

Protein Ordered Sequences are Formed by Random Joining of Amino Acids in Protein 0th-Order Structure, Followed by Evolutionary Process

Kenji Ikehara

Received: 26 September 2014 / Accepted: 3 November 2014 /
Published online: 14 January 2015
© Springer Science+Business Media Dordrecht 2015

Abstract Only random processes should occur on the primitive Earth. In contrast, many ordered sequences are synthesized according to genetic information on the present Earth. In this communication, I have proposed an idea that protein 0th-order structures or specific amino acid compositions would mediate the transfer from random process to formation of ordered sequences, after formation of double-stranded genes.

Keywords Protein 0th-order structure · [GADV]-amino acids · [GADV]-protein · Water-soluble globular protein · From random synthesis to ordered sequence

Only random joining of monomeric molecules could be carried out under prebiotic conditions on the primitive Earth, and random polymer synthesis is highly unlikely to produce two stochastically identical chains. Therefore, my answer to the question, “How can we make by prebiotic means ordered sequences of amino acids, or mononucleotides?”, as asked in the Open Question 1, is that ordered sequences or two identical copies never produced by prebiotic means. And my answer to the question: “Do you agree then that we do not know how to make macromolecular sequences in many identical copies under prebiotic conditions?” is: “Yes. It is impossible to make many identical copies of macromolecular sequences under prebiotic conditions”.

On the other hand many identical proteins with ordered amino acid sequences are produced in the modern genetic system. Therefore, my answer to a question posed in the Open Question 1: “Do we have to wait for this orderly sequence until the genetic code has been developed?”, is that “Yes. We have to wait for the synthesis of ordered sequence until not only the genetic code but also gene, or the genetic system, has been formed”. Therefore, the Open Question 1 should be rewritten as follows: “How and when did the conversion occur from random joining of amino acids to the synthesis of proteins with ordered sequence?”

The key to understanding how ordered sequences, such as proteins, were formed starting from random polymerization of monomeric units, is to appreciate the significance of a specific

Paper presented at the conference Open Questions on the Origin of Life, July 12-13, 2014, International Institute for Advanced Studies (IIAS), Kizugawa, Kyoto, Japan.

K. Ikehara (✉)

Nara Study Center, The Open University of Japan, Nara 630-8589, Japan
e-mail: ikehara@cc.nara-wu.ac.jp

amino acid composition, or protein 0th-order structure (Ikehara 2002, 2005, 2009, 2012), in which water-soluble globular proteins with surface structure slightly more flexible than that of extant proteins can be produced even by random joining of amino acids. The reasons are as follows. A large number of possible catalytic sites with weak but sufficient activities could be formed on the surface of water-soluble globular proteins. Based on the quasi-equivalence rule of globular virus structure proposed by Casper and Klug (1962), the number of surface amino acids can be estimated at around 40 even for a small protein composed of 100 amino acids, and every combination of two or more surface amino acids on a globular protein could form an active catalytic site. The number of possible catalytic sites should exceed two hundreds even for a small protein. In addition, flexible surface structure would increase the chance of forming a catalytic center on the protein, since wobbling of the surface amino acids should make it possible for amino acids to adjust their positions such that they could bind a new substrate.

Thus, if a water-soluble protein with flexible surface amino acids were synthesized from antisense sequence (anti-SS) of ds-(GNC)_n RNA gene, and once a weak but sufficiently catalytic activity necessary for life were found on the protein, the immature protein could evolve to mature enzyme with an ordered amino acid sequence, high catalytic activity, high selectivity for the substrate and a more rigid surface structure. This would happen upon accumulation of appropriate base substitutions onto the anti-SS (evolutionary process). Where G, C and N in GNC mean guanine, cytosine and either of four bases (G, C, A (adenine) and U (uracil)), respectively. So, GNC represents four genetic codes, GUC, GCC, GAC and GGC, encoding [V], [A], [D] and [G], respectively. A number of mature proteins with identical amino acid sequences could be produced by expression of the newly-born (GNC)_n gene evolved from the anti-SS (Fig. 1). Together with the evolution of an anti-SS to a newly-born gene, the original gene on sense strand becomes a new anti-SS.

It might be questioned if the random synthesis in the protein 0th-order structure or [GADV]-amino acids can be carried out after the formation of ds-(GNC)_n genes and that the flexible surface structure that would inevitably reduce catalytic activity of the protein, should become an obstacle in creation of entirely new enzymes. These questions can be answered as follows: After formation of ds-(GNC)_n gene, the essentially random synthesis could be carried out by expression from anti-SS, which codes for [GADV]-protein with a quite different amino acid sequence from, but with a similar amino acid composition to [GADV]-protein encoded by sense sequence of the gene. Moreover, low catalytic activity would be sufficient if a given catalytic function were necessary for the emerging living system and could not be found in the environment. Therefore, they would not be obstacles in creating new enzymes.

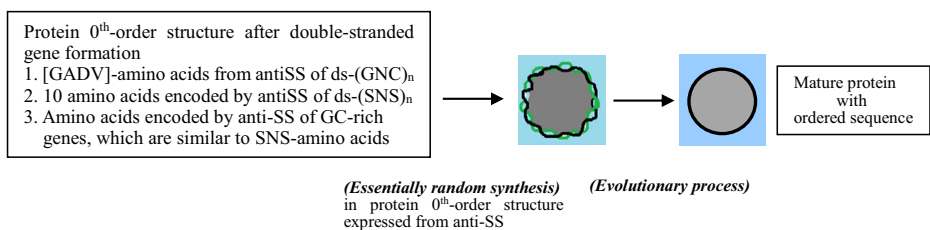


Fig. 1 After double-stranded gene formation, it is supposed that essentially random synthesis in protein 0th-order structure was carried out by expression from anti-SS, to produce protein with quite different sequence, but with similar amino acid composition to that from sense sequence. The immature protein could evolve to mature protein with an ordered sequence and a high catalytic activity. Wavy lines indicate flexible protein structure or wobbling of surface amino acids of a water-soluble globular protein

Whenever an entirely new enzyme was required and the corresponding catalytic function was not available, a newly-born gene expressing enzyme with the necessary catalytic function could be created from anti-SS of ds-(GNC)_n, ds-(SNS)_n and modern GC-rich gene, in the respective eras using GNC, SNS and the universal genetic code, along the evolutionary process from GNC primeval genetic code to the universal genetic code through SNS primitive genetic code. Where S means either G (guanine) or C (cytosine). Thus, SNS represents 16 genetic codes encoding 10 kinds of amino acids. That is, if a required, but generally weak, catalytic activity was detected in a protein expressed from the anti-SS, the immature gene would evolve to an entirely new gene encoding a mature enzyme with a high catalytic activity by accumulating appropriate base substitutions on the anti-SS in all three eras, as similarly to the case of ds-(GNC)_n gene shown in Fig. 1.

We have confirmed that all 12 average base compositions (4 (bases) times 3 (codon positions)=12) of microbial genes are well reproduced by simulation of gene evolution, taking an anti-SS of a GC-rich gene as the ancestor gene and the six conditions for water-soluble globular protein formation (Ikehara 2002, 2003). It is known that average compositions of 20 amino acids of proteins encoded by a microbial genome show a specific pattern. On the other hand, the gene simulation simultaneously corresponds to protein simulation, since genes always code for amino acid sequences of proteins. Then, we examined whether the average amino acid compositions of hypothetical proteins obtained by the simulation of gene evolution can reproduce those of proteins encoded by microbial genomes. From the results, it was found that the average amino acid compositions of simulated proteins are quite similar to those of proteins encoded by microbial genomes (Ikehara 2002). In addition, we have also showed that probabilities of two neighboring amino acids calculated from a multiplication of two average amino acid compositions are close to the average appearance frequencies that are actually detected in proteins encoded by microbial genomes (Ikehara 2002).

The results described above are all consistent with the following hypothesis: production of entirely new proteins originally started from direct random joining of [GADV]-amino acids in a protein 0th-order structure by prebiotic means, and was converted to the synthesis of proteins with ordered amino acid sequences that originated from immature proteins encoded by anti-SS genes, after formation of double-stranded genes. This provides an answer to the original question rewritten as follows: "How and when the conversion from random joining of amino acids to protein synthesis with ordered sequence could have happened?"

References

- Casper DL, Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol* 27:1–24
- Ikehara K (2002) Origins of gene, genetic code, protein and life: comprehensive view of life systems from a GNC-SNS primitive genetic code hypothesis. *J Biosci* 27(2):165–186
- Ikehara K (2003) Simulation of gene evolution: evidence for GC-NSF(a) hypothesis on the origin of genes. *Viva Origino* 31:201
- Ikehara K (2005) Possible steps to the emergence of life: the [GADV]-protein world hypothesis. *Chem Rec* 5(2): 107–118
- Ikehara K (2009) Pseudo-replication of [GADV]-proteins and origin of life. *Int J Mol Sci* 10(4):1525–1537
- Ikehara K (2012) "[GADV]-Protein world hypothesis on the origin of life" *Cellular Origin, life in extreme habitats and astrobiology 22, genesis-in the beginning: precursors of life, chemical models and early biological evolution* (ed. J. Seckbach), Springer, 107–121