Reprinted from the Proceedings of the NATIONAL ACADEMY OF SCIENCES
Vol. 48, No. 4, pp. 666–677. April, 1962.

*CHARACTERISTICS AND COMPOSITION OF RNA CODING UNITS**

BY J. HEINRICH MATTHAEI,† OLIVER W. JONES, ROBERT G. MARTIN, AND MARSHALL W. NIRENBERG

NATIONAL INSTITUTE OF ARTHRITIS AND METABOLIC DISEASES, BETHESDA

Communicated by Richard Roberts, February 27, 1962

The translation of a four-letter nucleotide code into a twenty-"word" amino acid dictionary has been the subject of much speculation. Although DNA and RNA polymers were inferentially involved in determining amino acid sequence, a cell-free protein synthesizing system dependent upon these polymers was not available. We have recently described such a system which is dependent upon the addition of template RNA.[1–3] This system affords a sensitive assay for both naturally occurring and synthetic template RNA.

Polyuridylic acid directed the synthesis of polyphenylalanine; thus, one or more uridylic acid residues in poly U appeared to be the coding unit corresponding to phenylalanine.[1, 2] Phenylalanine linked to soluble RNA is an intermediate in this process.[4] The finding that synthetic polyribonucleotides of known composition could be used to direct cell-free protein synthesis, suggested a reasonable experimental approach for establishing the characteristics of the genetic code, an approach which has been utilized by ourselves[5] and by others.[6, 7]

The purpose of this communication is to report further results concerning the influence of poly U in directing polyphenylalanine synthesis and the effects of randomly ordered copolymers upon the incorporation of other amino acids into protein. Some of these results have been reported in a preliminary communication.[5]

Methods and Materials.—The preparation of stable *E. coli* enzyme extracts (DNAase-treated, preincubated S-30 fractions) has been described.[2] Such extracts were dialyzed and stored under liquid nitrogen after preincubation. Reaction mixtures used in determining C^{14}-amino acid incorporation into protein contained the following components: 0.1 *M* tris(hydroxymethyl)-

aminomethane, pH 7.8; 0.01 M magnesium acetate; 0.05 M KCl; 6×10^{-3} M mercaptoethanol; 1×10^{-3} M ATP; 5×10^{-3} M potassium phosphoenolpyruvate; 20 μg/ml of crystalline phosphoenolpyruvate kinase (California Biochem. Corp.); $0.8 - 1.6 \times 10^{-4}$ M C^{14}-amino acids; 2×10^{-4} M each of 19 L-amino acids minus the C^{14}-amino acid; 3×10^{-5} M each of GTP, CTP, and UTP (except where stated); and *E. coli* extracts. Total volume was 0.50 ml except where specified. All assays were performed in duplicate. Techniques used in washing, plating, and counting protein precipitates have been reported.[2] Protein analyses were performed by a micro modification of the method of Lowry *et al.*[8] The lithium and sodium salts of nucleotide diphosphates were obtained from Schwarz Biochemical Corp. and Sigma Chemical Co., respectively. U-C^{14}-amino acids were obtained from Nuclear-Chicago Corp.

The purity of each C^{14}-amino acid was checked by high voltage electrophoresis followed by radioautography.[9] C^{14}-isoleucine was contaminated with C^{14}-leucine; therefore, C^{14}-isoleucine was purified electrophoretically before use. C^{14}-methionine was contaminated with C^{14}-methionine sulfoxide. All of the other C^{14}-amino acids were found to be free of C^{14}-contaminants. S^{35}-cysteine was reduced before use either electrolytically or with mercaptoethanol. The radioactive amino acids used, their source, and their respective specific radioactivities are as follows: U-C^{14}-glycine, U-C^{14}-L-isoleucine, U-C^{14}-L-tyrosine, U-C^{14}-L-leucine, U-C^{14}-L-proline, L-histidine-2(ring)-C^{14}, U-C^{14}-L-phenylalanine, U-C^{14}-L-threonine, L-methionine (methyl-C^{14}), U-C^{14}-L-arginine, and U-C^{14}-L-lysine obtained from Nuclear-Chicago Corporation, 5.8, 6.2, 5.95, 6.25, 10.5, 3.96, 10.3, 3.9, 6.5, 5.8, 8.3 mC/mM, respectively; C^{14}-L-aspartic acid, C^{14}-L-glutamic acid, C^{14}-L-alanine, obtained from Volk, 1.04, 1.18, 0.75 mC/mM, respectively; D-L-tryptophan-C^{14}, obtained from New England Nuclear Corporation, 2.5 mC/mM; S^{35}-L-cystine obtained from the Abbott Laboratories, 2.4 mC/mM; U-C^{14}-L-serine obtained from the Nuclear-Chicago Corporation, 0.2 mC/mM. The specific radioactivities of C^{14}-phenylalanine, valine, and leucine reported by Nuclear-Chicago Corp. were validated by assay. Amino acid concentration was determined by the ninhydrin method.[10] These assays agreed well with the commercially reported data. In a few instances the concentration of C^{14}-phenylalanine also was assayed by measurement of the absorption of the enol-borate complex of phenylpyruvic acid generated enzymatically from phenylalanine by L-amino acid oxidase.[11]

Polyribonucleotides were synthesized enzymatically with *Micrococcus lysodeikticus* polynucleotide phosphorylase purified by the method of Singer and Guss.[12] Reaction mixtures contained 0.15 M tris(hydroxymethyl)aminomethane pH 9.0; 0.01 M $MgCl_2$; 4×10^{-4} M ethylenediaminetetraacetate; 0.06 M nucleotide diphosphate; and polynucleotide phosphorylase purified through the acid ammonium sulfate fractionation step.[12] Approximately 1.4 units of polynucleotide phosphorylase were added per ml of reaction mixture. Reaction mixtures were incubated at 37° and the formation of polynucleotide was followed by determining phosphate liberation by the Fiske-SubbaRow method.[13] After the reaction had proceeded to 20–35% completion, the polynucleotides were precipitated by the addition of 3 volumes of cold absolute ethanol plus a few drops of M NaCl and the precipitates were collected by centrifugation. The polynucleotides were redissolved in H_2O and were reprecipitated as before. The pellet was dissolved in H_2O and the solution was deproteinized three times by the method of Sevag.[14] The aqueous phases obtained after deproteinization were combined and solid KCl was added to a final concentration of 0.025 M. The solution was dialyzed against 0.025 M KCl for 24 hr and against H_2O for an additional 48 hr. The polynucleotide solutions were then lyophilized and were stored at −15°. Some polynucleotides were the generous gifts of Drs. Leon Heppel, Maxine Singer, Daniel Bradley, David Davies, and Robert Steiner.

The base-ratio of each polynucleotide was determined by acid hydrolysis.[15] The bases were separated by descending chromatography on Whatman No. 1 paper for 18 hr using isopropanol-concentrated HCl-H_2O (130:33:37).

The base-ratio analyses obtained by acid hydrolysis were verified in some cases by comparison with base-ratio data obtained by alkaline hydrolysis[16] and by determination of phosphate.[17] Good agreement was found between base-ratio analyses obtained by the three methods.

Results.—Stoichiometry: In Table 1 the mμmoles of C^{14}-phenylalanine incorporated into protein are compared with the mμmoles of uridylic acid residues (pU) in poly U added to a reaction mixture. Limiting concentrations of poly U were

TABLE 1

RELATIONSHIP BETWEEN C^{14}-L-PHENYLALANINE INCORPORATED AND POLYURIDYLIC ACID PRESENT

Experiment no.	mμmole pU (in polyuridylic acid)	mμmole C^{14}-L-phenylalanine	mμmole pU / mμmole C^{14}-phenylalanine
1	11.35	9.12	1.24
2	17.7	10.25	1.73
3	25.7	16.9	1.52

The components of the reaction mixtures are presented under *Methods and Materials*. In addition, each reaction mixture contained 80 mμmole of C^{14}-L-phenylalanine with a specific activity of 1 mC per mM; 2.6 mg preincubated S-30 protein and the amounts of polyuridylic acid specified. In each experiment the amount of polyuridylic acid present was limiting. Final volume was 0.5 ml. Reaction mixtures were incubated for 60 minutes at 37°.

added and incubations were continued until phenylalanine incorporation had ceased. Approximately 1 mμmole of C^{14}-phenylalanine was incorporated into protein per 1.5 mμmole of uridylic acid residue in poly U. These data represent the lowest ratios obtained. The results of many similar experiments demonstrated that widely different stoichiometry ratios could be obtained depending upon the enzyme and poly U preparations used. In some experiments more than 50 mμmoles of uridylic acid residues in poly U were required to direct the incorporation of 1 mμmole of C^{14}-phenylalanine into protein. It should be noted that the presence of trace amounts of RNAase in the system, the molecular weight of poly U, etc., are variables which affect the incorporation of phenylalanine. For these reasons, it is likely that smaller ratios than those presented in Table 1 will be obtained. Since it is not known whether one molecule of poly U directs the synthesis of one or many molecules of polyphenylalanine, these data alone cannot be used to determine the number of uridylic acid residues in one phenylalanine coding unit (coding ratio).

Effect of Molecular Weight upon Template Activity of Poly U.—Poly U was separated into fractions of different molecular weights by sucrose density-gradient centrifugation.[18] Sixteen fractions were collected and both the absorbancy at 260 millimicrons of each fraction and its ability to stimulate C^{14}-phenylalanine incorpora-

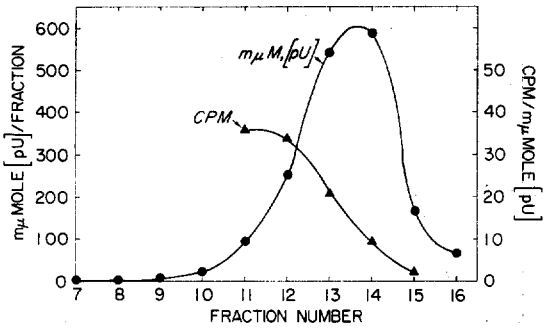

FIG. 1.—The relationship between the molecular weight of polyuridylic acid and its activity in stimulating C^{14}-phenylalanine into protein. Polyuridylic acid was separated into fractions of different sizes by sucrose density-gradient centrifugation. A linear gradient of sucrose concentration ranging from 16% at the bottom to 5% at the top of the tube was prepared. The sucrose solutions (4.4 ml total volume) contained 5×10^{-3} *M* imidazole, pH 6.8 and 0.1 *M* NaCl. 0.5 mg polyuridylic acid in 0.2 ml was layered on top of the tube which was centrifuged at 39,000 × *g* for 4.0 hours at 3°C in a swinging bucket rotor, Spinco type SW-39, using a Spinco Model L ultracentrifuge. Sixteen fractions containing 0.29 ml each were collected after piercing the bottom of the tube. Aliquots containing 25 mμmoles of uridylic acid residue in polyuridylic acid were used for C^{14}-phenylalanine incorporation assays. This amount of polyuridylic acid was limiting under the conditions of the assay. The components of each reaction mixture is presented under *Methods and Materials*. Total volume of each reaction mixture was 0.5 ml and incubations continued for 60 min at 37°.

tion into protein were determined (Fig. 1). Equal amounts of poly U were added to each reaction mixture and incubations were continued until the reactions had stopped. The concentration of poly U in reaction mixtures limited the rate of the reactions. The total activity of each fraction of poly U thus was determined. Poly U molecules of higher molecular weight were distributed toward the bottom of the tube; the lower molecular weight molecules were nearer the top of the tube. It can be seen that fractions of poly U of higher molecular weight were more active in directing polyphenylalanine into protein than lower molecular weight fractions. Although the biologic activity of poly U is related to its molecular weight, the

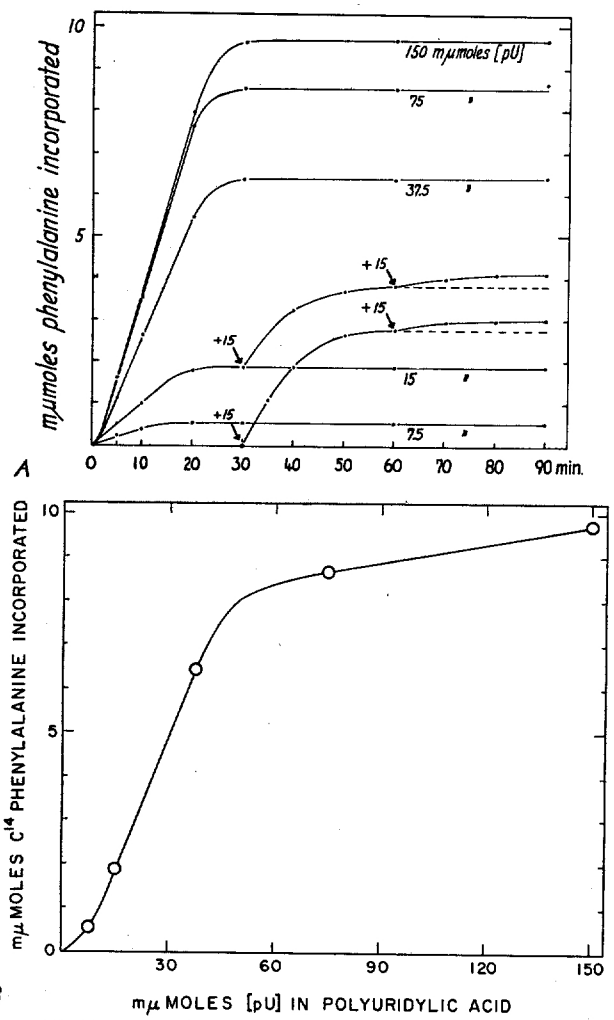

FIG. 2.—Kinetics of C^{14}-L-phenylalanine incorporation. (*A*) C^{14}-phenylalanine incorporation plotted as a function of time. The components of reaction mixtures are described under *Methods and Materials*. In (*A*) the mμmoles of uridylic acid residues (pU) in polyuridylic acid added to 0.5 ml of reaction mixture is shown. Each point represents a 0.5 ml reaction mixture containing 2.6 mg of preincubated S-30 protein. In (*B*) total C^{14}-phenylalanine incorporation (each final 90-minute analysis shown in (*A*)) is plotted as a function of polyuridylic acid concentration.

minimum molecular weight of poly U active as informational RNA is not known and experiments designed to answer this question are in progress.

Kinetics of Phenylalanine Incorporation.—In Figure 2 the rate of phenylalanine incorporation into protein is shown. From 7.5 to 150 mμmoles uridylic acid residue (pU) in poly U were added per 0.5 ml reaction mixture. C^{14}-phenylalanine incorporation ceased within 30 minutes after addition of poly U. Further addition of poly U after incorporation had stopped resulted in an additional stimulation of incorporation. After the reactions had again stopped, addition of more poly U did not appreciably stimulate C^{14}-phenylalanine incorporation. When the mixtures were incubated for 30 minutes in the absence of poly U, and at that time 15 mμmoles of pU were added, a somewhat greater incorporation of C^{14}-phenylalanine was observed compared with incorporation resulting from the addition of poly U at the onset of incubation. These experiments demonstrate that even though C^{14}-phenylalanine incorporation stops after 20–30 minutes of incubation, the enzyme extracts are fully active. The fact that C^{14}-phenylalanine incorporation stops after 30 minutes even though a large excess of poly U is present (150 mμmoles pU) shows that poly U is being inactivated. Such inactivation of excess poly U does not appear to be dependent upon protein synthesis.

The data of Figure 2 *B* show that phenylalanine incorporation is proportional to poly U concentration in the range of 10–50 mμmoles of pU in poly U per 0.5 ml reaction mixture. In the linear part of the curve about 6 mμmoles of pU were required to direct the incorporation of 1 mμmole of C^{14}-phenylalanine.

Stimulation of Amino Acid Incorporation by Polynucleotides Containing One Base.—The effect of "homopolynucleotides" upon amino acid incorporation into protein is presented in Table 2. As has been reported previously[12] poly U specifically stimulated phenylalanine incorporation into protein. Poly U also stimulated the incorporation of small amounts of C^{14}-leucine and C^{14}-valine into protein.

TABLE 2

SPECIFICITY OF AMINO ACID INCORPORATION STIMULATED BY HOMOPOLYNUCLEOTIDES AND YEAST RNA

C^{14}-L-amino acid	Minus polynucleotide	poly A	poly C	poly U	Yeast RNA
Alanine	27	18	21	26	82
Arginine	19	19	18	25	204
Aspartic acid	23	23	18	28	177
Cysteine	109	111	114	113	133
Glutamic acid	53	53	50	53	204
Glycine	32	31	28	31	284
Histidine	5	4	5	3	312
Isoleucine	50	22	55	74	229
Leucine	77	71	43	263	178
Lysine	14	22	11	26	129
Methionine	29	35	31	40	83
Phenylalanine	24	24	22	5424	128
Proline	22	20	336	27	118
Serine	250	250	250	281	687
Threonine	26	23	28	28	110
Tryptophan	245	239	238	262	514
Tyrosine	38	38	37	42	83
Valine	11	11	11	47	67

The figures represent the incorporation of C^{14}-amino acids in μμmoles. The components of reaction mixtures are presented under *Methods and Materials*. GTP, CTP, and UTP were omitted. Each 0.6 ml reaction mixture contained 80 mμmole of the appropriate C^{14}-amino acid (in addition to 19 C^{12}-amino acids); 2.6 mg preincubated S-30 protein; and either 100, 20, 10, or 500 μg of poly A, poly C, poly U or yeast RNA (prepared by the method of Crestifield *et al.*[14]) respectively. Samples were incubated for 60 minutes at 37°.

Base-ratio analyses (see *Methods and Materials*) showed that this polymer contained 2.5% guanylic acid present as an impurity. Thus, the small stimulation of leucine and valine incorporation by poly U probably is due to the presence of guanylic acid (see next section).

Poly C stimulated the incorporation of proline.[2] However, the effectiveness of poly C in stimulating proline incorporation varied with different preparations of poly C; for example, many preparations stimulated 5- to 10-fold and three preparations stimulated 75- to 100-fold. Since the effectiveness of poly C in stimulating proline incorporation was not fully reproducible, the purity of several preparations was examined (see *Methods and Materials* section). These preparations contained 4–14 per cent uridylic acid. Commerical preparations of CDP were analyzed and were found to contain no UDP. As will be seen in the next section, randomly-mixed poly UC markedly stimulated proline incorporation, therefore, the observed activity of poly C may result from the presence of small amounts of U in the polynucleotide preparations.

Poly A did not stimulate the incorporation of any amino acid into protein. The addition of smaller concentrations of poly A also gave similar results. The molecular weight of the poly A was approximately 30,000. In contrast, poly U of similar molecular weight markedly stimulated phenylalanine incorporation. The ineffectiveness of poly A in coding for any amino acid could be due to either its double-stranded structure in solution or to the possibility that a sequence of A does not specify any amino acid.

Previously, it was found that yeast ribosomal RNA prepared by the method of Crestfield *et al.*,[19] stimulated amino acid incorporation in this system.[2] Yeast RNA was used as a control to show that the system was active with respect to every amino acid. These experiments also demonstrate that the addition of naturally occurring template RNA stimulated the incorporation of every amino acid in contrast to the specificity displayed by poly U.

Since poly G is difficult to prepare enzymatically, it was not available to us in sufficient quantities to test with each individual amino acid. However, when poly G was tested with a C^{14}-algal protein hydrolysate which contained 16 C^{14}-amino acids, it did not stimulate the incorporation of any C^{14}-amino acid. An oligonucleotide primer (tetraadenylic acid) was required for the enzymic synthesis of poly G and primarily low molecular weight polynucleotides having an average chain length of 15 nucleotides were obtained. Therefore, the fact that poly G was unable to stimulate amino acid incorporation should be interpreted with caution.

The Effects of Randomly Ordered Polynucleotides upon Amino Acid Incorporation.—The data of Table 3 demonstrate that randomly ordered polyribonucleotides direct the incorporation of amino acids into protein in a highly specific manner. Since each preparation of polynucleotide will differ in molecular weight, it is difficult to compare directly the incorporation of activity of one polymer with another. Therefore, the figures in Table 3 represent the per cent of any amino acid incorporated compared to phenylalanine incorporation stimulated by the same polynucleotide. The counts/minute of C^{14}-L-phenylalanine incorporated due to the addition of each polynucleotide is presented in the legend accompanying Table 3; therefore, the percentage figures in the table readily may be converted into counts/minute. The

base-ratio of each polynucleotide, determined experimentally, is given also. If a polynucleotide containing two bases stimulated incorporation of an amino acid, inclusion of a third base in the polynucleotide did not prevent this stimulation.

Poly UG and poly UGC previously were found to direct the incorporation of small amounts of methionine and glutamic acid respectively.[5] Further experiments have shown that poly UGA codes for these amino acids more effectively; thus, the coding units for methionine and glutamic acid contain U, G and A. Lysine incorporation was stimulated by the addition of poly UA ($^1/_4$ ratio) but not by randomly mixed poly UA containing lower proportions of A. The coding unit for lysine therefore probably contains UAA.... The stimulation of phenylalanine incorporation by poly UA ($^1/_4$ ratio) was negligible, so these data have not been expressed in Table 3 as the ratio of lysine to phenylalanine incorporation.

A series of polynucleotides each containing a different ratio of U to A were prepared. Similar series of polynucleotides containing either U and C or U and G were synthesized also. It was found that stimulation of incorporation of a given amino acid by a polynycleotide varied with the base-ratio of the polymer.[22] Thus the relative amounts of two nucleotides in a coding unit could be estimated. These data are too extensive to be reported here and will be the subject of a future communication.

The coding ratio is not known definitively. However, assuming a triplet code, the probability of a triplet occurring in a polynucleotide, relative to UUU, may be calculated from the base-ratio data presented in Table 3. For example, if poly UG had a base-ratio of 3U/1G, the probability of obtaining the sequence UUU would be $^3/_4 \times {}^3/_4 \times {}^3/_4 = {}^{27}/_{64}$. The probability of obtaining the sequence UUG would be $^3/_4 \times {}^3/_4 \times {}^1/_4 = {}^9/_{64}$. Thus, 3 UUU would occur for 1 UUG and, assuming the frequency of UUU to be 100 per cent, the frequency of UUG would be 33 per cent. The theoretical frequency of each possible triplet, relative to UUU is presented in Table 3. Amino acid incorporation agreed with predictions based upon probability theory in most cases. The nucleotide composition of coding units corresponding to each amino acid may be derived from these data and a summary is presented in the last column of Table 3. Dots after each "word" are used to indicate the possible presence of additional uridylic acid residues.

Discussion.—Since the coding units of arginine, alanine, and glutamic acid each contain three different nucleotides, the minimum coding ratio would appear to be three. Poly GC does not code for arginine or alanine and poly AG does not code for glutamic acid.[22] Possibly the coding ratio is larger, but these data rule out the possibility of singlet and doublet codes. In this preliminary study we will assume that every amino acid has the same coding ratio[20] and that in this system, as in cellular systems,[21] an overlapping triplet code is improbable.

It is important to determine whether one molecule of poly U directs the synthesis of either one or several molecules of polyphenylalanine, i.e., whether poly U functions stoichiometrically or catalytically. If, for example, each molecule of poly U directed the synthesis of only one molecule of polyphenylalanine and a nonoverlapping sextuplet code were operative, six uridylic acid residues in poly U would be necessary to mediate the incorporation of one phenylalanine. Instead, the data of Table 1 show that almost one mμmole of phenylalanine was incorporated into

TABLE 3

AMINO ACID INCORPORATION INTO PROTEIN STIMULATED BY RANDOMLY MIXED POLYNUCLEOTIDES

Polynucleotide	UA	UC	UG	UAC	UGC	UGA	Composition of coding units*
Base-ratio	U = 0.87 A = 0.13	U = 0.39 C = 0.61	U = 0.76 G = 0.24	U = 0.834 A = 0.050 C = 0.116	U = 0.341 G = 0.152 C = 0.502	U = 0.675 G = 0.291 A = 0.034	
Probability of triplet relative to phenylalanine (UUU) = 100%	UUU—100 UUA— 13 UAA— 2.2 AAA— 0.3	UUU—100 UUC—157 UCC—244 CCC—382	UUU—100 UUG— 32 UGG— 10.6 GGG— 3.4	UUU—100 UUA— 6.0 UAA— 0.4 AAA— 0.02 UUC— 13.9 UCC— 1.9 CCC— 0.3 UAC— 0.8 AAC— 0.05 ACC— 0.12	UUU—100 UUG— 46.2 UGG— 21.0 GGG— 1.0 UUC—147 UCC—218 CCC—322 UGC— 68.1 GGC— 31.7 GCC—101	UUU—100 UUG— 43 UGG— 19 GGG— 8.1 UUA— 5.1 UAA— 0.26 AAA— 0.01 UGA— 2.2 GGA— 0.1 GAA— 0.01	
Amino Acid							
Phenylalanine	100	100	100	100	100	100	UUU...
Arginine	0	0	1.1	0	49.3	2.9	UCG...
Alanine	1.9	0	0	1.0	40.4	0.9	UCG...
Serine	0.4	160	3.2	3.6	170	2.3	UUC.. + UCG...
Proline	0	285	0	0	188	0	UCC...
Tyrosine	13	0	0	8	1.0	8.6	UUA...
Isoleucine	12	1.0	1.0	4.8	5.4	8.4	UUA...
Valine	0.6	0	37	0.4	29.8	75	UUG...
Leucine	4.9	79	36	5.1	157	44	UUC.. + UUG...
Cysteine	4.9	0	35	0	5.4	46	UUG... or UGG
Tryptophan	1.1	0	14	0	1.6	23	UGG...
Glycine	4.7	0	12	0.5	9.7	15	UGG...
Methionine	0.6	0	0	0.6	1.5	8	UGA...
Glutamic acid	1.5	0	0	1.2	0.44	6.2	UGA...
Lysine							UAA... (?)

The figures in the main part of the table represent the incorporation of any amino acid compared to phenylalanine incorporation expressed as percentages (mμmoles amino acid incorporated/mμmoles phenylalanine incorporated × 100). Underlined figures refer to the polynucleotide containing the nucleotides necessary to stimulate the incorporation of a given amino acid. The components of the reaction mixtures are presented under *Methods and Materials*. 0.1 μmole each of 19 L-amino acids minus the appropriate C^{14}-amino acid, 0.075 μmole of the C^{14}-amino acid, and approximately 25 μg of each polynucleotide were added to each 0.5 ml reaction mixture. Samples were incubated at 37° for 15 minutes. Incorporation of C^{14}-phenylalanine in counts per minute due to the addition of polynucleotides UA, UC, UG, UAC, UGC, and UGA were 731, 2900, 714, 804, 2144, and 2744, respectively. The reproducibility of the above percentage figures was ±3.

* Sequence of nucleotides in a coding unit is not specified.

protein per mμmole of uridylic acid residue in poly U. Therefore, each phenylalanine coding unit appears to function catalytically for a limited time.

A degenerate code is one in which two or more different coding units can direct the incorporation of the same amino acid into protein. In a completely degenerate code, every permutation of nucleotides would code for an amino acid. In a partially degenerate code, certain coding units would direct amino acid incorporation, whereas others would not. Coding units which would not direct amino acids into protein will be called "nonsense words." Up to now it has not been possible to determine directly whether the code contains nonsense units.

The data of Table 2 demonstrate that poly A does not direct the incorporation of any amino acid into protein. Other polynucleotides such as poly AG also do not direct incorporation of amino acids.[22] Such experiments indicate that nonsense words exist and thus would preclude the possibility of a completely degenerate code. Nonsense regions in template RNA may be functionally important, for it is possible that these nucleotide sequences serve as periods, i.e., may specify C- or N-terminal groups in proteins.

The data of Table 3 demonstrate that a coding unit corresponding to leucine can contain either U and C or U and G. Since two words containing different nucleotides correspond to leucine, the code is partially degenerate when synthetic polynucleotides are used to direct amino acid incorporation.

Whereas the U content of RNA viruses is not excessive, a surprisingly high proportion of U has been found in coding units thus far (Table 3). This dichotomy cannot be explained at the present time. However, it is probable that additional degenerate code words will be found which will not contain U. Trichloracetic acid insoluble peptides only have been measured in this study. Since polyphenylalanine becomes insoluble when four or five phenylalanine residues are linked together, such an insoluble "handle" may ensure the precipitation of a polypeptide which otherwise would be soluble. If a code with much degeneracy is assumed, it is possible that coding units containing U have been selected for by the assay method. However, other explanations are possible and are being considered.

The foregoing data demonstrate that the code is partially degenerate, the minimum coding ratio is three and that nonsense coding units exist. Crick *et al.*, on the basis of ingenious genetic experiments employing the rII region in T_4 phage, have come to similar conclusions.[23] It is important to note that such information, obtained by means of both biochemical and genetic approaches, is in complete accord with one another.

The treatment of tobacco mosaic virus (TMV) RNA with nitrous acid results in the deamination of nucleotides[24] and the formation of mutant strains of TMV with altered plaque morphology.[25] Cytidine is converted to uridine, adenosine to hypoxanthine and guanosine to xanthosine. The recent striking work of Wittmann[28] at Tübingen and Tsugita and Fraenkel-Conrat[26, 27] at Berkeley has shown that the protein isolated from such mutant strains is, in many cases, different from the protein produced by wild type TMV. The amino acid sequences of both wild type and mutant TMV proteins were determined and certain amino acids were found to be replaced by others rather frequently. A comparison has been made in Table 4 between nitrous acid-induced amino acid substitutions and the nucleotide compositions of corresponding coding units. Only amino acid substitutions which have

TABLE 4

COMPARISON OF NITROUS ACID INDUCED REPLACEMENTS IN TOBACCO MOSAIC VIRUS PROTEIN WITH THE NUCLEOTIDE COMPOSITION OF RNA CODING UNITS

No. of mutant strains A	No. of mutant strains B*	Amino acid replacement	Nucleotide composition of corresponding coding units	Possible nucleotide changes
1	2	Ser	UUC...	C
		↓	↓	↓
		Phe	UUU...	U
1	2	Glu	UAG...	A
		↓	↓	↓
		Gly	UGG...	G
1	2	Prol	UCC...	C
		↓	↓	↓
		Leu	UUC...	U
	2	Isoleu	UUA...	A
		↓	↓	↓
		Val	UUG...	G
5		Arg	UCG...	C
		↓	↓	↓
		Gly	UGG...	G

* The mutant amino acid replacement data cited and obtained either by (*A*) Tsugita and Fraenkel-Conrat[26, 27] or (*B*) Wittmann.[28]

been found more than once have been cited. Thus, the chance of a substitution occurring spontaneously rather than by deamination is lessened. Only those amino acid substitutions are included which correspond to RNA coding units with experimentally determined nucleotide compositions. We are indebted to Drs. Tsugita and Fraenkel-Conrat for allowing us to cite some of their unpublished replacement data.

Replacement of serine by phenylalanine, glutamic acid by glycine, proline by leucine, glutamic acid by valine, and arginine by glycine have been observed. Most of the nucleotide conversions corresponding to the amino acid replacements thus would be the conversion of either cytidine to uridine or of adenine to guanine.

As has been predicted by Freese,[29] a fairly high proportion of nitrous acid induced amino acid substitutions should be the result of conversion of cytidine to uridine in a coding unit. Wittmann has also shown that nitrous acid will convert adenine to guanine.[28] Deamination of guanosine has no mutagenic effect. Comparison of amino acid substitution data with the proposed nucleotide compositions of coding units shown in Table 4 strikingly confirm these predictions.

Previously we have shown that sRNA is an intermediate in polyphenylalanine synthesis.[4] It is likely that amino acids lose their identity after being linked to sRNA and that aminoacyl-sRNA molecules recognize template RNA coding units by base-pairing. During the course of evolution, mutations affecting the nucleotide sequence of sRNA or the specificity of amino acid activating enzymes might occur and might result in a series of phylogenetically related codes rather than one code universal to all species.

From an evolutionary point of view, a mutation resulting in a changed code wherein one amino acid substitutes for another in all proteins probably would be lethal to a highly organized cellular system. The use of synthetic polynucleotides affords a unique opportunity to determine whether the code is universal. The data of Table 4 demonstrate that both TMV coding units directing protein synthesis in tobacco plants, and coding units functioning in *E. coli* have similar nucleotide compositions. It is not improbable that minor changes in the code will be found in

different species; however, these data strongly suggest that at least part of the code may be universal.

Summary.—The activity of polyuridylic acid in directing polyphenylalanine synthesis was found to vary with the molecular weight of polyuridylic acid; longer polynucleotide chains were more active than shorter ones. Evidence was presented which suggested that one molecule of polyuridylic acid may direct the synthesis of a number of molecules of polyphenylalanine. Randomly mixed polynucleotides as well as "homopolynucleotides" were used to direct cell-free amino acid incorporation. Using this technique the nucleotide compositions of RNA coding units corresponding to 15 amino acids were determined. Characteristics of the code such as degeneracy, the existence of nonsense words, the minimum coding ratio, and the universality of the code were investigated. Two coding units corresponding to leucine were found; thus, part of the code was shown to be degenerate. Certain nucleotide sequences in a polynucleotide did not code for any amino acid; therefore, the presence of nonsense coding units was suggested. The *minimum* number of nucleotides per coding units appeared to be three. Comparison between the composition of RNA coding units in *E. coli* and amino acid replacement data in tobacco mosaic virus suggested that at least part of the code may be universal.

The authors wish to express their appreciation to Mrs. Linda Greenhouse for valuable technical assistance. It is also a pleasure to express our gratitude to Drs. Leon Heppel and Maxine Singer for invaluable advice and helpful discussions.

* The following abbreviations were used: Poly U, polyuridylic acid; poly A, polyadenylic acid; poly C, polycytidylic acid; poly G, polyguanylic acid; poly UA, polyuridylic-adenylic acid; poly UC, polyuridylic-cytidylic acid; poly UG, polyuridylic-guanylic acid; poly UAC, polyuridylic-adenylic-cytidylic acid; poly UCG, polyuridylic-cytidylic-guanylic acid; poly UGA, polyuridylic-guanylic-adenylic acid; ADP, adenosine diphosphate; UDP, uridine diphosphate; GDP; guanosine diphosphate; CDP, cytidine diphosphate, pU, uridylic acid residue in poly U; TMV, tobacco mosaic virus.

† NATO Postdoctoral Research Fellow.

[1] Nirenberg, M. W., and J. H. Matthaei, *V International Congress of Biochemistry*, Moscow (August, 1961).

[2] Nirenberg, M. W., and J. H. Matthaei, these PROCEEDINGS, **47**, 1588 (1961).

[3] Matthaei, J. H., and M. W. Nirenberg, these PROCEEDINGS, **47**, 1580 (1961).

[4] Nirenberg, M. W., J. H. Matthaei, and O. W. Jones, these PROCEEDINGS, **48**, 104 (1962).

[5] Martin, R. G., J. H. Matthaei, O. W. Jones, and M. W. Nirenberg, *Biochem. Biophys., Research Commun.*, **6**, 410 (1962).

[6] Lengyel, P., J. F. Speyer, and S. Ochoa, these PROCEEDINGS, **47**, 1936 (1961).

[7] Speyer, J. F., P. Lengyel, C. Basilio, and S. Ochoa, these PROCEEDINGS, **48**, 63 (1962).

[8] Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall, *J. Biol. Chem.*, **193**, 265 (1951).

[9] Dreyer, W. J., and E. Bynum, in preparation.

[10] Spies, Joseph R., in *Methods in Enzymology*, ed. S. P. Colowick and N. O. Kaplan (New York: Academic Press, 1957), vol. 3, p. 467.

[11] La Du, B., and P. J. Michael, *J. Lab. and Clin. Med.*, **55**, 491 (1960).

[12] Singer, M. F., and J. K. Guss, *J. Biol. Chem.*, **237**, 182 (1962).

[13] Fiske, C. H., and Y. SubbaRow, *ibid.*, **66**, 375 (1925).

[14] Sevag, M. G., D. B. Lackman, and J. Smolens, *ibid.*, **124**, 425 (1938).

[15] Smith, J. D., and R. Markham, *Biochem. J.*, **46**, 509 (1950).

[16] Davidson, J. H., and R. M. S. Smellie, *Biochem. J.*, **52**, 594 (1952).

[17] Ames, B., and D. T. Dubin, *J. Biol. Chem.*, **235**, 769 (1960).

[18] Britten, R. J., and R. B. Roberts, *Science*, **131**, 32 (1960).
[19] Crestfield, A. M., K. C. Smith, and F. W. Allen, *J. Biol. Chem.*, **216**, 185 (1955).
[20] Gamow, G., *Nature*, **173**, 318 (1954).
[21] Brenner, S., these PROCEEDINGS, **43**, 687 (1957).
[22] Nirenberg, M. W., J. H. Matthaei, and O. W. Jones, unpublished data.
[23] Crick, F. H. C., L. Barnett, S. Brenner, and R. J. Watts-Tobin, *Nature*, **192**, 1227 (1961).
[24] Levene, P. A., and R. S. Tipson, *J. Biol. Chem.*, **111**, 313 (1935).
[25] Gierer, A., and K. W. Mundry, *Nature*, **182**, 1457 (1958).
[26] Tsugita, A., and H. Fraenkel-Conrat, *J. Mol. Biol.*, in press.
[27] Tsugita, A., and H. Fraenkel-Conrat, personal communication.
[28] Wittmann, H. G., *Naturwissenschaften*, **48**, 729 (1961).
[29] Freese, E., in *Structure and Function of Genetic Elements*, Brookhaven Symposia in Biology, no. 12 (1959), p. 63.