CODING TRIPLETS AND THEIR POSSIBLE EVOLUTIONARY IMPLICATIONS

Thomas H. Jukes

Space Sciences Laboratory, University of Calif., Berkeley, Calif.*

Received March 5, 1965

It has been proposed that mutational changes involving a single amino acid residue in proteins originate predominantly from single-base changes in messenger RNA, including viral RNA (1,2,3,4,5). The changes produce an alteration in the coding properties of a "triplet" formed by three consecutive bases, so that the net result is the change of an amino acid in a polypeptide chain. The single-amino-acid mutations, (table 1) were compared (6) with the messenger RNA coding triplets proposed by various investigators who have studied the effect of synthetic polyribonucleotides on the incorporation of amino-acids into polypeptides. (7,8,9). Most of these mutations correspond to single-base changes between two ordered coding triplets. Most of the triplets could be arranged into six possible series of sequences as first proposed for the triplets containing U (4). Later it was reported that AUU appeared to be the sequence in a code for tyrosine, and one of the six series was selected including AUU as a code for tyr and GUU for cys (5). It was proposed that in a number of cases the 2 or 3 alternate codes for an amino acid contained homologous sequences with the same pair of bases occupying the same sites in 2 of the 3 loci in each triplet (6,7).

The use of sequentially characterized trinucleotides in a system for binding specific sRNAs to ribosomes in vitro was reported (10,11). The findings were not concordant with the suggestion that AUU and GUU are the codes for tyr and cys respectively. Instead it was reported

* Supported by grant NsG 479 to the University of California by the National Aeronautics and Space Administration.

Table 1 Single-amino-acid Mutations in Proteins

Amino Acid Change	Protein and Location of Residue, if known	Reference
Ala/val	Lgb,AP	13,14
Val/ala	TS	15
Arg/gly	TMV 46,61,122,134	16
Gly/arg	TS	15
Arg/lys	TMV 46	17
Arg/ser	TS	15
Arg/thr	TS	15
AsN/arg	TMV 33	18
AsN/asp	TMV 126	18
Asp/asN	Hb β79	19
AsN/lys	Hb α68, TMV 140	18,20
AsN/ser	TMV 25,33,73,126	17,18
Asp/ala	TMV 19 (?)	18
Asp/gly	TMV 66,Lgb	13,18
Gly/asp	Hb α22, α57, β16, TS	15,21,22,23
GIN/arg	Hb α54, TMV 99	17,24
Glu/ala	TS	15,37
Glu/gIN	Hb α30, β121	25,26
Glu/gly	Hb β7, TMV 97	18,27
Gly/glu	TS	15
Glu/lys	Hb α116, β6, β7, β26, β121	27-31
Glu/val	Hb β6, TMV 22*(?), TS	15,18,32,33
Val/glu	Hb β67	23
Gly/ala	TMV	34
Gly/val	TS	15
His/arg	Hb β63	35
His/gIN	Lgb	36
His/tyr	Hb α58,87 β63	38,39,40
Ilu/met	TMV	18
Ilu/thr	TMV 21,129	16,17,18
Thr/ilu	TMV 5,59,TS	15,18
Ilu/val	TMV 21,24,125,129	17,18
Leu/arg	TS	15
Leu/phe	TMV	18
Lys/asp	Hb α16	41
Met/leu	Cyt.c	42
Pro/leu	TMV 20,156	18,34
Pro/ser	TMV 63	18
Pro/thr	TMV 20	17
Ser/gly	TMV 65	17
Ser/leu	TMV 55,TS	15,18
Ser/phe	TMV 138,148	16,17,18
Thr/ala	TMV 81	17
Thr/met	TMV 107	18
Tyr/cys	TS	15
Tyr/phe	TMV	33
Val/met	TMV 11	17

*Formerly gIN/val

AP, alkaline phosphatase, E. coli; Lgb, bovine βlactoglobulin; Cyt c; human cytochrome c; TS, tryptophan synthetase, E. coli; TMV, tobacco mosaic virus coat protein; Hb, hemoglobin.

that GUU=val, UGU=cys and UUG=leu, (10). These sequences occur if the table published previously (6), is rearranged so that each base formerly in position 3 is now placed in position 1 so that GUU=cys becomes UGU=cys; UUG=val becomes GUU=val, and so on. The revision is in Table 2, which also includes further experimental findings as follows: (a) UUC=phe, UCU and UCC=ser, CUU and CUC=leu (tentative) and CCU=pro (tentative) (12); (b) AAC is a code for asN (Wahba, A.J., et al, personal communication); (c) AUA is a code for ilu, and UAU for tyr, (Yamane, T., Cheng, T.Y., Grunberg-Manago, M., and Fresco, J.R., in the press); (d) AAG is a code for lys, GAA for glu, GUG for val, and CUC for leu (Khorana, et al, in the press).

The first investigations with synthetic polyribonucleotides and the amino acid code were with copolymers that were high in U. It was striking that 22 triplet codes each containing at least one U were obtained for 19 of the 20 amino acids. It was noted (6) that if one U was subtracted from each triplet to form "doublets", one doublet could be assigned to each of 12 amino acids. The subsequent discovery of many codes not containing U (8,9) tended to divert attention from this. The 16 possible doublets are in column (b) of Table 2. Assignments proposed for them, on the basis of triplets in column (c), are in column (a). The triplets are assigned as a result of direct experimental evidence (underlined triplets) or of inferences from mutations (table 1). According to this proposal, 15 different amino acids correspond to the 16 doublets. It is logical to assume that 5 of the amino acids (asN, glN, met, try, and tyr) are of later evolutionary origin than the other 15, since these 5 are formed biosynthetically from asp, glu, cys, ser, and phe respectively. Therefore AA is assigned to lys rather than asN, and UA to the necessary "gap" between protein molecules rather than to tyr, since it is speculated that the doublets in column (b) were an archetypal code for the 15 amino acids to which they

Table 2. Relations between base pairs and possible coding triplets.

(a) Amino acid	(b) Doublet ("Archetypal Code")	Triplets (c) from doublets	Triplets (d) "Secondary" Codes	(e) Transitional mutations from (b)
Lys	AA	AAA, AAG		Glu, asp
Thr	AC	ACU, ACA, ACC		Ala, Ilu
Asp	AG	AGU, AGC		Gly
Ilu	AU	AUU, AUA, AUC		Val
His	CA	CAU, CAC		Tyr, arg
Pro	CC	CCU, CCA, CCC		Ser, leu
Arg	CG	CGU, CGC	AGA	(Cys)
Leu	CU	CUU, CUC	UUA, UUG	Phe
Glu	GA	GAU, GAA		Gly
Ala	GC	GCU, GCA, GCC		Val
Gly	GG	GGU, GGA, GGC		
Val	GU	GUU, GUG		
gap	UA			
Ser	UC	UCU, UCC	GAC	Phe
Cys	UG	UGU		
Phe	UU	UUU, UUC		
AsN	-		AAU, AAC, UCA	
GlN	-		AGG, CAA, CUA	
Met	-		AUG	
Try	-		UGG	
Tyr	-		UAU, UAC	

correspond in column (a), and that the vanished forms of life which used this code contained proteins whose amino acid content was confined to these 15 amino acids, at least 11 of which may be of abiogenic origin (43). The change from a doublet code to a triplet code is suggested as an evolutionary event which led to the vestigial survival of the remnants of the doublet code indicated by the triplets shown in column (c). This event was marked by the appearance of coding functions for the triplets in columns (c) and (d). The triplets in column (c) consist of the doublet plus another base in the third position, and these triplets retained the coding function which is assumed to have belonged to the doublet. The third base varies from amino acid to amino acid without any definite pattern: for phe the third base is U or C (but not A or G); for lys it is A or G; for Ilu it is A,C or U (but not G), and so on. Certain unassigned triplets, such as UGC (a

possible code for cys), may also belong in this column. Column (d) contains triplets which are presumed to have acquired their coding function later, including the codes for the 5 "new" amino acids. For want of a better name, these are termed "secondary" codes.

The 15 doublets assigned to amino acids in column (b) are convertible by transitional single-base changed (A⇋G and C⇋U) into 15 other doublets; 14 of these changes correspond to mutations in Table 1. The fact that the exception, arg/cys, is so far undescribed may be due to the rarity of mutations involving cys; only one has been recorded.

REFERENCES

1. Speyer, J.F., Lengyel, P., Basilio, C., & Ochoa, S., Proc. Nat. Acad. Sci. 48:63 (1962).
2. Wittmann, H.C., Naturwissenschaft, 48:729 (1961).
3. Tsugita, A., & Fraenkel-Conrat, H., Proc. Nat. Acad. Sci., 46:636 (1960).
4. Smith, E.L., Proc. Nat. Acad. Sci., 48:859 (1962).
5. Jukes, T.H., Biochem. Biophys. Res. Comm., 7:497 (1962).
6. Jukes, T.H., Biochem. Biophys. Res. Comm., 10:155 (1963).
7. Wahba, A.M., Gardner, R.S., Basilio, C., Miller, R.S., Speyer, J.F., & Lengyel, P., Proc. Nat. Acad. Sci. 49, 116 (1963).
8. Nirenberg, M.W., Jones, O.W., Leder, P., Clark, B.F.C., Sly, W.S., & Pestka, S.: Cold Spring Harbor Symp. Quant. Biol., 28:549 (1963).
9. Speyer, J.F., Lengyel, P., Basilio, C. Wahba, A.J., Gardner, R.S., & Ochoa, S.: Cold Spring Harbor Symp. Quant. Biol. 28:559 (1963).
10. Leder, P. & Nirenberg, M.W.: Proc. Nat. Acad. Sci. 52:420, 1521 (1964).
11. Nirenberg, M., & Leder, P., Science 145:1399 (1964).
12. Bernfield, M.R. & Nirenberg, M.W., Science 147:479 (1965).
13. Gordon, W.G., Basch, J.J. & Kalan, E.B., J. Biol. Chem., 236:2900 (1961).
14. Garen, A., Levinthal, C., Rothman, F., "Deoxyribonucleic Acid," p. 190, N.Y. Macmillan (1961).
15. Yanofsky, C., in The Bacteria, Vol. 5, edited by Gunsalus, I.C., & Stanier, R.Y., (N.Y. Academic Press, 1964), p. 373.
16. Tsugita, A., J. Mol. Biol., 5:284 (1962).
17. Funatsu, G. & Fraenkel-Conrat, H., Biochemistry 3, 1356 (1964).
18. Wittmann, H.G., & Wittmann-Liebold, B., Cold Spring Harbor Symp. Quant. Biol. 28,589 (1963).
19. Lehmann, H., Beale, D., & Bio-Daku, F.S., Nature 203:363 (1964).
20. Baglioni, C. & Ingram, V.M., Biochim. et Biophys. Acta; 48,253 (1961).
21. Baglioni, C., J. Biol. Chem., 237:69 (1962).
22. Gottlieb, A.J., Restrepi, A., & Itano, H.A., Fed. Proc. 23:172 (1964).
23. Chernoff, A.I., & Perrillie, P.E., Biochem. Biophys. Res. Comm., 16, 368 (1964).
24. Hanadu, M. & Rucknagel, D., Biochem. Biophys. Res. Comm., 11:229 (1963).
25. Swenson, R.T., Hill, R.L., Lehmann, H., & Jim, R.T.S., J. Biol. Chem. 237:1517 (1962).
26. Baglioni, C., Biochim. et Biophys. Acta, 59:437 (1962).

27. Baglioni, C. & Lehmann, H., Nature, 196:229 (1962).
28. Hunt, J.A., & Ingram, V.M., Nature, 184,640 (1959).
29. Hill, R.L., & Schwartz, H.C., Nature, 184:641 (1959).
30. Hunt, J.A., & Ingram, V.M., Nature, 184:870 (1959).
31. Pierre, L.E., Rath, C.E., & McCoy, K., New England J. Med., 268:862 (1963).
32. Ingram, V.M., Nature, 180:326 (1957).
33. Tsugita, A., cited by Speyer, et al, Proc. Nat. Acad. Sci. 48:441 (1962).
34. Tsugita, A., & Fraenkel-Conrat, H. J. Mol. Biol., 4:73 (1962).
35. Muller, C.J., & Kingma , S., Biochim. Biophys. Acta, 50:595 (1961).
36. Kalan, E.B., Greenberg, R., Walter, M., & Gordon, W.C., Biochem. & Biophys. Res. Comm., 16:199 (1964).
37. Yanofsky, C., Cold Spring Harbor Symp. Quant. Biol. 28:581 (1963).
38. Gerald, P.A.& Efron, M.L., Proc. Nat. Acad. Sci., 47:1758 (1961).
39. Jones, R.T., Coleman, R.D., & Heller, P., Fed Proc. 23:173 (1964).
40. Ingram, V.M., In The Molecular Control of Cellular Activity (N.Y., McGraw-Hill, 1962, p. 179).
41. Murayama, M., Fed. Proc. 19:78 (1960).
42. Smith, E.L., Fed. Proc. (in the press) presented at Bay Area Enzyme Club, San Francisco, March, 1964.
43. Harada, K. and Fox, S.W., Nature, 201:335 (1964).