Constructive Approach to Protocell: Theory and Experiments

Kunihiko Kaneko

Department of Basic Science, Graduate School of Arts and Sciences,University of Tokyo

ERATO Complex Systems Biology Project

University of Tokyo, Komaba, Meguro, Tokyo 153-8902, JAPAN

Abstract

As an example of constructive biology, recursive production of a protocell is studied. Theory of minority control in a mutually catalytic reaction system is presented, which gives a basis on the origin of bio-information. Recursive production states as well as switching among them in a catalytic reaction network system are studied numerically, where compositional information, evolvability, and universal fluctuation law are discussed. As an experimental approach, construction of protocell is outlined. It consists of in-vitro autonomous replication system, replication of liposome, and RNA transcription and protein synthesis in liposome. Relevance of the minority control theory to sustain and evolve catalytic activity is demonstrated.

1 Introduction

To understand what life is, we need to reveal universal features that all life systems have to satisfy at minimum, irrespectively of detailed biological processes. The present organisms, however, include detailed elaborated processes that are captured through the history of evolution. Then, for our purpose, it is desirable to set up a minimal biological system, to understand universal logic that organisms necessarily should obey. Hence, our approach that should be taken will be 'constructive' in nature. This is an approach what we call "constructive biology", which has been carried out both experimentally and theoretically[Kaneko 2003b,2004].

There are three levels, in the constructive biology. (1)gedanken experiment (logic) (2)computer model, and (3)real experiment. The first one is theoretical study, reveling a logic underlying universal features in life processes, essential to understand the logic of 'what is life'. Still, life system has a complex relationship among many parts, which constitute the characteristic feature as a whole, which then influences the process of each part[Kaneko 1998, Kaneko & Tsuda 2000]. We have not gained sufficient theoretical intuition to such complex system. Then it is relevant to make computer experiments and heuristically find some logic. This is the second approach mentioned above, i.e., construction of artificial world in a computer[Kaneko & Yomo 1997,1999, Furusawa & Kaneko 1998].

Still, in a system with potentially huge degrees of freedom like life, the construction in a computer may miss some essential factors. Hence, we need the third experimental

approach, i.e., construction in a laboratory. In this case again, one constructs a possible biology world in laboratory, by combining several procedures. For example, this experimental constructive biology has been pursued by Yomo and his collaborators (see e.g., [Matsuura et al.2002, Ko et al.1994, Kashiwagi et al. 2002] at the levels of biochemical reaction, cell, and ensembles of cells.)

Taking this synergetic approach from the three levels for constructive biology, we have been working on constructing minimal replicating cells, multi-cellular organism, evolution, symbiosis, and so forth, both theoretically and experimentally[Kaneko 2003b]. The construction of a replicating system with compartment, of course, is essential to consider the origin of a protocell, which is the theme of the present volume, and will be discussed here. Note, however, that we do not intend to reproduce what occurred in the earth. We do not try to guess the environmental condition of the past earth. Rather we try to construct such replication system from complex reaction network under a condition preset up by us.

2 Theoretical Issue:1 — Origin of Bioinformation

2.1 Question: Origin of Heredity

In a cell, among many chemicals, only some chemicals (e.g., DNA) are regarded to carry genetic information. Why do only some specific molecules play the role to carry the genetic information? How has such separation of roles in molecules between genetic information and metabolism progressed? Is it a necessary course of a system with internal degrees and reproduction?

Eigen, following the experimental study of Spiegelman on replication of RNA[Millis et al. 2000], considered how recursive production of catalytic molecules is possible[Eigen & Schuster 1979]. For replication to progress, catalysts are necessary, and information in a polymer to produce it must be preserved. However, error rate in replication must have been high at a primitive stage of life, and accordingly, it is recognized that the information to carry catalytic activity will be lost within few generations. To resolve this problem of inevitable loss of catalytic activities through replication errors, Eigen and Schuster [1979] proposed hypercycle, where, replicating chemicals catalyze each other forming a cycle, as A catalyzes B, B catalyzes C, C catalyzes D, and D catalyzes A. With this hypercycle, the original problem of error accumulation is avoided. However, the hypercycle itself turned out to be weak against parasitic molecules, i.e., those which are catalyzed by a molecule in the cycle, but do not catalyze those in the cycle. Later the possibility was discussed that compartmentalization by a cell structure [Eigen 1992,.Szathmary and Maynard-Smith 1997] or spatially localized structure in reaction-diffusion system may suppress the invasion of parasitic molecules [Boerilst & Hogeweg, 1991, Altmeyer & McCaskill 2001].

On the other hand, Dyson[1985] discussed that a set of a large number of chemical species may continue reproduction, sustaining catalytic activity. Although accurate replication of such variety of chemicals is not possible, chemicals, as a set, may continue reproducing themselves loosely, while keeping catalytic activity.

It is important to study if such loose reproduction as a set is possible in a mutually catalytic reaction network. If this is possible, and if these chemicals also include molecules forming a membrane for compartmentalization, reproduction of a primitive cell will be

possible. In fact, from chemical nature of lipid molecules, it is not so surprising that a comparmet structure is formed from lipid.

Still, in this reproduction system, any particular molecules carrying information for reproduction do not exist, in contrast to the present cell which has specific molecules (DNA) for it. As for a transition from early loose reproduction to later accurate replication with genetic information, Dyson only referred to 'genetic take-over'[Cairns-Smith 1982], while its mechanism is not discussed.

Now, it is important to study how recursive production of a cell is possible, with the appearance of some molecules to play a specific role for heredity. For this study, let us consider a simple protocell that consists of mutually catalyzing molecule species whose growth in number leads to cell reproduction[Kaneko & Yomo 2002]. In this protocell, the molecules that carry the genetic information are not initially specified. The first question we discuss here is if some specific molecules start to carry information for heredity, to realize continual reproduction of such protocell.

In the present cells, it is generally believed that DNA is a carrier of heredity, which controls the behavior of a cell. Still, even in these cells, proteins and DNA both influence their replication process each other. At this point, we need to first clarify what 'heredity' really means. Here, one might point out that DNA molecules would be suited to encode many bits of information, and hence would be selected as an information carrier. Although this 'combinatorial' capacity of DNA molecule is important, what we are interested here is a basic property that has to be satisfied prior to that, i.e., carrier of heredity at the minimum, for which the following two features are necessary.

(1) Such molecules are preserved well over generations. The number of such molecules exhibits smaller fluctuations than that of other molecules, and their chemical structure (such as polymer sequence) is preserved over a long time span. We refer to this as the **'preservation property'**.

(2) If this molecule is replaced by some other type of molecule, there is much stronger influence on the behavior of the cell than the case when other molecules are changed. We refer to this as the **'control property'**.

The question we address is as follows. Under what conditions, recursive production of a protocell continues maintaining catalytic activities? Are molecules carrying heredity necessary for it? Under what conditions, does one molecule species begin to satisfy the conditions (1) and (2) so that the molecule carries heredity?

2.2 Minority Controlled State

A theoretical study to answer the question in §2.1 was presented in [Kaneko & Yomo 2002]. By setting up a condition for prototype of a cell consisting of mutually catalyzing molecules, it was shown, under rather general conditions, that symmetry breaking between two kinds of molecules takes place. Through replication and selection, one kind of molecule comes to satisfy the conditions (1) and (2) in §2.1. Here, we outline the logic of the study.

First, assume a prototype of cell, consisting of molecules that catalyze each other (see Fig.1). As the reaction progresses, the number of molecules will increase. Then, considering physical nature of membrane, this cell will divide when its volume (the total number of molecules) is beyond some threshold. Then the molecules split into two 'daughter cells". Here we first consider the simplest case: Only two kinds of molecules X and Y exist in this protocell, and they catalyze each other for the synthesis of the molecules. Without

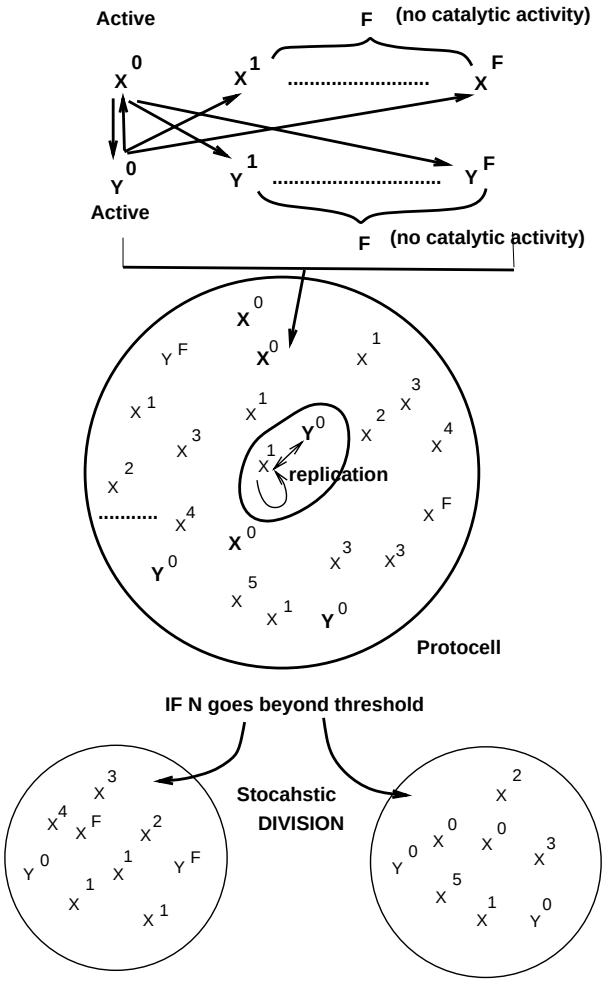

Figure 1: Schematic representation of our model

losing generality, one can assume that the synthesis speed of X is faster than that of Y, or in other words, one can say that the catalytic activity of Y is much stronger than that of X.

In chemical replication process of complex polymers, some structural changes in molecules can occur, that may be termed as 'replication error'. These structural changes in each kind of molecules often result in the loss of catalytic activity, and the molecules without catalytic activity are common. Then, this protocell is taken over by such non-catalytic molecules, as discussed in the parasite problem for the hypercycle. Hence, maintenance of reproduction of this protocell is not so easy.

As a simple 'gendaken experiment' consider the case that each kind of X and Y has several types $0, 1, ..., F-1$, as X^j or Y^j, and only the type 0 has catalytic activity. For example X and Y are different kinds of molecules, while the index X^j with different j represents a polymer with different configuration of monomers. Then, the question is how the active form 0 is maintained through reproduction, even if F is large.

Here, due to the difference in the speed of synthesis, the number of X molecules will get larger than that of Y. As long as the number of molecules in a cell is not large, eventually, there comes to a stage that the catalytically active Y molecule (i.e., Y^0) is extinct. Then, X molecules are no longer synthesized. Inactive Y molecules (Y^j; $j > 0$) may still be synthesized as long as X^0 molecules remain. However, after each division, the number of X^0 molecules becomes half, and sooner or later, the cell stops division. Hence, once the number of Y^0 becomes 0, the reproductivity of the cell will be lost.

Recall that the number of molecules in a (proto)cell is not huge. Due to fluctuations in molecule numbers, some cells may keep active Y molecules. Since there is little room for Y molecules in protocells, due to its slow synthesis, the number of Y^j molecules (for all $j = 0, 1, .., F-1$) should be small. Hence, typically, to keep the active Y molecules, the number of all the other Y molecules should be suppressed to zero. Once the inactive Y molecules go extinct, they do not reappear so often, since both the number of Y molecules and the error rate are small. Hence a cell state with very few Y^0 molecules and almost zero Y^j ($j > 0$) molecules can keep reproduction. Note that such state does not exist if the total number of molecules is very large. Due to the finiteness in molecule numbers, fluctuations for such special initial condition can occur, and such rare fluctuations, once they occurred, are preserved, since the cell with such compositions can continue reproduction (see Fig.2).

The above argument was numerically confirmed through simulations of the above model. The state with few active Y molecules (Y^0) and almost zero inactive Y molecules (Y^j ;$j > 0$)) is established, under the following conditions: (i)the number of molecules in a cell is not too large, (ii) the number of types of inactive molecules (F) is large, and (iii) there is sufficient difference between the growth speeds of the two kinds of molecules (X and Y). In this state with few Y^0 molecules, the active Y molecule is a carrier for heredity, in the sense that the molecule has the following properties [Kaneko & Yomo 2002].

'Preservation property': The active Y molecules are preserved well over generations. Indeed, selected and preserved is a state with the number of active Y molecules $2 \sim 6$, and with almost zero inactive Y molecules. The realization of such state is very rare from the calculation of probability, but it is preserved over generations.

'Control property': Consider a structural change in Y molecule, that may occur as a replication error and causes a change of catalytic activity. Since the number of active Y molecules is few, and all the X molecules are catalyzed by them, this influence is

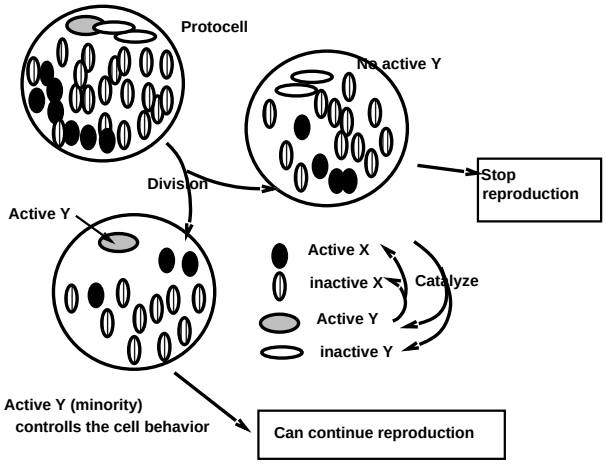

Figure 2: Schematic representation of the minority controlled state

enormous. The synthesis speed of a protocell should drastically change. On the other hand, a change to X molecules has a weaker influence, since there are many active X molecules, and influence of change in each molecule is averaged out. Hence, the change of Y molecule has a crucial influence on the cell behavior, compared with that of X molecules.

Summing up the argument in this section, the molecule species with slower replication speed and (accordingly) with minor population, comes to possess the properties for the heredity. The state controlled by minority molecule species is termed as minority controlled state (MCS).

2.3 Significance of Minority Controlled State

Since the change of the minority molecules is not smeared out by the law of large numbers, important characteristics of the MCS is evolvability. As the replication process is entirely facilitated by catalytic activity, growth speed of the protocell depends on catalytic activities of molecules in it. With some change to molecules, the protocells including greater catalytic activities will be selected through evolution, leading to the selection of molecules with higher catalytic activities. Because only a few Y^0 molecules exist in the MCS, a change to one of them strongly influences the catalytic activity of the protocell, as supported by the control property. On the other hand, a change to X molecules has a weaker influence, on the average. Hence the MCS is important for evolvability of a protocell.

This MCS has a positive feedback process to stabilize the state itself. Since the preservation of minority molecules is essential to a cell of MCS, a new selection pressure emerges to further ensure the preservation of the minority molecule into the offspring cells. (Otherwise the reproduction of the cell is highly damaged.) A machinery to guarantee the faithful transmission of the minority molecule should evolve, which further strengthens the preservation of the minority molecule. Hence, heredity evolves just as a result of kinetic phenomenon in a reproducing protocell consisting of mutually catalytic molecules.

Once this faithful transmission of a minority molecule is evolved, then other chemicals that are synthesized in connection with it are also probable to be transmitted. Then, more

molecules are transmitted, in conjunction with the minority molecule. With this evolution having more molecules catalyzed by the minority molecule, the machinery to better take care of minority molecules will also evolve, since this minority molecule is involved in reactions for the synthesis of many other molecules. Hence the MCS allows for co-evolution for better transmission of minority molecules and for coding of more information, leading to the separation of genetic information form other molecules carrying metabolism.

3 Theoretical Issue:2 — Origin of Recursive Production and Evolvability

Question: A cell consists of several replicating molecules that mutually help in their synthesis and keep some synchronization for replication. How is such recursive production maintained, while keeping diversity of chemicals? This recursive production is not complete, and there appears a slow 'mutational' change over generations, which leads to evolution. How are evolvability and recursive production compatible?

In the discussion of §2, we considered a system consisting of two kinds of molecules for simplicity. To study the general features of a system with mutually catalyzing molecules, however, it is important to consider a system with a variety of chemicals (k molecule species), forming a mutually catalyzing network (see Fig.3). The molecules replicate through catalytic reactions, so that their numbers within a cell increase. Again, when the total number of molecules exceeds a given threshold (here we used $2N$), the cell divides into two, with each daughter cell inheriting half of the molecules of the mother cell, chosen randomly. Here we choose a random catalytic network, i.e., chemical species catalyzes the synthesis of some other randomly chosen chemical as

$$X^i + X^j \to 2X^i + X^j. \tag{1}$$

with $i, j = 1, \cdots, k$. The connection rate of the catalytic paths is given by p per each chemical. Again, replication is accompanied by some 'error', and instead of the replication of the molecule i, one of other k molecule species is synthesized with an error rate μ.

In this model there are four basic parameters; the total number of molecules N, the total number of molecule species k, the mutation rate μ, and the reaction path rate p. Catalytic activity depends on each molecule, which gives the rate of the above reaction (1). We assume that the catalytic activity $c(j)$ is chosen again from random number over $[0, 1]$. By carrying out simulations of this model, choosing a variety of parameter values N, k, μ, p, also by taking various random networks, we have found that the behaviors are classified into the following three types[Kaneko, 2002,2003a,2005]:

(1) Fast switching states without recursiveness
(2) Achievement of recursive production with similar chemical compositions
(3) Switch over several quasi-recursive states

3.1 Significance of Minority Controlled State

In the first phase, there is no clear recursive production and the dominant molecule species changes frequently. At one time step, some chemical species are dominant but only a few

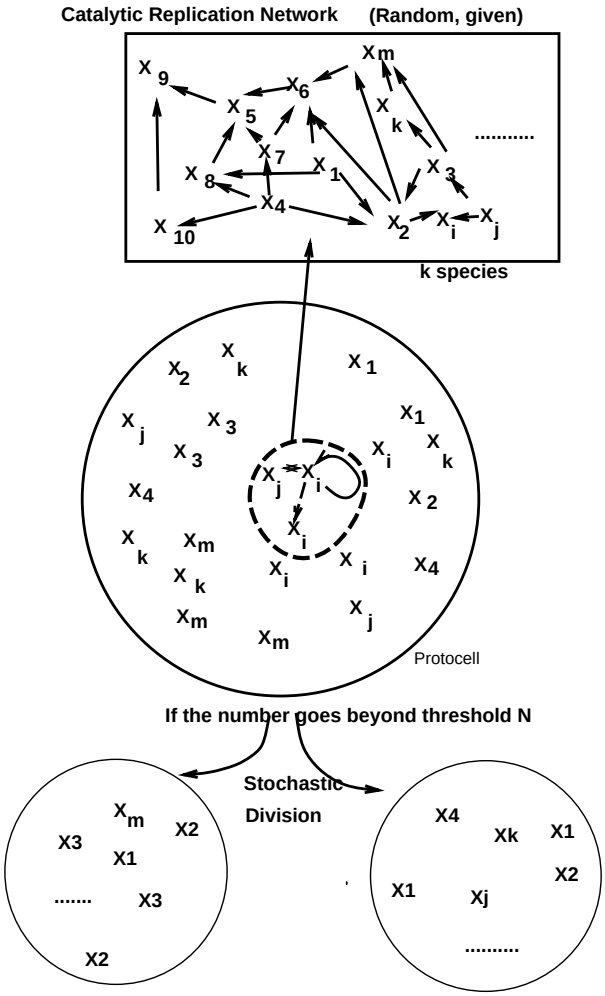

Figure 3: Schematic representation of our model

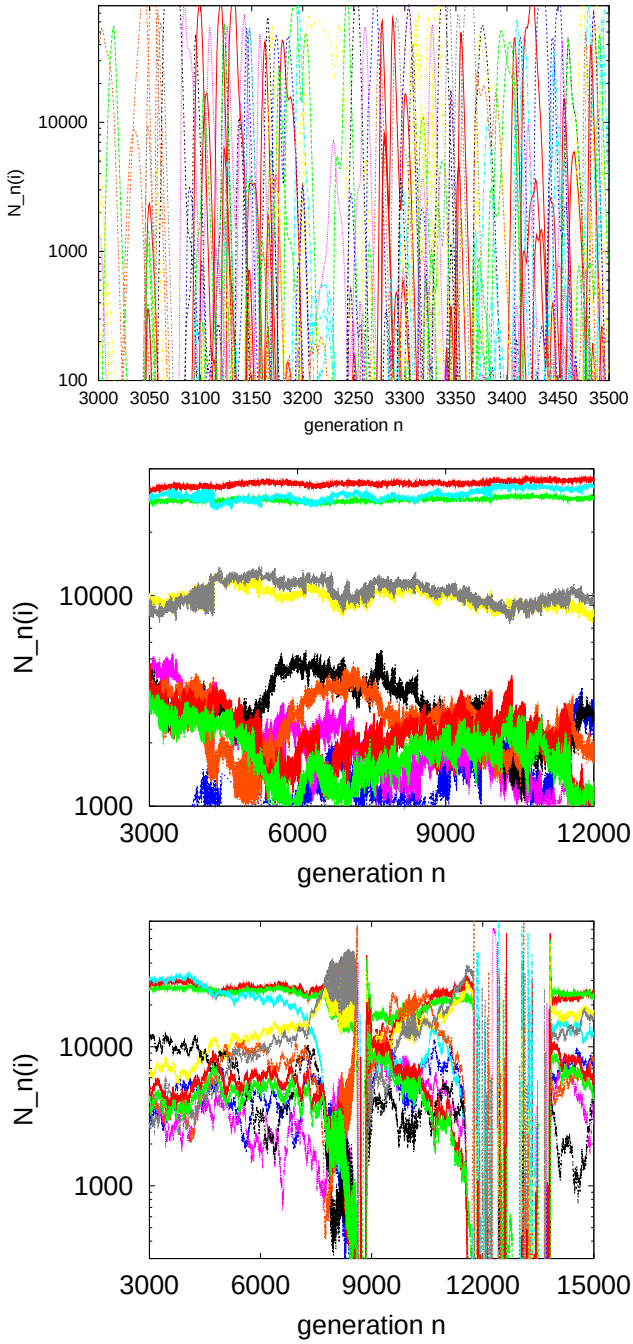

Figure 4: The number of molecules $N_n(i)$ for the species i is plotted as a function of generation n of cells, i.e., at each successive division event n. In (a), a random network with $k = 500$ and $p = .2$, and in (b) that with $k = 200$ and $p = .2$ was adopted, with $N = 64000$ and $\mu = 0.01$. Only some species (whose population get large at some generation) are plotted. In (a), dominant species change successively in generation, while in (b) three quasi-recursive states are observed.

generations later, this information is lost, and the number of the molecules in this species goes to zero. (see Fig.4a). Here, the time required for reproduction of a cell is huge compared with the case (2), and will go beyond the time scale accessible experimentally. Hence this phase is not suitable for protocell.

In the second phase, on the other hand, a recursive state is established, and the chemical composition is stabilized such that it is not altered much by the division process (see Fig.4b). Generally, all the observed recursive states consist of 5-10 species, except for those species with few molecule numbers, which exist only as a result of replication errors. These 5-10 chemicals form a mutually catalytic network as will be discussed later (see Fig.5). The member of these 5-10 species do not change by generations, and the chemical compositions are transferred to the offspring cells. Once reached, this state is preserved over the whole range of simulation steps.

In the third phase, after one recursive state lasts over many generations (typically a thousand generations), a fast switching state appears until a new (quasi-)recursive state appears. As shown in Fig.4c, for example, each (quasi-)recursive state is similar to that in the phase (2), but in this case, after many generations, it is replaced by a fast switching state as in the phase (1). Then the same or different (quasi-)recursive state is reached again, which lasts until the next switching occurs.

Although the behavior of the system depends on the choice of the network, there is a general trend with regards to the phase change, from (1), to (3), and then to (2) with the increase of N, or with the decrease of k or p. We also note that the state with (quasi-)recursive production corresponds to the "composome" in the model by Segre and Lancet[Segre et al. 2000].

3.2 Maintenance of Recursive Production

How is the recursive production sustained in the phase (2)? We have discussed already the danger of parasitic molecules that have lower catalytic activities and are catalyzed by molecules with higher catalytic activities. As discussed in §2.1, such parasitic molecules can invade the hypercycle. Indeed, under the structural changes and fluctuations, the recursive production state could be destabilized. To answer the question on the itinerancy and stability of recursive states, we have examined several reaction networks. The unveiled logic for the maintenance of recursive state is summarized as follows.

(a) **Stabilization by intermingled hypercycle network**:

The 5-12 spices in the recursive state form a mutually catalytic network, for example, as shown in Fig. 5. This network has a *core hypercycle network*, as shown in thick arrows in the figure. Such core hypercycle has a mutually catalytic relationship, as " A catalyzes B, B catalyzes C, and C catalyzes A". However, they are connected with other hypercycle networks such as $G \to D \to B \to G$, and $D \to C \to E \to D$, and so forth. The hypercylces are intermingled to form a network. Coexistence of core hypercycle and other attached hypercycles are common to the recursive states we have found in our model.

This intermingled hypercycle network (IHN) leads to stability against parasites and fluctuations. Assume that there appears a parasitic molecule to one species in the member of IHN (say X as a parasite to C in Fig.5). The species X may decrease the number of the species C. If there were only a single hypercycle $A \to B \to C \to A$, the population of all the members A, B, C would be easily decreased by this invasion of parasitic molecules,

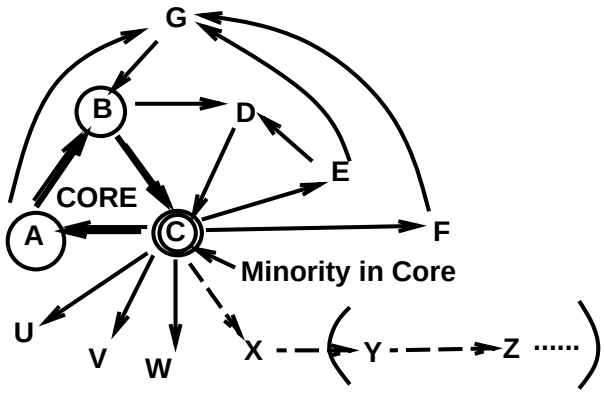

Figure 5: An example of mutually catalytic network in our model. The core network for the recursive state is shown by circles, while parasitic molecules (X,Y,..) connected by broken arrows, are suppressed at the (quasi-)recursive state.

resulting in the collapse of the hypercycle. In the present case, however, other parts of the network, compensate the decrease of the population of C by the parasite, so that the population of A and B are not so much decreased. Hence the complexity in the hypercycle network leads to stability against the attack of parasite molecules. It is also shown that IHN is also relevant to the stability against fluctuations, from the analysis of dynamical systems [Kaneko 2005].

Here minority molecule in the core network has relative importance to sustain the recursive state. Consider a core catalytic cycle with $A \to B \to C \to A$. Here, the number of molecules N_j of molecule species j, is in the inverse order of their catalytic activity $c(j)$, i.e,, $N_A > N_B > N_C$ for $c_A < c_B < c_C$. Because a molecule with higher catalytic activity helps the synthesis of others more, this inverse relationship is expected. Here, the C molecule is catalyzed by a molecule species with higher activities but larger populations (A). Hence, the parasitic molecule species cannot easily invade to disrupt this mutually catalytic network. The stability in the minority molecule is also accelerated by the complexity in IHN. In the IHN, this minority molecule species is involved in several hypercycles as in C, as it has higher catalytic activity. This, on the one hand, demonstrates the prediction in §2.5, that more species are catalyzed by the minority molecules, while on the other hand, leads to the suppression of the fluctuation in the number of minority molecules.

3.3 Switching and Evolvability

Next, we discuss the mechanism of switching. When the total population of molecules in a cell is not so large, the fluctuations are relatively large, especially with regards to the minority molecule. When the number of the minority molecule in the core may decrease due to fluctuations, some parasitic molecules may increase the number. Since the total number of molecules in a cell is limited, the number of minority molecules may decrease, and it may be taken over by the new parasitic molecule species. If this happens, the other molecule species in the original network loses the main source that catalyzes their synthesis. Then successive switching of dominant parasite molecules occur within a few generations, as in the phase (1). After the switching stage, another (or possibly the same)

catalytic network is formed, where existence of a minority molecule species with a higher catalytic activity stabilizes this (quasi-)recursive state. This is the process that occurs in the phase (3).

Note that the phase (3) gives a basis for evolvability, since a novel, quasi-recursive state with different chemical compositions is visited successively. To see this point, we have also studied the present model by including the evolution[Kaneko,2005]. Starting from a network of molecules with low catalytic activity, the evolution to a network with a higher activity is found to occur, through successive switching of (quasi-)recursive states as realized in the phase (3). In this case, this evolution is triggered by extinction of minority molecules in the core network. Evolution from a rather primitive cell consisting of low catalytic activities to that with higher activities is possible, accordingly.

3.4 Statistical Law

Since the number of molecules is not huge, there are relatively large number fluctuations. Is there some statistical law for such fluctuations?

Since the total number of molecules N is not large, say the number fluctuation of each molecule is inevitable even for a recursive production state. Although the fluctuations in the core network (A,B,C in Fig.5) are typically small, for other species in the hypercycle network, the number fluctuation is much larger, and the distribution is close to log-normal distributions

$$P(N_i) \approx \exp(-\frac{(\log N_i - \overline{\log N_i})^2}{2\sigma}), \tag{2}$$

rather than the normal distribution. The origin of the log-normal distributions can be understood by the following rough argument: for a replicating system, the growth of the molecule number N_m of the species m is given by

$$dN_m/dt = AN_m, \tag{3}$$

where A is the average effect of all the molecules that catalyze m. We can then obtain the estimate

$$d\log N_m/dt = \overline{a} + \eta(t), \tag{4}$$

by replacing A with its temporal average $\overline{a}$ plus fluctuations $\eta(t)$ around it. If $\eta(t)$ is approximated by a Gaussian noise, the log-normal distribution for $P(N_m)$ is suggested. This argument is valid if $\overline{a} > 0$. As such this equation diverges with time, but here, the cell divides into two before the divergence becomes significant. Although we need further elaboration of this rough argument [Kaneko 2003a,Furusawa et al.2005], this log-normal distribution seems to be rather universal.

Furthermore, Furusawa and the author[Furusawa and Kaneko 2003, Furusawa et al. 2005] have studied several models of minimal cell consisting of catalytic reaction networks, without assuming the replication process itself. In this class of catalytic reaction network, a huge number of chemical species coexists to form a recursive production. There again, the number of molecules of each chemical species over all cells generally obey the log-normal distribution, for a state with recursive production. Existence of such log-normal

distributions is also experimentally confirmed[Furusawa et al. 2005]. (Furthermore, there is a universal statistical law on the average number, over all molecule species. The rank-abundance law obeys a universal power law, as is also confirmed in the data of gene expression of the present cells[Furusawa and Kaneko 2003].

The fluctuations in log-normal distribution, however, are generally very large, and ranges over digits in the magnitude. This is in strong contrast to our naive impression that a process in a cell system must be well controlled. To have a more precise replication process, some molecules that may deviate from this log-normal distribution should be necessary. Indeed, the minority control mechanism suggests the possibility to suppress the fluctuation, as discussed in §2.2. For a recursive production system, some mechanism to decrease the fluctuation in minority molecule may be evolved. Also, the fluctuation in the chemical in the core network in the IHN is highly suppressed to be deviated from the log-normal distribution. Elucidation of mechanisms to suppress the fluctuation in a reproduction system is an important issue not only theoretically, but also in constructing a stable protocell.

4 Experiment

4.1 Steps to replicating protocell

Recently, there have been some experiments to construct minimal replicating systems in vitro. Let us consider synthesis of a protocell that reproduces itself. For it, we need the following steps;

- (1) A system consisting of chemicals (polymers) with some catalytic activity, which reproduces itself as a set, even though the reproduction may not be precise.
- (2) A compartment structure that splits its inside from the outside as a membrane. This membrane also grows through chemical reactions, so that it divides when its size is large.
- (3) Within the membrane the reaction system (1) works, while the synthesis of membrane is coupled to this internal reaction system.
- (4) The internal reaction process and the synthesis of membrane work in some synchrony, so that a system with membrane and internal chemicals are reproduced recursively.

In our complex systems biology project, together with Tetsuya Yomo and Tadashi Sugawara, the following steps have been achieved.

- i) In-vitro autonomous replication of DNA and protein for (1).
- ii) Repetitive replication of liposome for (2)
- iii) Protein synthesis from RNA as well as the amplification of RNA within a liposome for (3).

We discuss these three topics briefly.

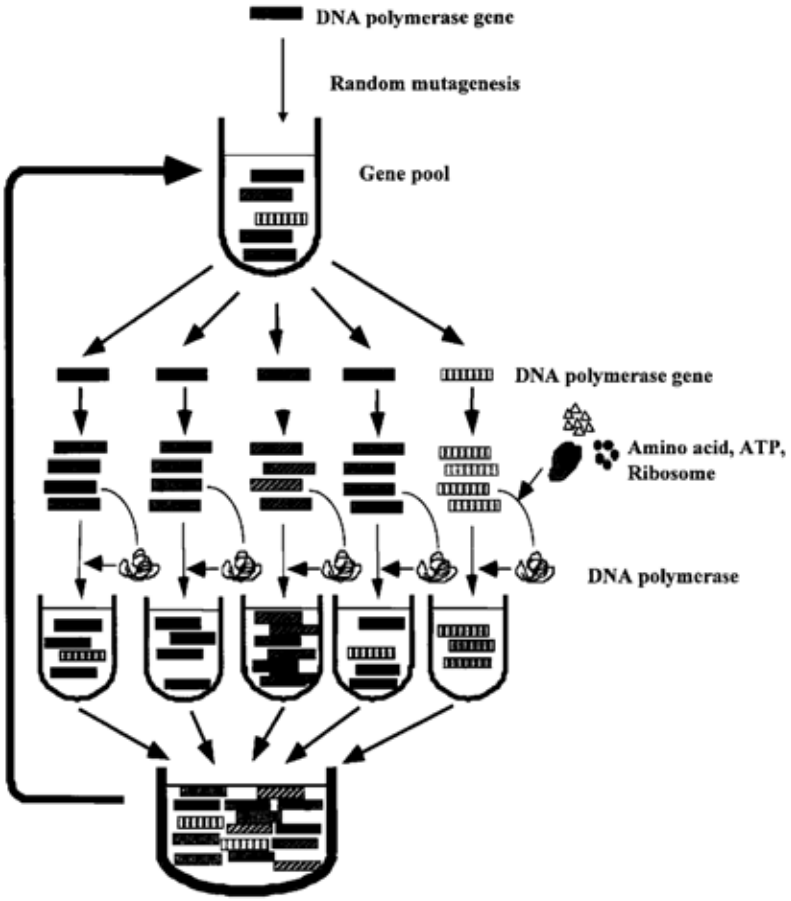

Figure 6: Schematic illustration of in-vitro autonomous replication system consisting of DNA and DNA polymerase. See [Matsuura et al. 2002] for details. Supplied with the courtesy of Yomo, Matsuura et al.

4.2 In-vitro autonomous replication; mutual synthesis of DNA and protein

As an experiment corresponding to this problem, we describe an in-vitro replication system, constructed by Yomo's group[Matsuura et al., 2002].

In general, proteins are synthesized from the information on DNA through RNA, while DNA are synthesized through the action of proteins. As a set of chemicals, they autonomously replicate themselves. Now simplifying this replication process, Matsuura et al[2002] constructed a replication system including DNA and DNA polymerase. This DNA polymerase is synthesized by the corresponding gene in the DNA, while it works as the catalyst for the corresponding DNA. Through this mutual catalytic process the chemicals replicate themselves. Roughly speaking, the polymerase in the experiment corresponds to X in our model, while the polymerase gene corresponds to Y.

As for the amplification of DNA, PCR is a standard tool for molecular biology. In this case, however, enzymes that are necessary for the replication of DNA must be supplied

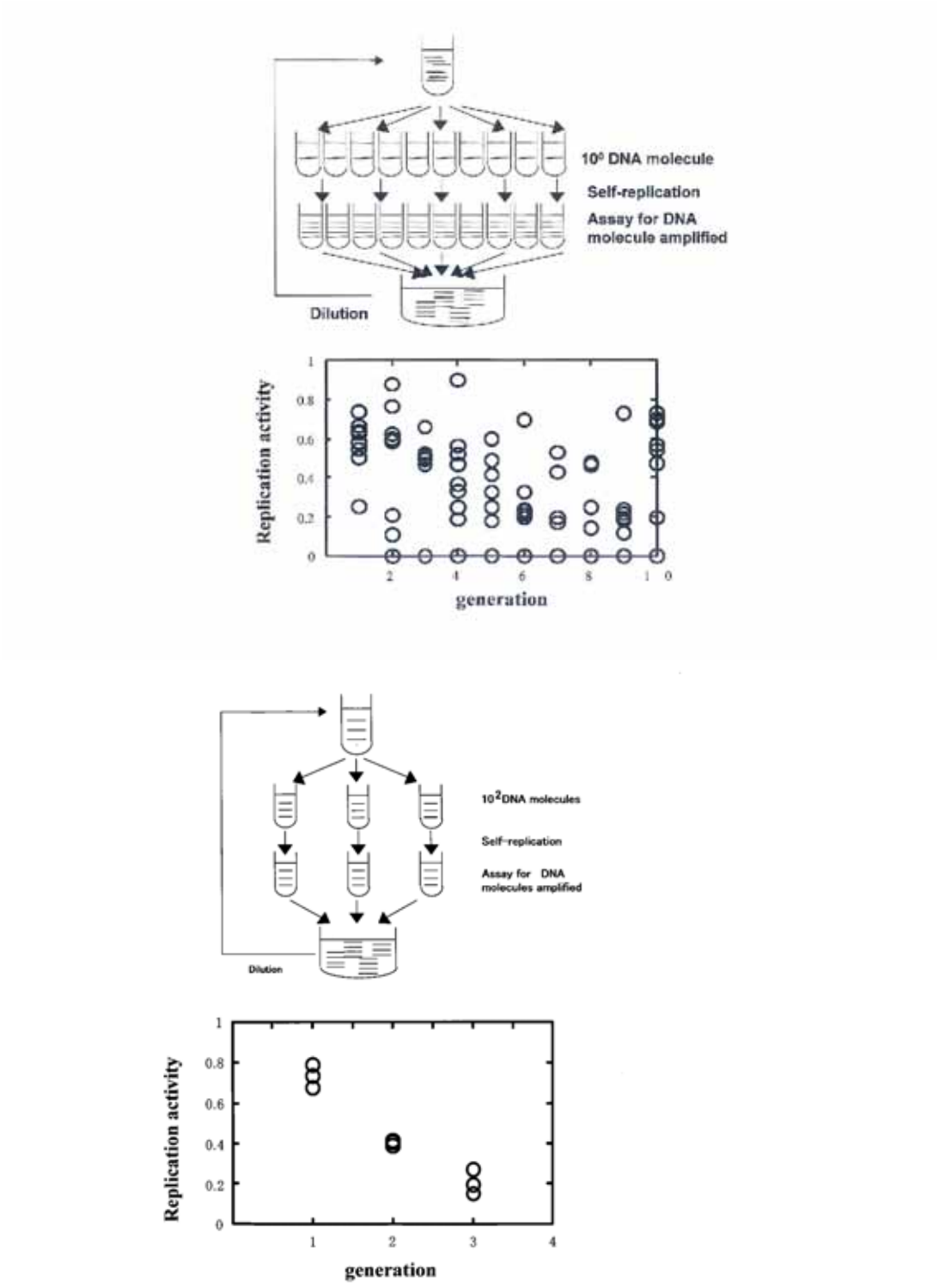

Figure 7: Self-replication activities for each generation, measured as described in the text. The activities for 10 tubes are shown. : Upper: result from a single DNA, where the next generation is produced mostly from the top DNA. Although activities vary by each tube, higher ones are selected, so that the activities are maintained. Lower: result from 100 DNA molecules. Activity decays within 3 generations. Provided with the courtesy of Yomo, Matsuura et al[2002].

externally, and it is not a self-contained autonomous replication system. In the experiment here, while the PCR is adopted as one step of experimental procedures, the enzyme (DNA polymerase) for DNA synthesis also replicates in vitro within the system. Of course, some material, such as amino acid or ATP, have to be supplied, but otherwise the chemicals are replicated by themselves. (see Fig. 6 for the experimental procedure).

Now, at each step of replication , about $2^{30} \sim 2^{40}$ DNA molecules are replicated. Here, of course there are some errors. These errors can occur in the synthesis of enzyme, and also in the synthesis of DNA. With these errors, there appear DNA molecules with different sequences. Now a pool of DNA molecules with a variety of sequences is obtained as a first generation.

From this pool, the DNA and enzymes are split into several tubes. Then, materials with ATP and amino acids are supplied, and this replication process is repeated. In other words, the 'test tube' here plays the role of "cell compartmentalization". Instead of autonomous cell division, split into several tubes are operated externally.

In this experiment, instead of changing the synthesis speed of molecules or N in the theoretical model of §2.1, one can control the number of genes, by changing the condition how the pool is split into several test tubes. Indeed, they studied the two distinct cases, i.e., split to tubes containing a single DNA in each and to tubes containing 100 DNA molecules, and compare the results to test the minority control theory.

First, we describe the case with a single DNA in each tube. Here, the sequence of DNA molecules could be different by each of separated 10 tubes, since there is replication error. Then the activity of DNA polymerase, and accordingly the copy number of DNA therein is also different. Then the contents of each tube are mixed. This soup of chemicals is used for the next generation. Then in this soup, the DNA molecules (or their direct mutants) that have higher replication rate occupy a larger fraction. Then, a (different) single DNA is selected from the soup for each of 10 tubes, and the same procedures are repeated. Accordingly, the probability that a DNA molecule with a higher reproduction activity is selected for the next generation is higher. The self-replication activity from this soup is plotted successively over generations in Fig.7. As shown, the self-replication activity is not lost (or can evolve in some case), although it varies by each tube in each generation.

One might say that the maintenance of replication is not surprising at all, since a gene for the DNA polymerase is included in the beginning. However, enzyme with such catalytic activity is rare. Indeed, with mutations some proteins that lost such catalytic activity but are synthesized in the present system could appear, which might take over the system. Then the self-replication activity would be lost. In fact this is nothing but the error catastrophe by Eigen, discussed in §2.1. Then, why is the self-replication activity maintained in the present experiment?

The answer is clear according to the theory of §2. There (and also in the experiment), mutants that lost the catalytic activity are much more common (i.e., F times larger in §2.1). Still, the number of such molecules is suppressed. This was possible first because the molecules are in a cell. In the experiment also, they are in a test tube, i.e., in a compartment. Now the selection works for this compartment, not for each molecule. Hence the tube (cell) with a lower activity produces a smaller number of offspring. Hence, the compartmentalization is one essential factor for the maintenance of catalytic activity (see also [Boerilst & Hogeweg, 1991,Eigen 1992, Szathmary and Maynard-Smith 1997, Altmeyer & McCaskill 2001], while another important factor is that in each compartment

(cell) there is a single (or very few) DNA molecules.

Recall that in the theory, if the number of Y molecules is larger, inactive Y molecules surpass the active one in population. To check this point, Matsuura et al. split the chemicals in the soup so that each tube has 100 DNA molecules instead of a single one. Otherwise, they adopt the same procedure. In other words, this corresponds to a cell with 100 copies of genome. Change of self-replication activity in the experiment is plotted in the lower column of Fig.7. As shown, the self-replication activity is lost by each generation, and after the fourth generation, capability of autonomous replication is totally lost.

When there are many DNA molecules, there should be variation in these molecules. In each tube, the self-replication activity is given by the average of the enzyme activities from these 100 DNA molecules. Although each catalytic activity of molecules varies, the variance of the average activity by tube is reduced, since the variance of the average of N variables decreases in proportion to $1/N$, according to the central limit theorem of probability theory. Hence the average catalytic activity does not differ much by tube. Since, the mutant with a higher catalytic activity is rare, and most changes in the gene lead to smaller or null catalytic activity, the average, catalytic activity after mutations gets smaller. Since the variation by tube is small, the selection does not work effectively. Hence deleterious mutations remain in the soup, and the self-replication activity will be lost by generations. In other words, the selection works only when the number of information carrier in a replication unit is very small, as is consistent with the theory.

Summary: In the experiment, it was found that replication is maintained even under deleterious mutations, only when the number of DNA polymerase genes is small. Then, the information containing in the DNA polymerase genes is preserved, while the system has also evolvability. This is made possible by the maintenance of rare fluctuations. These experimental results are consistent with the minority control theory described.

4.3 Replicating liposome

All the present cells are surrounded by membrane that consists of bilayer lipid. With this membrane, the in- and out-side of cells are distinguished, while within the membrane catalytic reactions progress. Now it should be important to synthesize such cell structure. This unit with lipid-layer structure is called liposome or vesicle.

In general, oil molecules often form a bilayer membrane with the part inside of the membrane. This membrane often forms a closed spherical structure. If the resource molecules are supplied, this membrane surface increases. Due to the balance with the surface tension, this growth cannot continue forever, and a large membrane will be destabilized. In some case, this results in division of this closed membrane, liposome (or sometimes called vesicle). Indeed Luisi succeeded in synthesizing such division process of liposomes[1] . [Bachmann et al. 1992].

Quite recently, Sugawara's group succeeded in stable replication process that continues from daughter to grand-daughter. By a suitable setup of chemical conditions, liposome (giant vesicle) increases its size absorbing a nutrient chemical, and then it divides into two, and the daughter again continues division [Takakura et al., 2003], while further repetition of divisions is estimated from the analysis of flow-cytometry[Toyota et al.2005].

[1] the terms vesicle or liposome is used in the same meaning, mostly adopted according to the field of research

4.4 Protein Synthesis within a liposome

Then, to achieve the next step, nucleic acids (such as RNA or DNA) as well as the proteins have to be synthesized within the liposome. Since the environment within the liposome is very much 'oily', it is not so easy to achieve this step.

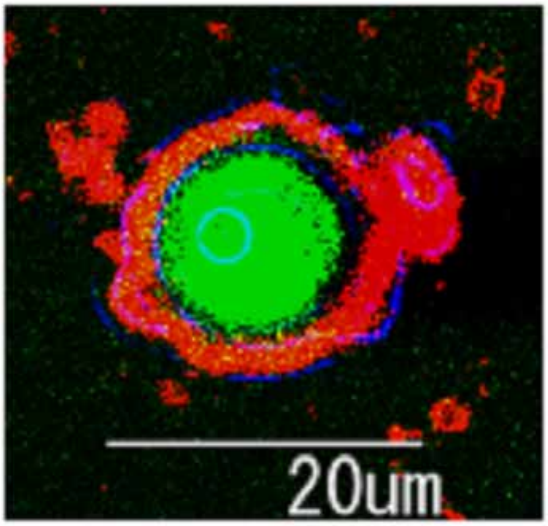

Figure 8: Synthesis of fluorescent protein in a liposome. By using the (green) fluorescent protein, it is shown that the protein is synthesized in a liposome, as observed by the green domain in the figure, which is the inside of the liposome, dyed with red. Supplied with the courtesy of Yomo, Yu, Sato et al. [2002]

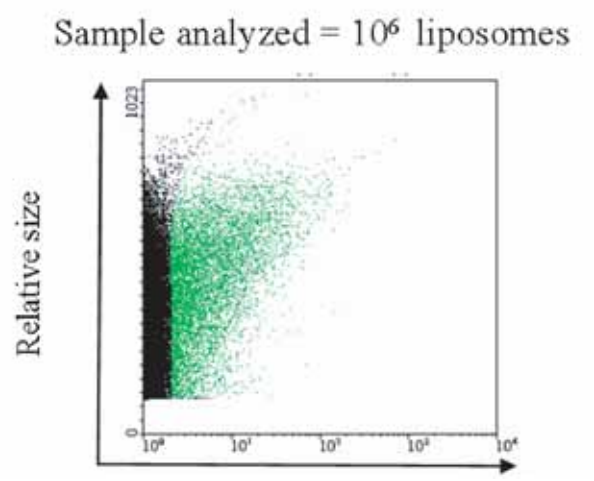

Figure 9: Distribution of abundance of fluorescent protein. Horizontal axis shows the degree of fluorescence, corresponding to the abundances of synthesized protein, and the vertical axis shows the intensity of the forward scatter from the flow cytometry, corresponding to the size of liposome. Supplied by the courtesy of T.Yomo and K. Sato.

Recently Yomo's group have succeeded in the amplification of RNA and the synthesis of protein from mRNA [Yu et al. 2001]. Within the liposome of the diameter $1\mu m$, transcription of RNA, i.e., the synthesis of proteins from the mRNA, is shown to occur. By making the corresponding protein fluorescent (by using Green Fluorescent Protein) , Yu et al. [2001] have shown that the synthesis indeed occurs within the liposome by the measurement of fluorescence. This rate of protein synthesis is high enough to measure the fluorescence of protein, by using the flow cytometry. In Fig.9, plotted are two-dimensional distribution of the size of liposome and fluorescence, with longitudinal axis as size and horizontal axis as fluorescence intensity. The liposomes at the left region correspond to those that could not synthesize proteins well. Hence by selecting liposomes by using the cell sorter, one can select 'good' RNA to have a higher protein synthesis within the liposome. Now the evolution to a better RNA is possible.

On the other hand, this selection process can work for choosing a better 'oil' molecule for liposome. Then evolution to have a liposome that is robust including the protein synthesis system within. By repeating the process the third step (of §4.1) for artificial cell will be achieved.

Finally, we study the distribution of protein abundances. Since the fluorescent protein is synthesized with the aid of other molecules (such as RNA), the fluctuations should be large, and the distribution will be log-normal, if the argument of §3.4 is valid. Indeed, the distribution obtained from the experiment is rather close to log-normal distribution.

4.5 For the synthesis of artificial replicating cell

The last step for the synthesis of artificial replicating cell is to combine the replication of liposome and replication of protein within. Here it is important to balance the two processes. If the replication of membrane is faster, then the inside ingredients will be sparse, while if the protein synthesis is faster the density of molecules will be too high to destroy the liposome. Hence it is important to balance the two processes to make the recursive production. To achieve this balanced replication, some link between the liposome growth and the internal reaction will be required, which has not yet been achieved. For achieving the balance, some condition discussed in §3.4 may be required, while the regulation of large fluctuations will be also crucial.

Another diffusivity lies in the interference of several processes. Success of each step in (1)-(3) in §4.1, assumes the separation of each processes. When we try to combine the processes, they interfere each other, resulting in the collapse of each process. To achieve the separation of processes, minority control mechanism and the switching process in §3.3 may be relevant.

One might think that there will be a long way to construct an autonomous replication of artificial cell. However, once a system with loose reproduction at any form is realized, then such 'cell' can be an object for Darwinian selection process. By selecting a relatively 'better' reproducing cell through cell sorter, a more reliable replication system will be obtained.

Acknowledgments

I would like to thank Tetsuya Yomo, Kanetomo Sato and Chikara Furusawa for discussions and collaborations.

References

S.Altmeyer and J.S. McCaskill "Error threshold for Spatially resoved Evolution in the Quasispecies Model", Phys. Rev. Lett. 86 (2001) 5819-5822

A.G. Cairns-Smith, Clay Minerals and the Origin of Life(1982), Cambridge Univ. Press.

Bachmann P.A, Luisi P.L. ,Lang J., Autocatalytic Self-Replicating Micelles As Models For Prebiotic Structures; Nature 357: (1992) 57-59

M. Boerlijst and P. Hogeweg, Physica 48D (1991) 17; Spiral Wave Structure in Prebiotic Evolution – Hypercycles Stable against Parasites

Freeman Dyson, *Origins of Life*, Cambridge Univ. Press., 1985

M. Eigen, P. Schuster, *The Hypercycle* (Springer, 1979).

M. Eigen, *Steps towards Life*, Oxford Univ. Press., 1992

Furusawa C. & Kaneko K., "Emergence of Rules in Cell Society: Differentiation, Hierarchy, and Stability" Bull.Math.Biol. 60; 659-687 (1998)

C. Furusawa and K. Kaneko ; Origin of Multicellular Organisms as an Inevitable Consequence of Dynamical Systems, Anatomical Record, 268 (2002) 327-342

C. Furusawa and K. Kaneko ; Zipf's Law in gene expression, Phys. Rev. Lett., 90 (2003) 088102

C. Furusawa T. Suzuki, A. Kashiwagi, T. Yomo, and K. Kaneko ; Ubiquity of Log-normal Distributions in Intra-cellular Reaction Dynamics, BIOPHYSICS 1, (2005) 25

K. Kaneko, "Life as Complex Systems: Viewpoint from Intra-Inter Dynamics", Complexity, 3 (1998) 53-60

K. Kaneko "Kinetic Origin of Heredity in a Replicating System with a Catalytic Network J Biol. Phys., 28 (2002) 781-792

K. Kaneko, "Recursiveness, Switching, and Fluctuations in a Replicating Catalytic Network", Phys. Rev.E. 68 (2003a) 031909

K. Kaneko, "What is Life?: A complex Systems Appreoach", Univ of Tokyo Press (2003b) (in Japanese, translation in progress)

K. Kaneko; "Constructive and Dynamical Systems Approach to Life" in *Function and Regulation of Cellular Systems: Experiments and Models* 213-224, ed. by A.Deutsch, M.Falcke, J.Howard, W.Zimmerman (Birkhaeuser Verlag Basel, 2004)

K. Kaneko; On Recursive Production and Evolvabilty of Cells: Catalytic Reaction Network Approach; Adv. Chem. Phys. 130 (2005) 543-598

K. Kaneko and T. Yomo, "Isologous Diversification: A Theory of Cell Differentiation ", Bull.Math.Biol. 59 (1997) 139-196

Kaneko K. & Yomo T, "Isologous Diversification for Robust Development of Cell Society ", J. Theor. Biol., 199 243-256 (1999)

Kashiwagi A., Noumachi W., Katsuno M., Alam M.T., Urabe I., and Yomo T. "Plasticity of Fitness and Diversification Process During an Experimental Molecular Evolution", J. Mol. Evol., (2001) **52** 502-509 .

E. Ko, T.Yomo, I. Urabe, Physica 75D (1994) 81-88 Dynamic Clustering of bacterial population

Matsuura T., Yomo T., Yamaguchi M, Shibuya N., Ko-Mitamura E.P., Shima Y., and Urabe I. "Importance of compartment formation for a self-encoding system", Proc. Nat. Acad. Sci. USA 99 (2002) 7514-7517

D.R. Mills, R.L. Peterson, and S. Spiegelman, An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule, Proc. nat. Acad. Sci. USA 58 (1967) 217

Segre D, Ben-Eli D, Lancet D., "Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies", Proc Natl Acad Sci USA 97 (2000)4112

E. Szathmary and J. Maynard Smith, "From Replicators to Reproducers: the First Major Transitions Leading to Life", J. Theor. Biol. 187 (1997) 555-571

K. Takakura, T. Toyota and T. Sugawara, A Novel System of Self-Reproducing Giant Vesicles, J. Am. Chem. Soc., 125 (2003) 8134-8140

T. Toyota, K. Takakura, T. Yomo, K. Kaneko, and T. Sugawara, in prepartion (2005)

Yu,W., Sato, K., Wakabayashi, M., Nakaishi, T., Ko-Mitamura, E. P., Shima, Y., Urabe, I., Yomo, T. (2001) Synthesis of functional protein in liposome. J. Biosci. Bioeng., 92(6): 590-593.