ALTERNATIVE CODES AND TEMPLATES

By Richard B. Roberts

DEPARTMENT OF TERRESTRIAL MAGNETISM, CARNEGIE INSTITUTION OF WASHINGTON, D. C.

Communicated April 18, 1962

The problem of deducing the "biological code" from the composition of various natural RNA's and their product proteins is difficult if not impossible. The converse problem of calculating the composition of a template from the known composition of its protein product according to an assumed code requires only a few minutes. Such a calculation to determine the composition of the template RNA which would be required to direct the synthesis of bacterial proteins according to recently published triplet codes[1-7] is given in Table 1. This hypothetical RNA has a high uracil content (45%) and is quite unlike any natural RNA.

On the other hand, the triplet code can be converted to a doublet by discarding the U common to all of the code words. A calculation of the hypothetical template RNA on this basis shows a composition which resembles ribosomal RNA. Table 2 shows a comparison of the template material (calculated according to the doublet code) with observed RNA's of several bacteria in which the proteins[8] and RNA[9] have been carefully analyzed. In all cases, the calculated template is similar to the 50S ribosomal RNA. No exact agreement can be expected since the efficiencies of different templates may vary.

The doublet code obtained by discarding the U common to the triplet words is shown in Table 3. The order is based on the amino acid replacement data.[6, 7] The complete use of all the possible doublet symbols is perhaps significant. In writing out this code, the apparent degeneracies were omitted. These can be attributed to errors in the attachment of certain amino acids to the S-RNA which acts as carrier in the cell-free system.[10] Leucine, for example, might be attached to the carriers for valine and isoleucine. These three amino acids also confuse the entry mechanism of the cell.[11-13]

In several respects, the two codes are equivalent. They predict the same incorporation of amino acids relative to phenylalanine, as the ratio XUU/UUU is the same as XU/UU. They are equally satisfactory in fitting the amino acid replacement data.

They differ markedly in predicting the stimulation of phenylalanine incorporation to be expected from synthetic polymers of reduced U content. The doublet code has the possible theoretical advantages that it uses fewer letters, contains no "nonsense," and agrees with the correlations found by Sueoka.[8]

The three-letter code has the serious failing that it predicts template RNA unlike any natural nucleic acid yet observed. This difficulty will be resolved if it turns out that the present symbols are only a small and little-used part of a highly degenerate code which includes many symbols lacking U.

The doublet code provides no reason that certain synthetic polymers (poly A, poly C, poly G, poly AC, poly GC, poly AG) do not act as templates. All of these possible combinations should provide sites. In this respect, it has the same flaw as the triplet code since symbols lacking U have not been detected.[6, 7] In both cases, the difficulty can be attributed to special properties of polyphenylalanine or poly U which are essential in cell-free systems.

TABLE 1

CALCULATION OF HYPOTHETICAL TEMPLATE

	Proportions in *E. coli* proteins*	Code	Common U	Proportion of Bases Expected in Template			
				U	G	A	C
Ala	14.7	UCG	14.7		14.7		14.7
Arg	7.0	UCG	7.0		7.0		7.0
Asp	14.4	UAG	14.4		14.4	14.4	
Cys	0.9	UUG	0.9	0.9	0.9		
Glu	15.6	UAG	15.6		15.6	15.6	
Gly	12.0	UGG	12.0		12.0		
His	2.9	UAC	2.9			2.9	2.9
Ileu	7.5	UUA	7.5	7.5		7.5	
Leu	12.4	UUC	12.4	12.4			12.4
Lys	8.5	UAA	8.5			8.5	
Met	4.3	UAG	4.3		4.3	4.3	
Phe	4.8	UUU	4.8	4.8			
Pro	5.8	UCC	5.8				5.8
Ser	6.5	UUC	6.5	6.5			6.5
Thr	7.8	UAC	7.8			7.8	7.8
Try	—	UGG	—				
Tyr	3.9	UUA	3.9	3.9		3.9	
Val	10.0	UUG	10.0	10.0	10.0		
			1390.	508.	909.	734.	629.
Composition of template (including common U)				45.1	21.8	17.6	15.1
Composition of template (excluding common U)				18.3	32.7	26.4	22.6
Composition of *E. coli* 50S ribosomal RNA†				19.6	33.5	25.4	21.5

* Data of Sueoka.[8]
† Data of Midgley.[9]

The doublet has the serious failing that it provides only 16 combinations. Perhaps asparagine and glutamine could be converted from aspartic acid and glutamic acid after incorporation, but it is difficult to extend this reasoning to the other ambiguities, methionine and tryptophan. An unlikely possibility is that unusual bases or missing bases provide a few needed code words. A more plausible escape from this dilemma lies in a mixed code which includes a few three-letter symbols. If, for example, the combinations AA and GG indicated the start of a three-letter word, there would be a sufficiency of combinations. The mixed code would also provide a mechanism which could occasionally produce the results of Crick *et al.*[14] Another interpretation is that the cell can distinguish two kinds of purine pairs (e.g. parallel and anti-parallel).

The finding that the hypothetical templates calculated according to the doublet code resemble ribosomal RNA raises another question. In the growing cell, newly formed RNA can be distinguished by chromatography or sedimentation.[15, 16] Roughly three per cent of the RNA is in this form, one per cent being DNA-like in composition and two per cent being ribosome-like.[9] Experiments with cell free systems suggest that the newly formed material is the most likely template,[1] but there is no evidence as to which component is active. In Table 2, the predicted template material shows a slight correlation with changes in the DNA composition but the correlation is less than would be expected if the DNA-like component were fully active as a protein-forming template. Thus, the ribosome-like component of the newly formed RNA would appear to act as the template for most of the cell's proteins.

This view cannot be ruled out at present. During one generation, the growing

TABLE 2

COMPARISON OF RNA'S

Organism	GC content of DNA (%)	Type of RNA	Mole Per cent U	G	A	C
B. subtilis	42	Template*	18.5	32.1	27.5	21.8
		50S†	19.3	32.0	26.5	22.5
		Newly synthesized†	23.7	27.3	25.5	23.5
E. coli	50	Template	18.3	32.7	26.4	22.6
		50S	19.6	33.5	25.4	21.5
		Newly synthesized	22.6	29.5	25.0	22.9
A. aerogenes	57	Template	18.1	33.4	25.2	23.2
		50S	21.2	31.2	25.6	22.0
		Newly synthesized	21.5	30.3	24.8	23.4
Ps. aeruginosa	65	Template	17.4	33.9	24.5	24.1
		50S	21.3	31.2	26.3	21.2
		Newly synthesized	20.5	31.9	21.4	26.2

* Hypothetical template calculated according to doublet code from amino acid analyses of Sueoka.[8]
† Observed RNA compositions (Midgley[9]).

TABLE 3

A DOUBLET CODE

		Second Letter U	C	G	A
First letter	U	phe	ser	cys	tyr
	C	leu	pro	arg	his
	G	val	ala	try	glu
				gly	glu N
	A	ileu	thr	asp	lys
				met	asp N

Letters based on amino acid incorporation; order based on amino acid replacements.

cell might make two RNA copies of all its DNA, thereby providing the observed rate of synthesis of the DNA-like RNA. Such a rate could be characteristic of the production of templates for uninduced (or repressed) enzymes or for RNA copies of nonstructural genes. This DNA-like RNA is degraded and reutilized to form stable nucleic acids,[17] possibly after serving as template for a small part of the protein.

At the same time, a limited group of DNA sites (perhaps 1,000 of a possible 15,000) could be on the average 30 times more active in synthesis, as these sites provide the templates for induced (or unrepressed) enzymes. Such selected material could well be different in composition from the average DNA; thus, there is no reason to eliminate ribosome-like RNA as possible template material on the basis of its composition. In fact it is difficult to visualize how the DNA-like RNA of highly variable composition could act as templates for proteins of relatively constant composition.

The fraction of ribosome-like compositions is ultimately incorporated into ribosomes but its lifetime is sufficient to let it serve as template for 20–40 polypeptide strands.[16] Thus, there is no kinetic evidence against its possible role as template.

At this time, it is not possible to choose with certainty among the alternatives presented, whether the code is triplet or mainly doublet, whether DNA-like or ribosome-like RNA, or both act as templates. As each alternative has advantages and failings, all deserve consideration until definitely eliminated.

[1] Nirenberg, M. W., and J. H. Matthaei, these PROCEEDINGS, **47**, 1588 (1961).
[2] Lengyel, P., J. F. Speyer, and S. Ochoa, these PROCEEDINGS, **47**, 1936 (1961).
[3] Speyer, S. F., P. Lengyel, C. Basilio, and S. Ochoa, these PROCEEDINGS, **48**, 63 (1962)

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

[5] Lengyel, P., J. F. Speyer, C. Basilio, and S. Ochoa, these PROCEEDINGS, **48**, 282 (1962).

[6] Speyer, J. F., P. Lengyel, C. Basilio, and S. Ochoa, these PROCEEDINGS, **48**, 441 (1962).

[7] Matthaei, J. H., O. W. Jones, R. G. Martin, and M. W. Nirenberg, these PROCEEDINGS, **48**, 666 (1962).

[8] Sueoka, N., these PROCEEDINGS, **47**, 1141 (1961).

[9] Midgley, J. E., *Carnegie Institution of Washington Year Book*, No. 60, 303 (1961), and personal communication (1962).

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

[11] Cohen, G. N., and H. B. Rickenberg, *Ann. Inst. Pasteur*, **91**, 693 (1956).

[12] Cohen, G. N., and J. Monod, *Bact. Rev.*, **21**, 169 (1957).

[13] Britten, R. J., and F. T. McClure, *Bact. Rev.* (in press).

[14] Crick, F. H. C., L. Barnett, S. Brenner, and R. J. Watts-Tobin, *Nature*, **192**, 1227 (1961).

[15] Gros, F., W. Gilbert, H. Hiatt, C. Kurland, R. W. Risebrough, and J. D. Watson, *Nature*, **190**, 581 (1961).

[16] McCarthy, B. J., R. J. Britten, and R. B. Roberts, *Biophys. J.*, **2**, 57 (1962).

[17] McCarthy, B. J., *Biophys. Soc. Abst.* (1962).