EXPERIENTIA

Vol. XX – Fasc. 2 Pag. 57–112 15. II. 1964

Chemical Basis of Heredity, the Genetic Code*

By SEVERO OCHOA

Department of Biochemistry, New York University School of Medicine, New York (U.S.A.)

Work of recent years has thrown considerable light on the way in which the genetic information stored in DNA[1] is transmitted to the protein making machinery of the cell, so that a specific nucleotide sequence in DNA gives rise to a unique amino acid sequence in the polypeptide chains of proteins. It may be said that DNA contains a coded message, the genetic code, with instructions for the manufacture of specific proteins. We know now that RNA participates in this process as a messenger[2,3] between the DNA and the protein. DNA directs the synthesis of a specific messenger RNA and this in turn directs the synthesis of a specific protein.

The relation between DNA, messenger RNA, and protein is illustrated schematically in Figure 1, in which portions of the nucleotide and amino acid chains of these compounds are shown as strips of recording tape. Moreover, under the assumption that the genetic code is a triplet code, the continuous sequence of nucleotide bases in DNA and RNA is represented as a sequence of separate triplets for better visualization of the correspondence between the individual base triplets of messenger RNA and the various amino acids. The amino acids on the protein strip of Figure 1 are known to correspond to the triplets having the base composition, but not necessarily the sequence, shown on the RNA strip. An account of the experimental work on which this correlation is based is the main purpose of this lecture. When the base composition and sequence of the coding units for all of the amino acids in proteins is known, the genetic code will have been deciphered. We may not be far from this goal.

Code Read Out Mechanism. RNA nucleotidyl transferase, the enzyme catalyzing the synthesis of messenger RNA with DNA as template, operates through a Watson-Crick base pairing mechanism and yields, as shown in Figure 1, complementary copies – or negatives – of one of the DNA strands[4]. In this first step the DNA code is transcribed into an RNA code represented by a unique sequence of bases. In what follows the words genetic code or amino acid code will be used to denote the messenger RNA code.

In the read out step the messenger RNA serves as a template for the alignment of the amino acids in a sequence prescribed by its own base sequence. For this purpose each amino acid is linked through its carboxyl group to one end of a specific amino acid acceptor or transfer RNA (s-RNA) of small molecular weight (about 25,000). This reaction is catalyzed by a specific enzyme which recognizes both the amino acid and its s-RNA. Energy is provided each time by the splitting of one molecule of ATP to AMP and inorganic pyrophosphate. There are as many amino acid activating enzymes (aminoacyl-s-RNA synthetases) and at least as many species of s-RNA as there are different amino acids, i.e. at least twenty. The coding triplets of messenger RNA are recognized by complementary adaptor

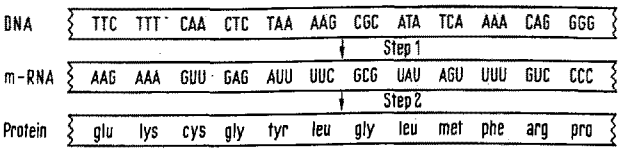

Fig. 1. Two-step transcription of genetic code. RNA nucleotidyl transferase, which catalyzes the DNA-dependent synthesis of RNA, is the key enzyme in the transcription of the genetic message from DNA to RNA (step 1). The code read-out step (step 2) can be considered as a translation from the four character language of RNA into the twenty character language of proteins. m-RNA stands for messenger RNA.

* Fifth PAUL KARRER Lecture, held at the University of Zürich (Switzerland), June 26, 1963.

[1] The following abbreviations are used: RNA, ribonucleic acid; DNA, deoxyribonucleic acid; A, C, G, U, T, the bases adenine, cytosine, guanine, uracil, thymine, also the corresponding nucleosides or nucleotides; in the case of polynucleotides, poly A means polyadenylic acid, poly UG, polyuridylic-guanylic acid, etc.; H, hypoxanthine; poly I, polyinosinic acid; ADP, CDP, GDP, UDP, the ribonucleoside 5′-diphosphates of adenosine, cytidine, guanosine, uridine; ATP, adenosine 5′-triphosphate; GTP, guanosine 5′-triphosphate; TMV, tobacco mosaic virus. The amino acids are abbreviated by using the first three letters of their names (e.g., thr, threonine; try, tryptophan; asp, aspartic acid), with the exception of isoleucine asparagine, and glutamine which are abbreviated as ile, asn and gln, respectively. In the shorthand writing of polynucleotides, as in pApUpUpU...pU, the letter p to the left of the nucleoside initial indicates a 5′-phosphate; the letter p to the right, a 3′-phosphate.

[2] F. JACOB and J. MONOD, J. mol. Biol. *3*, 318 (1961).

[3] S. BRENNER, F. JACOB, and M. MESELSON, Nature *190*, 576 (1961).

[4] J. MARMUR, Cold Spring Harbor Symp. Quant. Biol., in press (1963).

triplets (e.g. AAA for UUU) in s-RNA[5,6]. The adaptors are believed to be responsible for the specific attachment, through a base-pairing mechanism, of the amino acid bearing s-RNA to positions prescribed by the base sequence of the messenger RNA which, on interaction with the ribosomes, becomes a template for protein synthesis.

As shown for hemoglobin synthesis[7–9] the polypeptide chain is assembled sequentially from the N-terminal end. The s-RNA is released stepwise from the amino acid (*a*) last linked to the growing polypeptide chain, as a new aminoacyl-s-RNA (*b*) is placed on the next position of the template and a peptide bond is formed involving the acyl group of *a* (with cleavage of the aminoacyl-s-RNA linkage) and the amino group of the amino acid carried by *b* (Figure 2).

Recent work has shown[10–12] that during protein synthesis the ribosomes form aggregates (polysomes) which are apparently held together by messenger RNA for they are dissociated by ribonuclease. These observations have been interpreted as meaning that several ribosomes are simultaneously engaged in making protein on the same messenger as they move along its chain or as the messenger moves over the ribosomal surface. This is illustrated diagrammatically in Figure 3.

As discussed elsewhere[13] it has until recently been a matter for speculation whether the ribosomes deliver ready made protein or only the unfolded polypeptide chains, i.e. the protein's primary structure. The demonstration[14] that ribonuclease and other enzymes can assume their native globular conformation in aerated buffer solutions, following unfolding through reduction of their disulfide bridges in urea solutions, proves that the secondary and tertiary structure of proteins may be formed spontaneously and are determined by their primary structure, i.e. the amino acid composition and sequence. Folding of the polypeptide chains might occur partially on the ribosomes and be completed after release.

Another question of interest is the maximum size of individual polypeptide chains manufactured by the ribosomes. There is increasing evidence that native proteins are aggregates of subunits of relatively small

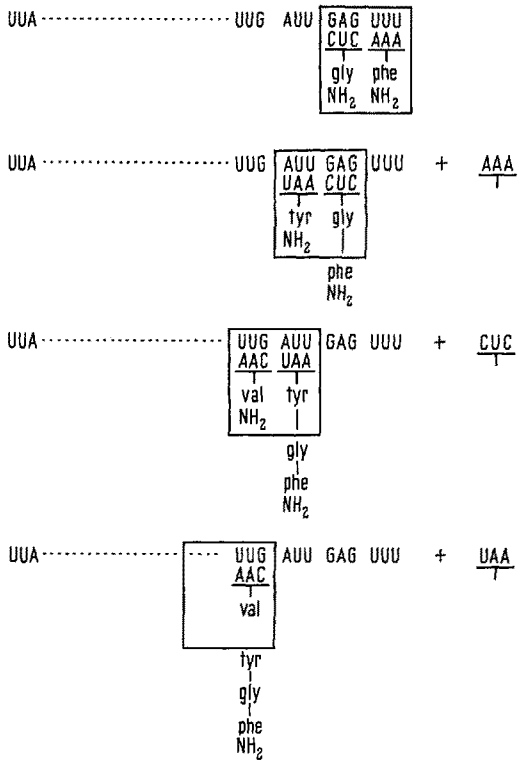

Fig. 2. Scheme illustrating the assembly of polypeptide chains and messenger read-out. The free amino group of the amino acids attached to s-RNA is shown for easier visualization of the polarity of chain growth. The s-RNA is represented as a T with the adaptor triplet on top of the horizontal limb and the amino acid linked (through its carboxyl group) at the free end of the vertical limb. The rectangles represent the ribosomal binding site. In the diagram the polypeptides are assumed to be assembled on the ribosomes as they move along the messenger RNA chain but it is equally possible that the messenger itself moves over the surface of static ribosomes.

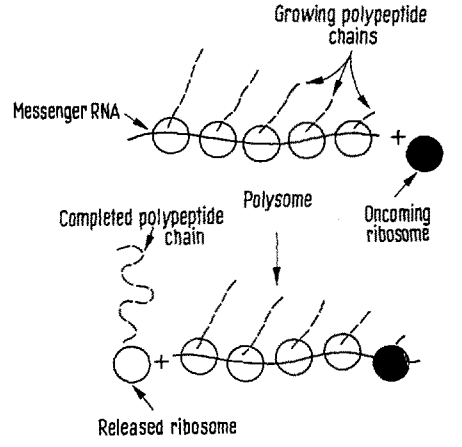

Fig. 3. Schematic model of polysome function (courtesy of Dr. A. RICH).

[5] F. H. C. CRICK, Symp. Soc. exp. Biol. *12*, 138 (1958).
[6] F. CHAPEVILLE, F. LIPMANN, G. VON EHRENSTEIN, B. WEISBLUM, W. J. RAY, JR., and S. BENZER, Proc. Nat. Acad. Sci. U.S. *48*, 1086 (1962).
[7] J. BISHOP, J. LEAHY, and R. SCHWEET, Proc. Nat. Acad. Sci. U.S. *46*, 1030 (1960).
[8] H. M. DINTZIS, Proc. Nat. Acad. Sci. U.S. *47*, 247 (1961).
[9] A. GOLDSTEIN, Biochim. biophys. Acta *53*, 468 (1961).
[10] J. R. WARNER, P. M. KNOPF, and A. RICH, Proc. Nat. Acad. Sci. U.S. *49*, 122 (1963).
[11] A. GIERER, J. mol. Biol. *6*, 148 (1963).
[12] W. GILBERT, J. mol. Biol. *6*, 374 (1963).
[13] S. OCHOA, in *Informational Macromolecules* (H. J. VOGEL, V. BRYSON, and J. O. LAMPEN, eds., Academic Press, New York 1963), p. 3.
[14] C. B. ANFINSEN, in *Informational Macromolecules* (H. J. VOGEL, V. BRYSON, and J. O. LAMPEN, eds., Academic Press, New York 1963), p. 153.

size. The molecular weight of the subunits of various proteins (hemoglobin, apoferritin, proteins of plant and bacterial viruses), from 17,000 to about 25,000, would correspond in the upper range to chains of 200 amino acid residues. For a triplet code this would require a messenger 600 nucleotides long (molecular weight about 180,000).

No small subunits have yet been detected in enzymes. However, several enzymes and various other proteins are known to be polymers of two to four units each ranging in molecular weight from 40,000 to 250,000[13]. It appears entirely possible that these units may in turn consist of a number of primary subunits. If so, most proteins would be formed by the – spontaneous or catalyzed – aggregation of small subunits produced under genetic control by the protein making machinery of the cell. In some cases the primary subunits might aggregate first into larger, second order units which would finally polymerize to the native protein.

Messenger RNA, like DNA, has only four different bases whereas proteins may have up to twenty different amino acids. Therefore, at least three bases (triplet code) would be needed to specify one amino acid. Combination of four elements three at a time yields 4^3 or 64 triplets, more than enough to determine 20 amino acids. A doublet code would not supply enough coding units, since it would yield only 4^2 or 16 doublets. It appears, therefore, that at least a code of triplets is required for transcription of the genetic message. Genetic experiments of CRICK et al.[15] have provided elegant evidence in favor of a triplet code.

Synthetic Polynucleotides as Messengers. The question arose whether synthetically prepared polyribonucleotides would substitute for natural messenger RNA in cell-free systems of protein synthesis and would promote the incorporation of amino acids into proteins. If so, any correlation between the base composition of such polynucleotides and the nature of the amino acids incorporated might open the way to the deciphering of the genetic code. This proved to be the case.

Key to the experiments with artificial messengers was the enzyme polynucleotide phosphorylase discovered in our laboratory in 1955[16]. This enzyme catalyzes the synthesis of polynucleotides from ribonucleoside 5'-diphosphates with liberation of orthophosphate. The overall, reversible reaction catalyzed by polynucleotide phosphorylase may be represented by the equation below, in which P∼P stands for the pyrophosphate moiety of nucleoside diphosphates.

$$n\ \text{Base-ribose-P}\sim\text{P} \xrightleftharpoons{(Mg^{++})} (\text{base-ribose-P})_n + P_i$$

In the direction to the left the enzyme catalyzes the phosphorolysis of polyribonucleotides with formation of the corresponding ribonucleoside diphosphates. Homopolynucleotides, i.e. polymers of only one kind of nucleotide (poly A, poly U, etc.) are produced from individual nucleoside diphosphates (ADP, UDP, etc.). Copolynucleotides containing two or more kinds of nucleotide residues (poly AU, poly AUG, etc.), are formed from mixtures of nucleoside diphosphates (ADP and UDP; ADP, UDP, and GDP, etc.). The chemical structure of the polynucleotides synthesized by polynucleotide phosphorylase is the same as that of natural RNA. In fact, the copolymer poly AGUC, containing adenylic, guanylic, uridylic, and cytidylic acid residues, is indistinguishable from RNA except for the random distribution of nucleotides[17,18]. Random distribution of nucleotides permits the calculation of the relative abundance of different base triplets in copolynucleotides of known base composition and is an essential condition for use of these compounds in the code studies discussed below.

A further convenient feature of polynucleotide phosphorylase is that the base composition of copolynucleotides can be predetermined by using for their preparation mixtures of nucleoside diphosphates in the same ratio as the desired base ratios. Many polynucleotides of preselected, widely varying base composition have thus been prepared in our laboratory. The actually determined base ratios were in all cases very close to the expected ones[19]. Divergent results of other laboratories[20] might be due to the presence of contaminating enzymes (nucleases, nucleoside mono- and diphosphokinases, etc.) in the polynucleotide phosphorylase preparations.

The effect of synthetic polynucleotides in promoting the incorporation of amino acids into protein-like products by a cell-free system from *Escherichia coli* was first reported by NIRENBERG and MATTHAEI[21]. These workers observed that poly U specifically promoted the incorporation of phenylalanine and obtained indications that it directed in fact the synthesis of polyphenylalanine. Thus UUU would be a triplet code letter for phenylalanine. Investigations in our laboratory[22] (see also [23]) next showed that copolynucleotides such as poly AU stimulated the incorporation of phenylalanine, isoleucine, leucine, and tyrosine. Poly

[15] F. H. C. CRICK, L. BARNETT, S. BRENNER, and R. J. WATTS-TOBIN, Nature *192*, 1227 (1961).
[16] S. OCHOA, Angew. Chem. *72*, 225 (1960).
[17] L. A. HEPPEL, P. J. ORTIZ, and S. OCHOA, J. biol. Chem. *229*, 695 (1957).
[18] P. J. ORTIZ and S. OCHOA, J. biol. Chem. *234*, 1208 (1959).
[19] R. S. GARDNER, A. J. WAHBA, C. BASILIO, R. S. MILLER, P. LENGYEL, and J. F. SPEYER, Proc. Nat. Acad. Sci. U.S. *48*, 2087 (1962).
[20] M. S. BRETSCHER and M. GRUNBERG-MANAGO, Nature *195*, 283 (1962).
[21] M. W. NIRENBERG and J. H. MATTHAEI, Proc. Nat. Acad. Sci. U.S. *47*, 1558 (1961).
[22] P. LENGYEL, J. F. SPEYER, and S. OCHOA, Proc. Nat. Acad. Sci. U.S. *47*, 1936 (1961).
[23] S. OCHOA, Fed. Proc. *22*, 62 (1963).

UC stimulated the incorporation of phenylalanine, leucine, proline, and serine; poly UG that of phenylalanine, cysteine, leucine, valine, glycine, and tryptophan. Polynucleotides with three different kinds of nucleotide residues, e.g. poly UCG, stimulated the incorporation of arginine and alanine among other amino acids. Copolynucleotides other than those just mentioned were synthesized later and were found to promote the incorporation of still other amino acids until, eventually, the incorporation of each of the 20 amino acids could be brought about by one or another polymer.

Experiments with homopolymers other than poly U (poly A, poly C) were at first negative or gave equivocal results. However, later work in our laboratory[19,24] showed that poly A directed the synthesis of polylysine and poly C promoted the incorporation of proline into acid-insoluble material. The effect of poly A had previously been missed because polylysine is soluble in trichloroacetic acid but was eventually detected with use of tungstic acid, in which polylysine is insoluble, as the precipitating agent. Poly C was found to have low affinity for the ribosomes and the early negative results were due to the use of too small amounts of this polynucleotide.

Table I illustrates some of these findings. It shows portions of the chain of various homo- and copolynucleotides with some of the triplets occurring in them and the amino acids whose incorporation into protein-like products by the *E. coli* system was correlated with the triplets of the base composition shown.

Amino Acid Code. Most of the studies on the amino acid code have utilized cell-free preparations of *E. coli* consisting of suspensions of ribosomes in the supernatant fluid obtained by high speed centrifugation of bacterial extracts. The supernatant contributes s-RNA's, aminoacyl-s-RNA synthetases, and other enzymes concerned with the polymerization of the activated amino acid residues. Incubation of these preparations with ATP, potassium (or ammonium) and magnesium ions, a mixture of the twenty amino acids, and GTP, which is required in an as yet unknown manner for amino acid polymerization, leads to the incorporation of small amounts of amino acids into acid-insoluble products. This incorporation is stimulated to a variable degree by the addition of various kinds of RNA or synthetic polynucleotides. It is often further increased by supplemental s-RNA. By use of amino acid mixtures in which one of the twenty amino acids is labeled with C^{14} at a time, the effect of different polynucleotides on the incorporation of individual amino acids can easily be determined. The incubation is allowed to proceed at 37° for 30 or more min by which time the incorporation of amino acids has practically stopped. In this manner one measures the total amount of amino acid incorporated, not the incorporation rate.

Some examples of the effect of synthetic polynucleotides on the incorporation of various amino acids are shown in Tables II and III[19,24,25]. Table II shows the effect of poly UG (5:1) and poly UCG (6:1:1) on the incorporation of arginine, phenylalanine, glycine, and tryptophan. The use of polynucleotides containing a large excess of uridylic acid residues was prompted by the consideration that such polynucleotides would give rise to phenylalanine rich polypeptides which, even if of small size, would be insoluble in trichloroacetic acid. The small incorporation in the absence of added polynucleotides may be due to the presence of residual natural messenger RNA on the

Table I. Some synthetic polynucleotides and their effect on the incorporation of various amino acids into polypeptide chains. For better visualization of the correlation between individual base triplets and amino acids, the polynucleotide chains are arbitrarily represented as a sequence of nucleotide triplets rather than a continuous sequence of nucleotides

Polynucleotide	Polynucleotide chain	Amino acids incorporated
Poly A	...AAA AAA AAA AAA ...AAA ...	Lys (polylysine synthesis)
Poly C	...CCC CCC CCC CCC ...CCC ...	Pro (polyproline synthesis)
Poly U	...UUU UUU UUU UUU...UUU...	Phe (polyphenylalanine synthesis)
Poly UG	...GUU GUG UUU UUG...UGU...	Cys, gly, phe, val, leu, etc.
Poly ACG	...AGG CAG AAG AAA ...GAG ...	Ser, ala, glu, lys, gly, etc.

Table II. Amino acid incorporation in *E. coli* system with various polynucleotides. Values expressed as mμmoles/mg ribosomal protein. In this Table and elsewhere, the figures in parentheses following the abbreviated names of synthetic polynucleotides indicate the molar ratios of the individual nucleoside diphosphates used for preparation of the polymers [e.g. poly UG (5:1) was prepared from a mixture of UDP and GDP in molar ratio 5:1]. The actual base composition of the polynucleotides used in this work closely reflected these ratios

Amino acid	Polynucleotide added		
	None	Poly UG (5:1)	Poly UCG (6:1:1)
Phenylalanine	0.18	13.40	10.60
Arginine	0.12	0.04	0.47
Glycine	0.19	0.74	0.45
Tryptophan	0.03	0.70	0.46

[24] A. J. WAHBA, R. S. GARDNER, C. BASILIO, R. S. MILLER, J. F. SPEYER, and P. LENGYEL, Proc. Nat. Acad. Sci. U.S. *49*, 116 (1963).

[25] J. F. SPEYER, P. LENGYEL, C. BASILIO, and S. OCHOA, Proc. Nat. Acad. Sci. U.S. *48*, 441 (1962).

ribosomes. As expected from the abundance of uridylic acid residues, these polynucleotides promoted very effectively the incorporation of phenylalanine. Both poly UG and poly UCG stimulated the incorporation of glycine and tryptophan but only the latter stimulated the incorporation of arginine. From these results one can conclude that arginine has a code triplet which contains U, C, and G.

Most of the synthetic polynucleotides in the experiments of Table III contained no uridylic acid but had instead a large proportion of adenylic acid residues. This was because, from the effect of poly A on the incorporation of lysine mentioned earlier, one would expect A-rich polynucleotides to lead to the synthesis of lysine-rich polypeptides precipitable by tungstic acid. It may be seen that of all twenty amino acids, poly A was effective only in promoting lysine incorporation while poly AC, for example, was active in promoting among others the incorporation of asparagine, glutamine, lysine, and threonine.

Table IV[19,25,26] presents a summary of most of the experiments carried out to date in our laboratory. Three groups of polynucleotides were used with a large proportion of uridylic acid (U-rich), adenylic acid (A-rich), and cytidylic acid (C-rich) residues, respectively. Polynucleotides of the first group promoted the incorporation of phenylalanine to an extent greater than that of any other amino acid whereas this was true of lysine and proline for polynucleotides of the second and

Table III. Amino acid incorporation in *E. coli* system with various polynucleotides. Net incorporation values, after subtraction of small blank in the absence of added polynucleotides, expressed as mμmoles/mg ribosomal protein. Absence of numerical values indicates no polynucleotide effect on the incorporation of a given amino acid

Amino acid	Polynucleotide							
	A	AU (5:1)	AC (5:1)	AG (5:1)	ACG (4:1:1)	ACG (6:1:1)	C	CI (5:1)
Ala	–	–	–	–	0.12	0.06	–	0.11
Arg	–	–	–	0.05	0.55	0.35	–	0.09
Asn	–	0.13	0.30	–	0.35	0.39	–	–
Asp	–	–	–	–	0.08	0.06	–	–
Glu	–	–	–	0.11	0.43	0.47	–	–
Gln	–	–	0.44	0.02	0.62	0.64	–	–
Gly	–	–	–	0.02	0.07	0.03	–	0.02
His	–	–	0.09	–	0.26	0.32	–	–
Ile	–	0.10	–	–	–	–	–	–
Lys	1.2	0.47	0.99	0.36	2.06	2.92	–	0.44
Pro	–	–	0.05	–	0.18	0.08	0.72	–
Ser	–	–	–	–	0.16	0.10	–	–
Thr	–	–	0.23	–	0.49	0.57	–	0.07

[26] S. Ochoa, in *Informational Macromolecules* (H. J. Vogel, V. Bryson, J. O. Lampen, eds., Academic Press, New York 1963), p. 437.

Table IV. Effect of synthetic polynucleotides on the incorporation of amino acids into protein in *E. coli* system. The experimental results are expressed as % of the incorporation of the one amino acid (phenylalanine, lysine, proline) which is promoted maximally by polymers of each of the three groups (U-rich, A-rich, C-rich). The incorporation of each amino acid was individually tested with each of the polynucleotides shown except for the three cases marked with an asterisk. Absence of numerical values indicates that there was no effect

Amino acid	Polynucleotide													
	U-rich							A-rich					C-rich	
	U	UA (5:1)	UC (5:1)	UG (5:1)	UAC (6:1:1)	UCG (6:1:1)	UAG (6:1:1)	A	AU (5:1)	AC (5:1)	AG (5:1)	ACG (4:1:1)	C	CI[a] (5:1)
Ala	–	–	–	–	–	3	–	–	–	–	–	6	–	22
Arg	–	–	–	–	–	3	–	–	–	–	13	27	–	19
Asn	–	7	–	–	7	–	5	–	28	30	–	17	–	–
Asp	–	–	–	–	–	–	3	–	–	–	–	4	–	–
Cys	–	–	–	20	–	25	32	–	–	–	–	–	–	–
Glu	–	–	–	–	–	–	2	–	–	–	30	21	–	–
Gln	–	–	–	–	*	–	*	–	–	44	6	30	–	–
Gly	–	–	–	4	–	3	–	–	–	–	5	3	–	5
His	–	–	–	–	3	–	–	–	–	9	–	13	–	–
Ile	–	20	–	–	16	–	32	–	20	–	–	–	–	–
Leu	–	14	20	13	25	25	27	–	3	–	–	–	–	–
Lys	–	3	–	–	2	–	–	100	100	100	100	100	–	–
Met	–	–	–	–	–	–	4	–	–	–	–	–	–	–
Phe	100	100	100	100	100	100	100	–	–	–	–	–	–	–
Pro	–	–	8	–	3	3	–	–	–	5	–	9	100	100
Ser	–	–	25	–	25	26	–	–	–	–	–	8	–	–
Thr	–	–	–	–	9	–	–	–	–	23	–	24	–	–
Try	–	–	–	5	–	4	*	–	–	–	–	–	–	–
Tyr	–	25	–	–	25	–	20	–	3	–	–	–	–	–
Val	–	–	–	20	–	20	25	–	–	–	–	–	–	–

[a] Poly CI (polycytidylic inosinic acid) was used in place of poly CG because the former promotes amino acid incorporation somewhat more effectively than the latter. As shown previously with other polymers (e.g. poly UG and poly UI), hypoxanthine can be substituted for guanine without change in the coding characteristics of the polynucleotides.

Table V. Correlation of relative triplet abundance with the incorporation of amino acids promoted by some polynucleotides

Polynucleotide	Triplets	Relative abundance of each triplet	Relative amino acid incorporation
Poly UG (5:1)	UUU	100	Phe, 100
	UUG, UGU, GUU	20	Cys, 20; val, 20
	UGG, GUG, GGU	4	Gly, 4; try, 5
	GGG	0.8	
Poly AC (5:1)	AAA	100	Lys, 100
	AAC, ACA, CAA	20	Asn, 30; thr, 23, glu, 44
	ACC, CAC, CCA	4	Pro, 5
	CCC	0.8	
Poly CG (5:1)	CCC	100	Pro, 100
	CCG, CGC, GCC	20	Ala, 22; arg, 19
	CGG, GCG, GGC	4	Gly, 5
	GGG	0.8	

third group, respectively. It may be seen that the base composition of the polynucleotides had a marked effect on the kind and amount of amino acids incorporated [cf., for example, poly AC with poly AG, poly UC with poly UCG, poly UA (5:1) with poly AU (5:1)].

For each polynucleotide group in Table IV the incorporation of any one amino acid relative to that of a reference amino acid (e.g. phenylalanine), can be correlated with the calculated mean abundance of triplets of a given base composition relative to that of a reference triplet (e.g. UUU in U-rich polynucleotides). This enabled us to assign triplets of known base composition to individual amino acids.

Table V[26] gives some examples of matching triplet abundance to amino acid incorporation. The Table lists from left to right a polynucleotide of preselected base composition, the relative abundance of each triplet assuming random nucleotide distribution, and the relative incorporation of amino acids promoted by this polynucleotide. From the effect of poly U, poly A, and poly C in stimulating, respectively, the incorporation of phenylalanine, lysine, and proline to the exclusion of other amino acids, the activity of each of the three polynucleotides in Table V on the incorporation of the above amino acids is referable in each case to the high abundance of UUU, AAA, and CCC triplets. If the abundance of UUU triplets in poly UG (5:1) is taken as 100, *each* of the triplets having 2 U's and 1 G (UUG, UGU, and GUU) will occur with a mean relative abundance of 20 while *each* of those having 1 U and 2 G's (UGG, GUG, and GGU) will occur with a mean abundance of 4. The abundance of GGG triplets is too low to be of any significance.

Poly UG (5:1) promoted the incorporation of phenylalanine, cysteine, valine, glycine, and tryptophan to a relative extent of 100, 20, 20, 4, and 5, respectively. The incorporation agreed exceedingly well with the mean abundance of UUU (phenylalanine), each of the 2U1G (cysteine, valine), and each of the 1U2G (glycine, tryptophan) triplets. It was therefore concluded that cysteine and valine are each coded by triplets of base composition 2U1G, while glycine and tryptophan are each coded by triplets of base composition 1U2G. From the results (Table V) with poly AC (5:1) and poly CG (5:1) it was concluded that asparagine and threonine are coded by 2A1C, proline by 1A2C (besides CCC), alanine and arginine by 2C1G, and glycine by 1C2G triplets. It must be emphasized that, except for AAA, CCC, and UUU, the experiments so far described give the base composition but not the base sequence of the code triplets.

Experiments with a variety of polynucleotides, with use of C^{14}-labeled amino acids of high specific radioactivity, have led to the code triplet assignments for all twenty amino acids shown in Table VI[24,27]. The

Table VI. Genetic code triplets

Amino acid	Code triplets	Shared doublets
Alanine	CUG, CAG, CCG	C•G
Arginine	GUC, GAA, GCC	G•C
Asparagine	UAA, CUA, CAA	C•A
Aspartic acid	GUA, GCA	G•A
Cysteine	*GUU*	
Glutamic acid	AUG, AAG	A•G
Glutamine	UAC, AAC	•AC
Glycine	GUG, GAG, GCG	G•G
Histidine	AUC, ACC	A•C
Isoleucine	UUA, AAU, CAU	•AU
Leucine	UUC, CCU, UGU, UAU	U•U?
Lysine	AUA, AAA	A•A
Methionine	AGU	
Phenylalanine	UUU, UCU	U•U
Proline	CUC, CAC, CCC	C•C
Serine	CUU, ACG, UCC	
Threonine	UCA, ACA, CCA	•CA
Tryptophan	UGG	
Tyrosine	*AUU*, ACU	A•U
Valine	UUG	

[27] A. J. WAHBA, R. S. MILLER, C. BASILIO, R. S. GARDNER, P. LENGYEL, and J. F. SPEYER, Proc. Nat. Acad. Sci. U.S. *49*, 880 (1963).

order of the bases in this Table is purely tentative and probably wrong. It is mostly based on correlations[28] of the base composition of the various triplets with single amino acid replacements in proteins, assumed to be due to single base mutations, and on experimental evidence to be mentioned later that the one triplet so far disclosed for cysteine has the absolute sequence GUU and one of the tyrosine triplets the sequence AUU. These two triplets are in italics in the Table.

Degeneracy. Since proteins contain twenty different amino acids the genetic code might consist of only twenty meaningful triplets (non-degenerate code). On the other hand, each of several amino acids might be coded by more than one triplet (degenerate code). It appears from Table VI that the code is extensively degenerate. It may be further seen that degeneracy occurs according to a regular pattern whereby the triplets for a given amino acid often share a doublet, e.g. C and G for alanine and arginine, G and A for aspartic and glutamic acid, A and C for glutamine and histidine, etc. The occurrence of a regularity in the code has also been pointed out by ECK[29]. In a degenerate code, replacement of one base within a triplet would occasionally lead to one of the other triplets for the same amino acid. Thus, from an evolutionary point of view extensive degeneracy of the genetic code would tend to stabilize the species by decreasing the frequency of mutations.

Degeneracy of the code would appear to involve the existence of different s-RNA's for one and the same amino acid, each with the appropriate adaptor for each of the amino acid's code triplets. Several workers have reported on the separation by various methods of two or more s-RNA's for each of several amino acids[30,31], and WEISBLUM et al.[32] have found that one of two leucine-s-RNA fractions isolated from *E. coli* interacts predominantly with poly UG whereas the other does so mainly with poly UC.

To date, 46 out of the 64 possible triplets have been assigned. The base distribution for the sum total of code triplets at present is A 38, U 40, G 23, C 37; $A+U/G+C = 1.3$. It is not possible to say how many of the remaining 18 triplets will prove to be meaningful. It is not unlikely that in the code $A=U=G=C$ and, since at present $A \simeq U \simeq C > G$, several G-rich triplets would seem to be still missing. Detection of further G-rich triplets is difficult due to the ease with which G-rich polynucleotides assume secondary structure and to the fact, noted by various workers, that polynucleotides with a high degree of secondary structure do not stimulate amino acid incorporation in the cell-free system due perhaps to lack of interaction with the ribosomes. This situation is not helped by the fact that inosinic acid can replace guanylic acid in coding[33] for I-rich polynucleotides also acquire secondary structure readily. Since poly G, contrary to poly I, is not easily synthesized by polynucleotide phosphorylase[16], we have tried poly I for stimulation of amino acid incorporation, glycine and tryptophan in particular, under a variety of conditions but it was inactive. It cannot be decided whether this inactivity is due to secondary structure or to intrinsic meaninglessness of the III (GGG) triplet.

Table VII lists all of the 64 triplets, arranged in 20 series, and the amino acids assigned to 46 triplets with indication of the 'empty' spaces. The tentative base sequences of the previous Table have been used in this one. It is of interest to note that no one triplet series has a number of assigned amino acids higher than the number of triplets in it; this gives confidence on the experimental approach and the assumptions made in our triplet assignments. As noted by a footnote, the assignments for the triplet series 1, 2, 3, 5, 7, 8, 10, 12, 14, 15, and 18 are complete. From this it follows that all of the triplets in poly AU (series 5 and 8), poly AC (series 7 and 14), poly UC (series 10 and 15), and poly AUC (series 18), are meaningful. These polynucleotides should therefore direct the synthesis of long, unbroken polypeptide chains.

Work on the amino acid code has been carried out independently, and almost simultaneously, mainly in

Table VII. Distribution of amino acids among triplets

Triplet series	Triplets	Amino acids
1[a]	AAA	Lys
2[a]	UUU	Phe
3[a]	CCC	Pro
4	GGG	–
5[a]	AUU UAU UUA	Tyr, leu, ile
6	AGG GAG GGA	–, gly, –
7[a]	ACC CAC CCA	His, pro, thr
8[a]	UAA AUA AAU	Asn, lys, ile
9	UGG GUG GGU	Try, gly, –
10[a]	UCC CUC CCU	Ser, pro, leu
11	GAA AGA AAG	Arg, –, glu
12[a]	GUU UGU UUG	Cys, leu, val
13	GCC, CGC, CCG	Arg, –, ala
14[a]	CAA ACA AAC	Asn, thr, gln
15[a]	CUU UCU UUC	Ser, phe, leu
16	CGG GCG GGC	–, gly, –
17	AUG AGU UAG UGA GAU GUA	Glu, met, –, –, –, asp
18[a]	AUC ACU UAC UCA CAU CUA	His, tyr, gln, thr, ile, asn
19	AGC ACG GAC GCA CAG CGA	–, ser, –, asp, ala, –
20	UGC UCG GUC GCU CUG CGU	–, –, arg, –, ala, –

[a] Series with complete amino acid assignments.

[28] T. H. JUKES, Amer. Scientist *51*, 227 (1963).
[29] R. V. ECK, Science *140*, 477 (1963).
[30] D. P. DOCTOR, J. APGAR, and R. W. HOLLEY, J. biol. Chem. *236*, 1117 (1961).
[31] N. SUEOKA and T. YAMANE, in *Informational Macromolecules* (H. J. VOGEL, V. BRYSON, J. O. LAMPEN, eds., Academic Press, New York 1963), p. 205.
[32] B. WEISBLUM, S. BENZER, and R. W. HOLLEY, Proc. Nat. Acad. Sci. U.S. *48*, 1449 (1962).
[33] C. BASILIO, A. J. WAHBA, P. LENGYEL, J. F. SPEYER, and S. OCHOA, Proc. Nat. Acad. Sci. U.S. *48*, 613 (1962).

two laboratories – our own and that of NIRENBERG. In their code studies the Bethesda workers[34,35] followed the approach that had been started in our laboratory with use of copolynucleotides containing different proportions of various bases. With few exceptions their results have been in good agreement with ours.

Code Triplets and Mutations. The experimentally determined triplets of the genetic code are in good agreement with single amino acid replacements which take place in the protein coat of tobacco mosaic virus as a result of single-base substitutions brought about by treatment of TMV RNA with nitrous acid[36]. Two changes can be brought about in this way; one is the deamination of cytosine to uracil (C→U), the other the deamination of adenine to hypoxanthine (A→H). As previously mentioned, hypoxanthine can be substituted for guanine in coding, showing that it is equivalent to guanine in this respect. Moreover on replication of the RNA in the tobacco leaf guanine will take the place of hypoxanthine. Therefore, the net result of the deamination of adenine will be an A→G conversion.

Infection of tobacco plants with HNO_2-treated TMV RNA often results in the production of TMV mutants which can be characterized because the lesions to which they give rise are different from those produced by wild type TMV. As illustrated in Figure 4, in some of these mutants a single amino acid is replaced by a different amino acid (e.g. serine by phenylalanine, aspartic acid by glycine) at certain positions of the polypeptide chain of the viral protein[36,37], and such replacements can be correlated with the C→U or the A→G change produced by HNO_2.

As shown in Table VIII, in which the amino acid code triplets are given in the base sequence of the two preceding Tables, 12 out of the 17 replacements reported agree with the code triplets that have been deduced from experiments with the cell-free *E. coli* system. This agreement, to the extent of 70%, cannot be due to chance for random agreement could only occur to the extent of 16–17%. The reason for the lack of agreement of the 5 amino acid replacements listed in the legend to the Table is not known. Replacements observed only once, e.g. asp→ser, may have been the result of spontaneous rather than HNO_2-evoked mutation. It is also possible that some of the non-agreeing replacements involve still 'missing' triplets.

Universality. The question whether the genetic code is universal, i.e. whether there is but one code for all living beings, is of considerable interest. A number of facts suggest that this is indeed the case. The following may be mentioned: (a) The incorporation into hemoglobin of leucine from *E. coli* leucyl-C^{14}-s-RNA by a mixed system of *E. coli* supernatant and rabbit reticulocyte ribosomes[38]. This suggests that the adaptor triplets of *E. coli* leucine s-RNA 'fit' the leucine triplets of rabbit hemoglobin messenger RNA. (b) The agreement noted in the previous section of the amino acid replacements in the protein of HNO_2-induced mutants of TMV with the code triplets derived from experiments with the *E. coli* system. This suggests that the TMV code is the same as that of *E. coli*. (c) The agreement noted by various workers[28,39] of single amino acid replacements due to spontaneous – presumably single base – mutations in human hemoglobin and other proteins with the *E. coli* triplets.

...UUU CCC GUA *CUU* GUA UUU AAA... RNA chain (wild type)
...phe pro asp *ser* asp phe lys ... Peptide IV chain (wild type)

...UUU CCC GUA *UUU* GUA UUU AAA... RNA chain (mutant)
...phe pro asp *phe* asp phe lys ... Peptide IV chain (mutant)

...GCC UUG *GCA* GUA CAG UCA UUG... RNA chain (wild type)
...arg val *asp* asp ala thr val ... Peptide IX chain (wild type)

...GCC UUG *GCG* GUA CAG UCA UUG... RNA chain (mutant)
...arg val *gly* asp ala thr val ... Peptide IX chain (mutant)

Fig. 4. Scheme illustrating the relation of single amino acid replacements in TMV protein to single base changes in the RNA brought about by nitrous acid treatment. Upper half, C→U conversion; lower half, A→G conversion.

Table VIII. Amino acid replacements in nitrous acid mutants of tobacco mosaic virus. Of 17 different replacements, 12 (or 70%) agree with the code triplet assignments. The five non-agreeing replacements (number of times observed given in parentheses) are asn→ser (2), asp→ser (1), asp→ala (6), thr→met (3), and gln→val (2)

Replacement	Times observed	Code triplet change
C→U		
Thr→ile	8	UCA→UUA
Pro→ser	4	CCC→UCC
Pro→leu	3	CCC→CCU
Leu→phe	1	UUC→UUU
Ser→phe	7	UCC→UCU
Ser→leu	2	UCC→UUC
A→G		
Asp→gly	4	GCA→GCG
Ile→val	3	UUA→UUG
Ile→met	1	AAU→AGU
Glu→gly	2	AUG→GUG
Thr→ser	2	ACA→ACG
Arg→gly	5	GAA→GAG

[34] O. W. JONES, JR. and M. W. NIRENBERG, Proc. Nat. Acad. Sci. U.S. *48*, 2115 (1962).
[35] M. W. NIRENBERG and O. W. JONES, JR., in *Informational Macromolecules* (H. J. VOGEL, V. BRYSON, J. O. LAMPEN, eds., Academic Press, New York 1963), p. 451.
[36] H. G. WITTMANN, Z. Vererbungslehre *93*, 491 (1962).
[37] A. TSUGITA and H. FRAENKEL-CONRAT, J. mol. Biol. *4*, 73 (1962).
[38] G. VON EHRENSTEIN and F. LIPMANN, Proc. Nat. Acad. Sci. U.S. *47*, 941 (1961).
[39] E. L. SMITH, Proc. Nat. Acad. Sci. U.S. *48*, 677 (1962).

More recent experiments in several laboratories with cell-free systems from animal and bacterial sources[40–45], have shown that synthetic polynucleotides (poly A, poly U, poly AU, poly AC, poly UG, etc.) promote in general the incorporation into acid-insoluble material of the same amino acids as the *E. coli* system. Thus far these results also speak in favor of the universality of the genetic code.

Preliminary Experiments on Base Sequence of Code Triplets. With extensive degeneracy of the code, determination of the base sequence of the code triplets will at best be a long and tedious task. Only a modest beginning has been made. This was based on the use of polynucleotides with a triplet of predetermined sequence at one end of the chain. HEPPEL et al.[46] have demonstrated that polynucleotide phosphorylase adds nucleotide residues to short oligonucleotide chains. These serve as nuclei for growth of the polynucleotide and act therefore as primers of polynucleotide synthesis. As shown in Figure 5 when the dinucleotide pApU (AU) is added as primer for the synthesis of poly U, the polynucleotide pApUpUpU...pUpU (AUUU... UUU) is formed. It has an AUU triplet at the origin of the chain. In experiments with the *E. coli* system[47] this polynucleotide promoted, of course, extensive incorporation of phenylalanine into acid-insoluble material; it also promoted consistently a very small incorporation of tyrosine, but of no other amino acids, including isoleucine, asparagine, and lysine which are coded by 2U1A or 1U2A triplets. Similar experiments with a poly GUUU...UU, suggested the sequence GUU for the 2U1G cysteine triplet.

Attempts were made to locate the position of the incorporated tyrosine, which should be present at one end or another of a polyphenylalanine chain, by end group assays. The solubility of polyphenylalanine, however, made these attempts exceedingly difficult. Nevertheless, following solubilization of the product by sulfonation with concentrated sulfuric acid, experiments with either phenylalanine-C^{14} and cold tyrosine or tyrosine-C^{14} and cold phenylalanine, indicated the presence of N-terminal phenylalanine, but not tyrosine, as well as that of C-terminal tyrosine. N-terminal groups were assayed by Sanger's dinitrofluorobenzene and C-terminal groups by Akabori's hydrazinolysis method. From these preliminary results it would appear that the polypeptide synthesized by the *E. coli* system with poly AUUU...UU as messenger has the sequence shown in Figure 5. Were this indeed the case these experiments would disclose the direction in which the code is read since, as already mentioned, proteins are synthesized in the direction from the N-terminal to the C-terminal amino acid. If one writes the polynucleotide chain, as in Figure 5, with the end bearing a monoesterified 5′ phosphate to the left (pApUpUpU... pUpU), synthesis of the polynucleotide proceeds from left to right while that of the polypeptide would proceed in the opposite direction, i.e. the code would be read from right to left.

We have recently tried a different approach for the determination of triplet sequences. We have synthesized poly AC (cytidylic acid residues labeled with C^{14}) and poly AU (uridylic acid residues labeled with C^{14}) containing on the average 30 times as many A as C or U residues. These polynucleotides promoted effectively the incorporation of lysine into acid-insoluble material. Digestion with pancreatic ribonuclease yielded ApApAp...ApApCp or ApApAp...ApApUp from which the terminal, 3′-monoesterified phosphate was removed with phosphomonoesterase to yield ApApAp... ApApC and ApApAp...ApApU, respectively. The average chain length of these fragments, as determined by end group assay, was in the neighborhood of 30 residues. However, while these short poly A's with AAC and AAU triplets at the right-hand side end were 30 to 50% as active as the parent polynucleotides in promoting the incorporation of lysine, they proved to be inactive toward other amino acids including asparagine, glutamine and isoleucine (cf. Table VI). This failure was probably due to the presence in the *E. coli* supernatant of an exonuclease[48] which, like snake venom phosphodiesterase, cleaves polynucleotides by successive removal of nucleoside 5′-monophosphate residues starting from the end with two unesterified hydroxyls, i.e. the right-hand side end in our present convention. The radioactivity of our C-C^{14}- or U-C^{14}-ended polynucleotides was in fact rendered acid-soluble with extreme rapidity upon incubation with *E. coli* supernatant.

(1) pApU + nUDP → pApUpUpUpUpUpUpU...pUpUpU + nP

(2) pApUpUpUpUpUpUpUpUpUpU...pUpUpU

(3) (COOH) tyr phe phe phe...phe(NH_2)

Fig. 5. Scheme illustrating preliminary experiments on base sequence determination of 2U1A tyrosine triplet and direction of code reading. (1) Synthesis of poly AUUU...UUU template with polynucleotide phosphorylase. (2) Template chain. (3) Polypeptide formed.

[40] H. R. S. ARNSTEIN, R. A. COX, and J. A. HUNT, Nature *194*, 1042 (1962).

[41] E. S. MAXWELL, Proc. Nat. Acad. Sci. U.S. *48*, 1639 (1962).

[42] I. B. WEINSTEIN and A. N. SCHECHTER, Proc. Nat. Acad. Sci. U.S. *48*, 1686 (1962).

[43] A. C. GRIFFIN and M. A. O'NEAL, Biochim. biophys. Acta *61*, 496 (1962).

[44] I. B. WEINSTEIN, in *Informational Macromolecules* (H. J. VOGEL, V. BRYSON, and J. O. LAMPEN, eds., Academic Press, New York 1963), p. 246.

[45] J. J. PROTASS, J. F. SPEYER, and P. LENGYEL, Science, in press.

[46] M. F. SINGER, L. A. HEPPEL, and R. J. HILMOE, J. biol. Chem. *235*, 738 (1960).

[47] A. J. WAHBA, C. BASILIO, J. F. SPEYER, P. LENGYEL, R. S. MILLER, and S. OCHOA, Proc. Nat. Acad. Sci. U.S. *48*, 1683 (1962).

[48] P. F. SPAHR and D. SCHLESSINGER, J. biol. Chem. *238*, PC2251 (1963).

While attempts are being made to remove nucleases from the supernatant, other polynucleotides are being prepared, e.g. CAAA...AA, by priming the synthesis of poly A with C^{14}-labeled di- and trinucleotides of known base sequence, for further experiments of the type described above with AUUUU...UU. The presence of radioactivity in the nucleotides at the origin of the chain is expected to facilitate characterization of these polymers whereas the poly A sequence, by directing the synthesis of lysine-rich, soluble polypeptides, should facilitate the structural study of the products.

It might be mentioned that an alternating poly AU, prepared with RNA nucleotidyl transferase using poly dAT as template, was inactive in stimulating amino acid incorporation in the *E. coli* system, even though all triplets having A and U (and therefore AUA and UAU) are meaningful. This negative result may be attributed to the high degree of secondary structure in alternating poly AU.

Identification of Polypeptides Synthesized with Polynucleotide Messengers. Identification of the products formed by the incorporation of amino acids into acid-insoluble material, promoted in cell-free systems by synthetic polynucleotides, lagged behind the use of these polymers in studies on the amino acid code. However, since the assignment of code triplets is based on the assumption that homopolynucleotides direct the synthesis of homopolypeptides and, even more important, that copolynucleotides direct the synthesis of copolypeptides, evidence that this is indeed the case was badly needed.

In their first paper on the effect of poly U NIRENBERG and MATTHAEI[21] reported the partial characterization of the product as polyphenylalanine based mainly on its solubility properties. Characterization by means of proteolytic enzymes was precluded by its insolubility. With the demonstration[19] that poly A promoted the incorporation of lysine, and A-rich polynucleotides that of lysine with smaller amounts of other amino acids into tungstic acid-insoluble but water soluble products, it became possible to identify the products as polypeptides by conventional chemical and enzymatic methods[19,49].

Tryptic digestion of the tungstic acid-insoluble material formed on incubation of lysine-C^{14} with the *E. coli* system, in the presence of poly A, yielded radioactive di- and trilysine (Figure 6). These are the main products of cleavage by trypsin of chemically synthesized poly-L-lysine[50]. The product was, as expected, resistant to chymotrypsin. These experiments showed unequivocally that poly A directs the synthesis of poly-L-lysine by the *E. coli* system.

End group assay by the dinitrofluorobenzene method of samples of polylysine-C^{14}, isolated by precipitation with a mixture of trichloroacetic and tungstic acids, gave mean chain lengths varying from 17 to 30 residues in several experiments. Since the average chain length of the poly A used was in excess of 200 nucleotide residues – which for a triplet code should yield polylysine of a mean length of 70 or over – cleavage of the poly A by nucleases may have been responsible in part for the shorter length of the polylysine formed. There was in fact cleavage of C^{14}-labeled poly A on incubation with the *E. coli* system under the conditions of polylysine synthesis.

STEWART et al.[51] have developed an excellent method for the fractionation of lysine polypeptides on carboxymethylcellulose columns. Analysis by SMITH and STAHMANN[52] of total reaction mixtures after incubation with poly A and lysine-C^{14}, without prior isolation of lysine peptides, showed the presence of di-, tri-, tetralysine etc. up to polypeptides with 15–17 residues. Dilysine made up almost half of the peptides present with decreasing amounts of higher peptides as their chain length increased. There were no lysine peptides in samples incubated without poly A. These results suggest that the *E. coli* fractions contain proteolytic enzymes which hydrolyze polylysine to yield dilysine, trilysine, and some higher oligopeptides soluble in tungstic acid. This possibility was in part corroborated by the finding that biosynthetic polylysine-C^{14}, isolated by the tungstic acid procedure, was slowly degraded upon incubation with a mixture of *E. coli* supernatant and ribosomes.

Poly AU (5:1) was used for identification of copolypeptides[49]. As already pointed out all triplets in AU polymers are meaningful. These polymers promote the

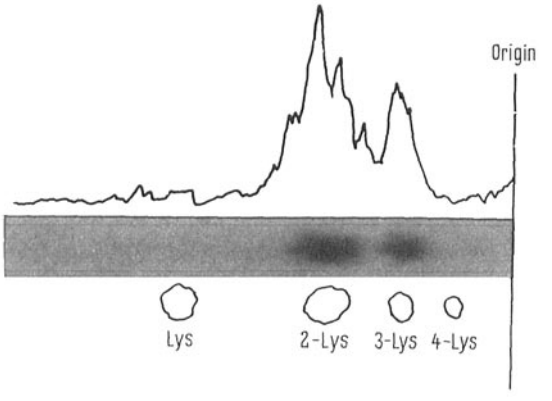

Fig. 6. Identification of the product of lysine incorporation in the presence of poly A as poly L-lysine. The products of tryptic digestion were separated by paper chromatography. Upper panel, actigraph tracing; middle panel, autoradiogram; lower panel, from left to right, lysine, dilysine, trilysine, and tetralysine markers.

[49] Y. KAZIRO, A. GROSSMAN, and S. OCHOA, Proc. Nat. Acad. Sci. U.S. *50*, 54 (1963).
[50] S. G. WALEY and J. WATSON, Biochem. J. *55*, 328 (1953).
[51] J. W. STEWART and M. A. STAHMANN, J. Chromatogr. *9*, 233 (1962).
[52] M. A. SMITH and M. A. STAHMANN, Biochem. biophys. Res. Comm., in press.

incorporation of six amino acids, namely asparagine, isoleucine, leucine, lysine, phenylalanine and tyrosine. The % incorporation relative to that of lysine and the % triplet abundance relative to that of the two lysine triplets (AAA, 2A1U) for poly AU (5:1) are shown in Table IX. In one experiment a mixture of isoleucine-C^{14} with cold asparagine, leucine, lysine, phenylalanine, and tyrosine, was incubated with poly AU, *E. coli* ribosomes and supernatant and the remaining usual additions. Following isolation of the tungstic acid-insoluble peptides and digestion with trypsin, high voltage paper electrophoresis yielded four main radioactive peptide peaks (Figure 7, upper tracing). It might be noted in passing that omission of cold lysine markedly reduced the incorporation of isoleucine-C^{14} into trichloroacetic-tungstic acid-insoluble material. A similar pattern was obtained in another experiment with a mixture of asparagine-C^{14} and cold isoleucine, leucine, lysine, phenylalanine, and tyrosine. The pattern of radioactive peptides was analogous when lysine was the only labeled amino acid, except for the presence of two additional peaks moving closer to the cathode in the position of dilysine and trilysine. When two of the six amino acids were labeled, e.g. isoleucine and lysine (Figure 7, lower tracing) or asparagine and lysine, the pattern was very much the same as when only lysine was labeled, with six major radioactive peaks of which peaks 5 and 6 moved in the position of di- and trilysine, respectively.

The radioactive peptide peaks were eluted, hydrolyzed with hydrochloric acid, and the free amino acids separated by paper chromatography. The actigraph tracings of Figure 8 show the distribution of radioactivity in the HCl hydrolysates of peptide peaks 1 and 3 from the experiments with both isoleucine and lysine (sections A and B) and asparagine and lysine labeled (sections C and D). Asparagine is of course converted to aspartic acid by hydrolysis. Both peaks yielded radioactive lysine and isoleucine in one case and lysine and aspartic acid in the other. It may be seen, however, that the proportion of lysine relative to that of the other amino acid was higher in peak 3 (the peak closer to the cathode) than in peak 1. This suggests that whereas peak 1 peptides may contain labeled lysine and isoleucine (or asparagine) in a 1:1 ratio, peak 3 peptides probably contain these amino acids in a ratio of 2:1 or higher.

Further work aimed at closer identification of the polypeptides formed with synthetic copolynucleotide messengers is in progress. In the meantime the above, preliminary experiments leave little doubt that copolynucleotides do indeed direct the cell-free synthesis of copolypeptides and lend firm support to the results and deductions of the work on the genetic code reviewed in this lecture.

Concluding Remarks. The fact that synthetic polyribonucleotides can take the place of natural messenger

Table IX. Relative triplet abundance and amino acid incorporation with poly AU (5:1)

Amino acid	Code triplets	Triplet abundance %	Incorporation %
Lysine	AAA, 2A1U	100	100
Isoleucine	2A1U, 1A2U	20	20
Asparagine	2A1U	16.6	28
Leucine	1A2U	3.3	3.2
Tyrosine	1A2U	3.3	3.4
Phenylalanine	UUU	0.7	–

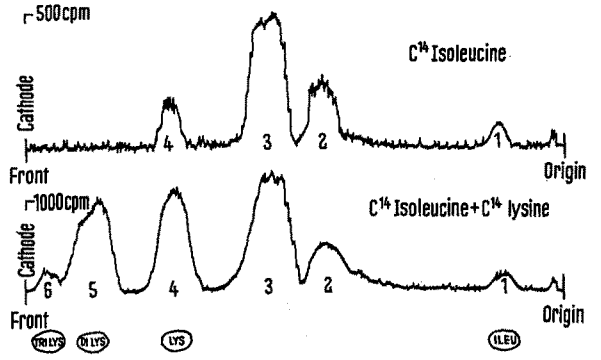

Fig. 7. Tryptic peptides from products of incorporation of various amino acids in presence of poly AU (5:1). Actigraph tracings of peptides separated by high voltage paper electrophoresis. Upper tracing, the incubation mixture contained isoleucine-C^{14} and cold asparagine, leucine, lysine, phenylalanine, and tyrosine. Lower tracing, the incubation mixture contained isoleucine-C^{14}, lysine-C^{14}, and cold asparagine, leucine, phenylalanine, and tyrosine. Bottom panel, from left to right, trilysine, dilysine, lysine, and isoleucine markers.

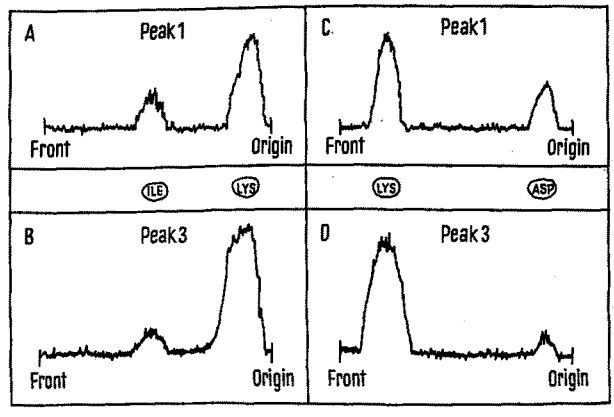

Fig. 8. Amino acids released by acid hydrolysis of tryptic peptides from products of incorporation of various amino acids in the presence of poly AU (5:1). Actigraph tracings of paper chromatograms. A and B, peaks 1 and 3 from experiment with isoleucine-C^{14} and lysine-C^{14}. C and D, peaks 1 and 3 from experiment with asparagine-C^{14} and lysine-C^{14}. The position of amino acid markers is shown on the middle strip.

RNA and direct the cell-free synthesis of polypeptides, the amino acid composition of which depends on the base composition of the polynucleotides, has brought us surprisingly close to the solution of the genetic code puzzle. The enzyme polynucleotide phosphorylase has, as we have seen, been the key to this development. Progress in our understanding of the code, together with the spectacular advances in our knowledge of the basic, chemical and enzymatic processes underlying the replication of DNA, the synthesis of messenger RNA, and the biosynthesis of protein, have given us deep insight into the molecular mechanisms of heredity.

There is evidence that the genetic code is of the non-overlapping type[36]. For correct transcription the message must be read through from one end of the messenger to the other. For a triplet code the coding ratio, i.e. the ratio of nucleotide residues of the polynucleotide template to amino acid residues of the polypeptide synthesized, should have a value of 3. Direct determination of this ratio is barred at present due to degradation of polynucleotides and polypeptides by nucleases and peptidases present in the crude cell fractions available for protein synthesis studies. As already pointed out these enzymes also interfere with the experimental determination of base sequence of the coding units. It thus appears that the preparation of purified systems of protein synthesis will be a prerequisite for successful experimental attack of some of the most important problems still awaiting solution in the study of the genetic code[53].

Zusammenfassung. Die Nukleotidreihenfolge der DNS, des Trägers der genetischen Information, wird in der Zelle in die komplementäre Nukleotidreihenfolge der «messenger» RNS übertragen. Letztere wirkt als Matrize für die Synthese von Proteinen, deren primäre Struktur, das heisst deren Aminosäuresequenz von der Nukleotidreihenfolge der «messenger» RNS bestimmt wird. Auf diese Weise wird die Sprache der Nukleinsäuren, die 4 Buchstaben (die 4 Nukleotidbasen) umfasst, in diejenige der Proteine, welche aus 20 Buchstaben (den 20 Aminosäuren) besteht, übersetzt. Es wird angenommen, dass je drei aufeinanderfolgende Nukleotidbasen der Matrizen-RNS eine Aminosäure bestimmen, dass also der genetische Code ein «Triplettcode» sei.

Polyribonukleotide, die man mittels des Enzyms Polynukleotidphosphorylase synthetisch aus den Ribonukleosid-5′-diphosphaten leicht darstellen kann, können natürliche Matrizen-RNS im zellfreien proteinsynthetisierenden System ersetzen und den Einbau von Aminosäuren in Polypeptidketten leiten. Die Tatsache, dass die Art der eingebauten Aminosäuren von der Basenzusammensetzung der synthetischen Polynukleotide abhängt, hat zur Entschlüsselung der Basenzusammensetzung einer Anzahl der den verschiedenen Aminosäuren entsprechenden Tripletts geführt. Im *Escherichia coli*-System bewirken Homopolynukleotide wie Polyadenylsäure, Polycytidylsäure und Polyuridylsäure den Einbau von Lysin, Prolin und Phenylalanin. Unter diesen Bedingungen werden Polylysin, resp. Polyprolin und Polyphenylalanin gebildet. Copolynukleotide, die zwei oder mehr verschiedene Nukleotidbasen enthalten, bestimmen die Synthese von Polypeptiden, die aus mehreren verschiedenen Aminosäuren bestehen; zum Beispiel bedingt Polyadenyl-cytidylsäure die Synthese von lysin-, isoleucin-, asparagin-, leucin-, tyrosin- und phenylalaninhaltigen Peptiden.

Durch Korrelation der Basenzusammensetzung einzelner synthetischer Polynukleotide, die dem zellfreien *E. coli*-System zugesetzt wurden, mit der Art und Menge der Aminosäuren, deren Einbau sie bedingten, gelang es bisher die Basenzusammensetzung von 46 von 68 möglichen Tripletts der Matrizen-RNS festzulegen, die den 20 Aminosäuren der Proteinen entsprechen. Wahrscheinlich sind noch nicht alle «sinnvollen» Tripletts bekannt. Aus der Zahl der bis jetzt bekannten Codetripletts folgt, dass in manchen Fällen eine gewisse Aminosäure von mehreren Tripletts bestimmt wird; man spricht hier von einem «degenerierten» Code. Mit wenigen Ausnahmen ist bisher nur die Basenzusammensetzung, nicht aber die Folge der Basen in den Tripletts bekannt. Es gibt Hinweise dafür, dass der genetische Code universell ist, das heisst dass es nur einen Code für alle Lebewesen gibt.

[53] *Acknowledgments.* Various phases of the work at New York University reported in this paper were carried out by Drs. P. LENGYEL, J. F. SPEYER, C. BASILIO, A. J. WAHBA, W. SZER, Y. KAZIRO, and A. GROSSMAN, and Messrs. R. S. GARDNER, J. J. PROTASS, and R. S. MILLER, with the technical help of M. C. SCHNEIDER and H. LOZINA. We are grateful to Drs. R. C. WARNER and C. GILVARG for many helpful discussions and suggestions. The work was aided by grants AM-01845-07 and 1-S01-FR-05099-01 from the National Institutes of Health, U.S. Public Health Service and by fellowships of the U.S. Public Health Service (C.B.), and the Jane Coffin Childs Fund for Medical Research (A.J.W., W.S.).