Doctoral Dissertation
博士論文

In vitro selection and studies on tRNA-recognizing ribozymes derived from T-box motifs

(T-box モチーフを基にした tRNA 認識リボザイムの試験管内選択およびその解析)

A Dissertation Submitted for Degree of Doctor of Philosophy
December 2018
平成 30 年 12 月博士（理学）申請

Department of Chemistry, Graduate School of Science,
The University of Tokyo
東京大学大学院理学系研究科
化学専攻

Satoshi Ishida
石田 啓

Abstract

Aminoacylation reaction is a key step that bridges RNA to protein. This reaction is solely conducted by protein enzymes called aminoacyl-tRNA synthetases (ARSs) in the current world. The high specificity toward both tRNA and amino acid of ARSs ensures the right connection between anticodon and amino acid thus accurate translation of mRNA to protein is achieved. It is thought that in the early stage of life, RNA was responsible not only for genomic storage but also for catalyzing chemical reaction and this hypothesis is called the RNA world. In the transition era between the RNA world and the current world, the aminoacylation reaction is possibly conducted by RNA. Such RNA responsible for aminoacylation would be called aminoacylation ribozyme. This ribozyme is expected to have the characteristics of natural ARS that is the specificity toward both amino acid and tRNA. Since no such ribozyme has yet been discovered, the aim of this thesis is to discovery such ARS-like ribozyme.

Chapter 1 describes the general introduction of this thesis. The RNA world hypothesis is explained in detail and the riboswitch which is a possible relic of RNA world is explained as well. Also, the history of aminoacylation ribozyme discovery is presented.

In chapter 2, the discovery of aminoacylation ribozyme using T-box riboswitch as starting structure is described. A ribozyme which has specificity toward both amino acid and tRNA was obtained. The detailed characterization of this ribozyme is also explained. This aminoacylation ribozyme was coupled to the *in vitro* translation mixture to demonstrate the feasibility of this ribozyme.

In chapter 3, the amino acid specificity against other amino acid was attempted using the similar approach as chapter 2. The finding of aminoacylation ribozyme as well as serendipitous biotinylation ribozyme are described.

In the last chapter, the achievement of this thesis is discussed. The implication of the discovery of aminoacylation ribozyme from riboswitch is also pointed out.

Contents

Chapter 1

General introduction

1.1 RNA world hypothesis

In the current world, life is governed by three types of biomolecules; DNA, RNA and protein. Genetic information is stored in the form of DNA which is transcribed into mRNA and then translated into protein and all of these processes are done with the help of protein enzymes (Figure 1.1a). It is unlikely that the complicated biological system nowadays suddenly appeared and it is thought that the life on early earth was using a different system. The idea that an RNA-based system was predominant before the rise of the current DNA/RNA/protein based system is a popular hypothesis which is also called the RNA world hypothesis. In this hypothesis, RNA was responsible for both genomic storage and chemical reaction catalysis (Figure 1.1b). This hypothesis was coined by Walter Gilbert in 1986[3]. The discovery of RNA bearing catalytic activity, ribozyme, in 1980s have enhanced the plausibility of this hypothetical RNA world. These ribozymes include the self-splicing RNA found in *Tereahymena thermophila*[4] and the M1 RNA which is the RNA component of bacterial RNase P responsible for tRNA maturation[5]. It is important to note that these ribozymes have catalytic activity even in the absence of protein. Since then many RNAs bearing catalytic function were both discovered in nature or artificially engineered in laboratory. Especially the engineering of ribozyme was boosted by the development of a methodology to selecting RNAs from large random libraries also known as Systematic Evolution of Ligands by Exponential enrichment (SELEX)[6,7].

Another evidence of the possible RNA world is the fact that the ribosome which plays a critical role in the modern translation system is also a ribozyme. Although ribosome is a ribonucleoprotein which is composed of both ribosomal RNAs and ribosomal proteins, the central core; peptidyl transfer center (PTC) is composed only of RNAs which was indisputably determined by X-ray structure[8,9]. These findings pose the possibility that the primitive ribosome could have been solely composed of RNA as coined by Crick[10].

In the course of evolution, the hypothetical ribozymes present in the RNA world would have been taken over by protein enzymes, and it is unlikely to find ancient ribozymes in the contemporary world. However, as the current ribosome is regarded as a ribozyme, the relic of the ancient ribozymes could be present or even experimentally engineered from random RNA library. Such ribozymes include the RNA polymerase ribozyme[11] which is able to replicate RNA in a protein-free fashion. Since replication of RNA in the absence of protein enzymes is an important feature in the RNA world, the successful engineering of such ribozyme, although not found in nature, is an evidence that supports the RNA world hypothesis.

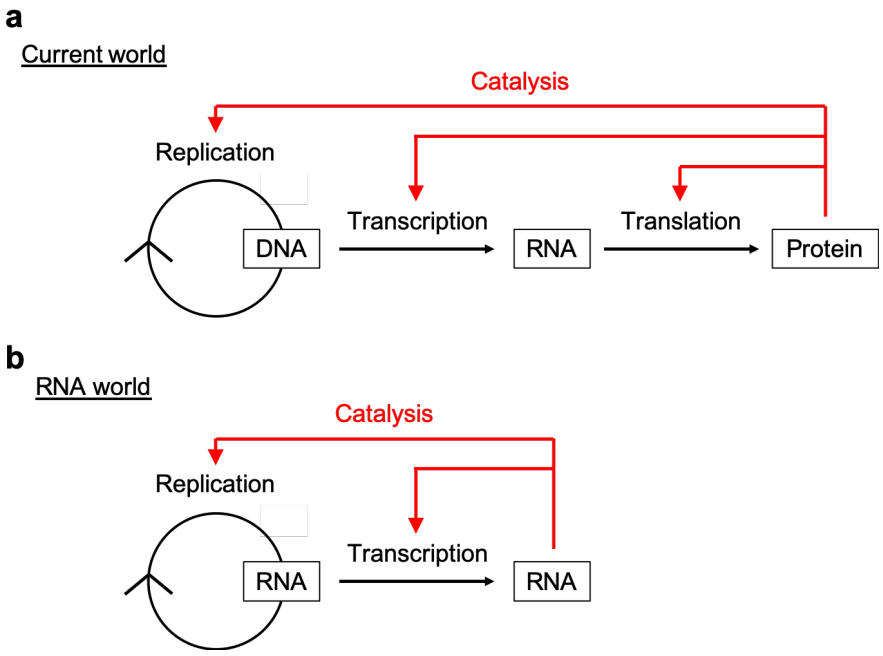

Figure 1.1 | Schematic illustration of the central dogma
(a) The interplay of DNA, RNA and protein in the current world. (b) The schematic image of how RNA functioned in the RNA world.

1.2 Riboswitch

Riboswitches are class of RNA motif that regulates the expression of target gene in the presence of its ligand. Riboswitches reside upstream of the gene they regulate and commonly found in 5'-untranslated region (UTR) of bacteria mRNA. Although most of the riboswitches are discovered in bacteria, some are spread in all three domains of life and there are variety of ligands that function as a gene expression switch for the riboswitch. These ligands include; coenzymes[12-15], nucleotides[16-19], amino acids[20-22], sugar[23], ions[24-26] or even RNA[27-29]. Upon binding of its ligand, the riboswitch undergoes a structural change and controls the expression of the target gene at the level of either transcription attenuation or translation initiation.

In most cases, riboswitch comprises two domains, where one domain is responsible for ligand sensing and the other domain is responsible for gene regulation. The sequence overlapping between these two domains contributes to the switching of riboswitch upon ligand binding. In terms of transcriptional attenuation, the gene regulation domain commonly forms a helix, followed by poly-uridine sequence. The weak interaction with poly-uridine sequence and DNA results in the release of RNA polymerase thus transcription termination is achieved. As for translation initiation, the gene regulation domain forms a helix that include the Shine-Dalgarno sequence. This would prevent ribosomal binding which hinders translation initiation. Ligand binding will either induce or prevent this helix formation and translation initiation is regulated.

The intriguing fact about the discovery of riboswitch is that RNA alone can sense the surrounding environment and undergoes structural change as well as has the ability to selectively bind to metabolites or even other RNAs. This has led us the possibility that RNAs in the possible RNA world could have been using a similar recognition mechanism

as seen in current riboswitches or riboswitches could be the reminiscent of ancestral ribozymes present in the RNA world[30].

1.3 Strategy to obtain functional oligonucleotides

SELEX or also known as *in vitro* selection is an established screening method for enriching nucleotides from a pool of random library based on desired functionality. This method was first reported from two different groups in 1990 and they succeeded in identifying specific RNA sequences that bind to T4 DNA polymerase[31] and organic dyes[7] respectively. The establishment of this method in 1990s has boosted the discovery of numerous ribozymes and RNA sequences binding to specific target compounds (aptamers). *In vitro* selection of functional RNA comprises of mainly four steps (Figure 1.2). First, DNA library is constructed where T7 promoter domain for transcription and constant regions at both 5' and 3' end are included. These constant regions are necessary for PCR and reverse transcription in the downstream procedure. The prepared DNA library was then *in vitro* transcribed and usually gel purified. The obtained library was then applied to screening based on affinity or catalytic activity. The active species that are selectively recovered from the random pool are then reverse transcribed followed by PCR amplification to obtain another set of a DNA library. By repeating the above cycle for multiple times, RNA sequence bearing the desired functionality would be enriched from a random library. Using this method, ribozymes catalyzing various chemical reactions have been obtained. These ribozymes, include the aforementioned RNA polymerase ribozyme[11,32,33], alcohol dehydrogenase ribozyme[34], ribozyme that catalyzes Diel-Alder reaction[35,36] or aminoacylation ribozymes[2,37-41].

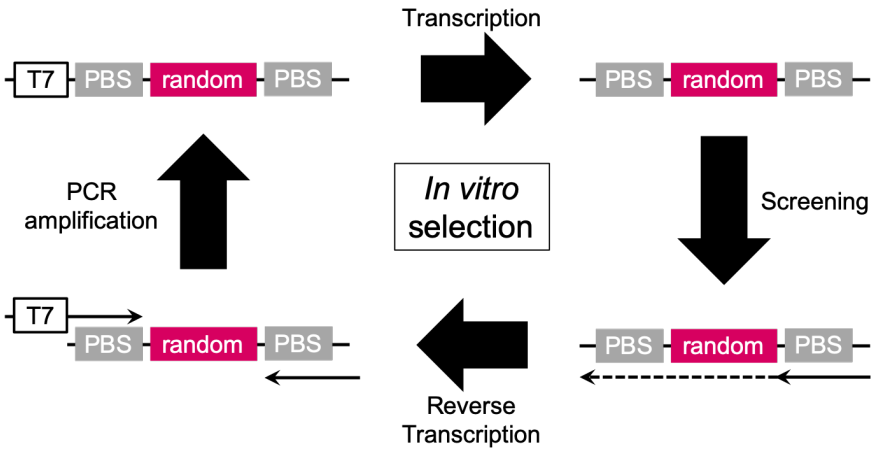

Figure 1.2 | Schematic illustration of the *in vitro* selection
PBS denotes primer binding site circled with gray. T7 denotes T7 promoter sequence. Random region is colored magenta.

1.4 History of aminoacylation ribozyme

1.4.1 Aminoacy-tRNA synthetase (aaRs)

Aminoacylation is a key step that bridges amino acid and RNA in the current world. This reaction is catalyzed by protein enzymes called aminoacyl-tRNA synthetases (aaRs). It is proposed that this reaction could have been catalyzed by RNA especially in the transition era between the RNA world and the current RNA/DNA/protein world. The reaction catalyzed by aaRs has two steps; (i) amino acid activation and (ii) aminoacyl transfer (Figure 1.3). The activation of amino acid is achieved by adenylation of the carboxylic acid using ATP as substrate. The adenylated amino acid is reacted with the substrate tRNA and 3′ end of tRNA is aminoacylated. Many researchers have attempted to engineer aminoacylation ribozymes that catalyze the above reaction but all of them can only catalyze either of the two reactions.

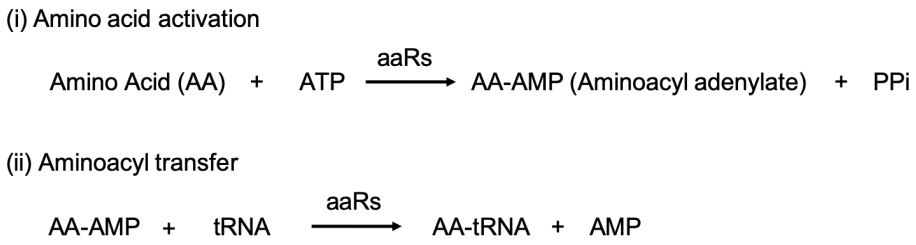

Figure 1.3 | Two reactions carried out by aminoacyl-tRNA synthetases (aaRs)
(i) Amino acid activation is achieved with the help of ATP and aminoacyl adenylate (AA-AMP) is formed while releasing pyrophosphate (PPi) after activation. (ii) aaRs recognizes the substrate tRNA and transfers AA-AMP to the 3′-end of tRNA.

1.4.2 Amino acid activating ribozyme

In 2001, Yarus group succeeded in developing an amino acid activation ribozyme, KK13[1]. This ribozyme utilizes the 5′-triphospahte of its own sequence to activate amino acid. The selection strategy employed in the discovery of KK13 is shown in Figure 1.4. They used 3-mercapto-propionic acid as the initial substrate and mixed it with the RNA pool having 5′-triphospate. After incubation, the reacted RNA was treated with thiopropyl

sepharose 6B matrix to selectively capture active RNAs. The active RNAs were eluted using DTT, ethanol precipitated, reverse transcribed, PCR amplified, and transcribed. After repeating this step for 8 times, they succeeded in identifying several sequences and among them, KK13 was further characterized and activation using amino acid substrates was confirmed. The secondary structure of KK13 is shown in Figure 1.5. The 5′-terminal triphosphate was necessary for activation since 5′-pKK13 and 5′-OH KK13 did not react with the substrate. However, this ribozyme functions best at pH 4-4.5 due to the instability of the product.

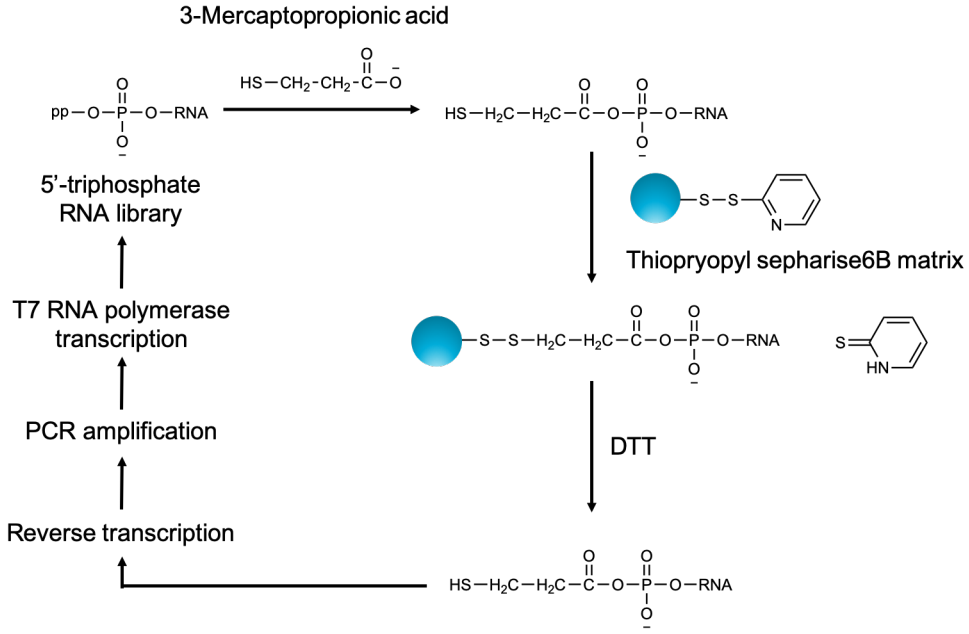

Figure 1.4 | Schematic procedure of selecting carboxyl-activating ribozymes.
This figure is adapted from previous paper[1]. 3-mercaptopropionic acid was used as the substrate and subjected to incubation with the RNA library. The active species were selectively recovered by the thiopyropyl sepharose6B matrix.

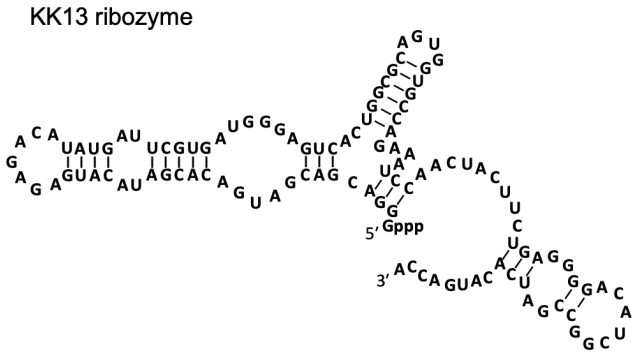

Figure 1.5 | Secondary structure of KK13
This figure is adapted from previous paper[1]. The 5′-terminal triphosphate attacked by the carboxylate oxygen to give activated amino acid.

1.4.3 Aminoacyl transfer ribozyme

All aminoacyl transfer ribozymes discovered so far utilizes an already activated amino acid. The activated amino acid could be either aminoacyl adenylate which the current aaRs utilizes or other artificially activated amino acid such as cyanomethyl esters (CME). The first aminoacylation ribozyme was discovered by Yarus group in 1995 using Phe-AMP as substrate[2]. The selection procedure employed is shown in Figure 1.6. The RNA library was mixed with Phe-AMP and then reacted with naphthoxyacetyl-N-hydroxysuccinimidyl ester to convert the active RNA more hydrophobic and purifiable using reverse-phase HPLC. The selection resulted in isolating isolate #29 RNA which aminoacylate Phe-AMP at the 3′-teminal guanosine. They also succeeded in minimizing isolate #29 RNA sequence (95 mer) to 29-mer (Figure 1.7a,b)[39]. Both the minimized isolate #29 or original isolate #29 RNA had no amino acid specificity and required both Mg^{2+} and Ca^{2+} for their activity. Yarus group also engineered a tiny aminoacylation ribozyme called C3 ribozyme (Figure 1.7c)[41]. This ribozyme was engineered utilizing an RNA library that includes hammerhead ribozyme (HDV) at the 3′-end. Since the HDV does not have any requirement for the upstream nucleotide, the 3′-end nucleotide for the aminoacylation site can be anything as opposed to the previous selection, where 3′-end nucleotide

was restricted because it was included in the primer binding site. This C3 ribozyme, as compared to #29 ribozyme, does not need divalent ions such as Mg^{2+} or Ca^{2+} for reactivity, and highly specific for L-amino acids.

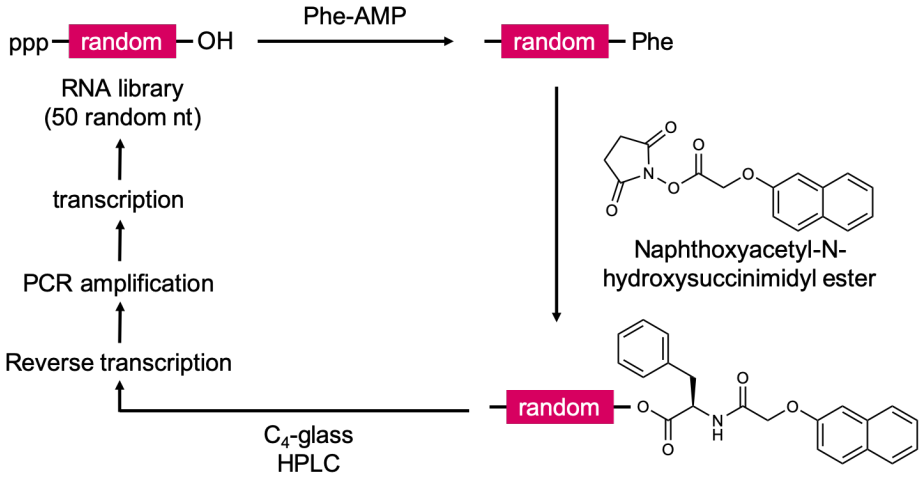

Figure 1.6 | Selection scheme for discovering isolate #29 ribozyme
This figure is adapted from previous paper[2]. The RNA pool was reacted with Phe-AMP followed by naphthoxyacetyl-N-hydroxysuccinimidyl ester mixing. Beacuase of the hydrophobilicy of naphthalene moiety, this resulted in retention time diference compared to unreacted RNA in the HPLC purification.

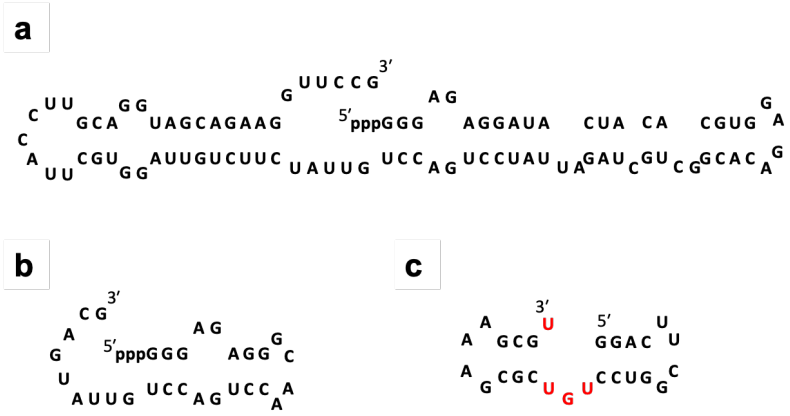

Figure 1.7 | Secondary structures of isolate #29 RNA derivative and C3 ribozyme.
(a) Secondary structure of isolate #29 RNA. (b) Secondary structure of minimized isolate #29 RNA. (c) Secondary structure of C3 ribozyme. Red nucleotides are critical for its activity.

Several aminoacylation ribozymes have been developed that utilizes artificially activated amino acids as well. The first of such aminoacylation ribozyme is ATRib which was developed in Szostak's group (Figure 1.8a)[42,43]. ATRib uses aminoacylated hexanucleotide as an aminoacyl donor and self-aminoacylates itself. The selection strategy was using Biotin-Met as an amino acid substrate and recovering active species through streptavidin agarose. Using ATRib, aminoacylation to tRNA was also achieved using a ping-pong like process (Figure 1.8b). Briefly speaking, ATRiB first binds to aminoacylated hexanucleotide and the amino acid is tranfered to ATRiB. The free-hexanucleotide is detached and tRNA-3′-end binds to the same site as hexanuleotide. The amino acids that was attached to ATRib itself then transfers to the bound tRNA which will result in the release of aminoacylated tRNA. Although all of these processes are in equilibrium, they succeeded in obtaining 30-40% of aminoacylated tRNA in 2hrs. Also, another *in vitro* selection was conducted to give ATRib additional tRNA specificity, which resulted in the

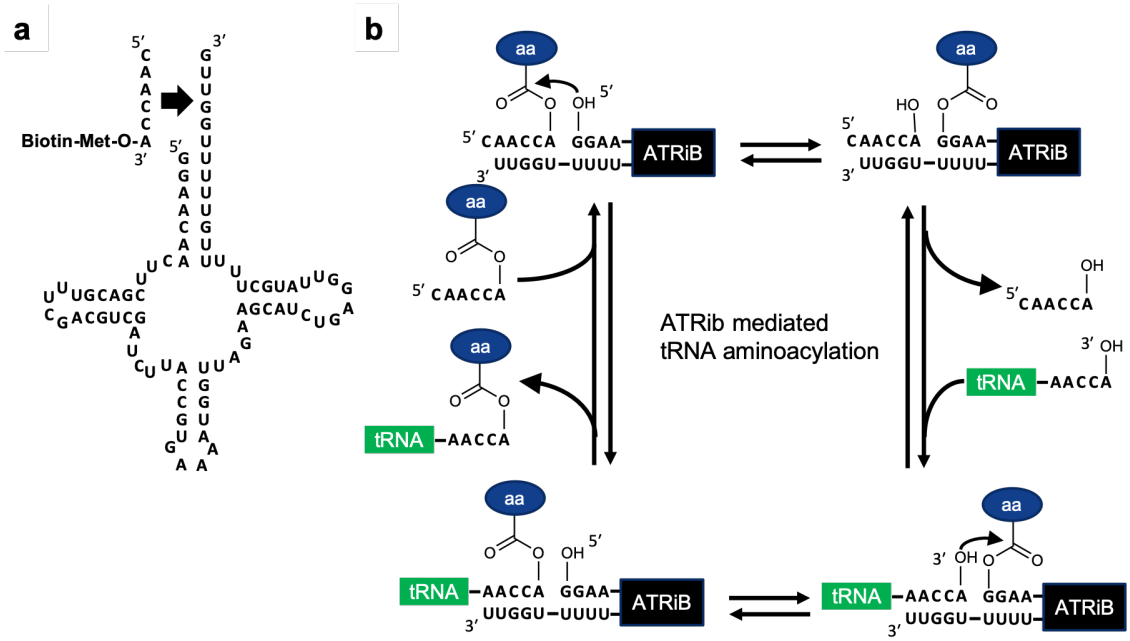

Figure 1.8 | Characterization of ATRib and ATRib mediated aminoacylation
(a) Secondary structure of ATRib. (b) Schematic illustration of ATRib mediated tRNA aminoacylation.

engineering of aminoacylation ribozyme BC28[44]. This ribozyme has an accessory domain to ATRib that recognizes tRNA anticodon.

Nearly two decades ago, our lab also developed an aminoacylation ribozyme based on the hypothesis that 5′-leader sequence of tRNA before maturation by RNaseP could have been responsible for aminoacylating the downstream tRNA[38]. An RNA library having 70-nt random sequence upstream of tRNA was prepared and those that can self-aminoacylate biotin-Phe-CME was selectively recovered through streptavidin agarose. Total of 17 rounds of screening was conducted, which eventually resulted in obtaining self-aminoacylating ribozyme, pre-24. This ribozyme also retained its activity when treated with RNaseP which means that this ribozyme functions *in trans* as well. Minimization and further engineering of r-24 (tRNA deleted version of pre-24) resulted in an aminoacylation ribozyme Fx3[45]. This ribozyme base-paris with the CCA-end of tRNA and could accept Phe derivative amino acids. Fx3 was further engineered to accept more variety of amino acid substrates as well as enhanced activity and these ribozymes were called flexizymes[40,46] (Figure 1.9). eFx accepts aromatic amino acids activated with cyanomethyl esters or chlorobenzyl thioester. dFx accepts non-aromatic amino acids that are activated with 3,5-dinitrobenzyl ester and aFx accepts amino acids with (2-aminoethyl)- amidocarboxybenzyl thioester. The versatility of flexizymes enabled us to mischarge any tRNA of interest with non-natural amino acids which lead to studies on genetic code reprogramming[47-50] or even discovery of functional non-natural peptides binding to specific protein(s) of interest[51-59]. For Fx3, the crystal structure has been elucidated and the binding pocket for the aromatic ring as well as the importance of hydrated magnesium for stabilizing the active site has been shown[60].

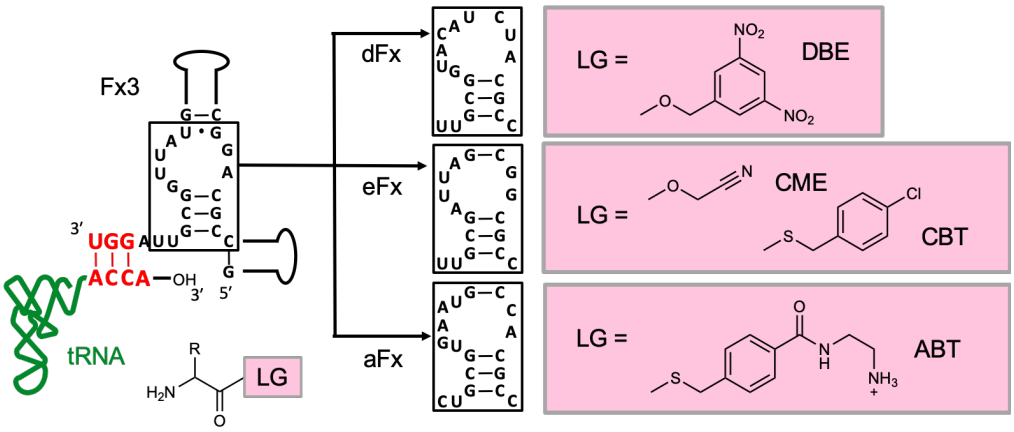

Figure 1.9 | Schematic illustration of Fx3 and flexizymes.
Red letters represent the positions of base-pairs between Fx3 and substrate tRNA. LG denotes leaving group. DBE denotes 3,5-dinitrobenzyl-ester, CME denotes cyanomethyl-ester, CBT denotes chlorobenzyl-thioester and ABT denotes (2-aminoethyl)- amidocarboxybenzyl thioester.

Chapter 2

In vitro selection of aminoacylation ribozyme using T-box motif

2.1 Introduction

As described in the first chapter, aminoacylation of tRNA is a key step that bridges the hypothesized RNA world with the current world and is conducted solely by a protein enzyme called aminoacyl-tRNA synthetase (aaRS). AaRS recognizes both the substrate amino acid and the substrate tRNA conjugating the right pair of amino acid and anticodon ensuring the fidelity of translation. In the transition era between RNA world and current protein, the aminoacylation reaction would be conducted by ribozyme. Nonetheless, aminoacylation ribozyme bearing both specificity toward amino acid and tRNA has not yet been found. In order to find such ancestral ribozyme of current aaRS, a specific riboswitch called T-box riboswitch caught the attention.

T-box riboswitch is an RNA motif that is found in gram-positive bacteria. It resides in the upstream of the gene it regulates, and monitors amino acid status by monitoring the aminoacylation status of the cognate tRNA. Figure 2.1 shows the common mechanism of gene regulation by the T-box riboswitch. When a specific tRNA is charged with the corresponding amino acid, there is no change in the T-box riboswitch structure. However, when a specific uncharged tRNA is present, T-box binds with this tRNA, undergoes a structural change and starts the transcription of the downstream gene. Therefore, T-box undergoes transcription antitermination mechanism for gene regulation. There are some cases, however, where upon binding of target tRNA, T-box riboswitch undergoes a structural change to expose the Shine-Dalgarno region of the mRNA and initiate translation. The genes that T-box riboswitch regulates are amino acid related, such as aminoacyl-tRNA synthetase or amino acid metabolism related proteins.

In this research, a specific T-box riboswitch from the *Bacilus Subtilis* species that is responsible for recognizing *Bacilus Subtilis* $tRNA^{Gly}_{GCC}$ ($BStRNA^{Gly}_{GCC}$) called *Bacillus subtilis glyQS* T-box[61,62] was chosen because this riboswitch is one of the well-characterized T-box riboswitch. The partial crystal structure of *Bacillus subtilis glyQS* T-box in

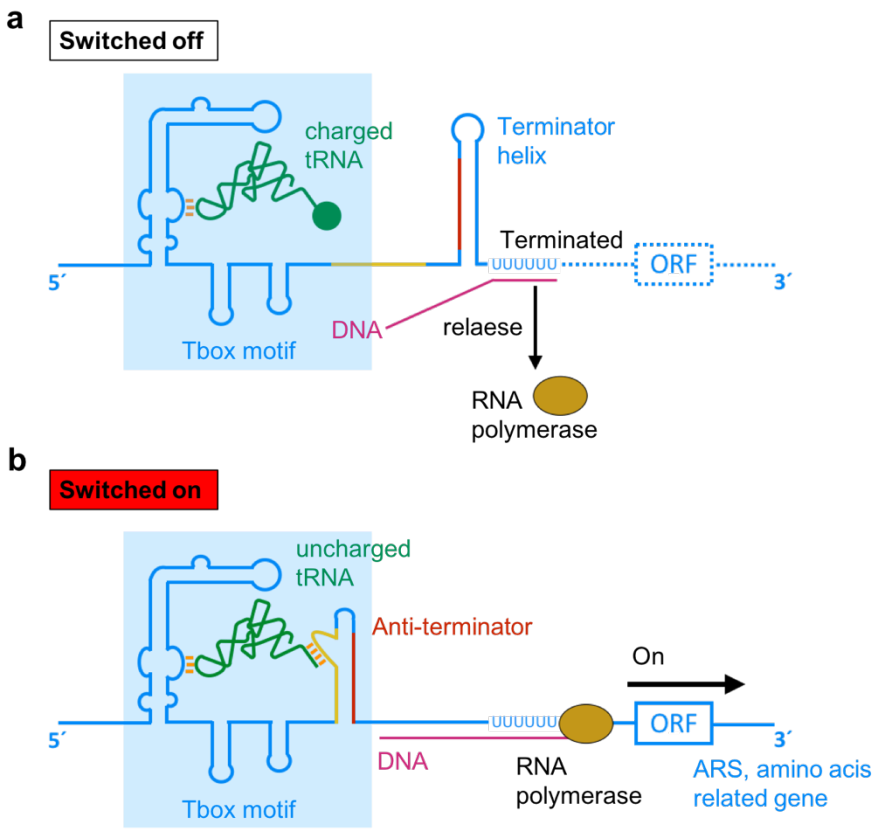

Figure 2.1 | Schematic representation of gene regulation by T-box riboswtich
a) When charged tRNA is present the T-box riboswitch is switched off and the downstream transcription is terminated by release of RNA polymerase. b) When a specific uncharged tRNA is present, T-box riboswitch undergoes a structural change that breaks the terminator helix and switch on the transcription.

complex with BStRNA$^{Gly}_{GCC}$ is shown in Figure 2.2 where tRNA base pairs with the two conserved domain of T-box riboswitch, Stem I domain and antiterminator domain. The specific sequence in the Stem I domain called the specifier sequence is where the base pairing is formed with the tRNA anticodon. The antiterminator domain is predicted to compromise with two helices, helix A1 and helix A2. The coaxial stacking between helix A1 of T-box, apical loop in the Stem I domain and tRNA is predicted to form which results in the better affinity toward tRNA. Helix A2 seems not to affect the binding ability toward tRNA[63]. Thus, this tRNA specificity of T-box riboswitch and engineered this RNA sequence was chosen so that it would give additional aminoacylation activity while

maintaining the tRNA recognition element. By doing so the discovery of a new tRNA recognizing aminoacylation ribozyme was attempted.

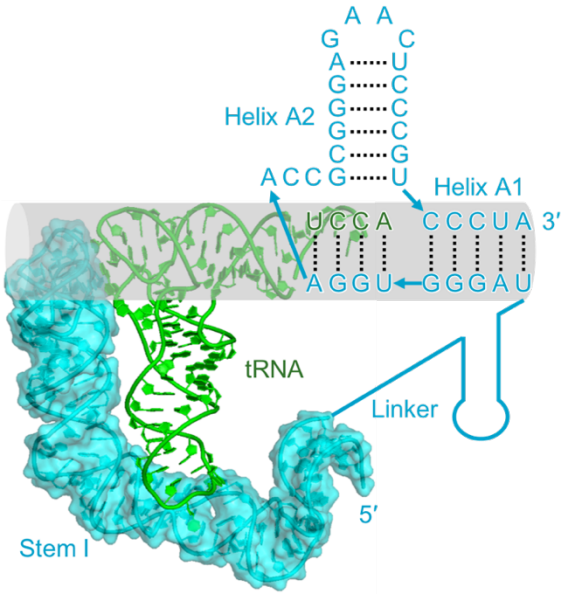

Figure 2.2 | Co-crystal structure and prediction model of $tRNA^{Gly}_{GCC}$ binding to the T-box riboswitch.
Gray cylinder shadow indicates the possible co-axial stacking between tRNA and T-box riboswitch.

In order to find such aminoacylation ribozyme, SELEX (Systematic Evolution of Ligands by EXponential enrichment) method or in other words the *in vitro* selection was conducted. The RNA library used in this study was based on the *Bacillus subtilis glyQS* T-box sequence. Because the helix A2 domain of *glyQS* T-box riboswitch is said to not affect the binding capability[63] and that it is positioned near tRNA-CCA end, this domain was randomized by substituting with forty random nucleotides (N40). Figure 2.3 shows the two libraries used for this study. Library 1 was designed based on the assumption that 3′-end of tRNA would be hindered by the helix A1 domain and that is should be freely exposed, resulting in the deletion of 5 nucleotides (CCCUA) sequence from the wildtype sequence indicated by a dashed rectangle. Library 2 was designed based on the full wildtype sequence with helix A2 randomized with forty nucleotides. For both of the

libraries, the 3′-end of T-box sequence was fused to the 5′-end of substrate tRNA by poly adenine linker.

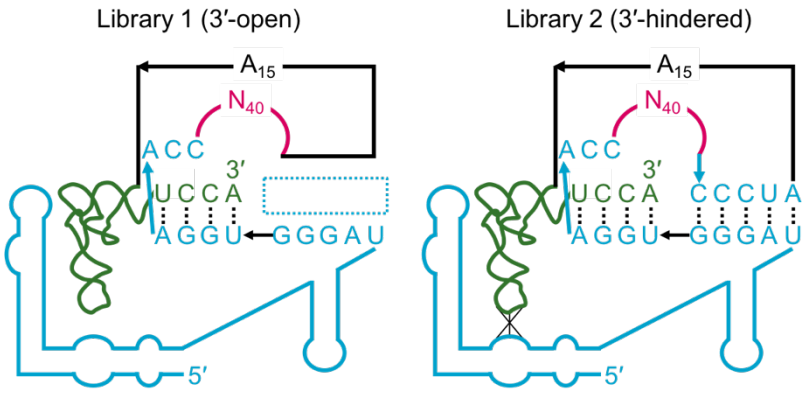

Figure 2.3 | Schematic representation of libraries used for in vitro selection
Blue line indicates the T-box sequence, green line shows the tRNA, black line indicates the poly adenine linker and pink line indicates the random forty nucleotides. Dashed line indicates the domain deleted for preparing library 2.

Using these two libraries, *in vitro* selection was conducted. Figure 2.4a indicates the flow of *in vitro* selection. For the amino acid substrate, *N*-biotinyl-L-phyenylalanyl cyanomethyl ester (bio-Phe-CME) was used as a representative example. Biotin residue was needed for the handle to recover active species. The complete chemical structure of bio-Phe-CME is shown in Figure 2.4b. The RNA library was treated with bio-Phe-CME for aminoacylation reaction. Those with aminoacylation activity will then be selectively recovered with streptavidin magnetic beads with the high affinity of biotin residue and streptavidin. The selected RNA sequence will then be reverse transcribed, PCR amplified and transcribed again to afford a new set of RNA library.

After seven rounds of selection, the increase in the recovery of active RNA sequence was observed. Figure 2.5 shows the progress of selection and the aminoacylation reaction time and condition for each round. Because the recovery of RNA increased in the 6th round, the aminoacylation reaction was conducted in a harsher condition with only 30 minutes for aminoacylation reaction compared with the initial 2 hours.

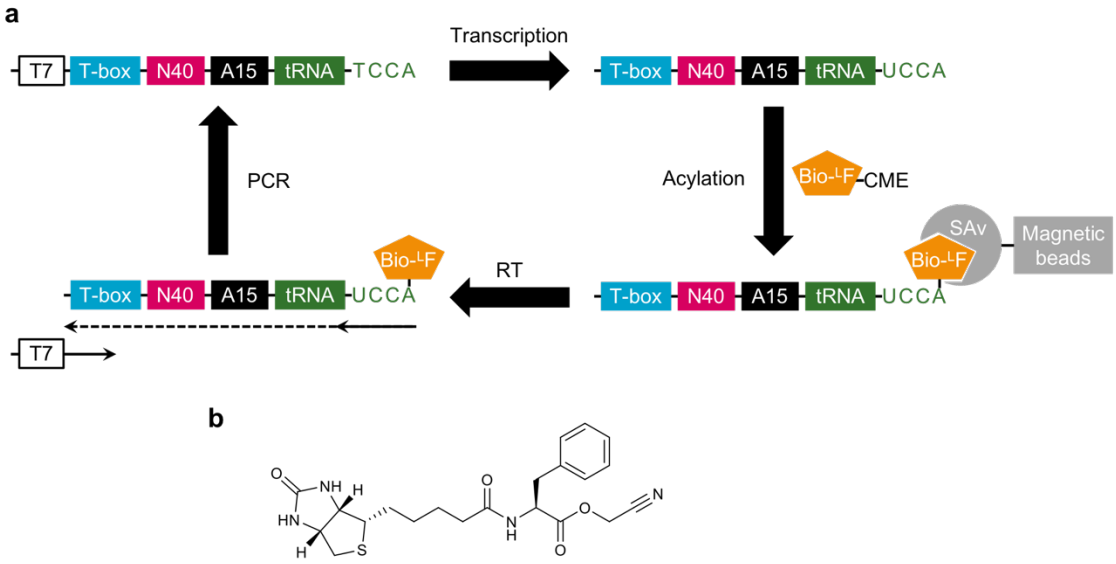

Figure 2.4 | Schematic representation of *in vitro* selection

a) *In vitro* selection strategy used in this research. b) Chemical structure of bio-Phe-CME

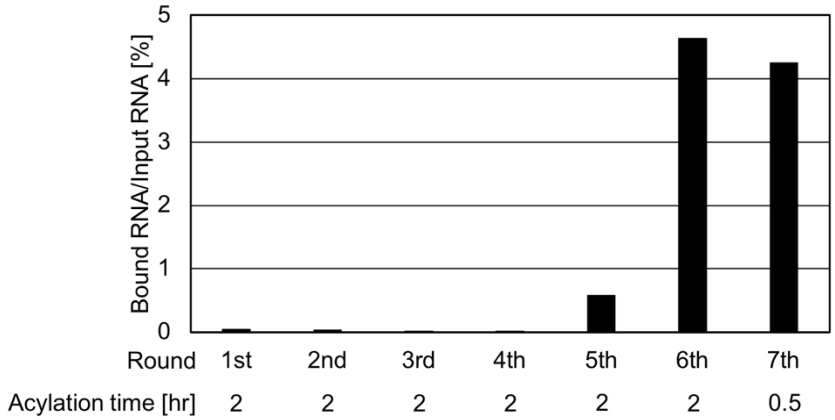

Figure 2.5 | Progress of *in vitro* selection

2.2 Results and discussion

The study in this chapter 2 was published as an article in the scientific journal and the figures used in this chapter are adapted from the published paper.

2.2.1 Sequencing of enriched RNA library

Since the recovery of RNA increased after round 6th, the RNA sequence recovered after round 6 and round 7 was analyzed. Total of 47 clones (28 clones from 6th round and 19 clones from 7th round) were sequenced. 5 clones appeared more than once and the exact sequence in the random region is summarized in Figure 2.6. From library 1 three clones were identified and from library 2 two clones were identified. Interestingly, there is no homology between these five sequences and all of them are unique.

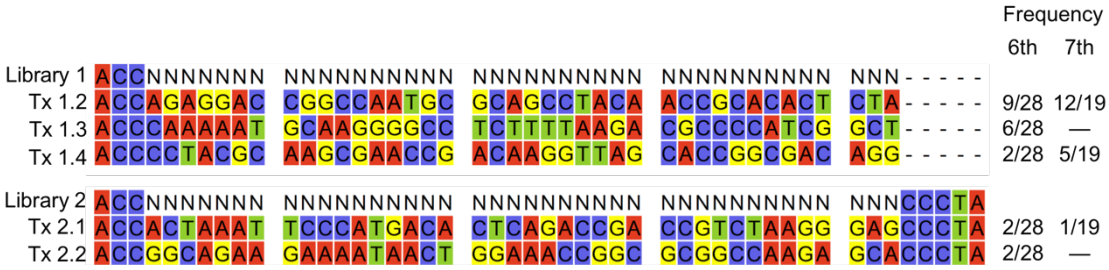

Figure 2.6 | Clones identified after sequencing round 6th and round 7th

2.2.2 Self-aminoacylation check of selected clones with bio-^{L}Phe-CME

To confirm the identified sequences with aminoacylation activity, each of the RNA sequence having the specific randomized domain was transcribed and aminoacylation reaction was conducted. For these experiments, the $tRNA^{Gly}_{GCC}$ sequence is conjugated to the 3′-ends of T-box sequence through polyadenine linker and named as Tx-tRNA. Using streptavidin dependent mobility gel shift assay[42,43], the activity against bio-^{L}Phe-CME was checked (Figure 2.7). At the same time, to confirm the activation site is the hydroxyl groups at the 3′-end ribose, periodate oxidation treated Tx-tRNA sequence was also tested for aminoacylation activity. For all five selected clones, aminoacylation activity was observed and also abolished for periodate oxidation treated ones, meaning that all of the aminoacylation reaction is happening at the 3′-CCA end of tRNA. The activity of each ribozyme was ranging from 13% to 48%.

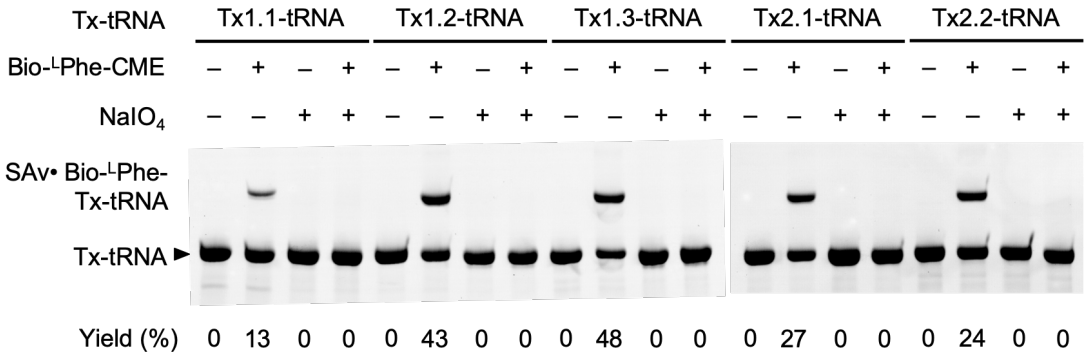

Figure 2.7 | Self-aminoacylation activity of five identified clones against bio-^{L}Phe-CME. The reaction was conducted both with and without the amino acid substrate, bio-^{L}Phe-CME as well as with and without periodate oxidation treatment.

2.2.3 Self-aminoacylation activity check using other amino acids

With the previous experiment, the activity against the original substrate used in the *in vitro* selection was confirmed for all identified clones. For the next step, the substrate scope of these ribozymes was examined using other biotinylated and non-biotinylated substrates (Figure 2.8). The substrates used in this experiment are, *N*-biotinyl-D-phenylalanyl-cyanomethyl ester (bio-^{D}Phe-CME), *N*-biotinyl-L-tyrosyl-cyanomethyl ester (bio-^{L}Tyr-CME), *N*-biotinyl-glycyl-cyanomethyl ester (bio-Gly-CME) and L-phenylalanyl-cyanomethyl ester (^{L}Phe-CME). The activity was determined by the same streptavidin dependent gel shift assay. As a result, Tx2.1-tRNA and Tx2.2-tRNA could only charge other biotinylated substrates. Tx1.1-tRNA charged bio-^{D}Phe-CME at a similar efficiency as bio-^{L}Phe-CME and bio-^{L}Tyr-CME at approximately 50 % efficiency compared with the original substrate. Tx1.2-tRNA and Tx1.3-tRNA showed similar tendency; slightly charging bio-^{D}Phe-CME and bio-^{L}Tyr-CME. None of the identified ribozymes was able to aminoacylate bio-Gly-CME showing that the aromatic residue is essential for aminoacylation acitivity. For substrate ^{L}Phe-CME, after the aminoacylation reaction, the ethanol precipitated RNA was treated with biotin-sulfo-NHS to selectively biotin label the ^{L}Phe acylated product. For this experiment, Tx1.1-tRNA and Tx1.3-tRNA showed comparable activity but others showed decrease activity with this non-biotinylated product. These findings indicate that the five clones identified have different amino acid selectivity.

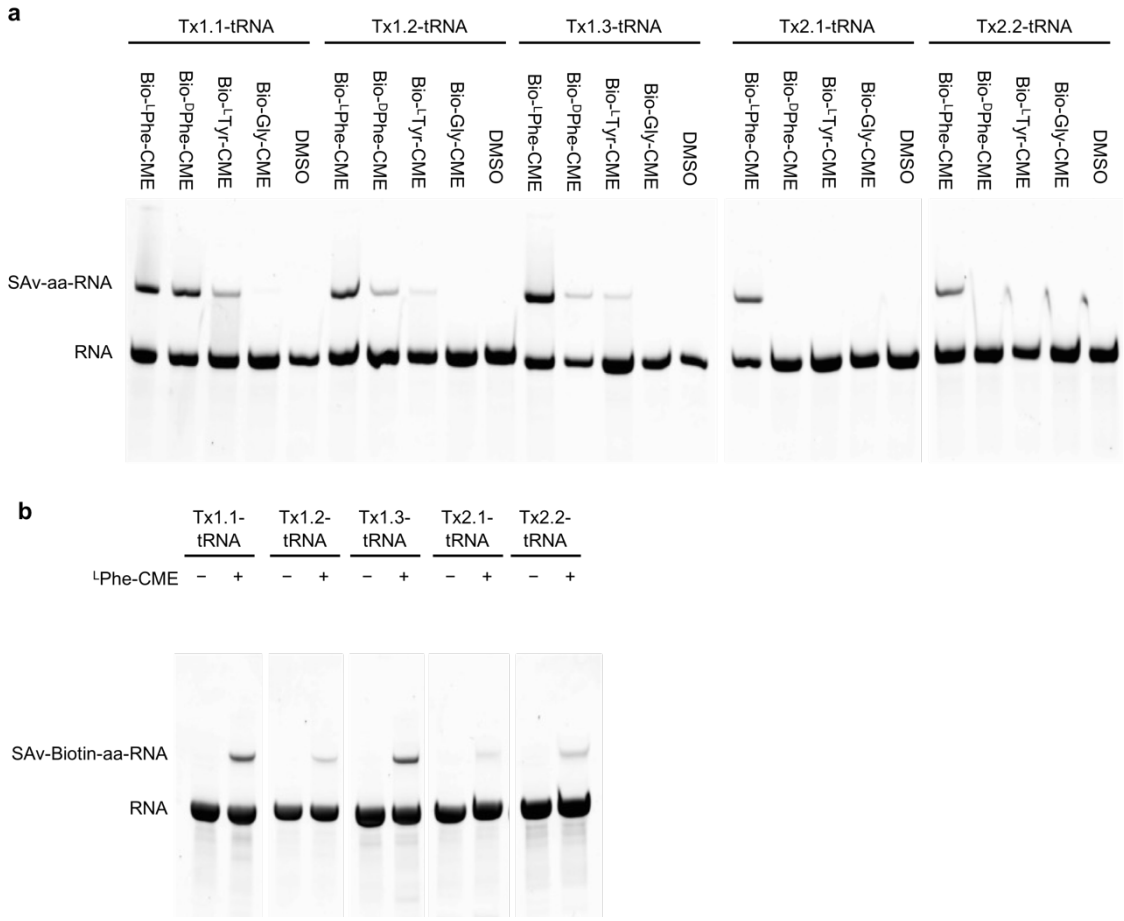

Figure 2.8 | Self-aminoacylation activity of five identified clones with variety of amino acid substrates

a) Self-aminoacylation activity check with biotinylated substrates. b) Self-aminoacylation activity check with ^{L}Phe-CME. After the aminoacylation reaction, the product was reacted with sulfo-NHS-biotin to biotin label the aminoacylated RNA.

2.2.4 Trans-aminoacylation activity of selected clones

For the above experiments, the substrate tRNA was covalently linked with the identified ribozymes. Therefore, the aminoacylation activity against separate $tRNA^{Gly}_{GCC}$ was examined[64]. All of the ribozymes were transcribed again without the polyadenine linker or the substrate $tRNA^{Gly}_{GCC}$ sequence at the 3′-end and checked for aminoacylation activity against bio-^{L}Phe-CME (Figure 2.9a). At the same time, the periodate oxidation treated $tRNA^{Gly}_{GCC}$ was also tested for aminoacylation to check whether aminoacylation is happening at the 3′-CCA end. As a result, out of the five ribozymes tested, Tx1.3 and Tx 2.1 could aminoacylate the terminal 3′-CCA end with bio-^{L}Phe-CME. In order to validate whether the activities of these two ribozymes were dependent on the interaction between the tRNA anticodon and ribozymes, trans-aminoacylation activity was measured against $tRNA_{UAA}$ bearing non-cognate UAA anticodon (Figure 2.9b). Tx1.3 was able to aminoacylate both $tRNA^{Gly}_{GCC}$ and $tRNA^{Gly}_{UAA}$ while Tx2.1 had specific activity against $tRNA^{Gly}_{GCC}$.

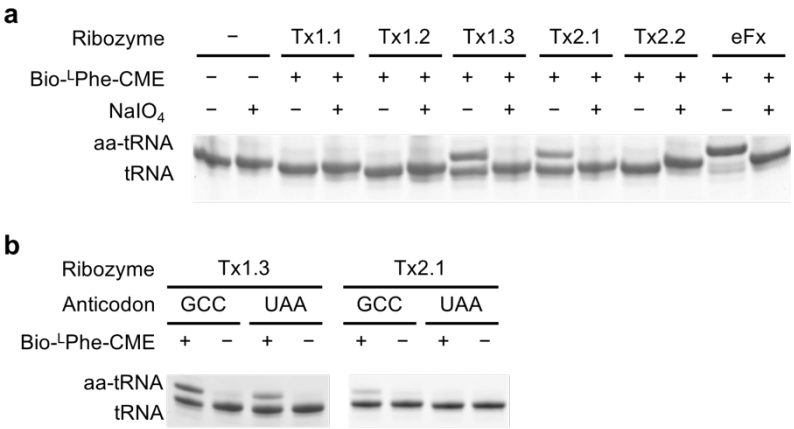

Figure 2.9 | Trans-aminoacylation activity of identified ribozymes.
a) Trans-aminoacylation of identified ribozymes with or without periodate oxidation treatment against mature $BStRNA^{Gly}_{GCC}$. b) Trans-aminoacylation activity against anticodon mutated $BStRNA^{Gly}_{UAA}$.

2.2.5 Activity check of T-box body sequence deleted Tx1.3 and Tx2.1

The activity of Tx1.3 and Tx2.1 with only the randomized domain in the initial library construction was also measured to check the importance of T-box body sequence for the aminoacylation activity. The transcribed sequence along with the possible secondary structure is shown in Figure 2.10a,d. The trans-aminoacylation activity of these two truncated ribozymes revealed that Tx1.3 retained its aminoacylation activity even in the absence of T-box body sequence while Tx2.1 lost its activity when T-box body sequence was deleted (Figure 2.10c). This result is in accordance with the result in the previous section 2.2.4 where Tx2.1 activity in dependent on interaction with the tRNA anticodon. Surprisingly, the predicted secondary structure of truncated Tx1.3 was similar to that of flexizyme, which only recognizes the tRNA CCA-end for its activity. Especially, the purine/pyrimidine positions in the catalytic core of flexizyme (J1a/2 and J2/1a)[65] are completely the same compared to the corresponding region of Tx1.3 (Figure 2.10d). This suggests that Tx1.3 could be a flexizyme derivative that does not need the T-box body sequence for its activity against substrate tRNA. Discovery of such ribozyme is somewhat predictable given that the amino acid substrate used in this study was the same as when the original flexizyme was obtained and that the same in vitro selection strategy was used. However, since the goal of this research was to obtain tRNA-recognizing aminoacylation ribozyme, Tx2.1 which seems to have specificity toward tRNA anticodon was further characterized.

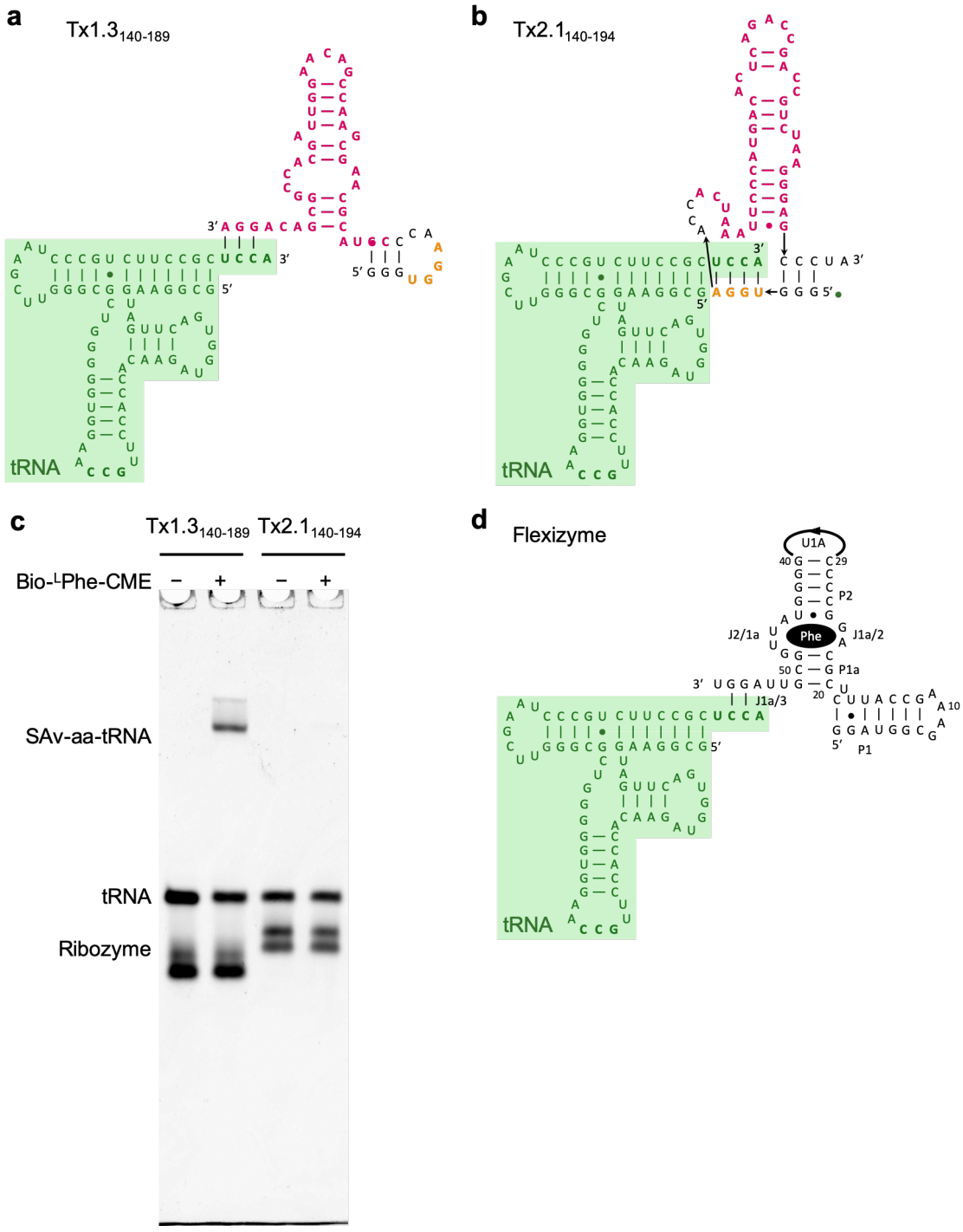

Figure 2.10 | Trans-aminoacylation activity of truncated Tx1.3 and Tx2.1.
a) Possible secondary structure of truncated Tx1.3 along with substrate tRNA. b) Possible secondary structure of truncated Tx2.1 along with substrate tRNA. c) Trans-aminoacylation activity of truncated Tx1.3 and Tx2.1. d) Secondary structure of Fx3 with substrate tRNA.

2.2.6 Characterization of Tx2.1

First, the amino acid selectivity of Tx2.1 was examined (Figure 2.11). Various kinds of amino acids substrates were examined including, *N*-acetyl-L-phenylalanyl-cyanomethyl ester (Ac-L-Phe-CME), L-Phe-CME, bio-^{L}Phe-CME, bio-^{D}Phe-CME, bio-^{L}Tyr-CME and bio-Gly-CME. Among them, only the original amino acid substrate, bio-^{L}Phe-CME could be aminoacylated. Phenylalanine derivatives were not aminoacylated indicating the important role of biotin residue. Even the other biotinylated substrates were not aminoacylated, showing that both the configuration and the structure of side chain as phenylalanine is strictly recognized by the Tx2.1 ribozyme.

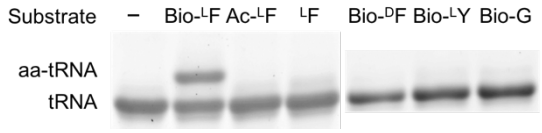

Figure 2.11 | Amino acid selectivity of Tx2.1 ribozyme

Next, a plot of initial rates for the self-aminoacylation of Tx2.1-tRNA against various bio-^{L}Phe-CME concentrations were conducted to determine the kinetic parameters. The plot revealed Michaelis-Menten behavior with kinetic parameters of k_{cat} = 0.0577 ± 0.0047 min^{-1} and K_m = 3.0 ± 0.51 mM (Figure 2.12). These results are comparable to a previously discovered aminoacylation ribozyme, pre-24 which is a prototype of flexizyme[38] that does not have tRNA specificity with the values k_{cat} = 0.13 ± 0.014 min^{-1} and K_m = 2.8 ± 0.61 mM. pre-24 was discovered using the same amino acid substrate, bio-^{L}Phe-CME. Since pre-24 ribozyme was further engineered to give enhanced activity and result in the discovery of flexizymes, similar engineering could be made to Tx2.1 as well to get better efficiency for aminoacylation. For the kinetics experiment, bio-^{L}Phe-CME concentration above 5 mM were not conducted due to its low solubility in aqueous solvent.

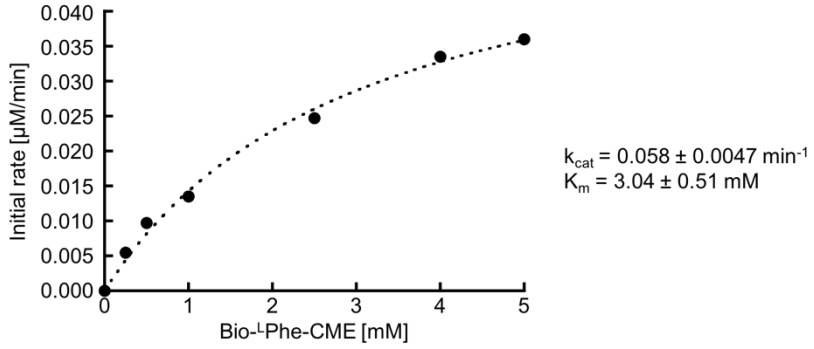

Figure 2.12 | Plot of initial rate for self-aminoacylation activity of Tx2.1-tRNA
Each data point is the mean of three independent experiments. The plot was fitted to the Michaelis-Menten non-linear curve by KaleidaGraph (HULINKS)

Detailed anticodon specificity was then examined for Tx2.1. It is reported that for the natural *glyQS* T-box, the binding toward the substrate $tRNA^{Gly}_{GCC}$ and the anti-termination activity of the riboswitch is reduced by introducing mutations to the tRNA anticodon and the T-box specifier sequence[66]. Therefore, the effect of mutation(s) in the $tRNA^{Gly}$ anticodon to the aminoacylation activity was examined. The apparent initial rate of seven non-cognate tRNAs with different anticodons as well as the original $tRNA^{Gly}_{GCC}$ were measured. The actual anticodon sequences tested are as follow, 3-base pairing anticodon (GCC), three 2-base paring anticodons (GCA, GAC, UCC), three 1-base pairing anticodons (GAA, UCA, UAC) and no-base pairing anticodon (UAA). Over 60-fold decrease was observed with tRNA having no base pairing with the specifier sequence. In the case of tRNAs having 1-base pairing anticodons, $tRNA_{GAA}$ showed 40-fold decrease in the initial rate and the other two exhibited 10-fold decrease. The initial rate for tRNAs with 2-base paring anticodons revealed that the 2nd base in the anticodon greatly affected the initial rate (Figure 2.13a). The reason for the importance of 2nd base is possibly the disruption of adjacent stacking interaction when the 2nd base is not forming base-pairing, which would in turn disrupt the whole base-pairs. It should also be noted that when mutation is introduced to compensate with anticodon UAA, the initial rate was recovered by

20%. However, compared with the wildtype pair of GCC anticodon and specifier sequence GGC, the initial rate is 5-fold lower. This decrease can be attributed to the weaker Watson-Crick base pairs between adenine and uracil compared with guanine-cytosine. Not only the apparent initial rate but also aminoacylation efficiency at the endpoint was measured (Figure 2.13b). The endpoint measurement was conducted with 2 hours of aminoacylation reaction. This result shows the same trend as what was observed with the initial rate. However, the recovery in the aminoacylation efficiency with the mutated specifier sequence to compensate with UUA anticodon tRNA was more apparent with the endpoint analysis.

The interaction of T-box riboswitch with the substrate tRNA is not just base-pairing between specifier sequence and anticodon. It has been reported that the apical loop of stem I region interacts with an elbow of ligand tRNA[62,67-70]. Since the length of tRNA T-loop is conserved as 7 nt for all *Bacilus Subtilis* tRNAs[71], I focused on the D-loop length which is 7nt for the wildtype BtRNA$^{Gly}_{GCC}$ and 7-11 nt for the other BstRNAs. Aminoacylation against D-loop mutated BStRNAs$^{Gly}_{GCC}$ revealed that increasing the D-loop size resulted in the decrease in aminoacylation efficiency, suggesting that Tx2.1 also interacts with the tRNA D-loop in the same manner as the natural T-box (Figure 2.14).

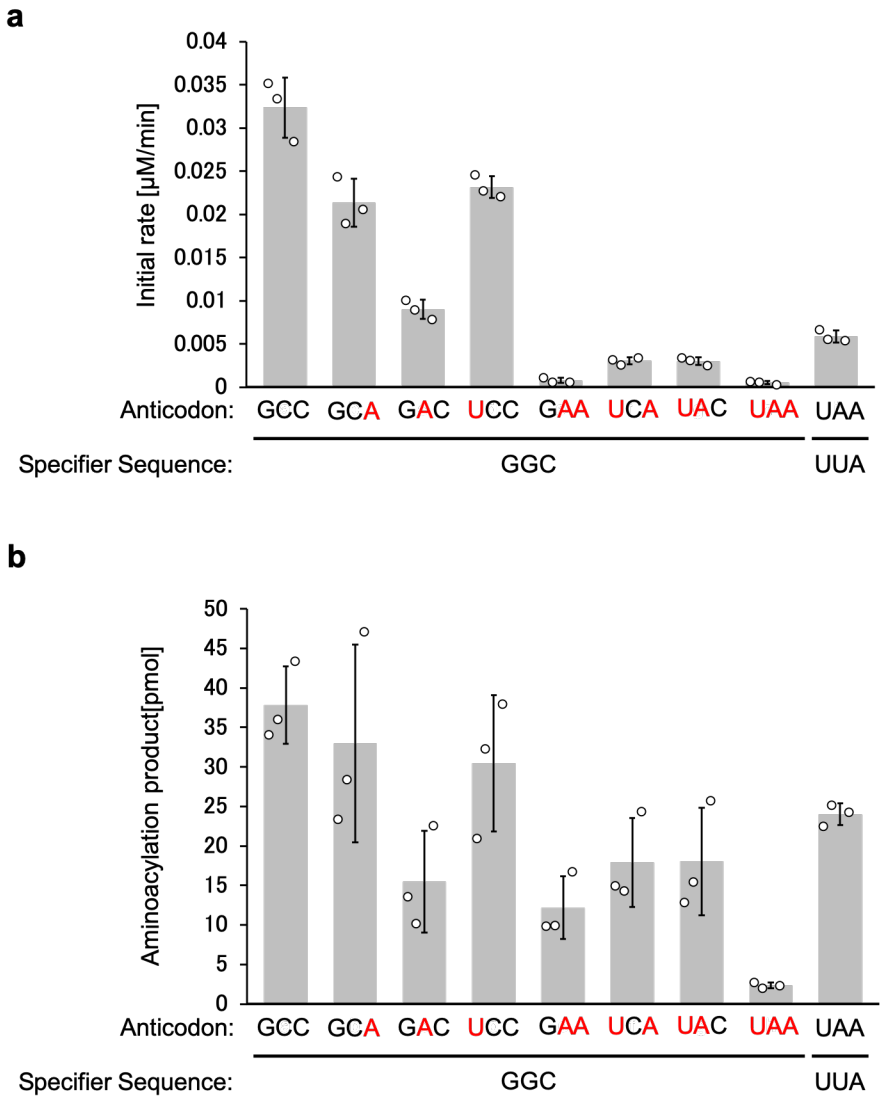

Figure 2.13 | Anticodon specificity analysis for Tx2.1
a) Apparent initial rates for trans-aminoacylation of Tx2.1 against various anticodon mutant tRNAs. Red letters denote the mismatched positions. Each data point represents the mean of three independent experiments ± s.d. (Standard deviation). b) Endpoint analysis for transaminoacylation against various anticodon mutant tRNAs using Tx2.1. Red letters denote the mismatched positions. Each data point represents the mean of three independent experiments ± s.d. (Standard deviation).

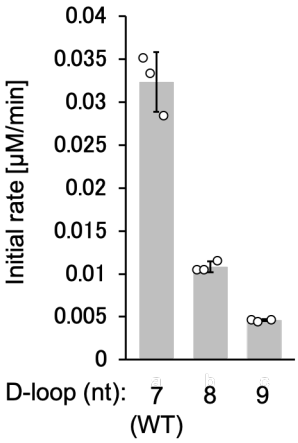

Figure 2.14 | The effect of D-loop size in the aminoacylation of Tx2.1
Each data point represents the mean of three independent experiments ± s.d. (Standard deviation).

2.2.7 Validation of Tx2.1 secondary structure

Figure 2.15 shows the predicted secondary structure of Tx2.1 randomized domain produced by CentroidFold[72]. To validate this secondary structure of Tx2.1, deep mutational scanning of Tx2.1 was conducted[73-75]. Using this technique, the effect of mutation at each position of the catalytic core to the ribozyme activity can be assessed. Single point mutation of Tx2.1-tRNA at every position of the Tx2.1 catalytic domain was prepared and mixed together with the wildtype sequence and Tx2.1-tRNA library was constructed. This library was then mixed with bio-^{L}Phe-CME and mutants that retained their activity were selectively recovered using streptavidin magnetic beads. The recovered library (an output library) as well as Tx2.1-tRNA library before mixing with bio-^{L}Phe-CME (an input library) was reverse transcribed and deep sequencing was conducted. Using the sequencing data, the enrichment score (E) for each mutant was calculated[76] and compared with that of the wild type to give W_{socre}. This score was plotted as a heatmap to see the tolerance of each mutation towards its activity (Figure 2.16a). The heatmap revealed that the bases in two regions, 150A to 154A and 170A to 174C, were tolerated whereas other bases were indispensable because these positions were possibly either structurally and/or functionally important.

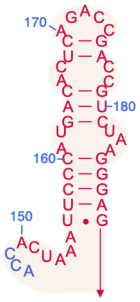

Figure 2.15 | Possible secondary structure of Tx2.1 random domain
Red number nucleotides correspond to the random region. Blue number denotes the nucleotide number in terms of Tx2.1.

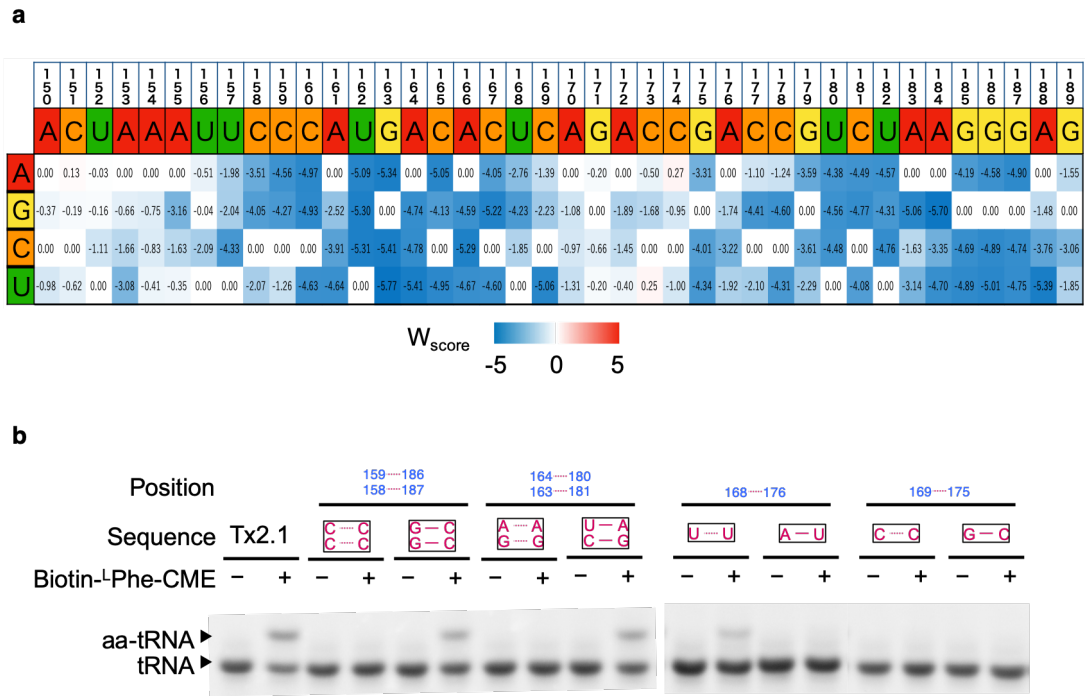

Figure 2.16 | Mutational scanning of Tx2.1 random domain
a) Heatmap of W score for all mutants tested in the deep mutational scanning. Top lane corresponds to the nucleotide position in Tx2.1, followed by the actual nucleotide sequence in each position. b) Mutational study to check the base paring in predicted secondary structure in Figure 2.15.

This trend actually is in accordance with the predicted secondary structure since both 150A to 154A and 170A to 174C regions correspond as loops while other bases are forming base-paring or positioned at internal small loops between stems. To further confirm the predicted base-pairing, ribozyme mutants having base-disrupting or base-flipping mutations in the predicted base-pairing position at C158C159/G187G186, G163A164/C181U180, U168–A176, or C169–G175 were prepared and checked for their activities. For C158C159/G187G186 and G163A164/C181U180 regions, the loss and recovery of activity for the corresponding mutants were observed confirming the formation of base-paring. On the other hand, for U168–A176 base pair, the disruptive mutation retained weak aminoacylation activity while the flipping mutation abolished the activity, suggesting that either U168–A176 is not formed or this base-paring is indispensable for the activity. Furthermore, for C169–G175 base pair, both disruptive and flipping mutation

lost its activity, suggesting that this base-paring is also indispensable. I also chose representative two tolerant mutations, C151A and C174A, and two deleterious mutations, A166C and A184G, based on the W_{score} mapping and checked whether these mutants would show the retainment or elimination of their activity as seen in the deep mutational scanning (Figure 2.17). As a result, these mutations did show the expected retainment or loss of activity. These results strongly support that the predicted secondary structure is the most plausible one and the total sequence of Tx2.1 is presented in Figure 2.18.

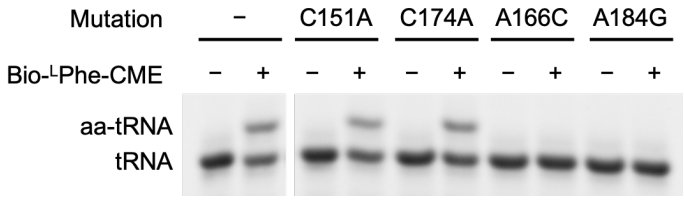

Figure 2.17 | Trans-aminoacylation activity of representative point mutation.
C151A, C174A, A166C and A184G was chosen as representative point mutations for either acceptable or deleterious mutation in terms of aminoacylation activity.

Since the base-paring position has been determined, another set of mutational study was conducted. The purpose of this mutational study was to evaluate which types of base-paring does Tx2.1 tolerate. Since there are ten base-paring positions in Tx2.1 random domain, Tx2.1-tRNA sequences whose two nucleotides corresponding to each base-pair position were completely randomized and mixed together (Figure 2.19). This gave another set of Tx2.1-tRNA library and the same procedure of mixing the library with bio-^{L}Phe-CME followed by reverse transcription and deep sequencing was conducted. The enrichment score (E) as well as W_{score} was again calculated and plotted as a heatmap (Figure 2.19b). For base-pair position 1, almost all type of nucleotide composition was tolerated. For base-pair position 2 and 3, as long as this position is forming base-pair either through C-G, A-U or G-U wobble base-paring, the mutant seemed to retain aminoacylation activity. For base-pair position 4, the original pair was C-G and flipping these bases as well as substituting with U-A base pair or G-U wobble base-pair seemed to retain

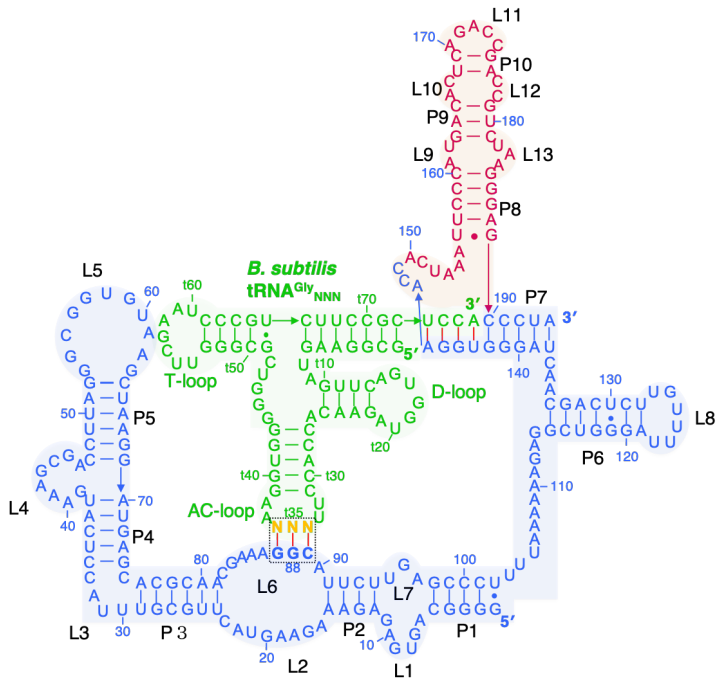

Figure 2.18 | Secondary structure of Tx2.1
Blue nucleotides denote T-box body sequence, green nucleotides correspond to tRNA and red letter denote for random region. For the numbering of tRNA, small "t" is placed before the number.

its activity except for the case of A159-U185 substitution. Interestingly, for base-pair position 5, although the original pair is C160-G184, only the A160-U184 was tolerated. For base-pair position 6, flipping the two nucleotides pattern was accepted while all types of base-paring except for wobble base-paring was tolerated for base-pair position 7 and 8. For base-pair position 9, as long as position 168 is U, it retained aminoacylation activity. For the last position 10, most combinations were not tolerated except for A169, G175 combination. This knowledge would be beneficial for further engineering of Tx2.1 aminoacylation activity in the future experiments.

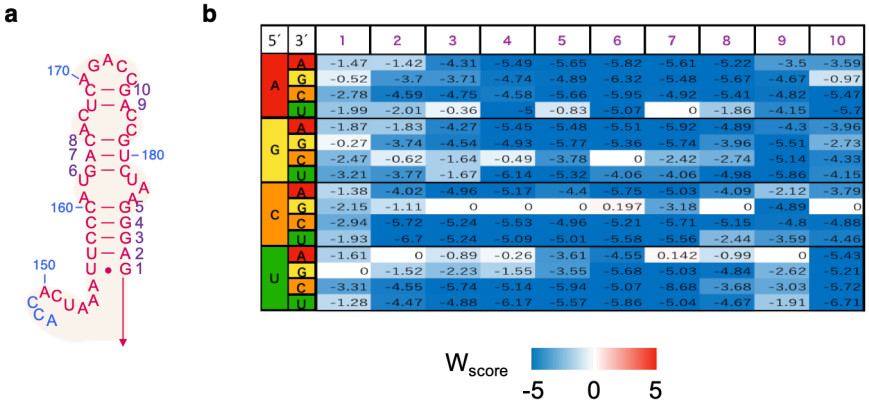

Figure 2.19 | Deep mutational scanning for base-paring formation
a) Secondary structure of Tx2.1 random domain with purple numbers indicating the assigned base-paring positions. b) Heatmap of Wscore calculated from deep sequencing.

2.2.8 *In situ* aminoacylaiton and translation using Tx2.1

To check whether aminoacylation by Tx2.1 can function in parallel with translation, I coupled the Tx2.1 ribozyme with the *in vitro* translation system. The expression of a model peptide was evaluated in the presence of Tx2.1, $tRNA^{Gly}_{GCC}$, Bio-^{L}Phe-CME, and a DNA template within our FIT (Flexible In-vitro Translation) cocktail[77] (Figure 2.20a). If Tx2.1 can aminoacylate $tRNA^{Gly}_{GCC}$ using Bio-^{L}Phe-CME in the FIT cocktail, a model peptide initiating with Bio-^{L}Phe should be observed. Since Tx2.1 was most active toward tRNA bearing GCC anticodon, the start codon was engineered to GGC rather than the canonical AUG to initiate the translation with Bio-^{L}Phe (Figure 2.20a), following the approach previously reported in our group using FIT system[78]. Therefore, the DNA template used in this study had a T7 promoter sequence followed by Shine-Dalgarno sequence and GGC as the start codon. After the GGC start codon, the DNA template coded a model peptide sequence of KKKDYKDDDDK (the underlined sequence is Flag tag, Figure 2.20b).

The T-boxzyme/tRNA components included the DNA template (for peptide expression), $tRNA^{Gly}_{GCC}$, Tx2.1, and Bio-LPhe-CME. The components in the FIT system included the ribosome, T7 RNA polymerase, translation factors, total ARSs with limited amino acids (Lys, Asp, and Tyr), total E. coli tRNAs, dNTPs, NTPs, and other factors required for translation. First, aminoacylation of $tRNA^{Gly}_{GCC}$ using flexizyme was conducted and Bio-^{L}Phe-$tRNA^{Gly}_{GCC}$ was isolated via ethanol precipitation[77]. The prepared aminoacylated tRNA was then added to the FIT component along with the DNA template and the expression of target peptide, Bio-LPhe-KKK-Flag, was confirmed by MALDI-TOF-MS analysis. This indicates that Bio-^{L}Phe-$tRNA^{Gly}_{GCC}$ could be used as an initiator in the FIT system (Ctrl2 in Figure 2.21c).

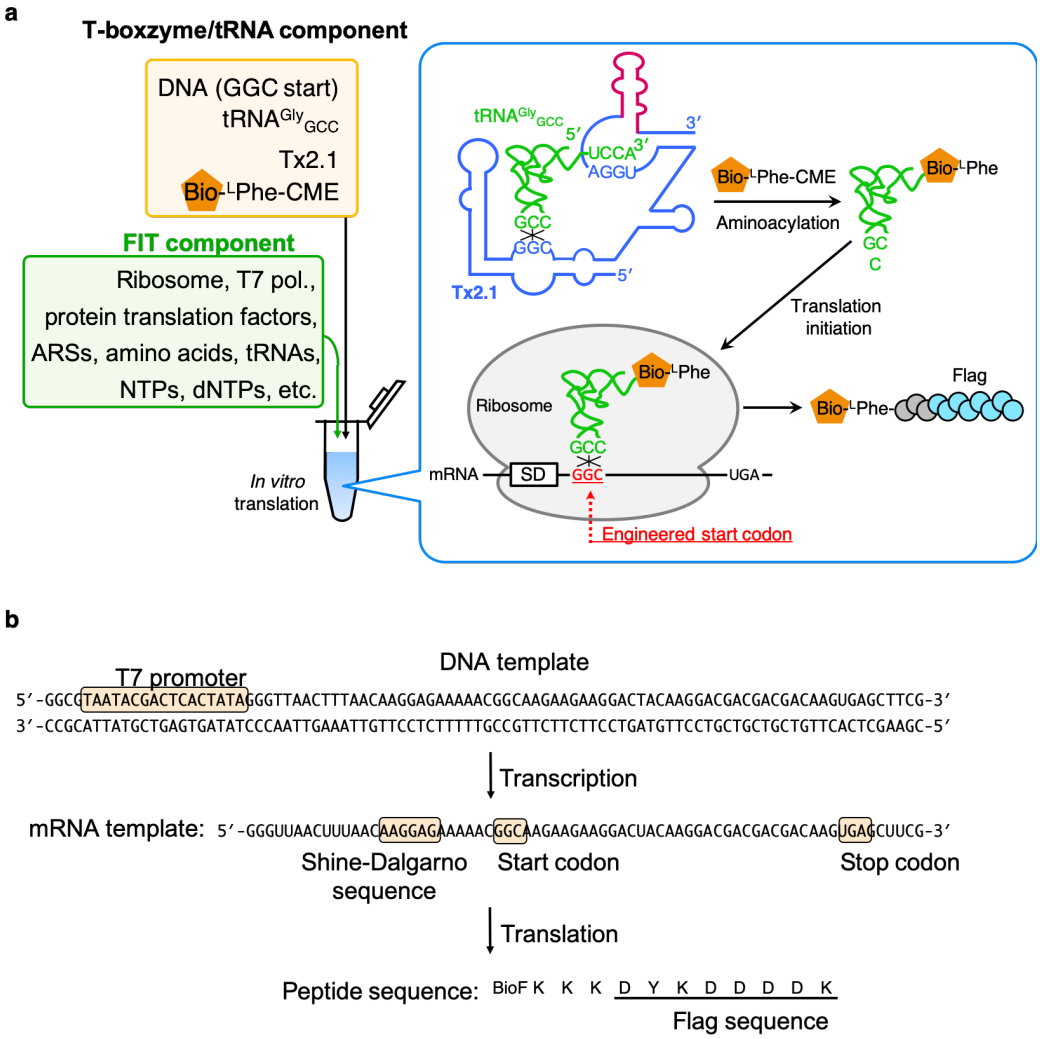

Figure 2.20 | *In situ* aminoacylation and translation by coupling Tx2.1 and FIT system
a) Schematic representation of the in situ aminoacylation/translation experiment conducted in this study. In addition to the normal components of the *in vitro* translation mixture with limited amino acids (covered in green), Tx2.1, $tRNA^{Gly}_{GCC}$, Bio-^{L}Phe-CME, and a DNA template bearing a GCC initiation codon were added (covered in orange). Rectangle SD denotes the Shine-Dalgarno sequence in the mRNA. b) The design of DNA template used in this study. The transcribed RNA as well as the sequence of desired model peptide is also presented.

Next, transcription/translation of the target peptide from the DNA template coupled with the *in situ* Tx2.1-catalyzed aminoacylation of $tRNA^{Gly}_{GCC}$ was attempted. The peptide was expressed by mixing both the FIT and T-boxzyme/tRNA components, where all tRNA charging events, including ARS-catalyzed tRNA aminoacylations and Tx2.1-catalyzed aminoacylation of $tRNA^{Gly}_{GCC}$ with Bio-^{L}Phe occurred simultaneously. The

expression level of the translated peptide was evaluated by autoradiographic detection of [^{14}C]-Asp on Tricine SDS-PAGE gel (Figure 2.21a). Two positive control sample was used in the Tricine SDS-PAGE experiment for band assignment. The first control is the peptide initiating with the second amino acid and the second control is the peptide initiating with Bio-^{L}Phe which was prepared by pre-charging tRNA$^{Gly}_{GCC}$ with flexizyme and adding to the FIT component. To check the background level of peptide expression, four types of control experiments were conducted that lack the substrate and/or Tx2.1 (Figure 2.21a lane 1-4). A faint band was observed only when both Bio-^{L}Phe-CME and tRNA$^{Gly}_{GCC}$ judging from the Tricine SDS-PAGE gel but MALDI-TOF-MS analysis of non-radiolabeled peptide expressed under identical conditions revealed that the model peptide was expressed in all controls that included Bio-^{L}Phe-CME (Figure 2.21 lane 2-4). This suggests that Bio-^{L}Phe-CME can non-enzymatically react with tRNA$^{Gly}_{GCC}$ at background level. An additional control was conducted where catalytically inactive but "binding active" *glyQS* T-box riboswitch was mixed with Bio-^{L}Phe-CME and tRNA$^{Gly}_{GCC}$ (Figure 2.21a lane 5). The expression level of the peptide in lane 5 decreased to the background level, implying that the above background aminoacylation was inhibited by binding of *glyQS* T-box to tRNA$^{Gly}_{GCC}$. However, when Tx2.1 was added along with Bio-^{L}Phe-CME and tRNA$^{Gly}_{GCC}$, the expression of the full-length peptide increased more than 6-fold than lane 4 and more than 15-fold compared to lane 5. The peptide identity in lane 6 was confirmed by MALDI-TOF-MS clearly showing that peptide initiating with Bio-^{L}Phe was indeed expressed (Figure 2.21c).

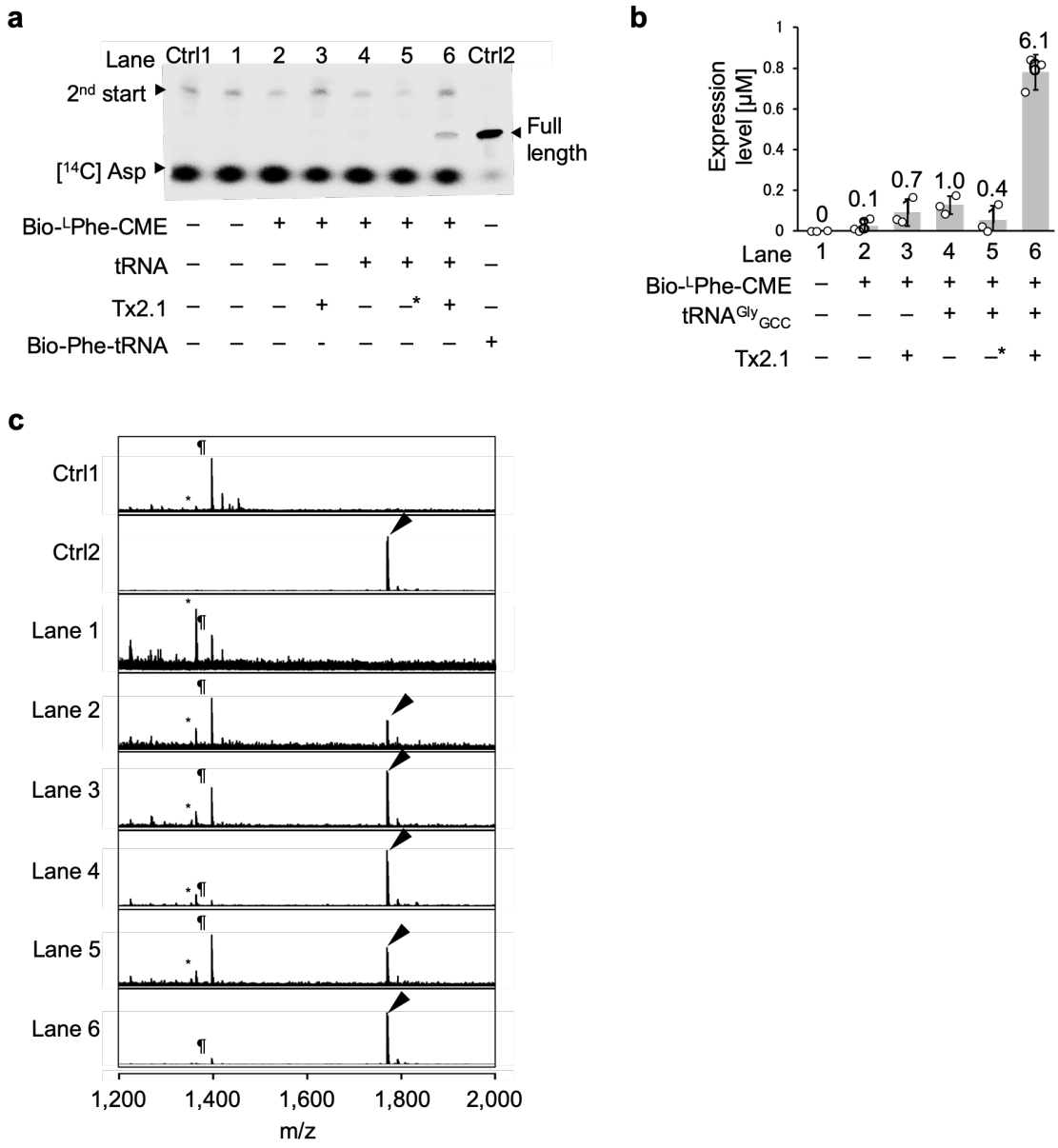

Figure 2.21 | Gel and MALDI spectra of in situ aminoacylation and translation experiments.

a) Tricine-SDS PAGE gel image of *in vitro* translation experiments under various conditions. Ctrl1 indicates the control experiment of translated peptide initiating from the second amino acid. Ctrl2 indicates the control experiment of full-length peptide by adding pre-charged Bio-^{L}Phe-tRNA prepared by eFx. In lane 5, the glyQS T-box riboswitch was added instead of Tx2.1, which is represented as –*. b) Quantification of full length product from each lane in a) except for the Ctrl lanes. c) MALDI-TOF-MS spectra of translation products of each lane from a. Black triangle indicates the full-length product. ¶ denotes the peptide initiating from the second amino acid. * denotes background peak present in the translation system.

2.3 Conclusion

In this chapter, using T-box motif as the base structure, an aminoacylation ribozyme, Tx2.1 which retains tRNA specificity was obtained. On the course of this discovery, flexizyme derivative ribozyme, Tx1.3 was also obtained but since the *in vitro* selection was conducted with the same amino acid substrate as with the discovery of flexizyme prototype, it is to some extent expected. For Tx2.1, the aminoacylation reaction was clearly anticodon dependent, suggesting that this ribozyme does recognize the anticodon triplet upon aminoacylation. Although tRNA whose anticodon does not base pair at all with the specifier sequence of T-box domain, was not aminoacylated, this reaction efficiency can be recovered by introducing compensatory mutations to the T-box specifier sequence to completely match with substrate tRNA anticodon. Because the aminoacyl-tRNA prepared by this ribozyme can be used for translation reaction with the ribosome, taken together with the ease of tuning tRNA specificity of Tx2.1 by introducing mutations to the specifier sequence, Tx2.1 could become an excellent aminoacylation ribozyme for genetic code reprogramming. However, there is still a lot to learn from this ribozyme such as how Tx2.1 specifically recognize the amino acid substrate, biotin-^{L}Phe-CME. In order to elucidate this question, crystal structure study is in high demand. Currently, a collaboration with a laboratory specialized in RNA crystallography is underway. The high specificity toward biotin-^{L}Phe-CME is a characteristic of this Tx2.1 the resemble the natural amioacyl-tRNA synthase which also recognizes the amino acid substrate in high specificity. However, the biotin residue attached to the backbone amine hinders the use to apply genetic code reprogramming at the elongation position. In order to obtain another ribozyme with other amino acid selectivity, another approach is needed and this is discussed in the next chapter.

2.4 Materials and methods

2.4.1 Materials

All DNA oligomers used were purchased from Eurofin Genomics except from oligomers having 2′-*O*-Methyl guanosine which was purchased from Gene Design.

Here, G(M) denotes 2′-*O*-methylguanosine and N denotes all nucleotides(A,T,G,C).

Name	Sequence (5′-3′)	Used for
Oligo 1	GTACTTGCGTTTACCTCATGAAAGCGACCTTAGGGCGGTGTAAGCTAAGGATGAGCACGCAACGAAAGGCATTCTTGAGCCCTTTTAAAAAAGAGGCTGGGATTTTGTTCTCAGCAACTAGGGTGGAACC	Library construction
Oligo 2	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTAGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGTTCCACCCTAGTTGCTGA	Library construction
Oligo 3	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGTTCCACCCTAGTTGCTGA	Library construction
Oligo 4	GGCGTAATACGACTCACTATAGGGGCAGTGAGAGAAAGAAGTACTTGCGTTTACCTCATGAAAG	Library construction
Oligo 5	TG(M)GAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTGGCAAGGTGGT	Library construction
Oligo 6	TG(M)GAGCGGAAGACGGG	RT PCR
Oligo 7	GGCGTAATACGACTCACTATAGGGGCAGTGAGAGAAAGAA	PCR amplification
Oligo 8	GCCTCTTCGCTATTACGCCAGC	Sequencing
Oligo 9	TGTTGTGTGGAATTGTGAGCGG	Sequencing
Oligo 10	CTTAGGGCGGTGTAAGCTAAGGATGAGCACGCAACGAAAGGCAT	Tx1.3 and Tx2.1
Oligo 11	GCGTTTACCTCATGAAAGCGACCTTAGGGCGGTGTAAGCT	Tx1.3 and Tx2.1
Oligo 12	GCAGTGAGAGAAAGAAGTACTTGCGTTTACCTCATGAAAGC	Tx1.3 and Tx2.1
Oligo 13	GGCGTAATACGACTCACTATAGGGGCAGTGAGAGAAAGAAGTAC	Tx1.3 and Tx2.1
Oligo 14	GCCTCTTTTTTAAAAGGGCTCAAGAATGCCTTTCGTTGCGTGCT	Tx1.3 and Tx2.1
Oligo 15	TAGTTGCTGAGAACAAAATCCCAGCCTCTTTTTTAAAAGGGCTC	Tx1.3 and Tx2.1
Oligo 16	CGCTTGCGTAGGGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	Tx1.3
Oligo 17	CGCCGGTGCTAACCTTGTCGGTTCGCTTGCGTAGGGGTTC	Tx1.3
Oligo 18	TCCTGTCGCCGGTGCTAACCTTG	Tx1.3
Oligo 19	TGGGAATTTAGTGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	Tx2.1
Oligo 20	CGGTCGGTCTGAGTGTCATGGGAATTTAGTGGTTCCAC	Tx2.1

Name	Sequence (5′-3′)	Used for
Oligo 21	TAGGGCTCCCTTAGACGGTCGGTCTGAGTGTCA	Tx2.1
Oligo 22	AGAGTTCTCCCGCGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	glyQS T-box riboswitch
Oligo 23	TAGGGACGAGAGTTCTCCCGCGGTTC	glyQS T-box riboswitch
Oligo 24	CTTAGGGCGGTGTAAGCTAAGGATGAGCACGCAACGAAATTAAT	Tx2.1 (UUA)
Oligo 25	GCCTCTTTTTTAAAAGGGCTCAAGAATTAATTTCGTTGCGTGCT	Tx2.1 (UUA)
Oligo 26	GGCGTAATACGACTCACTATAG	Tx, tRNA and DNA template
Oligo 27	GAAGTAGTTCAGTGGTAGAACACCACCTTGCCAAGGTGGGGGTC	$tRNA^{Gly}_{GCC}$
Oligo 28	GGCGTAATACGACTCACTATAGCGGAAGTAGTTCAGTGGTAGAA	$tRNA^{Gly}$
Oligo 29	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTGGCA	$tRNA^{Gly}_{GCC}$
Oligo 30	TG(M)GAGCGGAAGACGGGA	$tRNA^{Gly}$
Oligo 31	GAAGTAGTTCAGTGGTAGAACACCACCTTTAAAAGGTGGGGGTC	$tRNA^{Gly}_{UAA}$
Oligo 32	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTTTAA	$tRNA^{Gly}_{UAA}$
Oligo 33	GAAGTAGTTCAGTGGTAGAACACCACCTTGCAAAGGTGGGGGTC	$tRNA^{Gly}_{GCA}$
Oligo 34	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTTGCA	$tRNA^{Gly}_{GCA}$
Oligo 35	GAAGTAGTTCAGTGGTAGAACACCACCTTGACAAGGTGGGGGTC	$tRNA^{Gly}_{GAC}$
Oligo 36	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTGTCA	$tRNA^{Gly}_{GAC}$
Oligo 37	GAAGTAGTTCAGTGGTAGAACACCACCTTTCCAAGGTGGGGGTC	$tRNA^{Gly}_{UCC}$
Oligo 38	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTGGAA	$tRNA^{Gly}_{UCC}$
Oligo 39	GAAGTAGTTCAGTGGTAGAACACCACCTTGAAAAGGTGGGGGTC	$tRNA^{Gly}_{GAA}$
Oligo 40	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTTTCA	$tRNA^{Gly}_{GAA}$
Oligo 41	GAAGTAGTTCAGTGGTAGAACACCACCTTTCAAAGGTGGGGGTC	$tRNA^{Gly}_{UCA}$
Oligo 42	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTTGAA	$tRNA^{Gly}_{UCA}$

Oligo 43	GAAGTAGTTCAGTGGTAGAACACCACCTTTACAAGGTGGGGGTC	$tRNA^{Gly}_{UAC}$
Oligo 44	TGGAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTGTAA	$tRNA^{Gly}_{UAC}$
Oligo 45	AAGTAGTTCAGCTGGTAGAACACCACCTTGCCAAGGTGGGGGTC	$tRNA^{Gly}_{GCC}$ 8nt d-loop
Oligo 46	GGCGTAATACGACTCACTATAGCGGAAGTAGTTCAGCTGGTAGA	$tRNA^{Gly}_{GCC}$ 8nt d-loop
Oligo 47	AGTAGTTCAGCCTGGTAGAACACCACCTTGCCAAGGTGGGGGTC	$tRNA^{Gly}_{GCC}$ 9nt d-loop
Oligo 48	GGCGTAATACGACTCACTATAGCGGAAGTAGTTCAGCCTGGTAG	$tRNA^{Gly}_{GCC}$ 9nt d-loop
Oligo 49	GGTCGGTCTGAGTGTCATGGGAATTTAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 50	GGTCGGTCTGAGTGTCATGGGAATTTAGVGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 51	GGTCGGTCTGAGTGTCATGGGAATTTAHTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 52	GGTCGGTCTGAGTGTCATGGGAATTTBGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 53	GGTCGGTCTGAGTGTCATGGGAATTVAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 54	GGTCGGTCTGAGTGTCATGGGAATVTAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 55	GGTCGGTCTGAGTGTCATGGGAAVTTAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 56	GGTCGGTCTGAGTGTCATGGGABTTTAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 57	GGTCGGTCTGAGTGTCATGGGBATTTAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 58	GGTCGGTCTGAGTGTCATGGHAATTTAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 59	GGTCGGTCTGAGTGTCATGHGAATTTAGTGGTTCCACCCTAGTTGC TGA	Deep sequencing
Oligo 60	CTCCCTTAGACGGTCGGTCTGAGTGTCATHGGAATTTAGTGGTTCC ACCC	Deep sequencing
Oligo 61	CTCCCTTAGACGGTCGGTCTGAGTGTCAVGGGAATTTAGTGGTTCC ACCC	Deep sequencing
Oligo 62	CTCCCTTAGACGGTCGGTCTGAGTGTCBTGGGAATTTAGTGGTTCC ACCC	Deep sequencing
Oligo 63	CTCCCTTAGACGGTCGGTCTGAGTGTDATGGGAATTTAGTGGTTCC ACCC	Deep sequencing
Oligo 64	CTCCCTTAGACGGTCGGTCTGAGTGVCATGGGAATTTAGTGGTTCC ACCC	Deep sequencing
Oligo 65	CTCCCTTAGACGGTCGGTCTGAGTHTCATGGGAATTTAGTGGTTCC ACCC	Deep sequencing
Oligo 66	CTCCCTTAGACGGTCGGTCTGAGVGTCATGGGAATTTAGTGGTTCC ACCC	Deep sequencing

Name	Sequence (5′-3′)	Used for
Oligo 67	CTCCCTTAGACGGTCGGTCTGAHTGTCATGGGAATTTAGTGGTTCCACCC	Deep sequencing
Oligo 68	CTCCCTTAGACGGTCGGTCTGBGTGTCATGGGAATTTAGTGGTTCCACCC	Deep sequencing
Oligo 69	CTCCCTTAGACGGTCGGTCTHAGTGTCATGGGAATTTAGTGGTTCCACCC	Deep sequencing
Oligo 70	TTTTTTTTAGGGCTCCCTTAGACGGTCGGTCVGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 71	TTTTTTTTAGGGCTCCCTTAGACGGTCGGTDTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 72	TTTTTTTTAGGGCTCCCTTAGACGGTCGGVCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 73	TTTTTTTTAGGGCTCCCTTAGACGGTCGHTCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 74	TTTTTTTTAGGGCTCCCTTAGACGGTCHGTCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 75	TTTTTTTTAGGGCTCCCTTAGACGGTDGGTCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 76	TTTTTTTTAGGGCTCCCTTAGACGGVCGGTCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 77	TTTTTTTTAGGGCTCCCTTAGACGHTCGGTCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 78	TTTTTTTTAGGGCTCCCTTAGACHGTCGGTCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 79	TTTTTTTTAGGGCTCCCTTAGADGGTCGGTCTGAGTGTCATGGGAATTTAGTGG	Deep sequencing
Oligo 80	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCCTTAGBCGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 81	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCCTTAHACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 82	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCCTTBGACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 83	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCCTVAGACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 84	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCCVTAGACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 85	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCDTTAGACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 86	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCDCTTAGACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 87	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTDCCTTAGACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 88	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCVCCCTTAGACGGTCGGTCTGAGTGTCAT	Deep sequencing
Oligo 89	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGDTCCCTTAGACGGTCGGTCTGAGTGTCAT	Deep sequencing

Name	Sequence (5′-3′)	Used for
Oligo 90	TCAGCAACTAGGGTGGAACCACTAAATTCC	Deep sequencing
Oligo 91	TCAGCAACTAGGGTGGAACCACTAAATTCCCATGACACTC	Deep sequencing
Oligo 92	TCAGCAACTAGGGTGGAACCACTAAATTCCCATGACACTCAGACC GACCG	Deep sequencing
Oligo 93	TACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCCTTA GACGGTCGGTCTGAGTGTCATG	Deep sequencing
Oligo 94	CTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTTAGGGCTCCCTT AGACGGTCGGTCT	Deep sequencing
Oligo 95	CCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTT TTTTTTTAGGGCTCCCTTAGA	Deep sequencing
Oligo 96	CCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGC	Deep sequencing
Oligo 97	GAGGCTGGGATTTTGTTCTCA	scanning (RT-PCR)
Oligo 98	GGTGGTGTTCTACCACTGAACTACTT	scanning (RT-PCR)
Oligo 99	CACTCTTTCCCTACACGACGCTCTTCCGATCTGAGGCTGGGATTTT GTTCTCAGCA	scanning (for miseq)
Oligo 100	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGTGGTGTTCTACC ACTGAACTACTTCCG	scanning (for miseq)
Oligo 101	GGTGGAACCCCTACGCAAGCGAACCGACAAG	$Tx1.3_{140\text{-}190}$
Oligo 102	GGCGTAATACGACTCACTATAGGGGTGGAACCCCTACGCAA	$Tx1.3_{140\text{-}190}$
Oligo 103	TCCTGTCGCCGGTGCTAACCTTGTCGGTTCGCTTGC	$Tx1.3_{140\text{-}190}$
Oligo 104	TCCTGTCGCCGGTGCT	$Tx1.3_{140\text{-}190}$
Oligo 105	GTGGAACCACTAAATTCCCATGACACTCAGACCG	$Tx2.1_{140\text{-}194}$
Oligo 106	GGCGTAATACGACTCACTATAGGGTGGAACCACTAAATTCCCA	$Tx2.1_{140\text{-}194}$
Oligo 107	TAGGGCTCCCTTAGACGGTCGGTCTGAGTGTCATGGG	$Tx2.1_{140\text{-}194}$
Oligo 108	TAGGGCTCCCTTAGACGG	$Tx2.1_{140\text{-}194}$
Oligo 109	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACGGC	DNA template
Oligo 110	GTCGTCGTCCTTGTAGTCCTTCTTCTTCGGGTTTTTCTC	DNA template
Oligo 111	CGAAGCTCACTTGTCGTCGTCGTCCTTGTAGTC	DNA template
Oligo 112	GTAATACGACTCACTATAGGATCGAAAGATTTCCGC	eFx
Oligo 113	ACCTAACGCTAATCCCCTTTCGGGGCCGCGGAAATCTTTCGATCC	eFx
Oligo 114	GGCGTAATACGACTCACTATAG	eFx
Oligo 115	ACCTAACGCTAATCCCCT	eFx

2.4.2 Synthesis of amino acid substrates.

Bio-^{L}Phe-CME was synthesized as previously described[38]. For the study of amino acid selectivity of Tx2.1 in aminoacylation, three bio-aa-CMEs (^{D}Phe, ^{L}Tyr, Gly) were synthesized following the same procedure as bio-^{L}Phe-CME.

2.4.3 *In vitro* transcription of RNAs

The RNAs were prepared by in vitro transcription with T7 RNA polymerase and purified by denaturing PAGE (8M Urea, 1 × TBE). DNA templates were modified with 2′-O-methylation at the second last nucleotide of the 5′ termini to reduce non-templated nucleotide addition by T7 RNA polymerase39. The primers for preparing transcription templates and RNAs are shown in the material. All non-methylated primers were purchased from Operon Biotechnology (Japan) and methylated primers were purchased from Gene Design (Japan). The concentrations of RNAs were determined by absorbance at 260 nm.

2.4.4 Sequencing

The cDNAs of RNAs after in vitro selection were cloned into pGEM®-T Easy Vectors (A1360, Promega K.K.). These vectors were used to transform E. coli DH5α cells in LB medium with 100 μg/mL ampicillin. Inserted cDNAs in the vectors were amplified by colony-PCR using primers oligo 8 and oligo 9, and sequenced from both 5′- and 3′-end by FASMAC Co., Ltd. (Japan) Preparation of DNA templates for *in vitro* translation

2.4.5 Aminoacylation

Aminoacylation reactions by flexizymes were performed as previously described[15]. The conditions of self-aminoacylation by T-box ribozymes are the same as those described in a method of *in vitro* selection. For trans-aminoacylation, 83.3 μM tRNA and 166.7 μM Tx were first heated at 95°C for 2 min and cooled on ice over 5 min. This RNA

solution was mixed 66.7 % volume of 25 mM amino acid substrates in DMSO followed by equal volume of 2×reaction buffer [100 mM HEPES-KOH (pH 7.5), 1 M KCl and 20 mM $MgCl_2$]. The reaction mixture was incubated on ice for 2 hours. After the reaction, aminoacyl-RNA was precipitated with ethanol as previously described[15].

2.4.6 Analysis of acylation using streptavidin

aa-RNA which was ethanol precipitated was dissolved in SAv-shift buffer [4 M Urea, 1 mM EDTA, 1 mM Tris, 333 μM NaOAc (pH 5.3) and 10 mg/mL Streptavidin] and then loaded on urea 6% polyacrylamide gels (8 M urea).

For the analysis of acylation with Phe-CME, ethanol-precipitated aa-RNA was first dissolved in 1 mM NaOA (pH 5.3) and then 300 mM HEPES-KOH (pH 8.0) was added. Biotin-3-sulfo-N-hydroxylsuccinimide ester was immediately added at 4°C with the finial concentration of 15mM. After 1 hour, the biotinylation reaction was stopped by ethanol precipitation twice. Ethanol-precipitated bio-aa-RNA was then dissolved in loading buffer [4 M Urea, 1 mM EDTA, 1 mM Tris, 333 μM NaOAc (pH 5.3)] and heated for 30 seconds at 95°C, and then cooled to 25°C. After addition of SAv solution (10 mg/mL), the sample was analyzed by urea 6% polyacrylamide gels (8 M urea).

2.4.7 Preparation of 3′-dial-RNA

Ribose at the 3′-end of RNA was oxidized to dialdehyde group (named as dial-RNAs) under the following conditions; 20 μM RNA was incubated with 30 mM $NaIO_4$ on ice for 60 min and purified by NAP-5 column (17085301, GE Healthcare) with NAP-5 Buffer [100 mM NaOAc (pH 5.3) and 10 mM $MgCl_2$] following manufacture's protocol. The eluted RNA solution was precipitated with ethanol and dissolved in water. The concentrations of RNAs were determined by absorbance at 260 nm.

2.4.8 Analysis of acylation by acid-PAGE

Ethanol-precipitated aa-tRNA was dissolved in acid-PAGE loading buffer (150 mM NaOAc, pH 5.2, 10 mM EDTA, and 93% (v/v) formamide) and then loaded on acid-urea 12% polyacrylamide gels (8 M urea, 50 mM NaOAc, pH 5.2). Electrophoresis was performed using 300 V (approximately 20 Vcm^{-1}) for 6 h at 4°C. The gels were stained with ethidium bromide and analyzed using a Typhoon FLA 7000 fluorescent image analyzer (Fujifilm). Aminoacylation efficiency was calculated based on the band intensity of aa-tRNA (A) and free tRNA (T) and is obtained as (A)/[(A) + (T)].

2.4.9 Measurement of initial aminoacylation rate

RNA used in the measurement of initial aminoacylation rate was transcribed *in vitro* in the presence of [α-^{32}P]UTP for body-radiolabeling. The conditions of self-aminoacylation were the same as those described in the method of *in vitro* selection except for the pH of the buffer, which was changed to pH 8.5 from 7.5. At each time point, an aliquot of the reaction was mixed to ethanol precipitation solution to stop the reaction. The solution was ethanol precipitated twice and the pellet was dissolved in SAv-shift buffer [4 M Urea, 1 mM EDTA, 1 mM Tris, 333 μM NaOAc (pH 5.3) and 10 mg/mL Streptavidin]. The sample was loaded on urea 6% polyacrylamide gels (8 M urea) and run in 4°C at 230 V for 90 min. The gel was exposed to IP cassette (Fujifilm) and analyzed by Typhoon FLA 7000 fluorescent image analyzer (Fujifilm). The initial rate at each substrate concentration was determined by a slope of the liner region in a plot of time versus fraction of product with five time points. Data were fitted to the Michaelis-Menten non-linear curve by KaleidaGraph (HULINKS).

2.4.10 Measurement of apparent initial aminoacylation rate and endpoint aminoacylation yield against various tRNA mutants.

tRNA mutants used in the measurement of apparent initial aminoacylation rate and endpoint aminoacylation yield was transcribed *in vitro* in the presence of [α-^{32}P]UTP for

body-radiolabeling. The condition in this experiment was as follow: 50 mM HEPES-KOH (pH 8.5), 500 mM KCl, 10 mM $MgCl_2$, 10 μM Tx2.1, 1 μM tRNA mutant. At each time point, an aliquot of the reaction was mixed to ethanol precipitation solution to stop the reaction. The solution was ethanol precipitated twice and the pellet was dissolved in SAv-shift buffer [4 M Urea, 1 mM EDTA, 1 mM Tris, 333 μM NaOAc (pH 5.3) and 10 mg/mL Streptavidin]. The sample was loaded on urea 8% polyacrylamide gels (8 M urea) and run in 4°C at 230 V for 40 min. The gel was exposed to IP cassette (Fujifilm) and analyzed by Typhoon FLA 7000 fluorescent image analyzer (Fujifilm). The initial rate at each substrate concentration was determined by a slope of the liner region in a plot of time versus fraction of product with five time points. For endpoint aminoacylation yield, the aminoacylation time was 2.5 hrs and for each tRNA mutant the experiment was conducted three times.

2.4.11 Deep mutational scan

DNA templates for the transcription of the RNA point mutant library were prepared by 4-step PCR, using oligos 49-96. Oligo 49-59, oligo 60-69, oligo 70-79, and oligo 80-89 were mixed equally to prepare mixed primer A, B, C, and D respectively. Extension reaction was carried out with the following combination in 50 μL scale (oligo 1 and primer A, oligo 90 and primer B, oligo 91 and primer C, and oligo 92 and primer D). 1/20 of the extension product was PCR amplified using oligo 4 and 93 for the primer A extension product, oligo 1 and 94 for the primer B extension product, oligo 1 and 95 for the primer C extension product, and oligo 1 and 96 for the primer D extension product. 2nd PCR amplification was conducted using 1/20 of the previous PCR amplified product using oligo 7 and 96 for the primer A derived PCR product and oligo 4 and 96 for the primer B derived PCR product. The final PCR amplification was conducted using 1/100 of the previous PCR amplified product using oligo 5 and 26 for PCR product derived from primer A and B, and oligo 4 and 5 for PCR product derived from primer C and D. The DNA

libraries were extracted with phenol-chloroform mixture and precipitated with ethanol. The obtained DNA template was mixed evenly and in vitro transcribed and purified as previously described. For the recovery of point mutant library after aminoacylation reaction using streptavidin beads was the same as the in vitro selection procedure. Reverse transcription and PCR amplification of the input library as well as the output library was also the same as in vitro selection procedure except oligo 98 was used for reverse transcription and oligo 97 and 98 was used from PCR amplification. The amplified library was prepared for Illumina Miseq sequencing using oligo 99 and 100 followed by manufacture's protocol and Miseq reagent Kit v3 (150 cycle) was used. The DNA reads were analyzed by calculating the enrichment score (E) for each Tx2.1 point mutant38. Each score was then converted into W score (W) by the following equation:

W= 〖log〗 _2 (E_mutant/E_Tx2.1)

2.4.12 *In vitro* translation

The FIT system[77] comprised a mixture of all desired components for translation. The composition of the FIT system was as follows: 50 mM HEPES-KOH (pH 7.6), 12.3 mM magnesium acetate, 100 mM potassium acetate, 2 mM spermidine, 20 mM creatine phosphate, 2 mM DTT, 2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 0.1 mM 10-formyl-5,6,7,8-tetrahydrofolic acid, 0.5 mM Tyrosine and Lysine, 50 μM Aspartic acid and 1.5 mg/mL *E. coli* total tRNA along with 0.73 μM AlaRS, 0.03 μM ArgRS, 0.38 μM AsnRS, 0.13 μM AspRS, 0.02 μM CysRS, 0.06 μM GlnRS, 0.23 μM GluRS, 0.09 μM GlyRS, 0.02 μM HisRS, 0.4 μM IleRS, 0.04 μM LeuRS, 0.11 μM LysRS, 0.03 μM MetRS, 0.68 μM PheRS, 0.16 μM ProRS, 0.04 μM SerRS, 0.09 μM ThrRS, 0.03 μM TrpRS, 0.02 μM TyrRS, 0.02 μM ValRS, 0.6 μM MTF, 2.7 μM IF1, 3 μM IF2, 1.5 μM IF3, 0.26 μM EF-G, 10 μM EF-Tu, 10 μM EF-Ts, 0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 0.1 μM T7 RNA polymerase, 4 μg/mL creatine kinase, 3μg/mL myokinase, 0.1 μM pyrophosphatase, 0.1 μM nucleotide-diphosphatase kinase, and 1.2

μM ribosome. For additional components, the final concentrations were as follow; 4 mM Bio-^{L}Phe-CME (in DMSO 8% v/v), 100 μM Tx2.1, 50 μM $tRNA^{Gly}_{GCC}$, and 200 nM DNA template. Translation reactions were conducted at 37 °C for 1hr. For control peptides, aminoacyl-tRNA prepared using flexizyme as previously described (for Ctrl 2) or same volume of water (for Ctrl 1) was added to the translation mixture at 50 μM final concentration.

2.4.13 Tricine-SDS-PAGE analysis of translation products.

The translation reaction (2.5 μL) was performed in the presence of 50 μM [^{14}C]-Asp (> 200 mCi/mmol, 0.1 mCi/mL, PerkinElmer cat. No. NEC268E) and mixed with equal volume of 2× tricine-SDS-PAGE loading buffer [0.9 M Tris-HCl (pH 8.45), 8% SDS, 30% glycerol and 0.001% xylene cyanol] after translation and heated for 5 min at 95°C. The sample was then run through 15% tricine-SDS-PAGE for 45 min with 150V and then analyzed by autoradiography using a Typhoon FLA 7000 (GE Healthcare).

2.4.14 MALDI-TOF analysis

The translated peptide solution was desalted using a C18-tip (Nikkyo Technos) and analyzed using MALDI-TOF mass in reflector positive modes. All MALDI-TOF analysis was performed using an ultrafleXtreme (Bruker Daltonics) with external calibration (Peptide Calibration Standard II, Bruker Daltonics).

Chapter 3

In vitro selection of T-box motif derived aminoacylation ribozyme using various substrates

3.1 Introduction

In chapter 2, a novel aminoacylation ribozyme, Tx2.1 which recognizes tRNA anticodon and amino acid substrate at the same time was discovered and analyzed. However, the strict specificity toward biotin-^{L}Phe-CME hinders the usage of aminoacyl-tRNA prepared by Tx2.1 to elongation position. Although the discovered Tx2.1 indeed possess the important features of natural aminoacyl-tRNA synthetase nowadays, the fact that different aminoacyl-tRNA synthetase exists for each proteinogenic amino acid encouraged me to design new T-box riboswitch-based ribozyme that recognize different amino acid substrate as well. Such aminoacylation ribozyme would have the possibility to be used for preparing aminoacyl-tRNA in elongation position. Therefore, in this chapter, using the T-box riboswitch as the base structure, the discovery of aminoacylation ribozyme with different amino acid specificity was attempted.

The strategy that was used in the previous chapter to selectively recover active species from random library, was by the strong affinity between biotin and streptavidin. In this chapter, the same biotin-streptavidin interaction will be used but the amino acid substrate structure is different in way that biotin residue is introduced after the aminoacylation reaction or biotin residue is positioned in the side chain of the substrate. Because this Tx2.1 recognized the aromatic ring of the amino acid substrate, the aromatic residue was moved to the leaving group of amino acid, by introducing di-nitrobenzyl group to activate the amino acid. Furthermore, in order to conduct the aminoacylation in harsh condition for amino acid substrate, hydroxy acid structure was employed for one of the amino acid substrates. Figure 3.1 presents the actually amino acid substrates structure used in this study.

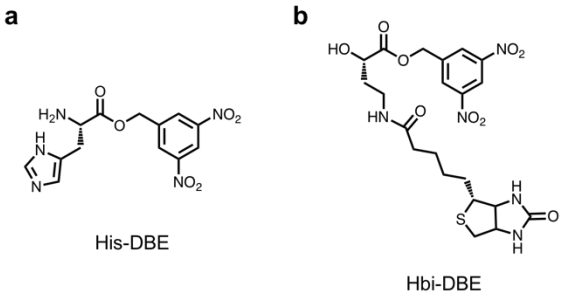

Figure 3.1 | Amino acid substrates used in this study
a) Chemical structure of histidine 3,5-dinitrobenzyl ester (His -DBE). b) Chemical structure of δ -(N-biotinyl)amino-α-(S)-hydroxy- butanoic acid 3,5-dinitrobenzyl ester (Hbi-DBE)

3.2 Results and discussion

3.2.1 *In vitro* selection of aminoacylation ribozyme using His-DBE

The strategy is the same as described in Figure 2.4. However, since the amino acid substrate do not contain biotin residue, in order to selectively recover the active species from the library, biotin labeling of the active species was performed using biotinyl-*N*-hydroxysulfosuccinimde (biotin-NHS). After the aminoacylation reaction, the reaction mixture was ethanol precipitated to remove excess amount of amino acid substrate, His-DBE. Then the RNA was treated with biotin-sulfo-NHS so that active RNA attached with histidine moiety will selectively and irreversibly react with biotin-sulfo-NHS through primary amine and NHS. The biotin labeled RNA was selectively recovered using streptavidin coated magnetic beads followed by reverse transcription, PCR and transcription to yield a new set of RNA library. In this experiment, the same library design that was used to obtain Tx2.1 was adopted but library 1 and library 2 was conducted separately (Figure 3.2). The selection cycle was repeated for total of eight rounds. For the first seven rounds, the aminoacylation reaction was conducted for 2 hours and the last round was conducted in harsh condition with only 30 minutes for aminoacylation. Figure 3.2 shows the progress of *in vitro* selection.

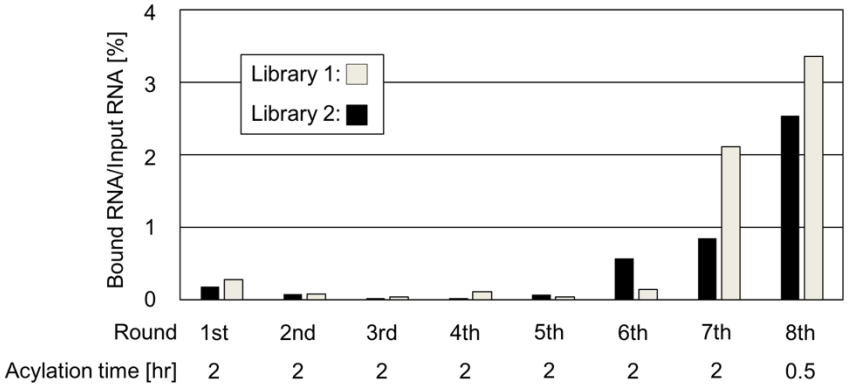

Figure 3.2 | Progress of *in vitro* selection using His-DBE as substrate

The RNA sequence recovered after round 8 was analyzed by deep sequencing and total of five sequences were identified from both libraries, three from library 1 and two from library 2 (Figure 3.3). These sequences do not share any consensus motif indicating that they are all unique candidates.

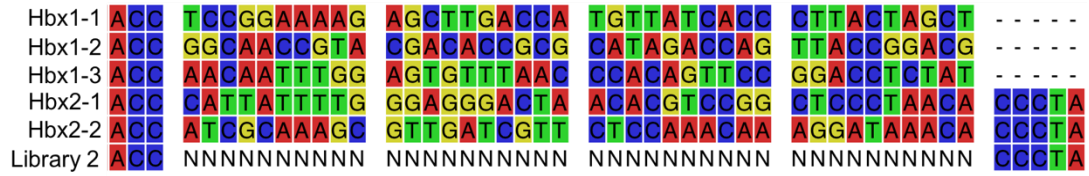

Figure 3.3 | Identified clones after deep sequencing

3.2.2 Trans-aminoacylation of selected sequence using His-DBE

The identified clones were checked for aminoacylation activity against His-DBE (Figure 3.4). The efficiency was analyzed using acid-Urea-PAGE analysis. To the disappointment, no clones were able to aminoacylate Hid-DBE. Since the increase of recovery for the RNA was observed in the *in vitro* selection and enrichment of sequence was observed, the sequence is expected to have some kind of activity. Therefore, I focused on the reaction condition in this experiment. I envisioned that these identified clones could be catalyzing the biotinylation reaction rather than the aminoacylation reaction due to the use of biotin-sulfo-NHS after aminoacylation.

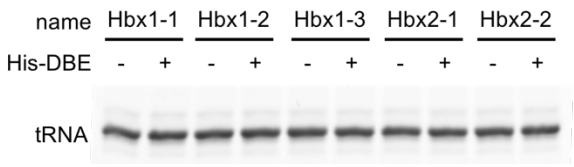

Figure 3.4 | Trans-aminoacylation activity check identified clones

3.2.3 Biotinylation reaction of selected clones using biotin-NHS

Biotinylation of the selected clones were monitored using biotin-sulfo-nhs and streptavidin. First, the self-biotinylation was measured where the substrate tRNA is fused to downstream of the RNA sequence (Figure 3.5). This condition is the same as the *in vitro* selection condition. The result shows that four clones actually did biotinylate itself when treated with biotin-sulfo-nhs for 1.5 hrs at 4°C. The biotinylation efficiency was ranging from 20 % up to 47 %. This result indicates that even though I aimed at obtaining aminoacylation ribozyme, the recovered clones were actually biotinylation ribozyme. Although the result was not what I expected, further characterization was conducted.

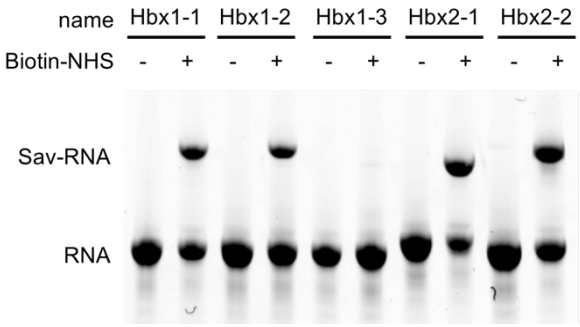

Figure 3.5 | Self-biotinylation of recovered clones

The trans-biotinylation was next conducted where the substrate tRNA was detached from the RNA sequence to see whether these ribozymes act without the tRNA sequence within the ribozyme (Figure 3.6). To our surprise, for Hbx1-1, 1-2 and 2-2, biotinylation was observed and even without the tRNA. This indicates that these ribozymes are selfbiotinylating, rather than biotinylating the tRNA. The biotinylated product was treated for hydrolysis under basic condition and the biotinylated product decreased indicating that the biotin conjugation is formed as ester bond rather than amide bond.

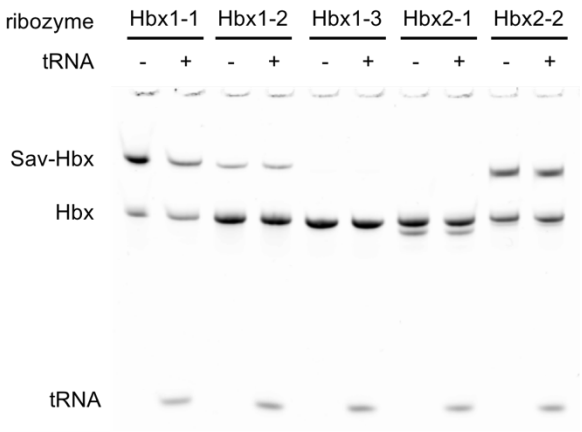

Figure 3.6 | Trans-biotinylation of recovered clones with and without tRNA.

3.2.4 Determining the biotinylation site by deletion of T-box domains

The identified clones were analyzed as self-biotinylating ribozyme and the top most active ones, Hbx1-1 and Hbx2-2, where further analyzed and the important domain for this reaction was determined by partially deleting certain domain of the sequence. Total of three domians were deleted and the effect to biotinylation was checked (Figure 3.6). When the apical loop was deleted (ΔApical), the biotinylation reaction was abolished. Also, when the total Stem I domain was deleted and transrcription was started from the stem III region(fmStemIII), the RNA sequence did not have biotinylation activity, indicating that the total Stem I domain is necessary for biotinylaiton. When the Stem III reigon was deleted(ΔStemIII), the biotinylation was preserved, suggesting this domain does not contribute to the activity. These results indicate that the pink and magenta colored region in Figure 3.7 is necessary and possible reaction site for the biotinylation reaction.

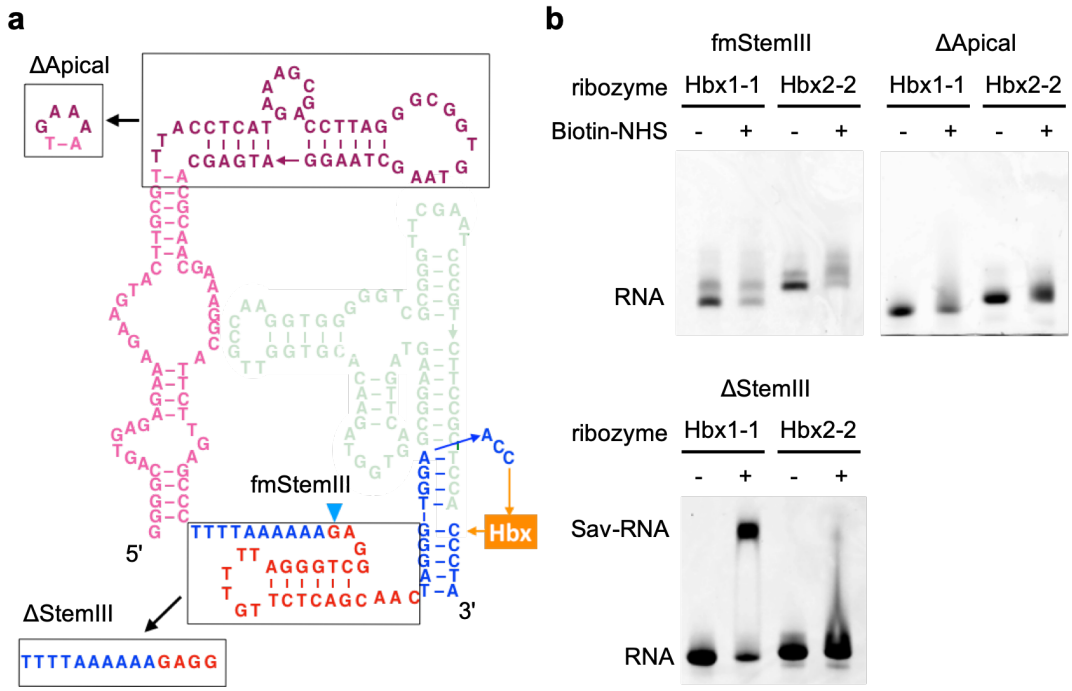

Figure 3.7 | Investigation of important domain for biotinylation.
a) Secondary structure of T-box body sequence and the corresponding partially deleted T-box body sequence. Orange rectangle denoted with Hbx corresponds to the random domain. b) Biotinylation experiments with apical domain deleted, stemIII domain deleted or from stemIII domain.

3.2.5 *In vitro* selection using Hbi-DBE as substrate

In order to obtain Tx2.1 analogue that recognize different amino acid substrate, an RNA library was designed focused on Tx2.1 sequence. For each of the 40 nucleotide residues specific for Tx2.1, the library was designed as to have 90% the original nucleotide of Tx2.1 and 3.3% each for the other three nucleotides. For example, since the first nucleotide of Tx2.1 random domain is adenine, the library used in this study contains 90% adenine, 3.3% cytosine, 3.3% guanine and 3.3% uracil at the first position. The same principle was applied to every 40 nucleotides in the Tx2.1 specific sequence region. In this library 56% contains at least three, four or five mutations compared to Tx2.1. Using this focused library, *in vitro* selection was conducted using Hbi-DBE as substrate. Total of 9 rounds was conducted and sufficient increase in the recovery of RNA was observed (Figure 3.8). For the first five rounds of selection, the aminoacylation reaction was conducted for 2 hours. For the last four rounds, aminoacylation was conducted for only 30 minutes to harshen the condition. The recovered DNA after round 9 was sequenced and total of 7 clones were identified (Figure 3.9). Interestingly, these clones contain a certain motif, "AAGCTTGGATAAC". Although not reported, this could be the domain that recognize biotin residue.

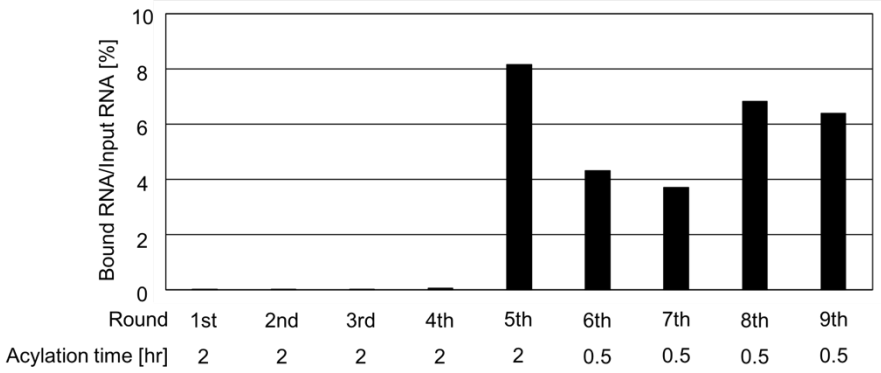

Figure 3.8 | Progress of *in vitro* selection using Hbi-DBE as substrate

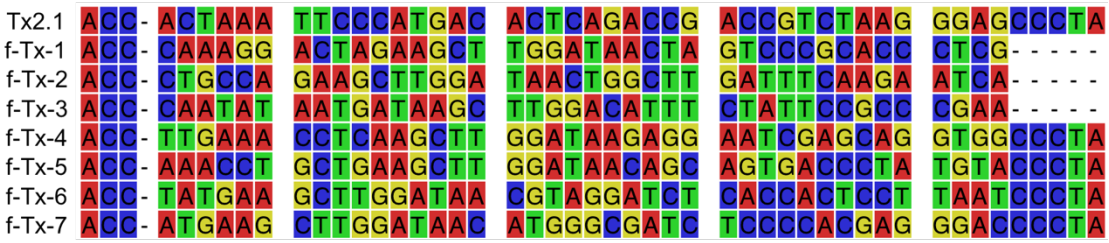

Figure 3.9 | Identified clones from sequencing round 5, 6 and 9

3.2.6 Aminoacylation activity of individual clones

First, the self-aminoacylation activity of selected seven clones were checked using Hbi-DBE as substrate followed by mixing with streptavidin. The result showed that all of the clones exhibited aminoacylation activity ranging from 7% to 38% (Figure 3.10). Since tRNA is present without anything attached in the 5' end, the trans-aminoacylaiton activity against native tRNA was also tested (Figure 3.11). To my disappointment, none of the collected clones showed aminoacylation activity against mature tRNA. Although none of the identified clones showed transaminoacylation activity, I focused on the most abundant f-Tx-2 clone for further characterization. The first thing that was checked was whether the self-aminoacylation was happening at the 3-terminal tRNA-CCA end. The self-aminoacylation was checked with sodium periodate treated f-Tx-2 where the terminal ribose was converted to dialdehyde (Figure 3.12). The aminoacylation activity was abolished for sodium periodate treated f-Tx-2 meaning that Hbi-DBE was acylated at the 3'-end of tRNA. Since the difference between self-aminoacylaiton and transaminoacylation is the presence of poly-A linker at the 3'-end of ribozyme sequence, I envisioned that poly-A linker could be contributing to the aminoacylation reactivity and prepared f-Tx-2 having poly-A linker at its 3'-end. The way to prepare such f-Tx-2 having poly-A linker was treating the RNA sequence used for self-aminoacylation which is fished out from in vitro selection with RNase P. RNase P is a haloenzyme that recognizes the tertiary structure of pre-tRNA to produce mature tRNA 5'-end. Thus, the RNase P treated RNA would be

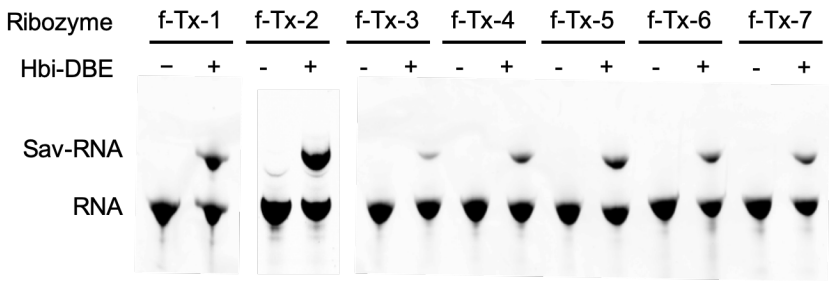

Figure 3.10 | Self-aminoacylation of identified clones

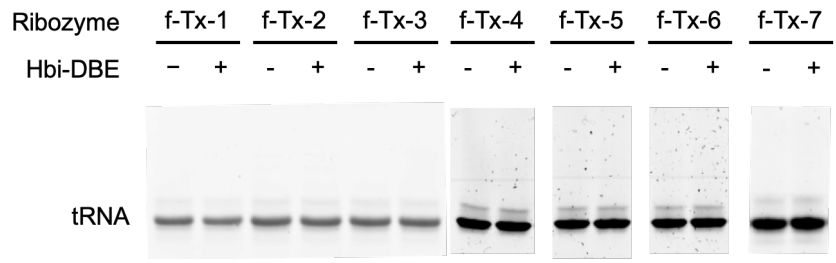

Figure 3.11 | Trans-aminoacylation of identified clones

divided in to f-Tx-2 having poly A tail and mature tRNA. The RNase P treated product was ethanol precipitated and dissolved in aminoacylation buffer and then mixed with Hbi-DBE to see if aminoacylation of mature tRNA with f-Tx-2 having poly A was possible (Figure 3.13). As a result, even with poly A tail, f-Tx-2 did not show any transaminoacylation activity. It seems that the ribozymes enriched in the in vitro selection was only active for self-aminoacylation.

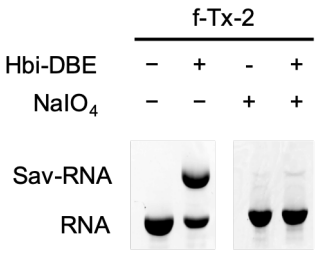

Figure 3.12 | Trans-aminoacylation of f-Tx-2 with sodium periodate treatment

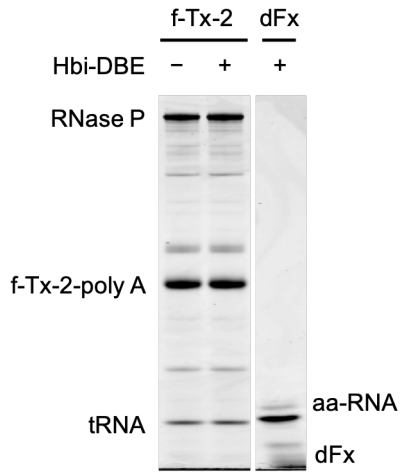

Figure 3.13 | Trans-aminoacylation of f-Tx-2 coupled with RNase P
For positive control, Hbi-DBE acylated with dFx was used.

3.3 Conclusion

Because Tx2.1 obtained in chapter 2 recognizes the biotin residue of amino acid strongly, in this chapter the amino acid substrates used had no biotin residue or biotin residue installed in the side chain of the amino acid in the hope that aminoacylation ribozymes without biotin recognition could be obtained. Using RNA library that was used in obtaining Tx2.1 or Tx2.1-focused RNA library, in vitro selection to fish out aminoacylation ribozyme was again conducted.

As a result, even though enrichment in RNA sequence was observed for both amino acid substrates, none of the identified ribozymes exhibited aminoacylation toward mature tRNA. Instead, the ones identified using His-DBE as amino acid substrate turned out to be biotinylation ribozyme instead of aminoacylation ribozyme. Although biotinylation ribozymes have been previously identified, the random sequence does not show any similarity and that T-box body sequence is to some extent necessary for its activity. As for in vitro selection using Hbi-DBE, RNA sequences that exhibited self-aminoacylation activity was obtained. In order to obtain aminoacylation ribozymes with trans-aminoacylation activity, extending the poly A linker region to make substrate tRNA and ribozymes domain further apart could be one way to tackle this problem.

3.4 Materials and methods

3.4.1 Materials

All DNA oligomers used were purchased from Eurofin Genomics except from oligomers having 2′-*O*-Methyl guanosine which was purchased from Gene Design.

Name	Sequence (5′-3′)	Used for
Oligo 1	GTACTTGCGTTTACCTCATGAAAGCGACCTTAGGGCGGTGTAAGCTAAGGATGAGCACGCAACGAAAGGCATTCTTGAGCCCTTTTAAAAAAGAGGCTGGGATTTTGTTCTCAGCAACTAGGGTGGAACC	Library construction
Oligo 2	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTAGGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGTTCCACCCTAGTTGCTGA	Library construction
Oligo 3	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGGTTCCACCCTAGTTGCTGA	Library construction
Oligo 4	GGCGTAATACGACTCACTATAGGGGCAGTGAGAGAAAGAAGTACTTGCGTTTACCTCATGAAAG	Library construction
Oligo 5	TG(M)GAGCGGAAGACGGGATTCGAACCCGCGACCCCCACCTTGGCAAGGTGGT	Library construction
Oligo 6	TG(M)GAGCGGAAGACGGG	RT PCR
Oligo 7	GGCGTAATACGACTCACTATAGGGGCAGTGAGAGAAAGAA	PCR amplification
Oligo 8	GCCTCTTCGCTATTACGCCAGC	Sequencing
Oligo 9	TGTTGTGTGGAATTGTGAGCGG	Sequencing
Oligo 10	CTTAGGGCGGTGTAAGCTAAGGATGAGCACGCAACGAAAGGCAT	Hbx, f-Tx
Oligo 11	GCGTTTACCTCATGAAAGCGACCTTAGGGCGGTGTAAGCT	Hbx, f-Tx
Oligo 12	GCAGTGAGAGAAAGAAGTACTTGCGTTTACCTCATGAAAGC	Hbx, f-Tx
Oligo 13	GGCGTAATACGACTCACTATAGGGGCAGTGAGAGAAAGAAGTAC	Hbx, f-Tx
Oligo 14	GCCTCTTTTTTAAAAGGGCTCAAGAATGCCTTTCGTTGCGTGCT	Hbx, f-Tx
Oligo 15	TAGTTGCTGAGAACAAAATCCCAGCCTCTTTTTTAAAAGGGCTC	Hbx, f-Tx
Oligo 16	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTAGCTAGTAAGGGTGATAACATGGTCAAGCTCTTTTCCGGAGGTTCCACCCTAGTTGCTGA	cis Hbx1-1
Oligo 17	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTCGTCCGGTAACTGGTCTATGCGCGGTGTCGTACGGTTGCCGGTTCCACCCTAGTTGCTGA	cis Hbx1-2
Oligo 18	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTATAGAGGTCCGGAACTGTGGGTTAAACACTCCAAATTGTTGGTTCCACCCTAGTTGCTGA	cis Hbx1-3
Oligo 19	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTAGGGTGTTAGGGAGCCGGACGTGTTAGTCCCTCCCAAAATAATGGGTTCCACCCTAGTTGCTGA	cis Hbx2-1

Name	Sequence (5′-3′)	Used for
Oligo 20	CCACCTTGGCAAGGTGGTGTTCTACCACTGAACTACTTCCGCTTTTTTTTTTTTTTAGGGTGTTTATCCTTTGTTTGGAGAACGATCAACGCTTTGCGATGGTTCCACCCTAGTTGCTGA	cis Hbx2-2
Oligo 21	TCTTTTCCGGAGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	Hbx1-1
Oligo 22	GGGTGATAACATGGTCAAGCTCTTTTCCGGAGGTTCCAC	Hbx1-1
Oligo 23	AGCTAGTAAGGGTGATAACATGGTCAAGC	Hbx1-1
Oligo 24	TACGGTTGCCGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	Hbx1-2
Oligo 25	GGTCTATGCGCGGTGTCGTACGGTTGCCGGTTCCAC	Hbx1-2
Oligo 26	CGTCCGGTAACTGGTCTATGCGCGGTGTCG	Hbx1-2
Oligo 27	CTCCAAATTGTTGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	Hbx1-3
Oligo 28	CCGGAACTGTGGGTTAAACACTCCAAATTGTTGGTTCCAC	Hbx1-3
Oligo 29	ATAGAGGTCCGGAACTGTGGGTTAAAC	Hbx1-3
Oligo 30	CCCAAAATAATGGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	Hbx2-1
Oligo 31	CCGGACGTGTTAGTCCCTCCCAAAATAATGGGTTCCAC	Hbx2-1
Oligo 32	TAGGGTGTTAGGGAGCCGGACGTGTTAGTCCCT	Hbx2-1
Oligo 33	GCTTTGCGATGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	Hbx2-2
Oligo 34	TTTGTTTGGAGAACGATCAACGCTTTGCGATGGTTCCAC	Hbx2-2
Oligo 35	TAGGGTGTTTATCCTTTGTTTGGAGAACGATCAAC	Hbx2-2
Oligo 36	CGTAATACGACTCACTATAGGAACCTCCGGAAAAGAG	fmStemIII Hbx1-1
Oligo 37	AGCTAGTAAGGGTGATAACATGGTCAAGCTCTTTTCCGGAGGTTCC	fmStemIII Hbx1-1
Oligo 38	AGCTAGTAAGGGTGATAACATGGT	fmStemIII Hbx1-1
Oligo 39	GGCGTAATACGACTCACTATAGGAACCATCGCAAAGCGTTGATC	fmStemIII Hbx2-2
Oligo 40	TAGGGTGTTTATCCTTTGTTTGGAGAACGATCAACGCTTTGCGATGGTTCC	fmStemIII Hbx2-2
Oligo 41	TAGGGTGTTTATCCTTTGTTTG	fmStemIII Hbx2-2
Oligo 42	CCTCTTTTTTAAAAGGGCTCAAGAATGCCTTTCGTTGCGTGCT	ΔStemIII
Oligo 43	CTTTTCCGGAGGTTCCACCCTACCTCTTTTTTAAAAGGGCTCAA	ΔStemIII Hbx1-1
Oligo 44	GCTTTGCGATGGTTCCACCCTACCTCTTTTTTAAAAGGGCTCAA	ΔStemIII Hbx2-2
Oligo 45	GAAAACGCAACGAAAGGCATTCTTGAGCCCTTTTAAAAAAGAGG	ΔStemI
Oligo 46	GCAGTGAGAGAAAGAAGTACTTGCGTGAAAACGCAACGAAAGGC	ΔStemI
Oligo 47	TCTAGTCCTTTGGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	f-Tx-1
Oligo 48	GCGGGACTAGTTATCCAAGCTTCTAGTCCTTTGGGTTCCAC	f-Tx-1
Oligo 49	TTTCGAGGGTGCGGGACTAGTTATCCAAGCT	f-Tx-1
Oligo 50	CTTCTGGCAGGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	f-Tx-2
Oligo 51	TGAAATCAAGCCAGTTATCCAAGCTTCTGGCAGGGTTCCAC	f-Tx-2
Oligo 52	TTTGATTCTTGAAATCAAGCCAGTTATCCA	f-Tx-2

Name	Sequence (5′-3′)	Used for
Oligo 53	CATTATATTGGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	f-Tx-3
Oligo 54	CGGAATAGAAATGTCCAAGCTTATCATTATATTGGGTTCCACCC	f-Tx-3
Oligo 55	TTTTCGGGCGGAATAGAAATGTCCAAGC	f-Tx-3
Oligo 56	TTGAGGTTTCAAGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	f-Tx-4
Oligo 57	GCTCGATTCCTCTTATCCAAGCTTGAGGTTTCAAGGTTCCAC	f-Tx-4
Oligo 58	TAGGGCCACCTGCTCGATTCCTCTTATCCAAG	f-Tx-4
Oligo 59	CAGCAGGTTTGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	f-Tx-5
Oligo 60	GGGTCACTGCTGTTATCCAAGCTTCAGCAGGTTTGGTTCCAC	f-Tx-5
Oligo 61	TAGGGTACATAGGGTCACTGCTGTTATCCAAG	f-Tx-5
Oligo 62	AAGCTTCATAGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	f-Tx-6
Oligo 63	GAGTGGTGAGATCCTACGTTATCCAAGCTTCATAGGTTCCACCC	f-Tx-6
Oligo 64	TAGGGATTAAGGAGTGGTGAGATCCTACGTTATC	f-Tx-6
Oligo 65	AGCTTCATGGTTCCACCCTAGTTGCTGAGAACAAAATCCC	f-Tx-7
Oligo 66	GGAGATCGCCCATGTTATCCAAGCTTCATGGTTCCACCC	f-Tx-7
Oligo 67	TAGGGGTCCCTCGTGGGGAGATCGCCCATGTTATC	f-Tx-7
Oligo 68	ACCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCC	dFx
Oligo 69	ACCTAACGCCATGTACCCT	dFx
Oligo 70	TACCTAACGCCATGTACCCT	dFx

3.4.2 Synthesis of amino acid substrates.

His-DBE and Hbi-DBE was synthesized as previously described[38].

3.4.3 Preparation of DNA templates for RNA libraries

DNA templates for transcription of RNA library were prepared by 2-step PCR. First, 500 μL-scale DNA polymerase extension of the two DNA primers (oligo 1 and 2 for library 1, oligo 1 and 3 for library 2) was performed using Taq polymerase Hot Start Version (R007A, TaKaRa) at 95°C for 30 sec, followed by 5 cycles of 55°C for 30 sec and 72°C for 1 min. Next, all amount of the extension products were extended to full-length and amplified with oligo 4 and 5 by 5 mL-scale PCR using Taq polymerase Hot Start Version (R007A, TaKaRa) by 6 cycles of 95°C for 30 sec, 55°C for 30 sec and 72°C for 1 min. The DNA libraries were extracted with phenol-chloroform mixture and precipitated with ethanol.

3.4.4 *In vitro* transcription of RNAs

The RNAs were prepared by in vitro transcription with T7 RNA polymerase and purified by denaturing PAGE (8M Urea, 1×TBE). DNA templates were modified with 2′-O-methylation at the second last nucleotide of the 5′ termini to reduce non-templated nucleotide addition by T7 RNA polymerase39. The primers for preparing transcription templates and RNAs are shown in Supplementary Table 2 and 3. All non-methylated primers were purchased from Operon Biotechnology (Japan) and methylated primers were purchased from Gene Design (Japan). The concentrations of RNAs were determined by absorbance at 260 nm.

3.4.5 *In vitro* selection

In vitro selection was carried out under the following condition: 33.3 μM (first round only) or 3.33 μM RNA library was heated at 95°C for 2 min and cooled on ice over 5 min. This RNA solution was mixed with 66.7% volume of 25 mM His-DBE or Hbi-DBE in DMSO followed by equal volume of 2×reaction buffer [100 mM HEPES-KOH (pH 7.5), 1 M KCl and 20 mM MgCl2]. Total reaction volume is 200 μL (the first round only) or 100 μL. Then the mixture was incubated on ice for 2 hours or 30 min. The reaction was stopped by adding 4 times volume of cold 0.3 M NaOAc (pH 5.3) and 10 times volume of cold ethanol. Then the RNA was ethanol precipitated twice. For His-DBE selection, the pellet was dissolved in 10 μL of 1 mM NaOAc (pH5.3) followed by addition of 13.3 mg/mL Biotin-sulfo-NHS and 60 mM HEPES-KOH (pH 8.5). Biotinylation was conducted on ice for 1.5 hours and ethanol precipitated twice. The RNA pellet was dissolved into SAv binding buffer [50 mM HEPES-KOH (pH 7.5), 500 mM KCl, 5 mM EDTA and 0.05% tween-20], then incubated with 20 μL (200 μL for the first round) of Dynabeads® My One Streptavidin C1 (65001, Life Technologies), which were prewashed with SAv binding buffer three times before use, for 30 min using rotary shaker at 4°C. After the incubation, the supernatant was discarded and the beads were washed with

200 μL or 500 μL (the first round only) of 1×SAv binding buffer, 200 μL or 500 μL (the first round only) of urea wash buffer (4 M Urea and 0.05% tween-20) and 200 μL or 500 μL (the first round only) of 0.05% tween-20. The beads were suspended in 20 μL or 200 μL (the first round only) SAv elution buffer [50 mM HEPES-KOH (pH 7.5), 300 mM NaCl, 5 mM EDTA, 1 mM biotin and 0.05% tween-20] and incubated at 95°C for 10 min using rotary shaker. The supernatant was recovered and this elution cycle repeated three times. All of the supernatants were mixed and precipitated with ethanol.

RNA pellet was dissolved into 5 μL water and reverse transcribed with oligo 6 by SuperScript® III (18080044, ThermoFisher Scientific) following manufacture's protocol.

To quantify the amount of recovered RNAs from beads, quantitative PCR was performed as follows; 1% of reverse transcription reaction solution was mixed with qPCR mix [1 μM oligo 6, 1 μM oligo 7, 0.2 mM dNTP, 1/2,00,000 SYBR Green I (50512, LONZA), Taq polymerase Hot Start Version (R007A, TaKaRa) and 1×PCR buffer] and qPCR was performed using LightCycler®Nano (Roche Diagnostics K.K.) by 25 cycles of 95°C for 30 sec (5°C/sec), 55°C for 30 sec (4°C/sec) and 72°C for 60 sec (0.5°C/sec) to measure Cq values. The remaining reverse transcription reaction solution not used for qPCR, was mixed with 200 μL of qPCR mix without SYBR Green I and DNA was amplified by PCR by Cq-1 cycle of 95°C for 30 sec, 55°C for 30 sec and 72°C for 60 sec. DNA was extracted with phenol-chloroform mixture and precipitated with ethanol, and then used for next round of selection.

3.4.6 Sequencing

A The cDNAs of RNAs after in vitro selection were cloned into pGEM®-T Easy Vectors (A1360, Promega K.K.). These vectors were used to transform E. coli DH5α cells in LB medium with 100 μg/mL ampicillin. Inserted cDNAs in the vectors were amplified by colony-PCR using primers oligo 8 and oligo 9, and sequenced from both 5′- and 3′-end by FASMAC Co., Ltd. (Japan) Preparation of DNA templates for *in vitro* translation

3.4.7 Aminoacylation

Aminoacylation reactions by flexizymes were performed as previously described[15]. The conditions of self-aminoacylation by T-box ribozymes are the same as those described in a method of *in vitro* selection. For trans-aminoacylation, 83.3 μM tRNA and 166.7 μM Tx were first heated at 95°C for 2 min and cooled on ice over 5 min. This RNA solution was mixed 66.7 % volume of 25 mM amino acid substrates in DMSO followed by equal volume of 2×reaction buffer [100 mM HEPES-KOH (pH 7.5), 1 M KCl and 20 mM $MgCl_2$]. The reaction mixture was incubated on ice for 2 hours. After the reaction, aminoacyl-RNA was precipitated with ethanol as previously described[15].

3.4.8 Analysis of acylation using streptavidin

aa-RNA precipitated with ehanol was dissolved in SAv-shift buffer [4 M Urea, 1 mM EDTA, 1 mM Tris, 333 μM NaOAc (pH 5.3) and 10 mg/mL Streptavidin] and then loaded on urea 6% polyacrylamide gels (8 M urea).

For the analysis of acylation with Phe-CME, ethanol-precipitated aa-RNA was first dissolved in 1 mM NaOA (pH 5.3) and then 300 mM HEPES-KOH (pH 8.0) was added. Biotin-3-sulfo-N-hydroxylsuccinimide ester was immediately added at 4°C with the finial concentration of 15mM. After 1 hour, the biotinylation reaction was stopped by ethanol precipitation twice. Ethanol-precipitated bio-aa-RNA was then dissolved in loading buffer [4 M Urea, 1 mM EDTA, 1 mM Tris, 333 μM NaOAc (pH 5.3)] and heated for 30 seconds at 95°C, and then cooled to 25°C. After addition of SAv solution (10 mg/mL), the sample was analyzed by urea 6% polyacrylamide gels (8 M urea).

3.4.9 Analysis of acylation by acid-PAGE

aa-tRNA which was ethanol precipitated was dissolved in acid-PAGE loading buffer (150 mM NaOAc, pH 5.2, 10 mM EDTA, and 93% (v/v) formamide) and then loaded on

acid-urea 12% polyacrylamide gels (8 M urea, 50 mM NaOAc, pH 5.2). Electrophoresis was performed using 300 V (approximately 20 Vcm^{-1}) for 7 h at 4°C. The gels were stained with ethidium bromide and analyzed using a Typhoon FLA 7000 fluorescent image analyzer (Fujifilm). Aminoacylation efficiency was calculated based on the band intensity of aa-tRNA (A) and free tRNA (T) and is obtained as (A)/[(A) + (T)].

3.4.10 RNase P cleavage followed by aminoacylation

C5 protein which is the protein component of RNase P was prepared as previously described[79]. The RNA was treated with 1 μM of RNase P at 37°C for 30min whose buffer condition was as follow; 50 mM HEPES-KOH (pH 7.5), 12 mM $Mg(OAc)_2$, 100 mM KOAc and then subjected to ethanol precipitation. The ethanol precipitated RNase P treated RNA was dissolved in water and then aminoacylation reaction was conducted.

Chapter 4

General conclusion

In the current world, the aminoacylation reaction is conducted solely by aminoacyl-tRNA synthetases (ARS) which recognizes its corresponding tRNA and amino acid specifically. The defined correspondence between tRNAs and acids defines the genetic code, which is essential for accurate translation of mRNAs to proteins in living organism. In the transition era between the current world and the possible RNA world, such aminoacylation reaction could be conducted by ribozyme. It is possible that ribozyme specifically recognizing both the amino acid and tRNA could have been present. In this thesis the recreation of such ribozyme was attempted.

In chapter 2, with the use of T-box riboswitch, a novel ribozyme named Tx2.1 was discovered. This ribozyme has the ability to specifically recognize biotin-^{L}Phe-CME for its amino acid substrate. Furthermore, Tx2.1 could distinguish the tRNA anticodon as well. The similar characteristics between natural ARS and Tx2.1 presents us the possibility that such ribozyme similar to Tx2.1 could have been present in the ancient time. Furthermore, the demonstration that Tx2.1 was able to function within the in vitro translation mixture was presented. This is to my knowledge the very first case to show that aminoacylation ribozyme functions in parallel with the translation mixture. This approach to discover ribozymes from existing riboswitch could be applicable to other riboswitches as well and those engineered ribozymes could also be resembling a possible ancient ribozyme existed in the RNA world. Since there is still room for improvement in efficiency, the discovery of Tx2.1 presents us with new apparatus for *in situ* aminoacylation for genetic code reprogramming as well.

In chapter 3, similar approach to chapter 2 was conducted against different amino acid substrates and His-DBE and Hbi-DBE was chosen as amino acid substrates. In the case of His-DBE selection, the identified RNA sequence turned out to be biotinylation

ribozyme instead of aminoacylation ribozyme. The study indicates the possible need of biotin residue directly attached to the amino acid substrate rather than conjugating afterwards to remove the risk of obtaining biotinylation ribozyme. The reduction of reaction time for biotinylation could also be applied to selectively recover aminoacylating ribozymes. For Hbi-DBE selection, ribozymes having selfaminoacyalation activity was obtained. These results present us the benefit of using T-box riboswitch as base structure rather than from totally random RNA sequence and that this T-box riboswitch approach to obtain aminoacylation ribozyme is to some extent feasible.

In this thesis, the main success is the discovery of aminoacylation ribozyme that recognizes the tRNA secondary structure and anticodon. As a technical point of view, this ribozyme presents a possible starting point for the engineering of new tRNA-recognizing ribozymes with different specificity to both amino acids and tRNAs. As in the point for origin of life, this ribozyme presents us the possible link between the RNA world and the current protein world governed by genetic code. Tx2.1 could be resembling the first ribozymes that was responsible for the emergence of genetic code that links amino acids and codons in the current world.

Reference

1. Kumar, R. & Yarus, M. RNA-catalyzed amino acid activation. *Biochemistry* **40**, 6998-7004 (2001).
2. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643-647 (1995).
3. Walter, G. Origin of life: The RNA world. *Nature* **319**, 618 (1986).
4. Kruger, K. et al. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-157 (1982).
5. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
6. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).
7. Ellington, A. & Szostak, J. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
8. Nissen, P., Hansen, J., Ban, N., Moore, P.B. & Steitz, T.A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-930 (2000).
9. Moore, P.B. & Steitz, T.A. The structural basis of large ribosomal subunit function. *Annu. Rev. Biochem.* **72**, 813-850 (2003).
10. Crick, F.H. The origin of the genetic code. *J. Mol. Biol.* **38**, 367-379 (1968).
11. Johnston, W.K., Unrau, P.J., Lawrence, M.S., Glasner, M.E. & Bartel, D.P. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325 (2001).
12. McDaniel, B.A., Grundy, F.J., Artsimovitch, I. & Henkin, T.M. Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3083-3088 (2003).
13. Fuchs, R.T., Grundy, F.J. & Henkin, T.M. The S(MK) box is a new SAM-binding RNA for translational regulation of SAM synthetase. *Nat. Struct. Mol. Biol.* **13**, 226-233 (2006).
14. Nahvi, A. et al. Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043 (2002).

15. Mironov, A.S. et al. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**, 747-756 (2002).

16. Mandal, M. & Breaker, R.R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.* **11**, 29-35 (2004).

17. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C. & Breaker, R.R. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **113**, 577-586 (2003).

18. Watson, P.Y. & Fedor, M.J. The ydaO motif is an ATP-sensing riboswitch in Bacillus subtilis. *Nat. Chem. Biol.* **8**, 963-965 (2012).

19. Kim, J.N., Roth, A. & Breaker, R.R. Guanine riboswitch variants from Mesoplasma florum selectively recognize 2'-deoxyguanosine. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 16092-16097 (2007).

20. Mandal, M. et al. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275-279 (2004).

21. Grundy, F.J., Lehman, S.C. & Henkin, T.M. The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 12057-12062 (2003).

22. Ames, T.D. & Breaker, R.R. Bacterial aptamers that selectively bind glutamine. *RNA Biol.* **8**, 82-89 (2011).

23. Winkler, W.C., Nahvi, A., Roth, A., Collins, J.A. & Breaker, R.R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286 (2004).

24. Furukawa, K. et al. Bacterial riboswitches cooperatively bind Ni(2+) or Co(2+) ions and control expression of heavy metal transporters. *Mol. Cell* **57**, 1088-1098 (2015).

25. Cromie, M.J., Shi, Y., Latifi, T. & Groisman, E.A. An RNA sensor for intracellular Mg(2+). *Cell* **125**, 71-84 (2006).

26. Baker, J.L. et al. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**, 233-235 (2012).

27. Grundy, F.J., Rollins, S.M. & Henkin, T.M. Interaction between the acceptor end of tRNA and the T box stimulates antitermination in the Bacillus subtilis tyrS gene: a new role for the discriminator base. *J. Bacteriol.* **176**, 4518-4526 (1994).

28. Saad, N.Y. et al. Two-codon T-box riboswitch binding two tRNAs. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 12756-12761 (2013).

29. Sherwood, A.V., Grundy, F.J. & Henkin, T.M. T box riboswitches in Actinobacteria: translational regulation via novel tRNA interactions. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1113-1118 (2015).

30. Breaker, R.R. Riboswitches and the RNA World. *Cold Spring Harb. Perspect. Biol.* **4**(2012).

31. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).

32. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209-212 (2011).

33. Sczepanski, J.T. & Joyce, G.F. A cross-chiral RNA polymerase ribozyme. *Nature* **515**, 440-442 (2014).

34. Tsukiji, S., Pattnaik, S.B. & Suga, H. An alcohol dehydrogenase ribozyme. *Nat. Struct. Biol.* **10**, 713-717 (2003).

35. Seelig, B. & Jäschke, A. A small catalytic RNA motif with Diels-Alderase activity. *Chem. Biol.* **6**, 167-176 (1999).

36. Agresti, J.J., Kelly, B.T., Jäschke, A. & Griffiths, A.D. Selection of ribozymes that catalyse multiple-turnover Diels-Alder cycloadditions by using in vitro compartmentalization. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 16170-16175 (2005).

37. Lee, N., Bessho, Y., Wei, K., Szostak, J. & Suga, H. Ribozyme-catalyzed tRNA aminoacylation. *Nat. Struct. Biol.* **7**, 28-33 (2000).

38. Saito, H., Kourouklis, D. & Suga, H. An in vitro evolved precursor tRNA with aminoacylation activity. *EMBO J.* **20**, 1797-1806 (2001).

39. Illangasekare, M. & Yarus, M. Specific, rapid synthesis of Phe-RNA by RNA. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 5470-5475 (1999).

40. Murakami, H., Ohta, A., Ashigai, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **3**, 357-359 (2006).

41. Chumachenko, N., Novikov, Y. & Yarus, M. Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *J. Am. Chem. Soc.* **131**, 5257-5263 (2009).

42. Lohse, P.A. & Szostak, J.W. Ribozyme-catalysed amino-acid transfer reactions. *Nature* **381**, 442-444 (1996).

43. Suga, H., Lohse, P.A. & Szostak, J.W. Structural and kinetic characterization of an acyl transferase ribozyme. *J. Am. Chem. Soc.* **120**, 1151-1156 (1998).

44. Bessho, Y., Hodgson, D.R.W. & Suga, H. A tRNA aminoacylation system for non-natural amino acids based on a programmable ribozyme. *Nat. Biotechnol.* **20**, 723-728 (2002).

45. Murakami, H., Saito, H. & Suga, H. A versatile tRNA aminoacylation catalyst based on RNA. *Chem. Biol.* **10**, 655-662 (2003).

46. Niwa, N., Yamagishi, Y., Murakami, H. & Suga, H. A flexizyme that selectively charges amino acids activated by a water-friendly leaving group. *Bioorg. Med. Chem. Lett.* **19**, 3892-3894 (2009).

47. Ohta, A., Murakami, H., Higashimura, E. & Suga, H. Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* **14**, 1315-1322 (2007).

48. Goto, Y. et al. Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS Chem. Biol.* **3**, 120-129 (2008).

49. Kawakami, T. et al. Diverse backbone-cyclized peptides via codon reprogramming. *Nat. Chem. Biol.* **5**, 888-890 (2009).

50. Kawakami, T., Murakami, H. & Suga, H. Messenger RNA-programmed incorporation of multiple N-methyl-amino acids into linear and cyclic peptides. *Chem. Biol.* **15**, 32-42 (2008).

51. Yamagishi, Y. et al. Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library. *Chem. Biol.* **18**, 1562-1570 (2011).

52. Hayashi, Y., Morimoto, J. & Suga, H. In vitro selection of anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors. *ACS Chem. Biol.* **7**, 607-613 (2012).

53. Morimoto, J., Hayashi, Y. & Suga, H. Discovery of macrocyclic peptides armed with a mechanism-based warhead: isoform-selective inhibition of human deacetylase SIRT2. *Angew. Chem. Int. Ed. Engl.* **51**, 3423-3427 (2012).

54. Tanaka, Y. et al. Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature* **496**, 247-251 (2013).

55. Ito, K. et al. Artificial human Met agonists based on macrocycle scaffolds. *Nat. Commun.* **6**, 6373 (2015).

56. Matsunaga, Y., Bashiruddin, N.K., Kitago, Y., Takagi, J. & Suga, H. Allosteric Inhibition of a Semaphorin 4D Receptor Plexin B1 by a High-Affinity Macrocyclic Peptide. *Cell Chem Biol* **23**, 1341-1350 (2016).

57. Jongkees, S.A.K. et al. Rapid Discovery of Potent and Selective Glycosidase-Inhibiting De Novo Peptides. *Cell Chem Biol* **24**, 381-390 (2017).

58. Kawamura, A. et al. Highly selective inhibition of histone demethylases by de novo macrocyclic peptides. *Nat. Commun.* **8**, 14773 (2017).

59. Yu, H. et al. Macrocycle peptides delineate locked-open inhibition mechanism for microorganism phosphoglycerate mutases. *Nat. Commun.* **8**, 14932 (2017).

60. Xiao, H., Murakami, H., Suga, H. & Ferré-D'Amaré, A.R. Structural basis of specific tRNA aminoacylation by a small in vitro selected ribozyme. *Nature* **454**, 358-361 (2008).

61. Grundy, F.J., Winkler, W.C. & Henkin, T.M. tRNA-mediated transcription antitermination in vitro: codon-anticodon pairing independent of the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 11121-

11126 (2002).

62. Yousef, M.R., Grundy, F.J. & Henkin, T.M. tRNA requirements for glyQS antitermination: a new twist on tRNA. *RNA* **9**, 1148-1156 (2003).

63. Zhang, J. & Ferré-D'Amaré, A.R. Direct evaluation of tRNA aminoacylation status by the T-box riboswitch using tRNA-mRNA stacking and steric readout. *Mol. Cell* **55**, 148-155 (2014).

64. Köhrer, C. & Rajbhandary, U. The many applications of acid urea polyacrylamide gel electrophoresis to studies of tRNAs and aminoacyl-tRNA synthetases. *Methods* **44**, 129-138 (2008).

65. Saito, H., Watanabe, K. & Suga, H. Concurrent molecular recognition of the amino acid and tRNA by a ribozyme. *RNA* **7**, 1867-1878 (2001).

66. Caserta, E., Liu, L.C., Grundy, F.J. & Henkin, T.M. Codon-Anticodon Recognition in the Bacillus subtilis glyQS T Box Riboswitch: RNA-dependent Codon Selection Outside the Ribosome. *J. Biol. Chem.* **290**, 23336-23347 (2015).

67. Grigg, J.C. et al. T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7240-7245 (2013).

68. Zhang, J.W. & Ferre-D'Amare, A.R. Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* **500**, 363-366 (2013).

69. Grigg, J.C. & Ke, A. Structural determinants for geometry and information decoding of tRNA by T box leader RNA. *Structure* **21**, 2025-2032 (2013).

70. Li, S. et al. Structural basis of amino acid surveillance by higher-order tRNA-mRNA interactions. *Nat. Struct. Mol. Biol.* (2019).

71. Chan, P.P. & Lowe, T.M. GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res.* **44**, D184-189 (2016).

72. Sato, K., Hamada, M., Asai, K. & Mituyama, T. CENTROIDFOLD: a web server for RNA secondary structure prediction. *Nucleic Acids Res.* **37**, W277-W280 (2009).

73. Kobori, S. & Yokobayashi, Y. High-Throughput Mutational Analysis of a Twister Ribozyme. *Angew. Chem. Int. Ed. Engl.* **55**, 10354-10357 (2016).

74. Dhamodharan, V., Kobori, S. & Yokobayashi, Y. Large Scale Mutational and Kinetic Analysis of a Self-Hydrolyzing Deoxyribozyme. *ACS Chem. Biol.* **12**, 2940-2945 (2017).

75. Rogers, J.M., Passioura, T. & Suga, H. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 10959-10964 (2018).

76. Fowler, D.M. et al. High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741-746 (2010).

77. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **6**, 779-790 (2011).

78. Goto, Y., Iseki, M., Hitomi, A., Murakami, H. & Suga, H. Nonstandard peptide expression under the genetic code consisting of reprogrammed dual sense codons. *ACS Chem. Biol.* **8**, 2630-2634 (2013).

79. Guo, X. et al. RNA-dependent folding and stabilization of C5 protein during assembly of the E. coli RNase P holoenzyme. *J. Mol. Biol.* **360**, 190-203 (2006).

List of accomplishments

【Publication(s) related to the thesis】

1. “Ribosomal Synthesis of Norbornene Containing Peptides for Peptide Hetero-Dimerization”, Satoshi Ishida, Takayuki Katoh, Hiroaki Suga, Peptide Science 2017, 2018, 66-67

2. “An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch”, Satoshi Ishida, Naohiro Terasaka, Takayuki Katoh, Hiroaki Suga, *in press*

【Oral presentation】

1. “Ribozymes with tRNA recognition and aminoacylation properties”, Satoshi Ishida, Naohiro Terasaka, Takayuki Katoh, Hiroaki Suga, 26th tRNA conference, Jeju, Korea, September 2016

2. “An aminoacylation ribozyme evolved from a natural tRNA sensing T-box riboswitch”, Satoshi Ishida, Naohiro Terasaka, Takayuki Katoh, Hiroaki Suga, RNA2019 Tokyo, Tokyo, Japan, July 2019

【Poster presentation】

1. “In vitro selection of tRNA-recognizing aminoacylation ribozyme derived from a T-box motif.”, Satoshi Ishida, Naohiro Terasaka, Takayuki Katoh, Hiroaki Suga, RNA 2016, Kyoto, Japan, April 2016

2. “T-box リボスイッチに基づいた tRNA 認識アミノアシル化リボザイムの開発”, 石田 啓, 寺坂 尚紘, 加藤 敬行, 菅 裕明, 第 5 回バイオ関連化学シンポジウム若手フォーラム, 東京, 2017 年 9 月

3. “*In vitro* selection 法による tRNA 認識とアミノアシル活性を併せ持ったリボザイムの開発”, 石田 啓, 寺坂 尚紘, 加藤 敬行, 菅 裕明, 第 11 回バイオ関連化学シンポジウム, 東京, 2017 年 9 月

4. “Construction of a hetero-dimeric macrocyclic pepetide library for the discovery of pepetide ligands that induce IL28RA-IL10R2 hetero-dimerization” ,Satoshi Ishida, Takayuki Katoh, Hiroaki Suga, 第 54 回ペプチド討論会, 大阪, 2017 年 11 月

5. “Preparing hetero-dimeric library for the discovery of peptide ligands that induce IL28RA-IL10RB hetero-dimerization.”, Satoshi Ishida, Takayuki Katoh, Hiroaki Suga, Gordon Research Conference, Ventura, California, February 2018

6. “Toward the discovery of heterodimeric-macrocylic peptides that induce IL28RA activation.”, Satoshi Ishida, Takayuki Katoh, Hiroaki Suga, 10th International peptide symposium, Kyoto, Japan, December 2018

【Awards and fellowships】

1. Advanced Leading Graduate Course for Photon Science (ALPS) (2015-2019)

2. Research Fellowship for young scientists, the Japan Society of the Promotion of Science (2016-2019).

3. Best presenter award in RNA2019 Tokyo (July 2019).

Acknowledgements

I greatly appreciate Professor Hiroaki Suga for his support, mentorship and offering me with the best environment for pursuing my Phd.
I also would like to acknowledge Associate Professor Takayuki Katoh and Yuki Goto for their advice and discussion. I would also like to thank Project Assistant Professor Naohiro Terasaka for not only his academic but also personal advices and countless discussion throughout my Phd program.

I am grateful to all the Suga lab members both alumni and current members for making me feel at home whenever I am in the lab.

I would like to thank the Japan Society for the Promotion of Science (JSPS) and Advanced Leading Graduate Course for Photon Science (ALPS) for a grant and financial support.
Lastly, I would like to thank my family member and friends for their support and cheering.

Satoshi Ishida