THE DEGENERACY RULE OF GENETIC CODE

LIAO FU LUO
Physics Department, Inner Mongolia University, Huhehote, China

(Received 27 January, 1987)

Abstract. The degeneracy rules of genetic code including the distribution of terminators have been deduced through the minimization of mutational deterioration (MD). The MD of a given group of codons is divided into three parts: transitional, transversional and wobble's. The averaged mutational deteriorations (AMD) of various amino acids have been proved in order of their degrees of irreplaceability.

The problem about the logic of the genetic code has been put forward recently [1] . They tried to give an answer to the problem from a mathematical principle of extreme and proposed that the degeneracies of genetic code (synonyms) obey the rule to resist best to the effects of mutations which are the equivalent of noise or errors inherent to all information systems. We will call these effects of mutations as mutational deterioration (dangerousness, insecurity) or MD in this paper. From the optimal resistance to the effect of mutations they deduce the mutational neighbourhood of synonyms under the assumption that all point mutations are equiprobable. The proposal is foresighted and encouraging. However, there are still problems in their discussions. Firstly, given the degree of degeneracy, there are too many groups of codons satisfying the same requirement of optimization. For example, why the quarternary degenerate codons corresponding to an amino acid generally take the form of XYU, XYC, XYA, and XYG in the table of 64 triplets but neither in form of UXY + CXY + AXY + GXY nor in form of XUY + XCY + XAY + XGY? Secondly the maximum of resistance index B for a hexamerous degenerate amino acid has been calculated and equals 18 but serine does not follow this optimization and takes $B = 14$. Thirdly, three terminators in contemporary code are distributed in the table of codons with $B = 4$ while the theoretical maximum of B for any degenerate 3-triplets is proved to be 6. Why is the distribution of terminators different from other triple degenerate codons corresponding to an amino acid? It seems that the assumption of equiprobability of all point mutations is too simplified. In fact the S_4-symmetry of mutations is seriously broken [2]. At first, the rates of transition and transversion mutations are largely different with each other [3]. Secondly, the mutation of third letter of a codon differs from first and second ones. According to Crick's wobble assumption [4] the complementary codon-anticodon pairing in the first two nucleotides of a codon is unique and standard while the third pairing is not. The nonunique and nonstandard binding is equivalent to a kind of mutation of nucleotide to which we shall refer as 'wobble mutation'. Define the

Origins of Life and Evolution of the Biosphere **18** (1988) 65–70.

mutation effect caused by transitions as transitional MD denoted by u and the mutation effect caused by transversions as transversional MD denoted by v. Denote the additional wobble MD of third letter of sense codons as w. We shall calculate total MD for each ideal group of degenerate codons and assume that total MD takes minimization for practical distributions in the table of contemporary genetic code which may be viewed as a result of optimality in the evolution against the effects of mutation.

Consider an ideal group of double degenerate codons CAU + CAC. The total MD is

$$\mathrm{MD}(2) = c_2 - 2u - 2w, \tag{1}$$

where c_2 is a constant

$$c_2 = 6u + 12v + 6w. \tag{2}$$

Compare with other distributions such as

$$\mathrm{CAU} + \mathrm{CGU} \quad (\mathrm{MD} = c_2 - 2u)$$
$$\mathrm{CAC} + \mathrm{CAA} \quad (\mathrm{MD} = c_2 - 2v - 2w)$$
$$\mathrm{CAU} + \mathrm{AGU} \quad (\mathrm{MD} = c_2)$$

etc., one obtains minimum of MD given by Equation (1) under the assumption $u > v$. Thus the double degenerate codons CAU + CAC are optimal. In fact they correspond to histidine. In contemporary genetic code the nine groups of double degenete codons all have like distributions as CAU + CAC and have minimum MD(2) given by (1) as compared with all other possible groups of comparable size.

Next, consider an ideal group of quarternary degenerate codons GCU + GCC + GCA + GCG. The total MD is

$$\mathrm{MD}(4) = c_4 - 4u - 8v - 12w \tag{3}$$

with constant c_4

$$c_4 = 12u + 24v + 12w. \tag{4}$$

Compare with other ideal partners

$$\mathrm{GUU} + \mathrm{GCU} + \mathrm{GAU} + \mathrm{GGU} \quad (\mathrm{MD} = c_4 - 4u - 8v)$$
$$\mathrm{GCU} + \mathrm{GCC} + \mathrm{GUU} + \mathrm{GUC} \quad (\mathrm{MD} = c_4 - 8u - 4w)$$
$$\mathrm{GCU} + \mathrm{GCC} + \mathrm{ACU} + \mathrm{CCU} \quad (\mathrm{MD} = c_4 - 4u - 4v - 2w)$$

etc., one finds that under the assumption $2w > u - 2v$ the distribution GCU + GCC + GCA + GCG takes minimum MD(4) given by Equation (3) and it gives the optimal codons against the mutations. In fact this 4-triplets correspond to alanine. It can be easily shown that other quarternary degenerate codons corresponding to Val, Pro, Thr, and Gly have similar distributions as Ala and they all take the minimum of MD described by (3) as compared with all other possible groups of comparable size.

Similarly, the MDs of hexamerous degenerate codons – Leu, Arg and Ser can be calculated in the same way and it gives

$$\text{MD (Leu)} = c_6 - 10u - 8v - 14w \tag{5}$$

$$\text{MD (Arg)} = c_6 - 6u - 12v - 14w \tag{6}$$

$$\text{MD (Ser)} = c_6 - 6u - 8v - 14w \tag{7}$$

with

$$c_6 = 18u + 36v + 18w. \tag{8}$$

For comparison one may consider other ideal hexamerous degenerate codons. For example, the structure

$$\text{UUG} + \text{CUU} + \text{CUC} + \text{CUA} + \text{CUG} + \text{AUG}$$

has MD $= c_6 - 6u - 12v - 12w$, the structure

$$\text{UUC} + \text{UUA} + \text{UUG} + \text{CUC} + \text{CUA} + \text{CUG}$$

has MD $= c_6 - 10u - 8v - 12w$ and the structure

$$\text{CUA} + \text{CUG} + \text{UUA} + \text{UUG} + \text{CCA} + \text{CCG}$$

has MD $= c_6 - 14u - 6w$, etc. Under the assumption $w > 2u > 2v$ these ideal distributions in the table of code all have larger MD than Leu, Arg and Ser given by (5)–(7).

Finally we consider triple degenerate codons. Isoleucine has

$$\text{MD}(3) = c_3 - 2u - 4v - 6w \tag{9}$$

with constant

$$c_3 = 9u + 18v + 9w. \tag{10}$$

Evidently Equation (9) takes minimum as compared with other 3-triplets distributions, for example,

$$\text{AUU} + \text{AUC} + \text{ACU} \quad (\text{MD} = c_3 - 4u - 2w)$$

$$\text{AUU} + \text{AUC} + \text{CUU} \quad (\text{MD} = c_3 - 2u - 2v - 2w)$$

etc. if $2w > u - 2v$. On the other hand the three terminators constitute a group of triple degenerate codons. We shall compare MDs of different distributions of these three nonsense codons. The structure UAG + UGA + UAA has

$$\text{MD}(\text{T}3) = 5u + 18v = c_3 - 4u - 9w \tag{11}$$

Instead, the structure UAG + UAA + UAC which belongs to the same category as Ile has

$$\text{MD} = 7u + 14v = c_3 - 2u - 4v - 9w. \tag{12}$$

Under the assumption $u > 2v$ one can show that Equation (11) takes minimum and the corresponding group UAG + UGA + UAA constitutes terminators from the minimization of MD and the selection against the appearance of nonsense codons. Note that, different from 61 sense codons, the terminators are not related to t-RNA and therefore, in calculating MD, for example, in the deduction of (11) and (12), the wobble rate is not necessary to be taken into account. This is why isoleucine and terminators are distributed in different way in the present code. Another point is the importance of S_4-symmetry breaking of mutations for the location of terminators in the table. Because if $u = v$ then (11) is not smaller than (12) and UAG, UGA and UAA should not constitute a group of terminators [2].

In summary, the distribution of the accepted mutations among the three nucleotide positions of codons is non-random [5]. Under the assumption of $u > 2v$ and $w > 2u$ we have explained all degeneracy rules of genetic code, including the distribution of terminators, by the minimization of MD. It seems that the contemporary genetic code is optimal against the mutations [6].

However one comment should be made.We have explained the degeneracy rules of each group of degenerate codons, but the distribution of degenerate multiplets (i.e. different groups of codons) in the table has not been given. Especially for hexamerous degenerate codons, it is proved that the MDs of Leu, Arg and Ser are smaller than all other possible groups of comparable size, but Equation (5) to (7) can never give an equal result of MDs.

Set MD of a group of codons corresponding to a given amino acid divided by the number of codons (degeneracy degree). We call the quantity as averaged mutational deterioration (AMD) of this amino acid. The AMDs of amino acid residues are listed in decreased order as follows:

Trp	Met		$3u + 6v + 3w$
Cys	Tyr	His	
Phe	Gln	Lys	$2u + 6v + 2w$
Asn	Asp	Glu	
Ile			$^7/_3u + {}^{14}/_3v + w$
Ser			$2u + {}^{14}/_3v + {}^2/_3w$
Arg			$2u + 4v + {}^2/_3w$
Leu			$^4/_3u + {}^{14}/_3v + {}^2/_3w$
Pro	Thr	Gly	
	Val	Ala	$2u + 4v$ (13)

On the other hand, using the sequence data of hemoglobins a table of mutual replaceabilities of amino acids can be obtained. On the basis of this table the degree of irreplaceability of amino acid residue was established [7]. The results are listed in the following table:

Trp	1.82	Met	1.25	Cys	1.12	Tyr	0.98	
His	0.94	Phe	0.86	Gln	0.86	Lys	0.81	
Asn	0.79	Asp	0.77	Glu	0.76	Ile	0.65	
Ser	0.64	Pro	0.61	Arg	0.60	Leu	0.58	
Thr	0.56	Gly	0.56	Val	0.54	Ala	0.52	(14)

By comparison of (13) with (14) it is found that the order of AMD agrees well with the empirical data of the degree of irreplaceability of amino acid residues. The only exception is Pro which should be clarified through further statistical analysis of replaceability of amino acids and more detailed theoretical investigation about the symmetry breaking property of mutations of nucleotides. The degeneracies of MD in Equation (13) can also be removed by introducing further breakings of mutations between nucleotides, for example, (rate U $\leftrightarrow$ C) $\neq$ (rate A $\leftrightarrow$ G) etc. The agreement of AMD with degree of irreplaceability means that the rarer are substitutions for the residue, the higher is its value of AMD. It is not surprising. Because the degree of irreplaceability is a response to mutational deterioration in the course of evolution of life. It is suggested that the degree of irreplaceability characterizes the value of information. So AMD may play an important role in the measure of the value of information.

Finally we point out that the genetic code used above is a standard one. For mitochondria some ambiguities have been found [8]: AGA and AGG are terminators, UGA corresponds to Trp an AUA corresponds to Met, etc. If these data are confirmed by further observations, then the new groups of codons can be explained by like analysis as above. AUA + AUG and UGA + UGG are examples of double degenerate codons. They have minimum MD(2) given by (1). Four terminators constitute two groups of double degenerate nonsense codons UAA + UAG and AGA + AGG. Their MDs take minimum

$$\mathrm{MD(T2)} = 4u + 12v = c_2 - 2u - 6w \tag{15}$$

respectively. Of course the problem about why they form two distinct groups but not an single optimal group is still open and should be clarified by further works.

Acknowledgement

The author would like to thank Mrs. Minhuei Tsai for her encouragement, support and kindness.

References

1. Labouygues, J. M. and Figureau, A.: 1984, *Origins of Life* **14**, 685.
2. Luo, L. F. and Li, Q. Z.: 1985, *Science Bulletin* **30**, 1056 (in Chinese).
   Luo, L. F. and Li, Q. Z.: 1985, *The China-Japan Bilateral Symposium on Biophysics,* China Academic Publishers, p. 409.

3. Vogel, F.: 1972, *J. Mol. Evol.* **1**, 334.
4. Crick, F.: 1966, *J. Mol. Biol.* **19**, 548.
5. Fitch, W: 1973, *J. Mol. Evol.* **2**, 123.
6. Milanov *et al.*: 1986, *Origins of Life* **16**, 529.
7. Volkenstien, M. V.: 1982, *Physics and Biology*, Academic Press, Inc., p. 134.
8. Gautier, C. *et al.*: 1981, *Nucleic Acid Sequences Handbook,* Praeger, New York.