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

Protein 0th-order Structure is encoded onto GC-NSF(a) base sequence

Kenji Ikehara[1,2], Ryoko Oi[1],
[1]G&L Kyosei Institute, [2]International Institute for Advanced Sciences
* ikehara@cc.nara-wu.ac.jp

Introduction: We have proposed [GADV]-protein world hypothesis (GADV hypothesis in short) on the origin of life, assuming that life originated from [GADV]-protein world, which was formed by pseudo-replication of [GADV]-protein. [GADV] means four amino acids; glycine [G], alanine [A], aspartic acid [D] and valine [V] [1, 2]. On the other hand, GC-NSF(a) hypothesis for entirely new (EntNew) gene formation in extant microorganisms triggered the discovery of the GADV hypothesis [1-3]. GC-NSF(a) is derived from non-stop frame on antisense strand of GC-rich gene. Therefore, it is quite important for both GC-NSF(a) hypothesis itself and GADV hypothesis to confirm whether EntNew gene is actually created from GC-NSF(a) or not. Oi and Ikehara have obtained direct evidence for the GC-NSF(a) hypothesis, as presented in Poster session in this conference.

Protein 0th-order structure on GC-NSF(a): We discuss on the reason why EntNew gene/protein is easily produced in this presentation, based on protein 0th-order structure or a specific amino acid composition, in which even random joining of amino acids produces water-soluble globular protein with slightly more flexible structure than extant or mature protein at a high probability. The reason is because protein 0th-order structure is written onto GC-NSF(a), as described below.

(1) A large number of different amino acid sequences, at least more than 10^{22}, could be encoded on one antisense strand of a GC-rich gene, because of degeneracy of the genetic code.
(2) Every amino acid sequence encoded by GC-NSF(a) is quite different from that of any previously existing proteins.
(3) One amino acid sequence encoded by a GC-NSF(a) can generate an extraordinary large number (more than 10^{24}) of different protein structures, owing to flexiblity of the protein. This means that a GC-NSF(a) carrying actually one amino acid sequence can encode substantially the large number of protein structures in protein 0th-order structure.
(4) Once a weak catalytic activity for a substrate was detected on a surface of one of the large number of protein structures, the immature protein gradually evolves to a mature enzyme, through introduction of necessary base replacements onto the GC-NSF(a) (Figure 1).
(5) Thus, a catalytic activity necessary to adapt for a new environment can be searched out from the protein 0th-order structure written onto one antisense sequence of a GC-rich gene.

We have considered that life emerged from [GADV]-protein world, which was established by pseudo-replication of [GADV]-protein in one of protein 0th-order structures, [GADV]-amino acids. Thus, the idea, protein 0th-order structure, is indispensable for understanding not only the origin of life but also evolution of fundamental life system composed of gene and protein.

References: [1] Ikehara K (2005) *Chemical Records* 5:107-118. [2] Ikehara K (2016) LAP LAMBERT Academic Publishing, Saarbrucken, Germany. [3] Ikehara K (2016) *Life (Basel)* 6: 6.

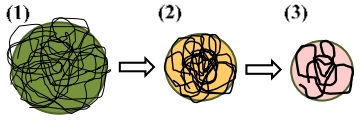

Figure 1. An immature protein (1) with a weak catalytic activity evolves gradually to a mature enzyme (3) with a higher activity and rigid structure through an intermediate (2), as accumulating necessary base replacements onto GC-NSF(a).