1961) 293.
25. J. D. Watson, in *The involvement of RNA in the synthesis of proteins*, in Nobel Lectures, *Physiology or Medicine 1942-1962*, Elsevier, Amsterdam, 1964, p. 785.
26. W.B. Wood and P.Berg, *Proc. Natl. Acad. Sci. (U.S.)*, 48 (1962) 94.
27. J. H. Matthaei and M. W. Nirenberg, *Proc. Natl. Acad. Sci. (U.S.)*, 47 (1961) 1580.
28. H. G. Khorana, *Federation Proc.*, 19 (1960) 931.
29. H. G. Khorana, *Some Recent Developments in the Chemistry of Phosphate Esters of Biological Interest*, Wiley, New York, 1961.
30. H.G.Khorana, T.M.Jacob, M.W.Moon, S.A.Narang and E.Ohtsuka, *J.Am. Chem. Soc.*, 87 (1965) 2954.
31. H. G.Khorana, *Federation Proc.*, 24 (1965) 1473.
32. H.G.Khorana, in *Proc. 7th. Intern.Congr.Biochem., Tokyo, August 19-25,1967*, 1968, p.17
33. H. G. Khorana, on *Proc. Plenary Lectures of IUPAC Symp. Natural Products, London, 1968*.
34. G.Gamow, *Nature*, 173 (1954) 318.
35. S. Ochoa, *Federation Proc.*, 22 (1963) 62.
36. M.W.Nirenberg, J.H.Matthaei, O.W.Jones, R.G.Martin and S.H.Barondes, *Federation Proc.*, 22 (1963) 55.
37. A.Falascchi, J.Adler and H.G.Khorana, *J.Biol.Chem.*, 238 (1963) 3080.
38. J.Hurwitz, J.J.Furth, M.Anders and A.Evans, *J.Biol.Chem.*, 237 (1962) 3752.
39. B.D.Mehrotra and H.G.Khorana, *J.Biol.Chem.*, 240 (1965) 1750.
40. A. Kornberg, L.L.Bertsch, J.F. Jackson and H.G.Khorana, *Proc. Natl. Acad. Sci. (U.S)*, 51 (1964) 315.
41. E. Ohtsuka, M. W. Moon and H. G. Khorana, *J. Am. Chem. Soc.*, 87 (1965) 2956.
42. T. M. Jacob and H. G. Khorana, *J. Am. Chem. Soc.*, 87 (1965) 2971.
43. S. A.Nrang and H. G.Khorana, *J.Am.Chem. Soc.*, 87 (1965) 2981.
44. S. A.Narang, T.M. Jacob and H.G.Khorana, *J.Am.Chem.Soc.*, 87 (1965) 2988.
45. S. A. Narang, T.M. Jacob and H.G.Khorana, *J.Am.Chem. Soc.*, 89 (1967) 2158.
46. S.A. Narang, T. M. Jacob and H. G. Khorana, *J.Am.Chem. Soc.*, 89 (1967) 2167.
47. H.Kössel, H.Büchi and H.G.Khorana, *J. Am.Chem.,Soc.*, 89 (1967) 2185.
48. E. Ohtsuka and H. G.Khorana, *J.Am.Chem.Soc.*, 89 (1967) 2195.
49. H.G.Khorana, H.Büchi, T.M. Jacob, H.Kössel, S.A.Narang and E.Ohtsuka, *J. Am.Chem.Soc.*, 89 (1967) 2154.
50. H. Kössel, A. R. Morgan and H. G. Khorana, *J.Mol.Biol.*, 26 (1967) 449.
51. H.Kössel, *Biochim.Biophys.Acta*, 157 (1968) 91.
52. Robert D. Wells, T. M. Jacob, S. A.Narang and H. G.Khorana, *J.Mol.Biol.*, 27 (1967) 237.
53. S.Nishimura, T. M. Jacob and H.G.Khorana, *Proc.Natl.Acad.Sci. (U.S.)*, 52 (1964) 1494.
54. C. Byrd, E. Ohtsuka, M. W. Moon and H.G.Khorana, *Proc. Natl. Acad. Sci. (U.S.)*, 53 (1965) 79.
55. R.D. Wells, E.Ohtsuka and H.G.Khorana, *J.Mol.Biol.*, 14 (1965) 221.
56. R.D. Wells, H.Büchi, H. Kössel, E. Ohtsuka and H. G.Khorana, *J.Mol.Biol.*, 27 (1967) 265.

57. R.D. Wells and J.E. Blair, *J.Mol.Biol.*, 27 (1967) 273.
58. S.Nishimura, D. S. Jones and H. G. Khorana, *J.Mol.Biol.*, 13 (1965) 302.
59. H. G. Khorana, in *Genetic Elements*, Federation of European Societies for Biological Chemists, April 1966, p. 209.
60. Unpublished work of A.R. Morgan.
61. R. Lohrmann, D. Söll, H. Hayatsu, E. Ohtsuka and H. G.Khorana, *J.Am. Chem. Soc.*, 88 (1966) 819.
62. D. S. Jones, S.Nishimura and H. G.Khorana, *J.Mol.Biol.*, 16 (1966) 454.
63. S. Nishimura, D. S. Jones, E. Ohtsuka, H.Hayatsu, T.M. Jacob and H. G.Khorana, *J.Mol.Biol.*, 13(1965)283.
64. A.R. Morgan, R.D. Wells and H.G. Khorana, *Proc. Natl. Acad. Sci. (U.S.)*, 56 (1966) 1899.
65. For references, see *The Genetic Code, Cold Spring Harbor Symp.Quant.Biol.*, 31 (1966).
66. H. Lamfrom, C. S.McLaughlin and A. Sarabhai, *J.Mol.Biol.*, 22 (1966) 359.
67. M. W. Nirenberg and P. Leder, *Science*, 145 (1964) 1399.
68. R.W.Holley, J.Apgar, G.A.Everett, J.T.Madison, M.Marquisee, S.H.Merrill, J.R.Penswick and A. Zamir, *Science*, 147 (1965) 1462.
69. U.L.RajBhandary, S.H. Chang, A. Stuart, R.D.Faulkner, R.M.Hoskinson and H. G.Khorana, *Proc. Natl. Acad. Sci. (U.S.)*, 57 (1967) 751.
70. F. Chapeville, F.Lipmann, G.von Ehrenstein, B. Weisblum, W. J.Ray Sr. and S. Benzer, *Proc. Natl. Acad. Sci. (U.S.)*, 48 (1962) 1086.
71. H. M. Goodman, J.Abelson, A.Landy, S.Brenner and J.D.Smith, *Nature*, 217 (1968) 1019.
72. B.F. C. Clark, S. K. Dube and K.A. Marcker, *Nature*, 219 (1968) 484.
73. F.H. C. Crick, *J. Mol.Biol.*, 19 (1966) 548.
74. D. Söll and U. L.RajBhandary, *J. Mol.Biol.*, 29 (1967) 97.
75. D. Söll, J. Cherayil, D. S. Jones, R.D.Faulkner, A.Hampel, R.M.Bock and H. G. Khorana, p. 51 in ref. 65.
76. D. Söll, J.D.Cherayil and R.M.Bock, *J. Mol.Biol.*, 29 (1967) 97.
77. W.Fuller and A.Hodgson, *Nature*, 215 (1967) 817.
78. B.F.C. Clark, B.P.Doctor, K. C.Holmes, A.Klug, K. A.Marcker, S. J.Morris and H.H.Paradies, *Nature*, 219 (1968) 1222.
79. S.Kim and A.Rich, *Science*, 162 (1968) 1381.
80. A.Hampel, M.Labanauskas, P.G.Connors, L.Kirkegaard, U.L.RajBhandary, P. B. Sigler and R.M.Bock, *Science*, 162 (1968) 1384.
81. F. Cramer, F.v.d.Haar, W. Saenger and E. Schlimme, *Angew. Chem.*, 80 (1968) 969.
82. H.H.Paradies, *FEBS Letters*, 2 (1968) 112.
83. J.R.Fresco, R.D.Blake and R.Langridge, *Nature*, 220 (1968) 1285.
84. H.P. Ghosh, D. Söll and H.G. Khorana, *J. Mol. Biol.*, 25 (1967) 275.
85. H.P.Ghosh and H. G. Khorana, *Proc. Natl. Acad. Sci. (U.S.)*, 58 (1967) 2455.
86. G.Mangiarotti and D. Schlessinger, *J.Mol.Biol.*, 20 (1966) 123; 29 (1967) 395; D. Schlessinger, G. Mangiarotti and D. Apirion, *Proc. Natl. Acad. Sci. (U.S.)*, 58 (1967) 1782.

87. M. Nomura and C. V. Lowry, *Proc. Natl. Acad. Sci. (U.S.)*, 58 (1967) 946; M. Nomura, C.V. Lowry and C. Guthrie, *ibid.*, 58 (1967) 1487.
88. N.K.Gupta and H.G.Khorana, *Proc. Natl. Acad. Sci. (U.S.)*, 56 (1966) 772.
89. J. Carbon, P. Berg and C. Yanofsky, *Proc. Natl. Acad. Sci. (U.S.)*, 56 (1966) 764.
90. M.R. Capecchi and G.N. Gussin, *Science*, 149 (1965) 417.
91. D. L. Engelhardt, R. Webster, R. C. Wilhelm and N. D. Zinder, *Proc. Natl. Acad. Sci. (U.S.)*, 54 (1965) 1791.
92. B. J. McCarthy and J.J. Holland, *Proc. Natl. Acad. Sci. (U.S.)*, 54 (1965) 880.
93. A. R. Morgan, R.D. Weds and H. G. Khorana, *J. Mol. Biol.*, 26 (1967) 477.
94. A. Kumar and H. G. Khorana, *J. Am. Chem. Soc.*, in the press.