WHY TERMINATORS ARE UAA UAG AND UGA ?

L. F. Luo, and Q.Z. Li
Dept. of physics, Inner Mongolia University, Huhehote, China

According to the assumption that the contemporary code is selected by optimization against the occurence of nonsense codons we shall discuss why terminators are UAA UAG and UGA by the calculation of the number of nonsense mutations (S) and Hamming distance D(1) (Luo and Li,1985). The concept of S and D(1) had been introduced by Cullmann (Cullmann,1980) and Figureau et al.(Figureau and Labouygues ,1980). But the transition and transversion mutations have not been distinguished in these works. In fact, if there is rigorous permutation symmetry S_4 between bases U,C,A and G then the 64 codons are equipollent and it is impossible to find out the terminators from them. However, the S_4 symmetry between the frequencis of mutations is not valid rigorously. One may decompose the frequencies of single mutations into two parts— the part of S_4 symmetry and S_4 symmetry-breaking part. For the purpose of determination of terminators it is sufficient to consider the S_4 symmetry breaking mutation only. In the following we shall discuss the S_4 symmetry breaking transition alone since the frame shift mutation is S_4 symmetrical while the transversion may be put into the S_4 symmetrical part. Set **S** — the number of single nonsense transition mutations

$$S = \sum_{ijk} \left(m_{ijk}(1) + m_{ijk}(2) + m_{ijk}(3) \right) \qquad (1)$$

Here ijk is empty in the 4×4×4 matrix of genetic code which corresponds to a nonsense triplet, $m_{ijk}(1)$ is the number of sense codons the first letter of which can be transfereed to i by transition mutation and the 2nd and 3rd letter remain unchanged (jk), etc. Define Hamming's distance D(1) -- the number of pairs of sense codons which are related each other by a single transition mutation. By tedious enumeration of triplet codes in different 'dic--tionaries' one can show that (Luo and Li, 1985)

$$S = -2\Delta D(1) \qquad (2)$$

So the minimum of S corresponds to the maximum of D(1) under the shifting of the codes in the 'dictionary'.

From the requirement of minimization of S or maximization of $D(1)$ under the condition of 61 normal codes and 3 senseless ones, by comparing the number S of different cases it is proved that the minimum of S is $S_{min} = 5$ and that the code with nonsense codons UAA UAG and UGA has $S = 5$ exactly. Therefore under the pressure of S_4-breaking mutations the contemporary code is one of the best which resist against the appearance of terminators. However, there is degeneracy between codes with the same S_{min}. To eliminate the degeneracy one should assume further breaking of S_4 symmetry. Denote the frequencies of the single transition mutations by a_i, b_i, and c_i for first, second and third 'letter' in triplet codes and $i = 1,2,3,4$ —corresponding to the mutations $U \rightarrow C$, $C \rightarrow U$, $A \rightarrow G$ and $G \rightarrow A$ respectively. For example, the frequency of transition mutation of CAA to UAA is a_2, UGA to UAA is b_4, UAG to UAA is c_4 and the other codons to UAA is zero. Easily shown that

$$S = 3a_2 + b_4 + c_4 \qquad (3)$$

for present code. If $a_i \sim b_j \sim c_k \sim 1$ and $a_i < b_j$, $a_i < c_k$ $(i,j,k = 1 \cdots 4)$ (it is consistent with Crick's hypothesis of wobbles) and if a_2 is the smallest one in a_i, b_4 - the smallest in b_j and c_4 - the smallest in c_k in the period of formation of code, then the minimization of S would lead to the non--sense codens UAA, UAG and UGA unambiguosly.

Finally, why the number of nonsense codons is three ? We assume that under the influence of random mutation, the direction of evolution is to minimize the nonsense codons but guarantee the least erronous reading. Set the number of terminators P and consider the cases of $P = 1,2$ and 3. Easily shown that $S_{min} = 3$ for $P = 1$, $S_{min} = 2+2$ for $P = 2$ and $S_{min} = 2+2+1$ for $P=3$. The terminator with the number of single nonsense mutations $S=1$ is called precise terminator. Evidently, there is a precise terminator only in the case of $P=3$. If at least one precise terminator is required then $P=3$ is selected from the cases of oligo-terminators. In fact, for the present code, UAA is a precise terminator. The probability of erronous reading of it only 1- 5%. But for the other two, UGA and UAG, the probability of erronous reading is larger than 50 %.

References

Cullmann, C.G.: 1980, Proceedings of the 3rd ISSOL Meeting & the 6th ICOL Meeting (Jerusalem). 405.

Figureau, A. and Labouygues, J.: ibid. 431.

Luo, L.F. and Li, Q.Z.: 1985, Kexue Tongbao 30, 1056.