M ARSHALL N IRENBERG

The genetic code

Nobel Lecture, December 12, 1968

Genetic memory resides in specific molecules of nucleic acid. The information is encoded in the form of a linear sequence of bases of 4 varieties that corresponds to sequences of 20 varieties of amino acids in protein. The translation from nucleic acid to protein proceeds in a sequential fashion according to a systematic code with relatively simple rules. Each unit of nucleic acid defines-the species of molecule to be selected, its position relative to the previous molecule selected, and the time of the event relative to the previousevent. The nucleic acid therefore functions both as a template for other molecules and as a biological clock.

The information is encoded and decoded in the form of a one-dimensional string. The polypeptide translation product then folds upon itself in a specific manner predetermined by the amino acid sequence, forming a complex, three-dimensional protein.

The Concept of a Gene-Protein Code

The advances in biochemical genetics are due to the efforts of investigators from virtually every field of science. Among the milestones are the identification of DNA as the genetic material by Avery, MacLeod, and McCarty[1], the "one gene-one enzyme" concept of Beadle and Tatum[2], and the pioneering experiments of Brachet[3] and Caspersson[4] on the relation of RNA to protein synthesis. In addition, the puzzle of protein synthesis was unraveled, bit by bit, and *in vitro* systems for protein and nucleic acid synthesis were developed.

The concept of a simple code relating the base sequence of nucleic acid to the amino acid sequence of protein originated on three independent occasions during the early 1950's. Intuition was spectacular when considered in the context of the information then available. Caldwell and Hinshelwoods suggested that RNA is composed of five kinds of units, the four bases and ribose phosphate, and that two adjacent units in RNA correspond to one amino acid

in protein Dounce[6] proposed that three adjacent bases in RNA correspond to one amino acid in protein. In addition, the concepts of polarity of translation and activation of amino acids were formulated in considerable detail. Dounce's conviction that templates are required for the synthesis of protein originated during his Ph.D. oral examination when he was asked by James Sumner to consider the problem of how proteins synthesize other proteins.

Concurrently, George Gamow[7] suggested that a double-strand of DNA contains binding sites for amino acids, each site defined by one base-pair and adjacent non-complementary bases on opposite strands of DNA. Gamow conceived the idea upon reading the article by Watson and Crick on the pairing of bases in DNA[8]. Other speculations concerning the nature of the code were advanced by many investigators during the latter part of the 1950's (*cf.* the recent review of Woese[9]).

Although the concept that RNA is a template for protein was well established, direct biochemical evidence was lacking. However, Hershey's[10] finding that a fraction of RNA is rapidly synthesized and then degraded in *E. coli* infected with T2 bacteriophage, and the demonstration by Volkin and Astrachan[11] that the composition of this RNA fraction resembles phage DNA rather than *E. coli* DNA were exciting, because the data suggested that the unstable RNA fraction might function as templates for the synthesis of phage protein.

I plunged into the problems of protein synthesis after I had obtained postdoctoral training. My graduate studies were in biochemistry under the guidance of James Hogg; I obtained postdoctoral training with Dewitt Stetten and so with William Jakoby at the National Institutes of Health. Then I joined Gordon Tompkins' department and began to study the steps that relate DNA, RNA, and protein. The training in enzymology and the stimulating enviroment greatly influenced the future course of my work.

Extensive studies on the mechanism of protein synthesis had yielded much information and it seemed likely then that it would be possible within the coming decade to obtain the synthesis of an enzyme in cell extracts. Since a system of this kind would provide many opportunities to study questions pertaining to the flow of information from nucleic acid to protein, I decided to work on the cell-free synthesis of penicillinase. Pollock and his colleagues[12,13] had obtained much information on the regulation of penicillinase synthesis *in vivo,* and had shown that the molecular weight of the enzyme is relatively low, and that the enzyme lacks cysteine. It seemed likely that one might selectively inhibit the synthesis of proteins that require cysteine and at the same

time stimulate penicillinase synthesis *in vitro* by the addition of nucleic acid templates to cell extracts.

During the next 2 years I studied the properties of the system, particularly the effect of reaction conditions, nucleic acids and other factors upon the rate of cell-free protein synthesis. During this period results of great interest dealing with protein synthesis in *E. coli* extracts were reported by Lamborg and Zamecnik [14]; Tissieres, Schlessinger and Gros [15], and others. Tissieres, Schlessinger and Gros [15]; Kameyama and Novelli [16]; and Nisman and Fukuhara [17] reported that DNAase inhibited *in vitro* amino acid incorporation into protein. I had also observed this phenomenon and was greatly interested in it because the results strongly suggested that the cell-free synthesis of protein was dependent, ultimately, upon DNA templates.

Heinrich Matthaei then joined me in these studies. We soon showed that RNA prepared from ribosomes stimulates amino acid incorporation into protein [18]. However, amino acids were incorporated into protein rapidly without added RNA [19], so RNA-dependent protein synthesis was difficult to detect. This problem was solved, as shown in Fig. 1, by incubating *E.coli* extracts with the components required for protein synthesis and DNAase in order to reduce the level of endogenous RNA templates. After a brief incubation period, the synthesis of protein stops and further protein synthesis is then dependent upon the addition of template RNA. Transfer RNA does not replace template RNA.

A rapid assay was devised based on the filtration of [14C]protein precipitates that reduced the time required for each experiment about four-fold. Preparations of RNA from many sources were obtained to determine the specificity and activity of each RNA preparation as templates for protein synthesis. RNA from yeast, ribosomes, and from tobacco mosaic virus were found to be highly active in stimulating the incorporation into protein of every species of amino acid tested. In contrast, poly-U stimulated phenalanine incorporation into protein rather specifically, and the product was shown to be polyphenylalanine. Single-stranded poly-U was an active template for phenylalanine incorporation, but double- or triple-stranded poly-U · poly-A helices did not serve as templates for protein synthesis [18].

These results showed that RNA is a template for protein, that residues of U in poly-U correspond to phenylalanine in protein, and that the translation of mRNA is affected by both the primary and the secondary structure of the RNA.
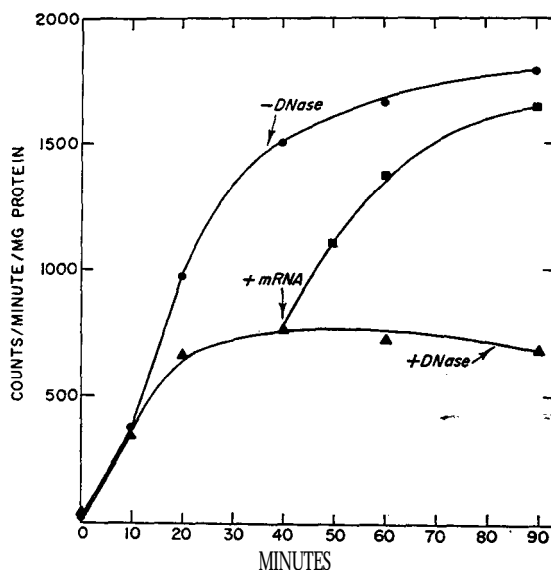
Fig.1. The effect of DNAase and mRNA upon the incorporation of [^{14}C]valine into protein in *E. coli* extracts. The symbols represent the following: ●, no addition; ▲, 10 µg DNAase added per ml of reaction; ■, 10 µg DNAase and 0.5 mg of an mRNA fraction added per ml of reaction.

In 1961, the role of tRNA was still controversial. Most investigators assumed that tRNA participated in the synthesis of protein, but direct proof that tRNA is required for this process was lacking. Lipmann and Nathans generously gave us a purified preparation of transfer enzymes and we found that Phe-tRNA is an *obligatory* intermediate in polyphenylalanine synthesis and that transfer enzymes and GTP are also required for the synthesis of this polypeptide [20].

Base Composition of Codons

The genetic code was deciphered in two experimental phases over a period of approximately six years. During the first phase, the base composition of codons and the general nature of the code were explored by directing cell-free protein synthesis with randomly-ordered RNA templates containing different combinations of bases. Such polymers were synthesized with the aid of polynucleotide phosphorylase that had been discovered by Grunberg-Manago, Ortiz and Ochoa[21].

A summary of data obtained by Ochoa and associates[23] and by ourselves[24] is shown in Table I. Only polynucleotides containing the minimum species of bases required to stimulate an amino acid into protein are shown. Poly- U, poly-C, and poly-A stimulate the incorporation into protein of phenylalanine, proline, and lysine, respectively. No template activity was detected with poly-G. In later studies Maxine Singer, Bill Jones, and I showed that poly-(U,G) preparations rich in G contain a high degree of secondary structure in solution and do not serve as templates for protein synthesis[22].

Poly-(U,C), poly-(C,G), and poly-(A,G) are templates for 2 additional amino acids per polynucleotide, whereas poly-(U,A), poly-(U,G), and poly-(CA) are templates for 4 additional amino acids per polynucleotides. Each polynucleotide composed of 3 species of bases is a template for 10 or more amino acids.

Table 1

Minimum species of bases required for mRNA codons

The specificity of randomly ordered polynudeotide templates in stimulating amino acid incorporation into protein into *E. coli* extracts is shown. Only the minimum species of bases necessary for template activity are shown, so many amino acids responding to polymers composed of two or more kinds of bases are omitted.

Polynucleotides	*Amino acids*			
U	PHE			
C	PRO			
A	LYS			
G	—			
UC	LEU	SER		
UA	LEU	TYR	ILE	ASN
UG	LEU	VAL	CYS	TRP
CA	HIS	THR	GLN	ASN
CG	ARG	ALA		
AG	ARG	GLU		
UAG	ASP	MET		
CAG	ASP	SER		

Randomly-ordered polynucleotides composed of 1, 2, 3, or 4 kinds of bases contain 1, 8, 27, and 64 kinds of triplets, respectively. The relative abundance of each kind of triplet can be calculated easily if the base-ratio of a randomly-ordered polynucleotide is known. One can derive both the kinds of bases that correspond to an amino acid and the *number* of bases of each kind, because the amount of each species of amino acid that is incorporated into protein due

to the addition of a polynucleotide preparation and the base-ratio of the polynucleotide can be determined experimentally. In this manner the base compositions of approximately 50 codons were assigned to amino acids[23,24]. The results showed that multiple codons can correspond to the same amino acid; hence the code is highly degenerate. In most cases synonym codons differ by only one base; therefore, it was assumed that the non-variable bases occupy the same relative positions within each synonym word. By means of genetic studies, Crick, Barnett, Brenner and Watts-Tobin[25] showed that the code is a triplet code, and the biochemical studies confirmed this conclusion. Analysis of the coat protein of mutant strains of tobacco mosaic virus provided evidence that triplets in mRNA are translated in a non-overlapping fashion, because the replacement of one base by another in mRNA usually results in only one amino acid replacement in protein[26].

Base Sequence of Codons

Although base compositions of codons were determined, the order of bases within codons was not known. We investigated many potential methods for determining base sequence of codons. A clue to the solution of the problem stemmed from the important finding by Arlinghaus, Favelukes and Schweet[27] and by Kaji and Kaji[28] that Phe-tRNA attaches to ribosomes in response to poly-U prior to peptide bond formation. Perhaps trinucleotides or hexanucleotides of known base sequence would also stimulate binding of AA-tRNA to ribosomes. To test this possibility, Philip Leder and I devised a rapid method for separating ribosomal-bound AA-tRNA from unbound AA-tRNA that depends upon the selective retention of the ribosomal intermediate by discs of cellulose nitrate and then found that trinucleotides function as specific templates for AA-tRNA binding to ribosomes[29]. As shown in Table II, the trinucleotide, AAA, stimulates Lys-tRNA binding to ribosomes and is as active a template for Lys-tRNA as the tetra-or penta-nucleotide. The doublet, AA, has no effect upon Lys-tRNA binding; hence, 3 *sequential* bases in mRNA correspond to I amino acid in protein.

This experimental approach provided a relatively simple means of determining base sequence of codons. Fractionation of poly-(U, G) digests yielded 3 trinucleotides, GUU, UGU, and UUG, which were shown to be codons for valine, cysteine, and leucine, respectively[75].

Trinucleotide synthesis proved to be our major experimental problem. At

Table II

[^{14}C]Lys-tRNA binding to ribosomes

The effect of oligo A preparations upon the binding of *E. coli* Lys-tRNA to ribosomes. The assay for AA-tRNA binding to ribosomes is described elsewhere[29]. Each 50 μl reaction contained 0.4 mμmoles of oligonucleotide as specified; 7.0 $\mu\mu$moles of [^{14}C] - Lys-tRNA (0.150 A^{260} units); 1.1 A^{260} units of *E. coli* ribosomes; 0.05 *M* Tris acetate, pH 7.2; 0.03 *M* magnesium acetate; and 0.05 *M* potassium chloride. In the absence of oligo A, 0.49 $\mu\mu$moles of [^{14}C]Lys-tRNA bound to ribosomes; this amount has been subtracted each value shown above.

Addition	[^{14}C]Lys-tRNA bound due to oligo A ($\mu\mu$moles)
ApA	0.01
ApApA	1.92
ApApApA	1.92
ApApApApA	1.92
ApApApApApA	2.71

that time only 20 to 30 of the 64 trinucleotides were described in the literatu Philip Leder began to explore possible enzymatic methods for synthesizing trinucleotides and we sought the advice of Leon Heppel and Maxine Singer. Throughout the course of our studies on the code Heppel and Singer advised us on problems pertaining to nucleic acids. Each visit to their laboratories became, for me, something akin to a pilgrimage to Delphi. The major difference was that the advice from either oracle was invariably clear and accurate. Marianne Grunberg-Manago was visiting the National Institutes of Health for a few days, and both she and Maxine Singer joined Philip Leder in studying oligonucleotides synthesis catalyzed by primer-dependent poly-nucleotide phosphorylase (Fig. 2). Eventually, conditions for oligonucleotide synthesis were found by Leder, Singer and Brimacombe[30] and by Thatch and Doty[31].

$$\text{ApG} + \text{UDP} \underset{\begin{subarray}{c}\text{Polynucleotide}\\\text{phosphorylase}\\\text{Mg}^{2+}\end{subarray}}{\rightleftharpoons} \text{ApGpU} + \text{ApG(pU)}_n + \qquad (1)$$

$$\text{ApG} + \text{Uridine-2',3'-}\underset{\text{RNAase A}}{\rightleftharpoons} \text{UpApG} + \text{(UP)}_n\text{A} \qquad (2)$$
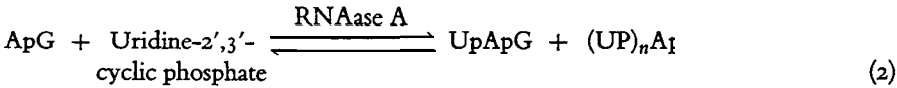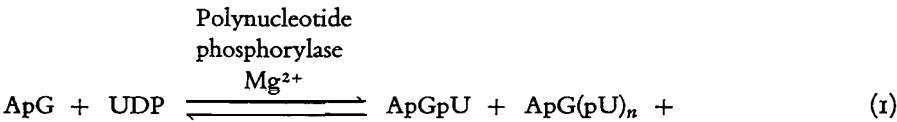cyclic phosphate

Fig.2. Trinucleotide synthesis catalyzed by polynucleotide phosphorylase and by pancreatic RNAase A are shown in reactions 1 and 2, respectively.

Heppel suggested another synthetic method that he, Whitfield and Markham[32] had discovered that depends upon the ability of pancreatic RNAase A to catalyze the synthesis of oligonucleotides from pyrimidine 2',3'-cyclic phosphates and mono- or oligo-nucleotide accepter moieties. Merton Bernfield studied various aspects of the reaction and synthesized many trinucleotides with this enzymes[33-35].

In a remarkable series ofstudies over many years, Khorana and his associates established chemical methods for oligo- and poly-nucleotide synthesis[38]. They were able to synthesize the 64 trinucleotides by chemical methods whereas enzymatic methods were used in our laboratory.

Codon- base sequences were established both by stimulating the binding of AA-tRNA to ribosomes with trinucleotides of known sequence[36,37] and by stimulating in *vitro* protein synthesis with polyribonucleotides containing repeating doublet, triplet, or tetramers of known sequence as described by Khorana in the accompanying article.

THE GENETIC CODE

△ BASE SEQUENCE. (AA-tRNA-TRINUCLEOTIDE-RIBOSOME) COMPLEX
○ BASE COMPOSITION. RNA TEMPLATES FOR PROTEIN SYNTHESIS

Fig. *3.* The symbols represent the following: A, base sequences of mRNA codons determined by stimulating binding of *E. coli* AA-tRNA to *E. coli* ribosomes with trinucleotide templates; 0, base compositions of mRNA codons determined by stimulating the incorporation of amino acids into protein with randomly-ordered polynucleotide templates in extracts of *E. coli.* TERM corresponds to terminator codons (terminator- and initiator-codons are shown in Table III).

The genetic code is shown in Fig. 3. Most triplets correspond to amino acids. Codons for the same amino acid usually differ only in the base occupying the third position of the triplet. Therefore, synonym codons are *systematically* related to one another. Five patterns of codon degeneracy are found,

each pattern determined by the kinds of bases that occupy the third positions of synonym triplets. The third base of each degenerate triplet is shown below; the dashes correspond to the first and second bases of each triplet.

$$— — G \qquad (1)$$

$$— — U \qquad (2)$$
$$— — C$$

$$— — A \qquad (3)$$
$$— — G$$

$$— — U \qquad (4)$$
$$— — C$$
$$— — A$$

$$— — U \qquad (5)$$
$$— — C$$
$$— — A$$
$$— — G$$

The last pattern (discussed in a later section) corresponds to the sum of two patterns.

Results with trinucleotides confirm 43 of the 50 base compositions of codons that were estimated previously on the basis of studies with randomly-ordered polynucleotides and the cell-free protein synthesizing system.

From 1 to 6 codons may correspond to one amino acid, depending upon the amino acid in question. One consequence of systematic degeneracy is that the replacement of one base by another in DNA often does not result in the replacement of one amino acid by another in protein. Many mutations, therefore, are silent ones. The code appears to be arranged so that effects of base replacements in DNA, or erroneous translations of bases in mRNA, often are minimized. Amino acid replacements in protein that occur due to the replacement of one base by another in nucleic acid can be read in Fig. 3 by moving horizontally or vertically from the amino acid in question, but not diagonally

Punctuation

Punctuation of transcription and translation is illustrated schematically in Figs. 4 and 5. RNA polymerase attaches to specific site(s) on DNA and thereby selects the strand of DNA to be transcribed, the direction of transcription, and
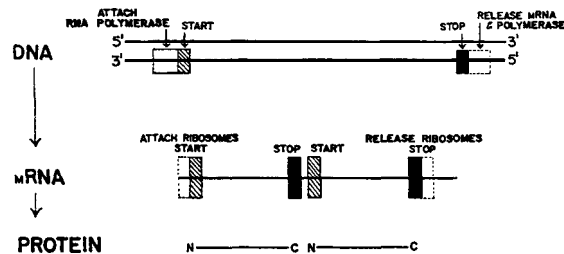
PUNCTUATION

Fig. 4. The punctuation of transcription and translation is illustrated diagramatically. Ribosomal subunits attach to mRNA near the 5'-terminus of the mRNA and are released near the 3'-terminus of the mRNA. Speculations are indicated by the dotted lines. N and C represent the N- and C-terminal amino acid residues of protein, respectively.
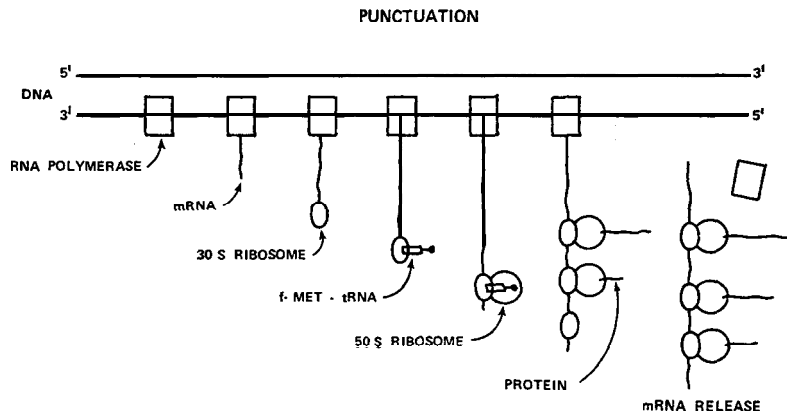
PUNCTUATION

Fig. 5. Diagrammatic illustration of early steps of protein synthesis.

the first base to be transcribed. Many questions remain to be answered about the initiation of RNA synthesis.

The direction of mRNA synthesis is opposite to that of the DNA strand being read. The first base to be incorporated into the nascent mRNA chain is the 5'-terminus of the mRNA, the last base is the 3'-terminus. Similarly, the RNA template is translated during protein synthesis starting at or near the 5'-terminus of the RNA and proceeding three bases at a time, sequentially, toward the 3'- terminus of the RNA. Therefore, mRNA is synthesized and then translated with the same polarity. The first amino acid corresponds to the N-terminus of the peptide chain; the C-terminal amino acid is the last amino acid incorporated.
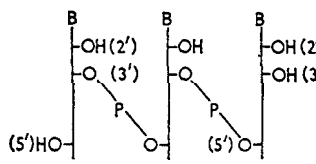
Initiation

Protein synthesis is initiated in *E. coli* by a unique species of tRNA, N-formyl-tRNAf, discovered by Marcker and Sanger[39]. A 30S ribosomal particle attaches to the nascent chain of mRNA near its 5'-terminus before the mRNA detaches from the DNA template. At least three non-dializable factors and GTP are required for the initiation of protein synthesis. The reactions have not been clarified fully; however, the available evidence suggests that one factor (F3) participates in the attachment of a 30S ribosomal subunit to a nascent chain of mRNA and that other factors (FI and F2) and GTP are required for the binding of N-formyl-met-tRNA to the 30S ribosome-mRNA complex in response to an initiator coda (AUG or GUG[62,63]). The 50S ribosomal subunit then attaches to the 30S ribosomal complex before the next codon is recognized by AA-tRNA. N-Formyl-met-tRNA thus selects the first codon to be translated and phases the translation of subsequent codons.

Another species of tRNA from *E. coli,* Met-tRNAm, does not accept formyl moieties, responds only to AUG, and corresponds to methionine at internal positions in protein.

The pattern of degeneracy observed with IV-formyl-met-tRNA differs from the patterns observed with other species of AA-tRNA because initiator codons have alternate first bases rather than alternate third bases.

Each triplet can occur in three structural forms: as 5'-terminal-, 3'-terminal-, or internal-codons. Substituents attached to ribose hydroxyl groups of codons can influence codon template properties profoundly. The relation between codon structure and template activity was investigated by my colleague, Fritz Rottman[83] (Fig. 6). Relative template activities of oligo-U preparations, at limiting oligonucleotide concentrations, are as follows :
p-5'-UpUpU,   >UpUpU,   >CH_3O-p-5'-UpUpU,   >UpUpU-3'-p,
>UpUpU-3'-p-OCH_3>UpUpU- 2',3'-cyclic phosphate. Trimers with
(2',5') phospho diester linkages, (2',5')- UpUpU and (2',5')-ApApA, do not serve as templates for Phe- or Lys-tRNA, respectively. The relative template efficiencies of oligo-A preparations are as follows: p-5'-ApApA > ApApA >ApApA-3'-p >ApApA-2'-p. Ikehara and Ohtsuka[70] showed that N^6-DiMeApApA does not stimulate Lys-tRNA binding to ribosomes; whereas the tubercidin (7-deazaadenosine) analog, TupApA, serves as a template for Lys-tRNA.

RNA polymerase catalyzes the synthesis of mRNA with 5'-terminal triphosphate. Also, many enzymes have been described that catalyze the transfer

Relative template activity of substituted oligonucleotides

Oligonucleotide	Template activity† relative to UpUpU
p-5'-UpUpU	510
UpUpU	100
CH₃O-pUpUpU	74
UpUpU-3'-p	48
UpUpUp-OCH₃	18
UpUpU-2',3'-cyclic p	17
(2'-5')-UpUpU	0
Oligodeoxy T‡	0

Oligonucleotide	Template activity† relative to ApApA
p-5'-ApApA	181
ApApA	100
ApApA-3'-p	57
ApApA-2'-p	15
(2'-5')-ApApA	0
Oligodeoxy A‡	0

Fig. 6. Relative template activities + of substituted oligonucleotides are approximations obtained by comparing the amount of AA-tRNA bound to ribosomes in the presence of limiting concentrations of oligonucleotides compared to either UpUpU for [¹⁴C]Phe-tRNA or ApApA, for [¹⁴C]Lys-tRNA (each designated at 100%). The data are from Rottman and Nirenberg[83].

of molecules to or from hydroxyl groups of nucleic acids. It is possible, therefore, that certain modifications of ribose or deoxyribose hydroxyl groups of nucleic acids provide a means of regulating the rate of transcription or translation.

Since mRNA and protein synthesis are transiently coupled *via* the formation of a (DNA-mRNA-ribosome) intermediate, it is possible that the synthesis of certain species of mRNA may be regulated selectively by events at the level of the ribosome[76-79].

Termination

The first evidence for "nonsense" codons was reported in 1962 by Benzer and Champe[40]. They obtained a mutant of bacteriophage T4 with a deletion span-

ning part of the A gene and part of the contiguous B gene of the rII region. Presumably, the remaining segments of gene A and B are joined and thus form one gene. Nevertheless, a functional B gene product was found. However, a second mutation that mapped in the A gene resulted in the loss of a functional B gene product. These results suggested that a "sense" codon is converted by mutation to a "nonsense" codon that cannot be read; hence subsequent regions of the gene also are not read. Sarabhai, Stretton, Brenner and Bolle[41] then showed that "nonsense" mutations at various sites within the gene for the head protein of bacteriophage T4 determine the chain length of the corresponding polypeptide. These dramatic results showed that "nonsense" codons correspond to the termination of protein synthesis. Additional evidence obtained by Brenner[42,43] and by Garen[44,45] and their colleagues showed that 3 codons, UAA, UAG and UGA, correspond to the termination of protein synthesis (also *cf.* the recent review of Garen[46]).

   The mechanism of peptide-chain termination was investigated by stimulating cell-free protein synthesis with randomly-ordered polynucleotides[47-49], oligonucleotides[50] and polynucleotides[51,52] of known sequence, and viral RNA[53,54]. Capecchi showed that the release of peptides from ribosomes is dependent upon both a release factor and a terminator codon[53]. The codons, UAA, UAG, and UGA, do not stimulate binding of AA-tRNA to ribosomes (although mutant strains of bacteria have been found that contain species of AA-tRNA that respond to terminator codons).

   Recently my colleagues, Caskey, Tompkins, and Scolnick[55,56] found that the process of termination can be studied with trinucleotides. Incubation of terminator trinucleotides and the release factor with the [N-formyl-Met-tRNA-AUG-ribosomal] complex results in the release of free N-formyl-methionine from the ribosomal intermediate. The release factor of *E. coli* then was separated into two components that correspond to different sets of codons : RI, active with UAA or UAG; and R2, active with UAA or UGA. It is clear, therefore, that terminator codons are recognized by specific molecules. The simplest hypothesis is that RI and R2 interact with terminator codons on ribosomes; however, the codon recognition step and the mechanism of termination have not been clarified thus far.

   As shown in Table III, the pattern of codon degeneracy found with RI (UAA and UAG) resembles that found with some species of AA-tRNA; *i.e.,* A is equivalent to G at the 3rd position of codons. However, the degeneracy pattern found with R2 (UAA and UGA) is different from that of AA- tRNA because A and G are equivalent at the 2nd but not at the 3rd position of triplets.

Table I I I

Codons corresponding to the initiation or termination of protein synthesis in *E. coli* are shown. Release factors 1 and 2 are required for termination with the codons indicated, but it is not known whether they interact directly with terminator codons.

Initiation (*N*–formyl–Met–tRNA)	AUG or GUG
Termination (release factor-1)	UAA or UAG
Termination (release factor-2)	UAA or UGA

Redundancy

By 1962, studies with randomly-ordered RNA templates had shown that the code is extensively degenerate and that synonym codons often differ by only one base. It was assumed that the non variable bases occupy the same relative positions within synonym triplets. A systematic form of degeneracy seemed probable because often U was equivalent to C, and A was equivalent to G. Attempts were made to deduce the rules governing degeneracy from the available data on base compositions of codons and amino acid replacements in protein[57,58].

Two species of Leu-tRNA were found that respond to different mRNA codons[59]. However, further work was required to determine whether one species of tRNA responds only to one codon, or to 2 or more codons.

As the order of bases within codons was established, it became abundantly clear that synonym codons are systematically related to one another. As discussed earlier, alternate bases occupy the third position of synonym triplets. Since only a few kinds of degeneracy patterns were found for the 20 amino acids, it seemed likely that correspondingly few codon recognition mechanisms were operative[60].

Evidence that one molecule of AA-tRNA can respond to two kinds of codons was provided by the demonstration that most molecules of Phe-tRNA respond both to UUU and to UUC[61]. Further evidence was obtained by determining the specificity of purified tRNA fractions for trinucleotide codons. The results showed that a purified species of tRNA responds either to 1, 2 or 3 codons[62-67,37]. A summary of our studies with purified fractions of tRNA from *E. coli* is shown in Table IV. Four, possibly 5, kinds of synonym codon sets were found, as shown below. The third base of each synonym triplet is shown; the dashes represent the first and second bases of each triplet.

Table IV

Codons recognized by species of *E. coli* AA-tRNA

Aminoacyl-tRNA preparations from *E. coli* were fractionated by reverse phase column chromatography and their response to trinucleotide templates was determined[67]. Additional results have been obtained by Khorana and his colleagues[64-66]. A dash, —, represents Leu-tRNA fractions that do not respond to trinudeotide codons. Numerals within parentheses indicate the number of redundant peaks of AA-tRNA found.

Amino acid	AA–tRNA		Species		
	1	*2*	*3*	*4*	*5*
			mRNA codons		
LEU	CUU CUC	CUA CUG	CUG	UUG	—(2)
SER	UCU UCC	UCA UCG(2)	UCG	AGU AGC	
ARG	CGU CGC CGA	AGA AGG	CGG		
ALA	GCU GCC	GCA GCG			
VAL	GUU GUC	GUA GUG			
TRP	UGG				
MET	AUG				
ILE	AUU AUC(2)				
PHE	UUU UUC(2)				
TYR	UAU UAC(2)				
CYS	UGU UGC(3)				
HIS	CAU CAC				
LYS	AAA AAG(3)				
GLU	GAA GAG				

— — G   (1)

-   - U   (2)
-    - C

— —A   (3)
-    - G

-    - U   (4)
-    - C
-    - A

— — A   (5)
— — G
— — U

The fifth pattern of degeneracy was found with *E. coli* Ser-tRNA (possibly also with Val-tRNA) but has not been found thus far with AA-tRNA from other organisms.

The number of words, or sets of words, in the code corresponds to the number of tRNA anticodons rather than the number of amino acids. Since multiple species of tRNA for the same amino acid often respond to different sets of codons, the tRNA code consists of more word-sets than the amino acid code.

Redundant fractions of AA-tRNA for the same amino acid were found that differ in chromatographic mobility but respond similarly to codons. Such AA-tRNA fractions may be products of the same gene that have been altered in different ways by enzymes in *vivo* or perhaps have been altered *in vitro* during the fractionation procedure. Alternatively, redundant AA-tRNA fractions may be products of different genes.

Crick suggested that codon degeneracy is due to the formation of alternate base pairs between a base in a tRNA anticodon and alternate bases occupying the third positions of synonym mRNA anticodons[68]. Presumably, the first and second bases of mRNA codons form antiparallel, Watson-Crick base-pairs with corresponding bases in the tRNA anticodon. Alternate base-pairs proposed by Crick are shown in Table V; U in the tRNA anticodon pairs alternately with A or G occupying the third position of synonym mRNA codons; C pairs with G; G pairs with C or U; and I pairs with U, C, or A.

The elucidation of the base sequence of Ala-tRNA from yeast by Holley et *al.*[69] provided an opportunity to relate the base sequence of the   anti-

Table V

Alternate   base-pairing

Alternate base pairing between a base in a tRNA anticodon, shown in the left hand
column, and the base(s) in the third position of synonym mRNA codons. Relationships
are antiparallel "wobble" hydrogen bonds suggested by Crick[68].

tRNA Anticodon	mRNA Codons (3rd base)
U	A
	G
C	G
G	C
	U
I	U
	C
	A

codon with the mRNA codons. Holley generously gave us a preparation of
tRNA[Ala] of known sequence and of high purity and Philip Leder and I and,
concurrently, Söll et al.[65] found that the Ala-tRNA responds to GCU, GCC,
and GCA. The results confirmed Holley's prediction that the sequence, IGC,
serves as the tRNA[Ala] anticodon. Inosine in the anticodon, therefore, pairs
alternately with U, C, or A, in the third position of the mRNA codons. The
base sequences of other species of tRNA have been defined and in every case,
codon-anticodon relationships are in accord with wobble base-pairing.

Universality

The results of many studies suggest that different forms of life use essentially
the same genetic language. However, the fidelity of codon translation can
change quite dramatically due to alterations that affect components required
for protein synthesis. Thus cells sometimes differ in the specificity of codon
translation.

Richard Marshall, Thomas Caskey, and I studied the responses of bacterial,
amphibian, and mammalian AA-tRNA (*E. coli, Xenopus laevis,* and guinea
pig liver, respectively) to trinucleotide codons. Almost identical translations

of nucleotide sequences to amino acids were found with bacterial, amphibian, and mammalian AA-tRNA[71]. However, *E. coli* AA-tRNA preparations do not respond appreciably to certain codons that are active templates with metazoan AA-tRNA.

Our interest in species-dependent variation in codon recognition was stimulated by the possibility that such phenomena might serve as regulators of cell differentiation. Therefore, AA-tRNA preparations were fractionated by column chromatography and responses of tRNA fractions to trinucleotide codons were determined[67]. A summary of our results is shown in Fig.7.

Fig. 7. A summary of the results obtained in this laboratory with purified aminoacyl-tRNA fractions[67] is shown above[67]. Additional results have been reported by Söll *et al.*[64-66]. Synonym codon sets were determined by stimulating the binding of purified fractions of *E. coli,* yeast, or guinea pig liver aminoacyl-tRNA fractions to *E. coli* ribosomes with trinucleotide codons. The joined symbols adjacent to the codons represent synonym codons recognized by one purified aminoacyl-tRNA fraction from ●, *E. coli;* ▲, yeast; or ■ , guinea pig liver. Numerals between symbols represents the number of redundant peaks of aminoacyl-tRNA (aminoacyl-tRNA fractions with the same specificity for codons).

Additional information has been reported by Söll *et al.*[64-66]. Many "universal" species of AA- tRNA were found. However, 7 species of mammalian tRNA were found that were not detected with *E. coli* tRNA and, conversely, 5 species of *E. coli* tRNA were not detected with mammalian tRNA preparations.

Large differences were observed in the concentration of tRNA corresponding to certain codons. Some organisms apparently do not contain AA-tRNA

for certain codons. For example, mammalian Ile-tRNA responds well to AUU, AUC, and AUA; whereas *E. coli* Ile-tRNA responds only to AUU and AUC (AUA-deficient). Also, a species of mammalian Arg- tRNA was found responding to ACG but no Arg-tRNA was found corresponding to AGA (AGA-deficient).

Although some variation in codon translation clearly does occur, the remarkable similarity in codon-base sequences recognized by bacterial, amphibian, and mammalian AA-tRNA suggest that most, perhaps all, forms of life on this planet use essentially the same genetic language, and that the language is translated according to universal rules.

Fossil records of microorganisms estimated to be $3.1 \cdot 10^9$ years old have been reported[72]. The first vertebrates appeared approximately $0.5 \cdot 10^9$ years ago; amphibians and mammals appeared 350 and 180 million years ago, respectively. Thus the genetic code probably originated more than $0.6 \cdot 10^9$ years ago. Hinegardner and Engelberg[73] and Sonneborn[74] suggested that the code was frozen after organisms as complex as bacteria had evolved because major alterations in the code would affect the amino acid sequence of most proteins synthesized by the cell and probably would be lethal.

Reliability of Translation

When one considers the number of species of molecules that are required for the synthesis of a single molecule of protein and the fact that the cellular machinery that participates in the assembly process is complex, heterogeneous, and not reliable, the problem of synthesizing protein with precision seems formidable. To synthesize one molecule of protein composed of 400 amino acid residues, 400 AA-tRNA molecules must be selected in the proper sequence. For the synthesis of the corresponding molecule of mRNA, at least 1206 molecules of ribonucleoside triphosphate must be selected in sequence.

One must distinguish between serial operations, that is, successive steps, and parallel, i.e., simultaneous steps. Usually the overall precision of a multistep process deteriorates rapidly as the number of serial steps increases. Two or more serial steps are required for the synthesis of each molecule of AA-tRNA because an AA-tRNA ligase first catalyzes the synthesis of an aminoacyl-adenylate and then catalyzes the transfer of the aminoacyl moiety to an appropriate species of tRNA, yielding AA-tRNA. Many molecules of AA-tRNA can be synthesized in parallel. Although hundreds of sequential selec-

tions are required for the synthesis of one molecule of protein, the process of protein synthesis is organized within the cell so that each amino acid usually is selected *independently* of other amino acids. Thus, one translational error usually does not influence the accuracy of other codon translations, and errors usually are not cumulative. However, if an error in translation alters the phase of reading or results in premature termination, subsequent selections obviously will be affected.

Baldwin and Berg[80] have shown that Be-tRNA ligase from *E. coli* catalyzes the synthesis of AA-tRNA only if both amino acid and tRNA species are selected correctly. If an erroneous aminoacyl-adenylate is synthesized, the enzyme corrects the error by catalyzing the hydrolysis of the aminoacyl-adenylate.

In 1960 Yanofsky and St.Lawrence[81] suggested that certain mutations might result in the production of structurally modified tRNA or AA-tRNA synthetases with altered specificity for amino acid incorporation into protein. Much information is now available concerning suppressor mutations that affect components required for protein synthesis[82]. In addition, factors that influence the precision of protein synthesis have been studied extensively with synthetic polynucleotide templates and *in vitro* protein-synthesizing systems and by determining the binding of AA-tRNA to ribosomes in response to tri- or poly-nucleotide templates. The results show that the precision of codon recognition is affected by the temperature of incubation, pH, concentration of various species of tRNA, concentration of Mg^{2+}, aliphatic amines such as putrescine, spermidine, spermine, streptomycin and related antibiotics, and other compounds.

Most codons probably are translated with relatively littleerror (0.1-0.01% error or less); however, the level of error can be as high as 50% with certain codons. Hence, the precision of translation can vary from one codon to another at least 5000-fold.

Most errors in codon translation do not result in random amino acid replacements in protein because two out of three bases per codon usually are recognized correctly (*i.e.,* when the precision of translation deteriorates, a codon such as UUU may be translated 80% of the time as phenylalanine, 15% as isoleucine, and 5% as leucine). One codon then is translated by relatively few species of AA-tRNA.

One can only speculate about the biological significance of a flexible, easily modifiable codon-translation apparatus. One extremely interesting possibility is that the codon-recognition apparatus is modified in an orderly, pre-

dictable way at certain times during cell growth and differentiation and that such modifications selectively regulate the rate of synthesis of certain species of protein.

Rate of Translation

The *E. coli* chromosome is composed of approximately $3 \cdot 10^6$ base pairs; sufficient information is present to determine the sequence of $1 \cdot 10^6$ amino acids in protein (equivalent to approximately 2500-3000 species of protein or less since duplicate copies of the same gene may be present).

Approximately 20-80 mRNA triplets are translated per second per ribosome at 37°. One cell may contain 1000-15000 ribosomes per chromosome, depending upon the rate of growth; therefore, proteins are synthesized at many sites simultaneously. Parallel operations greatly enhance the efficiency of the cell in synthesizing protein.

Concluding Remarks

The genetic code is now essentially deciphered. I have been fortunate in having the collaboration of many enthusiastic associates during the course of our studies. To do justice to the years of effort and the important contributions made by associates and numerous colleagues throughout the world is virtually impossible in the available time. One has only to refer to the comprehensive reviews in the *Cold Spring Harbor Symposium on Quantitative Biology* of 1963 and 1966 to view the breadth of the field and the extent of information now available. Additional information can be found in the recent books by Woese[9] and by Jukes[84].

1. O.T. Avery, C. M. MacLeod and M. McCarty, *J.Exptl.Med.,* 79 (1944) 137.
2. G. W.Beadle and E.L.Tatum, *Proc.Natl. Acad. Sci. (U.S.), 27* (1941) 499.
3. J.Brachet, *Arch.Biol. (Liège), 53* (1942) 207.
4. T. Caspersson, *Naturwiss.,* 29 (1941) 33.
5. P. C. Caldwell and C. Hinshelwood, *J.Chem. Soc.,* Pt. 4 (1950) 3 156.
6. A.L.Dounce, *Enzymologia,* 15 (1952) 251.

7. G.Gamow, *Nature,* 173 (1954) 318.

8. J.D. Watson and F.H.C.Crick, *Nature,* 171 (1953) 737.

9. CR. Woese, *The Genetic Code,* Chapter 2, Harper and Row, New York, 1967.

10. A.D.Hershey, J.Dixon and M. Chase, *J. Gen. Physiol.,* 36 (1953) 777.

11. E.Volkin and L.Astrachan, *Virology,* 2 (1956) 149.

12. M.R.Pollock, *Proc.Roy.Soc. (London), Ser.B,* 148 (1958) 340.

13. M.R.Pollock, *in* I. *C.* Gunsalus and R. *Y.* Stanier (Eds.), *The Bacteria,* Vol.4, 1962, p.121.

14. M. R. Lamborg and P. *C. Zamecnik, Biochim.Biophys. Acta,* 42 (1960) 206.

15. A.Tissieres, D. Schlessinger and F. Gros, *Proc. Natl. Acad. Sci. (U.S.),* 46 (1960) 1450.

16. T.Kameyama and G.D. Novelli, *Biochem.Biophys. Res.Commun.,* 2 (1960) 393.

17. B.Nisman and H.Fukuhara, *Compt. Rend.,* 249 (1959) 2240.

18. M. W.Nirenberg and J.H.Matthaei, *Proc.Natl.Acad.Sci.(U.S.),* 47 (1961) 1588.

19. J. H. Matthaei and M. W. Nirenberg, *Proc. Natl. Acad. Sci. (U.S.),* 47 (1961) 1580.

20. M. W. Nirenberg, J. H. Matthaei and O. W. Jones, *Proc. Natl.Acad. Sci. (U.S.), 48* (1962) 104.

21. M. Grunberg-Manago, P. J. Ortiz and S. Ochoa, *Biochim.Biophys. Acta,* 20 (1956) 269.

22. M. Singer, O. W. Jones and M. W. Nirenberg, *Proc. Natl. Acad. Sci. (U. S.),* 49 (1963) 392.

23. J.F. Speyer, P.Lengyel, C.Basilio, A. J. Wahba, R. S. Gardner and S. Ochoa, *Cold Spring Harbor Symp.Quant.Biol.,* 28 (1963) 559.

24. M. W.Nirenberg, O. W. Jones, P.Leder, B.F. C.Clark, W. S. Sly and S.Pestka, *Cold Spring Harbor Symp.Quant.Biol.,* 28 (1963) 549.

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

26. H. G. Wittmann and B. Wittmann-Leibold, *Cold Spring Harbor Symp.Quant.Biol.,* 28 (1963) 589.

27. R. Arlinghaus, *G.* Favelukes and R. Schweet, *Biochem.Biophys. Res. Commun.,* 11 (1963) 92.

28. A.Kaji and H.Kaji, *Biochem.Biophys. Res.Commun.,* 13 (1963) 186.

29. M. W.Nirenberg and P.Leder, *Science,* 145 (1964) 1399.

30. P.Leder, M.F. Singer and R.Brimacombe, *Biochemistry, 4* (1965) 1561.

31. R.E.Thach and P.Doty, *Science,* 147(1965) 1310.

32. L.A.Heppel, P.R. Whitfield and R.Markham, *Biockem.J.,* 60 (1955) 8.

33. M.Bemfield, *J.Biol.Chem.,* 240 (1965) 4753.

34. M. Bernfield, *J.Biol. Chem.,* 241 (1966) 2014.

35. M.Bernfield and F.M.Rottman, *J.Biol.Chem.,* 242 (1967) 4134.

36. D. Söll, J. Cherayil, D. S. Jones, R.D. Faulkner, H. Hampel, R. M. Bock and H. G. Khorana, *Cold Spring Harbor Symp.Quant.Biol.,* 31 (1966) 51.

37. M. W. Nirenberg, T. Caskey, R.Marshall, R.Brimacombe, D.Kellog, B.Doctor, D. Hatfield, J. Levin, F. Rottman, S. Pestka, M. Wilcox and F. Anderson, *Cold Spring Harbor Symp.Quant.Biol.,* 31(1966) 11; B.P.Doctor, J.E.Loebel and D. A.Kellogg, *ibid.,* 31 (1966) 543; D.Hatfield, *ibid.,* 31 (1966) 619; S.Pestka and M.Nirenberg, *ibid.,* 31 (1966) 641.

38. H. G. Khorana, H.Buchi, H. Ghosh, N. Gupta, T.M. Jacob, H. Kössel, R.Morgan, S. A. Narang, E. Ohtsuka and R. O. Wells, *Cold Spring Harbor Symp.Quant.Biol.,* 31 (1966) 39.

39. K.Marcker and F.Sanger, *J.Mol.Biol., 8* (1964) 835.

40. S.Benzer and S.P.Champe, *Proc.Natl.Acad.Sci.(U.S.),* 48 (1962) 1114.

41. A. S. Sarabhai, A. O. W. Stretton, S.Brenner and A.Bolle, *Nature,* 201 (1964) 13.

42. S.Brenner, A. O. W. Stretton and S.Kaplan, *Nature,* 206 (1965) 994.

43. S.Brenner, L.Barnet, E.R. Katz and F. H. C. Crick, *Nature,* 213 (1967) 449.

44. M. G. Weigert and A. Garen, *Nature, 206* (1965) 992.

45. M. G. Weigert, E.Lanka and A.Garen, *J.Mol.Biol.,* 23 (1967) 391.

46. A.Garen, *Science,* 160 (1968) 149.

47. M.Takanami and Y.Yan, *Proc. Natl. Acad. Sci. (U.S.),* 54 (1965) 1450.

48. M. S.Bretscher, H.M.Goodman, J.R.Menninger and I.D. Smith, *J.Mol.Biol.,* 14 (1965) 634.

49. M.C.Ganoza and T.Nakamoto, *Proc. Natl. Acad. Sci. (U.S.),* 55 (1966) 162.

50. J.A.Last, W.M.Stanley Jr., M.Salas, M.B.Hille, A.J.Wahba and S.Ochoa, *Proc. Natl. Acad. Sci.(U.S.),* 57 (1967) 1062.

51. A.RMorgan, R.D. Wells and H.G.Khorana, *Proc.Natl.Acad. Sci. (U.S.), 56* (1966) 1899.

52. H.Kössel, *Biochim.Biophys. Acta,* 157 (1968) 91.

53. M.R. Capecchi, *Proc. Natl. Acad. Sci. (U.S.),* 58 (1967) 1144.

54. M. S.Bretscher, *J. Mol.Biol.,* 34 (1968) 131.

55. CT. Caskey, R. Tompkins, E. Scolnick, T. Caryk and M.Nirenberg, Science, 162 (1968) 135.

56. E. Scolnick, R.Tompkins, T. Caskey and M. Nirenberg, *Proc. Natl. Acad. Sci. (U.S.),* 61, (1968) 768.

57. C.Woese, *Nature,* 194 (1962) 1114.

58. R.Eck, Science, 140 (1963) 477.

59. B.Weisblum, S.Benzer and R. W. Holley, *Proc. Natl. Acad. Sci. (U.S.),* 48 (1962) 1449.

60. M. W.Nirenberg, P.Leder, M.Bernfield, R.Brimacombe, J.Trupin, F.Rottman and *C.* O'Neal, *Proc. Natl. Acad. Sci. (US.),* 53 (1965) 1161; J.Trupin, F.Rottman, R. Brimacombe, P. Leder, M.Bernfield and M. Nirenberg, *ibid.,* 53 (1965) 807; R. Brimacombe, J.Trupin, M. Nirenberg, P. Leder, M.Bernfield and T.Jaouni, ibid., 54 (1965) 954.

61. M.R.Bernfield and M. W.Nirenberg, *Science,* 147 (1965) 479.

62. B. F. C. Clark and K.A.Marcker, *J.Mol.Biol.,* 17 (1966) 394.

63. D. A.Kellogg, B. P.Doctor, J. E. Loebel and M. W. Nirenberg, *Proc. Natl.Acad. Sci. (U.S.),* 55 (1966) 912.

64. D. Söll, D. S. Jones, E. Ohtsuka, R.D. Faulkner, R.Lohrmann, H.Hayatsu, H.G. Khorana, J.D. Cherayil, A. Hampel and R. M. Bock, *J. Mol.Biol.,* 19 (1966) 556.

65. D. Söll, J.D. Cherayil and R. M.Bock, *J.Mol.Biol.,* 29 (1967) 97; D. Söll E. Ohtsuka, D. S. Jones, R.Lohrmann, H.Hayatsu, S. Nishimura and H.G.Khorana, *Proc. Natl. Acad.Sci. (U.S.),* 54 (1965) 1378.

66. D. Söll and U. L.RajBhandary, *J. Mol.Biol., 29* (1967) 113.

67. C. T. Caskey, A. Beaudet and M. Nirenberg, *J. Mol. Biol.,* 37 (1968) 99.

68. F.H. *C.* Crick, *J. Mol.Biol.,* 19 (1966) 548.

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

70. M.Ikehara and E. Ohtsuka, *Biochem.Biophys.Res.Commtun,* 21 (1965) 257.

71. R. E. Marshall, C. T. Caskey and M. W. Nirenberg, *Science,* 155 (1967) 820.

72. E.Barghoorn and J.Schopf, Science, 152 (1966) 758.

73. R. Hinegardner and J. Engelberg, *Science,* 144 (1964) 1031.

74. T. M. Sonneborn, *in* V. Bryson and H. J. Vogel (Eds.), Evolving *Genes and Proteins,* Academic Press, New York, 1965, p. 377.

75. P.Leder and M. W.Nirenberg, *Proc. Natl. Acad. Sci. (U.S.),* 51 (1964) 420, 1521.

76. H.Bremer and M. W.Konrad, *Proc. Natl. Acad. Sci. (U.S.),* 51 (1964) 801.

77. G. S. Stent, *Science,* 144 (1964) 816.

78. R. Byrne, J. G. Levin, H. A. Bladen and M. W. Nirenberg *Proc. Natl. Acad. Sci.(U.S.),* 52 (1964) 140.

79. H. A.Bladen, R.Byrne, J.G.Levin and M. W. Nirenberg, *J.Mol.Biol.,* 11 (1965) 78.

80. A.N.Baldwin and P.Berg, *J.Biol.Chem.,* 241 (1966) 839.

81. C.Yanofsky and P.St.Lawrence, *Ann.Rev.Microbiol.,* 14 (1960) 311.

82. L. Gorini and J.R.Beck with, *Ann. Rev. Microbial.,* 20 (1966) 401.

83. F.Rottman and M.Nirenberg, *J. Mol.Biol.,* 21 (1966) 555.

84. T. Jukes, *Molecules and Evolution,* Columbia University Press, New York, 1967.