RNA CODEWORDS AND PROTEIN SYNTHESIS, VI. ON THE NUCLEOTIDE SEQUENCES OF DEGENERATE CODEWORD SETS FOR ISOLEUCINE, TYROSINE, ASPARAGINE, AND LYSINE

By Joel S. Trupin,* Fritz M. Rottman,† Richard L. C. Brimacombe, Philip Leder, Merton R. Bernfield, and Marshall W. Nirenberg

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

Communicated by C. B. Anfinsen, February 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 12 trinucleotides have been described and nucleotide sequences have been suggested for RNA codons corresponding to phenylalanine, valine, leucine, cysteine, proline, serine, and lysine.[1-4] In this report, the template activities of seven additional trinucleotides are described. ApUpU and ApUpC serve as RNA codons for isoleucine, UpApU and UpApC for tyrosine, ApApU and ApApC for asparagine, and ApApG and ApApA for lysine (cf. ref. 1). These findings are discussed in terms of the recognition of specific bases at different positions within a trinucleotide.

Materials and Methods.—Components of reactions: *E. coli* W 3100 ribosomes and sRNA were prepared by modifications of methods described previously.[5-7] Each C^{14}-aminoacyl-sRNA was prepared in the presence of 19 C^{12}-amino acids.

The assay for ribosomal bound C^{14}-aminoacyl-sRNA has been described.[1] Each 50-μl reaction mixture contained 0.10 *M* Tris-acetate, pH 7.2; 0.05 *M* KCl; 0.03 *M* magnesium acetate; 1.5-2.0 A^{260} units of washed *E. coli* W 3100 ribosomes; and the amount of C^{14}-AA-sRNA indicated in Table 1. Other data concerning each C^{14}-AA-sRNA preparation are also given in Table 1.

Uniformly labeled C^{14}-amino acids were purchased from New England Nuclear Corp. or Nuclear-Chicago Corp. Radioactivity was determined in a liquid scintillation counter (Packard Inst. Co.) with a C^{14}-counting efficiency of 55–65% as described previously.[1]

Trinucleotides: A derivative of bovine pancreatic ribonuclease was used to synthesize UpApC.[8] All other trinucleotides were prepared with a highly purified preparation of polynucleotide phosphorylase.[9] The purity of each trinucleotide preparation was assessed as described previously.[1-4, 9] The UpApC preparation contained approximately 8% cytidine-2′(3′)-phosphate. Other trinucleotide preparations appeared to be homogeneous. Analyses of the ApApA preparation have been reported.[1] The base ratio, base sequence, and chain length of each oligonucleotide were established as shown in Table 2. The methodology employed has been described previously.[1-4, 9]

Results and Discussion.—In Table 3 are shown the effects of eight trinucleotides upon the binding to ribosomes of 17 C^{14}-AA-sRNA preparations, each acylated with a different C^{14}-amino acid (radioactive Cys-, Met-, and Tryp-sRNA were not used). Each trinucleotide markedly stimulated the binding to ribosomes of only one C^{14}-AA-sRNA preparation. ApUpU and ApUpC stimulated C^{14}-Ileu-sRNA binding to ribosomes; UpApU and UpApC stimulated C^{14}-Tyr-sRNA binding; ApApC and ApApU stimulated C^{14}-Asp-NH_2-sRNA binding; and ApApA and

TABLE 1

C^{14}-AMINOACYL-sRNA PREPARATIONS

C^{14}-L-amino acid	Specific radioactivity (μcuries/μmole)	C^{14}-AA-sRNA Added to Each Reaction: A^{260} units	$\mu\mu$Moles of C^{14}-amino acid accepted	Origin of sRNA* *E. coli* strain
Ileu	240	0.26	17.3	W3100
Tyr	360	0.98	5.8	W3100
Asp-NH_2	29.8	2.50	15.3	W3100
Lys	240	0.25	8.5	W3100
Arg	193	0.34	17.8	W3100
Asp	106	1.17	8.4	W3100
Ala—Expt. 1	88	0.33	8.5	W3100
Expt. 2	88	0.75	32.6	W3100
Glu	205	0.56	14.9	B
Glu-NH_2	32.2	0.84	26.4	B
Gly	66	0.47	9.2	W3100
His	240	0.62	5.5	B
Leu	160	0.21	20.4	W3100
Phe	333	0.53	9.5	W3100
Pro	200	0.70	6.4	B
Ser—Expt. 1	120	1.03	6.3	W3100
Expt. 2	120	0.42	18.6	B
Thr	165	0.43	14.6	W3100
Val	205	0.26	17.6	W3100

* *E. coli* W3100 is a K-12 strain. Aminoacyl-sRNA synthetase preparations were from *E. coli* W3100.

TABLE 2

TRINUCLEOTIDE CHARACTERIZATION

Trinucleotide	Digestion* Enzyme†	Products	Base ratio
ApUpU	T_2	Ap, Up, U	1.0/0.9/1.0
	SVD	A, pU	1.0/2.0
ApUpC	T_2	Ap, Up, C	1.0/0.8/1.0
	SVD	A, pU, pC	1.1/0.9/1.0
UpApU	T_2	Up, Ap, U	1.0/1.0/1.0
	SVD	U, pA, pU	1.0/1.0/1.0
UpApC	T_2	Up, Ap, C	1.0/1.0/1.0
	SVD	U, pA, pC	1.0/0.9/1.2
ApApU	T_2	Ap, U	2.0/0.9
	SVD	A, pA, pU	1.0/1.0/1.0
ApApC	T_2	Ap, C	2.0/0.9
	SVD	A, pA, pC	1.0/1.0/1.0
ApApG	T_2	Ap, G	2.0/1.0
	SVD	A, pA, pG	1.0/0.9/1.1

* Procedures have been reported elsewhere.[1-4, 9]

† T_2 (T_2 RNase) was generously provided by Dr. George Rushizky. SVD (*Crotalus adamanteus* venom phosphodiesterase) was purchased from Worthington Biochemical Corp. and was further purified to remove phosphomonoesterase activity.[20]

ApApG stimulated the binding of C^{14}-Lys-sRNA. The template activity of the first trinucleotide indicated for each amino acid was higher than that of the second. Although the specificity of trinucleotides for C^{14}-AA-sRNA appeared to be high, it seems probable that slight stimulations, such as the effect of ApUpC upon C^{14}-Tyr-sRNA binding, may represent partial recognition of a trinucleotide, by perhaps 2 out of 3 bases (cf. refs. 2 and 4).

As noted previously,[1] a trinucleotide which stimulated the binding to ribosomes of one C^{14}-AA-sRNA generally decreased background binding of other C^{14}-AA-sRNA preparations. Although the effect of tri- and polynucleotides upon background binding is not fully understood, it is possible that ribosomal binding sites may be occupied nonspecifically by sRNA if template RNA is not present, whereas in the

TABLE 3

TEMPLATE SPECIFICITY OF TRINUCLEOTIDES FOR C^{14}-AMINOACYL-sRNA

The specificity of trinucleotides in directing the binding of C^{14}-aminoacyl-sRNA to ribosomes. Reactions contained the components described under *Materials and Methods*, the amount of C^{14}-aminoacyl-sRNA stated in Table 1, and 0.15 A^{260} units of trinucleotide, as specified, in a final volume of 50 μl. C^{14}-Asp-NH_2-sRNA was assayed in 100-μl reactions; amounts of all components were doubled (the amount of sRNA used in 100-μl reactions is shown in Table 1). Marked stimulations due to the addition of trinucleotides are underlined.

C^{14}-Amino-acyl-sRNA added	Δ $\mu\mu$Moles C^{14}-AA-sRNA Bound Due to Addition of Trinucleotides*								$\mu\mu$Moles C^{14}-AA-sRNA bound minus trinucleotide*
	ApUpU	ApUpC	UpApU	UpApC	ApApU	ApApC	ApApA	ApApG	
Ileu	0.72	0.59	0.01	−0.04	−0.04	−0.01	−0.04	−0.02	0.10
Tyr	−0.05	0.07	0.81	0.56	−0.03	0.05	0.02	−0.03	0.32
Asp-NH_2	−0.01	−0.03	−0.02	0.02	1.19	1.50	0.02	0.02	0.19
Lys	−0.04	0.02	−0.09	−0.06	0.03	−0.12	1.77	1.00	0.70
Arg	−0.01	0.09	−0.41	−0.37	−0.02	−0.03	−0.02	−0.17	1.42
Asp	−0.01	−0.07	−0.04	−0.02	0	0.02	−0.02	−0.01	0.31
Ala—Expt. 1	−0.01	0	−0.02	−0.01	−0.05	—	—	−0.05	0.20
Expt. 2	—	—	—	—	—	−0.19	−0.23	—	0.53
Glu	0	−0.01	0	−0.03	0	−0.05	−0.05	0.01	0.17
Glu-NH_2	−0.08	−0.03	−0.11	−0.16	−0.03	0.05	−0.14	0.03	1.65
Gly	−0.34	−0.15	−0.23	−0.78	−0.24	−0.16	−0.11	−0.02	1.10
His	−0.02	0	−0.04	−0.04	0.02	−0.02	−0.03	−0.03	0.24
Leu	−0.18	−0.20	−0.24	−0.21	−0.25	−0.20	−0.43	−0.26	1.03
Phe	0.03	−0.10	−0.27	−0.23	−0.34	−0.05	−0.33	−0.19	0.48
Pro	−0.02	0.03	0	0.02	−0.02	−0.02	−0.04	−0.02	0.14
Ser—Expt. 1	−0.02	−0.02	−0.02	0	−0.08	—	—	−0.03	0.24
Expt. 2	—	—	—	—	—	−0.21	−0.19	—	0.94
Thr	−0.02	−0.06	−0.05	−0.11	0	−0.02	−0.04	−0.02	0.44
Val	−0.01	0.01	0.03	0	−0.03	0.01	−0.04	−0.01	0.25

* Background binding of C^{14}-aminoacyl-sRNA to ribosomes in the absence of trinucleotides is expressed in $\mu\mu$moles. All other values (Δ $\mu\mu$moles) are obtained by subtracting background binding of C^{14}-aminoacyl-sRNA from binding obtained upon addition of a trinucleotide preparation.

presence of an RNA template, such sites may be occupied by sRNA corresponding to the template.

Background binding observed with certain C^{14}-AA-sRNA preparations was higher than with others. As discussed elsewhere,[1-4, 7] the amount of background binding 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 (deacylated sRNA also binds to ribosomes in response to RNA templates[10]), and the amount of template RNA on the ribosomes or in the sRNA preparations.

The specificity of each trinucleotide for C^{14}-AA-sRNA suggests that the sequences ApUpU and ApUpC serve as RNA codons for isoleucine, UpApU and UpApC for tyrosine; ApApC and ApApU for asparagine; and ApApA and ApApG for lysine. These results agree well with estimates of base compositions of RNA codons obtained previously with randomly ordered synthetic polynucleotides and cell-free protein synthesizing systems.[11,12] Such estimates were as follows: isoleucine, (AUU) and (AAU); tyrosine, (AUU) and (ACU); asparagine, (AAC), (AAU), and (ACU); lysine, (AAU) and (AAA).

Thus far, template functions of 19 of the 64 trinucleotide sequences have been investigated in this system. The base sequences of other RNA codons can be predicted, as shown in Table 4, by utilizing the base sequences suggested for RNA codons, the general pattern of degeneracy which has been observed, and amino acid replacement data reported for *E. coli*[13] and TMV mutants.[14, 15] The limitations of the predictions deserve to be emphasized, for codon recognition is subject to modification, and 5′-terminal, 3′-terminal, and internal codons probably are recognized in somewhat different ways. (For example, it is possible that CpUpU and CpUpC resemble nonsense terminal codons but serve as Leu-codons at internal positions.[4])

Studies with randomly ordered polynucleotides and cell-free protein synthesizing

TABLE 4

PREDICTED NUCLEOTIDE SEQUENCES OF RNA CODONS

The template activities of the trinucleotides marked with an asterisk (*) have been studied experimentally in *E. coli* W3100 extracts. Trinucleotides with the same sequence in the first and second positions and either U or C as the 3′-terminal nucleoside (for example, XpYpU and XpYpC) are arranged in pairs, each triplet in a pair corresponding to the indicated amino acid. Trinucleotides with 3′-terminal A or G are also arranged in pairs; however, with the exception of ApApA and ApApG, predictions need not apply to both members of the pair. Amino acid replacement data used for these predictions were obtained with *E. coli* by Yanofsky,[13] or were induced by HNO_2 in TMV by Wittmann and Wittmann-Liebold[14] or Tsugita.[15]

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

systems have shown that multiple codons corresponding to one amino acid often differ in base composition by only one base.[11, 12, 16, 17] As reported previously, trinucleotide pairs containing identical bases in the first and second positions and U or C 3′-terminal nucleosides (such as XpYpU and XpYpC) correspond to the same amino acid.[4] This general pattern of degeneracy has been found with every trinucleotide with 3′-terminal U or C examined (14 of the 32 possible trinucleotides[4]). It should be noted that Woese, on the basis of the relatively few codon assignments then available, predicted a code in which A, C, G, and U would be independently recognized at one position in a triplet, C = U at a second position, and A = C and G = U at a third position.[18] A modification of this type of code was suggested by Eck in which U = C and A = G at an unspecified position in a triplet.[19]

Although U and C clearly are similar at 3′-terminal positions, the trinucleotides ApApA and ApApG represent the first pair with 3′-terminal A and G which we have studied. Since both trinucleotides stimulate the binding of C^{14}-Lys-sRNA to ribosomes, A and G appear to be functionally similar at 3′-terminal positions and are distinguished from U and C during codon recognition. *However, the template activity of ApApA for C^{14}-Lys-sRNA was considerably higher than that of ApApG, especially when lower concentrations of trinucleotides were used (data not shown).* Further studies are necessary to determine whether each member of other trinucleotide pairs containing A or G 3′-terminal nucleosides also corresponds to the same amino acid. Since the available evidence suggests that one molecule of Phe-sRNA may recognize two codons (UpUpU and UpUpC),[4] experiments also are in progress to determine whether ApApA and ApApG are recognized by different species of Lys-sRNA.

Summary.—To investigate nucleotide sequences of RNA codons, the trinucleotides ApApU, ApApC, ApApA, ApApG, ApUpU, ApUpC, UpApU, and UpApC were used as templates to direct the binding of C^{14}-aminoacyl-sRNA to ribosomes. The results suggest that ApApU and ApApC serve as RNA codons for asparagine; ApApG and ApApA, for lysine; ApUpU and ApUpC, for isoleucine; and UpApU and UpApC, for tyrosine. Thus, two types of synonym codon pairs were observed: (1) synonym codon pairs such as XpYpU and XpYpC, and (2) a synonym pair such as XpYpA and XpYpG.

It is a pleasure to thank Miss Norma Zabriskie, Mrs. Theresa Caryk, and Mr. Taysir M. Jaouni for their invaluable assistance.

The following abbreviations are used: Ala-, alanine-, Arg-, arginine-, Asp-NH_2-, asparagine-, Asp-, aspartic acid-, Cys-, cysteine-, Glu-, glutamic acid-, Glu-NH_2-, glutamine-, Gly-, glycine-, His-, histidine-, Ileu-, isoleucine-, Leu-, leucine-, Lys-, lysine-, Met-, methionine-, Phe-, phenylalanine-, Pro-, proline-, Ser-, serine-, Thr-, threonine-, Tryp-, tryptophan-, Tyr-, tyrosine-, and Val-, valine-sRNA; AA-sRNA, aminoacyl-sRNA; sRNA, transfer RNA; U, uridine; C, cytidine; A, adenosine; G, guanosine. For brevity, trinucleoside diphosphates are referred to as trinucleotides. Internal phosphates of trinucleotides are (3′, 5′)-phosphodiester linkages. Trinucleotides whose nucleoside initials are enclosed within parentheses are of unspecified sequence.

* Supported by American Cancer Society postdoctoral fellowship PF 201.

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

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

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

[3] *Ibid.*, p. 1521.

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

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

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

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

[8] Bernfield, M. R., and M. W. Nirenberg, *Abstracts*, 148th National Meeting, American Chemical Society, Chicago, Illinois, August 1964, p. 35c.

[9] Leder, P., M. F. Singer, and R. L. C. Brimacombe, *Biochemistry* (in press).

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

[11] 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.

[12] 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.

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

[14] 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.

[15] Tsugita, A., personal communication.

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

[17] 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.

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

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

[20] Keller, E. B., *Biochem. Biophys. Res. Commun.*, **17**, 412 (1964).