On the Origin of Genes

Otsuka J*

JO Institute of Biophysics, Tokyo, Japan

*Corresponding author: Jinya Otsuka, JO Institute of Biophysics, Higashi-Ohizumi 6-49-18, Nerima-ku, Tokyo, 178-0063, Japan, Tel: +813 3921 1466; Email: jin.otsuka@kyj.biglobe.ne.jp

Abstract

The character of a present-day organism is determined by the set of genes. Thus, the origin of genes is one of the most important problems in the origin of life. From the aspect of the second law of thermodynamics, some systematic force is needed to produce such special arrangement of nucleotide bases as genes. In the present paper, such systematic force is proposed to be the selection acting on self-reproduced cells. As a molecular fossil of the self-reproduction of the primitive cell before the appearance of the ribosome, it is indicated in the prokaryotes that the polypeptide of a cell wall element is formed as the mixture of L- and D-types of amino acid residues by the catalytic actions of proteins and the matured element is transferred to the outside across the membrane by the orientation change of a lipid derivative bound to the element. This strongly suggests that the self-reproduction of a primitive cell has started from the antagonistic balance between lipid synthesis and cell wall construction. At this early stage, some of the RNAs contained in the cell would have catalyzed the polymerization of amino acids and some of the random sequences of polypeptides would have catalyzed the replication of RNAs and others catalyzed the lipid synthesis and cell wall construction. The self-reproducing cells obey the equation of replicator dynamics containing the change term. This equation not only shows Darwinian evolution that selects the nucleotide base changes upon the replication of RNAs to raise the increase rate of the cell but also leads to the convergence of RNA contents to the translation apparatus and RNA genes of proteins through the intermediate variant cells with the increase rate transiently lowered by the interference between RNAs. Lastly, the innovation process from such self-reproducing cells in the RNA-protein world to the present-day organisms is also briefly summarized.

Keywords: Cell wall; Lipid vesicle; Negative entropy; Selection; Self-reproduction

Introduction

The biological phenomena of a present-day organism are determined by the genes, and the origin of genes cannot be separated from the problems about the origin of life. However, most studies on the origin of life hitherto discussed have not reached the origin of genes.

In the early studies of biochemistry, Oparin proposes the first organism to have been the heterotroph utilizing the organic compounds synthesized non-biologically from the fact that the metabolic pathways of assimilation are added to the metabolic pathways of dissimilation, also noticing the intrinsic property of lipid molecules to form vesicles each consisting of a lipid bilayer in water [1]. As the biochemical studies reveal the important role of protein functions in an organism, the experiment of high voltage sparks passing through the gaseous phase of CH_4, NH_3, H_2O and H_2 is carried out to produce non-biologically different kinds of amino acids [2]. In other experiments,

the nucleic acid base adenine is synthesized by refluxing a concentrated solution of hydrogen cyanide in ammonia, polymerization of HCN also gives rise to orotic acid, which can be photochemically decarboxylated by sunlight to yield uracil, and various sugars are synthesized from formaldehyde [3].

In parallel to these trials of chemical syntheses, the studies of molecular biology reveal the central dogma in the present-day free living organisms; the proteins are translated from messenger RNAs (mRNAs) by the aid of ribosome consisting of ribosomal RNAs (rRNAs) and proteins, and of transfer RNAs (tRNAs), and the three kinds of RNAs are all transcribed from the respective DNA genes, which are replicated upon self-reproduction [4]. The transcription and replication are carried out by the specific proteins, DNA-dependent RNA and DNA polymerases, respectively. This raises the question "which came first, proteins or nucleic acids?". The discovery of the self-splicing activity in the intervening sequence of rRNA [5] proposes that the life has started from the self-replicative RNA called the RNA replicase [6].

In the field of physics, it has been discussed whether the life is compatible with the second law of thermodynamics [7,8], or how the organism maintain the negative entropy (the special arrangement of nucleotide bases in genes). Recently, the self-reproduction and the succeeding selection are proposed to resolve this problem on the basis of the concept of biological activity [9,10]. The RNA replicases certainly satisfy this condition [9,11] but they only form a closed system of RNAs. One role of RNA replicases toward the higher organization would have been to raise the concentration of variant RNAs as well as of the RNAs with the optimum replicase activity. Some of the variant RNAs would have catalyzed the polymerization of amino acids [12] and some of such polypeptides would have catalyzed the lipid synthesis and cell wall construction to form the self-reproducing primitive cells. Although the lipid vesicle itself is too unstable to evolve, the vesicle covered with a cell wall is not only protected from the osmotic pressure and/or external disturbance but also gives rise to the potentiality to self-reproduce. Such self-reproducing cells can be the candidate to have started the evolution of the unicellular organisms, which is mathematically formulated in the previous paper [13].

In the present paper, the metabolic pathways of lipid synthesis and cell wall construction in the present-day organisms are first examined to ascertain the above possibility of forming the self-reproducing primitive cells. This examination confirms the antagonistic balance between lipid synthesis and cell wall construction in the prokaryotes as the molecular fossil and proposes a model for the self-reproduction of primitive cells. On the basis of this model, it is explained in detail that the systematic force of selection acting on such self-reproduced cells converges their RNA contents to the translation apparatus and RNA genes of proteins.

Lipid Synthesis and Cell Wall Construction in the Present-Day Organisms

The enzyme proteins that catalyze the reaction steps leading to the biosynthesis of diacyl-sn-glycerol 3-phosphate from glyceraldehyde 3-phosphate and acetyl-CoA are homologous between prokaryotes and eukaryotes, although each of the catalytic activities is assigned to a single polypeptide chain in the prokaryote while the activities are attributed to a smaller number of multifunctional proteins in the eukaryote [14-19]. These similarities strongly suggest that the genes of proteins for the synthesis of phospholipid in this section have appeared before the divergence of prokaryotes and eukaryotes, although diacyl-sn-glycerol 3-phosphate is further modified differently to various lipid derivatives in different branches of organisms.

The biosynthesis of cell wall element and the cell wall construction revealed in the prokaryote (*Escherichia coli*) are especially noticeable in considering the self-reproduction of primitive cells. First, UDP-N-acetylglucosamine is derived from fructose 6-phosphate, acetyl-CoA and UTP. In the main pathway, UDP-N-acetylglucosamine is then converted into UDP-N-acetylmuramic acid by the reaction with phosphoenolpyruvate and further into UDP-MurNAc-LAla-DGlu-mesoDAP-DAla-DAla by the successive attachment of L-alanine, D-glutamic acid and meso-diaminopimelic acid under the catalytic actions of enzyme proteins. This N-acetylmuramoyl-pentapeptide is then bound to the undecaprenyl-phosphate lying across the lipid bilayer of the cytoplasmic membrane by the aid of enzyme protein MraY, and it is further attached to another molecule of N-acetylglucosamine, resulting in GlcNAc-MurNAc-pentapeptide moiety by the enzyme protein MurG [20,21]. At this stage, the undecaprenyl-phosphate changes its orientation to transfer the GlcNAc-MurNAc-pentapeptide moiety from the cytoplasm to periplasmic space across the lipid bilayer. Then, five molecules of glycines are successively bound to the moiety, and this moiety is then released from undecaprenyl-phosphate into the periplasm as a cell wall element. The incorporation of such elements into the cell wall is carried out by the penicillin binding proteins with

transglycosidase and transpeptidase activities [22]. Although these proteins contain the derivatives to determine the cell shape and septum [23-26], such derivation would have appeared in the DNA-RNA-protein world where replicating DNAs are attached to the membrane [27,28].

The process of constructing cell wall in the eukaryotes is not simple, but the biosynthetic process of mannoprotein, the main component of yeast cell wall, is fairly well revealed [29]. In this eukaryote, N-acetylglucosamine is first bound to dolichyl phosphate lying across the lipid bilayer of the endoplasmic reticulum by the catalytic action of UDP-N-acetylglucosamine-1-P-transferase (ALG7) [30], and another N-acetylglucosamine molecule is further bound to it. Five molecules of mannoses are then successively bound to the N-acetyleglucosamine secondarily attached to dolichyl pyrophosphate. At this stage, the oligosaccharide is transferred from cytoplasm into the endoplasmic reticulum by the change of oligosaccharide-dolichyl pyrophosphate complex in its orientation. After this change in orientation, other saccharides such as mannoses and glucoses, which are also transferred from the cytoplasm into the endoplasmic reticulum by the other molecules of dolichyl phosphates, are further bound to the oligosaccharide-dolichyl pyrophosphate complex, resulting in the assembly Glc3Man9GlcNAc2-pp-dolicol. This oligosaccharide is then combine with the polypeptide chain, which is synthesized on a mRNA template by the ribosome at the surface of endoplasmic reticulum. After the releasing of glucoses, the oligosaccharide-protein is transferred to the Golgi apparatus, where the mannose chain is further extended. The mannoprotein thus synthesized is then transferred into a secretory vesicle and is finally secreted out of the plasma membrane upon the fusion of the secretory vesicle to the membrane.

Although the biosynthetic and transferring process of cell wall elements in the eukaryote becomes complicated probably in association with the evolution of its intracellular structure, considerable similarities between *E. coli* and *S. cerevisiae* are recognized in the biomolecules participating in the essential part of this process. First of all, the dolichyl phosphate is similar to the undecaprenyl phosphate in the chemical structure and the biosynthesis of the former takes place in an extended pathway of polyprenyl phosphate. Second, ALG7 shows considerable similarity to MraY especially in their C-terminal halves of amino acid sequences. Third, the enzyme proteins catalyzing the reaction steps from glyceraldehyde 3-phosphate to glucosamine 6-phosphate are highly homologous between *E. coli* and *S. cerevisiae*.

A Model Proposed for the Self-Reproduction of Primitive Cells

The facts in the preceding section indicate that the lipid synthesis and cell wall construction have occurred before the divergence of prokaryotes and eukaryotes. Moreover, the polypeptide chain synthesized not by the ribosome but by the catalytic actions of proteins in the prokaryote strongly suggests that the cell wall construction began before the appearance of the ribosome. The ribosome in the present-day organism consists of large and small subunits of RNAs covered with many kinds of small proteins. While no conserved amino acid residue is found in the homologous alignment of any kind of ribosomal proteins, the conserved nucleotide bases are found in several borders of stem and loop regions of rRNAs. This implies that the catalytic action to polymerize the amino acid residues is attributed to some of the conserved nucleotide bases in the rRNAs [4].

From the above facts, the following route to form primitive self-reproducing cells will be proposed. Some of the variant RNAs derived from the RNA replicases catalyzes the polymerization of amino acids each wrapped by other variant RNA [4,12]. Although the polypeptide chains thus generated would have each consisted of only several amino acid residues, they amount to 20^5 kinds of polypeptide chains even when 20 kinds of L-type amino acids are randomly polymerized. Such many kinds of polypeptides sufficiently cover the amino acid sequences in the active centres of enzyme proteins of the organisms at the present time. Thus, it is reasonable that some of these polypeptides exhibits the RNA polymerase activity and others catalyze the syntheses of lipid molecules and cell wall elements, also using the D-type of amino acids. These lipid molecules and cell wall elements would have formed the primitive cells, each encompassing various polypeptides and RNAs in the vesicle covered with the cell wall. If the primitive cell can take material and energy sources for the syntheses of lipid and cell wall element as well as nucleotides and amino acids from the outside through the hydrophobic polypeptides lying across the lipid bilayer, such primitive cells can self-reproduce by the following properties of cell wall construction inferred from the prokaryote.

(i)   The cell wall element is first synthesized within the lipid vesicle, then transferred to the outside across the lipid bilayer by the orientation change in the special derivative of lipid molecule bound to the cell wall element, and the element is incorporated into the cell wall by the catalytic action of polypeptides on

the surface of the lipid vesicle.

(ii) The cell wall is relatively rigid to suppress the swelling of the lipid vesicle due to the osmotic pressure but is flexible in the sense that it is being partially broken and steadily reconstructed.

(iii) The rate of transferring a cell wall element passing through the lipid vesicle depends on the density of lipid molecules in the bilayer, that is, the transfer easily takes place at the lower density but becomes harder at the higher density.

With the above properties of cell wall construction, the primitive cell self-reproduces if its ability to synthesize lipid molecules and cell wall elements is sufficiently high. Four stages in the process of cell growth and self-reproduction are schematically shown in Figures 1(a)-(d).
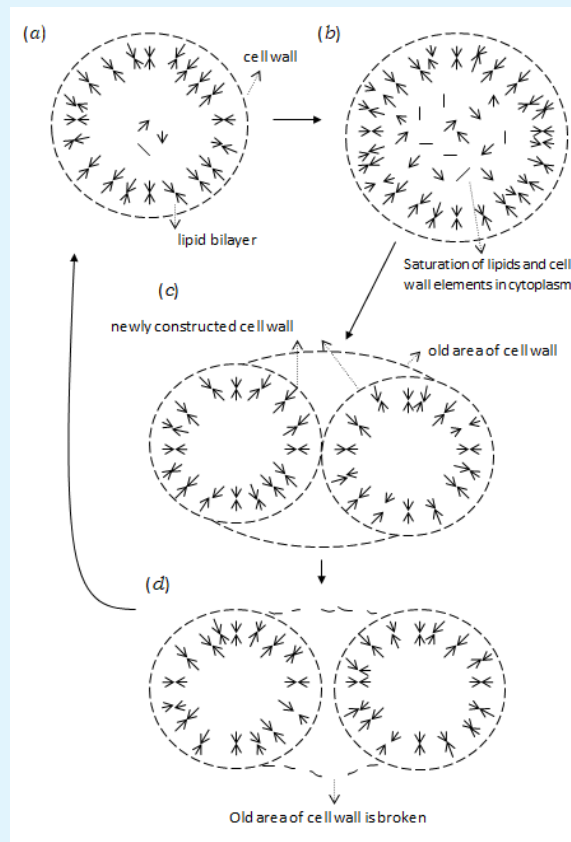

Figure 1: The schematic representation of self-reproduction a primitive cell due to the antagonistic balance between lipid synthesis and cell wall construction. Each lipid molecule is denoted by $\rightarrow$ and a cell wall element is by $-$. (a) The cell of a lipid bilayer vesicle covered with a cell wall at the initial stage of cell cycle. During this period of low density of lipid bilayer ($\rightarrow\!\!\leftarrow$), the cell wall elements synthesized in the cytoplasm are transferred to the outside of the vesicle across the lipid bilayer and they are used for the construction of new area of cell wall to enlarge the cell. The newly synthesized lipid molecules are also penetrated into the bilayer of the vesicle. (b) The stage of cell maturation. The lipid density in the bilayer becomes higher to prevent the newly synthesized cell wall elements from passing through the bilayer, and the lipid molecules as well as the cell wall elements are accumulated within the vesicle. (c) The division of lipid vesicle. The increased concentration of lipid molecules in the cytoplasm raises the free energy of the vesicle and the division of lipid vesicle spontaneously occurs to lower its free energy by increasing the ratio of surface to volume. (d) The separation of divided vesicles each covered with cell wall. The lowered lipid density in the bilayer again makes it possible for the cell wall elements to pass through the membrane, but the old area of the cell wall outside the newly constructed network is finally broken because this area is not contact with the catalytic polypeptides on the surface of either of divided vesicle. Although the lipid vesicle is not necessarily divided into equivalent daughter vesicles, the multiple sets of RNAs and polypeptides in the parent cell mostly guarantee the continuation of above cycle.

At the first stage (Figure 1(*a*)), the synthesized cell wall elements are transferred to the outside of the lipid vesicle (properties (i) and (iii)) and they are used for expanding the cell wall network as well as for repairing the broken area (property (ii)). If the synthesis of lipids progresses faster than the enlargement of the cell, the lipid density of the bilayer becomes gradually higher to prevent the newly synthesized cell wall elements from passing through the lipid bilayer. Then, the increase in the surface dimension of the cell wall is stopped (property (iii)) and the concentration of lipid molecules is also increased within the vesicle (Figure 1(*b*)). This raises the free energy of the vesicle, leading to the spontaneous division of the lipid vesicle to increase the ratio of surface to volume (Figure 1(*c*)). Because of the property (i), the network of cell wall is newly constructed especially in the interface between the divided vesicles to be connected with the old areas of network distant from the interface. On the other hand, the old area of cell wall outside the newly constructed network is finally broken. Thus, the two lipid vesicles, each covered with the cell wall, are ultimately separated from each other (Figure 1(*d*)).

The self-reproduction of cells is essentially important in improving the cell contents under the systematic force of selection. In the next section, this process unique to self-reproducing cells will be mathematically shown to describe the changes in the cell contents of RNAs and polypeptides.

The Changes in the Contents of Self-Reproducing Cells

The primitive self-reproducing cell proposed in the preceding section is characterized by the contained RNA assembly x_i, because its behaviour depends on the polypeptides synthesized under x_i. The number $n(x_i)$ of such self-reproducing cells x_i, each taking material and energy source M, obeys the following time-change equation.

$$\frac{d}{dt} n(x_i;t) = \{R(M;x_i) - D(x_i)\}n(x_i;t) + \sum_j q(x_i \leftarrow x_j;t)R(M;x_j)n(x_j;t)$$
(1)

Here, $R(M;x_i)$ and $D(x_i)$ are the self-reproducing rate and the death rate of the cell x_i, respectively, and $q(x_i \leftarrow x_i; t)$ is the change term of the cell x_i to other kinds of variant cells, i. e.,

$$q(x_i \leftarrow x_i;t) = -\sum_{k(\neq i)} q(x_k \leftarrow x_i;t)$$
(2)

Although the genes are not present at this stage, the RNA contents are considerably transmitted to the descendant cells through the replication of RNAs and is gradually changed by miss in the replication of RNAs.

The population behaviour of such self-reproducing cells becomes transparent by transforming Equation (1) into two equations; one concerning the total number of self-reproducing cells $B(t) = \sum_i n(x_i;t)$ in the population and another concerning the fraction $f(x_i; t)$ of variant cells x_i defined by $f(x_i;t) \equiv n(x_i;t)/B(t)$. By simple calculation, these equations are obtained from equation (1) in the following forms, respectively.

$$\frac{d}{dt} B(t) = W_{av}(M;t)B(t)$$
(3)

and

$$\frac{d}{dt} f(x_i;t) = \{W(M;x_i) - W_{av}(M;t)\}f(x_i;t) + \sum_j q(x_i \leftarrow x_j;t)R(M;x_i)f(x_i;t)$$
(4)

Here, the increase rate $W(M; x_i)$ of the variant cell x_i is defined by

$$W(M;x_i) \equiv R(M;x_i) - D(x_i)$$
(5)

and the average increase rate of the cells in the population is defined by

$$W_{av}(M;t) \equiv \sum_i W(M;x_i)f(x_i;t)$$
(6)

When the change term $q(x_i \leftarrow x_j; t)$ indicates the change in RNA contents arising from the nucleotide base changes upon replication of RNAs and from less frequent elongation and/or deletion of RNAs, Equation (4) shows the following Darwinian evolution of self-reproducing cells. If the increase rate $W(M; x_i)$ of the variant cell x_i spontaneously arisen from other variant cells is higher than the average increase rate $W_{av}(M; t)$, the fraction $f(x_i; t)$ of the new variant cells x_i increases with time t according to the first term on the right hand side of Equation (4). This raises the average increase rate, resulting in the increase in the total number $B(t)$ of the self-reproducing cells in the population, according to Equation (3), although this increase is ultimately stopped because of the decrease in available material and energy source M. On the other hand, the fraction $f(x_i; t)$ of variant cell x_i decreases if its increase rate is lower than the average increase rate. By the recurrence of such changes

in RNA contents and selection, the self-reproducing cells x_o with the optimum increase rate $W(M; x_o)$ become dominant in the population.

The Darwinian evolution at this stage of self-reproducing cells mainly enhances the activity of RNAs to polymerize amino acids and thus promotes the production of primitive RNA polymerase as well as of polypeptides for the lipid synthesis and cell wall construction. However, this not only increases the concentration of RNAs but also elongates some of RNAs in the cell. As the result, the variant cells with the increase rate lowered by the interference between RNAs also appear. Such interference occurs by transiently forming hydrogen bonds between complementary bases of different RNAs. The fraction of such variant cells suffering the interference is also derived from Equation (4) after the primitive self-reproducing cells with the optimum increase rate $W(M; x_o)$ have become dominant in the population. For this purpose, Equation (4) will be formally integrated with respect to time t.

$$f(x_i;t) = \exp \int_0^t \{W(M;x_i) - W_{av}(M;\tau)\}d\tau [\int_0^t \sum_j q(x_i \leftarrow x_j;\tau)R(M;x_j)f(x_j;\tau)$$

$$\exp[-\int_0^\tau \{W(M;x_j) - W_{av}(M;\tau')\}d\tau']d\tau + f(x_i;0)]$$

(7)

Among the fractions $f(x_i; t)$'s in this expression, we focus on the fraction $f(x_{I1}; t)$ of the variant cells x_{I1} whose increase rate $W(M; x_{I1})$ is lowered by the first step of RNA interference. Then, the average increase rate $W_{av}(M; \tau)$ and $W_{av}(M; \tau')$ are approximated to be $W(M; x_o)$ and the fractions of other variant cells except for x_o are neglected on the right hand side of Equation (7). The change terms $q(x_{I1} \leftarrow x_o; t)$ are averaged over a sufficiently long time to be regarded as the change rate from x_o to x_{I1}, i. e.,

$$q_{xI1,xo} = \frac{1}{t} \int_0^t q(x_{I1} \leftarrow x_o;\tau)d\tau \quad (8)$$

In this large time scale, Equation (7) leads to the following relation between $f(x_{I1})$ and $f(x_o)$ in a stationary state.

$$f(x_{I1}) = \frac{q_{xI1,x0}R(M;x_o)}{W(M;x_o) - W(M;x_{I1})} f(x_o) \quad (9)$$

By focusing on the change term $q(x_{I2} \leftarrow x_{I1}; t)$ of $f(x_{I1};t)$ on the right hand side of Equation (7), the fraction $f(x_{I2})$ of the variant cells x_{I2} suffering second step of RNA

interference is also expressed as the following relation with the fraction $f(x_{I1})$.

$$f(x_{I2}) = \frac{q_{xI2,xI1}R(M;x_{I1})}{W(M;x_o) - W(M;x_{I2})} f(x_{I1}) \quad (10)$$

By the repetition of similar procedures, the fraction $f(x_{Im})$ of variant cells x_{Im}, which have experienced m steps of RNA interference, is expressed as

$$f(x_{Im}) = \frac{q_{Im,xIm-1}R(M;x_{Im-1})}{W(M;x_o) - W(M;x_{Im})} f(x_{Im-1}) \quad (11)$$

Throughout the innovation of the above series of RNA interference and base changes, a new style of self-reproducing cell appears with the following types of RNA interference. (α) Generation of anticodons. A part of ancestral tRNA originally embracing the side chain of an amino acid residue bound by the primitive aminoacyl-tRNA synthetase alternatively attaches to the non-functional RNAs, which become later the RNA genes of proteins. This occurs for every pair of L-type amino acid and ancestral tRNA. (β) Contact of such ancestral RNA genes with the third type of RNAs, which become later 30S initiation complex. (γ) The interaction between the ancestral RNA gene carrying tRNAs each charged with an amino acid residue and the primitive rRNA under the assistance of primitive initiation complex. (δ) The enlargement of rRNA to form the A and P sites for the successive acceptance and polymerization of amino acid residues carried by tRNAs aligned on ancestral RNA genes of proteins.

From Equations (9) - (11), the probability $P(y \leftarrow x)$ that the new style self-reproducing cell y appears is given by

$$P(y \leftarrow x) = q_{y,xIm} \prod_{\mu=1}^m \frac{q_{xI\mu,xI\mu-1}R(M;x_{I\mu-1})}{W(M;x_o) - W(M;x_{I\mu})} \quad (12)$$

Here, x_o is set to x_{Io} and $q_{y,xlm}$ is change probability from x_{Im} to y mainly due to the changes in nucleotide bases of RNAs towards the new style of polymerization of amino acids in which L-type of amino acid residues carried by the corresponding tRNAs aligned on the ancestral protein gene are polymerized.

The fraction $f(y_k;t)$ of the new style variant cells y_k obeys the following time change equation, first coexisting with the original style of self-reproducing cells x.

$$\frac{d}{dt}f(y_k;t) = \{W(M;y_k) - \overline{W}(t)\}f(y_k;t) + \sum_l q(y_k \leftarrow y_l;t)f(y_l;t)$$

(13)

Here, the average increase rate $\overline{W}(t)$ of new and original styles of self-reproducing cells is defined by

$$\overline{W}(t) \equiv \sum_i W(M;x_i)f(x_i;t) + \sum_k W(M;y_k)f(y_k;t)$$

(14)

The total number $B(t)$ of self-reproducing cells is redefined as

$$B(t) = \sum_i n(x_i;t) + \sum_k n(y_k;t)$$ (15)

and obeys the following time-change equation.

$$\frac{d}{dt}B(t) = \overline{W}(t)B(t)$$ (16)

The increase rate of new style self-reproducing cells is gradually raised by Darwinian evolution according to Equation (13). First, the new style of amino acid polymerization by the evolved ribosome not only raises the rate of producing polypeptides but also produces the longer amino acid sequences to strengthen the specificity for substrates, in comparison with the random collision of the primitive rRNAs and ancestral tRNAs charged with amino acid residues in the original cells x. Second, the nucleotide bases in ancestral RNA genes of proteins are converged to encode the proteins more useful for self-reproduction of cells. Such changes in nucleotide bases of respective RNAs and selection make the optimal increase rate of the new style self-reproducing cells y_o larger than that of the original style self-reproducing cell x_o, compelling the latter to be extinct.

In the new style of self-reproducing cell y_o, the negative entropy - S_N is more extended than that of the primitive self-reproducing cell x_i first formed. According to the second law of thermodynamics, the entropy production arising from the difference between the acquired energy $E_a(M; y_o, S_N)$ and the energy $E_s(y_o, S_N)$ stored in the biomolecules of the cell must compensate for the negative entropy - S_N. In practice, the biological activity, which is defined by $E_a(M; y_o, S_N) - E_s(y_o, S_N) - TS_N$, is considered to be proportional to the self-reproducing rate of the cell y_o as the first approximation [9,10]. Thus, the RNA genes of proteins are mostly directed to raise the

acquired energy as well as to improve the cell division. In other words, the available organic compound synthesized non-biologically must have been sufficiently present in the environment under which the self-reproducing cells have established the translation apparatus and the RNA genes of proteins.

Conclusions and Discussion

Without the self-reproduction and selection, the polymerization of ribonucleotides only yields the random sequences of nucleotide bases because the binding energy of ribonucleotides hardly depends on their bases. In this case, the special arrangements of nucleotide bases such as those in the translation apparatus and RNA genes of proteins are hardly generated and their encounter is still more difficult. Although the RNA replicases form a self-reproducing system, they tend to be extinct when the concentration of ribonucleotides synthesized non-biologically decreases. On the contrary, the RNAs containing those catalyzing the polymerization of amino acids in self-reproducing cells are organized towards the higher systematization of RNAs and proteins by the systematic force acting on the self-reproducing cells. In the population of such self-reproducing cells, the variant cells with the increase rate transiently lowered by multiple steps of RNA interference are also allowed to exist as the minor members of population, although the variant cell with the higher increase rate arising from the first order of change term tends to be selected by Darwinian evolution. The selection among the minor members of population leads to the new style self-reproducing cells having the translation apparatus and RNA genes of proteins. Such new style self-reproducing cells have the potentiality to overcome the decrease in organic compounds synthesized non-biologically by the following innovation and evolution, as mathematically formulated already [13,31].

The first innovation is the deoxydization of RNAs due to the protons released from the glycolysis. This leads to the divergence of prokaryotes and eukaryotes by the ways to evolve the equi-partition of replicated DNAs to the daughter cells. In the prokaryote, DNA genes are fused to a single circular DNA and the replicating DNAs are directly attached to the membrane, while multiple chromosomes are attached to the membrane through the contractive microtubules upon their replication in the eukaryote. On the basis of stable DNA genes, the prokaryotes have first evolved various chemical syntheses, O_2- respiration and photosynthesis. The eukaryotes, which have first evolved nuclear membrane, endocytosis and exocytosis probably as the predator of

prokaryotes, have then acquired the mitochondria as the endosymbionts of O_2-respiratory eubacteria and some of them have further acquired the photosynthetic plastids as the endosymbionts of photosynthetic eubacteria. Some of such eukaryotes have further evolved multicellularity and cell differentiation by expanding protein repertoire under the sufficient supply of energy and materials from the endosymbionts. These innovation and evolution are also ascertained by the analyses on the neutral base changes in the DNA genes of rRNAs [32-36]. Meanwhile, the self-reproducing cells in the RNA-protein world have probably turned their strategy to survive as RNA phages and viruses. Although they have lost the ribosome and their origin cannot be traced by the study of molecular evolution, they utilize the ribosome in the host cell to translate their own RNA genes with the universal codon usage. This strongly suggests that the translation apparatus has been established in the self-reproducing cells in the RNA-protein world.

However, the RNA phages and viruses are too degenerate to be inquired into the origin of genes. Some of them only consists of RNA genes enveloped by capsule proteins. For confirming experimentally the origin of genes, therefore, it is necessary to advance the laboratory experiments of synthesizing organic compounds towards the creation of circumstance under which the primitive self-reproducing cells are generated in the RNA-protein world. After the formation of self-reproducing cells, the computer simulation is also useful for reproducing the convergence of RNA contents to the translation apparatus and RNA genes of proteins in a short time.

References

1. Oparin AI (1957) The origin of life on the earth. 3rd(Edn.), Academic Press, New York, pp: 528.

2. Miller SL, Orgel LE (1974) The origin of life on Earth. Englewood Cliffs, N. J. Prentice-Hall.

3. Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM (1987) Molecular biology of the gene. 4th (Edn.), volume II, The Benjamin/Cumming Publishing Inc., California, pp: 1101-1103.

4. Watson JD, Hopkins NH, Roberts JW, Steitz JA, Weiner AM (1987) Molecular biology of the gene. 4th (Edn.), volume I, The Benjamin/Cumming Publishing Inc. California.

5. Zaug AJ, Cech TR (1986) The intervening sequence RNA *Tetrathymena* is an enzyme. Science 231(4737): 470-475.

6. Cech TR (1986) A model for the RNA catalyzed replication of RNA. Proc Natl Acad Sci USA 83(12): 4360-4363.

7. Maxwell JC (1871) Theory of heat. Longmans, Green and Company, London, New York and Bombay, pp: 328.

8. Brillouin L (1951) Maxwell's demon cannot operate: information and entropy. I. J Appl Phys 22(3): 334-337.

9. Otsuka J (2017) The concept of biological activity and its application to biological phenomena. J Phys Chem & Biophys 7(1): 235-241.

10. Otsuka J (2018) The negative entropy in organisms; its maintenance and extension. J Mod Phys 9(12): 2156-2168.

11. Otsuka J, Nozawa Y (1998) Self-reproducing system can behave as Maxwell's demon; theoretical illustration under prebiotic conditions. J Theor Biol 194(2): 205-221.

12. Weiner AM, Maizel N (1987) tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. Proc Natl Acad Sci USA 84(21): 7383-7387.

13. Otsuka J (2017) A mathematical formulation of evolution and innovation. I. Unicellular organisms. Phys Sci & Biophy J 1(1): 1-10.

14. Schweizer M, Roberts LM, Holke H, Takabayashi K, Hollerer E, et al. (1986) The pentafunctional FASI gene of yeast; its nucleotide sequence and order of the catalytic domains. Mol Gen Genet 203(3): 479-486.

15. Seyama Y, Kawaguchi A (1987) Fatty acid synthesis and the role of pyrimidine nucleotide. *In:* Dolphin D, Poulson R, Avramovic O (Eds.), Pyrimidine nucleotide coenzymes: biochemical and medical aspects. Wiley, New York, vol 2B, pp: 381-431.

16. Harwood JL (1988) Fatty acid metabolism. Annu Rev Plant Physiol Plant Mol Biol 39: 101-138.

17. Li S, Cronan JE (1992a) The gene encoding the biotin carboxylase subunit of *Escherichia coli* acetyl-CoA carboxylase. J Biol Chem 267(2): 855-863.

18. Li S, Cronan JE (1992b) The genes encoding the two carboxylase subunits of *Escherichia coli* acetyl-CoA carboxylase. J Biol Chem 267(24): 16841-16847.

19. Al-Feel W, Chirale SS, Wakil SJ (1992) Cloning of the yeast FAS3 gene and primary structure of yeast acetyl-CoA carboxylase. Proc Natl Acad Sci USA 89(10): 4534-4538.

20. Izaki K, Matsuhashi M, Strominger JL (1968) Biosynthesis of the peptidoglycan of bacterial cell wall. XIII Peptidoglycan transpeptidase and D-alanine carboxypeptidase: penicillin-sensitive enzymatic reaction in strains of *Escherichia coli*. J Bio Chem 243(11): 3180-3192.

21. Van Heijenoort Y, Gemez M, Derrien M, Ayala J, Van Heijenoort J (1992) Membrane intermediates in the peptidoglycan metabolism of *Escherichia coli:* possible roles of PBP1b and PBp3. J Bact 174(11): 3549-3557.

22. Bupp K, Van Heijenoort J (1993) The final step of peptidoglycan subunit assembly in *Escherichia coli* occurs in the cytoplasm. J Bact 175(6): 1841-1843.

23. Spratt BG (1975) Distinct penicillin-binding proteins involved in the division, elongation and shape of *Escherichia coli* K12. Proc Natl Acad Sci USA 72(8): 2999-3003.

24. Begg KJ, Donachie WD (1985) Cell shape and division in *Escherichia coli*: experiments with shape and division mutants. J Bact 163(2): 615-622.

25. Cooper S (1988) Rate and topology of cell wall synthesis during the division cycle of *Salmonella typhmurium*. J Bact 170(1): 422-430.

26. Matsuhashi M, Wachi M, Ishino F (1990) Machinery for cell growth and division: penicillin-binding proteins and other proteins. Res Microbiol 141(1): 89-103.

27. Jacob F, Brenner S, Cuzin F (1963) On the regulation of DNA replication in bacteria. Cold Spr Harb Symp Quant Biol 28: 329-348.

28. Ogden GB, Pratt MJ, Schaechter M (1988) The replication origin of the *Escherichia coli* chromosome binds to cell membrane only when hemimethylated. Cell 54(1): 127-135.

29. Ballou CE (1982) The yeast cell wall and cell surface. *In:* Strathern JN, Jones EW, Broach JR (Eds.), The molecular biology of the yeast *Sacharomyces*, metabolism and gene expression. Cold Spring Harbor, New York, Cold Spring Harbor Laboratory, pp: 335-360.

30. Criscuolo BA, Kragg SS (1982) Selection of tunicamycin resistant Chinese hamster ovary cells with increased N-acetylglucosaminyl transferase activity. J Cell Biol 94(3): 586-591.

31. Otsuka J (2018) A mathematical formulation of evolution and innovation. II. From unicellular monoploid eukaryotes to multicellular diploid eukaryotes. Phy Sci & Biophy J 2(2): 1-11.

32. Yang D, Oyaizu Y, Olsen GJ, Woese CR (1985) Mitochondrial origin. Proc Natl Acad Sci USA 85(13): 4443-4447.

33. Woese CR (1987) Bacterial evolution. Microbiol Rev 51(2): 221-271.

34. Van den Eynde H, De Baere R, De Boeck E, Van de Peer Y, Vandenberghe A, et al. (1988) The 5S ribosomal RNA sequences of a red algal rhodoplast and gymnosperm chloroplast: implication for the evolution of plants and cyanobacteria. J Mol Evol 27(2): 126-132.

35. Otsuka J, Nakano T, Terai G (1997) A theoretical study on the nucleotide change under a definite functional constraint of forming stable base-pairs in the stem regions of ribosomal RNAs; its application to the phylogeny of eukaryotes. J Theor Biol 184(2): 171-186.

36. Otsuka J, Terai G, Nakano T (1999) Phylogeny of organisms investigated by the base-pair changes in the stem regions of small and large subunit RNAs. J Mol Evol 48(2): 218-235.