OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Adaptation, diversification, and increase in complexity of an artificial RNA genome replication system
Author(s)	水内, 良
Citation	大阪大学, 2018, 博士論文
Version Type	VoR
URL	https://doi.org/10.18910/69727
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka

Adaptation, diversification, and increase in complexity of an artificial RNA genome replication system

Submitted to
Graduate School of Information Science and Technology, Osaka University

January 2018

Ryo Mizuuchi

Accomplishments Related to this Dissertation

1. Publications

Journal articles

- **Ryo Mizuuchi**, Norikazu Ichihashi, Tetsuya Yomo (2016). Adaptation and diversification of an RNA replication system under initiation- or termination-impaired translational conditions. *ChemBioChem*, 17 (13), 1229-1232. **(Chapter 3)**

- **Ryo Mizuuchi**, Norikazu Ichihashi, Kimihito Usui, Yasuaki Kazuta, Tetsuya Yomo (2015). Adaptive Evolution of an Artificial RNA Genome to a Reduced Ribosome Environment. *ACS Synthetic Biology*, 4 (3), 292-298. **(Chapter 2)**

Conference papers

- **Ryo Mizuuchi**, Norikazu Ichihashi, Kimihito Usui, Tetsuya Yomo, "Evolution and Adaptation of the RNA Coupled with an Artificial Life-Like Self-Replication System to a Severe Translational Environment" *Proceeding of International Astrobiology Workshop 2013*, 2013, 1006. **(Chapter 3)**

2. Conference Presentations

(Delivered by the author of this dissertation at international conferences)

Oral presentations

- **Ryo Mizuuchi,** Norikazu Ichihashi "Darwinian evolution of mutualistic RNA replicators with different genes" The XVIIIth International Conference on the Origin of Life, #4077, La Jolla, California, United States, July 2017. **(Chapter 4)**

- **Ryo Mizuuchi,** Norikazu Ichihashi "Darwinian evolution of two distinct cooperative RNA replicators" Astrobiology Science Conference 2017, #3030, Mesa, Arizona, United States, April 2017. **(Chapter 4)**

- **Ryo Mizuuchi** "Adaptation, diversification, and evolution of complexity of a simple artificial RNA self-replication system" 3rd NoR HGT & LUCA conference, Milton Keynes, United Kingdom, November 2016. **(Chapter 3)**

- **Ryo Mizuuchi**, Norikazu Ichihashi, Kimihito Usui, Yasuaki Kazuta, Tetsuya Yomo, "Adaptive Evolution of an Artificial RNA through in vitro Translation-Coupled Replication System to a Severer Translational, Less Ribosome Environment" Origins2014: 2nd ISSOL – The International Astrobiology Society and Bioastronomy Joint International Conference, CE007, Nara, Japan, July 2014. **(Chapter 2)**

- **Ryo Mizuuchi**, Norikazu Ichihashi, Kimihito Usui, Yasuaki Kazuta, Tetsuya Yomo, "Evolution

and adaptation of a life-like complex system to severe environments" OWECS2013: OIST Winter Course "Evolution of Complex Systems", WS13, Okinawa, Japan, December 2013. **(Chapter 2)**

Poster presentations

- **Ryo Mizuuchi**, Norikazu Ichihashi, "Darwinian evolution of two distinct cooperative RNA replicators" Humanware International symposium, 15, Osaka, Japan, January 2017. **(Chapter 4)**

- **Ryo Mizuuchi**, Norikazu Ichihashi, "Construction and evolution of an RNA replication system with two distinct replicators linked in a hypercyclic network" ELSI 5th international symposium "Expanding Views on the Emergence of Biospheres", S2-P2, Tokyo, Japan, January 2017. **(Chapter 4)**

- **Ryo Mizuuchi**, Norikazu Ichihashi, Tetsuya Yomo, "Adaptive Evolution and Diversity of an Artificial RNA Self-Replication System" Astrobiology Science Conference 2015, #7163, Chicago, Illinois, United States, June 2015. **(Chapter 3)**

- Tetsuya Yomo, **Ryo Mizuuchi***, Yohsuke Bansho, Norikazu Ichihashi, "An evolvable artificial cell model" ELSI 3rd international symposium"Life in the Universe", P2-09, Tokyo, Japan, January 2015. **(*speaker) (Chapter 2)**

- **Ryo Mizuuchi**, Norikazu Ichihashi, Kimihito Usui, Yasuaki Kazuta, Tetsuya Yomo, "Adaptive evolution of a genomic RNA in an artificial cell-like model" CEITEC Annual Conference "Frontiers in Material and Life Sciences", 102, Brno, Czech Republic, October 2014. **(Chapter 2)**

- **Ryo Mizuuchi**, Norikazu Ichihashi, Kimihito Usui, Tetsuya Yomo, "Evolution and Adaptation of the RNA Coupled with an Artificial Life-Like Self-Replication System to a Severe Translational Environment" The International Astrobiology Workshop 2013, #1006, Sagamihara, Japan, November 2013. **(Chapter 3)**

3. Awards and Honors

- Excellent poster presentation award, 19th Annual Meeting of the Society of Evolutionary Studies, Japan, P-03, Kyoto, Japan, August 2017. **(Chapter 4)**

- The NASA Astrobiology Program award (Best poster presentation, 3rd place), Astrobiology Science Conference 2015, #7163, Chicago, Illinois, United States, June 2015. **(Chapter 3)**

- PhD students honorable (honorable poster presentation), CEITEC Annual Conference "Frontiers

in Material and Life Sciences“, 102, Brno, Czech Republic, October 2014. **(Chapter 2)**

Other Accomplishments

1. Publications

- Hiroshi Kawase, **Ryo Mizuuchi**, Hirofumi Shin, Yuki Kitajima, Koh Hosoda, Masahiro Shimizu, Daisuke Iwai, Shigeru Kondo (2017). Discovery of an earliest-stage “mystery circle” and development of the structure constructed by pufferfish, Torquigener albomaculosus (Pisces: Tetraodontidae). *Fishes*, 2 (3), 14.

2. Conference Presentations

(Delivered by the author of this dissertation at international conferences)

- **Ryo Mizuuchi**, Milena Popović, Andrew Pohorille, Mark A. Ditzler, “Exploring length dependence in RNA evolution through in vitro selection of ligase activity” RNA 2016, 377, Kyoto, Japan, June 2016.

3. Awards and Honors

- Overseas Research Fellowship, Japan Society for the Promotion of Science, Japan, 2018 April – 2020 March
- Research Fellowship for Young Scientists (DC1), Japan Society for the Promotion of Science, Japan, 2015 April – 2018 March.
- Award of the Graduate School of Information Science and Technology of Osaka University (graduation top of the class), March 2015.
- Osaka University Engineering Award (graduation top of the class), School of Engineering, Osaka University, March 2013.

Abstract

Reconstituting biological abilities from known molecules is a direct approach to understand sufficient factors for biological phenomena. It also provides a novel biological system to advance biotechnology and contributes to understanding the origin and the early evolution of life. To date, various important biological functions were reconstituted in vitro, such as genome replication, translation of genetic information, and dynamic boundary to encapsulate biological components. However, an artificial system that has sufficient ability of evolution has not been achieved.

Evolution is one of the most remarkable characteristics of living organisms, which has enabled them to persist for four billion years and caused the emergence of diverse and complex species. It is believed that through the evolution, the organisms diversified by adapting to different environments and gradually increased their complexity by acquiring new components and genes. Therefore, a reconstituted evolutionary ability must enable adaptation and the increase in complexity.

In a previous study, a translation-coupled RNA replication (TcRR) system was constructed by combining a translation system and an RNA genome encoding a replication enzyme. A subsequent study found that the RNA genome self-improved the TcRR activity through repetitive replications in a cell-like compartment, but the experiment was performed only in an optimized condition. Therefore, its ability to adapt to various severe conditions and the possibility of becoming a more complex system are unknown. On this basis, I investigated those potentials using the TcRR system.

In the first half of the dissertation, I examined the adaptation ability of the TcRR system. First, I repeated the replications in a severe translation condition, reduced ribosome concentrations, and proved that the RNA genome gradually improved the TcRR activity and accumulated mutations in the condition, which is an evidence of adaptation. I further showed that a possible adaptation mechanism was a simple RNA structural modification. Next, I repeated the replications in newly prepared four different translation-impaired conditions and found that the RNA genome accumulated condition-specific mutations and diversified toward distinct genome sequences. Taken together, I proved that the TcRR system has a certain ability of adaptation.

In the latter half of the dissertation, I focused on the other major evolutionary phenomenon, the increase in complexity by acquiring new cooperative replicating entities, which has been considered unstable due to spontaneously appearing parasitic replicators that lost functional genes. By mimicking a plausible evolutionary process, I added another replicating RNA encoding a metabolic enzyme to the TcRR system so that two RNAs cooperatively replicate through translation of the different genes. Through a long-term replication, I found that the increased complexity can be evolutionary sustained in a certain range of RNA concentrations by overcoming the parasite problem, the biggest barrier for the evolution of complexity. I also proved that the complex system kept the ability to further self-improve the replication. Overall, a simple artificial RNA replication system has considerable evolutionary potential to adapt to various conditions and become more complex.

Table of contents

Chapter 1. Introduction.....1

1-1. Introduction.....1

1-2. Construction of the Dissertation.....7

Chapter 2. Adaptation of an RNA Replication System.....9

2-1. Introduction.....9

2-2. Materials and Methods.....11

2-2-1. Materials.....11

2-2-2. RNA Preparation (*in vitro* transcription).....13

2-2-3. Translation-coupled RNA Replication (TcRR) in Water-in-oil Emulsion.....14

2-2-4. Translation-coupled RNA Replication (TcRR) in Bulk Conditions.....14

2-2-5. Quantitative RT-PCR.....14

2-2-6. Evolution Experiment.....15

2-2-7. Preparation of RNA clones.....16

2-2-8. Sequence Analysis.....16

2-2-9. Mathematical Model of the Translation-Coupled Minus RNA Synthesis.....17

2-2-10. Analysis of Translation Activities.....18

2-2-11. Analysis of Replication Activities.....18

2-2-12. Replication Activity of the Clones under Uncoupled Conditions.....18

2-2-13. Analysis of Encoded Replicase Activities.....18

2-2-14. SHAPE Analysis.....18

2-3. Results and Discussion.....20

2-3-1. Evolution Experiment under Reduced Ribosome Conditions.....20

2-3-2. Mutation Analysis.....20

2-3-3. Self-replication Activities of the Evolved RNA Clones.....21

2-3-4. Biochemical Analysis of the Evolved RNA Clones.....23

2-3-5. Encoded Replicase Activities of the evolved Clones.....25

2-3-6. Analysis of the RNA Structure of the Evolved RNA Clones.....25

2-4. Chapter Summary.....29

Chapter 3. Diversification of the RNA Replication System.....30

3-1. Introduction.....30

3-2. Materials and Methods.....31

3-2-1. Evolution Experiment.....31

3-2-2. Sequence Analysis.....31

3-2-3. Phylogenetic analysis.....31
3-3. Results and Discussion 32
3-3-1. Evolution Experiment under Multiple Translation-impaired Conditions.....32
3-3-2. Mutation Analysis.....33
3-3-3. Phylogenetic Tree Analysis 34
3-3-4. Mutation Sites and the Potential Effect.....36
3-4. Chapter Summary 38

Chapter 4. Expansion of the RNA Replication System and Sustainable Evolution 39
4-1. Introduction 39
4-2. Materials and Methods 41
4-2-1. Materials 41
4-2-2. Modification of NDK-RNA Structures 42
4-2-3. TcCRR Reaction of the Cooperative RNA Replication System.....42
4-2-4. Simulation 43
4-2-5. Long-term Replication Experiment.....45
4-2-6. Detection of the Parasitic RNAs.....46
4-2-7. Sequence Analysis.....46
4-2-8. Construction of the isolate clones with dominant mutations (Evo). 46
4-2-9. Analysis of Replication Activities 46
4-2-10. Analysis of Enzyme Activities.....46
4-3. Results and Discussion 48
4-3-1. Instability of Cooperative RNA Replication System, Proved by Computer Simulation.....48
4-3-2. Construction of a Translation-coupled Cooperative RNAs Replication (TcCRR) System.....50
4-3-3. Long-term Replication Experiment. (1) Destruction of the TcCRR System.....53
4-3-4. (2) Sustainable Replication of the TcCRR System.....54
4-3-5. Mutation Analysis.....57
4-3-6. Evolution of the TcCRR System.....57
4-3-7. Enzyme Activity of all the Obtained Clones......60
4-3-8. Coordinated RNA Concentration and Chromosome Structures 61
4-4. Chapter Summary 63

Chapter 5. Summary 64
5-1. Perspective of Constructive Biology.....64

5-2. Perspective of Origin-of-Life Research.....67

Appendix.....70
A2-1. The Sequence of the Original RNA Clone 70
A2-2. The Sequence of VectorF 71
A2-3. Mutation List of the Evolved Clones (round 11 (R11), 15 (R15), and 30 (R30)).....73
A2-4. Replication Activity of the Clones under Uncoupled Conditions 75
A2-5. Translation and Replication Activities at High Ribosome Concentrations.....76
A3-1. Mutation List of the Evolved Clones in the condition A.....78
A3-2. Mutation List of the Evolved Clones in the condition B.....79
A3-3. Mutation List of the Evolved Clones in the condition C 80
A3-4. Mutation List of the Evolved Clones in the condition D 81
A4-1. The Sequence of the Original NDK-RNA.....82
A4-2. Mutation List for Improvement of NDK-RNA Translation and Replication Activities 83
A4-3. Mutation List of the Evolved Rep-RNA Clones 84
A4-4. Mutation List of the Evolved NDK-RNA Clones 86

Reference 87

Acknowledgements 95

Chapter 1. Introduction

1-1. Introduction

Reconstructing biological functions from defined biological molecules allows us to understand the essential factors for life[1,2,3,4], advance biotechnology[1,2,3,5], and provide insights into the origin and early evolution of life[3,6,7,8,9,10]. Even in *Escherichia coli*, the best studied model organism, the roles of about 50 % of the genes are unidentified[11]. This unknown property of natural organisms is a major barrier for understanding crucial biological mechanisms and free design of microbes for various applications. On the other hand, an artificial system that imitates biological abilities can be constructed from only defined molecules, which means we have a detailed blueprint to generate biological phenomena. We can therefore understand what kinds of molecules and conditions are sufficient to achieve target biological functions. Detailed analysis of the minimal system also provides insights into the basic mechanism of integrating distinct molecules to work in concert, giving rise to specific biological phenomena.

The artificial systems would develop new biotechnology such as useful material production and medical treatment. Bacteria are used most commonly for organism-based production of desired molecules, but bacterial modification technology has not been fully developed for suitable and adequate production. For example, the massive production of antibodies and membrane proteins in bacteria has remained challenging because it requires complicated genetic modification techniques and specialized nutrients media. Moreover, bacteria are usually incompatible with the production of toxic molecules, which limits possible applications of modified bacteria. In contrast, in principle, artificial systems can be customized to produce any desired chemical materials. The *in vitro* systems also have expansibility to acquire new mechanisms to modify target materials, simultaneously produce multiple materials, self-regulate the productions, and spontaneously improve the productions by coupling with evolutionary technology. A good example of such an artificial system for material production is reconstituted translation systems, developed by gathering purified translation proteins, including the PURE system, derived from *E.coli*[12,13]. The reconstituted translation has many advantages over bacterial translation such as RNase/protease-free reaction, flexible incorporation of unnatural amino acids, and compatible production of toxic proteins. It has also been shown to perform reactions inside cell-like compartments[14,15], which enabled contamination-free reactions and evolutionary optimization. Furthermore, life-like systems, such as drug delivery system (drugs in cell-like compartments)[16] and artificial blood cells[17], can be used for risk-free medical treatment because we can easily control and predict their behaviors inside human bodies, difficult for living organisms.

Another important role of reconstructing biological functions is to provide insights into the origin and the early evolution of life. Primitive life-forms are generally considered as much simpler replicating systems than any living organisms as we know. Therefore, artificial life-like systems,

imitating biological functions with small number of components, could be a model for possible forms of primitive life. Thus, although it may not have truly existed on the early Earth, analyzing the simple model gives ideas about possible scenario of the early evolution of life. For example, many studies constructed simple membrane-based compartments and showed that they can spontaneously form, grow, and divide without enzymes[7,18,19,20,21,22]. Thus, the replication of compartments was found to be possible before the emergence of proteins. Other studies constructed various types of replications, such as RNA-catalyzed RNA replications[23,24], proteins-based RNA replications[25,26], and an unnatural DNA replication[27]. These examples highlighted diverse possibilities of early replication mechanisms.

What system comes next? One of the challenges is to reconstruct evolutionary ability, a remarkable characteristic of life, through which replicating systems can survive in various environments. Evolution mainly occurs as a result of three types of features: genome replication, generation of inheritable variation through replication, and selection in an environment[28]. Through the most common type of genome replication, mutations are spontaneously introduces into the genome as replication errors, generating inheritable genetic variation, and expressing the varied genetic information produces corresponding phenotypic variation. Then, mutant replicating systems that fit in the environment (e.g., grow faster in the environment) have more likely to survive (be selected) because they have more chance to pass down their traits into descendants than other unfit replicating systems.

Through the long evolutionary history of life, it is believed that there were two major recurring phenomena: diversification of species, mainly achieved by adaptation to diverse environments, and increase in complexity through acquisition of new genes and components, caused by integrating distinct replicating species. Adaptation allows living things to survive under various conditions by changing genome sequences and differentiate into distinct species. By comparing genome sequences among diverse living species, it was found that living organisms can be classified into three domain of life, named bacteria, archaea, and eukaryote[29]. The bacteria and archaea are unicellular organisms that lacks nucleus and are included in prokaryotes. The sequence comparison also revealed that the emergences of each domain were independent events in the life's evolutionary history, and after the emergences, their genomes differentiated in each domain toward forming highly diverse species in the contemporary life. This diversification process, however, does not explain the emergence of each domain. For example, it is hard to consider that eukaryotic cells, containing mitochondria with a distinct genome or other organelle, appeared through genome diversification by natural selection acting on mutations; eukaryotic cells were generally considered to have originated through integrating two or more distinct cell lineages by symbiotic interaction (called endosymbiosis)[30]. This process is known as one of the major transitions in evolution: establishing cooperative relationships

between distinct replicating species so that each replicator would be a part of a larger and more complex unit, increasing in complexity through acquisition of new genes and components[31,32]. Other examples of the emergence of higher order of organization includes the appearance of multicellular organisms from unicellular organisms, the emergence of chromosome structure (encoding multiple genes) from independently replicating genomes with different genes, and integration of RNA replicators in the early evolution of life. It should be noted that there is a seemingly differently categorized phenomena, evolution of degenerate genomes (genome reduction or loss of genes) through long-term symbiosis. However, this event can be explained by co-adaptation of hosts and symbionts to maximize their fitness by reducing redundant genes[33], and thus, the process can be categorized in diversification by adaptation described above. Therefore, the evolutionary phenomena can be divided into the two processes, diversification of species by adaptation and increase in complexity. Reconstituting such evolutionary phenomena in a test tube greatly contributes to the understanding of the core of life and possible life's history. The development of evolvable systems would also provide a new platform for material production because the products could self-improve through evolution. Thus, in the present study, I focused on adaptation and increase in complexity from the perspective of reconstitution.

It has been considered that an important driving force to achieve the evolution of complexity is cooperation between independently replicating systems or molecules[31,32,34]. For example, eukaryotic cells or multicellular organisms can be evolved by cooperation between unicellular organisms so that their replication synchronizes as a unit. However, such evolution of complexity is unstable by spontaneously appearing selfish replicators (or called defector, parasite, cheater etc. depending on scientific fields) that maximize their own growths by exploit cooperation from surrounding replicators. Because natural selection favors fast replicators, such selfish replicators are likely to be selected, disrupting newly established complex systems (co-existence of distinct replicators). Therefore, one of the most important things to demonstrate whether evolution of complexity is possible is to elucidate possible conditions in which cooperative replicators can be sustained by overcoming selfish replicators through natural selection.

In a previous study, a simple translation-coupled RNA genome replication (TcRR) system was constructed by combining an RNA genome and about a hundred translation proteins (an *E.coli* reconstituted translation system, the PURE system[12])[35] (Fig. 1-1). The TcRR system was the first artificial genome replication system in which a genome replicates in the same manner as living organisms, through translation of a self-encoded protein. In the TcRR system, the RNA genome encodes a subunit of an RNA-dependent RNA polymerase (RNA replicase) which forms a complex with EF-Tu and EF-Ts in the PURE system, and the genome replicates through translation of the self-encoded RNA replicase. Through the replication, mutations are spontaneously introduced into

the genome as replication errors, generating inheritable variation. Despite the great simplicity, a subsequent research showed that if encapsulated in cell-like compartments to link genotypes and phenotypes, the artificial system spontaneously improved the growth through repetitive replications, as selected in a condition[36]. Thus, the TcRR system has the ability of evolution, although it was an experiment performed under only one optimized condition. As a next challenge, I attempted to study adaptation ability and possibility to increase complexity of the TcRR system. In particular, I investigated whether the TcRR system can adapt to severe conditions through repetitive replications, whether the RNA genome diversified depending on the conditions where it repeated replication, and whether the TcRR system can be expanded to acquire another gene and continue sustainable replication and further improve replication.

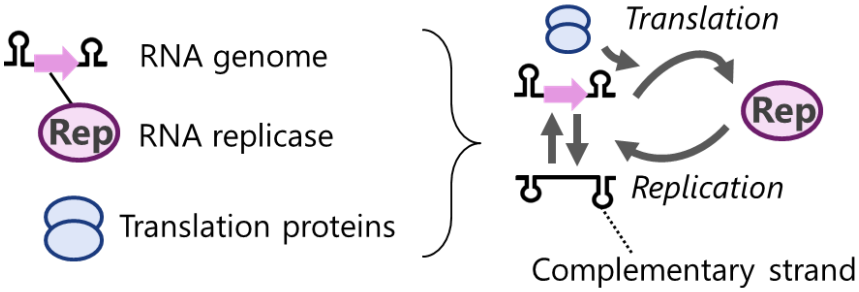


Figure 1-1. Translation-coupled RNA replication (TcRR) system

The RNA genome replicates through the translation of the self-encoded RNA replicase. The synthesized complementary strand is also recognized by the replicase to produce the original strand.

Some previous studies constructed artificial evolvable RNA replication systems[25,26]. They repetitively replicated RNAs and showed that mutations were accumulated and clones with higher replication activities were selected. Subsequent studies also showed that such replication system can adapt to severe environments, including in the presence of inhibitors[37,38]. However, in their replication systems, the RNAs did not encode genes and replicated through external supply of replication polymerase enzymes necessary for their replications, which is not the replication scheme of living organisms: genome replicates through translation of self-encoded polymerase enzymes. Requirement of translation may restrict possible evolution compared to the replication systems that were not coupled with translation because genomes must express functional proteins, and many mutations are deleterious to the proteins' functions. Therefore, whether an artificial replication system coupled with translation, including the TcRR system, can adapt to various environments is not explicitly explained. Furthermore, the RNA genome in the TcRR system is only 2 kb, much shorter than living organisms, and it encodes only one gene of RNA replicase. Whether or not evolutionary ability of translation-coupled RNA replication systems can be achieved with only one

gene is another important question to understand sufficient components for evolutionary ability of life. Taken together, in the present study, I attempted to elucidate whether an artificial replication system that replicates through translation of a self-encoded protein, only one gene of RNA replicase, can adapt to various environments and evolve to diversify the genomes sequences.

Another important feature of the TcRR system is that it is a minimal replication system consisting of RNAs and proteins, and the RNA genome replicates through the translation of a self-encoded RNA replicase enzyme. Such replication systems lacking DNA is generally considered to have existed before the advent of modern DNA world, where living organisms use DNA as their genomes, and after the RNA world, in which RNA catalyzes the replication of RNA genomes (Fig. 1-2)[39,40,41]. This period is called the RNA-protein world. It has been considered that replication systems during the RNA-protein world contained an RNA genome encoding replication-related proteins, and the genome replicates via translation, through which inheritable variation generated as mutations, allowing evolution through selection in environments. Thus, the TcRR system can be considered as a model for a possible replicating system in the RNA-protein world in terms of RNA genome replication through the translation of the self-encoded RNA replicase. In the present study, I investigated the evolutionary ability of the TcRR system, a replication system coupled with translation, contributing to understanding sufficient conditions to achieve evolutionary ability and the possible evolutions of primitive replication systems in the RNA-protein world.

A previous study constructed a similar artificial RNA replication system by combining only an RNA and proteins, and it showed that the RNA spontaneously evolved to replicate faster through repetitive replications[25]. However, in the replication system, RNA replicase was manually supplied to replicate the RNA, which is not a plausible RNA replication scheme in the RNA-protein world. Moreover, although a subsequent research showed that the system adapted to a sever condition[37], whether the system can adapt to diverse conditions and the RNA sequence accordingly diversifies was not investigated. Therefore, my research on adaptation of the TcRR system made at least two progress in the origin-of-life field: investigation of (1) adaptation ability of an RNA replication system that replicates through the translation of a self-encoded protein (a plausible replication mechanism in the RNA-protein world) and (2) adaptation and diversification of an RNA genome in various conditions. It should be noted that other research constructed different types of evolvable replication systems, in which RNA replicates through the conversion to DNA, or DNA replicates, although evolution experiments were performed under limited conditions [26,42,43].

For understanding the evolution of complexity in primitive replication systems, especially in the RNA world or the RNA-protein world, research has been mainly conducted from the theoretical perspective[44,45,46,47,48,49,50,51,52,53,54]. Experimental approaches were hindered because of the lack of an artificial replication system that resembles the following two theoretical assumptions. (1) Multiple

RNAs with different genetic information cooperatively replicate, thorough which parasitic RNAs spontaneously appear by mutations that make the cooperative RNAs better templates to be replicated faster by exploiting the cooperation, which has been considered the major problem in the increase in complexity. (2) A set of the cooperative replicators are compartmentalized or spatially divided, which has been considered to solve the parasite problem. The TcRR system is known to spontaneously produce such parasitic RNAs by replication errors[55,56] and can be encapsulated in water-in-oil emulsion. Therefore, expanding the system by adding another cooperative RNAreplicator encoding a different gene would recapitulate the theoretical ideas by experiments for the first time. Furthermore, the experiments would provide direct evidence about what parameters of the RNAs evolve if their replications continue, another interest in theoretical studies[57,58].

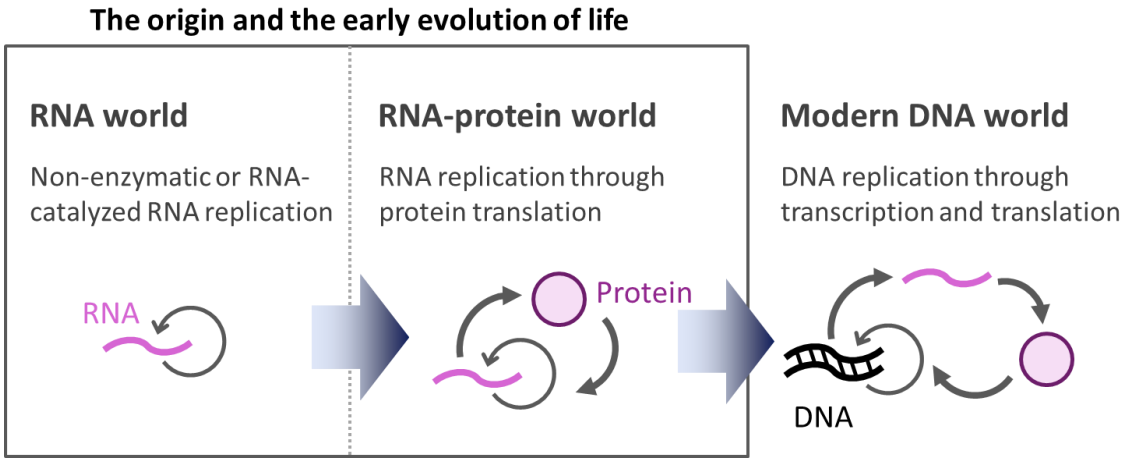


Figure 1-2. The origin and the early evolution of life

It is generally considered that replicating RNA molecules emerged in the origin of life before the development of proteins (RNA world). The RNA replicators may have evolved to gain proteins and start replicating through protein translation (RNA-protein world), and further evolution established DNA and modern genetic systems (modern DNA world).

1-2. Construction of the Dissertation

I summarized the organization of the dissertation in (Fig. 1-3). I used the TcRR system and independently investigated the two major evolutionary phenomena, genome diversification achieved by adapting to various environments (Chapter 2 and 3) and increase in complexity achieved by integrating new genes into the system (Chapter 4).

In the Chapter 2, I focused on adaptation ability of the simple artificial genome replication system. I conducted a long-term replication experiment using the TcRR system in a severe translation condition to answer whether the simple replication system has adaptation ability. I used a clone RNA sequence that evolved in a rich environment as the starting RNA genome. I prepared the condition by reducing ribosome in the translation proteins. I found that the RNA genome in the TcRR system accumulated mutations through the repetitive replications and greatly improved the TcRR activities in the translation-impaired condition, which is an evidence of adaptation. I also showed that this adaptation partly resulted from a simple modification of RNA structures.

In the Chapter 3, I investigated the adaptation ability of the TcRR system more deeply. Specifically, I examined whether the RNA genome adapts to multiple severe conditions and diversifies by accumulating condition-specific mutations. I prepared four different translation-impaired conditions, including initiation-impaired and termination-impaired conditions. I conducted long-term replication experiments in each condition and analyzed evolved RNA sequences. In all the conditions, the TcRR activities were improved with mutations. By comparing sequences evolved in different conditions, I found that the RNA genome accumulated mutations depending on the conditions, which can be considered a process of genome diversification.

In the Chapter 4, I focused on the other major evolutionary phenomenon and investigated the possibility of the evolution of complexity by reconstituting the process. To mimic a plausible process of the increase in complexity, I expanded the TcRR system to one with two cooperative RNA replicators encoding a replication enzyme or a metabolic enzyme. The two RNAs replicated only in the presence of their partners. Next, I performed long-term replication experiments to find conditions allowing the cooperators to sustainably replicate. I found that it can be sustained only in a specific condition, in which RNA concentrations were adjusted in a certain range. Moreover, detailed analysis revealed that the two RNA replicators evolved to improve the cooperative replication activities. Therefore the increased complexity can be sustained even in an artificial genome replication system, while maintaining evolutionary ability. Furthermore, from different perspectives, the newly constructed replication system is a novel platform for cell-free production of metabolic enzymes as well as an experimental model for origin-of-life research.

In the Chapter 5, I summarized the findings and discussed further prospects.

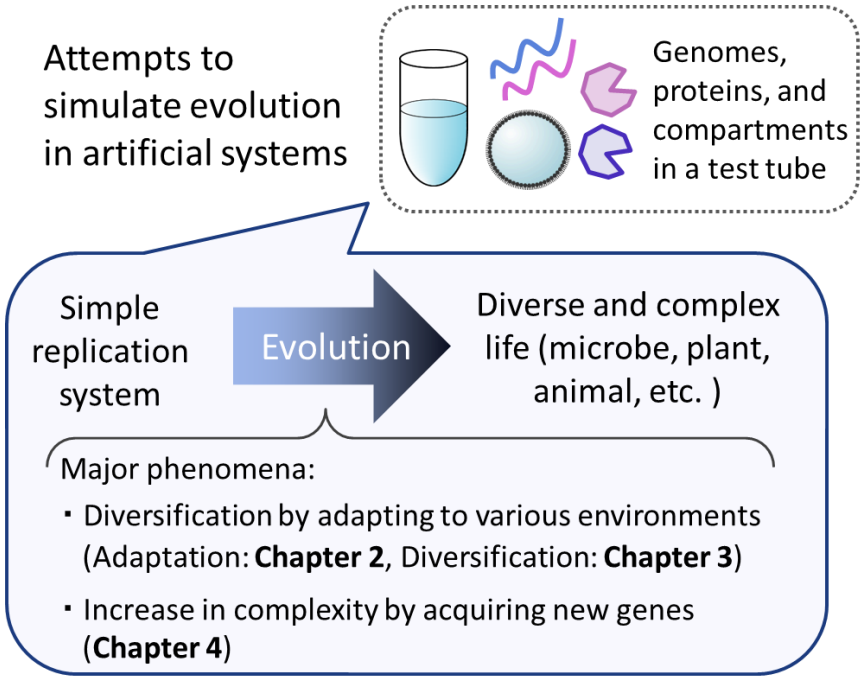


Figure 1-3. Relationships between key three chapters (2, 3, and 4) in the dissertation

The detail was described in the main text.

Chapter 2. Adaptation of an RNA Replication System

2-1. Introduction

Adaptive evolution is one of the key functions of living organisms. Through adaptation, they can change their sequences or gene expressions to survive in response to environments. Adaptation occurs as a result of selection of replicating systems that fit (i.e., replicate faster) in an environment where the systems replicated. Because genome replication produces inheritable genetic variation by mutations, if the mutations are beneficial in the environment, mutants containing the mutations became fitter in the environment, hence selected. Even the simple model bacterium Escherichia coli were experimentally demonstrated to can adapt to various severe conditions such as poor nutrition[59], high temperature[60], and the presence of antibiotics (e.g. inhibiting translation)[61]. What creates this remarkable ability? Reconstituting the ability *in vitro* may provide certain insights into the origin and mechanisms of the ability to adapt. Previous studies showed that artificial RNA replication systems can adapt to severe environments though external supply of RNA replicase or a set of reverse transcriptase, transcriptase, and DNA polymerase enzymes[37,38]. However, these replication mechanism are not natural as living organisms do: replication through translation of self-encoded proteins, which restricts the types of mutations so that the proteins keep functioning. Therefore, whether replication systems coupled with translation reaction can adapt to a severe environment should be investigated independently.

In a previous study, a translation-coupled RNA replication (TcRR) system was constructed[35]. This system consists of only 100 kinds of proteins and RNAs, and the RNA genome is only 2 kb. It is therefore much simpler than any living organisms. Despite the great simplicity, the TcRR system was previously found to evolve at the optimal, ribosome-rich condition[62]. However, with such a limited evolutionary experiment, we cannot conclude that the system has the capacity of adaptation to various environments as living things do. Thus, in this chapter, I investigated whether the artificial RNA replication system adapted to a severe condition: reduced ribosome concentration. Ribosome is known as an essential factor for translation reaction, and therefore reducing ribosome inhibits translation in the TcRR reaction.

Another important feature of the TcRR system is that the RNA genome replication depends on both the translation of replicase from the genome and the replication of the genome by the replicase. This type of replication is seen in some natural self-replicating systems including single-stranded RNA virus. However, how the 2-step replication is regulated during evolution has not been fully understood because of unknown factors involved in the replication. In contrast, all the components of the TcRR system are defined. Therefore by dissecting change in the 2-step replication through adaptation, I could provide detailed insights about how the relationship between the two distinct reactions evolve.

In this chapter, firstly, I performed repetitive replication experiments of the TcRR system under reduced ribosome concentrations and showed that the system gradually adapted to the condition. Next, I characterized evolved RNAs and the mutations, then showed how translation and replication ability were changed. Lastly, I provided a possible answer to how the simple system adapted to the translation-impaired condition by analyzing RNA structures.

2-2. Materials and Methods

2-2-1. Materials

Original RNA

The starting RNA for the experimental evolution was the evolved clone obtained at the round 128 in a previous evolution experiment performed at high ribosome concentration[62]. The sequence was shown in (Appendix A2-1).

The list of primers

110314_1stRTannealF2 (TAAGCGAATGTTGCGAGCACGGCCCATTCTGTGTACCTCAAG)
100306_Minus_RT2 (AAAGCGCTAGCCCGTGCTCTAGC)
100310_mdv-11 (CGTACGGGAGTTCGACCGTG)
pUC-primer+2 (CGTTGTAAAACGACGGCCAG)
pUC-primer-1 (CCGGCTCGTATGTTGTGTGG)
110314_MBTR2 (GATCCACCCGCGGTTTTTC)
2nd TPCR primer (TAAGCGAATGTTGCGAGCAC)
100915_16_IR#12 (TACGAACGGTTATTCGTTGTTG)
100626_seq_beta+4 (GAACGCTCGTCTCTATAGGCCTG)
pYUQb-p_3913b (GGGCGATTGCCTGACGGTAGTG)
100915_17_IF#15 (CGTCGGATCGGTCCTAATC)
111213_MBTF3 (GCTGCCTAAACAGCTGCAAC)
111213_MBTR3 (CGCTCTTGGTCCCTTGTATG)

The sequence of VectorF

The sequence is shown in (Appendix A2-1).

Reconstituted translation system (PURE system)

The composition of the PURE system was listed in (Table 2-1).

Saturation buffer

The composition of the PURE system was listed in (Table 2-2).

Table 2-1 Composition of the PURE system (solA (buffer) + solB (translation proteins))

solA	final
18 AA	0.36 mM
Tyr	0.3 mM
Cys	0.3 mM
Hepes (pH 7.6)	100 mM
Glu-K	70 mM
Spermidine	0.375 mM
$Mg(OAc)_2$	18 mM
CP	18 mM
DTT	6 mM
FD	10 ng/μL
ATP	3.75 mM
GTP	2.5 mM
CTP	1.25 mM
UTP	1.25 mM
tRNA mix	1.56 μg/μL

solB	final (nM)		final (nM)		final (nM)
IF1	25000	MTF	590	ProRS	170
IF2	1000	AlaRS	730	SerRS	80
IF3	4900	ArgRS	30	ThrRS	80
EF-G	1100	AsnRS	420	TrpRS	30
EF-Tu	80000	AspRS	120	TyrRS	150
EF-Ts	3300	CysRS	20	ValRS	20
RF1	50	GlnRS	60	Ribosome	1000
RF2	50	GluRS	230		
RF3	170	GlyRS	90		
RRF	3900	HisRS	90		
CK	250	IleRS	370		
NDK	20	LeuRS	40		
Ppiase	40	LysRS	120		
Tig	20	MetRS	110		
HrpA	10	PheRS	130		

18 AA: amino acids except for tyrosine and cystein

Hepes: 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Glu-K: L-Glutamic acid hydrogen 1-potassium
$Mg(OAc)_2$: magnesium acetate
CP: Phosphocreatine
DTT: dithiothreitol
FD: 5-formyl-5.6.7.8.-tetrahydrofolic acid
IF: initiation factor
EF: elongation factor
RF: release factor
RRF: ribosome recycling factor
CK: creatine kinase
NDK: nucleoside diphosphate kinase
Ppiase: pyrophosphatase
MTF: methionyl-tRNA formyltransferase
ARS: aminoacyl-tRNA synthetase

Table 2-2 Composition of the saturation buffer

	final		final
18 AA	0.36 mM	Spermidine	0.375 mM
Tyr	0.3 mM	Mg(OAc)2	7.25 mM
Cys	0.3 mM	CP	2.5 mM
Hepes(pH 7.6)	100 mM	DTT	36 mM
Glu-K	70 mM	FD	10 ng/ul

2-2-2. RNA Preparation (*In Vitro* Transcription)

The plus-strand of the original RNA was inserted in the pUC19 plasmid with a T7 promoter and a SmaI site at the 5'- or 3'- terminus respectively. The plasmid (1 μg) was mixed with 0.1 vol. 10xT buffer (Takara), 0.01% BSA (Takara), 1.6 U/μl SmaI (Takara) and digested at 37°C for 2 hours, then the products were purified using QIAquick PCR Purification Kit (QIAGEN). Next, the product DNA was mixed with 0.1 vol. 10xT7 RNA polymerase buffer (Takara), 5 mM DTT (Takara), 2 mM ATP (Takara), 2 mM GTP (Takara), 2 mM CTP (Takara), 2 mM UTP (Takara), 1 mM Sp-GTP-α-S (Guanosine- 5'- O- (1- thiotriphosphate), BioLog) 0.8 U/μl RNasin (Promega), 1 U/μl T7 polymerase (Takara), and RNA was synthesized at 37°C for 2-3 hours. The RNA products were purified using RNeasy Mini Kit (Takara). Next, they were further treated with 0.1-0.2 U/μl DNaseI (Takara) in 0.1 vol. 10xDNaseI buffer (Takara) at 37°C for 30 minutes, followed by purification using RNeasy Mini Kit. The product was used as the starting RNA for the evolution experiment. It should be noted that

in a previous study, the average mutation rate in the in vitro transcription process was calculated as less than 2 x 10^{-4}/nt for the RNA genome (2 x 10^3 nt)[35]. Therefore, it is considered that RNA genomes prepared for the starting RNA population through this process contain less than one mutation per genome, much lower than mutation numbers observed in evolution experiments performed in this dissertation.

Evolved RNA clones obtained during the evolution experiment were prepared by the same method.

2-2-3. Translation-coupled RNA Replication (TcRR) in Water-in-oil Emulsion

Oil phase was prepared by mixing mineral oil (95%, SIGMA), Span80 (2%, Wako), Tween80 (3%, Wako). Then saturation buffer (Table 2-2) was added to it and the mixture was incubated at 37°C for 10 minutes after 30 seconds mixing using vortex. The saturated oil phase was collected by centrifuging it (15 krpm, 5 min, 25°C) and collecting the supernatant. To perform TcRR reaction, 10 μl RNA solution: PURE system (Table 2-1), 1 U/μl RNasin, plus-strand RNA (each concentration) and 1 ml saturated oil was filtered through a multipore membrane (20 mm hydrophilic SPG pumping filter, SPG Techno, Japan) 50 times to prepare water-in-oil emulsion that was approximately 2 μm in diameter. The emulsion was incubated at 37°C to start the TcRR reaction. RNA concentrations after the reaction was quantified as described in (2-2-5).

2-2-4. Translation-coupled RNA Replication (TcRR) in Bulk Conditions

TcRR reactions were performed as described in (2-2-3) except that the reaction mixture was directly incubated at 37°C to start the reactions. RNA concentrations after the reaction was quantified as described in (2-2-5).

2-2-5. Quantitative RT-PCR

To measure minus-strand RNA concentrations of after TcRR reactions, incubated emulsion was sampled into 1 mM EDTA (Wako) and 100-fold diluted. If the TcRR reactions were conducted without emulsion, the sample was 10000-fold diluted by 1 mM EDTA. After incubating at 37°C for 5 minutes, the solution was put on ice for 5 minutes. Next, to attach a primer for reverse transcription, 2 μl of the solution was mixed with 3 μl of 10 mM each dNTP Mix (Clonetech) and 1 μM primer 110314_1stRTannealF2, then incubated at 65°C for 5 minutes and placed on ice for 5 minutes. To conduct reverse transcription, 5 μl of 0.2 vol. 5×PrimeScript buffer (Takara), 1 U/μl RNasin (Promega), 10 U/μl PrimeScript Reverse Transcriptase (Takara) were added to the solution and incubated at 50°C for 30 minutes, followed by further incubation at 70°C for 15 minutes. The synthesized cDNA was 5-fold diluted and the 5 μl was mixed with 0.5 vol. SYBR Premix Ex Taq II (Takara) and 0.4 μM primers 110314_MBTR2 and 2nd TPCR primer to prepare 20 μl reaction

mix. Quantitative PCR was conducted with the mixture using Mx3005P (STRATAGENE) or Applied Biosystems StepOne™ / StepOnePlus™ real-time PCRsystem (Life Technology). The concentrations were calculated from a standard curve.

For plus-strand RNAs, reverse transcription and quantitative PCR were conducted using PrimeScript® One Step RT-PCR Kit (Takara) and primers MBTF3 and MBTR3 directly after dilution of the TcRR reaction mix with EDTA. The concentration was measured in the same way as minus-strand RNAs.

2-2-6. Evolution Experiment

The summary of the 1 round of the experiment was shown in (Fig. 2-1). First, the TcRR reaction with 0.1 nM starting RNA was conducted in emulsion at 37°C for 4 hours at reduced ribosome concentrations. Because ribosome is an essential factor for translation reaction, decreasing ribosome concentration inhibits the TcRR reaction. Instead of 1 μM (Table 1), 200 nM ribosome was used for the first 15 rounds and 50 nM for the next 15 rounds except for the round 1 (50 nM) and the round 18 (12.5 nM). After the reaction, the emulsion (300 μl) was centrifuged (15 krpm, 5 min, 25°C) to collect separated water phase. To maximize the recovered amount of the RNAs, ribosome was added up to 1 μM of the total concentration to the emulsion at the centrifugation. The water phase was mixed with 0.4 vol. (120 μl) of diethyl ether and centrifuged (10 krpm, 5 min, 25°C) again. After removing the ether phase, the sample was placed on ice for 5 minutes. Next, to degrade the original RNA templates with Sp-GTP-α-S, 90 μl of 10 mM (round 1-5) or 20 mM (round 6-30) iodine and 1M Tris-HCl (pH 7.4) were added to the sample and incubate it at 37°C for 5 minitues. Then, the RNA was purified using RNeasy Mini Kit (QIAGEN).

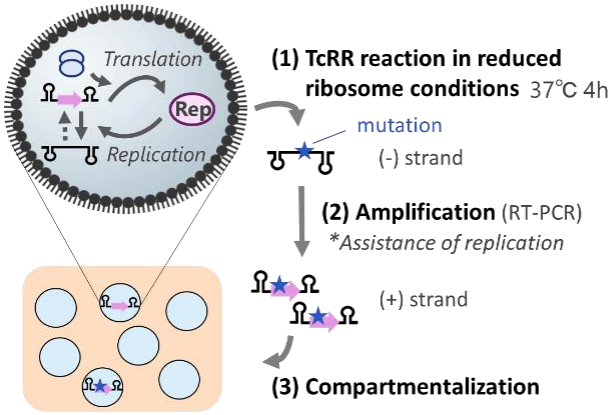


Figure 2-1. Scheme of evolution experiment.

(1) TcRR reactions at reduced ribosome concentrations, (2) collection and amplification of synthesized minus-strand RNAs, (3) compartmentalization of converted plus-strand RNAs were repeated. Further details are in the main text.

An aliquot (0.75 μl) of purified RNA was subjected to reverse transcription as described in (2-2-5) with some exceptions: the primer was 100306_Minus_RT2, the first temperature for the reverse transcription reaction was 54°C. Next, after the 10 μl solution was treated with 0.5 μl of 60 U/μl RNaseH (Takara) at 37°C for 20 minutes, an aliquot (0.5 μl) was mixed with KOD-FX PCR solution (Takara) with the primers 100306_Minus_RT2 and 100310_mdv-11, and the cDNA was amplified by PCR. The DNA products were purified using MinElute PCR Purification Kit (QIAGEN) and subjected to gel extraction (E-Gel® CloneWell 0.8% SYBR Safe™ gel and E-Gel® iBase™ Power System (Life Technology)) to selectively collect around 1860 bp DNA. After purification using MinElute PCR Purification Kit, the DNA was subjected to another PCR reaction with VectorF, primers pUC-primer+2 and pUC-primer-1, and Prime STAR HS PCR solution (Takara). The DNA products were purified using MinElute PCR Purification Kit and subjected to gel extraction by the same method described above to collect 2200 bp DNA. The DNA was converted to RNA for next rounds by *in vitro* transcription as described in (2-2-2), and the next round was started with 0.1 nM of the RNA. These procedures were repeated for 30 rounds.

2-2-7. Preparation of RNA clones

DNA population at the round 11, 15, and 30 obtained by the second gel extraction (2-2-5) were inserted into a vector by infusion method. The 25-50 ng DNAs, mixed with 80 ng VectorF and In-Fusion HD enzyme premix (Clontech) (total 10 μl), were incubated 50°C for 15 minutes, then purified using MinElute PCR Purification Kit. Next, the obtained plasmids containing the target DNAs were transformed into competent cells of *E.coli*. The purified DNA samples (0.5 μl) were added to a 5 μl competent cells (JM109, Takara) and placed on ice for 10 minutes, followed by 45 second incubation at 42°C. The cells were placed on ice for 2 minutes, then mixed with 45μl of SOC mediums (Takara) and incubated at 37°C with shaking (160 rpm). The precultures were inoculated on LB medium plates with 50 μg/ml ampicillin, followed by overnight incubation at 37°C. Next, 8 colonies were randomly picked up from each plate and put into LB liquid mediums, followed by overnight incubation at 37°C with shaking (160 rpm). The plasmids were purified from the cultured cells using QIAprep Spin Miniprep Kit (QIAGEN), then subjected to *in vitro* transcription as described in (2-2-2) to prepare RNA clones.

2-2-8. Sequence Analysis

The sequence analysis of the obtained plasmids was outsourced to Eurofins Genomics (Operon at the time of the experiments) with primers 100915_16_IR#12, 100626_seq_beta+4, pYUQb-p_3913b, 100915_17_IF#15. The obtained waveform data was analyzed using the CLC Main Workbench (CLC bio) software and mutations from the original sequence were determined.

2-2-9. Mathematical Model of the Translation-Coupled Minus RNA Synthesis

The model of minus-RNA synthesis through TcRR reactions was shown in (Fig. 2-2). The model consists of the translation of replicase from plus-strand RNA and minus-strand RNA synthesis from plus-strand RNA by replicase. Translation reaction can be assumed to be a first-order reaction with the rate constant k_{trans} because ribosome concentration was much higher than plus-strand RNA. When replicase concentration and plus-strand RNA concentration are [Rep] and [(+) RNA] respectively, the translated replicase concentration over time (t) is described as

$$\frac{\mathrm{d[Rep]}}{\mathrm{d}t} = k_{\mathrm{trans}}[(+)\ \mathrm{RNA}] \qquad \text{(eq. 1)}$$

By solving the (eq. 1), the following equation is obtained.

$$[\mathrm{Rep}] = k_{\mathrm{trans}}[(+)\ \mathrm{RNA}]t \qquad \text{(eq. 2)}$$

Here, in the experimental condition (the TcRR reactions were conducted in a short time), plus-strand RNA concentration can be assumed to be constant over the reaction time. Similarly, since the plus-strand RNA concentration is much higher than replicase concentration in the same condition, minus-strand RNA concentration can be assumed to be a first-order reaction with the replication rate constant k_{rep}. If minus-strand RNA concentration is [(-) RNA], the [(-) RNA] over time is described as

$$\frac{\mathrm{d[(-)\ RNA]}}{\mathrm{d}t} = k_{\mathrm{rep}}[\mathrm{Rep}] \qquad \text{(eq. 3)}$$

By solving the (eq. 3), the following equation is obtained.

$$[(-)\ \mathrm{RNA}] = \frac{1}{2}k_{\mathrm{rep}}k_{trans}[(+)\ \mathrm{RNA}]t^2 \qquad \text{(eq. 4)}$$

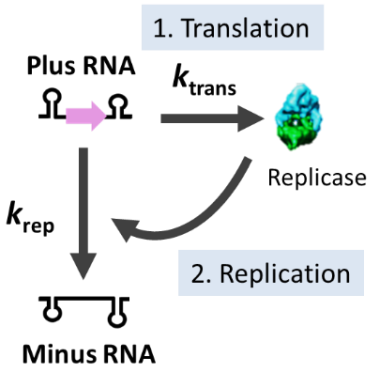


Figure 2-2. Kinetic model of TcRR reaction

Details are in the main text.

2-2-10. Analysis of Translation Activities

100–200 nM plus-strand RNA clones in a modified TcRR reaction mixture (no UTP, 1/10 methionine, a small amount of [^{35}S]-methionine, 200 nM or 1 μM ribosome) were incubated at 37°C for 2 hours. After incubation, synthesized replicase was separated by SDS-polyacrylamide gel electrophoresis (separate gel: 10% acrylamide, 350 mM Tris-HCl (pH8.8), 0.2% SDS, 1 mg APS, 0.001 vol. TEMED; condensation gel: 3% acrylamide, 125 mM Tris-HCl (pH6.8), 0.2% SDS, 0.5 mg APS, 0.0025 vol. TEMED). After fixation in 10% acetic acid and 30% methanol solution, the gel was aspirated to dry and subjected to autoradiography. The film was processed using Typhoon FLA 7000 (GE Healthcare), and the concentration of synthesized replicase was calculated from a standard curve. k_{trans} was also calculated by (eq. 2)

2-2-11. Analysis of Replication Activities

$k_{trans}k_{rep}$ was obtained by a curve fitting of minus-strand RNA synthesis (eq. 4, Fig. XX). Then, the $k_{trans}k_{rep}$ was divided by k_{trans} to calculate k_{rep}.

2-2-12. Replication Activity of the Clones under Uncoupled Conditions

First, 1 nM plus-strand RNA clones in the TcRR reaction mixture (without UTP) were incubated in the condition of 200 nM or 1 μM ribosome to start only translation reactions. Next, to start replication reactions, 1.25 mM UTP and 30 μg/mL streptomycin were added to the reaction mixtures. After 30 minutes (200 nM ribosome condition) or 5 minutes (1 μM ribosome condition) incubation at 37°C, the minus-strand RNA concentration was quantified as descried in (2-2-5).

2-2-13. Analysis of Encoded Replicase Activities

To synthesize replicase of the clones, 100 nM plus-strand RNA clones in the TcRR reaction mixture (without UTP) at 1 μM ribosome concentration were incubated at 37°C for 2 hours. After incubation, an aliquot added to 9 vol. of the solution for replication: [^{32}P]-UTP, 125 mM NTPs, 100 nM s222 RNA (222 nt RNA with replicase recognition sites)[63], 25 μg/mL chloramphenicol, 125 mM Tris-HCl (pH 7.8), 10 mM magnesium chloride, and 0.01% BSA. After 30 minutes incubation, the replicated s222 was separated by polyacrylamide gel electrophoresis (8% acrylamide, 1xTBE, 0.1 vol Glycerol, 0.2%SDS, 0.7 mg APS, 0.0014 vol. TEMED). The gel was fixed in 7% acetic acid and subjected to autoradiography in the same method as (2-2-10). The s222 concentration was calculated from a standard curve.

2-2-14. SHAPE Analysis

The analysis was mainly conducted by Dr. Kimihito Usui. The SHAPE experiment follows two-capillary protocol[64,65] with some modifications. First, to stabilize clone RNA structures, 1 pmol

clone RNAs were incubated in the solA of PURE system (Table 2-1) without glutamate potassium (Glu-K), NTP, and tRNA at 37°C for 2 hours. Then, 0.1 vol. of neat DMSO or 10 mM 1-methyl-7-nitroisatoic anhydride (1M7)[66] dissolved in DMSO was added to the sample and conducted further incubation at 37°C for 70 seconds, followed by RNA purification using RNeasy mini kit. The purified RNA (9 µl) was mixed with 1 pmol umodified RNA. The 2'-O-adducts were detected by primer extension with 5'- fluorescently labeled primer (5'-GTCGAATCTCGGGCTGAATG-3'). All the primers used in this experiment was labeled with VIC (used for the sequencing channel, Applied Biosystems) or NED (used for the (+) and (-) reagent channel, Applied Biosystems). The fluorescently labeled primers (3 µl, 0.3 mM) were added to the (+) and (-) 1M7 reactions and sequencing reactions. The mixture of the primers and RNA templates were incubated at 65°C for 5 minutes, then further incubated at 37°C for 1 minute and placed on ice. For the sequencing reactions, ddGTP (1 µl, 10 mM) was also added. To start primer extension reactions, 6 µl of Superscript III enzyme mix[64] or 1 µl of SuperScript III (200 units, Invitrogen) was added to (+) and (-) reagent reaction or the sequencing reaction mixtures. The solutions were incubated at 45°C for 2 minutes, 52°C for 20 minutes, and 65°C for 5 minutes. After the primer extension, each (+) and (-) reagent reaction mixture was mixed with the sequencing reaction mixtures. Obtained cDNA was purified using Performa DTR Gel Filtration Cartridges (EdgeBio systems) and sequenced using Genetic Analyzer 3130 (Applied Biosystems). The electropherograms of the fluorescence intensity over elution time were analyzed using QuShape[65]. The SHAPE reactivities were normalized by model-free statistics[67,68].

2-3. Results and Discussion

2-3-1. Evolution Experiment under Reduced Ribosome Conditions

Through the Long-term replication experiment (Fig. 2-1) in 200 nM ribosome concentration (1/5 concentration of the optimum), the amount of replication of RNA population had gradually increased approximately 30-fold for 15 rounds (Fig. 2-3A). Next, I further decreased the ribosome concentration to 50 nM to make the condition severer and started repeating the RNA replication experiment again. The RNA replication dropped once due to the reduction of ribosome but had increased approximately 2-fold for another 15 rounds even in the severer condition (Fig. 2-3B). These results suggested that the original RNA evolved through the long-term replication experiments.

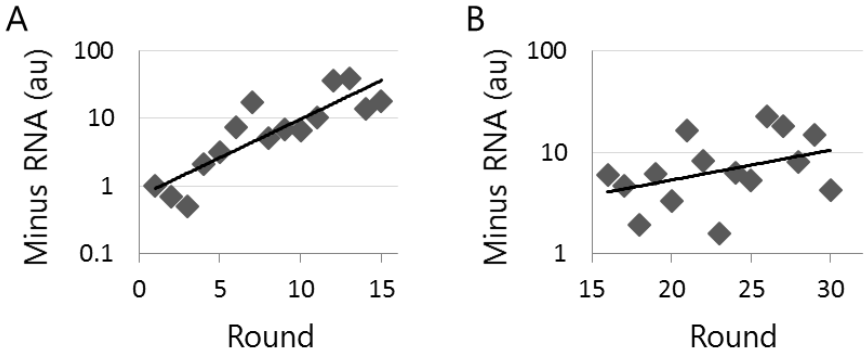


Figure 2-3. Adaptation of the RNA genome to reduced ribosome concentrations.
Minus-strand RNA replications at each round in the evolution experiment were measured by quantitative RT-PCR and normalized to that of the round 1. The experiment was performed at 200 nM (A, except for 50 nM at the round 1) or 1 μM ribosome concentration (B, except for 12.5 nM at the round 18).

2-3-2. Mutation Analysis

I randomly chose 8 plus-RNA clones at the 11, 15, and 30 round of the long-term replication experiment and sequenced them, and mutations from the original sequence were determined. The The average mutation number were constantly increased through the cycle of replications (Fig. 2-4, Total), up to 19 mutations on average, much greater than possible mutation numbers accumulated through the preparation for the starting population as described in (2-2-2), suggesting that mutations were introduced through the evolution experiment. It should be noted that as examined in the previous study[35], some mutations may have been introduced experimental steps apart from the TcRR reaction because the mutation rate of the RNA replicase is about the half of that throughout the steps. I also examined the number of "fixed mutations", which I defined as mutations observed in more

than 50% of the clones. The fixed mutations were also accumulated with the passage of rounds (Fig. 2-4, Fixed). This result indicated that the RNA genome constantly accumulated mutations through replication, and some of them were selected in the condition and spread through the RNA population, which is an evidence of adaptive evolution.

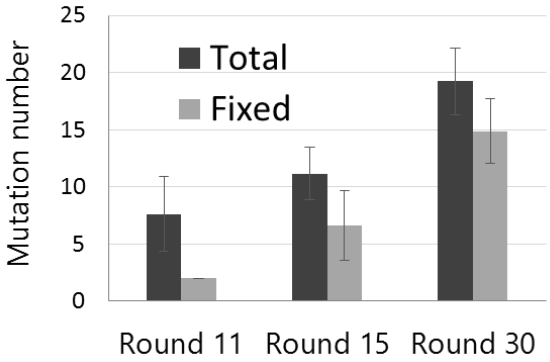


Figure 2-4. Mutation analysis.

8 RNA clones were analyzed for each round. The average of total mutation number (Total) or dominant mutations observed in more than 50% clones (Fixed) were shown. The Error bars indicate standard deviation.

2-3-3. Self-replication Activities of the Evolved RNA Clones

To investigate how the RNA genome adapted to the reduced ribosome condition, I measured self-replication activities of all the obtained clones in the presence of 200 nM ribosome (Fig. 2-5) and chose the clones with highest activities at each round (clones are names R11, R15, and R30).

Next, to compare the original RNA (R0) with the selected evolved RNA clones, I performed TcRR reactions in the same manner as the evolution experiment with 200 nM (Fig. 2-6A) or 50 nM (Fig. 2-6B) ribosome. As a result, I confirmed that minus-strand synthesis through the TcRR reaction was increased over rounds, supporting the idea that the RNA genome adapted to the reduced ribosome concentrations. Furthermore, plus-strand synthesis of the clone R30 through the TcRR reaction was also higher than the original RNA genome, indicating that the ability of complete self-replication (plus-strand → minus-strand → plus-strand…) evolved (Fig. 2-7).

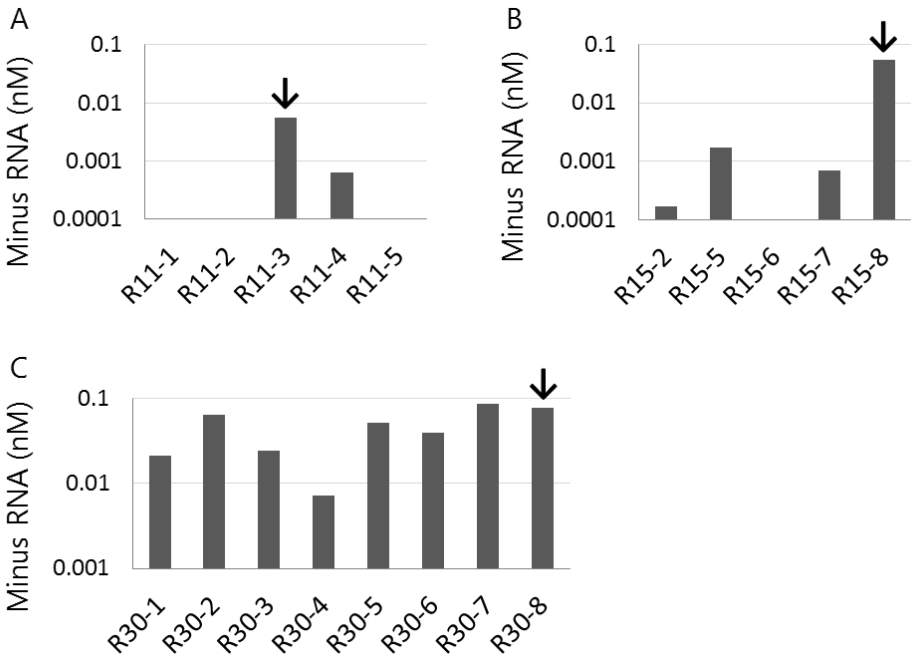


Figure 2-5. TcRR reaction of all of the obtained RNA clones.

Clones that had frame shift mutations were omitted. Analyzed cloned were obtained at the (A) round 11, (B) round 15, or (C) round 30. The minus-strand RNA synthesis through 4-hour TcRR reaction at 200 nM ribosome was measured by quantitative RT-PCR. The clones marked with an arrow were chosen for further biochemical analysis.

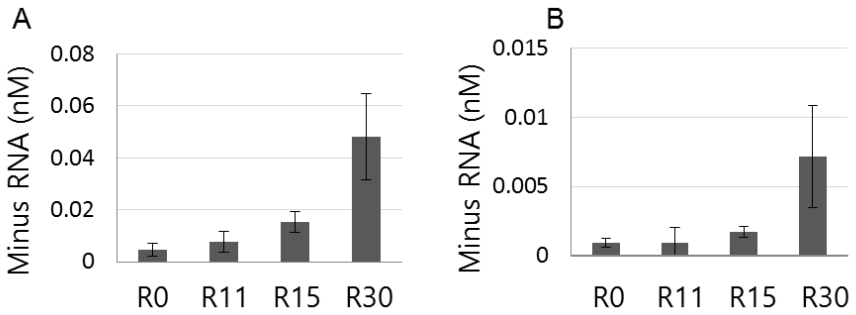


Figure 2-6. Evolved TcRR activities (minus-strand synthesis).

TcRR reactons were conducted with 0.1 nM plus-strand RNA clones at 200 nM (A) or 50 nM (B) ribosome. The synthesized minus-strand RNA concentration was measured by quantitative RT-PCR. The error bars indicate standard deviation.

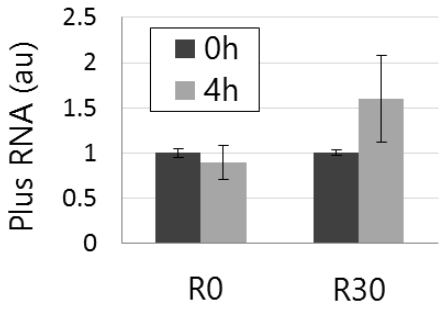


Figure 2-7. Evolved TcRR activities (plus-strand synthesis).

TcRR reactons were conducted with 0.1 nM plus-strand RNA clones at 200 nM ribosome. The synthesized plus-strand RNA concentration was measured by quantitative RT-PCR and normalized to the initial concentration at 0 h. The error bars indicate the standard deviation.

2-3-4. Biochemical Analysis of the Evolved RNA Clones

To investigate how the self-replication activity evolved, I analyzed kinetic parameters in the TcRR reactions. I constructed a mathematical model of the TcRR reaction as described in (Fig. 2-2). In the model, a minus-strand RNA replication depends on two reactions: translation of replicase from a plus-strand RNA and the synthesis of minus-RNA by the replicase. The former was assumed to be a single-order reaction depending on the plus-strand RNA concentration, and the latter was assumed to be a single-order reaction depending on the translated replicase concentration. I defined the rate constant of the reactions as k_{rep} and k_{trans} respectively.

First, I conducted the TcRR reaction of 10 nM plus-strand RNA clones in the presence of 200 nM ribosome and fitted the curve of minus-strand concentrations against reaction time with (eq. 4) (Fig 2-8A). To minimize the effect of plus-strand RNA replications, the reaction was performed in a short time, resulting in the well-fitted curves. The increased value of estimated $k_{trans}k_{rep}$ over rounds (Fig. 2-8B) was consistent with the evolution of TcRR activities (Fig. 2-6), and it finally reached 23-fold increase at the round 30.

Next, I performed translation reaction in the absence of UTP to stop replication and calculated the rate constant k_{trans} by (eq. 2) (Fig. 2-8C). The k_{trans} increased as the rounds proceeded and finally reached 19-fold increase at the round 30. This result indicated that translation ability was improved though the long-term replication experiment. Next, I divided $k_{trans}k_{rep}$ by k_{trans} to obtain k_{rep}. Although k_{rep} was slightly increased at the round 15 or 30 (1.8-fold or 1.2-fold) compared with the original clone (R0), the increases were much lower than the increases in k_{trans} (9-fold or 19-fold respectively),

especially at the round 30 (Fig. 2-8D). Another experiment further confirmed the result by dividing TcRR reaction into translation and replication, and measuring the replication rate constants of the clones in the presence of known concentration of translated replicase (Appendix A2-4). Taken together, the improvement of the self-replication activities and adaptation to low ribosome concentrations were mainly achieved by the increase in translation activities.

I also investigated the activities of the evolved RNAs at the higher ribosome concentrations (1 μM, optimum concentration for the original RNA genome (R0)). I described and discussed results in (Appendix A2-4).

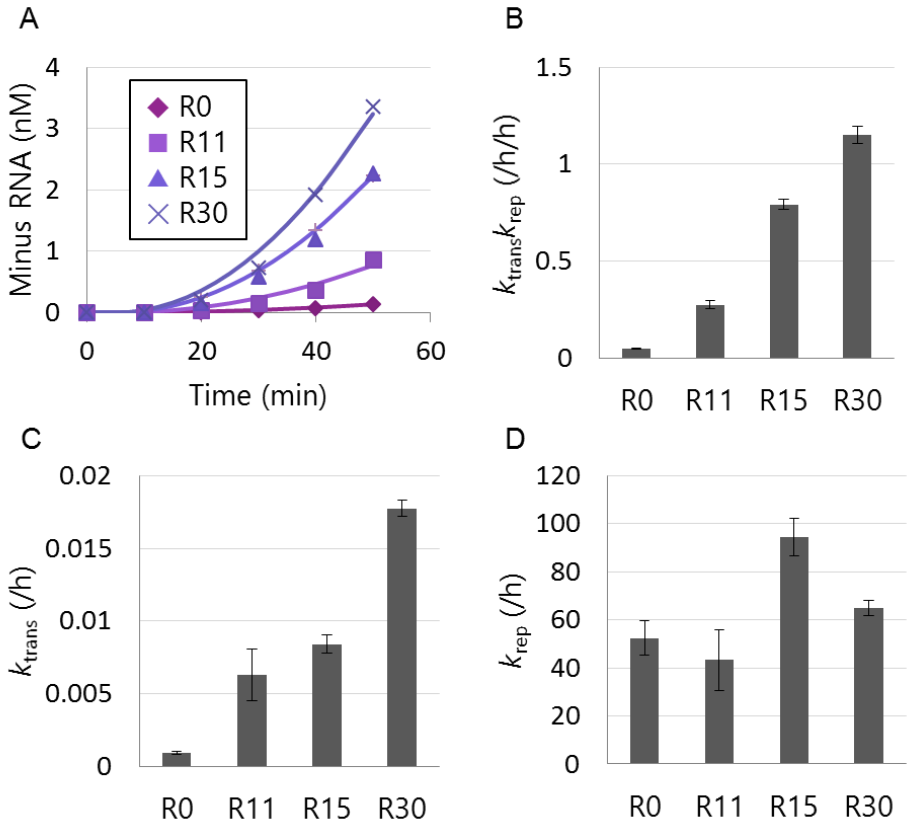


Figure 2-8. Kinetic parameters at reduced ribosome concentration.

(A) Minus-strand RNA synthesis through TcRR reaction of 10 nM plus-strand RNA clones at 200 nM ribosome in bulk condition (without emulsion). The plots were fitted with the (eq. 4). The lag-time for unreacted phase was set 5 minutes. (B) Translation and replication rate constants ($k_{trans}k_{rep}$), estimated by the curve fit of A. (C) Translation rate constants (k_{trans}). 200 nM plus-RNA clones were translated at 200 nM ribosome and the constants were calculated with the (eq. 2). (D) Replication rate constants (k_{rep}) were calculated from obtained $k_{trans}k_{rep}$ and k_{trans}.

2-3-5. Encoded Replicase Activities of the evolved Clones

The replication rate constant consists of encoded replicase activities and RNA's template activities to be replicated by the replicase. Although k_{rep} were not significantly changed, one of these parameters could have changed. I therefore measured the replicase activities for each clone RNA. Briefly, after plus-strand RNA clones were translated, an aliquot of the reaction mixture was transferred to a solution containing [^{32}P]-UTP, NTP, and s222 RNA, which have the replicase recognition site and has high replication activity [63], then I separated newly synthesized s222 by polyacrylamide gel electrophoresis and measured the radioisotope incorporation. This measurement value reflects replicase activities and translation activities of each clone. I therefore divided the value by translation activities obtained in the same condition and calculated the replicase activities. As a result, replicase activities were not significantly changed throughout the evolution (Fig. 2-9). Because unchanged replicase rate constants (Fig. 2-8D) consist of change in encoded replicase activities and RNA's template activities, this result suggested that the template activities were not also significantly changed by evolution.

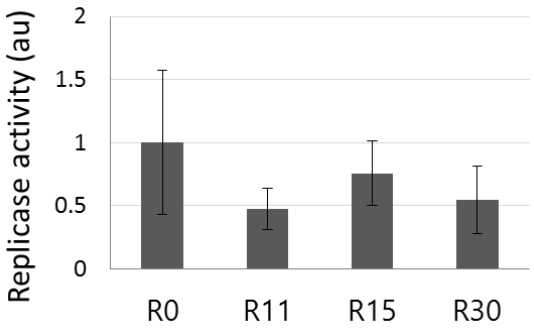


Figure 2-8. Encoded replicase activities.

After plus-strand RNA clones were translated in the absence of UTP for 2 hours, s222 and UTP were added to the synthesized replicase to replicate s222. The replicase activities were obtained by normalizing s222 amplification by the amount of translated replicase. The error bars indicate standard deviation.

2-3-6. Analysis of the RNA Structure of the Evolved RNA Clones

To investigate how the translation activity was evolved, I focused on RNA structure of the clones. Relaxed RNA structures around ribosome binding sites, such as Shine-Dalgarno (SD) sequences and initiation codons, have been known to enhance translation efficiency by recruiting more ribosome to the RNAs[69,70,71]. The clone RNA's structures around the SD sequences (Fig. 2-9) were analyzed by the SHAPE (selective 2'-hydroxyl acylation analyzed by primer extension) method, which detects

RNA sequences that do not form a base pair by selectively attaching a reagent to the sites. (Fig. 2-10A) showed SHAPE reactivity of the clones around the SD sequences (216GGAG219), representing the frequency of unpaired nucleotides in the analyzed region. All the evolved clones (R11, R15, R30) exhibited higher SHAPE reactivities at the SD sequences than the original RNA (R0), and the average reactivities of the R30 clone was 7-fold higher than the original (Fig. 2-10B). These results demonstrated that evolved clones' SD sequence form relaxed structures. The relaxed structures were confirmed by secondary structure predictions using the Vienna RNA software (Fig. 2-11). I also calculated the probability of ribosome binding to the RNAs around SD regions by the RBS Designer software [72], and I found that the probability was increased in a similar manner to the increase in the SHAPE reactivities (Fig. 2-12). These results can partly explain how the translation activity was improved: by recruiting more ribosome even at reduced-ribosome concentrations.

Despite clear increase in translation activities from R11 to R30 (Fig. 2-8B), the SHAPE reactivity was not significantly increased during the rounds (Fig. 2-10B), implying that translation was improved by another mechanism. There are at least three possibilities: improved codon usage to accelerate translation [73], improved elongation speed during translation by reducing internal large or GC-rich stem structures [717475], and decrease in internal SD-like structures which could stop ribosome and impair translation [76]. These mechanisms remain to be seen in future studies.

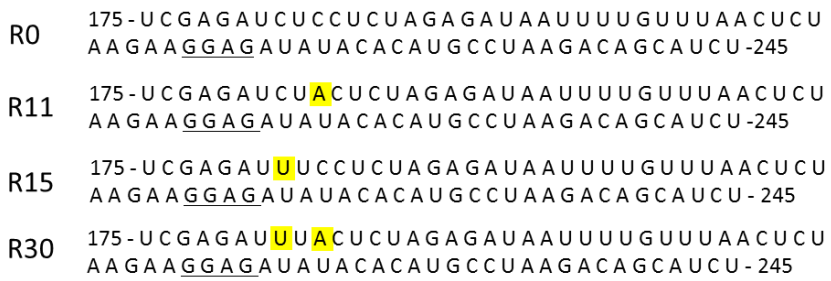


Figure 2-9. Sequences around the SD region.

Mutation sites are marked in yellow. SD sequence are underlined.

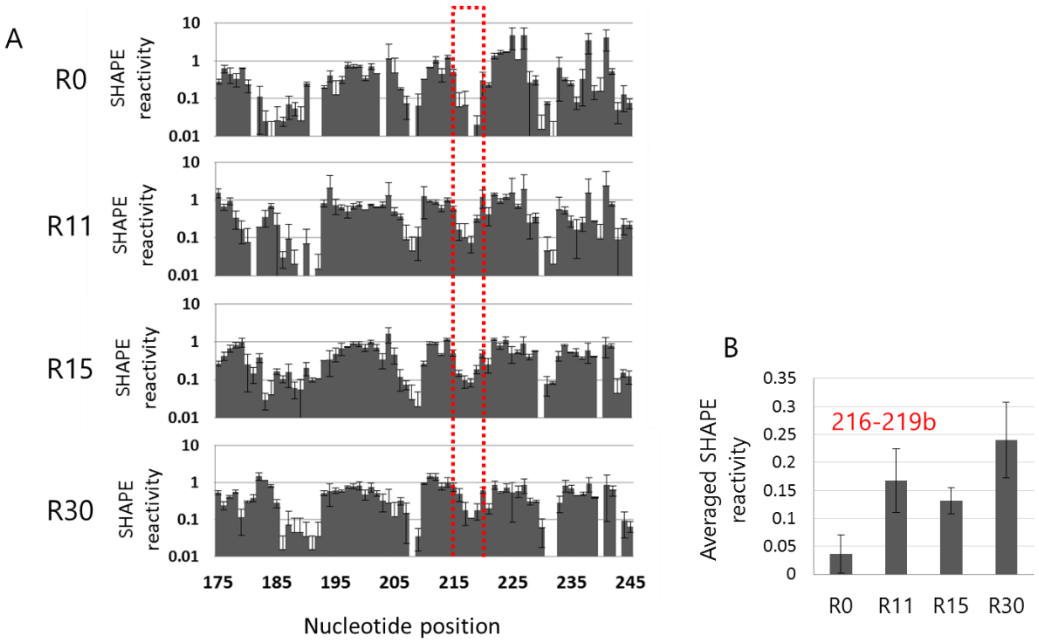


Figure 2-10. SHAPE reactivities of the clones around SD sequences.

(A) Raw SHAPE reactivities around the SD sequence of the RNA clones. The higher reactivities indicate more relaxed structures; the lower reactivities indicates more rigid structures. The SD sequence is shown within the red dotted square (216−219). (B) Average SHAPE reactivities at the SD sequence of the clones. The error bars indicate the standard deviation.

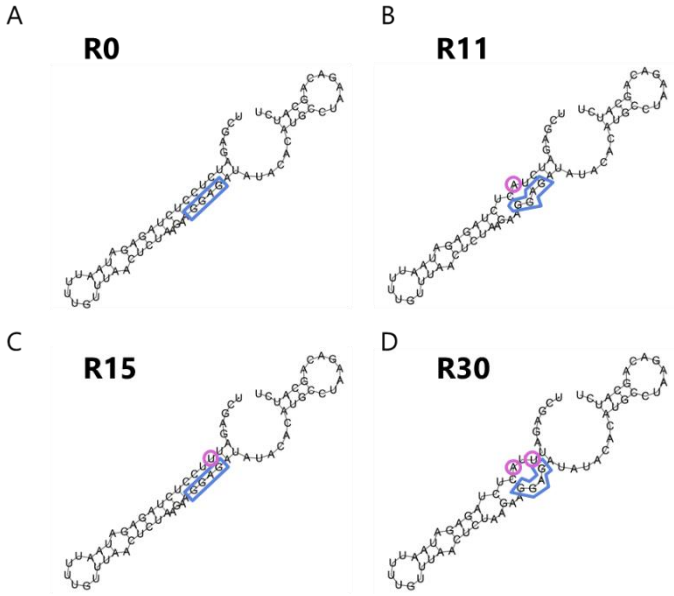


Figure 2-11. Predicted RNA structures around SD sequences.

The structures were predicted using the Vienna RNA (centroid structure) software. The mutation sites are colored in pink. The SD sequences are colored in blue.

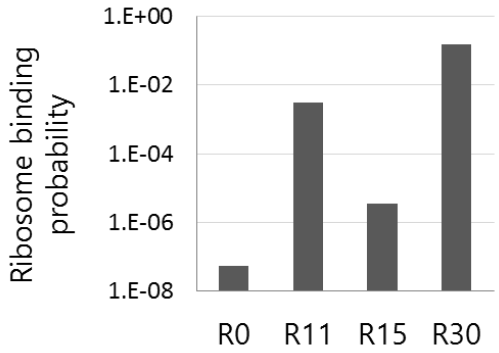


Figure 2-12. Ribosome-binding probability of the clones

The probabilities for 175−245 nucleotide sites of the clones were calculated using RBS Designer.

2-4. Chapter Summary

In this chapter, I examined whether the TcRR system adapted to a severe translation condition. Through the repetitive replication in reduced ribosome conditions for 30 rounds, the RNA genome gradually increased the TcRR activities. The mutation analysis revealed that the RNA genome also gradually fixed mutations as particular mutations, generated by replication errors, were selected in the condition. These results suggested that the RNA genome adapted to the reduced ribosome concentration.

Next, I deeply investigated biochemical properties of some representative evolved RNA clones at the round 11, 15, and 30. The kinetic analysis of the TcRR reaction revealed that their increase in the TcRR activities mainly resulted from the increase in translation activities. Extensive investigation of their RNA structures by SHAPE analysis, structure prediction, and calculation of ribosome-binding probability suggested that the relaxation of RNA structures around the SD sequence may have accelerated ribosome recruitment at the low ribosome concentrations, which may consequently enhanced translation. These results indicated that despite the lack of complex biological functions or components, the simple artificial RNA replication system has a certain ability to adapt to severe conditions, and the adaptation can be possibly achieved by a simple mechanism, at least in part, such as partial RNA structure modification. In the TcRR system, the RNA genome replicates through the translation of self-encoded genetic information as living things do for their replication. Although previous studies showed that artificial RNA replication systems can adapt to severe environments though the external supply of RNA replicase or a set of enzymes related to replication[37,38], whether such adaptation ability can be reconstituted in an artificial RNA replication under restrictions that a self-encoded enzyme must be translated. Therefore, my study provided the first experimental evidence that an artificial translation-coupled RNA genome replication system can adapt to a severe environment. Furthermore, the RNA genome in the TcRR system encodes only the gene of RNA replicase, indicating that the ability to adapt to a severe environment can be achieved with much smaller number of genes than living organisms. Overall, the adaptation ability of living organisms were partly reconstituted *in vitro* in this study.

Another important result from this study is that the evolved RNA genome showed less requirement of ribosome for the TcRR reaction. The decrease in ribosome directly leads to minimizing the construction cost because the complex ribosome structure and the high density necessary for the reaction make the ribosome purification take huge time and money. Moreover, less complex systems with miminum ribosome could be used as a model of the simplest life-form. Further evolution experiments for the reduction of ribosome and other translation proteins would push the TcRR system toward such a minimal replication system and allow us to use the artificial system for various purposes, from applications to basic research.

Chapter 3. Diversification of the RNA Replication System

3-1. Introduction

In the Chapter 2, I demonstrated that the simple TcRR system can adapt to a severe condition, reduced ribosome concentrations. However, whether the system can adapt to diverse severe conditions remains unknown. *Translation may constrain possible mutations so that encoded information can be expressed in a functional form. Furthermore, the single gene in the TcRR system may also be limitation of possible adaptation compared with living organisms that have more than thousands of genes. It should be noted that* previously constructed *in vitro* genome replication systems also showed a certain evolutionary ability, but the experiments were performed under limited conditions *and without translation of genetic infromation*[25,26,37,42]. Therefore our knowledge of *the TcRR system's* adaptation was limited to specific conditions, and hence to what extent the adaptation ability was reconstituted has not been understood *and has to be tested in more diverse conditions*.

Living organisms can adapt to diverse environments and eventually differentiate into distinct species. To understand minimal components to create such a remarkable adaptation ability of life, we must know whether a reconstituted replication system can independently adapt to various conditions, and whether an identical RNA genome treated in different conditions diversify into different sequences by accumulating condition-specific mutations. To answer this question, in this chapter, I conducted experimental evolution of the TcRR system under multiple translation-impaired conditions independently, with an identical starting RNA genome.

Firstly, I newly prepared three translation initiation-impaired conditions by omitting initiation-related proteins: (A) no initiation factors 1 (IF1) and 3 (IF3), (B) no methionyl-tRNA formyltransferase (MTF), IF1, and IF3, (C) no initiation factor 2 (IF2), and one termination-impaired condition by omitting termination-related protiens: (D) no release factors 1 (RF1), 2 (RF2), 3 (RF3), and ribosome recycling factor (RRF). Next, I conducted repetitive replication experiment using the TcRR system under these conditions to see whether the RNA genome adapted to each condition. I compared the result with my previous evolution experiment performed in reduced ribosome concentrations (Chapter 2), where the entire process of translation was inhibited. Lastly, I analyzed accumulated mutations in each condition and described high adaptation ability of the simple artificial system by revealing that the adaptations occurred in a condition-dependent manner, leading to genome diversification.

3-2. Materials and Methods

3-2-1. Evolution Experiment

The general method of the experiment was same (2-2-6). The conditions for TcRR reactions were different from the previous experiments. Four different translation-impaired conditions were prepared; from the PURE system (Table 2-1), (A) IF1 and IF3, (B) MTF, IF1, and IF3, (C) initiation factor 2 (IF2), and (D) RF1, RF2, RF3, and RRF were omitted. The conditions A, B, and C were translation initiation-impaired conditions; the condition D was a translation termination-impaired condition. I perform the repetitive replication experiment (Fig. 2-1) with the same starting RNA as (2-2-6) but in newly prepared conditions (instead of reduced-ribosome conditions) for about 30 rounds in each condition. Only the exception was that Sp-GTP-α-S was not used for in vitro transcription at 1–12 rounds experiments performed in the condition D. Synthesized minus-strand RNA concentration during TcRR reactions at each round was quantified as described in (2-2-5).

3-2-2. Sequence Analysis

8 RNA genomes at the final rounds under each condition were randomly cloned as described in (2-2-7). The sequences were analyzed as described in (2-2-8).

3-2-3. Phylogenetic analysis

The phylogenetic tree was constructed with the original RNA genome, all the 32 clones obtained in the translation initiation- or termination-impaired conditions, and 8 clones obtained at the final round of the previous evolution experiment under reduced ribosome conditions (Chapter 2). The software Molecular Evolutionary Genetics Analysis (MEGA) ver. 5.2[77] was used for sequence alignment and tree construction. The sequences were aligned in ClustalW[78].

3-3. Results and Discussion

3-3-1. Evolution Experiment under Multiple Translation-impaired Conditions

I performed long-term replication experiments under translation initiation- or termination-impaired conditions. For comparison, I included the result of the previous evolution experiment performed under low ribosome concentrations as the condition "E" at following results and discussions.

In all the conditions, replication (minus-strand concentration) of RNA population had gradually increased (Fig. 3-1), but the timing of distinct increase differed for the different conditions. In the conditions A, B, and D, the minus-strand concentrations were increased after approximately 20 rounds. On the other hand, in the condition C and E, the concentration started increasing soon after initiating the cycles of the experiments. These results indicated the RNA genome adapted to the conditions differently.

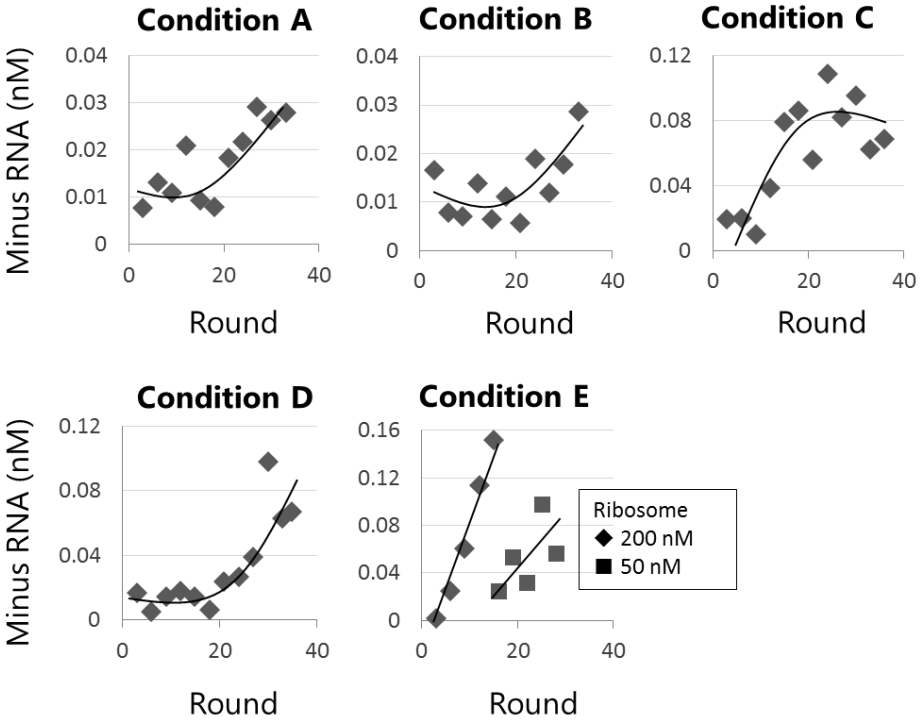


Figure 3-1. Average minus-strand RNA concentrations during the evolution experiments.
Minus-strand RNA concentrations at the TcRR reactions of the evolution experiment under each translation-impaired condition were measured by quantitative RT-PCR. The average concentration for every three round is plotted. The data of the condition E was obtained from (Chapter 2).

3-3-2. Mutation Analysis

Next, I obtained 8 RNA clones at the final round of each condition and analyzed their sequences. The number of dominant mutations (defined as common mutations observed in more than 4 of 8 RNA clones in each condition) were shown in (Fig. 3-2). These mutation numbers were smaller than expected from the mutation rate estimated in a previous research[62], indicating that some mutations were negatively selected through the evolution experiment. On the other hand, total mutation numbers (Appendix, A3-1, 2, 3, 4) are much greater than possible mutation numbers accumulated through the preparation for the starting population as described in (2-2-2), suggesting that mutations were introduced through the evolution experiment. The number of synonymous mutations, non-synonymous mutations, and mutations in untranslated regions varied by the conditions.

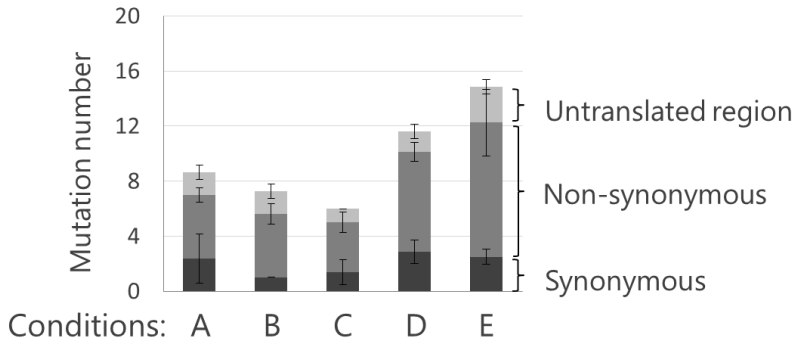


Figure 3-2. Average mutation numbers.

The average number of dominant mutations observed in 8 RNA clones for each condition. Each color of the bars represent the number of mutations in the untranslated regions, non-synonymous mutations in coding regions, and synonymous mutations in coding regions. The error bars show standard deviation.

The sequence analysis revealed that there were some dominant mutations (observed in more than 50% of the clones) in all the conditions. This result can be explained by adaptive evolution, selection of particular mutations, generated by replication errors through repetitive replication. The dominant mutations for different conditions were listed in (Fig. 3-3). Some mutations were specific to the initiation impaired conditions (A, B, and C) and not observed in the termination-impaired condition D (C184A A1603G). To the contrary, other mutations were specific to the termination-impaired condition (e.g., C721T, A825G, A1055G, C1612T, A1729G, and C1978A). This condition-dependent mutation pattern suggests that the RNA genome adapted depending on the

conditions.

Furthermore, the evolved RNA clones in the ribosome-reduced condition E had most of the dominant mutations present in other initiation- or termination-impaired conditions (Fig. 3-3). This may be because ribosome is related to the entire translation process, and its reduction therefore inhibits translation initiation and termination steps. This result is also suggests that the RNA genome adapted in dependence on impaired-translation processes.

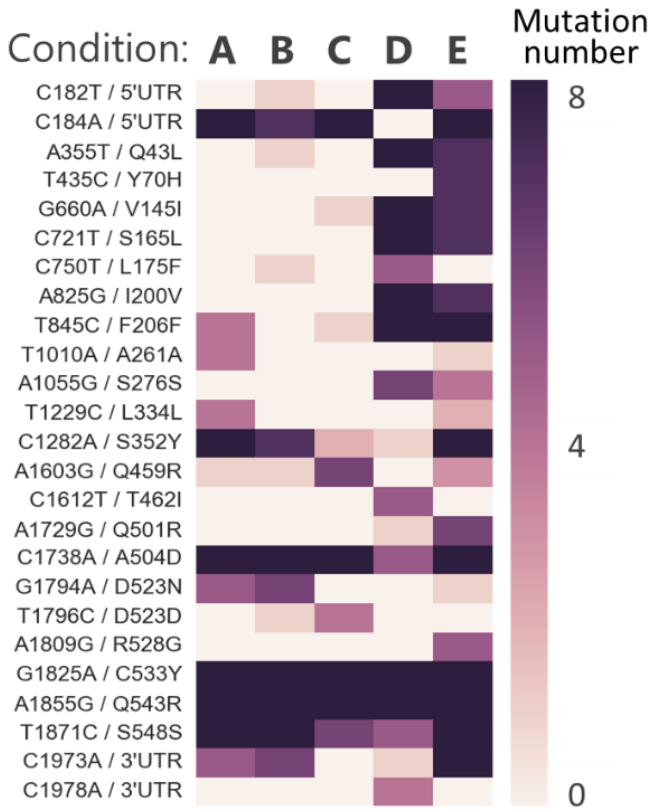


Figure 3-3. Dominant mutation list.

The frequency of mutations dominant at least in one of the five conditions were shown as a heat map.

3-3-3. Phylogenetic Tree Analysis

To illustrate the evolutionary distance of the evolved RNA clones, I constructed a phylogenetic tree with all the obtained clones and the original RNA clone (Fig. 3-4). All the evolved clones in the termination-impaired condition D formed a distinct branch, whereas evolved clones in the initiation-impaired conditions A, B, and C are more closely related and formed mixed branches. The bootstrap value at the center branch connecting clones from initiation-impaired,

termination-impaired, and reduced ribosome conditions (black arrow in Fig. 3-4) was 86% from 1000 replications. The high value ensures the accuracy of the constructed tree and therefore supports the idea that the RNA genome adapted in response to the conditions.

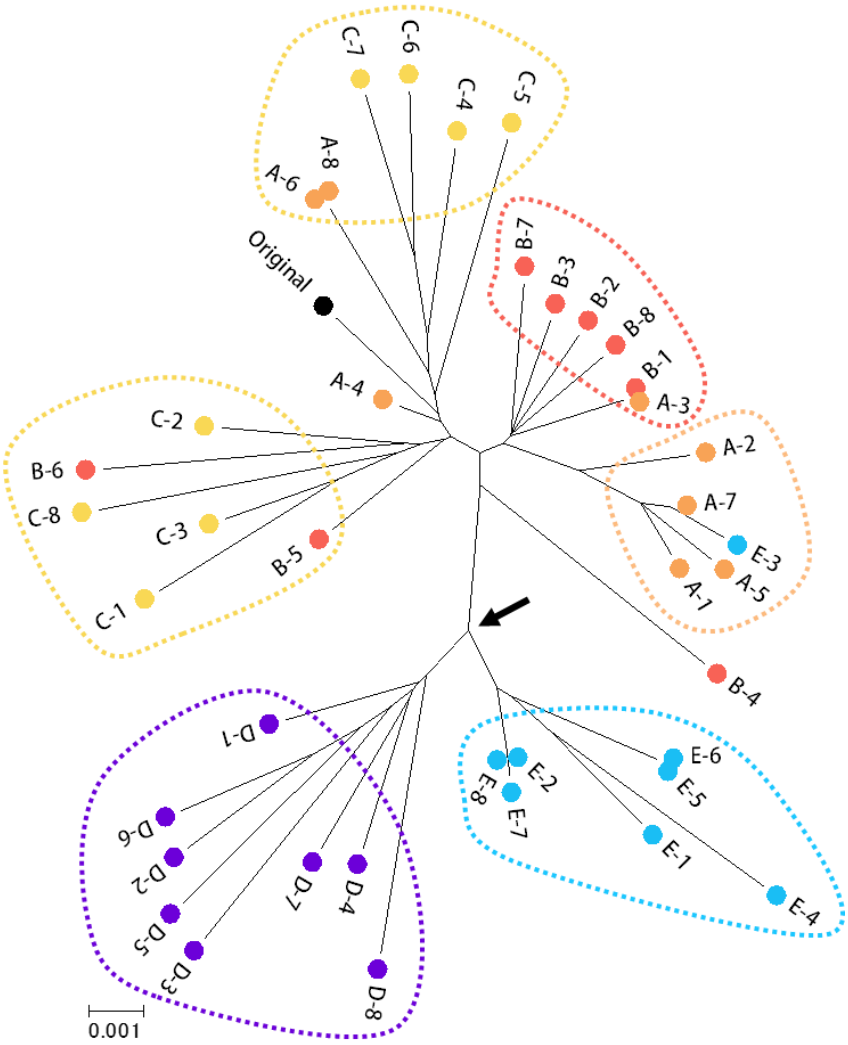


Figure 3-4. Phylogenetic tree of the evolved RNA clones.

The phylogenetic tree was constructed with all the obtained RNA clones and the original clone before evolution. Filled circle represent the sequence of the original RNA clone (black) or each RNA evolved in the conditions A (orange), B (red), C (yellow), D (purple), and E (light blue). The name of the clones was labeled as "condition"–"clone number". All the sequences are shown in (Appendix A2-1, A3-1, A3-2, A3-3, and A3-4). Close sequences were encircled by dotted lines in the color of dominant clones. The black arrow shows the branch of which the bootstrap value was calculated.

There was one identical RNA genome sequence obtained in different conditions (A-3 and B-1). This possibility can be explained by similarity between the condition A and the condition B. Both the conditions were initiation-impaired conditions, where IF1 and IF3 were omitted, and only the difference was the presence of MTF. In fact, the change in replication efficiency through rounds (Fig. 3-1), mutation numbers (Fig. 3-2), and mutation patterns (Fig. 3-3) were similar in the conditions A and B. The phylogenetic tree analysis (Fig. 3-4) also showed that many clones obtained in the two conditions had close sequences. Although A-3 and B-1 genomes contain 12 mutations against the original sequence, 9 in the 12 mutations were dominant mutations in the initiation-impaired conditions (A and B) or the reduced ribosome condition E (entire translation, including initiation, was impaired), and therefore the 9 mutation accumulation was not mysterious. The other 3 mutations were observed only in the clones A-3 and B-1, but it is possible if the three mutations epistatically improved the TcRR activities in the conditions, and hence rare and simultaneous enrichment on the same genome. Therefore, similar selection pressure in the two conditions may have caused enrichment of the same genome sequences in different conditions.

The above evolution experiments in each condition were conducted only once, and therefore I cannot deny the possibility that the results were accidentally obtained. However, the five independent experiments showed a common phenomenon that what mutations accumulated depended on the conditions where the RNA genome was exposed. Close mutation patterns were observed in the three independent experiments performed in the initiation-impaired condition, and a distinct pattern was observed in another experiment performed in the different, termination-impaired condition. Furthermore, in another experiment performed in the reduced ribosome condition, where entire translation was inhibited, detected mutations included most of the mutations observed in the initiation- or termination-impaired conditions probably because both initiation and termination were impaired in this condition. Therefore, although the detail of the mutation patterns may change in reproductive experiments, it is considered that the fact that mutation patterns depended on the conditions should be reproduced.

3-3-4. Mutation Sites and the Potential Effect

One possible effect of mutations is to enhance translation by changing RNA structures[70,79,80]. Under all the initiation-impaired conditions and the reduced ribosome condition, the mutation C184A was dominant. This mutation was known to relax secondary structures around the SD sequences and perhaps enhance translation by increasing ribosome recruitment as described in (Chapter2). Thus, the C184A mutation may have compensated for the impaired initiation.

To my knowledge, single nucleotides immediately following stop codons are only the site

known to affect termination efficiency[81]. Although such mutations were not accumulated in the termination-impaired condition, the TcRR activities in the condition were clearly improved (Fig. 3-1, Condition D). This result suggests that there may be other mechanisms to compensate for impaired termination. Further studies may reveal a new non-canonical termination mechanism independent from release factors (well-known termination-related proteins).

Furthermore, evolved clones in all the tested conditions accumulated mutations across the whole genome (Appendix A2-1, A3-1, A3-2, A3-3, and A3-4), indicating that sequences or structures that affect translation initiation or termination may not be located only the regions close to translation initiation or termination regions, and change in distant sites may have given rise to some improvement. In fact, such mechanisms to regulate translation was known in some virus[82]. Future studies of the RNAs obtained in this study may reveal a new mechanism of long-distance regulation of translation, and could improve the prediction of translation efficiency, which has been developed on the basis of only translation initiation[83,84].

3-4. Chapter Summary

In this chapter, I investigated whether and how the TcRR system adapted to multiple translation-impaired conditions: initiation-impaired conditions (A, B, C) and termination-impaired condition (D) prepared by omitting initiation factors or release factors. I also included the result from (Chapter 2): adaptation to the reduced ribosome condition (E), the entire process of translation was inhibited). In all the conditions, the TcRR activities gradually increased, but the timing of the increase was different. In some conditions, the clear increase was observed in an early phase of the long-term replication experiments, whereas a late phase in the other conditions. This result indicated that the RNA genome adapted to the conditions differently.

Mutation analysis revealed that different mutations became dominant in different conditions by selection from random mutations introduced by replication errors. Mutations in the three distinct initiation-impaired conditions were relatively similar and clearly different from mutations detected in the termination-impaired condition. Furthermore, most of those mutations were common in mutations accumulated in the reduced ribosome condition, where the entire translation process, including initiation and termination, was apparently inhibited. These condition-dependent pattern of mutation accumulation suggested that the RNA genome adapted to the multiple conditions differently, in response to impaired-translation steps.

There was no tendency of mutation positions among the conditions. Mutations specific to the initiation-impaired conditions were not located only at the initiation site but distributed across the genome. The same is true in mutations specific to the termination-impaired condition. This result indicates that even in the case of the short RNA genome (2 kb), RNA structure-based regulation of the translation initiation and termination may not be restricted to the sites where the reactions occur. Thus, the possibility of translation improvement in primitive replication systems may be much greater than we expected, which could be a source of robustness as seen in the adaptations to multiple distinct impaired conditions. Taken together, I demonstrated that a minimal genome replication system with only one gene has a certain adaptation ability and can potentially evolve toward distinct species.

Another important result from this study is that the TcRR system evolved to show higher TcRR activities without initiation factors or release factors, known as essential proteins in contemporary translation systems. Such a compensatory mechanism may have existed in ancient life using primitive translation system. Further evolution with deficient translation systems may allow the replication system to work with less translation proteins, which may be reverting the evolutionary process of the translation system and push the limits of our understanding about the history of translation systems. This line of study will also contribute to reducing the cost for the construction of the TcRR system by omitting required components and increase the possibility of application.

Chapter 4. Expansion of the RNA Replication System and Sustainable Evolution

4-1. Introduction

Complexity is a remarkable evolutionary consequence of living organisms. They have tremendous number of genes and complex gene networks. It has been believed that the evolution of the complexity was driven by the formation of cooperation between replicating species, from simple replicating molecules to replicating cells[32,34,44]. It is considered that in one of the earliest transitions, an RNA self-replicator acquired a new function through establishing cooperative relationship with another RNA replicator that has distinct genetic information[32,44,45,46], which can emerge from the other replicator by mutations or independently appear through evolution at different places. Previous studies mainly focused on whether the coexistence of cooperative RNA replicators is sustained from the theoretical perspective[44,45,46,47,48,49,50,51,52,53,54].

Theoretical studies found that intermolecular cooperation is difficult to sustain through evolution because natural selection usually favors selfish replicators that replicate only itself without helping other's replication. Therefore, a major concern in the increase in complexity is the appearance of selfish or parasitic RNAs, less or non-cooperative replicators, generated with mutations that increase their replications by exploiting other cooperators[32,44,85]. Theoretical research also showed that compartmentalizing the cooperative RNA replicators could stabilize their replications by excluding parasites but only in limited conditions[48,49,51] because in primitive replication, in addition to continuous parasite generation, mutation load and inaccurate distribution through division also prevent the growth of cooperators. Despite a growing body of theoretical literatures on this topic, we still lack empirical evidence for whether the RNA cooperators in compartments could be eventually sustained for a long-term replication. To fully understand this, we should construct an empirical model of evolvable cooperative RNA replicators and prove that they can be sustained by overcoming parasite existence.

Several minimal cooperative replicators were constructed at the molecular level, based on oligonucleotides[86], peptide[87], ssDNA[43], and RNA[23,24]. However, these systems were uncoupled with compartments, and their replication mechanisms, different from contemporary genome replication, did not generate parasites. Furthermore, polymers in their systems did not encode distinct genetic information that carry different functions. On the other hand, the TcRR system I have used in the previous chapters can work in a compartment, and the genome (Rep-RNA) replication spontaneously produces parasitic RNAs by replication errors[55,56]. The parasites had replicase recognition sites yet lacked the replicase coding regions, and they inhibited the Rep-RNA's replication. Therefore expanding this system to one containing two cooperative RNA replicators with different genes is important to create a new level of life-like systems and prove that the increase in complexity is evolutionary achievable.

Another important question about the evolution of complexity is how an RNA self-replicator acquired metabolic function. Metabolism is one of the most remarkable characteristics of living organisms, and it is believed that metabolic function was coupled with replication and boundary functions at the very beginning of life, making a chemical system living[88,89]. One possible pathway of the integration is that a compartmentalized RNA replicator that encodes replicase enzyme acquired another replicator that has metabolic enzyme. On the basis of this idea, I expanded the TcRR system by adding another RNA replicator encoding a metabolic enzyme so that two distinct RNA replicators played replication or metabolic role respectively, through which they interdependently replicate.

Furthermore, evolutionary ability should not be limited as the complexity increased; the cooperators should not only be sustained but also self-improved by evolution, which is what living organisms do. I therefore also focused on the adaptive evolution of the cooperative system, especially whether the self-replication activities of the system can be improved.

In this chapter, firstly, I attempted to expand the TcRR system so that two distinct RNA replicators cooperatively replicate through the translation of replication and metabolic enzymes in a cell-like compartment. For that purpose, I combined the TcRR system with another RNA replicator (NDK-RNA), which supplies CTP for polymerization through the translation of encoded nucleotide diphosphate kinase (NDK), an essential metabolic enzyme in living organisms. Both Rep-RNA and NDK-RNA were replicated by replicase translated from Rep-RNA with the synthesized CTP. Through the replication, mutations were introduced into both the RNAs by replication errors, and they are expected to be selected if beneficial in a condition. Secondary, to discover conditions in which the cooperative RNA replicators can be sustained by overcoming spontaneous parasite generation, I performed long-term replication experiments via random fusion-division of the compartments, which is a plausible replication mechanism of primitive compartments[6,90]. Lastly, I investigated how the system changed through evolution (e.g., improvement of the two RNAs' cooperative replication) to know whether the complex system kept evolutionary ability.

4-2. Materials and Methods

4-2-1. Materials

Original Rep-RNA

The R11 RNA clone obtained by the experimental evolution in the reduced ribosome conditions (Chapter 2) was used.

Original NDK-RNA

The sequence of the original NDK-RNA is a translation-improved *E.coli* NDK, attached with the replicase recognition sites. The RNA was given by Dr. Ichihashi. The sequence was shown in (Appendix A4-1). Mutations for the improvement of translation were listed in (Appendix A4-2).

The list of primers for the modification of NDK-RNA

160522_mdvNDK_mod2_C_F (for Mod2-C construction)
(CTTCTTTGGAGAAGGTGAGGTGTGCCCGCGAACCCGTTAGGTTTGCGGCCGCACTCGAG
ACATCTAGAGCATCACGGTCG)
160522_mdvNDK_mod2_C_R (for Mod2-C construction)
(CCTTCTCCAAAGAAGTAGGCGATTTCACGAGCTGCCGATTCAACAGAATCGGAACCGTG
AGTACCGTTTTCGGTAAGGC)
160522_mdvNDK_mod2_E_F1 (for Mod2-CE construction)
(AGGTCTCGATATATACTCTAGAGATAATTTTGTTTAACTC)
160522_mdvNDK_mod2_E_F2 (for Mod2-CE construction)
(GACATCTAGAGCTACACGGTCGAACTCCCGTACG)
160522_mdvNDK_mod2_E_R1 (for Mod2-CE construction)
(GACATCTAGAGCTACACGGTCGAACTCCCGTACG)
160522_mdvNDK_mod2_E_R2 (for Mod2-CE construction)
(TATATATCGAGACCTTCTAGAGCTAGAGCACGG)
160803_mdvNDK_ mod2CE_X_F (for Mod2-CE-X construction)
(AATGCGCTGGCTGGTACCCTACGGGCTGACTATGCAGACAGCCTTACCGAAAACGG)
160803_mdvNDK_ mod2CE_X_R (for Mod2-CE-X construction)
(ACCAGCCAGCGCATTTGCCGGATTGGTAGCACCCAACAGATCGCGGTGGCGCTGAAC)

The list of primers for the other experiments

qRT-PCR_Univ7B (GCAAGTGACTCAGGATTCGTACATAATATCGTCTCCGTAAACAGTG)
qRT-PCR_WildB (GGTAGTGTTGTTACCTACGAGAAG)
qRT-PCR_Univ6B (GCAAGTGACTCAGGATTCGTAC)
090519_mdv+3 (GGGTCACCTCGCGCAGC)

090602_mdv-3 (CCGGAAGGGGGGGACGAGG)
160706_mdvNDK_1stRTanneal_R
(GCAAGTGACTCAGGATTCGTACGGTTTTCCATCGTGTTCAGC)
160517_mdvNDK_F3 (AGGAGGATATACACATGGCTAT)
160517_mdvNDK_F6 (CCATCATTAAACCAAATGCAGTAGC)
161004_mdvNDK_1stRTanneal_F2
(TAAGCGAATGTTGCGAGCACAGGAGGATATACACATGGCTAT)
170608_mdvNDK_1stRTanneal_F3
(TAAGCGAATGTTGCGAGCACCCATCATTAAACCAAATGCAGTAGC)
160516_mdvNDK_qPCR_R (GGTTTTCCATCGTGTTCAGC)
170315_ZB_39_A_F (TGATGGTGCCCTCGTCGGATCGGTCCTAATCAATC)
170315_ZB_39_A_R (ACGAGGGCACCATCACCGTAACCGTCAGGTATAG)

4-2-2. Modification of NDK-RNA Structures

The preparation of Mod2 sequence was outsourced to FASMAC (Kanagawa, Japan). The sequence was constructed by adding 32 mutations to the original NDK-RNA. To create Mod2-CE sequence, first, 9 mutations were added to the Mod2 sequence and constructed Mod2-C sequence. The Mod2-C sequence was inserted in the pUC19 plasmid. The mutated DNA sequence was prepared by mutagenic PCR of the plasmid using the corresponding primers and circularized in self-infusion method. The 25 ng DNA and In-Fusion HD enzyme premix (Clontech) (total 5 μl), were incubated at 50°C for 15 minutes, then mixed with 20 μl TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Next, the obtained plasmids containing the target DNAs were transformed into competent cells of *E.coli*. as described in (2-2-7). The amplified plasmid was purified from the cultured cells using PureLink Quick Plasmid DNA Miniprep Kits (Invitrogen). Next, Mod2-CE was constructed from Mod2-C by adding 3 mutations. An insert DNA and vector DNA were amplified by mutagenic PCR using the corresponding primers. The 25–50 ng insert DNA was inserted into the 50–100 ng vector DNA in the Infusion method described above. Mod2-CE-X was constructed by adding 8 mutations to Mod2-CE by self-Infusion as described above. The RNAs were prepared by *in vitro* transcription as described in (2-2-2) with the exceptions that PureLink PCR Purification Kit (Invitrogen) or PureLink RNA Mini Kit (Invitrogen) was used for DNA or RNA purification respectively. The list of introduced mutations were shown in (Table S1). RNA structures were predicted by the Vienna RNA software (centroid structure).

4-2-3. TcCRR Reaction of the Cooperative RNA Replication System

Plus-strands of 10 nM Rep-RNA and/or 10 nM NDK-RNA were mixed with the customized PURE system (Table 2-1) with the exceptions: 0 nM NDK, 0 nM Myokinase, 25 nM Creatine Kinase, 0 nM

CTP, and 1.25 mM CDP, and kinase-free EF-Tu and ribosome (given by Dr. Ichihashi). To create water-in-oil emulsion, 10 μl of the mixture was strongly stirred in 1 ml saturated oil by homogenizer POLYTRON PT-1300D, KINEMATICA). The saturated oil was the same as one used in (2-2-3). After incubating the emulsion at 37°C, the minus-strand of Rep-RNA concentration was measured by quantitative RT-PCR as described in (2-2-5). The plus-strand concentration was also measured by the same method as the minus-strand measurement but with the primer qRT-PCR_Univ7B for the reverse transcription and the primers qRT-PCR_WildB and qRT-PCR_Univ6B for the PCR. For the measurement of NDK-RNA concentration, the primer 160706_mdvNDK_1stRTanneal_R or 161004_mdvNDK_1stRTanneal_F2 was used for the reverse transcription of the plus- or minus-strand, and the primer set 160517_mdvNDK_F3 and qRT-PCR_Univ6B, or 160516_mdvNDK_qPCR_R and 2nd TPCR primer was used for the following PCR. For the comparison of TcCRR activities between the designed NDK-RNAs (Fig. 4-5), the RNAs containing Sp-GTP-α-S, prepared as described in (2-2-2) were used, and after the reaction, the original RNAs were degraded with iodine before the quantitative RT-PCR as described in (2-2-6). For the TcCRR reaction using the evolved clones (Fig. 4-10B), the primer 160517_mdvNDK_F6 was used instead of the primer 160517_mdvNDK_F3 for the measurement of plus-strand NDK-RNA concentration.

4-2-4. Simulation

R_1: Number of RNA1 molecules in each compartment

R_2: Number of RNA2 molecules in each compartment

P: Number of parasitic RNA molecules in each compartment

k_1: Rate constant of RNA1 replication

k_2: Rate constant of RNA2 replication

k_p: Rate constant of parasite replication

N: Total number of RNA1 and RNA2

K: Fixed number of the sum of RNA1 and RNA2 numbers

L_1: Factor depending on the length of RNA1

L_2: Factor depending on the length of RNA2

L_p: Factor depending on the length of parasitic RNA

C: Carrying capacity (corresponding to NTP)

D: Fixed dilution rate

M: Number of compartments

S: Probability of parasite generation

F: Fusion-Division number

m: random number between 0 and 1 from uniform distribution

To know the dynamics of cooperative RNA replicators with the emergence of parasitic RNAs, I constructed the simplest computer simulation model for two cooperative RNA replicators in compartments and investigated conditions allowing the cooperators to survive. The summary of the simulation method was described in (Fig. 4-1).

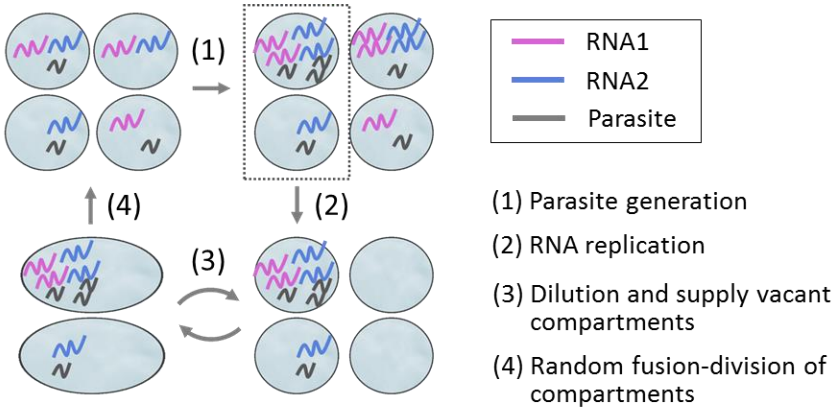


Figure 4-1. Summary of the simulation method for repetitive replication of cooperative replicators.

There are three types of RNAs: RNA1, RNA2, and parasites. RNA1 and RNA2 replicate interdependently, and the parasites are replicated by RNA1 and RNA2. The simulation was run by repeating the following processes: (1) parasites generation depending on the number of RNA1 and RNA2 molecules in each compartment, (2) RNA1, RNA2, and parasites replication, (3) dilution of the compartments (supply vacant compartments to keep the number of compartments), (4) random fusion-division of the compartments. The detailed method was described in the main text.

In this simulation, I assumed the simplest case of the replication of two cooperative replicators RNA1 and RNA2 in a compartment, where each replication depends on both the replicator and the other's concentrations. The ordinary differential equations of their concentrations are described as follows:

$$\frac{\mathrm{d}R_1}{\mathrm{d}t} = k_1 R_1 R_2 \left(1 - \frac{L_1 R_1 + L_2 R_2 + L_\mathrm{p} P}{C}\right) \qquad [S1]$$

and

$$\frac{\mathrm{d}R_2}{\mathrm{d}t} = k_2 R_1 R_2 \left(1 - \frac{L_1 R_1 + L_2 R_2 + L_\mathrm{p} P}{C}\right). \qquad [S2]$$

The parasitic replication depends on both the RNA1 and RNA2 concentration. The ordinary

differential equations of its concentration is described as follows:

$$\frac{dP}{dt} = k_p P R_1 R_2 \left(1 - \frac{L_1 R_1 + L_2 R_2 + L_p P}{C}\right). \quad [S3]$$

I assumed parasitic RNA stochastically appears at the beginning of each round in dependence on the number of RNA molecules in each compartment as follows.

$$P = P + 1 \qquad (SR_1R_2 \geq m)$$

and

$$P = P \qquad (SR_1R_2 < m)$$

Then the dynamics of total number of RNA1, RNA2, and parasites were simulated as the following processes, and their total numbers were calculated at the end of each round:

(1) 0.5*N* molecules of RNA1 and RNA2 were randomly distributed to M compartments according to the Poisson distribution.

(2) Parasites were generated in each compartment as described above.

(3) The RNA1, RNA2, and parasites were replicated according to the Eqs. S1, S2, and S3, solved by the python package scipy.integrate.odeint for a fixed reaction time.

(4) The compartments are diluted *N*/*K*-fold so that the total number of RNA1 and RNA2 is approximately constant at the beginning of each round. If *N*/*K* are less than *D*, the compartments are D-fold diluted.

(5) Vacant compartments were supplied to prepare *M* compartments.

(6) Each compartment was randomly fused with one of the *M* compartments allowing duplication, then divided into two compartments containing randomly distributed RNA molecules according to the binomial distribution. This process is repeated for *F* times.

(7) Repeat the processes (2) to (6) for rounds.

4-2-5. Long-term Replication Experiment

The TcCRR reaction was conducted in 1 ml emulsion using 10 nM Rep-RNA and/or 10 nM NDK-RNA. After 4 hours incubation at 37°C, an aliquot of the emulsion (200 µl (Fig. 4-7B, C, round 1–5 in D), 100 µl (Fig. 4-7D after round 5), 50 µl–200 µl (Fig. 4-7E from the round 18 and fig. 4-9 from the round 20)) was added to a fresh saturated oil containing 10 µl of the translation system to prepare 1 ml emulsion for the next round. The concentration of plus-strands Rep-RNA and NDK-RNA was measured at every round by quantitative RT-PCR. The NDK-RNA concentrations shown in (Fig. 4-7E, after the round 18) was the result of remeasurement using the different primer 160517_mdvNDK_F6 because one dominant mutation (U234G), enriched through 50 rounds replication experiment (Fig. 4-11A), was located at the original primer region.

4-2-6. Detection of the Parasitic RNAs

Water phase was recovered at each round by centrifuging (15000 rpm, 5 min) 300 μl of the reacted emulsion. The recovered solution was mixed with 4 vol. of diethyl ether and centrifuged (10000 rpm, 1 min). After vaporizing the ethyl phase, the RNA sample was purified by using PureLink RNA Mini Kit (Invitrogen) and subjected to 8% polyacrylamide gel electrophoresis. The gel was stained with SYBR Green II (Takara).

4-2-7. Sequence Analysis

32 clones of both the Rep-RNA and NDK-RNA were randomly cloned at the 50 round (Fig. 4-7E) in the same method to clone Rep-RNA as described in (2-2-7) with the exception that the primer 090602_mdv-3 was used for reverse transcription, and the primer set 090519_mdv+3 and 090602_mdv-3 was used for PCR. The same primers and vector were used for the cloning of NDK-RNA. Mutations were identified by comparing the sequences to the original Rep-RNA or NDK-RNA.

4-2-8. Construction of the isolate clones with dominant mutations (Evo).

To construct Evo of Rep-RNA, the mutations listed in (Fig. 4-10A) were introduced to the R-39 clone in the same method as described in (4-2-2) using the primers 170315_ZB_39_A_F and 170315_ZB_39_A_R. The N-1 clone was used for Evo of NDK-RNA.

4-2-9. Analysis of Replication Activities

1 nM plus-strand Rep-RNA or NDK-RNA was mixed with 10 nM purified replicase obtained in a previous study[91] and the PURE system (Table 2-1). The sample was incubated at 37°C for 15 minutes, and the concentration of synthesized minus-strands were measured by quantitative RT-PCR as described in (4-2-3) with the exception that the primer 170608_mdvNDK_1stRTanneal_F3 was used instead of the primer 161004_mdvNDK_1stRTanneal_F2 for the measurement of minus-strand NDK-RNA concentration.

4-2-10. Analysis of Enzyme Activities

For Rep-RNA, 10 nM plus-strand clones were incubated with the PURE system (Table 2) without UTP at 37°C for 2 hours. Next, 0.1 vol. of the solution containing translated replicase was mixed with the fresh PURE system, 0.03 mg/mL streptomycin, and 10 nM plus-strand NDK-RNA to stop translation and start replication of the NDK-RNA. After incubation at 37°C for 20 minutes, the concentration of synthesized minus-strand NDK-RNA was measured by quantitative RT-PCR as described in (4-2-3). For NDK-RNA, 100 nM plus-strand clones were incubated with the newly customized translation system (4-2-3) at 37°C for 2 hours. Next, 0.0001 vol. of the solution

containing translated NDK was mixed with 100 nM Rep-RNA, 1 μM purified replicase obtained in a previous study[91], 1.25 mM ATP 1.25 mM GTP, 1.25 mM UTP, 1.25 mM CDP, 25 μg/ml streptomycin, 125 mM Tris-HCl (pH 7.8), 10 mM magnesium acetate, 0.01% BSA. After incubation at 37°C for 45 minutes, the concentration of synthesized minus-strand Rep-RNA was measured by quantitative RT-PCR as described in (2-2-6).

4-3. Results and Discussion

4-3-1. Instability of Cooperative RNA Replication System, Proved by Computer Simulation

As a cooperative replication system, I considered two obligatory mutualistic RNA replicators, in which they can replicate only in the presence of the partner RNAs (Fig. 4-2A). Through the replication, a parasitic RNA spontaneously emerges by replication errors and dominantly replicate regardless of its cooperativity by exploiting other cooperative RNA replicators (Fig. 4-2B). Now, I consider the case that this system is compartmentalized and replicated via random fusion-division of the compartment (Fig. 4-2C). In high RNA concentrations, parasites rapidly appear and competitively inhibit the replication of cooperators, then the parasites spread in population through fusion-division of compartments (Fig. 4-2D). If the parasite concentration is much higher than cooperators, even though cooperators could escape from parasites into new compartments by sufficient dilution, they no longer replicate due to the lack of partners in the same compartment. Similarly, in low RNA concentrations, maintaining the pair of the cooperators through division of compartments is difficult, and therefore the RNAs do not sufficiently replicate. Therefore only the possible sustainable range of RNA concentrations is middle, in which the parasite generation may not be seriously frequent, and cooperators can coexist.

To prove this idea, I constructed the simplest computer simulation model for two cooperative RNA replicators in compartments and investigated conditions allowing the cooperators to survive. I conducted this simulation experiment to ensure that my model described above reproduces general results of instability of the replicators, found in previous theoretical studies that are slightly different from my model in the following points. First, some previous models assumed that compartments divide in accordance with internal RNA compositions[49,51,57], whereas in this study, they divide immediately after random fusion with other compartments, which may be relevant to a primitive compartment reproduction process[6,90]. Second, another study assumed constant RNA numbers in every compartment[48], whereas they varies in my model. In the simulation model, I prepared RNA1, RNA2, and parasites. RNA1 and RNA2 replicate interdependently, and the parasites are replicated by RNA1 and RNA2. I repeated RNA replication, dilution, and random fusion-division of the compartments, during which parasites stochastically appeared depending on the number of RNA molecules in each compartment (Fig. 4-1). I fixed total RNA molecules in population until dilution rate was below a certain value. In high number of RNA molecules, I found that their replications were inhibited by excess parasite replication (Fig. 4-3), and the population went extinct because RNA1 and RNA2 are finally separated into different compartments. I found that on rare occasion, they started replicating again by overcoming parasites in an oscillating manner but tended to be destroyed soon after reaching high RNA numbers again. Conversely, in low number of RNA molecules, even though the parasite generation was repressed, the cooperators did not replicate due

to the lack of partners in each compartment. On the other hand, in the middle number of RNA molecules, a tiny number of parasites occasionally emerge but they were washed out soon and did not spread in the population. The coexistence of the cooperators were therefore possible in the middle number of RNA molecules, basically consistent with a previous theoretical study[48].

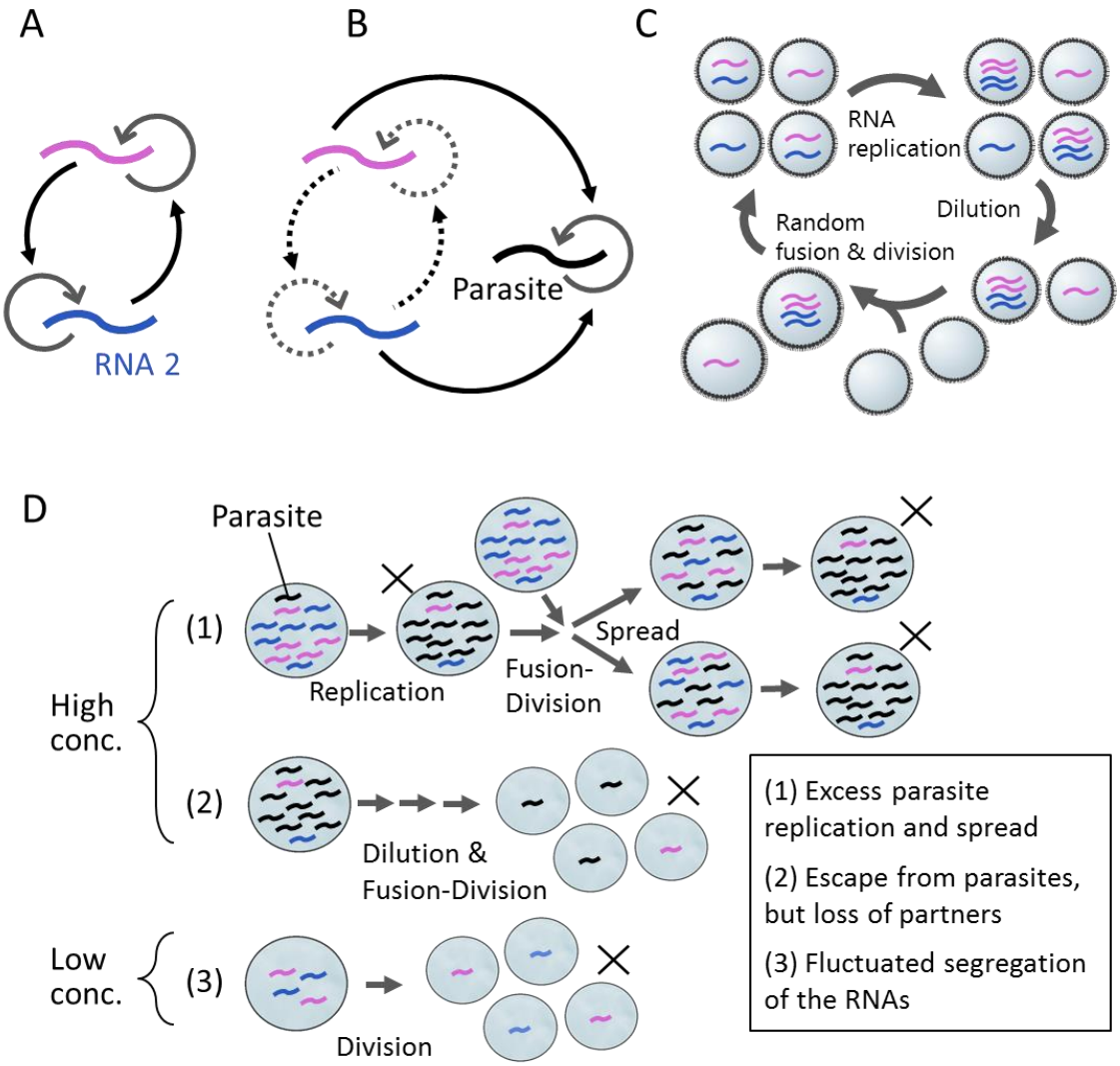


Figure 4-2. Cooperative RNA replication system.

(A) Obligatory mutualistic RNA replicators. The replication of each RNA depends on the existence of the other RNA. (B) The emergence of parasites destroy the mutualism by exploiting the cooperative behaviors. (C) The cycle of RNA replications. The cooperative RNA replicators replicated in compartments. Then the compartments were diluted and undergo fusion-division with new compartments containing fresh nutrients to start next replication. (D) Instability of the cooperative RNA replication system. The detail was described in the manuscript.

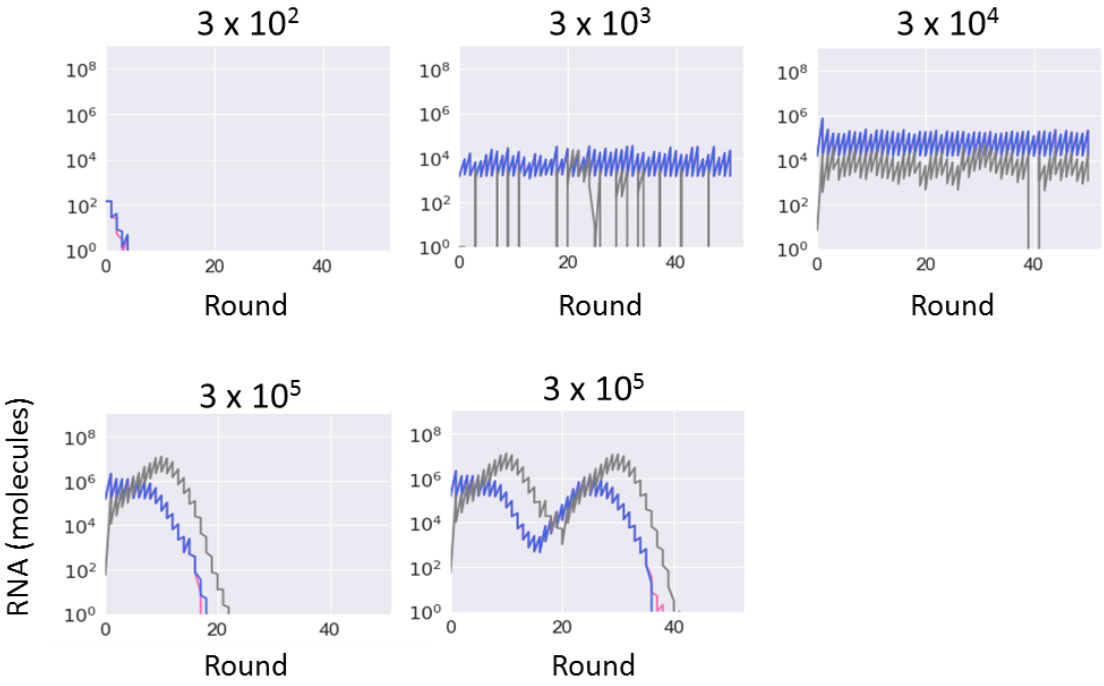


Figure 4-3. Simulation of the repetitive replication of cooperative replicators.

The simulation result with different number of initial RNA molecules (number above each figure). The total number of RNA1 and RNA2 were diluted to the initial number of molecules (K = 3 x 10^2, 3 x 10^3, 3 x 10^4, 3 x 10^5). The pattern of the result with 3 x 10^5 initial molecules was occasionally changed so that RNAs replication were recovered by overcoming parasite existence. k_1 = 0.15, k_2 = 0.15, k_p = 0.3, L_1 = 10, L_2 = 4, L_p = 1, C = 3000, D = 5, M = 10000, S = 0.0002, F = 4.

4-3-2. Construction of a Translation-coupled Cooperative RNAs Replication (TcCRR) System

I attempted to construct a translation-coupled cooperative RNAs replication (TcCRR) system by combining the TcRR system[62], consisting of the translation system and Rep-RNA which encode replicase enzyme, and NDK-RNA which encodes *E.coli* nucleotide diphosphate kinase (NDK) and has replicase recognition sites. I omitted NDK from the translation system so that newly synthesized NDK should be essential for the replication of the TcCRR system. In approximately 1 μm radius of compartments (water-in-oil emulsion), the reaction is expected to proceed in the following way: NDK is translated from NDK-RNA and synthesize CTP from CDP, then replicase, translated from Rep-RNA, replicates both the RNAs with the synthesized CTP (Figure 4-4). Through the replication, it is expected that mutations were spontaneously introduced into both Rep-RNA and NDK-RNA as inheritable variation, similar to the TcRR reaction, and mutants with higher fitness should be

selected in a condition. Therefore, both Rep-RNA and NDK-RNA are evolvable in the TcCRR system.

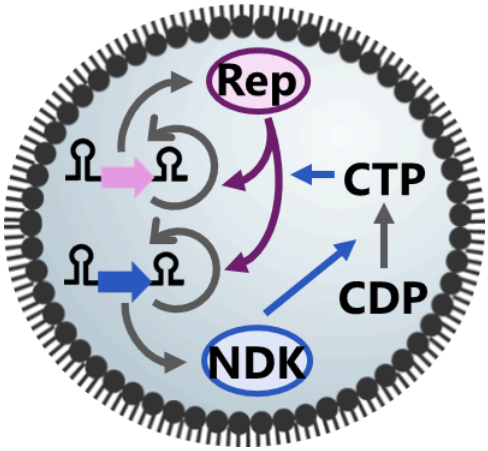


Figure 4-4. Translation-coupled cooperative RNA replication (TcCRR) system
The detail was described in the manuscript.

However, the original NDK-RNA sequence was poorly replicated by Qβ phage replicase. The replication efficiency is known to depend on secondary structures of template RNAs, and the structures can be modified to improve the replication with synonymous mutations without changing amino acids so that protein sequences are not changed [92]. I therefore designed the NDK-RNA structure step by step with total 49 mutations (Appendix A4-2) to make rigid structures (Fig. 4-5). By combining the designed NDK-RNAs and a translation-improved Rep-RNA (R11, obtained in Chapter 2), I found that the improvement of NDK-RNA replication during the TcCRR reaction finally reached more than 250-fold (Mod2-CE-X, Fig. 4-6). Neither Rep-RNA nor NDK-RNA replicated much in the absence of one of the RNAs; they sufficiently replicated when they coexisted (Fig. 4-7A). The complementary strands also sufficiently replicated only in the presence of the partners (Fig. 4-7B).

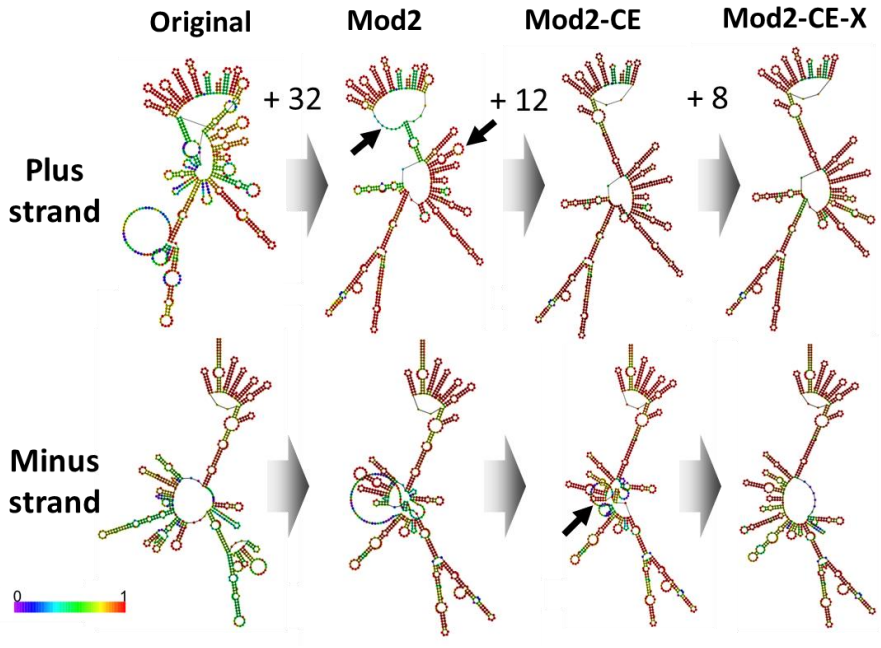


Figure 4-5. Stepwise addition of mutations to design a replicable NDK-RNA

The stepwise modifications of NDK-RNA structure of plus- and minus-strand were shown. Added mutation numbers were shown above gray arrows at each step of modification. The entire structures were predicted by the Vienna RNA software. The nucleotide color shows the probability to form the structures. The black arrows indicate the loop structures that were modified at the next steps. The original structure was the translation-improved RNA (Appendix A4-1). All the added mutations were listed in (Appendix A4-2).

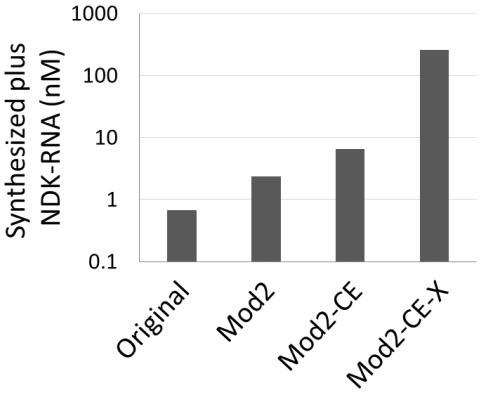


Figure 4-6. TcCRR reactions with designed NDK-RNA

The TcCRR reaction was conducted with 10 nM plus-strand Rep-RNA and 10 nM of each designed plus-strand NDK-RNA clone at 37°C for 4 h. The synthesize plus-strand NDK-RNA concentration was measured by quantitative RT-PCR after degrading the original NDK-RNA.

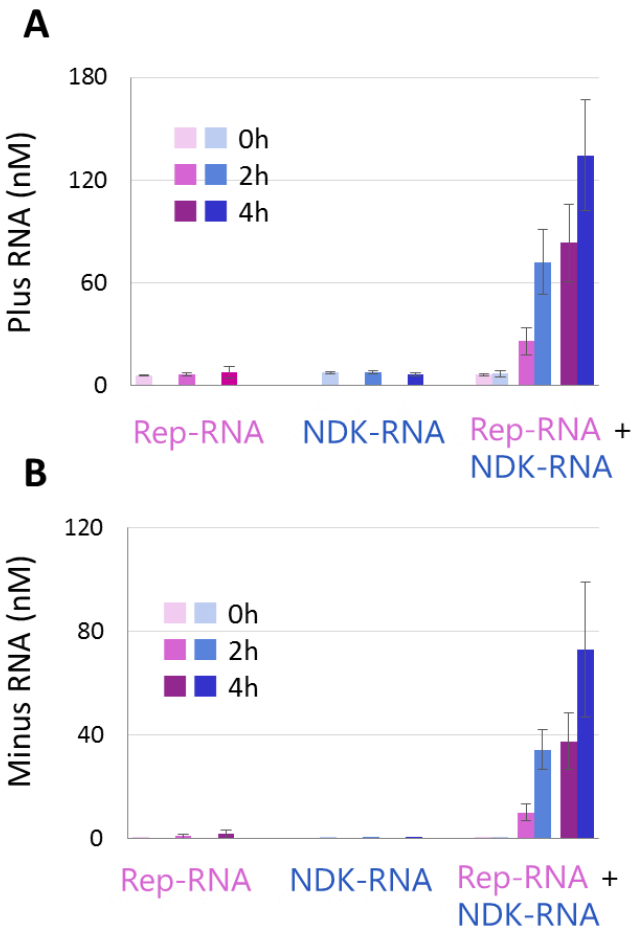


Figure 4-7 Interdependence of Rep-RNA and NDK-RNA replications.

The translation-coupled replication reaction was conducted in water-in-oil emulsion with 10 nM plus-strand Rep-RNA, 10 nM plus-strand NDK-RNA, or the 10 nM of the both RNAs (TcCRR reaction) at 37°C for 0, 2, or 4h. The plus-strand RNA (A) or minus-strand RNA (B) concentrations were measured by quantitative RT-PCR. The error bars indicate standard deviation (n=3).

4-3-3. Long-term Replication Experiment. (1) Destruction of the TcCRR System

I performed long-term replication experiments with the TcCRR system in a transfer method with various dilution rates so that the RNA concentrations were kept in a certain range. First, I performed the TcCRR reaction with 10 nM Rep-RNA (R11) and 10 nM NDK-RNA (Mod2-CE-X) at 37°C for 4 h (Fig. 4-8A). After the incubation, the compartments were diluted 5-fold, and fresh translation system was supplied through manual fusion-division with new compartments. During the first few rounds, both Rep-RNA and NDK-RNA sufficiently replicated. Afterwards, however, their replications stopped and the RNAs were serially diluted (Fig. 4-8B). In this phase, I observed a

significant generation of approximately 220 nt short RNAs (Fig. 4-9A), known as parasites in a previous study[56]. The excess parasites may have inhibited the cooperators' replications. I continued the experiment up to the 36 round, but the RNA concentrations remained undetectable.

Similarly, when I started the repetitive replications from low RNA concentrations (0.03 nM), at which the number of RNA molecules per compartment is below one, both Rep-RNA and NDK-RNA could not sufficiently replicate and were gradually diluted out (Fig. 4-8C).

Next, I increased the dilution rate from 5- to 20-fold until the replications were inhibited so that the RNA concentrations were kept lower than the first experiment (Fig. 4-8D). In this condition, the parasites appeared but the concentrations were smaller than the first experiment (Fig. 4-8D, 4-9A). Consequently, the cooperative replicators continue to replicate partly, and after sufficient dilution, they eventually overcame parasite existence and recovered their replications probably because of the existence of compartments that did not contain parasites but cooperators. However, once the excess parasites appeared (around round 26–29) again, the cooperator's replications stopped, and they were no longer recovered, at least up to 48 rounds. The oscillating RNA concentration was also observed in the simulation experiment, and the instability was confirmed (Fig. 4-3).

4-3-4. (2) Sustainable Replication of the TcCRR System

Next, I increased the dilution rate from the round 18 of the experiment (Fig. 4-8D) and kept the RNA concentration below 0.1 nM by controlling dilution at every round (Fig. 4-8E). During the replications, emerging parasites were lower compared to the replications at high RNA concentrations (Fig. 4-9B), and in consequence, both the Rep-RNA and NDK-RNA sustainably replicated for more 32 rounds (total 50 rounds). They replicated 2^{158}- or 2^{160}-fold in total. I repeated a similar experiment and such sustainable replication was reproduced (Fig. 4-10).

Above experiments showed that if the cooperative RNA replicators do not reach too high concentrations to produce excess parasites, and a number of compartments contains adequate RNA molecules, the replication of the TcCRR system is sustainable.

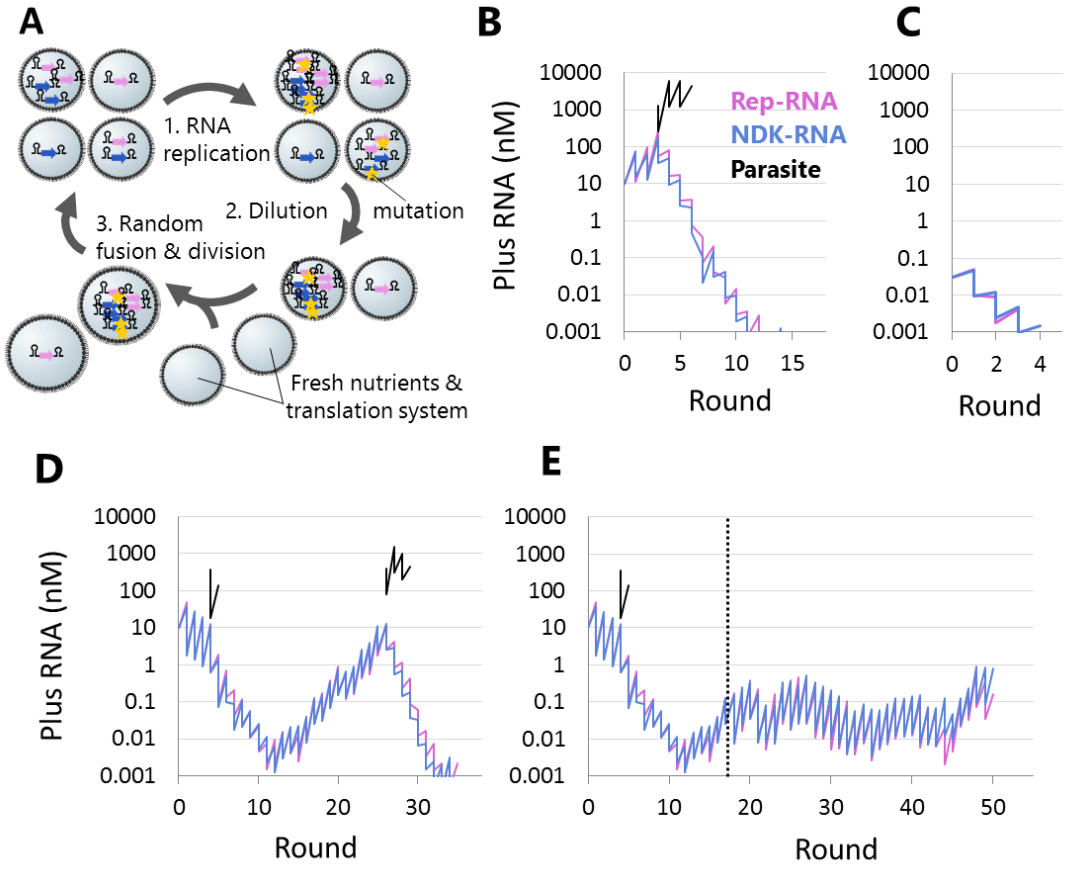


Figure 4-8 Long-term replication experiment

(A) Scheme for the transfer experiment. (1) The TcRR reaction in water-in-oil emulsion was conducted with Rep-RNA and NDK-RNA at 37°C for 4h. (2) The compartments were diluted, and (3) fresh nutrients and translation system in large emulsion were supplied through random fusion-division of the compartments for the next rounds. (B) The dynamics of RNA concentration starting from 10 nM plus-strand Rep-RNA and 10 nM plus-strand NDK-RNA, constantly 5-fold diluted after replication. The Rep-RNA (pink line) and NDK-RNA (blue line) concentrations were measured by quantitative RT-PCR at the end of each round. Parasite concentration (black line) was quantified from band intensities after polyacrylamide gel electrophoresis of the RNA samples (Fig. S5C). (C) The dynamics of RNA concentration starting from 0.03 nM plus-strand Rep-RNA and 0.03 nM plus-strand NDK-RNA, constantly 5-fold diluted after replication. (D) The dynamics of RNA concentrations when dilution rates of the first five rounds was changed from 5 to 20. (E) The dynamics of RNA concentrations when dilution was increased from the round 18 of the C (black dotted line) so that the concentrations were below 0.1 nM at the beginning of each round.

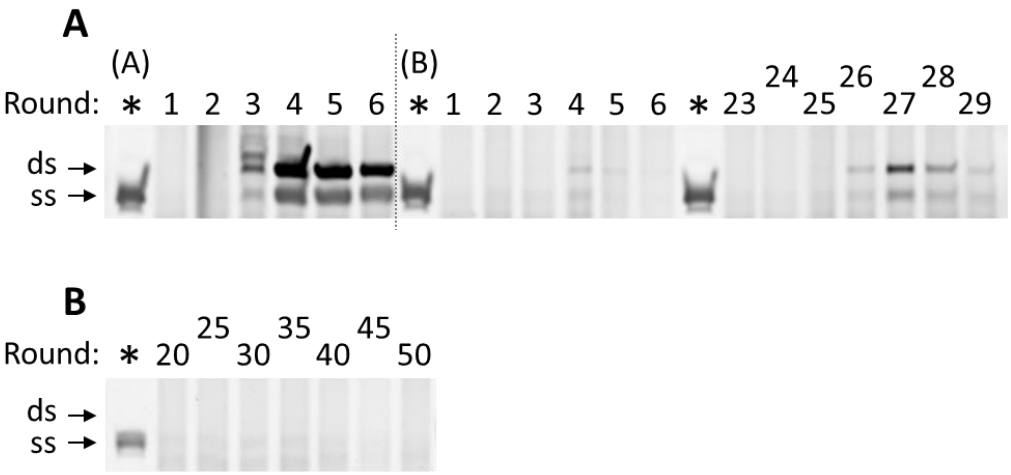


Figure 4-9. Parasite emergence and the excessive replications

(A) The native polyacylamide gel electrophoresis of the reaction mixtures during the long-term replication experiments (Fig. 4-7B, C). "ss" and "ds" represent single strand or double strand of the parasitic RNA. Asterisk indicates the size marker of the parasite (s222). Band intensity was quantified and used for the plots of Fig. 4-7B and 4-7C. (B) Another native polyacylamide gel electrophoresis using the reaction mixtures during the sustainable replication (Fig. 4-7E).

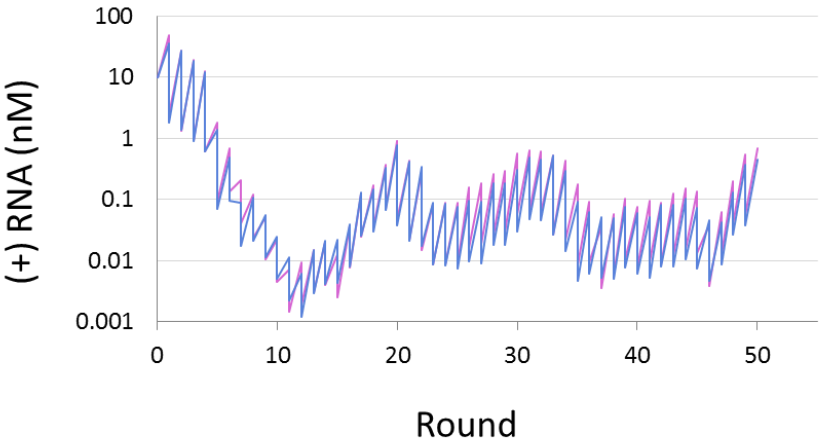


Figure 4-10. Reproducibility of sustainable replication

Dilution rate was increased from the round 20 of (Fig. 4-7D) and keep the RNA concentration below 0.1 nM by dilution at beginning of each round.

4-3-5. Mutation Analysis

Through the replication cycles, mutations are spontaneously introduced into both the Rep-RNA and NDK-RNA by replication errors. Thus, I expected some mutations, if beneficial, were accumulated on the cooperative RNA replicators through our transfer experiments. I cloned 32 RNAs for Rep-RNA and NDK-RNA at the round 50 of (Fig. 4-7E) respectively and analyzed their sequences. The average mutation numbers were 5.8 for Rep-RNA and 3.2 for NDK-RNA (Table 4-1), and some mutations were enriched in the populations. These mutation numbers cannot be explained by the mutation rate of RNA preparation, thus most of them were introduced by replications in the long-term replication experiment. Most frequent mutations were a set of A206G, U746C, and G975A for Rep-RNA, detected in 10 of the 32 clones (31%) and a set of C86U and U234G for NDK-RNA, detected in 24 of the 32 clones (75%). These dominant mutations may be evidence of adaptive evolution, because mutations, generated as replication errors, can become dominant in population by selection in a condition, leading to evolution.

Table 4-1. Mutation analysis

	Rep-RNA	NDK-RNA
Total mutations	5.8 ± 1.9	3.2 ± 1.4
Untranslated region	1.6 ± 0.94	1.3 ± 0.57
Synonymous	1.8 ± 1.2	1.5 ± 0.80
Nonsynonymous	2.4 ± 1.2	0.47 ± 0.84

4-3-6. Evolution of the TcCRR System

Next, I constructed isolate mutants (Evo), containing only the dominant mutations (Fig. 4-11A). Then I conducted the TcCRR reactions with all the four combinations: the original (Ori) or the evolved clone of Rep-RNA with those of NDK-RNA (Fig. 4-11B). The evolved pair exhibited higher TcCRR activities for both Rep-RNA and NDK-RNA than the original pair. In contrast, in the mixture of the evolved and the original clones, the evolved clones inhibited the replication of the original clones, indicating that the evolved clones were parasitic against the original clones. There are two factors affecting the TcCRR activities: how fast an RNA is replicated by replicase (replication activities, called selfishness) and how efficiently translated enzymes altruistically support the replication of surrounding RNAs (enzyme activities, called cooperativity). Both Rep-RNA and NDK-RNA increased their replication activities (Fig. 4-12), while maintaining or slightly increasing enzyme activities (Fig. 4-13). Therefore the change of the TcCRR activities can be mainly explained by the increase in the replication activities. These results indicated that the evolved clones inhibited the replication of the original partners by sequestering replicase as with parasites; the evolved pair can sufficiently replicate because both the RNAs increased their replication activities, which is an evidence for co-evolution. Whether the coordinated improvement was inevitable or not remains

unclear, but such evolution may be a reason why mutualism is prevailing in nature.

A

Rep-RNA (Evo)	A206G , U746C, G975A	31%
NDK-RNA (Evo)	C86U, U234G	75%

B

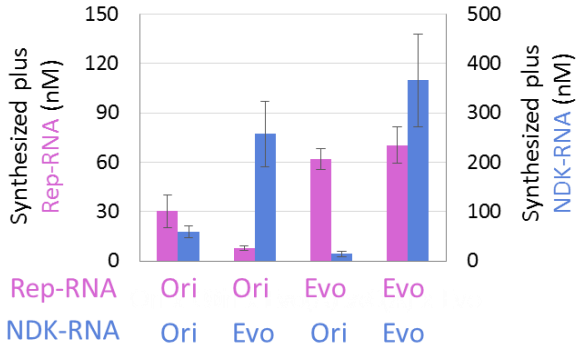
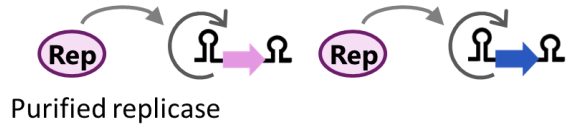


Figure 4-11. Improvement of the TcCRR activities

(A) Constructed Rep-RNA and NDK-RNA clones with only the dominant mutations. (B) TcRR reactions using various combinations: 10 nM of the plus-strand original (Ori) or the evolved (Evo) Rep-RNA clone with 10 nM of the plus-strand original or the evolved NDK-RNA. Pink and blue bars represent synthesized plus-strand Rep-RNA and NDK-RNA concentrations during the reactions respectively, measured by quantitative RT-PCR. The error bars show standard error of the mean (n = 6 for the Ori x Ori combination, n = 3 for the others).

A



B

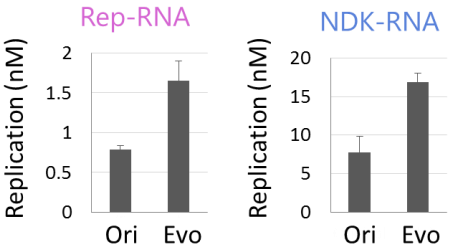


Figure 4-12. Change in replication activities

(A) Method for the measurement of replication activities. Plus RNA clones (Ori, original clone; Evo, evolved clone) were incubated with 10 nM purified replicase at 37°C for 15 min. (B) The concentration of synthesized minus-strand RNA was measured by quantitative RT-PCR. The error bars indicate standard error of the mean (n = 4 (Rep-RNA) or 3 (NDK-RNA)).

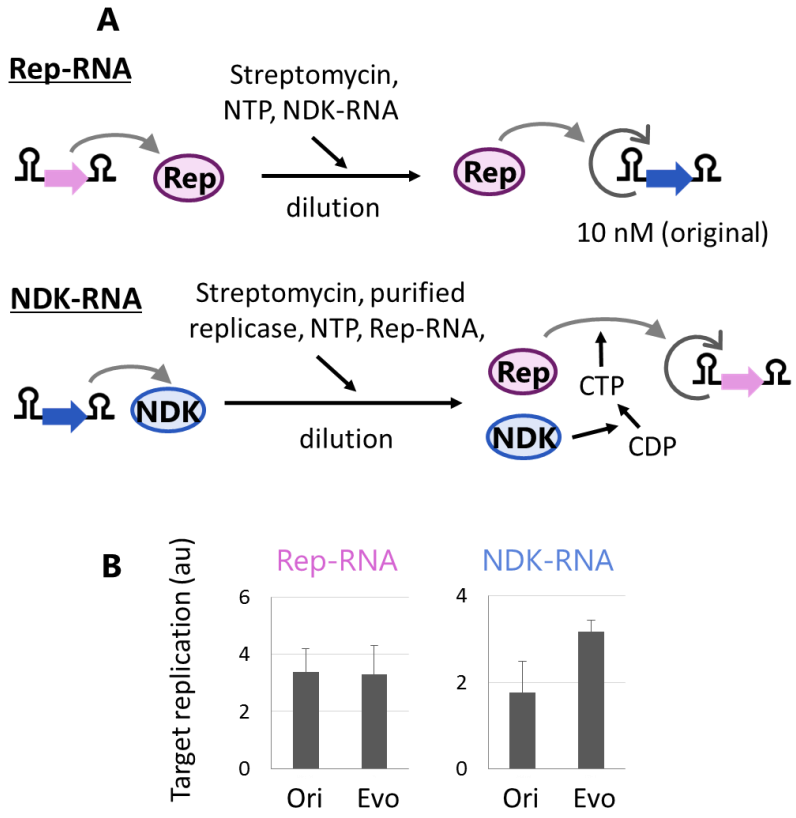


Figure 4-13. Change in enzyme activities

(A) Method for the measurement of enzyme activities. For Rep-RNA, 10 nM plus-strand RNA clone was translated without UTP to stop replication at 37°C for 2 h. An aliquot was mixed with streptomycin, UTP, and 10 nM plus-strand NDK-RNA clone to stop translation and start replicating NDK-RNA at 37°C for 20 min. The synthesized minus-strand NDK-RNA concentration was measured by quantitative RT-PCR. For NDK-RNA, 100 nM plus-strand NDK-RNA clone was translated at 37°C for 2 h. An aliquot was mixed with streptomycin, NTP, 100 nM plus-strand Rep-RNA clone, and 1 μM purified replicase to stop translation and start replicating Rep-RNA at 37°C for 45min. The synthesized minus-strand NDK-RNA concentration was measured by quantitative RT-PCR. The error bars indicate standard error of the mean (n = 3).

4-3-7. Enzyme Activity of all the Obtained Clones.

I analyzed enzyme activities of all the obtained clones. They showed varied activities from low to high (Fig. 4-14), and surprisingly, some of them were completely inactive. Therefore the replication of Rep-RNA and NDK-RNA in the TcCRR system was sustained even in the presence of loss-of-function mutants, which are essentially parasites because they replicate by exploiting surrounding cooperators, suggesting high robustness of the TcCRR system.

The varied enzyme activities could be explained by the concept of quasispecies[46,93]: a group of closely related sequences with several mutations, emerging with high mutation rate during replication. In particular, in NDK-RNA which has highly dominant mutations (75%), the clones formed such mutated sequences, and the decrease in the enzyme activities was correlated with mutation numbers.

Interestingly, the frequency of cooperative NDK-RNA mutants was much higher than Rep-RNA. Although no Rep-RNA clone has equivalent enzyme activity with the evolved isolate (Evo), 16 of 32 clones of NDK-RNA showed the same or higher activities than the evolved isolate. This high frequency of retained enzyme activities is consistent with low number of non-synonymous mutations (Table 4-1). These results indicated that NDK-RNA may have inherent properties to maintain their enzyme activities through evolution. There are at least two possible mechanisms to cause the difference: the 2.7-fold shorter NDK-RNA length than Rep-RNA, which reduce the frequency of mutations per replication, and the NDK property to form multimeric structures[94]. An impaired NDK monomer might fold with active monomers translated from other NDK-RNA in the same compartments and disrupt the whole complex. Such a dominant-negative property of impaired NDK-RNA mutants can also explain the low frequency of the non-synonymous mutations in the NDK-RNA clones (Table 4-1). This idea hypothesizes that primitive replication mechanism, such as random fusion-division of the compartments, may have been a driving force for the evolution of multimeric proteins.

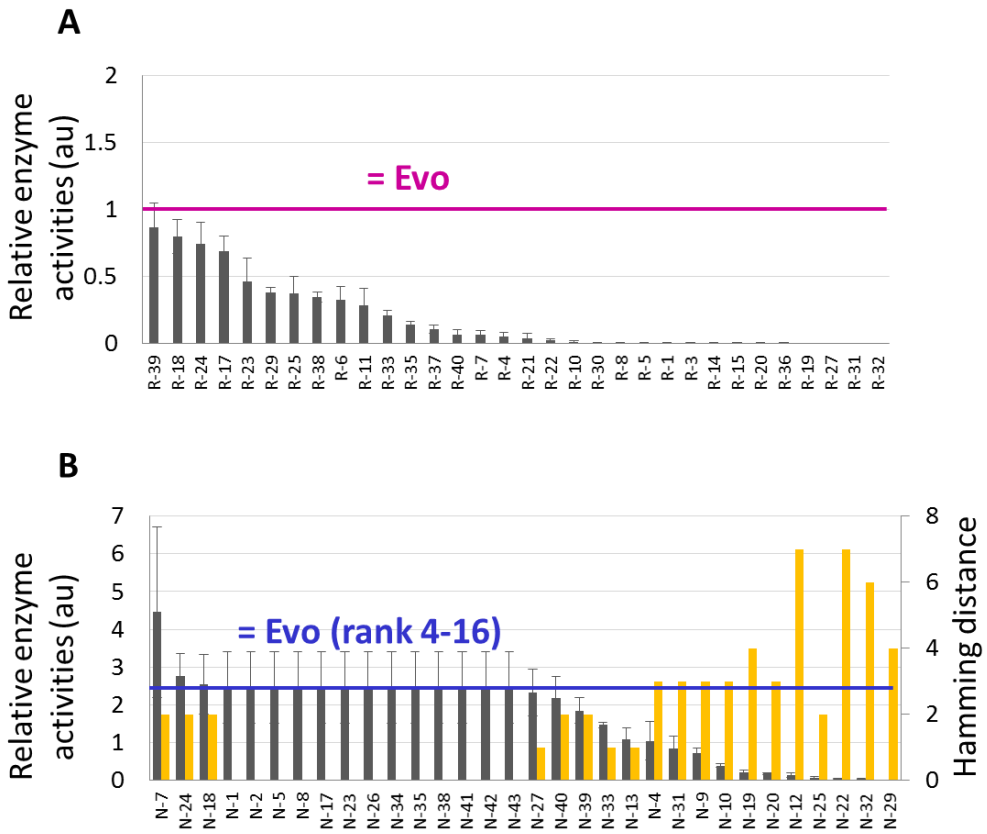


Figure 4-14. Enzyme activity of all the obtained clones.

The enzyme activities for Rep-RNA clones (A) and NDK-RNA clones (B) were measured in the same method as Fig. S9. The list of all the sequences were shown in (Table S3 and S4). The measured activities were normalized by that of the original clones and ranked in descending order. The horizontal lines show the enzyme activities of Evo. 13 NDK-RNA clones (rank 4-16) represents the activity of N-1 because they are an identical sequence. The error bars indicate s.e.m. (n = 3). The orange bars for NDK-RNA clones show hamming distance (number of different nucletides) from the dominant sequences (rank 4-16).

4-3-8. Coordinated RNA Concentration and Chromosome Structures

In this study, I found that the distinct RNA replicators can synchronously replicate by keeping their concentrations close in an evolutionary time scale (Fig. 4-8E, 4-11). The equal concentration of the mutualistic replicators is known to be the most beneficial because it minimizes internal competition for replicase, while maximizing the number of compartments containing complete replicator sets after division[95]. However, the replication of cooperators stochastically fluctuates, and replicators that

gain mutations to replicate faster are likely to be more amplified. It has been therefore considered that integrating distinct replicators as a chromosome structure encoding multiple genes is important to synchronize their concentrations[95]. On the other hand, our results showed that such synchronicity can be achieved without physically linking the replicators. Hence, the chromosome structures may have evolved by other selection pressures such as complete disruption of the cooperative replication system by parasites (Fig. 4-8B, D) or the load of deleterious mutations[96].

4-4. Chapter Summary

In this chapter, I constructed a new evolvable artificial genome replication system that has some extent of complexity: translation-coupled cooperative RNAs replication (TcCRR) system. I combined the TcRR system (an RNA replicator encoding replicase enzyme (Rep-RNA) and a reconstituted translation system) with another RNA replicator encoding NDK (NDK-RNA), which was artificially designed to be replicated. In the TcCRR system, Rep-RNA and NDK-RNA interdependently replicate through translation of their encoded replication and metabolic enzymes, and mutations are spontaneously introduced into both the cooperative RNAs by replication errors. The TcCRR system is more complex than the original TcRR system because the genetic information in the system level was expanded. The TcCRR system is the first experimental model of evolvable cooperative RNA replicators encoding distinct genetic information, and it is also the first example of successfully coupling autonomous replication and metabolism in an artificial genome replication system.

Using the TcCRR system in compartments, I investigated whether the complexity, coexistence of the two cooperative RNA replicators, were maintained in an evolutionary time-scale, especially by overcoming spontaneously appearing parasitic RNAs that replicate faster but do not cooperate. I found that there is a range of RNA concentrations that enables the cooperators to continuously replicate by simultaneously preventing excess parasite generations and loss of partners by fluctuated RNA distribution at compartment division. In this condition, the both the Rep-RNA and NDK-RNA replicated more than 2^{150}-fold (more than 150 generation). This is the first empirical demonstration of sustained cooperation between RNA replicators in compartments, and hence, the maintenance of cooperation is possible, confirming long-standing theories.

I also found that the TcCRR system kept a certain evolutionary ability to improve the TcCRR activities despite the limitation that the system must coordinate multiple RNA's replication. In fact, through the TcCRR reaction, mutations were introduced into both Rep-RNA and NDK-RNA, and particular mutants were selected in both the RNA populations. During the sustainable long-term replication, both the RNAs evolved to be faster replicators, while maintaining or slightly improving their cooperation, resulting in the improvement of the TcCRR activities. This result indicated that the increase in complexity does not limit the further evolution of a genome replication system.

Another interesting result from this chapter is different distribution of enzyme activities between Rep-RNA and NDK-RNA populations. The frequency of the loss of functions were lower in NDK-RNA than Rep-RNA, which could be explained by the change in their genome length and the NDK property to form multimeric structures. This finding proposes a new hypothesis that proteins that form multimeric complex may have been selected in early evolution of life, where compartments contained redundant genes. Further studies could confirm this theory and contribute to reconstructing a plausible scenario of the origin and early evolution of life.

Chapter 5. Summary

5-1. Perspective of Constructive Biology

In this dissertation, I investigated the evolutionary potentials of simple artificial RNA replication systems. In particular, I focused on two major evolutionary phenomena, adaptation to various conditions and the increase in complexity by acquiring another gene. In the first half of the dissertation (Chapter 2 and 3), I investigated adaptation of the TcRR system, consisting of a translation system and a short (2 kb) RNA genome encoding only a gene of RNA replicase, which is much simpler than any living things (e.g., *E. coli*, the "simple" model bacterial organism, containing 4000 genes). Despite the great simplicity, in the Chapter 2, I found that the RNA genome spontaneously adapted to a severe translation condition, where the concentration of ribosome in the translation system was reduced, and improved its TcRR activity by accumulating mutations. Previous studies also demonstrated evolution of artificial RNA replication systems in severe conditions, but with continuous supply of purified RNA replicase or a set of replication-related enzymes[37,38]. In contrast, the TcRR system adapted to a severe condition under restriction that functional RNA replicase must be translated from the RNA genome, similar replication manner as living things. Furthermore, I revealed that this adaptation partly resulted from a small change in RNA structures of the genome, increasing translation and TcRR efficiency, while keeping the function of RNA replciase. These result indicated that an artificial RNA replication system with only one gene (thus, the possibility of changes by mutations was limited and evolutionary potential was seemingly limited compared with living organisms) has a certain ability of adaptation. Moreover, the following evolution experiments in four different translation-impaired conditions, performed in the Chapter 3, revealed that the TcRR system adapted to all the conditions by accumulating different mutations depending on impaired steps of translation as the genome started to diversify, showing adaptation to diverse conditions compared with previous experimental evolution of artificial replication systems in limited conditions[25,26,37,42,38]. This result indicated that a replication system containing a short RNA with only one gene is sufficient to show great adaptation ability, allowing it to survive in various severe environments. Recently, another line of research has been conducted to investigate adaptive evolutionary ability of living organisms, that is, evolving microbes in fluctuating environments where particular parameters (e.g. temperature or resource availability) change through experimental time[97,98]**.** Such experiments contribute to understanding the environmental tolerance of evolving life and the effect of evolutionary history on future evolution. Although no research has been conducted in the same direction using artificial evolvable systems, the experiments should be important future works to understand the current evolutionary potential of artificial systems and achieve the same extent of evolutionary abilities as those of natural organisms from smaller set of molecules. It should be noted, however, that the experiments were performed only in various translation-impaired conditions and not in other impaired conditions such as different

temperatures and salt concentration. Therefore, future works include evolution experiments using the TcRR system in wider range of conditions to prove that the artificial system has the same extent of adaptation ability as living organisms.

In the Chapter 4, I focused on another major evolutionary phenomenon, the increase in complexity. In particular, I increased the complexity of the TcRR system, containing only one RNA replicator with one gene, by adding another RNA replicator encoding a different gene. I created a translation-coupled cooperative RNAs replication (TcCRR) system, in which two distinct RNA replicators, encoding different genes, can cooperatively replicate. The two cooperative RNAs are Rep-RNA and NDK-RNA, encoding a replication gene and a metabolic gene respectively. I investigated the sustainability and evolutionary ability of the compartmentalized TcCRR system and demonstrated that the increased complexity (coexistence of the cooperators) can be maintained without losing their cooperation in a certain range of RNA concentrations by overcoming spontaneously appearing parasitic RNAs. Thus the evolution of complexity is possible in a simple replication system. Furthermore, I found that the cooperative RNA replicators self-improved the TcCRR activities with mutations through repetitive replications in the stable condition, which is an evidence of adaptive evolution. Future research of comparing evolutionary outcomes between the TcCRR system and the TcRR system may lead to understanding how the evolutionary ability changed as the complexity increased. Overall, I found that even in a molecular level, unification of distinct cooperative replicators can be maintained, similar to recurring evolutionary transitions that originated higher order of organizations, such as eukaryotes from prokaryotes and multicellular organisms from unicellular organism. Finally, by combining this research outcome with ones obtained in Chapter 2 and 3, a simple artificial RNA replication system has considerable evolutionary potential to adapt to various conditions and become more complex.

One of the important challenges in the constructive biology is application of artificial systems in biotechnology industry. If combined with another replicating RNA that expresses useful enzymes, the TcRR system could be used as a platform for autonomous material production, which can be highly purified and improved through evolution. To achieve this goal, however, there is a large bottleneck: requirement of continuous supply of the translation proteins, especially ribosome, which is the largest and essential complex in the translation system and needed at more than micro-molar level. Efficient *in vitro* ribosome production is difficult because of the complicated pathway to the ribosome formation from three RNAs and more than 50 proteins, and thus we still have to obtain ribosome in a classical method using a tremendous amount of bacteria, which takes time and money. One possible strategy to overcome this hurdle is to modify the TcRR system to be as independent from ribosome as possible. The evolved RNA clone (R30) obtained in my research (Chapter 2) is

useful for this purpose because it showed lower ribosome requirement and can replicate at low ribosome concentration (200 nM). Future experimental evolution from the R30 clone should further reduce the ribosome requirement and approaching the development and application of handy and less expensive TcRR systems. Moreover, in the Chapter 3, I found that the RNA genome improved its TcRR activity in the absence of translation initiation- or termination-related proteins. By combining the finding in the Chapter 2 that the self-improved TcRR activity under reduced ribosome condition, the results indicated that the requirements of most of the translation proteins for the TcRR reaction were reduced. Further experimental evolution using the evolved RNAs may achieve complete loss of those translation proteins. Because such a system has lower complexity, the construction cost would decrease and more detailed analysis would become possible, leading to various applications. Furthermore, minimizing the complexity by reducing as translation proteins as possible leads to the construction of a minimal autonomous self-replicating system, in which all the required components are produced in the system. Such system can be considered as a minimal life-form, and thus it would help us to understand what life is. My study may be an appropriate starting point to realize this dream.

The TcCRR system, obtained in the Chapter 4, is the first artificial system in which genome replication and metabolism are autonomously occurred, opening the possibility for application of *in vitro* replication systems. In particular, the TcCRR system could be used for sustained yield of useful metabolic enzymes, including unnatural or toxic enzymes that are lethal to bacteria. Furthermore, the enzyme production can be improved by evolutionary engineering as I showed in my research. The TcCRR system is also the first step toward creating a self-sustained system, in which all the components are supplied by its own activities as living organisms do. To achieve this goal, all the proteins required for the self-replication must be encoded on replicating genomes, and the first step should be to encode multiple genes on genomes in a sustainable manner as I demonstrated. Furthermore, if combined with the findings in the Chapter 2 and 3, the TcCRR system would come closer to the goal because it would require less translation proteins for its sustainable replication. Overall, a series of studies described in this dissertation will be a new stream of research to create a living system from chemistry and push the limits of our understanding of what life is.

5-2. Perspective of Origin-of-Life Research

Constructing a simple life-like system is a standard approach to address problems related to the origin and early evolution of life[3,6,7,8,9,10]. Through the reconstruction of biological functions *in vitro*, we can understand conditions and ingredients required to achieve biology from chemistry. By investigating life-like systems that could have existed in the prebiotic era, such as the RNA world and the RNA-protein world, we can provide insights into a plausible pathway for the origin of living organisms. The TcRR system, which I used in this research, can be considered such a life-like system in the RNA-protein world because it consists of only an RNA genome encoding a single gene (RNA replicase) and a hundred of proteins, and its functions are only genome replication and translation. Needless to say, the system is much simpler than any living things (e.g., *E. coli*, the "simple" model bacterial organism, containing 4000 genes). Although previous studies constructed evolvable RNA replication systems by combining RNAs and proteins, the RNAs did not encode any genes, and enzymes for RNA replication reaction were externally supplied[37,38], which are not a plausible replication mechanism in the RNA-protein world, where enzymes for genome replication were translated[40,41]. In contrast, the TcRR system is the first experimental system in which an RNA genome replicates through the translation of a self-encoded protein necessary for the replication, and hence a model system for the RNA-protein world. My research in the Chapter 2 and 3 showed that such fairly simple TcRR system adapted to severe translation conditions and diversified depending on the impaired steps of translation. This great evolutionary ability indicates that primitive life in the RNA-protein world could have also survived in diverse environments and evolved toward known life. My research also suggested that unfitness in the environments could be overcome by simply changing RNA genome structures, and such a simple mechanism of adaptation implies the robustness of the primitive life-like systems. These ideas evoke a scenario that if the first microbe-like system had originated near hydrothermal vent in the deep sea[99], it would have easily evolved into cold ocean and dry lands, and "life" started to thrive.

The result of the adaptations to various impaired translation conditions suggested that the roles of translation proteins, especially initiation factors and release factors can be partly compensated for by the change in RNA sequences. Such a compensatory mechanism may have worked in ancient translation systems. Therefore further evolutionary research to completely replace the functions of translation proteins by RNAs may lead to elucidating the history of the translation apparatus, one of the biggest questions in the origin of life.

The formation of cooperation between RNA replicators has been considered the first evolutionary transition in the early evolution of life. The question, whether the transition is possible or not in the presence of emerging parasitic RNAs, has been investigated for four decades[44,45,46,47,48,49,50,51,52,53,54]. The large number of theoretical literatures suggested that in primitive replication in compartments,

excess parasite replications and fluctuated RNA distribution at compartment division quickly disrupted the cooperative relationship. However, a large knowledge gap remained due to the lack of empirical demonstrations. To the best of my knowledge, my research in the Chapter 4 provided the first direct empirical evidence for the possible transition using the newly constructed TcCRR system in which mutualistic RNA replicators proliferate and parasitic RNAs spontaneously appeared. The compartmentalized TcCRR system sustained in a certain range of RNA concentrations where parasites were not frequently produced even through random fusion-division of compartments and without regulated segregation mechanism of RNA molecules.

One interesting finding in the study is that Rep-RNA and NDK-RNA co-evolved to be a faster set of replicators, even though each of the evolved replicators showed parasitic behaviors to inhibit the replication of their original partner RNAs, probably by sequestering free replicase enzymes because their replication activities were increased. This result implied that cooperating RNA replicators inherently regulate their evolution not to excessively increase their replication activities so that the cooperation can be maintained for a long time. To prove this idea, I would like to suggest a future experiment that evolving Rep-RNA or NDK-RNA independently with replicase enzymes to demonstrate their replication significantly increases without restriction of coordinating their replications compared with their coupled replication as performed in this study.

It should be noted that sustainable replication observed in this study was achieved by manually regulating concentrations of cooperative RNA replicators, different from living organisms that can self-regulate to survive by overcoming the parasite appearance. Living organisms, especially cells, avoid the parasite problem by synchronizing genome replication and cell growth so that they can keep the number of their genome per cell approximately one. This facts suggest that continuing the evolution experiment with the TcCRR system may innovate a new mechanism to adjust the replication to the timing of fusion-division of the compartments, leading to the origin of the small number of genomes, common principles in cells.

On the other hand, the sustained population of the mutualists contain many mutants that showed impaired cooperative behaviors (i.e., parasitic behavior). This fact indicated that the complete exclusion of parasitic RNAs from the TcCRR system is not necessary to achieve the sustainable replication, and cooperators can coexist with parasitic mutants if they do not replicate too fast. This phenomenon is similar to stable host-parasite or predator-prey coexistence as seen in mathematical models and natural systems[100,101,102,103]. Therefore the detailed analysis of the simple TcCRR system is expected to lead to understanding conditions allowing the coexistence. Furthermore, the ratio of the mutants with impaired cooperative behavior was much lower for the NDK-RNA clones than the Rep-RNA clones. One possible reason for the robust NDK-RNA property to retain the enzyme activity is that NDK forms multimeric structures. This is because an inactive NDK monomer may fold with active monomers expressed form other NDK-RNAs in the same

compartments and disrupt the whole complex. This idea proposes a new testable hypothesis that multimeric proteins might have been preferentially selected in primitive replication system with redundant genes in compartments. Further studies could confirm this theory and provide a plausible scenario of the early evolution of proteins.

Furthermore, the TcCRR system is a powerful foundation to advance research on the origin and early evolution of life. For example, I can address the error threshold problem, a long-standing problem for primitive error-prone replication systems: a replication introduced too many errors on a long genome and immediately disrupt the system[46,93]. The separation of genes into distinct replicators was considered a possible solution but not empirically tested. By comparing individual cooperators with linked ones, I could understand efficacy of disjointed replicators. Similarly, I could find the conditions in which the linked one is selected and elucidate the origin of chromosome structures[95,96].

Another line of research is to address the origin of obligate mutualisms, a common event that recurred in evolutionary transitions and increased complexity of life[32]. How cooperative molecular "self-" replicators are ratcheting up to obligate mutualists has been investigated only theoretically[104]. By externally supplying some replicase or CTP to the TcCRR system and controlling the benefits from the mutualists, I might be able to reveal a threshold required to form the obligate symbiosis in the very beginning of life.

Appendix

A2-1. The Sequence of the Original RNA Clone

GGGAACCCCCCTTCGGGGGGTCACCTCGCGCAGCGGGCTACGCGAGGGAGCCACGCTG
CGAAGCAGCGTGGCGGTTCTCGTGCGTCACCGAAACGCACGAAGGTCGCGCCTCTTCAC
GAGGCGTCACCTGGGAGAGCGCGAAAGCGCTAGCCCGTGCTCTAGCTCTAGAAGGTCTC
GAGATCTCCTCTAGAGATAATTTTGTTTAACTCTAAGAAGGAGATATACACATGCCTAAGA
CAGCATCTTCGCGTAACTCTCTCAGCGCACAATTGCGCCGAGCCGCGAACACAAGAATT
GAGGCTGAAGGTAACCTCGCACTTTCCATTGCCAACGATTTACTGTTGGCCTATGGTCAG
TCGCCATTTAACTCTGAGGCTGAGTGTATTTCATTCAGCCCGAGATTCGACGGGACCCCG
GATGACTTTAGGATAAATTATCTTAAAGCCGAGATCATGTCGAAGTATGACGACTTCAGCC
TAGGTATTGATACCGAAGCTGTTGCCTGGGAGAAGTTCCTGGCAGCAGAGGCTGAATGT
GCTTTAACGAACGCTCGTCTCTATAGGCCTGACTACAGTGAGGATTTCAATTTCTCACTGG
GCGAGTCATGTATACACATGGCTCGTAGAAAAATAGCCAAGCTAATAGGAGATGTTCCGT
CCGTTGAGGATATGTTGCGTCACTGCCGATTTTCTGGCGGTGCTACAACAACGAATAACC
GTTCGTACAGTCATCCGTCCTTCAAGTTTGCACTTCCGCAAGCGTGTACGCCTCGGGCTT
TGAAGTATGTTTTAGCTCTCAGAGCTTCTACACATTTCGATATCAGAATTTCTGATATTAGC
CCTTTTAATAAAGCAGTTACCGTACCTAAGAACAGTAAGACAGATCGTTGTATTGCTATCG
AACCTGGTTGGAATATGTTTTTCCAACTGGGTATCGGTGGCATTCTACGCGATCGGTTGCG
TTGCTGGGGTATCGATCTGAATGATCAGACGATAAATCAGCGCCGCGCTCACGAAGGCTC
CGTTACTAATAACTTAGCAACGGTTGATCTCTCAGCGGCAAGCGATTCTATGTCTCTTGCC
CTCTGTGAGCTCTTATTGCCCCGAGGCTGGTTTGAGGTTCTTATGGACCTCAGATCACCTA
AGGGGCGATTGCCTGACGGTAGTGTTGTTACCTACGAGAAGATTTCTTCTATGGGTAACG
GTTACACATTCGAGCTCGAGTCGCTTATTTTTGCTTCTCTCGCTCGTTCCGTTTGTGAGAT
ACTGGACTTAGACTCGTCTGAGGTCACTGTTTACGGAGACGATATTATTTTACCGTCCCGT
GCAGTCCCTGCCCTCCGGGAAGTTTTTAAGTATGTTGGTTTTACGACCAATACTAAAAAG
ACTTTTTCCGAGGGGCCGTTCAGAGAGTCGTGCGGCAAGCACTACTATTCTGGCGTAGAT
GTTACTCCCTTTTACATACGTCACCGTATAGTGAGTCCTGCCGATTTAATACTGGTTTTGAA
TAACCTATATCGGTGGGCCACTATTGACGGCGTATGGGATCCTAGGGCCCATTCTGTGTAC
CTCAAGTATCGTAAGTTGCTGCCTAAACAGCTGCAACATAATACTATACCTGACGGTTACG
GTGATGGTGCCCTCGTCGGATCGGTCCTAATCAATCCTTTCGCGAAAAACCGCGGGTGGA
TCCGGTACGTACCGGTGATTACGGACCATACAAGGGACCAAGAGCGCGCTGAGTTGGGG
TCGTATCTCTACGACCTCTTCTCGCGTTGTCTCTCGGAAAGTAACGATGGGTTGCCTCTTA
GGGGTCCATCGGGTTGCGATTTTGCTGATCTATTTGCCATCGATCAGCTTATCTGTAGGAG
TGATCCTACGAAGATAAGCAGGCCTACCGGTAAATTCGATATACAGTACATCGCGTGCTGT
TGTTCGGGTTGTTGTTAGGCTTGCGGCCGCACTCGAGAGATCTAGAGCATCACGGTCGAA
CTCCCGTACGAGGTGCCCGCACCTCGTCCCCCCCTTCCGGGGGGGTCCCC

A2-2. The Sequence of VectorF

TCGAACTCCCGTACGAGGTGCCCGCACCTCGTCCCCCCCTTCCGGGGGGGTCCCCGGGG
ATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTAGCTTGGCGTAATCATGGTCATAGCTG
TTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATA
AAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCA
CTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGC
GCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTG
CGCTCGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTA
TCCACAGAATCAGGGGATAACGCAGGAAAGAACATGTGAGCAAAAGGCCAGCAAAAGG
CCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGAC
GAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAA
GATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGC
TTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCAATGCTCAC
GCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAAC
CCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGG
TAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGG
TATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGG
ACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGC
TCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAG
ATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGAC
GCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATC
TTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTA
AACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTA
TTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGATAACTACGATACGGGAGGGCT
TACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATT
TATCAGCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTA
TCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTA
ATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTG
GTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTT
GTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCG
CAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTA
AGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGG
CGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGCGCCACATAGCAGAACT
TTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCG
CTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTA
CTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGG

AATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGC
ATTTATCAGGGTTATTGTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACA
AATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGTCTAAGAAACCATTAT
TATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCTCGCGCGTTTC
GGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCT
GTAAGCGGATGCCGGGAGCAGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGG
TGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGTGCACCATATA
TGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCG
CCATTCAGGCTGCGCAACTGTTGGGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGC
CAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTC
CCAGTCACGACGTTGTAAAACGACGGCCAGTGAATTCTAATACGACTCACTATAGGGAAC
CCCCCTTCGGGGGGTCACCTCGCGCAGCGGGCTGCGCGAAGGAGCCACGCTGCGAAGC
AGCGTGGCGGTTCTCGTGCGTCACCGAAACGCACGAAGGTCGCGCCTCTTCACGAGGCG
TCACCTGGGAGAGCGCGAAAGCGCTAGCCCGTGCTC

A2-3. Mutation List of the Evolved Clones (round 11 (R11), 15 (R15), and 30 (R30))

		R11-								R15-								R30-							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
A169G	5'UTR															+									
C182T	5'UTR							+				+	+	+	+		+	+	+		+			+	+
C184A	5'UTR			+	+				+	+	+							+	+	+	+	+	+	+	+
C225T	5'UTR																					+	+		
C238T	T4I		+																						
T246C	S7P	+																							
C263T	S12S																				+				
A285G	N20D						+																		
A288T	T21S	+																							
C332T	N35N									+															
G341A	L38L									+															
A355T	Q43L							+				+	+	+	+		+	+	+		+	+	+	+	+
G356A	Q43Q																			+					
T371C	S48S																	+							
T381C	C52R													+											
T435C	Y70H																	+	+		+	+	+	+	+
T462C	Y79H	+																							
A483G	I86V														+										
insertion 630-1	-												+												
G660A	V145I							+			+	+	+	+	+		+	+	+		+	+	+	+	+
T662G	V145V																			+					
T701C	A158A						+																		
C721T	S165L										+	+	+	+	+		+	+	+		+	+	+	+	+
C750T	L175F	+								+					+	+									
C810T	H195Y				+																				
C821T	I198I																				+				
A825G	I200V							+			+	+	+	+			+	+	+		+	+	+	+	+
T845C	F206F							+			+	+	+	+			+	+	+	+	+	+	+	+	+
T922C	F232S														+										
C945A	L240I															+									
T1010A	A261A																			+					
C1022T	S265S						+																		
G1023A	V266I																					+	+		
A1032G	N269D		+																						
A1040G	A271A							+												+					
T1045A	V273D		+																						
A1055G	S276S										+	+	+	+			+	+			+	+	+		
deletion A1459	-								+																
C1085A	L286L														+										
A1117C	E297A					+																			
C1155T	P310S		+																						
G1158A	D311N																				+				
A1164G	S313G					+																			
C1174T	T316I																				+				
T1188C	S321P										+														
T1193C	S322S								+																
C1220T	L331L																							+	
C1225T	S333L															+									
T1229C	L334L					+	+													+	+				
T1272A	L349I						+																		
C1282A	S352Y							+				+	+	+				+	+	+	+	+	+	+	+
C1298T	Y357Y																					+	+		
deletion C1323	-									+															
G1324A	R366H											+													
deletion T1392	-											+													
A1411T	E395V							+																	
C1464T	R413C		+																						
C1469T	H414H	+																							

A2-3. Mutation List of the Evolved Clones (continue)

		R11-								R15-								R30-							
		1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
T1498G	L424R									+															
G1539A	V438I																	+							
C1578T	R451C																				+	+	+		
C1582T	Q459Stop						+																		
A1603G	Q459R																		+					+	+
C1612T	T462I		+			+											+								
C1679A	N484K										+														
A1719G	T498A																	+							
A1729G	Q501R																		+		+	+	+	+	+
C1738A	A504D							+	+		+	+		+				+	+	+	+	+	+	+	+
G1761T	D512Y																			+					
T1768C	F514S	+																							
T1782C	S519P														+	+									
T1790C	S521S		+																						
T1782C	S519P	+					+																		
C1784A	S519S																	+							
A1791G	N522D																				+				
G1794A	D523N																			+					
A1809G	R528G																		+			+	+	+	+
G1813A	G529D															+									
G1825A	C533Y	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
A1855G	Q543R	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
T1871C	S548S							+	+			+		+				+	+	+	+	+	+	+	+
A1890C	R555R																			+					
G1892A	R555R											+			+										
C1907T	F560F																	+			+				
G1957A	3'UTR		+																						
G1960A	3'UTR	+																+			+				
G1966A	3'UTR												+												
C1973A	3'UTR													+				+	+	+	+	+	+	+	+
G1978A	3'UTR									+							+								

A2-4. Replication Activity of the Clones under Uncoupled Conditions

The replication rate constant (k_{rep}) of the obtained clones (2-3-4) was estimated using a mathematical model (2-2-9). I therefore confirmed these results by experiments of TcRR reactions under uncoupled conditions. At 200 nM ribosome concentrations, I started only translation of each plus-strand RNA clone in the absence of UTP, then I stopped translation and started replication by adding streptomycin and UTP. I normalized the replicated RNA concentrations by translated replicase concentrations to obtain relative replication constants (Fig. A2-1).

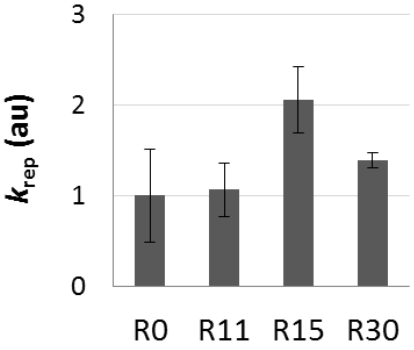


Figure A2-1. Replication rate constants estimated by the uncoupled experiment.

At 200 nM ribosome concentrations, plus-strand RNA clones were translated in the absence of UTP not to initiate replication at at 37°C for 1 hour. Next, the replication was started by adding streptomycin and UTP at 37°C for 30 min. The synthesized minus-strand RNA concentration was measured by quantitative RT-PCR. The RNA concentration was normalized by the translated replicase concentration obtained in the previous experiments (2-3-4) to obtain replication rate constants.

A2-5. Translation and Replication Activities at High Ribosome Concentrations

In (2-3-4), I demonstrated that the RNA genome evolved to improve translation and self-replication activities at the low ribosome concentrations (200 nM or 50 nM). In this section, I also examined whether the evolved RNAs also show improved activities at the higher ribosome concentrations (1 μM, optimum concentration for the original RNA genome (R0)). I analyzed rate constants $k_{trans}k_{rep}$, k_{trans}, and k_{rep} in the same manner as described in (2-3-4) yet at 1 μM ribosome concentration (Fig. A2-2). The self-replication rate constant ($k_{trans}k_{rep}$) were not significantly changed during the rounds (Fig. A2-2B). On the other hand, translation rate constant (k_{trans}) increased from the original clone (Fig. A2-2C), and in contrast, replication rate constant dropped (Fig. A2-2D). Taken together, the self-replication activity at the higher ribosome concentration was not changed even though the translation activity was increased because the increase was offset by the decreased replication activity.

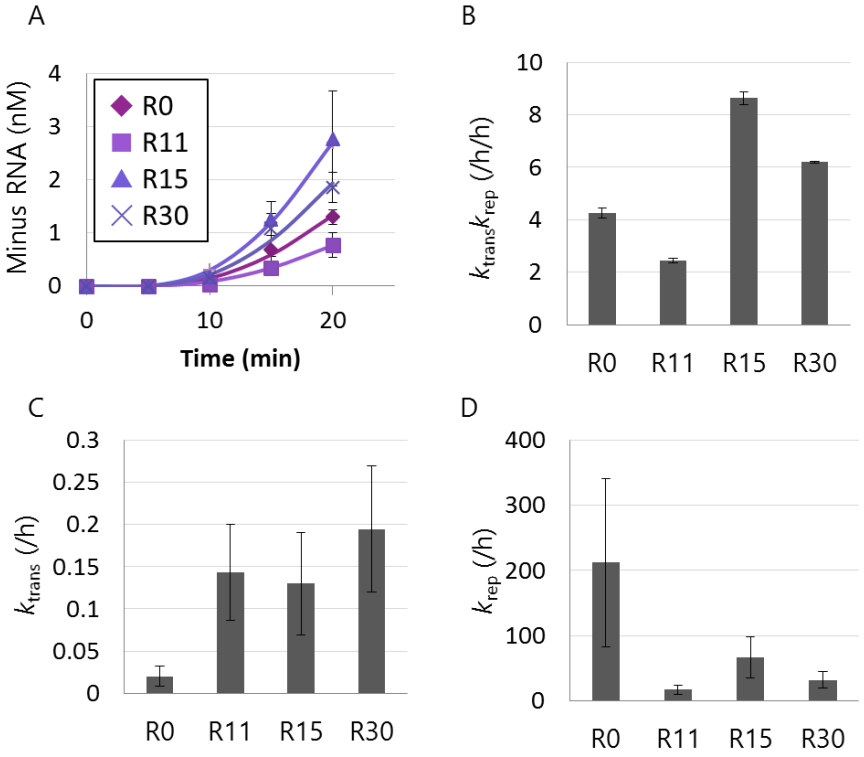


Figure A2-2. Kinetic parameters at high ribosome concentration.

(A) Minus-strand RNA synthesis through TcRR reaction of 10 nM plus-strand RNA clones at 1 μM ribosome in bulk condition. The plots were fitted with the (eq. 4). The lag-time for unreacted phase was set 5 minutes. (B) Translation and replication rate constants ($k_{trans}k_{rep}$), estimated by the curve fit of A. (C) Translation rate constants (k_{trans}). 100 nM plus-RNA clones were translated at 1 μM ribosome and the constants were calculated with the (eq. 2). (D) Replication rate constants (k_{rep}), calculated from obtained $k_{trans}k_{rep}$ and k_{trans}. The error bars show standard deviation.

The above result implied the existence of trade-off between translation and replication reactions. One possible mechanism is that the relaxed SD sequence of the evolved clones recruits too much ribosome to the RNA (Fig. A2-3). Since replication and translation simultaneously occur on the same RNA but toward opposite directions[105], frequent ribosome recruitment may cause serious competition and collision between the ribosome and the replicase on the RNA genome. Such a trade-off has been observed in several RNA viruses[105,106,107] and is a common problem in all systems that have a single-strand RNA as a genome. This phenomenon may prevent further evolution in the TcRR system. For example, the slower evolution from the round 15 to 30 compared with the first 15 rounds (Fig. 2-3) may have resulted from this trade-off. Also, no significant increase in translation efficiency in the previous evolution experiment at high ribosome concentration[62] could be explained by the trade-off. It is therefore important to overcome this trade-off to achieve higher evolutionary ability in the TcRR system. Possible solutions to this problem include introducing a mechanism of switching from translation to replication[106] or non-canonical translation[108] as seen in RNA viruses.

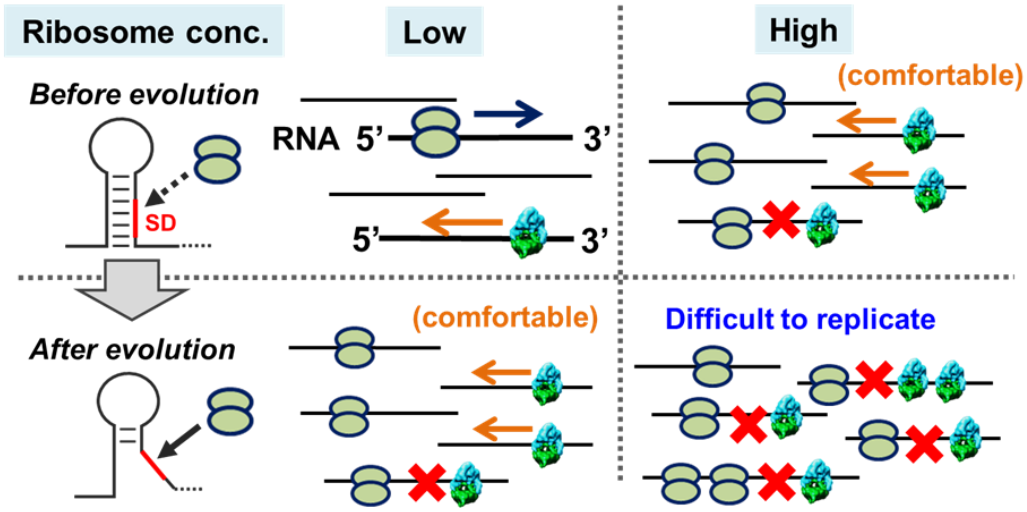


Figure A2-3. Possible mechanism to cause a trade-off between replication and translation
Details are in the main text.

A3-1. Mutation List of the Evolved Clones in the condition A

		A-							
		1	2	3	4	5	6	7	8
G171A	5'UTR			+					
C184A	5'UTR	+	+	+	+	+	+	+	+
A212G	5'UTR	+							
G356A	Q43Q	+				+		+	
G394A	S56N						+		+
T435C	Y70H			+					
A524G	A99A		+						
T566C	P113P						+		+
584-5 insertion	-	+							
T616C	M130T						+		+
G638A	K137K					+			
T662G	V145V	+				+		+	
C709T	T161M			+					
T731C	H168H				+				
T845C	F206F	+	+			+		+	
T953C	D242D					+			
T1010A	A261A	+	+			+		+	
A1015G	E263G	+							
A1040G	A271A							+	
G1100A	L291L						+		+
T1229C	L334L	+	+			+		+	
C1282A	S352Y	+	+	+	+	+	+	+	+
T1484A	P419P						+		+
A1603G	Q459R		+						
C1738A	A504D	+	+	+	+	+	+	+	+
G1772A	S515S		+						
G1774A	R516H		+						
A1788G	S521G						+		+
G1794A	D523N	+	+	+		+		+	
T1824C	C533R			+					
G1825A	C533Y	+	+	+	+	+	+	+	+
T1835C	A536A						+		+
A1855G	Q543R	+	+	+	+	+	+	+	+
T1871C	S548S	+	+	+		+	+	+	+
G1957A	3'UTR						+		+
C1961T	3'UTR					+			
C1973A	3'UTR	+	+	+		+		+	

A3-2. Mutation List of the Evolved Clones in the condition B

		B-							
		1	2	3	4	5	6	7	8
A167G	5'UTR						+		
G171A	5'UTR	+							
C182T	5'UTR				+				
C184A	5'UTR	+	+	+		+	+	+	+
G279A	A18T					+			
T302C	A25A						+		
G303A	E26K						+		
C323T	S32S								+
C339T	L38L					+			
A355T	Q43L				+				
T365C	F46F								+
C368A	N47K								+
T390C	F55L				+				
T435C	Y70H	+							
A511G	K95R						+		
G663A	E146K						+		
C709T	T161M	+							
C750T	L175F						+		
A791T	L188F						+		
A809G	T194T			+					
T848C	N207N			+					
A863G	V212V				+				
T896C	A223A				+				
C1002T	R259C							+	
T1046C	V273V				+				
C1104 deletion	-								+
C1149A	R308R					+			
G1165A	S313N				+				
C1282A	S352Y	+	+	+	+	+		+	+
A1346G	E373E						+		
A1346T	E373D								+
C1402T	P392L						+		
C1439T	G404G				+				
C1460T	Y411Y					+			
G1533A	D436N			+					
A1603G	Q459R						+		
T1613A	T462T			+					

		B-							
		1	2	3	4	5	6	7	8
C1639A	A471D							+	
A1700G	V491V							+	
A1710G	T495A		+						
C1738A	A504D	+	+	+	+	+	+	+	+
G1751A	S508S				+				
G1794A	D523N	+	+	+	+			+	+
T1796C	D523D					+			
G1799A	G524G						+		
A1817G	P530P		+						
T1824C	C533R	+							
G1825A	C533Y	+	+	+	+	+	+	+	+
G1851A	D542N							+	
A1855G	Q543R	+	+	+	+	+	+	+	+
T1871C	S548S	+	+	+	+	+	+	+	+
G1923A	A566T		+						
G1942A	C572Y							+	
T1964C	3'UTR		+						
C1973A	3'UTR	+	+	+	+			+	+
T1981C	3'UTR				+				

A3-3. Mutation List of the Evolved Clones in the condition C

		C-							
		1	2	3	4	5	6	7	8
A167G	5'UTR				+				
C184A	5'UTR	+	+	+	+	+	+	+	+
G201 deletion	5'UTR	+		+					
T202 deletion	5'UTR	+							
C227T	5'UTR						+		
G264A	A13T			+					
G300A	A25T					+			
T302C	A25A				+				
G420T	D65Y		+						
A494G	E89E					+			
A508G	E94G						+		
T566C	P113P						+	+	+
A615G	M130V	+							
T616C	M130T						+	+	
A630 deletic	-			+					
G660A	V145I					+			
T662G	V145V	+		+					
C680T	H151H					+			
T728C	S167S	+		+					
A749G	A174A	+							
G767A	T180T							+	
A798C	R191R								+
T845C	F206F								+
A849G	K208E						+		
A851G	K208K		+						
A863G	V212V							+	
T962G	R245R	+							
C1039T	A271V	+							
T1046C	V273V								+
T1049C	D274D		+			+			
T1076C	S283S		+						
C1092T	L289F							+	
T1095C	L290L								+
G1100A	L291L				+		+	+	
T1124C	L299L								+

		C-							
		1	2	3	4	5	6	7	8
G1167A	V314I				+		+	+	
C1178T	Y317Y					+			
T1238C	A337A		+						
A1270G	D348G				+		+	+	
C1277T	D350D								+
C1282A	S352Y	+		+					
1458-9 insertion	-					+			
T1501G	V425G			+					
T1521C	W432R				+				
C1538T	G437G						+		
A1558G	H444R						+		
A1603G	Q459R	+	+	+	+			+	+
C1617T	P464S		+						
C1622T	D465D	+							
G1652A	S475S	+							+
G1706T	V493V			+					
C1738A	A504D	+	+	+	+	+	+	+	+
G1751A	S508S					+	+	+	
T1796C	D523D	+		+				+	+
C1815A	P530T								+
G1825A	C533Y	+	+	+	+	+	+	+	+
A1855G	Q543R	+	+	+	+	+	+	+	+
G1870A	S548N				+	+			
T1871C	S548S	+	+	+			+	+	+
G1957A	3'UTR				+	+			
G1960A	3'UTR								+
A1962G	3'UTR	+							
T1964A	3'UTR			+				+	

A3-4. Mutation List of the Evolved Clones in the condition D

		D-							
		1	2	3	4	5	6	7	8
T175C	5'UTR				+				
C182T	5'UTR	+	+	+	+	+	+	+	+
A211G	5'UTR								+
A212G	5'UTR		+						
C256T	S10F			+					
A304G	E26G						+		
G344A	L39L					+			
A355T	Q43L	+	+	+	+	+	+	+	+
T377C	A50A								+
A609G	I128V			+					
C614T	H129H						+		
T652C	V142A	+							
G660A	V145I	+	+	+	+	+	+	+	+
G663A	E146K	+							
C721T	S165L	+	+	+	+	+	+	+	+
C725T	Y166Y					+			
C750T	L175F	+	+	+		+			+
T770G	P181P								+
814-5 insertion	-							+	
A825G	I200V	+	+	+	+	+	+	+	+
T845C	F206F	+		+	+	+	+	+	+
A854G	A209A				+	+		+	
C886A	R220H						+		
T957C	L244L						+		
A1012G	H262R				+				
C1013T	H262H			+					+
G1017T	G264C				+				
C1052T	L275L					+			
A1055G	S276S	+	+		+	+	+	+	
G1073A	M282I					+			
G1100A	L291L		+						
A1151G	R308R					+			
A1211C	T328T			+					
C1282A	S352Y					+			
A1301G	G358G			+					
G1319A	P364P		+						
A1384G	K386R							+	

		D-							
		1	2	3	4	5	6	7	8
C1427T	H400H		+						
A1461G	I412V								+
G1537A	G437D							+	
A1582G	K452R							+	
C1612T	T462I		+	+	+	+	+		
A1649G	G474G		+			+	+		
A1729G	Q501R								+
C1738A	A504D	+		+	+			+	+
T1801C	L525S			+					
G1825A	C533Y	+	+	+	+	+	+	+	+
G1845A	A540T	+							
A1855G	Q543L	+	+	+	+	+	+	+	+
C1862T	I545M			+					
T1871C	S548S	+			+	+		+	+
C1875T	P550S					+			
A1881G	K552E							+	
G1957A	3'UTR								+
C1973A	3'UTR	+			+				
C1978A	3'UTR		+	+			+	+	
T1981C	3'UTR			+					

A4-1. The Sequence of the Original NDK-RNA

GGGAACCCCCCTTCGGGGGGTCACCTCGCGCAGCGGGCTACGCGAGGGAGCCACGCTG
CGAAGCAGCGTGGCGGTTCTCGTGCGTCACCGAAACGCACGAAGGTCGCGCCTCTTCAC
GAGGCGTCACCTGGGAGAGCGCGAAAGCGCTAGCCCGTGCTCTAGCTCTAGAAGGTCTC
GAGATCTCCTCTAGAGATAATTTTGTTTAACTCTAAGAAGGAGATATACACATGGCTATTG
AACGTACTTTTTCCATCATCAAACCGAACGCGGTAGCAAAAAACGTCATTGGTAATATCT
TTGCGCGCTTTGAAGCTGCAGGGTTCAAAATTGTTGGCACCAAAATGCTGCACCTGACC
GTTGAACAGGCACGTGGCTTTTATGCTGAACACGATGGAAAACCGTTCTTTGATGGTCTG
GTTGAATTCATGACCTCTGGCCCGATCGTGGTTTCCGTGCTGGAAGGTGAAAACGCCGTT
CAGCGTCACCGCGATCTGCTGGGCGCGACCAATCCGGCAAACGCACTGGCTGGTACTCT
GCGCGCTGATTACGCTGACAGCCTGACCGAAAACGGTACCCACGGTTCTGATTCCGTCG
AATCTGCCGCTCGCGAAATCGCTTATTTCTTTGGCGAAGGCGAAGTGTGCCCGCGCACCC
GTTAGGCTTGCGGCCGCACTCGAGAGATCTAGAGCATCACGGTCGAACTCCCGTACGAG
GTGCCCGCACCTCGTCCCCCCCTTCCGGGGGGGTCCCC

A4-2. Mutation List for Improvement of NDK-RNA Translation and Replication Activities

All mutations are synonymous or located at untranslated regions.

*Mutations at the same sites.

Translation improvement

C182A
C184A
G219-220 (insertion)
} Original

Replication improvement

Original → mod2		mod2 → mod 2-CE	mod2-CE → mod2-CE-X
C258U	C426U	G179U	C496U
G264A	C453U	A576U*	C501U
C267U	U483C	C603A	G504U
G270A	C519U	C609U	U534C
C282U	A522G	U618C	G537A
C285U	C549U	U621C	C540G
U294C	G561U	U624C*	U546C
C297U	C576A*	C630A	U552A
G303U	U585C	C636U	
C306U	C591U	G681C	
C324U	C594U	A691U	
C339U	U600G	U692A	
C357U	C624U*		
A363G	A639G		
A369U	C651A		
U420C	C662U		

A4-3. Mutation List of the Evolved Rep-RNA Clones

		R-																															
		1	3	4	5	6	7	8	10	11	14	15	17	18	19	20	21	22	23	24	25	27	29	30	31	32	33	35	36	37	38	39	40
C54U	5'UTR																+		+														
A62C	5'UTR	+																															
A92G	5'UTR																											+			+		
A93G	5'UTR																						+		+	+	+						
C105A	5'UTR																																+
U114C	5'UTR								+																								
U124C	5'UTR																													+			
A161G	5'UTR									+																							
U164C	5'UTR																											+			+		
U166C	5'UTR																						+		+	+	+						
A184U	5'UTR					+			+					+			+		+			+								+			
U197C	5'UTR																	+															
insertion 200-1 U	5'UTR											+																					
U202G	5'UTR																	+															
del U204	5'UTR																												+				
A206G	5'UTR	+	+		+					+	+	+	+								+			+				+				+	+
A211G	5'UTR			+																													
A214G	5'UTR														+																		
A224G	5'UTR																+																
A224C	5'UTR																	+											+				
A226G	5'UTR															+																	
A239G	T4T														+																		
A288G	T21A				+																												
U296C	I23I																		+														
A305G	E26E							+																									
U326C	I33I					+																											
U348C	Y41H	+																															
A355G	Q43R														+																		
U357C	S44P					+																							+				
A367C	N47T						+																										
A367G	N47G																					+											
U386C	I53I								+											+													
U390C	F55L																						+		+	+	+				+		
A401G	R58R																						+		+	+	+						
U419C	D64D																													+			
C467U	D80D						+																										
U471C	F82L																							+									
C550U	A108V									+																							
C564U	P113S			+																													
C572G	Y115Stop										+																						
U620C	A131A						+																										
deletion A630	-															+																	
A636G	K137E									+																							
U652C	V142A																														+		
U657C	S144P												+		+																		
C678U	H151Y																						+										
A714G	N163D																													+			
C729U	H168Y																												+				
A730G	H168H		+																														
U744C	F173L																																+
U746C	F173F	+	+		+					+	+	+												+				+				+	+
A757G	Q177R																							+									
U776C	A183A																											+					
U778C	L184S					+																											
U783C	Y186H								+																								
U789C	L188L												+																				
U812C	H195H							+																									
A825G	I200V															+																	
A851G	K208K																													+			
A854G	A209A														+																		
C923U	F232F																									+	+						
C954U	R243W													+																			
G968U	W247C																+																
G975A	D250N	+	+		+					+	+	+	+								+			+	+			+				+	+

A4-3. Mutation List of the Evolved Rep-RNA Clones (Continue)

		R-																													
		1	3	4	5	6	7	8	10	11	14	15	17	18	19	20	21	22	23	24	25	27	29	30	31	32	33	35	36	37	38	39	40
U998C	N257N																+																
A1016G	E263E	+																															
A1033G	N269S																			+													
C1057U	A277V																		+														
A1071G	M282V			+																													
U1157C	P310P						+																										
U1166C	S313S			+																													
U1187C	I320I																									+							
C1208U	Y327Y					+																											
U1224C	S333P	+																															
U1232C	I335I																												+				
insertion 1235-6 UU	-											+																					
U1247C	A340A							+																									
U1278C	S351P					+																											
G1280A	S351S							+																									
G1287A	V354I						+																										
insertion 1352-3U	-																					+											
A1353U	K376Stop				+																												
U1361C	V378V																				+												
U1405C	F393S																+																
A1423G	K399R												+																				
G1437A	G404S						+																										
U1503C	L426L																											+					
U1521C	W432R														+																		
U1550C	P441P																						+		+	+	+						
G1554A	A443T																	+															
U1560C	S445P				+																												
U1562C	S445S							+																									
A1573G	K449R	+																															
A1597G	Q457R																				+												
U1637C	G470G																															+	+
A1658G	L477L				+																												
A1678U	N484I																						+										
insertion 1678-9A	-							+																		+							
U1718C	H497H						+																										
G1751A	S508S													+																			
U1775C	R516R																	+															
U1780C	L518P																								+								
G1798A	G524E												+																				
insertion 1813-4G	-																					+											
A1817G	P530P																								+		+						
U1824C	Y533H						+																										
U1832C	F535F				+																			+									
C1850U	I541I																									+							
A1884U	I553L																														+		
G1891C	R555T				+			+													+						+						+
U1912C	I562T																								+								
U1943G	C572W																						+										
U1994C	C589R																				+												
U2005C	3'UTR											+									+												
C2007U	3'UTR																	+															

A4-4. Mutation List of the Evolved NDK-RNA Clones

		N-																														
		1	2	4	5	7	8	9	10	12	13	17	18	19	20	22	23	24	25	26	27	29	31	32	33	34	35	38	39	40	41	42	43
C59G	5'UTR																	+															
G84A	5'UTR												+																				
C86U	5'UTR	+	+		+	+	+					+	+		+		+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+
G90A	5'UTR									+																							
C157U	5'UTR																													+			
U164C	5'UTR																						+										
A167G	5'UTR							+																									
G171U	5'UTR														+																		
A178G	5'UTR																													+			
U181C	5'UTR			+																													
U183C	5'UTR			+																													
U188C	5'UTR								+																								
A193U	5'UTR															+								+									
deletion U200	5'UTR																								+								
A227G	5'UTR																												+				
U234G	A2A	+	+	+	+	+	+	+	+	+	+	+	+	+	+		+	+	+	+	+	+	+		+	+	+	+	+	+	+	+	+
insertion 250-1 UU	-																					+											
A256G	I10V									+																							
A279G	K17K									+																							
G289A	G21S														+																		
U315G	A29A															+																	
U333C	V35V																	+															
A350G	H41R													+																			
C383U	A52V																												+				
A387G	E53E									+																							
C390U	H54H					+																											
A462G	E78E														+																		
U465C	G79G							+	+												+												
U477C	V83V																					+											
G482A	R85H													+																			
U510C	N94N												+																				
G513A	P95P																						+										
deletion A518	-																					+											
U524C	L99P			+										+																			
A537G	L103L									+																							
G550U	A108S															+								+									
U560C	L111P					+																											
A567G	E113E																		+														
A611G	E128G																		+														
A612G	E128E															+								+									
C615U	I129I															+								+									
A620G	Y131C									+																							
G644A	C139Y																					+											
A651G	R141R																						+										

Reference

1. Benner, S. A. & Sismour, A. M. Synthetic biology. *Nat. Rev. Genet.* **6,** 533–543 (2005).
2. Forster, A. C. & Church, G. M. Towards synthesis of a minimal cell. *Mol. Syst. Biol.* **2,** (2006).
3. Jewett, M. C. & Forster, A. C. Update on designing and building minimal cells. *Curr. Opin. Biotechnol.* **21,** 697–703 (2010).
4. Forlin, M., Lentini, R. & Mansy, S. S. Cellular imitations. *Curr. Opin. Chem. Biol.* **16,** 586–592 (2012).
5. Pohorille, A. & Deamer, D. Artificial cells: Prospects for biotechnology. *Trends Biotechnol.* **20,** 123–128 (2002).
6. Szostak, J. W. *et al.* Synthesizing life. *Nature* **409,** 387–390 (2001).
7. Hanczyc, M. M., Fujikawa, S. M. & Szostak, J. W. Experimental models of primitive cellular compartments: encapsulation, growth, and division. *Science (New York, N.Y.)* **302,** 618–22 (2003).
8. Ichihashi, N. *et al.* Constructing partial models of cells. *Cold Spring Harb. Perspect. Biol.* **2,** 1–10 (2010).
9. Stano, P. & Luisi, P. L. Semi-synthetic minimal cells: Origin and recent developments. *Curr. Opin. Biotechnol.* **24,** 633–638 (2013).
10. Ichihashi, N. & Yomo, T. Constructive Approaches for Understanding the Origin of Self-Replication and Evolution. *Life* **6,** 26 (2016).
11. Riley, M. *et al.* Escherichia coli K-12: A cooperatively developed annotation snapshot - 2005. *Nucleic Acids Res.* **34,** 1–9 (2006).
12. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19,** 751–5 (2001).
13. Zhou, Y., Asahara, H., Gaucher, E. A. & Chong, S. Reconstitution of translation from Thermus thermophilus reveals a minimal set of components sufficient for protein synthesis at high temperatures and functional conservation of modern and ancient translation components. *Nucleic Acids Res.* **40,** 7932–7945 (2012).
14. Ishikawa, K., Sato, K., Shima, Y., Urabe, I. & Yomo, T. Expression of a cascading genetic network within liposomes. *FEBS Lett.* **576,** 387–390 (2004).
15. Murtas, G., Kuruma, Y., Bianchini, P., Diaspro, A. & Luisi, P. L. Protein synthesis in liposomes with a minimal set of enzymes. *Biochem. Biophys. Res. Commun.* **363,** 12–17 (2007).
16. Langer, R. Drug delivery and targeting. *Nature* **392,** 5–10 (1998).

17. Chang, T. M. S. Future prospects for artificial blood. *Trends Biotechnol.* **17,** 61–67 (1999).

18. Schrum, J. P., Zhu, T. F. & Szostak, J. W. The Origins of Cellular Life. *Cold Spring Harb. Perspect. Biol.* **2,** a002212–a002212 (2010).

19. Monnard, P. A. & Deamer, D. W. Membrane self-Assembly processes: Steps toward the first cellular life. *Minimal Cell Biophys. Cell Compart. Orig. Cell Funct.* **207,** 123–151 (2011).

20. Terasawa, H., Nishimura, K., Suzuki, H., Matsuura, T. & Yomo, T. Coupling of the fusion and budding of giant phospholipid vesicles containing macromolecules. *Proc. Natl. Acad. Sci.* **109,** 5942–5947 (2012).

21. Shiomi, H., Tsuda, S., Suzuki, H. & Yomo, T. Liposome-based liquid handling platform featuring addition, mixing, and aliquoting of femtoliter volumes. *PLoS One* **9,** 5–10 (2014).

22. Bhattacharya, A., Brea, R. J. & Devaraj, N. K. De novo vesicle formation and growth: an integrative approach to artificial cells. *Chem. Sci.* **0,** 1–11 (2017).

23. Lincoln, T. A. & Joyce, G. F. Self-Sustained Replication of an RNA Enzyme. *Science (80-.).* **323,** 1229–1232 (2009).

24. Vaidya, N. *et al.* Spontaneous network formation among cooperative RNA replicators. *Nature* **491,** 72–77 (2012).

25. Mills, D. R., Peterson, R. L. & Spiegelman, S. An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *Proceedings of the National Academy of Sciences of the United States of America* **58,** 217–224 (1967).

26. Wright, M. C. & Joyce, G. F. Continuous in Vitro Evolution of Catalytic Function. *Science (80-.).* **276,** 614–617 (1997).

27. Sakatani, Y., Ichihashi, N., Kazuta, Y. & Yomo, T. A transcription and translation-coupled DNA replication system using rolling-circle replication. *Sci. Rep.* **5,** 10404 (2015).

28. Maynard Smith, J. *The problems of biology*. **144,** (Oxford: Oxford University Press, 1986).

29. Woese, C. R., Kandler, O. & Wheelis, M. L. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc. Natl. Acad. Sci.* **87,** 4576–4579 (1990).

30. Margulis, L. Symbiosis in cell evolution. , 1981. *New York: W.H. Freeman* (1981). doi:9780716770282

31. Szathmáry, E. & Smith, J. M. The major evolutionary transitions. *Nature* **374,** 227–232 (1995).

32. Maynard-Smith, J. & Szathmáry, E. The Major Transitions in Evolution. *Oxford Univ. Press* (1997).

33. McCutcheon, J. P. & Moran, N. A. Extreme genome reduction in symbiotic bacteria. *Nat. Rev.*

Microbiol. **10,** 13–26 (2012).

34. Nowak, M. A. Five Rules for the Evolution of Cooperation. *Science (80-.).* **314,** 1560–1563 (2006).

35. Kita, H. *et al.* Replication of genetic information with self-encoded replicase in liposomes. *Chembiochem : a European journal of chemical biology* **9,** 2403–10 (2008).

36. Ichihashi, N. *et al.* Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. *Nat. Commun.* **4,** (2013).

37. Kramer, F. R., Mills, D. R., Cole, P. E., Nishihara, T. & Spiegelman, S. Evolution in vitro: sequence and phenotype of a mutant RNA resistant to ethidium bromide. *J. Mol. Biol.* **89,** 719–736 (1974).

38. Paegel, B. M. & Joyce, G. F. Microfluidic compartmentalized directed evolution. *Chem. Biol.* **17,** 717–724 (2010).

39. Gilbert, W. The RNA world. *Nature* **319,** 618 (1986).

40. Goldman, A. D., Samudrala, R. & Baross, J. A. The evolution and functional repertoire of translation proteins following the origin of life. *Biol Direct* **5,** 15 (2010).

41. Noller, H. F. Evolution of protein synthesis from an RNA world. *Cold Spring Harb. Perspect. Biol.* **4,** 1–14 (2012).

42. Walter, N. G. & Strunk, G. Strand displacement amplification as an in vitro model for rolling-circle replication: deletion formation and evolution during serial transfer. *Proc Natl Acad Sci U S A* **91,** 7937–7941 (1994).

43. Ellinger, T., Ehricht, R. & McCaskill, J. S. In vitro evolution of molecular cooperation in CATCH, a cooperatively coupled amplification system. *Chem. Biol.* **5,** 729–41 (1998).

44. Higgs, P. G. & Lehman, N. The RNA World: molecular cooperation at the origins of life. *Nat. Rev. Genet.* **16,** 7–17 (2014).

45. Eigen, M. Self organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58,** 465–523 (1971).

46. Eigen, M. & Schuster, P. A principle of natural self-organization - Part A: Emergence of the hypercycle. *Naturwissenschaften* **64,** 541–565 (1977).

47. Takeuchi, N. & Hogeweg, P. Evolutionary dynamics of RNA-like replicator systems: A bioinformatic approach to the origin of life. *Phys. Life Rev.* **9,** 219–263 (2012).

48. Niesert, U., Harnasch, D. & Bresch, C. Origin of life between Scylla and Charybdis. *J Mol Evol* **17,** 348–353 (1981).

49. Szathmáry, E. & Demeter, L. Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128,** 463–486 (1987).

50. Boerlijst, M. C. & Hogeweg, P. Spiral wave structure in pre-biotic evolution: Hypercycles stable against parasites. *Phys. D Nonlinear Phenom.* **48,** 17–28 (1991).

51. Fontanari, J. F., Santos, M. & Szathmáry, E. Coexistence and error propagation in pre-biotic vesicle models: A group selection approach. *J. Theor. Biol.* **239,** 247–256 (2006).

52. Zintzaras, E., Santos, M. & Szathmary, E. Living under the challenge of information decay: the stochastic corrector model vs. hypercycles. *J. Theor. Biol.* **217,** 167–81 (2002).

53. Sardanyés, J. & Solé, R. V. Spatio-temporal dynamics in simple asymmetric hypercycles under weak parasitic coupling. *Phys. D Nonlinear Phenom.* **231,** 116–129 (2007).

54. Takeuchi, N., Kaneko, K. & Hogeweg, P. Evolutionarily stable disequilibrium: Endless dynamics of evolution in a stationary population. *Proc. R. Soc. London. Ser. B, Biol. Sci.* **283,** 20153109 (2016).

55. Bansho, Y. *et al.* Importance of parasite RNA species repression for prolonged translation-coupled RNA self-replication. *Chemistry & biology* **19,** 478–87 (2012).

56. Bansho, Y., Furubayashi, T., Ichihashi, N. & Yomo, T. Host-parasite oscillation dynamics and evolution in a compartmentalized RNA replication system. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 4045–4050 (2016).

57. Takeuchi, N. & Hogeweg, P. Multilevel selection in models of prebiotic evolution II: A direct comparison of compartmentalization and spatial self-organization. *PLoS Comput. Biol.* **5,** (2009).

58. Zintzaras, E., Santos, M. & Szathmáry, E. Selfishness versus functional cooperation in a stochastic protocell model. *J. Theor. Biol.* **267,** 605–613 (2010).

59. Lenski, R. E. & Travisano, M. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 6808–6814 (1994).

60. Kishimoto, T. *et al.* Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution. *PLoS Genet.* **6,** 1–10 (2010).

61. Oz, T. *et al.* Strength of Selection Pressure Is an Important Parameter Contributing to the Complexity of Antibiotic Resistance Evolution. *Mol. Biol. Evol.* **31,** 2387–2401 (2014).

62. Ichihashi, N. *et al.* Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. *Nat. Commun.* **4,** 1–7 (2013).

63. Hosoda, K. *et al.* Kinetic analysis of the entire RNA amplification process by Qβ replicase. *J. Biol. Chem.* **282,** 15516–15527 (2007).

64. McGinnis, J. L., Duncan, C. D. S. & Weeks, K. M. *High-throughput SHAPE and hydroxyl radical analysis of RNA structure and ribonucleoprotein assembly. Methods in enzymology* **468,** (Elsevier Inc., 2009).

65. Karabiber, F., McGinnis, J. L., Favorov, O. V & Weeks, K. M. QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *Rna* **19,** 63–73 (2013).

66. Mortimer, S. a & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129,** 4144–4145 (2007).

67. Chernick, M. R. & Friis, R. H. *Introductory biostatistics for the health sciences: modern applications including bootstrap*. (John Wiley & Sons, 2003).

68. Deigan, K. E., Li, T. W., Mathews, D. H. & Weeks, K. M. Accurate SHAPE-directed RNA structure determination a SHAPE experiment can be interpreted as a pseudo-free energy high accuracy . Free energy minimization , by using SHAPE pseudo-. *Sci. York* **2008,** 97–102 (2008).

69. Hall, M. N., Gabay, J., Débarbouillé, M. & Schwartz, M. A role for mRNA secondary structure in the control of translation initiation. *Nature* **295,** 616–8 (1982).

70. de Smit, M. H. & van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. *Proceedings of the National Academy of Sciences of the United States of America* **87,** 7668–72 (1990).

71. Kudla, G., Murray, A. W., Tollervey, D. & Plotkin, J. B. Coding-sequence determinants of gene expression in Escherichia coli. *Science (New York, N.Y.)* **324,** 255–8 (2009).

72. Na, D. & Lee, D. RBSDesigner: Software for designing synthetic ribosome binding sites that yields a desired level of protein expression. *Bioinformatics* **26,** 2633–2634 (2010).

73. Subramaniam, A. R., Pan, T. & Cluzel, P. Environmental perturbations lift the degeneracy of the genetic code to regulate protein levels in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 2419–24 (2013).

74. Qu, X. *et al.* The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475,** 118–21 (2011).

75. Chen, C. *et al.* Dynamics of translation by single ribosomes through mRNA secondary structures. *Nature structural & molecular biology* **20,** 582–8 (2013).

76. Li, G.-W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484,** 538–41 (2012).

77. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28,** 2731–2739 (2011).

78. Thompson, J. D., Higgins, D. G. & Gibson, T. J. ClustalW: improving the sensitivity of progressive multiple sequence aligment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res. Acids Res* **22,** 4673–4680 (1994).

79. Kozak, M. Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene* **361,** 13–37 (2005).

80. Studer, S. M. & Joseph, S. Unfolding of mRNA Secondary Structure by the Bacterial Translation Initiation Complex. *Mol. Cell* **22,** 105–115 (2006).

81. Poole, E. S., Brown, C. M. & Tate, W. P. The identity of the base following the stop codon determines the efficiency of in vivo translational termination in Escherichia coli. *EMBO J.* **14,** 151–8 (1995).

82. Nicholson, B. L. & White, K. A. Functional long-range RNA-RNA interactions in positive-strand RNA viruses. *Nat. Rev. Microbiol.* **12,** 493–504 (2014).

83. Na, D., Lee, S. & Lee, D. Mathematical modeling of translation initiation for the estimation of its efficiency to computationally design mRNA sequences with desired expression levels in prokaryotes. *BMC Syst. Biol.* **4,** 71 (2010).

84. Espah Borujeni, A., Channarasappa, A. S. & Salis, H. M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res.* **42,** 2646–2659 (2014).

85. Maynard Smith, J. Hypercycles and the origin of life. *Nature* **280,** 445–446 (1979).

86. Sievers, D. & von Kiedrowski, G. Self-replication of complementary nucleotide-based oligomers. *Nature* **369,** 221–4 (1994).

87. Lee, D. H., Severin, K., Yokobayashi, Y. & Ghadiri, M. R. Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature* **390,** 591–594 (1997).

88. Gánti, T. *The principles of life*. (Oxford University Press, 2003).

89. Szathmary, E. The origin of replicators and reproducers. *Philos. Trans. R. Soc. B Biol. Sci.* **361,** 1761–1776 (2006).

90. Kurihara, K. *et al.* A recursive vesicle-based model protocell with a primitive model cell cycle. *Nat. Commun.* **6,** 8352 (2015).

91. Ichihashi, N., Matsuura, T., Hosoda, K. & Yomo, T. Identification of two forms of Qβ replicase with different thermal stabilities but identical RNA replication activity. *J. Biol. Chem.* **285,** 37210–37217 (2010).

92. Usui, K., Ichihashi, N. & Yomo, T. A design principle for a single-stranded RNA genome that

replicates with less double-strand formation. *Nucleic Acids Res.* **43,** 8033–8043 (2015).

93. Eigen, M., Mccaskill, J. & Schuster, P. Molecular Quasi-Species. *J. Phys. Chem.* **92,** 6881–6891 (1988).

94. Lascu, I., Giartosio, S., Ransac, A. & Erent, M. Quaternary structure of nucleoside diphosphate kinases. *J. Bioenerg. Biomembr.* **32,** 227–236 (2000).

95. Maynard Smith, J. M., Szathmáry, E., Smith, J. M. & Szathmáry, E. The origin of chromosomes. I. Selection for linkage. *Journal of theoretical biology* **164,** 437–46 (1993).

96. Santos, M. Origin of chromosomes in response to mutation pressure. *Am. Nat.* **152,** 751–756 (1998).

97. Cooper, T. F. & Lenski, R. E. Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol. Biol.* **10,** 11 (2010).

98. Ketola, T. *et al.* Fluctuating Temperature Leads To Evolution Of Thermal Generalism And Preadaptation To Novel Environments. *Evolution (N. Y).* **67,** 2936–2944 (2013).

99. Martin, W., Baross, J., Kelley, D. & Russell, M. J. Hydrothermal vents and the origin of life. *Nat. Rev. Microbiol.* **6,** 805–814 (2008).

100. Maynard Smith, J. & Slatkin, M. The stability of predator-prey systems. *Ecology* **54,** 384–391 (1973).

101. Schrag, S. J. & Mittler, J. E. Host-Parasite Coexistence : The Role of Spatial Refuges in Stabilizing Bacteria-Phage Interactions IN STABILIZING BACTERIA-PHAGE INTERACTIONS populations for This failure to predict of Nicholson has led to a series of more comple. *Am. Nat.* **148,** 348–377 (1996).

102. Fussmann, G. F. *et al.* Crossing the Hopf Bifurcation in a Live Predator-Prey System Published by : American Association for the Advancement of Science Stable URL : http://www.jstor.org/stable/3078247 Crossing the Hopf Bifurcation in a Live Predator-Prey System. *Science (80-.).* **290,** 1358–1360 (2000).

103. Brockhurst, M. A., Buckling, A. & Rainey, P. B. Spatial heterogeneity and the stability of host-parasite coexistence. *J. Evol. Biol.* **19,** 374–379 (2006).

104. Frank, S. A. The origin of synergistic symbiosis. *J. Theor. Biol.* **176,** 403–410 (1995).

105. Kolakofsky, D. & Weissmann, C. Qβ replicase as repressor of Qb RNA-directed protein synthesis. *BBA Rep.* **246,** 596–599 (1971).

106. Gamarnik, A. V. & Andino, R. Switch from translation to RNA replication in a positive-stranded

RNA virus. *Genes Dev.* **12,** 2293–2304 (1998).

107. Lott, W. B. & Doran, M. R. Do RNA viruses require genome cyclisation for replication? *Trends Biochem. Sci.* **38,** 350–355 (2013).

108. Firth, A. E. & Brierley, I. Non-canonical translation in RNA viruses. *J. Gen. Virol.* **93,** 1385–1409 (2012).

Acknowledgements

Throughout my Ph.D. training, I was fortunate to meet excellent mentors, colleagues, and friends in Osaka University. First and foremost, I would like to express my sincere appreciation to my supervisor in the Symbiotic Network laboratory, Dr. Norikazu Ichihashi, for his dedication to training me for as many as six years. His guidance and passion for science has strongly encouraged me to be a “researcher”. I would also like to express my gratitude to Prof. Fumio Matsuda for his watchful mentorship for almost a year as the head of the laboratory. I am also indebted to my thesis advisors in Bioinformatic Engineering, Graduate School of Information Science and Technology, Prof. Hiroshi Shimizu, Prof. Taro Maeda, Prof. Hideo Matsuda, and Prof. Naoki Wakamiya for spending their precious time to criticizing my works and giving me suggestions to improve my dissertation. I would also like to thank Prof. Tetsuya Yomo at East China Normal University for many valuable discussion on my projects. To my colleagues in the lab, especially Taro Furubayashi and Atsushi Shibai, I could not enjoy life in the laboratory without you. Thank you for your kind support and friendly competition.

I cannot imagine my life during the Ph.D. program without Humanware Innovation Program. Through the innovative program, I had wonderful experiences of an interdisciplinary research. I would like to express my deep appreciation to Prof. Shigeru Kondo for supervising me for my main interdisciplinary research “simple rules to construct a geometric structure by pufferfish”, which was entirely different from my professional research, hence really inspiring. I would also like to show my gratitude to Dr. Miwako Doi at National Institute of Information and Communications Technology for giving me a lot of advice generously for five years. For Dr. Kazufumi Hosoda, the most influential person at the Humanware program to me, your way of science, thinking, and life always motivated me. Huge thanks.

Finally, to my family, I could not submit this dissertation without you. To my father and mother, thank you for giving me this chance to get a Ph.D. Your boundless support and love are always my energy to keep going forward. To my wife, Ami. I have been exhausted many times this past year. I could not move my thumb; I had difficulty in hearing. Yet, thanks to your warm support, I was able to make it. You are my oasis, now and forever.