OBSERVATIONS ON THE POSSIBLE NATURE OF THE GENETIC CODE

Thomas H. Jukes

American Cyanamid Company, Agricultural Division, Princeton, New Jersey

Received December 20, 1962

Reports of recent experimental work (1,2,3,4) have described the formation of polypeptides in cell-free systems in which synthetic polyribonucleotides containing adenine, cytosine and guanine (A, C and G), but no uracil (U), were used as templates. As a result 18 triplet codes containing no U were proposed (4). Earlier work had indicated that 19 of the 20 amino acids were coded by 22 triplets containing U (5,6). If one U is subtracted from each of these 22 triplets, 16 "doublets" are formed; a single doublet may be assigned to each of 12 amino acids and the other 4 doublets are each referable to two different amino acids, as pointed out by Roberts (7). This type of distribution is to be expected in view of the fact that the assignments for 19 amino acids were divided among 22 different U-containing triplets, and such triplets must differ from each other largely on the basis of what remains after subtracting a U. A problem is posed by assigning two different amino acids to a single doublet. This problem is increased by the existence of more than one code for each of most of the amino acids. For this reason, a triplet code is preferred. It was suggested by Wahba et al. (4) that "shared doublets" could be written for certain cases in which similar codes could be assigned to the same amino acid, for example, codes of CAG, CCG and CUG were proposed for alanine with a "shared doublet" of C*G. A proposal is made in the present communication to extend this principle to all of the coding triplets, as follows: It is postulated that in each triplet there is a "pivotal" base which, in a number of cases, has been shown to be subject to change without altering the coding function

of the triplet. This base appears to be U in the case of 11 codes containing this base (4). It is suggested that the "pivotal" base is U in the case of all codes containing this base, because it is deduced from examining the single-amino-acid mutations (8) that it is the bases other than U that participate in most mutational changes. The exceptions to this are the codes containing a terminal U, for it is also deduced from examining the single-amino-acid mutations that it is not possible to change the terminal base in any code without changing the coding function (8). Therefore, the U in positions 1 or 2 is designated as pivotal in all codes containing a terminal U, all of which contain two U's. The codes containing two U's followed by a third letter were allocated on the basis of which of the first two U's appeared to participate in mutations, this results in *UA for ileu, U*C and *UC for leu, U*G for val and U*U for phe. The remaining codes were AAC, ACG, GAA, AGG and CGC, which do not occur as variants of a code containing U for the same amino acid. These were resolved by allocating the first or only A to the pivotal position; this procedure seemed most consistent with the concept that the"non-pivotal" bases are involved in mutations. CGC was allocated to *GC because *GC remained unallocated. The results of these procedures are in Table 1. The term "shared doublet" was revised to "modified doublet" in Table 1 to indicate a broader scope for this designation.

Some interesting observations arise from examining Table 1, as follows (a) neither of two bases not designated by * in the triplets can be changed without changing the coding function of the triplet. This includes the implication that the base in position 3 is essential to the coding function. This may be related to the observation that the triplets are "read" from right to left in protein synthesis (9) so that position 3 may be the initial point of complementary attachment for the transfer RNA molecule. (b) 32 of the described single-amino-acid mutations can correspond to single changes in the bases not designated by * in the

Table 1. Relations between "modified doublets" and the proposed messenger RNA triplet codes for the amino acids.

		*=A	*=C	*=G	*=U			*=A	*=C	*=G	*=U
A*A	lys	AAA			AUA	G*A	arg/asp	GAA	GCA		GUA
*AA	aspN		CAA		UAA	*GA	met				UGA
A*C	his		ACC		AUC	G*C	arg		GCC		GUC
*AC	gluN	AAC				*GC	thre		CGC		
A*G	glu	AAG			AUG	G*G	gly	GAG	GCG		GUG
*AG						*GG	gluN/try	AGG			UGG
A*U	tyr				AUU	G*U	cys				GUU
*AU	leu				UAU	*GU	leu				UGU
C*A	aspN				CUA	U*A					
*CA	thre	ACA			UCA	*UA	ileu				UUA
C*C	pro	CAC	CCC		CUC	U*C	leu				UUC
*CC	ser				UCC	*UC	leu				UUC
C*G	ala	CAG	CCG		CUG	U*G	val				UUG
*CG	ser	ACG				*UG					
C*U	ser				CUU	U*U	phe				UUU
*CU						*UU					

codes in Table 1. These are changes such as AAA/UAA, AUA/GUA, etc. These mutations, written for brevity in one direction only, are lys/aspN, lys/asp, lys/glu, aspN/thre, aspN/ala, aspN/ser, aspN/arg, his/tyr, his/arg, glu/ala, glu/gluN, glu/gly, glu/val, tyr/cys, tyr/phe, leu/ser, leu/ileu, leu/phe, thre/ser, thre/met, thre/ileu, pro/leu, pro/ser, ala/gly, ala/val, ser/phe, asp/gly, arg/gly, arg/lys, arg/ser, ileu/met, and ileu/val. The other 2 described mutations, gluN/val and asp/ala, do not correspond to single-base changes in the triplets in Table 1; the reason for this is not apparent. References to the mutations are listed in (8, 10, 11, 12). (c) The column of the table in which *=G contains no triplets. The triplets containing G all appear

in the other 3 columns, thus implying that G in a code cannot be changed to another base without changing the coding function. The meaning of this is not clear. It is also evident that no coding triplets containing GG in positions 1 and 2 have yet been described. The right-hand column includes all the triplets containing U, a reflection of the observation that no coding triplet ends in U unless a second U is also present (8). It is of interest that only 2 of the 32 "modified doublets" in Table 1 are "shared" by 2 amino acids.

The coding triplets in Table 1 differ from those proposed by Wahba et al. (4) in that AAU for ileu is omitted for reasons mentioned previously (8) and UCC is suggested for ser. There is an absence of experimental evidence (4) for continuing to include UCG, and the existence of CCC as a code for proline removes the arguments proposed for UCC for pro (8).

It seems more probable that UCC is ser. This would account for the somewhat high values for ser incorporation with copolymers of U and C (8) and would also provide for arg/ser as GCC/UCC (12). The series of changes at a single locus described by Yanofsky and Henning (12) may be written as UUG/CUG/AUG/GUG/GCG/GCC/UCC for val/ala/glu/gly/gly/arg/ser; and the HNO_2 mutation ser/leu (10, 11) as UCC/UUC.

If Table 1 represents a valid classification, it contains the implication that functions might be predicted for coding triplets not yet described, for example, if CCA is a code, it probably should code either for asparagine or for threonine. The proposal in Table 1 helps to place the coding triplets into a series and it indicates a relationship between the "non-pivotal" bases and most of the single-amino-acid mutations.

References

(1) Bretscher, A. and Grunberg-Manago, M., Nature, 195: 283 (1962)
(2) Jones, O. W. and Nirenberg, M. W., Proc. Nat. Acad. Sci., 48 (1962) (in the press)
(3) Gardner, R. S., Wahba, A. J., Basilio, C., Miller, R. S., Lengyel, P. and Speyer, J. F., Proc. Nat. Acad. Sci., 48 (1962) (in the press)
(4) Wahba, A. J., Gardner, R. S., Basilio, C., Miller, R. S., Speyer, J. F. and Lengyel, P., Proc. Nat. Acad. Sci., 49 (1962) (in the press)

(5) Speyer, J. F., Lengyel, P., Basilio, C. and Ochoa, S., Proc. Nat. Acad. Sci., 48:441 (1962)
(6) Matthaei, J. H., Jones, O. W., Martin, R. G. and Nirenberg, M. W., Proc. Nat. Acad. Sci., 48:660 (1962)
(7) Roberts, R. B., Proc. Nat. Acad. Sci., 48:1245 (1962)
(8) Jukes, T. H., Proc. Nat. Acad. Sci., 48:1809 (1962)
(9) Wahba, A. J., Basilio, C., Speyer, J. F., Lengyel, P., Miller, R. S. and Ochoa, S., Proc. Nat. Acad. Sci., 48:1683 (1962)
(10) Wittmann, H. G., Proceedings of Symposium on Informational Macromolecules, Rutgers-The State University, New Brunswick, N. J. (1962) (in the press)
(11) Tsugita, A., J. Mol. Biol., 5:284,293 (1962)
(12) Henning, U. and Yanofsky, C., Proc. Nat. Acad. Sci., 48:183,1497 (1962)