REPRINTED FROM
SYMPOSIUM ON INFORMATIONAL MACROMOLECULES
©1963
ACADEMIC PRESS, INC., NEW YORK, N. Y.

The Current Status of the RNA Code[1]

MARSHALL W. NIRENBERG AND OLIVER W. JONES, JR.

National Heart Institute, National Institutes of Health, Bethesda, Maryland

Rather than review all of our work concerning the genetic coding problem, only one aspect which we have been investigating (up to September, 1962) will be presented; that is, the extent of degeneracy and its relationship to the general nature of the code. A degenerate genetic code was suggested a number of years ago by Gamow (10) and by Crick (4). In such a code, an amino acid may be directed into protein by two or more codewords. Previous work demonstrated that C^{14}-amino acids were directed into protein by synthetic polynucleotides in cell-free *Escherichia coli* extracts (19) and that leucine incorporation was stimulated by either poly UG,[2] UC, or UA (16, 17, 29). Thus the code was shown to be degenerate with respect to leucine (16, 17, 29).

Initially, all of the codewords found contained U. However, assuming a triplet code, the proportion of U compared with other nucleotides in codewords seemed unusually high, for natural template RNA, such as viral RNA, did not contain such a preponderance of U. To resolve this paradox, a more degenerate code was proposed with both non-U and U containing codewords (17). An alternative hypothesis was advanced by Roberts, who suggested a doublet code; for in such a code the pro-

[1] This report is limited to the data which were presented at the Symposium on Informational Macromolecules in September, 1962. Data obtained after this date are not included.

[2] The following abbreviations are used: poly U, polyuridylic acid; poly A, polyadenylic acid; poly C, polycytidylic acid; poly G, polyguanylic acid; poly UGAC, polyuridylic-guanylic-adenylic-cytidylic acid; poly ACG, polyadenylic-cytidylic-guanylic acid; poly AC, polyadenylic-cytidylic acid; poly CG, polycytidylic-guanylic acid; poly UG, polyuridylic-guanylic acid; poly UC, polyuridylic-cytidylic acid; poly UA, polyuridylic-adenylic acid; poly UCG, polyuridylic-cytidylic-guanylic acid; poly UAG, polyuridylic-adenylic-guanylic acid; G-G, guanylic-guanylic; A, adenylic acid; G, guanylic acid; C, cytidylic acid; U, uridylic acid.

TABLE I
BASE RATIO[a]
(Moles Per Cent)

Designation	Polymer	Input ratio of nucleotides	Base ratio of nucleotides
		U G A C	U G A C
Ap231	UGAC	40:20:20:20	55:32: 5: 8
Ap232	UGAC	58:14:14:14	56:25: 5:13
Ap233	UGAC	20:20:20:40	3:45: 9:43
Ap234	UGAC	12:12:12:64	23:21: 4:52
Ju 258	UGAC	29:13:29:29	27:22:22:29
Ju 2510	UGAC	29:29:29:13	27:43:21: 9
		A C G	A C G
J 251	ACG	60:20:20	46:32:22
M 76	ACG	7:86: 7	2:89: 9
M 75	ACG	10:80:10	4:77:19
M 74	ACG	30:60:10	16:72:12
		A C	A C
J 104	AC	9:91	3:97
J 103	AC	12:88	6:94
J 102	AC	20:80	12:88
J 101	AC	33:67	30:70
J 109	AC	75:25	67:33
J 108	AC	83:17	80:20
		C G	C G
M 141	CG	88:12	90:10
M 71	CG	92: 8	87:13
F 120	CG	88:12	82:18
F 135	CG	50:50	9:91
		A G	A G
J 106	AG	80:20	73:27
J 107	AG	66:33	48:52

[a] Polyribonucleotides were synthesized, as described previously, with the aid of polynucleotide phosphorylase partially purified from *Micrococcus lysodeikticus* according to the method of Singer and Guss (27).

The base-ratio of each polynucleotide preparation was determined by analysis. Polynucleotides were hydrolyzed by incubation in 0.4 *N* KOH at 25° for 18 hours. Under these conditions, little deamination occurred.[3] Such mild conditions were not sufficient to hydrolyze certain polymers; however, in such cases, incubation in 0.3 *N* KOH at 37° for 18 hours resulted in complete hydrolysis (5). Mononucleotide products were separated either by paper electrophoresis (Whatman No. 3 MM paper, 0.05 *M* ammonium formate, pH 3.7) or by descending paper chromatography (Whatman No. 3 MM paper and a solvent system containing 0.1 *M* sodium phosphate, pH 7.0 and 3 *M* ammonium sulfate). Two % or greater contamination of polynucleotides by U would have been detected. No contamination by U was found. Mononucleotides and appropriate blanks were eluted by shaking small paper

[3] We thank Dr. M. Grunberg-Manago for this protocol.

portions of nucleotides would be within the range found in viral RNA (25, 26).

The existence of non-U codewords was suggested when poly AC was found to direct small amounts of proline and threonine into protein (13, 21). Recently, in a careful study, Bretscher and Grunberg-Manago clearly demonstrated coding by non-U words (2). Several poly AC preparations were reported to code well for proline, threonine, histidine and, to a lesser extent, for glutamine. This work indicated that other non-U polynucleotides might have template activities. In this communication, further qualitative analysis of coding by such polynucleotides will be reported.

Results

Base-Ratio Analysis

The synthetic polynucleotides used in this study are listed in Table I. The base-ratio analysis of each polymer is compared with the ratio of nucleoside diphosphates present during the synthesis of each polynucleotide. In many cases, the base-ratio of the polymer product differed slightly from the input ratio of the substrates. In polymers containing two or three different nucleotides, preferential incorporation into polynucleotide of either G or C relative to A was observed. Bretscher and Grunberg-Manago have reported that *Azotobacter* polynucleotide phosphorylase also catalyzes a preferential incorporation of C and G into poly UC and UG (2).

Stimulation of Amino Acid Incorporation by Polynucleotides Containing Four Bases

The data of Table II demonstrate that synthetic polynucleotides containing four bases stimulate the incorporation of a large number of amino acids into protein. In the last column is given the basal level of C^{14}-amino acid incorporation obtained in the absence of polynucleotide; other figures refer to the net increase above basal incorporation due to addition of polynucleotide. The base-ratios of the polynucleotides vary widely. The fifth polynucleotide (Ju-258) contains approximately equal proportions of U, G, A, and C, whereas the other polynucleotides contain predominant amounts of two or three nucleotides. All of the polynucleotides were active in directing amino acid incorporation, except polynucleotide Ju-2510. Although 10 μg of polynucleotide were added to each reaction mixture, the total amount of C^{14}-amino acid directed into protein by each polynucleotide varied more than 50-fold. As we have shown previously, the template activity of polynucleotides is dependent upon factors other than nucleotide sequence. For example,

strips immersed in 0.1 *N* or 0.01 *N* HCl and determining UV absorption at appropriate wavelengths in a Beckman DU spectrophotometer.

TABLE II

STIMULATION OF AMINO ACID INCORPORATION BY POLY UGAC

Polynucleotide:	UGAC	UGAC	UGAC	UGAC	UGAC	UGAC	
Base ratio (moles per cent)	U 55 G 32 A 5 C 8	U 56 C 25 A 5 C 13	U 3 G 45 A 9 C 43	U 23 G 21 A 4 C 52	U 27 G 22 A 22 C 29	U 27 G 43 A 21 C 9	Minus polynucleotide control
Designation:	Ap231	Ap232	Ap233	Ap234	Ju258	Ju2510	
C^{14}-Amino acid	Incorporation above control Δ µµMoles[a]						
Alanine	110	127	62	152	31	5	10
Arginine	69	270	68	212	99	57	11
Aspartic acid ($—NH_2$?)	10	40	9	10	25	12	12
Glutamic acid ($—NH_2$?)	16	52	12	9	14	—	23
Glycine	62	663	25	40	12	6	13
Histidine	20	8	11	24	13	0	4
Isoleucine	68	301	60	90	0	0	22
Leucine	168	1,243	125	418	12	0	41
Lysine	10	21	0	3	25	—	4
Methionine	9	64	0	9	15	4	12
Phenylalanine	152	606	86	64	14	0	10
Proline	50	140	125	1,007	121	10	7
Serine	179	807	181	445	37	0	47
Threonine	15	54	19	78	44	0	7
Tryptophan	23	8	16	8	1	—	45
Tyrosine	14	80	11	12	6	0	17
Valine	100	602	57	70	43	13	7
Total	1,075	5,086	867	2,651	512	107	292

TABLE II (*Continued*)

. μμMoles represents the difference between C^{14}-amino acid incorporation into protein in the presence and absence of poly-
ides. Basal incorporations obtained when polynucleotides were omitted are presented in the last colum (minus poly-
ide).
ıction mixtures used to determine C^{14}-L-amino acid incorporation into protein contained the following components:
Tris (hydroxymethylaminoethane) pH 7.8; 0.01 *M* magnesium acetate; 0.05 *M* KCl; 6×10^{-3} *M* mercaptoethanol; 1 ×
I ATP; 5×10^{-3} *M* potassium phosphoenolpyruvate; 5 μg of crystalline phosphoenolpyruvate kinase (California Bio-
al Corporation); 0.8×10^{-4} *M* C^{14}-amino acid (approximately 30,000-150,000 counts/minute/reaction mixture); 3.2 ×
I each of 19 C^{12}-L-amino acids minus the C^{14}-amino acid; 10 μg of polynucleotide/reaction mixture, when specified; and
: preincubated S-30 extracts (1-2 mg protein/reaction mixture). Total volume of each reaction mixture was 0.3 ml. Re-
mixtures were incubated at 37° for 30 minutes; thus, total amino acid incorporation rather than rate of incorporation was
ed. A Nuclear-Chicago thin-window, gas flow counter was used.

large polymers of chain length greater than 100 units are considerably more active than shorter ones (17). Single-stranded polynucleotides are active, whereas double- or triple-stranded polymers are not (19). In addition, randomly-mixed copolymers which have a high degree of secondary structure are inactive in coding (28). In particular, polymers containing much G have little activity, possibly because of G-G interactions. Thus the relative inactivity of the last poly UGAC preparation (Ju-2510) should not be ascribed necessarily to the presence of a high proportion of nonsense nucleotide sequences. Such considerations make it difficult to compare with validity the relative abilities of different polynucleotides to code for the same amino acid; thus, *such comparisons should be made with caution.* The fact that polynucleotides containing four bases coded so well for so many amino acids strongly suggested that most nucleotide sequences could be read. In addition, a high proportion of U clearly was not required for messenger RNA activity.

Stimulation of Amino Acid Incorporation by Poly ACG

The coding activities of polymers which did not contain U are given in Table III. Base-ratio analyses of each poly ACG preparation failed to detect contamination by U. Poly ACG preparations stimulated the incorporation of many amino acids tested, including alanine, arginine, glutamic acid, lysine, proline, and threonine. Such high incorporations of glutamic acid, lysine, and threonine were not observed previously. A number of amino acids did not appear to be coded by any ACG preparations, which suggested that U may be an absolute requirement in coding for some amino acids. Since the template activities of some poly ACG preparations equaled those of our best synthetic template RNA preparations, U clearly was not required for coding other amino acids.

Stimulation of Amino Acid Incorporation by Polynucleotides Containing Two Bases

The data of Table IV demonstrate stimulation of amino acid incorporation by poly AC preparations. The polynucleotides are listed in order of decreasing C content. In accord with the findings of Bretscher and Grunberg-Manago (2), poly AC stimulated incorporation of proline, threonine, and histidine. In addition, poly AC was found to direct aspartic acid, glutamic acid, and lysine into protein. Bretscher and Grunberg-Manago (2) report that glutamine is coded by such polymers. We have not been able to obtain C^{14}-asparagine or C^{14}-glutamine and, thus, have not been able to study this point.[4] Although the addition of

[4] Recently, we have confirmed the finding of Bretscher and Grunberg-Manago (2) that glutamine rather than glutamic acid is directed into protein by poly CA. In addition, we find that poly CA codes for asparagine rather than aspartic acid.

C^{12}-aspartic acid and C^{12}-glutamic acid to reaction mixtures completely diluted the incorporation of C^{14}-aspartic and C^{14}-glutamic acids, respectively, the possibility of conversion of the free acid to the amide during incubation of reaction mixtures does not allow us to distinguish between the acid and amide forms. Many of the polynucleotides were found to have template activities equal to the most active poly U preparations tested. Poly AC (J-104) contained 97% C, yet actively directed proline into protein. Thus, it appears probable that one codeword for proline may contain only C. Relatively large amounts of lysine were directed into protein by AC (J-109) and (J-108), which contained 67 and 80% A, respectively. These data suggest that a codeword for lysine may contain only A.

TABLE III
STIMULATION OF AMINO ACID INCORPORATION BY POLY ACG

Polynucleotide:	ACG	ACG	ACG	ACG	Minus poly-nucleotide control
Base ratio (moles per cent)	A 46 C 32 G 22	A 2 C 89 G 9	A 4 C 77 G 19	A 16 C 72 G 12	
Designation:	J251	M 76	M 75	M 74	
C^{14}-Amino acid	Incorporation above control Δ μμMoles[a]				
Alanine	123	45	56	85	8
Arginine	128	30	40	74	9
Aspartic acid (—NH_2?)	167	0	0	24	13
Glutamic acid (—NH_2?)	326	0	0	33	21
Glycine	5	0	0	0	13
Histidine	71	6	9	95	5
Isoleucine	0	0	0	0	20
Leucine	0	10	0	0	40
Lysine	820	5	0	23	6
Methionine	1	4	0	0	10
Phenylalanine	0	0	1	6	9
Proline	147	320	185	41	8
Serine	182	24	30	55	45
Threonine	250	11	13	11	8
Tryptophan	1	0	0	0	43
Tyrosine	0	4	0	0	18
Valine	1	5	5	7	6
Total	2,222	464	339	454	282

[a] Δ μμMoles represents the difference between C^{14}-amino acid incorporation into protein in the presence and absence of polynucleotides.

Assay procedures are described in the footnote of Table II.

The data of Table V demonstrate the effects of poly CG and AG preparations in directing amino acids into protein. The first three CG

polymers contain high proportions of C and directed alanine, arginine, and proline into protein. The last poly CG preparation (F-135) contains 91% G and was inactive as template RNA. Poly AG directed incorporation of glutamic acid and lysine into protein.

TABLE IV
STIMULATION OF AMINO ACID INCORPORATION BY POLY AC

Polynucleotide:	AC	AC	AC	AC	AC	AC	Minus poly-nucleotide control
(moles per cent) Base ratio	A 3 C 97	A 6 C 94	A 12 C 88	A 30 C 70	A 67 C 33	A 80 C 20	
Designation:	J104	J103	J102	J101	J109	J108	
C^{14}-Amino acid	Incorporation above control Δ μμMoles[a]						
Alanine	0	1	0	0	0	0	11
Arginine	0	4	1	1	0	0	12
Aspartic acid (—NH_2?)	0	0	9	51	157	53	24
Glutamic acid (—NH_2?)	4	19	24	53	135	53	15
Glycine	5	16	0	2	0	0	6
Histidine	0	0	5	198	85	17	23
Isoleucine	6	0	0	0	0	0	42
Leucine	0	0	—	—	—	1	3
Lysine	5	10	14	47	909	441	5
Methionine	0	10	0	0	0	0	10
Phenylalanine	0	0	1	0	4	0	11
Proline	625	1,132	643	1,102	140	20	9
Serine	11	19	18	16	9	8	46
Threonine	30	65	75	170	176	105	9
Tryptophan	23	0	1	1	1	9	44
Tyrosine	14	2	2	0	0	0	19
Valine	0	0	0	0	0	0	5
Total	723	1,278	793	1,641	1,616	707	294

[a] Δ μμMoles represents the difference between C^{14}-amino acid incorporation into protein in the presence and absence of polynucleotides.

Assay procedures are described in the footnote of Table II.

Quantitative Aspects of Data

A comparative study of polynucleotides of varying base-ratios is helpful in evaluating amino acid incorporation data, for relative amino acid incorporations easily can be correlated with changes in base-ratio. Occasional inconsistencies and the significance of minor incorporations become apparent.

Isotope dilution experiments were performed routinely to detect the possible presence of radioactive impurities in C^{14}-amino acids. The

presence of C^{14}-impurities seemed unlikely, for incorporation of a C^{14}-amino acid was lowered sharply if the reaction mixture contained both a C^{14}-amino acid (0.05 μmoles) and the same C^{12}-amino acid (1.0 μmole). The purity of each C^{14}-amino acid also was determined by

TABLE V

STIMULATION OF AMINO ACID INCORPORATION BY POLY CG AND AG

Polynucleotide:	CG	CG	CG	CG	AG	AG	Minus poly-nucleotide control
Base ratio (moles per cent)	C 90 G 10	C 87 G 13	C 82 G 18	C 9 G 91	A 73 G 27	A 48 G 52	
Designation:	M141	M 71	F120	F135	J106	J107	
C^{14}-Amino acid	Incorporation above control Δ μμMoles[a]						
Alanine	30	20	63	0	0	0	14
Arginine	39	16	86	1	10	8	13
Aspartic acid (—NH_2?)	0	0	6	3	12	10	26
Glutamic acid (—NH_2?)	0	0	0	0	44	5	11
Glycine	5	0	8	0	2	0	4
Histidine	0	0	0	0	0	0	26
Isoleucine	0	0	0	0	1	0	39
Leucine	0	0	0	5	0	11	7
Lysine	2	0	0	0	110	8	3
Methionine	0	0	0	0	0	0	12
Phenylalanine	5	5	4	8	0	0	14
Proline	144	202	356	2	1	1	8
Serine	18	0	6	0	0	0	42
Threonine	0	0	1	0	1	0	5
Tryptophan	0	1	17	1	0	0	40
Tyrosine	2	6	2	0	0	0	14
Valine	1	1	0	0	0	0	4
Total	246	251	549	20	181	43	282

[a] Δ μμMoles represents the difference between C^{14}-amino acid incorporation into protein in the presence and absence of polynucleotides.

Details of the assay procedures are described in the footnote of Table II.

paper electrophoresis followed by radioautography as described previously (17).

Limiting amounts of polynucleotides were added to reaction mixtures and *total* amino acid incorporations were measured rather than rates of amino acid incorporations. *E. coli* extracts contain nucleases which rapidly degrade synthetic polynucleotides and the nuclease content may vary from one preincubated S-30 preparation to another. Since many different enzyme extracts were used in this study, *the data are*

not useful for quantitative analyses. Comparisons between theoretical frequencies of triplets, etc. in polynucleotides and relative amino acid incorporations have not been presented because the data do not permit such calculations to be made with accuracy. The data demonstrate only qualitative aspects of the code; that is, nucleotide compositions of code-words and the degree of code degeneracy.

Summary of Incorporation Data

Table VI summarizes all of the coding data previously published (19, 16, 17, 29, 15, 14, 30) and obtained in this study. Only polynucleotides containing the minimum number of bases capable of stimulating an amino acid into protein are given in Table VI. The coding of proline by poly C and lysine by poly A was suggested by the poly AC experiments presented in Table III. The fact that poly C and poly A code so weakly may be due either to inhibitory effects of secondary structure or to difficulty in precipitating peptides. At acid pH, poly A in solution is double-stranded (9, 24), and poly C also may have ordered structure (8).

TABLE VI
Summary of Coding Data[a]

C^{14}-Amino acid	Stimulated by poly-			
Phenylalanine	U(98)			
Proline	C(?)	CA(87)	CU(60)	CG(80)
Lysine	A(?)	AC(53)	AG(60)	AU(?)
Threonine	AC(15)			
Serine	UC(23)	UGG(23)?		
Valine	UG(15)			
Leucine	UG(14)	UC(13)	UA(?)	
Glycine	UG(5)			
Cysteine	UG(8-15)			
Glutamic acid (—NH_2?)	AC(7)	AG(20)		
Isoleucine	UA(8)			
Tryptophan	UG(6)			
Tyrosine	UA(9)			
Arginine	CG(15)			
Methionine	UAG(1)			
Histidine	AC(10			
Alanine	CG(11)			
Aspartic acid (—NH_2?)	AC(8)			

[a] Polymers used for these calculations represent optimal base-ratio directing C^{14}-amino acids into protein.

Numbers in parentheses refer to: $\frac{\text{Amino acid incorporated} \times 100}{\text{Sum of incorporation of 17 amino acids}}$

A surprising conclusion revealed by this summary is that almost every amino acid tested could be coded by a polymer containing only two bases. Methionine could be coded only by poly UGA as reported previously (17, 30), but the amount of methionine directed into protein was small; thus this codeword remains questionable.

Assuming a triplet code, a summary of codewords estimated thus far is presented in Table VII. Previously, poly UCG was found to direct alanine and arginine into protein, and codewords containing U, C, and G were proposed for these amino acids (16, 17, 14, 30). The observed frequencies of incorporations (17) suggest coding of alanine and arginine by *either* UCG or CCG, but not by both codewords. In addition, the data of Table V show that poly CG codes for alanine and arginine; thus, codewords corresponding to these amino acids do not appear to contain U.

TABLE VII
TENTATIVE SUMMARY OF CODEWORDS

C^{14}-Amino acid	Codewords[a]				
Alanine	CCG				
Arginine	CGC				
Aspartic acid (—NH_2?)	ACA				
Asparagine	UAC	or	UAA[b]		
Cysteine	UUG	or	UGG[c]		
Glutamic acid (—NH_2?)	ACA		AGA	AGU[d]	
Glycine	UGG				
Histidine	ACC				
Isoleucine	UUA				
Leucine	GUU		CUU	AUU[b]	(UUU)
Lysine	AAA		AAC	AAU	
Methionine	UGA[d]				
Phenylalanine	UUU				
Proline	CCC		CCU	CCA	CCG
Serine	UCG		UCU		
Threonine	CAC		CAA		
Tryptophan	UGG				
Tyrosine	UAU				
Valine	UGU				

[a] Nucleotide sequence in codewords is arbitrary.
[b] Proposed by Speyer *et al.* (30).
[c] We cannot differentiate between these possibilities at present.
[d] It is not entirely clear whether these codewords require U.

Since it is not possible at this time to distinguish between triplet and double codes, etc., the assignments in Table VII represent current approximations of codewords. It seems probable that additional codewords will be found.

Discussion

Codeword Specificity in Protein Synthesis

The term *degeneracy* refers to the phenomenon whereby one amino acid is coded by two or more codewords. This term is inadequate when applied to the mechanism of coding, for it does not indicate codeword *specificity*. A degenerate code may have high or low specificity depending upon the fidelity of protein synthesis. In most cases the fidelity of protein synthesis *in vivo* appears to be high, and amino acid replacements other than those due to mutation have not been found. However, although the amino acid sequence analyses would reveal mistakes at one site occurring with a frequency higher than 1 or 2%, they would not reveal occasional mistakes occurring at different sites. Thus, occasional coding errors of 1 or 2%, distributed at random over entire protein molecules, might not be detected.

In the *in vitro* system, codewords direct amino acids into protein with very striking specificity (21). In Table IV for example, poly AC preparations do not direct the incorporation into protein of alanine, arginine, glycine, isoleucine, leucine, methionine, phenylalanine, tryptophan, tyrosine, or valine. The specificity of coding by poly CG and AG preparations in Table V is equally apparent. Such negative data clearly demonstrate the very high fidelity of codeword recognition during protein synthesis in this cell-free system.

The codewords corresponding to both leucine and valine contain U and G (16, 17, 29). Although the nucleotide content of these codewords are identical, each word was shown to code only for the appropriate amino acid (21). Thus, nucleotide sequence as well as chemical structure confers specificity upon codewords.

However, one example of ambiguity has been found, but this occurs to a large extent in our experiments only under unusual conditions. Poly U directs about 3-5% as much leucine into protein as phenylalanine (17). Bretscher and Grunberg-Manago also have reported this phenomenon (2). In the absence of phenylalanine, using well-dialyzed *E. coli* extracts, poly U coded for leucine about 50% as well as it would code for phenylalanine (20). The molecular basis of this ambiguity is unknown. In the absence of phenylalanine, it is possible that leucine is attached to phenylalanine transfer RNA and then is coded like phenylalanine. On the other hand, the ambiguity may occur at the level of the coding units. It is important to note that phenomena of this type also may occur *in vivo* (3).

Efficiency of Synthetic RNA in Coding

In spite of the previously mentioned difficulties in comparing template activities of RNA preparations with different chain lengths and degrees of secondary structure, it seems clear that synthetic polynucleotides containing 4, 3, or 2 bases code as well in this system as natural template RNA obtained from viruses (19, 22, 32, 18). The efficiency in coding displayed by synthetic polynucleotides suggests that most nucleotide sequences direct amino acids into protein and that relatively few nonsense nucleotide sequences are present.

Although alternative explanations of coding efficiency, such as nonrandom polynucleotides or nonsequential reading of template RNA, may be considered, such efficiency cannot be ascribed simply to random error in directing amino acids into protein, for amino acids are coded with marked specificity.

Considerations such as these may be used to approximate the coding ratio. In a doublet code, only 16 base permutations are possible; thus, the information content would be insufficient to code specifically for all amino acids. Triplet and quadruplet codes would contain 64 and 256 codewords, respectively. Since almost every amino acid tested was found to be coded by polynucleotides containing only two bases, specific and efficient coding by quadruplet words would not seem likely. The data suggest either coding of all amino acids by triplet words, or coding of some by triplets and others by doublets (mixed doublet-triplet code).

Recently, Weisblum, Benzer, and Holley (33) have established a molecular basis of degeneracy by demonstrating that multiple species of transfer RNA recognize different codewords with specificity. Multiple peaks of transfer RNA corresponding to at least four amino acids have been found independently by Holley *et al.* (11), Sueoka *et al.* (31), and Doctor *et al.* (6).

If a triplet code is assumed, each cell would require almost 64 transfer RNA species. Alternatives which do not require so many transfer RNA species deserve consideration. For example, Donohue and others have described many models other than Watson-Crick pairing (7). The demonstrated interaction between poly A and poly I (23), and the type of base-pairing suggested by Hoogsteen (12) also might be cited. Theories which require recognition of either the 2- or 6-substituents of bases (34) are not supported by the demonstration that hypoxanthine functions in codewords like G (28, 1). The 2-amino group of G does not appear to be required for coding.

A triplet code may be constructed wherein correct hydrogen bonding between two out of three nucleotide pairs may, in some cases, suffice for

coding. Correct pairing of a base at one position in the triplet sometimes may be optional. It should be noted that a triplet code of this type in some respects would bear a superficial resemblance to a doublet code and would be in accord with all of the data available.

Any theory concerning the physical basis of the code must attempt to explain the following experimentally obtained data:

(a) High coding efficiency by synthetic polynucleotides.
(b) Marked codeword specificity.
(c) Degenerate codewords.
(d) The 2-amino group of G is not essential for proper coding.
(e) RNA with a high degree of secondary structure has little ability to code.
(f) Almost all amino acids tested can be coded by polynucleotides containing only two bases.

Summary

Synthetic polynucleotides containing 4, 3, or 2 bases have been found to direct amino acids into protein with high efficiency and specificity. Many additional RNA codewords which do not contain uridylic acid have been determined. Almost all amino acids could be coded by polynucleotides containing only 2 bases. These results have been discussed in terms of the general nature of the code.

References

1. Basilio, C., Wahba, A. J., Lengyel, P., Speyer, J. F., and Ochoa, S., *Proc. Natl. Acad. Sci. U.S.*, **48**, 613 (1962).
2. Bretscher, M. S., and Grunberg-Manago, M., *Nature*, **195**, 283 (1962).
3. Cohen, G. N., *Ann. Inst. Pasteur*, **94**, 15 (1958).
4. Crick, F. H. C., in "Structure and Function of Genetic Elements, Brookhaven Symposia in Biology, No. 12," 1959, p. 35.
5. Davidson, J. N., and Smellie, R. M. S., *Biochem. J.*, **52**, 594 (1952).
6. Doctor, B. P., Apgar, J., and Holley, R. W., *J. Biol. Chem.*, **236**, 1117 (1962).
7. Donohue, J., *Proc. Natl. Acad. Sci. U.S.*, **42**, 60 (1956).
8. Fresco, J. R., *Trans. N.Y. Acad. Sci., Series II*, **21**, 653 (1959).
9. Fresco, J. R., and Doty, P., *J. Am. Chem. Soc.*, **79**, 3928 (1957).
10. Gamow, G., *Nature*, **173**, 318 (1954).
11. Holley, R. W., Doctor, B. P., Merrill, S. H., and Saad, F. M., *Biochim. et Biophys. Acta*, **35**, 272 (1959).
12. Hoogsteen, K., *Acta Cryst.*, **12**, 822 (1959).
13. Jones, O. W., and Martin, R. G., *Federation Proc.*, **21**, 414 (1962).
14. Lengyel, P., Speyer, J. F., Basilio, C., and Ochoa, S., *Proc. Natl. Acad. Sci. U.S.*, **48**, 282 (1962).

15. Lengyel, P., Speyer, J. F., and Ochoa, S., *Proc. Natl. Acad. Sci. U.S.*, **47**, 1936 (1961).
16. Martin, R. G., Matthaei, J. H., Jones, O. W., and Nirenberg, M. W., *Biochem. Biophys. Research Communs.*, **6**, 410 (1962).
17. Matthaei, J. H., Jones, O. W., Martin, R. G., and Nirenberg, M. W., *Proc. Natl. Acad. Sci. U.S.*, **48**, 666 (1962).
18. Nathans, D., Notani, G., Schwartz, J. H., and Zinder, N. D., *Proc. Natl. Acad. Sci. U.S.*, **48**, 1424 (1962).
19. Nirenberg, M. W., and Matthaei, J. H., *Proc. Natl. Acad. Sci. U.S.*, **47**, 1588 (1961).
20. Nirenberg, M. W., Matthaei, J. H., and Jones, O. W., unpublished observations.
21. Nirenberg, M. W., Matthaei, J. H., Jones, O. W., Martin, R. G., and Barondes, S. H., *Federation Proc.*, **22**, 55 (1963).
22. Ofengand, J., and Haselkorn, R., *Biochem. Biophys. Research Communs.*, **6**, 469 (1962).
23. Rich, A., *Nature*, **181**, 521 (1958).
24. Rich, A., Davies, D. R., Crick, F. H. C., and Watson, J. D., *J. Mol. Biol.*, **3**, 71 (1961).
25. Roberts, R. B., *Proc. Natl. Acad. Sci. U.S.*, **48**, 897 (1962).
26. Roberts, R. B., *Proc. Natl. Acad. Sci. U.S.*, **48**, 1245 (1962).
27. Singer, M. F., and Guss, J. K., *J. Biol. Chem.*, **237**, 182 (1962).
28. Singer, M. F., Jones, O. W., Matthaei, J. H., and Nirenberg, M. W., unpublished observations.
29. Speyer, J. F., Lengyel, P., Basilio, C., and Ochoa, S., *Proc. Natl. Acad. Sci. U.S.*, **48**, 63 (1962).
30. Speyer, J. F., Lengyel, P., Basilio, C., and Ochoa, S., *Proc. Natl. Acad. Sci. U.S.*, **48**, 441 (1962).
31. Sueoka, N., and Yamane, T., *Proc. Natl. Acad. Sci. U.S.*, **48**, 1454 (1962).
32. Tsugita, A., Fraenkel-Conrat, H., Nirenberg, M. W., and Matthaei, J. H., *Proc. Natl. Acad. Sci. U.S.*, **48**, 846 (1962).
33. Weisblum, B., Benzer, S., and Holley, R. W., *Proc. Natl. Acad. Sci. U.S.*, **48**, 1449 (1962).
34. Woese, C. R., *Nature*, **194**, 1114 (1962).