July 16-21, 2017 at UC San Diego, CA, USA

Direct Evidence for GC-NSF(a) Hypothesis on Creation of Entirely New Gene/Protein

Ryoko Oi[1]*, Kenji Ikehara[1,2]
[1]G&L Kyosei Institute, [2]International Institute for Advanced Sciences
* santentaichikujoukou@ybb.ne.jp

Introduction: One of the most fundamental problems, which remains unsolved in the fields of Biochemistry and Molecular Biology, is how entirely new (EntNew) protein or the first family protein was produced which is totally different from any previously exisiting proteins. On the other hand, we have proposed GC-NSF(a) hypothesis on formation of EntNew protein [1, 2]. The hypothesis assumes that EntNew protein was generated from non-stop-frame on antisense strand of, not AT-rich gene, but GC-rich gene (GC-NSF(a)). GC-NSF(a) is codon sequence in the same frame with the corresponding gene on sense strand.

Results and Discussion: It is quite important to get direct evidence that EntNew gene has been actually produced as expected by the hypothesis. For the purpose, every amino acid sequence (AAS) of imaginary protein encoded by GC-NSF(a) of *Pseudomonas aeruginosa* PAO1 genome (GC content=66.6%) was homology-searched against all AASs of extant proteins encoded by the same genome. However, some difficulties were anticipated, when evidence for the hypothesis is searched for, as described below.

(1) Probability of mis-annotation between sense and antisense sequences increases beyond about 60% GC content, because all three stop codons are AT- or AU-rich.
(2) Base sequence of a gene would frequently change without amino acid substitution of a protein encoded by the gene, because of degeneracy of the genetic code, which should induce amino acid substitution in imaginary protein encoded by GC-NSF(a) [3].
(3) Moreover, base sequence of immature EntNew gene would also rapidly change upon development of the immature protein to get a higher catalytic activity and to evolve into mature protein, from just after the gene was newly born.

Nevertheless, it was found that AAS encoded by GC-NSF(a) of *tal* gene encoding transaldolase B has sufficient homology with AAS encoded by *ftsZ* gene encoding cell division protein FtsZ, after the results obtained were cautiously examined and judged whether it is correct evidence or not. In addition, several results supporting for GC-NSF(a) hypothesis were also obtained with 56 bacteria genomes having more than 50% of GC content. Thus, we have concluded that EntNew gene encoding EntNew protein has been generated from GC-NSF(a), according to the GC-NSF(a) hypothesis.

References: [1] Ikehara K et al. (1996) *Nuclic Acids Research* 24: 4249-4255. [2] Ikehara K (2003) *Viva Origino* 31: 201–15. [3] Ikehara K.(2016) *SOJ Genetic Science* 3: 1-3.

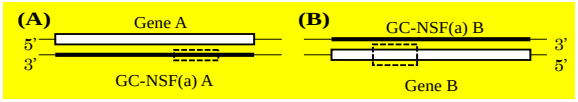

Figure 1. Amino acid sequence (A) of a protein encoded by GC-NSF(a) of a gene A showed sufficient homology with amino acid sequence（B）of an extant protein encoded by another gene B in the same genome. Therefore, it is concluded that (a part of) the gene was created from the GC-NSF(a), according to GC-NSF(a) hypothesis on formation of entirely new protein, which we have proposed.