[12] Setlow, R. B., and J. K. Setlow, these PROCEEDINGS, 48, 1250 (1962).
[13] Setlow, R. B., and W. L. Carrier, these PROCEEDINGS, 51, 226 (1964).
[14] Pettijohn, D., and P. Hanawalt, *J. Mol. Biol.*, 9, 395 (1964).

RNA CODEWORDS AND PROTEIN SYNTHESIS, VII. ON THE GENERAL NATURE OF THE RNA CODE

BY M. NIRENBERG, P. LEDER, M. BERNFIELD, R. BRIMACOMBE, J. TRUPIN,* F. ROTTMAN†, AND C. O'NEAL

NATIONAL HEART INSTITUTE, NATIONAL INSTITUTES OF HEALTH, BETHESDA, MARYLAND

Communicated by Robert J. Huebner, March 26, 1965

Nucleotide sequences of RNA codons have been investigated recently by directing the binding of C^{14}-AA-sRNA to ribosomes with trinucleotides of defined base sequence. The template activities of 19 trinucleotides‡ have been described and nucleotide sequences have been suggested for RNA codons corresponding to 10 amino acids.[1-5] In this report, the template activities of 26 additional trinucleotides are described and are related to the general nature of the RNA code.

Materials and Methods.—Components of reactions: *E. coli* W3100 ribosomes and sRNA were prepared by modifications of methods described previously.[6-8] Each C^{14}-aminoacyl-sRNA was prepared in the presence of 19 C^{12}-amino acids. The assay for ribosomal bound C^{14}-AA-sRNA and components of reaction mixtures have been described.[1] The characteristics and amounts of labeled AA-sRNA not described[5] are shown in Table 1.

Synthesis and characterization of oligonucleotides: ApG and UpG were obtained from a T-1 ribonuclease digest of RNA, and ApA was prepared by chemical synthesis.[9, 10] ApC, ApU, CpA, CpG, GpA, and GpC were obtained from Gallard Schlessinger Corp., but required extensive purification prior to use. GpGpU was obtained by digesting poly UG with pancreatic RNase A; treatment with alkaline phosphatase to remove terminal phosphate groups from the degradation products; and isolation by procedures similar to those described for GpUpU.[2] The remaining trinucleotides were synthesized from the appropriate dinucleoside monophosphate, using either

TABLE 1

RADIOACTIVE AMINOACYL-sRNA PREPARATIONS*

Radioactive amino acid†	Specific radioactivity $\mu c/\mu mole$	C^{14} or H^3-AA-sRNA Added to Each Reaction: A_{260} units	$\mu\mu$moles of C^{14}- or H^3-Amino acid accepted	Origin of sRNA *E. coli* strain‡
Lys	240	0.25	8.5	W3100
Expt. *b*	209	0.28	10.2	B
Ala	88	0.33	8.5	W3100
Expt. *b*	88	0.75	32.6	W3100
Glu	205	0.56	14.9	B
Expt. *b*	188	0.85	6.1	W3100
Gly-H^3	1130	0.15	11.72	A-23§
-C^{14} Expt. *b*	66	0.47	9.2	W3100
Pro	200	0.70	6.4	B
Expt. *b*	158	0.57	13.2	W3100
Ser	120	1.03	6.3	W3100
Expt. *b*	120	0.42	18.6	B
Trypt-H^3	3000	1.00	6.5	B

* Other AA-sRNA data have been described.[5]
† Amino acids stated were labeled with C^{14} with the exception of H^3-tryptophan, and H^3-glycine.
‡ *E. coli* W3100 is a K12 strain. Aminoacyl sRNA synthetase preparations were from *E. coli* W3100.
§ We thank Dr. Charles Yanofsky for this *E. coli* strain and Dr. Ray Byrne for the H^3-Gly-sRNA. A-23 sRNA and 100,000 × *g* supernatant fractions were used for the preparation of H^3-Gly-sRNA.

TABLE 2

CHARACTERIZATION OF TRINUCLEOTIDES

Compound	Synthetic method[a]	Digestion (T-2 ribonuclease) Products	Base ratio	Digestion (venom phosphodiesterase) Products	Base ratio
ApCpA	1	Ap,Cp,A	1.00/0.95/1.05	A,pC,pA	1.05/0.90/1.00
ApCpC	1	Ap,Cp,C	1.05/1.00/0.95	A,pC	1.15/2.00
ApCpG[b]	1	Ap,Cp,G	1.00/0.95/1.05	A,pC,pG	1.00/0.95/1.00
ApCpU	1	Ap,Cp,U	1.00/1.00/0.85	A,pC,pU[c]	1.00/0.90/1.10
ApGpA	1	Ap,Gp,A	1.00/1.00/0.95	A,pG,pA	1.00/1.00/1.00
ApGpC	1	Ap,Gp,C	1.05/1.00/0.95	A,pG,pC[c]	1.00/1.00/1.15
ApGpU[b]	1	Ap,Gp,U[c]	1.10/1.00/0.95	A,pG,pU[c]	1.00/1.00/1.00
ApUpG	1	Ap,Up,G	1.00/0.85/1.05	A,pU,pG[d]	1.00/0.90/1.10
CpApA	2	Cp,Ap,A	0.95/1.05/1.00	C,pA	1.05/2.00
CpApC	2	Cp,Ap,C	1.00/1.10/0.95	C,pA,pC	1.00/1.00/1.00
CpApG[b]	2	Cp,Ap,G	1.10/1.00/1.00	C,pA,pG	1.00/1.00/1.15
CpApU	2	Cp,Ap,U	1.00/1.10/0.90	C,pA,pU	1.00/1.05/0.95
CpCpA	2[e]	Cp,A	2.00/0.95	.....	.....
CpGpA[b]	2	Cp,Gp,A	1.10/1.00/0.95	C,pG,pA	1.05/0.95/1.00
CpGpC	2	Cp,Gp,C	1.05/1.00/0.80	C,pG,pC[d]	0.90/1.00/1.00
CpUpG	2	Cp,Up,G	0.95/1.00/1.10	C,pU,pG	0.95/1.00/1.10
GpApA[b]	1	Gp,Ap,A[d]	0.90/1.00/1.10	G,pA[d]	1.10/2.00
GpApC[bg]	1	Gp,Ap,C	1.05/1.00/0.90	G,pA,pC	1.05/1.00/0.90
GpApU	1	Gp,Ap,U	1.00/1.00/0.95	G,pA,pU[d]	1.20/1.00/0.90
GpCpU	1	Gp,Cp,U	1.05/1.00/0.90	G,pC,pU	0.85/1.00/1.00
GpGpU[bf]	3	Gp,U[f]	2.00/0.95	G,pG,pU[f]	0.85/1.00/1.05
UpApA	2	Up,Ap,A[d]	1.00/1.00/1.10	U,pA	0.95/2.00
UpApG	2	Up,Ap,G	1.00/0.95/1.00	U,pA,pG[c]	1.00/1.15/0.90
UpCpG	2	Up,Cp,G	1.00/0.90/1.00	U,pC,pG	0.95/1.00/1.05
UpGpA	2	Up,Gp,A	0.95/1.00/1.10	U,pG,pA	0.90/1.00/1.15
UpGpC[b]	2	Up,Gp,C	1.00/0.95/1.05	U,pG,pC[d]	1.10/1.00/0.90

[a] *Method 1:* primer-requiring polynucleotide phosphorylase.[12] *Method 2:* derivative of bovine pancreatic ribonuclease.[11] *Method 3:* isolation from a ribonuclease digest of poly-UG.[2]

[b] Trinucleotide contained a small amount (*ca.* 2%) of an unidentified contaminant. In every case, the chromatographic characteristics of the impurity in solvent systems A and B precluded the possibility of it being an oligonucleotide of chain-length greater than two base residues.

[c] The digestion mixture contained another component, 3–5% of the total nucleotide content; the electrophoretic mobility and UV spectrum of this component corresponded in each case to unreacted or partially digested oligonucleotide.

[d] As "*c*" above, except that here the contaminant was only 1–3% of the total nucleotidic material.

[e] Trinucleotide was synthesized directly from adenosine, by addition of two cytidylic acid moieties.

[f] Trinucleotide contained 5–10% of a contaminant which did not migrate in solvent A, but which ran with GpGpU in solvent B. The T-2 digestion mixture contained 8%, and the venom digestion mixture 14%, of undigested material; this was possibly due to aggregation of the trinucleotide.

[g] Trinucleotide contained 10% of pC; the base ratio of the venom digestion was adjusted accordingly.

primer-requiring polynucleotide phosphorylase and a nucleoside 5′-pyrophosphate,[12] or a derivative of bovine pancreatic ribonuclease with a nucleoside 2′,3′-cyclic phosphate[11] (Table 2). The products were isolated by paper chromatography and electrophoresis, as previously described[1–4, 12] and the purity of each preparation was assessed by two-dimensional chromatography of an aliquot (2.0 A^{260} units) on Whatman no. 40 paper. The first dimension (solvent A) was n-propanol/ammonia/water, 55/10/35; the second dimension (solvent B) was 0.10 *M* sodium phosphate, pH 7.0, containing ammonium sulfate (0.4 gm/ml). Chain-length and base composition (Table 2) were determined by digestion of 2.5 A^{260} units of each trinucleotide with T-2 ribonuclease and 2.5 A^{260} units with venom phosphodiesterase, as previously described.[4, 12]

Results and Discussion.—In Table 3 are shown the effects of 26 trinucleotides upon the binding to *E. coli* ribosomes of 19 C^{14}-AA-sRNA preparations, each acylated with a different C^{14}-amino acid (C^{14}-Cys-sRNA not used). In addition, near the bottom of the table are shown the effects of 18 trinucleotides previously described[1–5] upon the corresponding C^{14}-AA-sRNA (C^{14}-Cys-sRNA and UpGpU omitted). Many of these trinucleotides have not been isolated or synthesized previously.

Several factors should be mentioned which may be useful in assessing the data. (*a*) It is often difficult to compare *directly* the response of one C^{14}-AA-sRNA preparation to a template with that of another, for Kaji and Kaji[13] have shown that

both deacylated and acylated sRNA bind to ribosomes in response to polynucleotide templates. The extent of acylation of each C^{14}-AA-sRNA preparation must be considered (see *Methods and Materials*) as well as the relative response of each C^{14}-AA-sRNA to other trinucleotides. (*b*) A trinucleotide which stimulates the binding to ribosomes of one C^{14}-AA-sRNA generally decreases binding of other C^{14}-AA-sRNA preparations.[1] (*c*) Background binding of C^{14}-AA-sRNA to ribosomes appears to be a function of the sRNA species, the amount of sRNA added to a reaction, the proportion of sRNA acylated with a C^{14}-amino acid, and possibly the amount of template RNA on the ribosomes or in the sRNA preparations.[1, 4, 8] (*d*) Reactions contained limiting concentrations of ribosomes (as determined with ApApA, UpUpU, UpUpC, or GpUpU) and therefore were saturated with respect to these trinucleotides and C^{14}-AA-sRNA.

Most trinucleotides markedly stimulated the binding to ribosomes of only one C^{14}-AA-sRNA preparation; however, a number of trinucleotides displayed lower template specificity for C^{14}-AA-sRNA. For example, ApCpU, ApCpC, and ApCpG stimulated C^{14}-Thr-sRNA binding to ribosomes, but did not significantly stimulate the binding of 18 other C^{14}-AA-sRNA preparations. ApCpA also stimulated C^{14}-Thr-sRNA binding. This trinucleotide also stimulated C^{14}-Lys-sRNA binding. However, the template activity of ApCpA for C^{14}-Lys-sRNA was only 10 per cent that of ApApA. The disparity between the template activity of ApCpA and ApApA was even more apparent in reactions containing limiting concentrations of trinucleotides (data not shown). Such considerations suggest that the sequences ApCpG, ApCpU, ApCpC, and ApCpA correspond to threonine codons. It is possible that the template specificity of one synonym codon may differ from that of another; however, other alternatives, such as the possibility that C^{14}-Lys-sRNA may respond to an impurity in the ApCpA preparation which we have been unable to detect, also must be considered.

The data of Table 3 indicate that the sequence GpCpU corresponds to an RNA codon for alanine; CpCpA, CpCpU, and CpCpC correspond to proline (the template activity of CpCpA for C^{14}-Pro-sRNA was higher than that of pCpCpC (cf. ref. 4); UpCpG, UpCpU, and UpCpC correspond to serine (cf. ref. 4); GpApU and GpApC, to aspartic acid; GpApA, to glutamic acid; CpApU and CpApC, to histidine; CpApA and CpApG, to glutamine; CpGpC and CpGpA, to arginine; ApUpG, to methionine; and CpUpG and UpUpG, to leucine (CpUpU and CpUpC possibly serve as internal but not terminal Leu-codons[4]).

It seems clear that GpCpU serves as a codon for alanine, for this trinucleotide stimulated only the binding of C^{14}-alanine sRNA to ribosomes. This sequence is also in accord with predictions based upon amino acid replacement data. However, the weaker response of C^{14}-Ala-sRNA to ApGpC, CpGpC, and UpGpC suggests that recognition of 2 out of 3 bases, the GpC portion only of the latter trinucleotides, may permit C^{14}-Ala-sRNA binding. Similarly, C^{14}-Glu-sRNA responds best to GpApA, but also responds to a weaker extent to trinucleotides containing GpA, such as ApGpA, CpGpA, and UpGpA. C^{14}-Lys-sRNA responds best to ApApA but also recognizes ApApG,[5] GpApA, ApCpA, CpApA, UpApA, and CpCpA. Additional examples in Table 3 are readily apparent. These data indicate that one trinucleotide sometimes can direct the attachment of a limited group of C^{14}-AA-sRNA species to ribosomes. It is possible that correct re-

TABLE

TEMPLATE SPECIFICITY OF TRINUCLEOTIDES

Trinucleotide	Δ μμMoles of C^{14}- or H^3-Aminoacyl-sRNA Bound to: C^{14}-Ala	C^{14}-Arg	C^{14}-Asp	C^{14}-Asp-NH_2	C^{14}-Glu	C^{14}-Glu-NH_2	C^{14}- or H^3-Gly	C^{14}-His	C^{14}-Ileu
ApCpU	−0.16	−0.04	−0.08	−0.03	−0.02	−0.03	−0.27[b]	−0.02	0
ApCpC	−0.15	−0.15	0.03	0.05	−0.02	−0.22	−0.34	0.01	0.02
ApCpA	−0.04	−0.01	0.05	0.05	0.01	−0.03	−0.78	0	0
ApCpG	−0.15	−0.37	0.02	−0.03	−0.04	0.03	−0.18[b]	0.03	0.02
GpCpU	**0.71**	−0.18	−0.08	−0.05	0.01	−0.13	−0.13[b]	−0.02	0
CpCpA	−0.15	0.06	−0.03	0	0	0.04	0.04	0.01	0.01
UpCpG	−0.20	0	0	0.07	0.01	−0.23	−0.44	0.01	0.02
GpApU	−0.01[b]	−0.14	**1.29**	0.33†	0.05[b]	−0.02	−0.23[b]	−0.03	0.01
GpApC	−0.05[b]	−0.22	**1.32**	0.19†	0[b]	−0.10	−0.21[b]	0.02	−0.01
GpApA	−0.07[b]	−0.11	0.01	0	**0.62**	−0.32	−0.68	−0.04	−0.01
CpApU	0.01[b]	−0.31	−0.01	−0.03	−0.02	0	−0.13[b]	**0.52**	0
CpApC	−0.02[b]	−0.25	−0.01	0.04	−0.02	−0.14	−0.08[b]	**0.26**	−0.02
CpApA	0.02[b]	0.01	−0.06	−0.07	−0.02	**2.05**	−0.03[b]	0.02	−0.04
CpApG	−0.13	−0.01	0.02	−0.03	0.05	**2.60**	−0.12	−0.03	0
UpApA	−0.02[b]	−0.41	−0.01	−0.07	0.02	−0.30	−0.84[b]	−0.08	0
UpApG	−0.07	0.06	0.01	**0.12**	0.04	0	−0.01	−0.01	0.02
ApGpU	−0.07[b]	0.01	0	0.04	0	−0.03	−0.06[b]	0.03	−0.02
ApGpC	**0.43**[b]	−0.05	0.03	**0.10**	−0.02	0	−0.13[b]	0.03	−0.03
ApGpA	−0.06	**0.10**	0.01	0	**0.19**	0	−0.31	0	−0.03
GpGpU	−0.02[b]	−0.33	0.01	−0.02	0.04	−0.21	**3.04**	−0.03	−0.01
CpGpC	**0.14**	**1.63**	0.02	−0.13	−0.01	−0.07	−0.02	−0.03	−0.02
CpGpA	−0.20	**1.42**	−0.01	−0.01	0.07	−0.05	0.09	−0.03	0
UpGpC	**0.28**	−0.18	0.05	**0.12**	0.04	−0.05	−0.55	−0.02	−0.02
UpGpA	−0.12	−0.12	0.07	**0.14**	**0.10**	−0.13	−0.30	−0.03	0.02
ApUpG	−0.01[b]	−0.11	−0.06	0.02	−0.01	−0.14	−0.17[b]	−0.02	0
CpUpG	−0.14	−0.11	−0.01	0.04	0.01	0.11	−1.20	0.03	0.03
Minus trinucleotide (μμmoles)*	0.50 0.20[b]	1.16	0.21	0.21	0.12 0.14[b]	1.65	2.89 1.10[b]	0.25	0.08
Trinucleotides previously described[1–5] (Δ μμmoles)	...	...	...	**1.19** ApApU **1.50** ApApC	...	...	...	...	**0.72** ApUpU **0.59** ApUpC

The specificity of trinucleotides in directing the binding of C^{14}- or H^3-aminoacyl-sRNA to ribosomes. Reproducible stimulations of AA-sRNA binding due to the addition of trinucleotides are bold face. For comparison, the template activities of 18 trinucleotides previously described[1–5] are shown at the bottom of the table. Reactions contained the components described under *Materials and Methods*, the amount of C^{14}-AA-sRNA stated previously[5] or in Table 1, and 0.150 A^{260} units of trinucleotide, as specified, in a final volume of 50 μl. C^{14}-Asp-NH_2-sRNA was

cognition of 2 out of 3 bases in a trinucleotide, in or out of phase, or 2 bases in 1 trinucleotide and 1 base in an adjacent trinucleotide, often may suffice during protein synthesis. This striking phenomenon is often observed with trinucleotides containing 2 or 3 purines. Since the stability of codon-ribosome-AA-sRNA complexes may partially depend upon interactions between bases in codons and sRNA, the affinity of sRNA for a ribosome may be greater when each base in a codon is recognized correctly and in proper phase than when codon recognition is only partially correct.

The activity of *both* ApGpU and ApGpC in stimulating binding of C^{14}-Ser-sRNA may indicate that these sequences correspond to serine codons (in addition to UpCpU, UpCpC, and UpCpG). However, these assignments should be considered tentative. Although ApGpU and ApGpC have been proposed as asparagine codon sequences,[15] the data of Table 3 show that they do not significantly affect the binding of C^{14}-Asp-NH_2-sRNA under the conditions employed. We have previously reported that ApApU and ApApC stimulate binding of Asp-NH_2-sRNA with high specificity.[5]

ApGpA slightly stimulated the binding of both C^{14}-Arg- and C^{14}-Glu-sRNA. The sequence ApGpA was predicted for arginine on the basis of codon sequence data obtained earlier and amino acid replacement data reported by Yanofsky and

3

FOR C^{14}- OR H^3-AMINOACYL-sRNA

Ribosomes Due to Addition of Trinucleotides*

C^{14}-Leu	C^{14}-Lys	C^{14}-Met	C^{14}-Phe	C^{14}-Pro	C^{14}-Ser	C^{14}-Thr	H^3-Trypt	C^{14}-Tyr	C^{14}-Val
−0.12	0.01	0.03	0.02	−0.02	−0.05	**0.63**	−0.01	0.03	0
−0.09	−0.03	0	0	−0.01	0.03	**0.50**	−0.03	0.03	0
−0.09	**0.17**	−0.04	0.01	0.05[b]	−0.07	**0.45**	0.03	0.01	0.01
−0.22	0	0.02	−0.07	−0.01	−0.15	**0.75**	−0.03	0.03	0
−0.25	−0.10	−0.18	−0.04	−0.02	−0.18	0.06	0	0.02	0
0.02	**0.11**	0.03	0.03	**0.40**	−0.08	−0.02	0.03	0.03	0
0	0.04	−0.11	0.03	0.06[b]	**1.09**	−0.04	0.02	0.05	0.03
−0.10	−0.12	−0.09	−0.29	−0.03	0.01[b]	−0.04	0.04	...	0
−0.10	−0.10	−0.10	−0.24	−0.01	0.01[b]	0.01	0.02	0.03	0.03
−0.03	**0.73**	0	−0.03	−0.07[b]	−0.23	−0.08	−0.04	0	−0.04
−0.03	−0.07	−0.04	−0.15	−0.01	0.02[b]	−0.07	−0.01	0.03	0.02
0.01	−0.13	−0.01	−0.13	0.01	−0.03[b]	0.05	−0.03	0.04	0.03
0.03	**0.10**	−0.12	−0.19	−0.01	−0.02[b]	−0.04	0	−0.01	−0.01
0.03	−0.03	−0.02	0.01	−0.01	−0.09	−0.09	−0.03	0.09	0.02
−0.05	**0.10**	−0.09	−0.38	0.02	0.02	−0.03	0	0.01	0.03
−0.09	−0.05	−0.02	−0.17	0	0	−0.05	0.03	0.03	0
−0.16	−0.18	−0.05	−0.22	−0.04	**0.27**	−0.02	−0.04	−0.03	0.08
−0.11	−0.12	−0.05	−0.20	−0.01	**0.17**[b]	...	−0.05	0.01	0.02
−0.09	−0.01	−0.06	−0.13	0.01	0.03	−0.05	−0.05	0.02	0.02
−0.27	−0.10	−0.08	−0.34	−0.07	−0.07[b]	−0.11	−0.04	0.03	0.03
0.03	−0.16	0.01	−0.03	−0.05[b]	−0.12	0	−0.02	0	−0.01
−0.14	−0.11	−0.17	−0.04	−0.12	−0.22	−0.06	0	0.03	−0.02
−0.07	0	−0.18	0.04	0.02[b]	0.02	...	0.04	0.03	0.02
−0.04	−0.11	0	−0.27	0	0	−0.08	0.03	0.03	−0.05
−0.38	−0.06	**1.00**	−0.04	0	−0.05[b]	0.01	−0.06	0.03	−0.10
0.30	0.03	**0.10**	0.05	0.03	0.03	−0.06	−0.03	0.04	0.05
0.79	0.77	0.40	0.44	0.14	0.58	0.30	0.30	0.12	0.22
...	...	...	...	0.40[b]	0.24[b]	...	...	...	...
0.23	**1.77**	...	**1.05**	**0.28**	**1.27**	...	...	**0.81**	**1.54**
UpUpG	ApApA	...	UpUpU	pCpCpC	UpCpU	...	...	UpApU	GpUpU
0.04	**1.00**	...	**1.71**	**0.08**	**0.54**	...	...	**0.56**	...
CpUpU	ApApG	...	UpUpC	CpCpC	UpCpC	...	...	UpApC	...
0.03	...	...	...	**0.15**	...	...	...	...	...
CpUpC	...	...	...	CpCpU	...	...	...	...	...

assayed in 100-μl reactions; amounts of all components were doubled.

* Background binding of C^{14}-aminoacyl-sRNA to ribosomes in the absence of trinucleotides is expressed in μμmoles (shown near the bottom of the table). All other values (Δ μμmoles) were obtained by subtracting background binding of C^{14}-aminoacyl-sRNA from binding obtained upon addition of a trinucleotide preparation.

† sRNA may contain some C^{14}-Asp-sRNA.

co-workers.[14] The recognition of one triplet by sRNA corresponding to several amino acids again suggests partial codon recognition. Such observations should be considered in terms of *in vivo* studies, particularly those related to extragenic suppression.

Possible Base Sequences of Nonsense Codons.—Since nonsense codons may perform special functions in protein synthesis, we have been particularly interested in a small group of trinucleotides, UpApA, UpApG, UpGpA, CpUpU, CpUpC, and ApGpA, which either have little template activity, or have slight activity for two or more C^{14}-AA-sRNA. The possibility that CpUpU or CpUpC may serve as internal, but not as terminal, codons for leucine has been discussed previously.[4]

The striking results of Sarabhai, Stretton, Brenner, and Bolle[16] indicate that certain codons in "amber" mutants of T4 phage may correspond to an amino acid in certain strains of *E. coli* but may specify the terminus of a protein in other *E. coli* strains. Further analysis of mutant phages and amino acid replacement data have led Brenner and co-workers to suggest that UpApG or UpApA may specify the end of a polypeptide chain in some *E. coli* strains and serine in an additional strain which contains a suppressor gene.[17] Weigert and Garen also have found that mutations which lead to the formation of nonsense codons in the alkaline phosphatase gene can be related to amino acid substitutions at sites corresponding to

nonsense codons only if the base composition of the nonsense codon is (UAG).[18] This codon also corresponds to serine in a strain containing an appropriate suppressor gene.

We find that the sequences UpApG and UpApA have almost no template activity for C^{14}-amino acids under the conditions employed (only very small stimulation of C^{14}-Asp-NH_2- and C^{14}-Lys-sRNA binding was observed). These data are in full accord with the conclusions of Brenner and of Garen and their co-workers. In addition the sequences found for glutamine codons were CpApG and CpApA; the sequences found for serine codons were UpCpG, UpCpU, UpCpC, ApGpU, and ApGpC (UpCpA also predicted). These sequences demonstrate structural relationships between Terminator-, Glu-NH_2-, and Ser-codons and suggest mechanisms for alternate codon recognition in different strains of *E. coli*.

The General Nature of the Code.—Thus far, the template functions of 45 of the 64 trinucleotide sequences have been investigated in this system. A summary of the data and additional codon sequences which can be predicted from amino acid replacement data reported for *E. coli*[14] and TMV mutants,[19, 20] are shown in Table 4. Almost all of our earlier predictions were confirmed when the appropriate trinucleotide was tested.[2–5] Nevertheless, the summary shown in Table 4 should not be thought of as an invariant codon dictionary, since it is clear that codon recognition can be modified.

Previous studies with randomly ordered polynucleotides and cell-free protein synthesizing systems showed that synonym codons often differ in composition by only one base.[22, 23] This suggested that bases common to synonym codons occupy identical positions and, that either 2 out of 3 bases in a triplet sometimes may be recognized, or a base may be recognized correctly in 2 or more ways.[21, 24] On the

TABLE 4

NUCLEOTIDE SEQUENCES OF RNA CODONS

UpUpU ***UpUpC***	Phe	***UpCpU*** ***UpCpC***	Ser	***UpGpU*** UpGpC	Cys	***UpApU*** ***UpApC***	Tyr
UpUpA ***UpUpG***	Leu	UpCpA ***UpCpG***	Ser	***UpGpA*** UpGpG	Nonsense* or Trypt	***UpApA*** ***UpApG***	Nonsense†
CpUpU ***CpUpC***	Leu or Nonsense*	***CpCpU*** ***CpCpC***	Pro	CpGpU ***CpGpC***	Arg	***CpApU*** ***CpApC***	His
CpUpA ***CpUpG***	Leu	***CpCpA*** CpCpG	Pro	***CpGpA*** CpGpG	Arg	***CpApA*** ***CpApG***	Glu-NH_2
ApUpU ***ApUpC***	Ileu	***ApCpU*** ***ApCpC***	Thr	***ApGpU*** ***ApGpC***	Ser	***ApApU*** ***ApApC***	Asp-NH_2
ApUpA ***ApUpG***	Met	***ApCpA*** ***ApCpG***	Thr	***ApGpA*** ApGpG	Arg. or Nonsense*	***ApApA*** ***ApApG***	Lys
GpUpU GpUpC	Val	***GpCpU*** GpCpC	Ala	***GpGpU*** GpGpC	Gly	***GpApU*** ***GpApC***	Asp
GpUpA GpUpG	Val	GpCpA GpCpG	Ala	GpGpA GpGpG	Gly	***GpApA*** GpApG	Glu

* It is possible that these sequences are readable internal-, but nonreadable terminal-, codons.

† UpApA and UpApG may correspond to Terminator-, or Ser-codons in different strains of *E. coli* (see text or refs. 17 and 18).

Summary and predictions: The template activities of trinucleotides in BOLDFACE have been studied experimentally in this system. Other sequences are predicted. Although trinucleotides are arranged in pairs, one member of a pair may have greater template activity than the other. Estimates of relative template efficiencies are not indicated.

Amino acid replacement data used for these predictions were obtained with *E. coli* by Yanofsky,[14] or were induced by HNO_2 in TMV by Wittman and Wittman-Liebold[19] or Tsugita.[20]

basis of such data Woese suggested a code in which A, C, G and U are independently recognized at one position in the triplet, C = U at a second position, and A = C and G = U at a third position.[25] A modification was suggested by Eck in which U = C and A = G at an unspecified position in a triplet.[26] We have previously shown that each member of a trinucleotide pair with 3'-terminal pyrimidines, such as XpYpU and XpYpC, corresponds to the same amino acid,[4, 5] and each member of a codon pair with 3'-terminal purines, ApApA and ApApG, corresponds to lysine.[5]

Several generalizations can be made concerning the nature of the code. (*a*) Amino acids, which are structurally or metabolically related (such as synthesized *in vivo* from a common precursor) often have similar RNA codons. Such relationships would appear to reflect either the evolution of the code[24] or direct interactions between amino acids and bases in codons.[27, 28] (*b*) Many codons may be recognized partially, or may contain alternate acceptable bases at certain positions. (*c*) Recognition of the 3'-terminal base in a trinucleotide is most variable and fits several general patterns; U = C; G = A; or G = A = U = C. (*d*) In the case of Leu-codons, U = C at the 5'-terminal position. (*e*) In most cases the apparent template activity of one member of a synonym codon set differs from that of another.

These patterns apparently define the characteristics of several general recognition mechanisms. Although the molecular mechanisms which permit one codon to be distinguished from another are unknown, the pairing of bases in mRNA with bases in sRNA is an obvious one to consider. Enzymic modification of bases in certain doublet or triplet sequences in sRNA might affect codon recognition, as proposed by Ames and Hartman,[29] and also explain patterns of synonym codons sets. Interconversion of C and U (C $\rightleftarrows$ U) in an sRNA "anticodon" might result in a G = A pattern in mRNA codons and interconversion of A and I (A $\rightleftarrows$ I) in a U = C pattern in mRNA codons. However, the possibility of alternate acceptable base pairing is raised by observations which indicate that one molecule of Phe-sRNA may recognize *both* UpUpU and UpUpC.[4]

The nucleotide sequence found for an alanine codon (GpCpU found, GpCpC, GpCpA, and GpCpG predicted), together with the nucleotide sequence of an alanine sRNA isolated from yeast, reported by Holley *et al.*,[30] may provide clues to the recognition process. Two striking sequences in yeast Ala-sRNA, IpGpCpMeIpψ, and dihydroUpCpGpGpdihydroU, each potentially comprising a single-stranded loop at the end of a hairpin-like double-stranded segment, have been suggested as possible "anticodons".[30] The first sequence suggests *antiparallel* Watson-Crick base pairing between Ala-sRNA and a GpCpC Ala-codon mRNA (CGI)/GCC. The second sequence raises the possibility of *parallel* Watson-Crick pairing with an Ala-codon of GpCpC, (CGG/GCC). Further work is necessary to determine whether one molecule of Ala-sRNA is capable of recognizing two or more alanine codons.

It is a pleasure to thank Taysir Jaouni, Norma Zabriskie, and Theresa Caryk for invaluable assistance.

* Supported by American Cancer Society postdoctoral fellowship PF 201.

† Supported by USPHS postdoctoral fellowship 6 F2 AM-17, 108-01A1, and American Cancer Society PF 244.

‡ For brevity, trinucleoside diphosphates are referred to as trinucleotides.

[1] Nirenberg, M., and P. Leder, *Science*, **145**, 1399 (1964).

[2] Leder, P., and M. Nirenberg, these PROCEEDINGS, **52**, 420 (1964).

[3] Leder, P., and M. W. Nirenberg, these PROCEEDINGS, **52**, 1521 (1964).

[4] Bernfield, M. R., and M. W. Nirenberg, *Science*, **147**, 479 (1965).

[5] Trupin, J., F. Rottman, R. Brimacombe, P. Leder, M. Bernfield, and M. Nirenberg, these PROCEEDINGS, **53**, 807 (1965).

[6] Nirenberg, M. W., in *Methods in Enzymology*, ed. S. P. Colowick and N. O. Kaplan (New York: Academic Press, 1964), vol. 6, p. 17.

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

[8] Pestka, S., R. E. Marshall, and M. W. Nirenberg, these PROCEEDINGS, **53**, 639 (1965).

[9] Lapidot, Y., and G. Khorana, *J. Am. Chem. Soc.*, **85**, 3852 (1963).

[10] Chladek, S., and J. Smrt, *Collection Czech. Chem. Commun.*, **28**, 1301 (1963).

[11] Bernfield, M. R., and M. W. Nirenberg, *Abstracts*, 148th National Meeting, American Chemical Society, Chicago, Illinois, August 1964.

[12] Leder, P., M. F. Singer, and R. L. C. Brimacombe, submitted to *Biochemistry*.

[13] Kaji, H., and A. Kaji, these PROCEEDINGS, **52**, 1541 (1964).

[14] Yanofsky, C., in *Synthesis and Structure of Macromolecules*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 28 (1963), p. 581.

[15] Matthaei, J. H., H. Kleinkauf, and G. Schramm, *Angew. Chem.*, **76**, 717 (1964); *Angew. Chem. Intern. Ed. Engl.*, **3**, 590 (1964).

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

[17] Brenner, S., A. O. W. Stretton, and S. Kaplan, *Nature*, in press.

[18] Weigert, M. C., and A. Garen, in press.

[19] Wittmann, H. G., and B. Wittmann-Liebold, in *Synthesis and Structure of Macromolecules*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 28 (1963), p. 589.

[20] Tsugita, A., personal communication.

[21] Jones, O. W., and M. W. Nirenberg, these PROCEEDINGS, **48**, 2115 (1962).

[22] Nirenberg, M. W., and O. W. Jones, in *Symposium on Informational Macromolecules*, ed. H. Vogel, V. Bryson, and J. Lampen (New York: Academic Press, 1963), p. 451.

[23] Speyer, J. F., P. Lengyel, C. Basilio, A. J. Wahba, R. S. Gardner, and S. Ochoa, in *Synthesis and Structure of Macromolecules*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 28 (1963), p. 559.

[24] Nirenberg, M. W., O. W. Jones, P. Leder, B. F. C. Clark, W. S. Sly, and S. Pestka, in *Synthesis and Structure of Macromolecules*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 28 (1963), p. 549.

[25] Woese, C., *Nature*, **194**, 1114 (1962).

[26] Eck, R. V., *Science*, **140**, 477 (1963).

[27] Woese, C. R., *ICSU Review of World Science*, **5**, 210 (1963).

[28] Weinstein, I. B., in *Synthesis and Structure of Macromolecules*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 28 (1963), p. 579.

[29] Ames, B. N., and P. E. Hartman, in *Synthesis and Structure of Macromolecules*, Cold Spring Harbor Symposia on Quantitative Biology, vol. 28 (1963), p. 349.

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