SPECULATIONS ON THE EVOLUTION OF THE GENETIC CODE II

HYMAN HARTMAN

Department of Surgery Children's Hospital Medical Center, 300 Longwood Avenue, Boston, Mass. 02115

(Received 8 March, 1978)

Abstract. An evolutionary scheme is postulated in which a primitive code, involving only guanine and cytosine, would code for glycine (GG), alanine (GC), arginine (CG) and proline (CC). From each of these amino acids and their codons, there evolves a family of related amino acids as the code expands. The four families are: (1) *alanine* valine, leucine, isoleucine, phenylalanine, tyrosine, methionine and tryptophane; (2) *proline*, threonine and serine; (3) *arginine*, lysine, and histidine; (4) *glycine*, serine, cysteine, glutamic acid, glutamine, aspartic acid and asparagine. Except for the glycine relation to glutamic acid and aspartic acid, all amino acids are related by chemical similarities in their side chains. Glycine not having a side chain would permit a more complex set of substitutions.

The vocabulary expansion model [1] for the evolution of the genetic code argues that a primitive code would involve a few amino acids and that subsequent evolution would enlarge the number of amino acids. There are two variants of this theory: The first variant argues that the sixty-four triplet codons would initially code for a small number of amino acids. The new amino acids entering the code would then share codons with related amino acids. The second variant argues that the original code would involve only two bases (e.g. (A, U) or (G, C)). Both amino acids and nucleotides would be involved in the expansion of the genetic code.

In a recent paper, an evolutionary scheme was proposed for the genetic code [2]. If the code table is written as in Figure 1, one observes that a (G,C) code (e.g. glycine (GG), alanine (GC), arginine (CG) and proline (CC)) would be a possible primitive code. If one furthermore assumes that the code evolved by vocabulary expansion from this simple code, one observes four distinct families.

The reason for choosing a GC code is that the original postulated code contained only structural amino acids (e.g. gly, ala, pro, arg) [2].

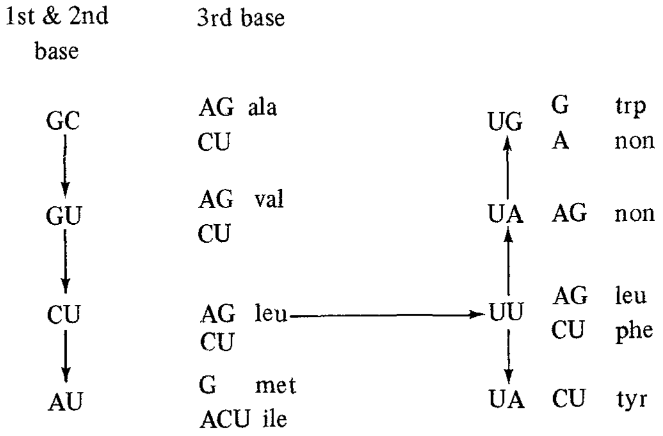
The considerations used in drawing the arrows were: (1) the codons must differ from each other by a single base change; (2) when amino acids differ from each other by a simple metabolic transformation, the unmodified amino acid came before the modified amino acid. Thus asp, glu, phe preceded asn, gln, tyr, respectively (3) Except for the case of glycine, all amino acids in a family must be related by chemical structure and function. Thus ala, val, leu, ile, phe are all related by having hydrophobic side chains; arg, hist, lys all have positively charged side chains; pro, thr, and ser are all neutral. (4) Finally met and trp are thought to be very late additions and hence merit special consideration.

Origins of Life 9 (1978) 133–136. *All Rights Reserved.*

1st	2nd				3rd
	G	C	A	U	
G	gly	ala	glu	val	G
	gly	ala	glu	val	A
	gly	ala	asp	val	C
	gly	ala	asp	val	U
C	arg	pro	gln	leu	G
	arg	pro	gln	leu	A
	arg	pro	his	leu	C
	arg	pro	his	leu	U
A	arg	thr	lys	met	G
	arg	thr	lys	ile	A
	ser	thr	asn	ile	C
	ser	thr	asn	ile	U
U	trp	ser	non	leu	G
	non	ser	non	leu	A
	cys	ser	tyr	phe	C
	cys	ser	tyr	phe	U

Figure 1.

The first family which we will consider is a family of hydrophobic amino acids which evolve from alanine



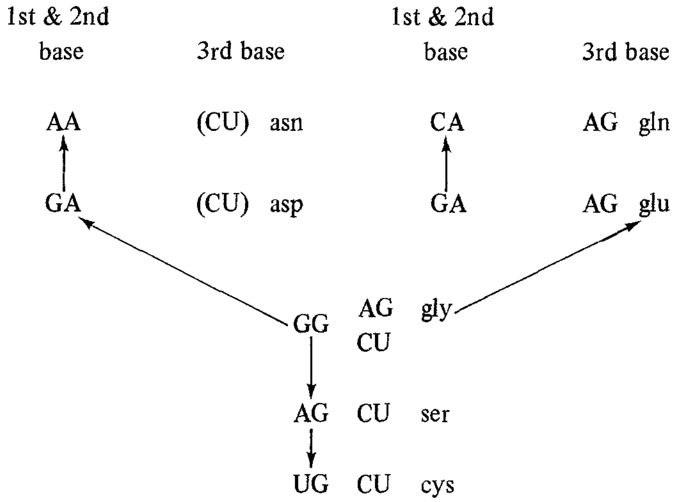
The second family of basic amino acids which evolve from arginine

1st & 2nd base	3rd base	
CA	(CU)	his
↑		
CG	(CU) (AG)	arg
↓		
AG	(AG)	arg
↓		
AA	(AG)	lys

The third family of neutral amino acids which evolve from proline

1st & 2nd base	3rd base	
CC	G A C U	pro
↓		
AC	G A C U	thr
↓		
UC	G A C U	ser

The final family is a complex family of amino acids which evolve from glycine



Glycine is unique in that it has no side chain. In an early polypeptide chain, especially if it is structural rather than globular, glycine can be substituted by unrelated amino acids (e.g. glut and asp) without a major disruption of the structure.

It is possibly this reason which allows glutamic and aspartic acids to enter the coding scheme in the glycine family.

The evolutionary scheme should be reflected in the recognition system. It has been observed that the transfer RNAs of the alanine family has an adenine at the discriminator site (3). The question of the recognition of t-RNA by the amino-acid synthetases is still unclear so that evidence for the other families must await further experimental clarification of this problem.

References

[1] Crick, F. H. C.: 1968, *J. Molec. Biol.* 38, 367.
[2] Hartman, H.: 1975, *Origins of Life* 6, 423.
[3] Crothers, D. M., Seno, T. and Soll, D.: 1972, *PNAS* 69, 3063.