学位論文

Applications of aminoacylation ribozymes that recognize the 3’-end of tRNA via two consecutive base pairs

（tRNA の 3’末端を塩基対合によって認識する

アミノアシル化リボザイムの応用）

平成 26 年 12 月博士（理学）申請

東京大学大学院理学系研究科

化学専攻

寺坂　尚紘

Abstract

In this thesis, I applied aminoacylation ribozyme “flexizyme” to discover the novel interaction between microRNA precursor and metabolite, and to develop the orthogonal translation machinery consisting of the pair of ribosome and tRNAs bearing mutations at the peptidyl transferase center. Flexizymes have following unique characteristics; (i) substrate RNA is recognized by two consecutive base pairs between 3′-end of substrate RNA and 3′-end of flexizyme, (ii) these base pairs can be substituted with other base pairs and (iii) various activated amino acids can be used as substrates including both canonical and noncanonical amino acids. This flexible aminoacylation of RNAs by flexizymes were used not only as a method for labeling endogenous tRNAs to be removed but also as a tool to prepare various aminoacyl-tRNAs bearing mutations at 3′-end to engineer the translation machinery.

Chapter 1 is a general introduction of this thesis. Transfer RNA, translation and small non-coding RNA such as miRNA are firstly introduced, and the problems and questions in these research field are also described. Then aminoacylation ribozyme “flexizyme” having unique characteristics is introduced, which has a potential to overcome these problems.

In chapter 2, novel interaction between human precursor microRNA 125a and folic acid is described. This interaction was discovered by small RNA transcriptomic SELEX (Systematic Evolution of Ligands by EXponential enrichment) using tRNA-depleted human small RNA library. Depletion of tRNAs from small RNA fraction was achieved by aminoacylation of tRNAs with biotin-conjugated phenylalanine using flexizymes. Measurement of binding affinity of precursor microRNA 125a mutants revealed essential motif for binding to folic acid. To my knowledge, this is the first example of microRNA precursors binding to metabolite.

In chapter 3, orthogonal translation machinery consisting of ribosome-tRNAs pair bearing compensatory mutations at the peptidyl transferase center is reported. The method to easily prepare various aminoacyl tRNA bearing mutations at 3′-end was developed by using compensatory mutated flexizymes. Next, translation activity of several mutant ribosome and tRNAs were investigated using FIT (Flexible In vitro Translation) system. Consequently, it was discovered that a certain mutant pair of ribosome and tRNAs had comparable translation activity and orthogonality to the wild-type pair of ribosome and tRNAs. Finally two different peptides was expressed simultaneously in one pot from a single mRNA template using the mutant and wild-type

ribosome–tRNA pair.

Chapter 4 is the conclusion of the entire thesis. Achievements of the studies and new insights obtained in this thesis are included. Finally, perspective of the potential application of aminoacylation by flexizymes for other studies, regulation of biogenesis of microRNAs by small molecules and engineering of translation machinery are discussed.

Table of contents

Chapter 1

General introduction

1.1. Transfer RNA

Genetic information is transcribed from DNA into mRNA (messenger RNA) and mRNA is translated into protein, and this process is called as central dogma (Figure 1). By human genome project, it was reported that only the 2% region of human genomic DNA is transcribed and translated to about 20,000-25,000 proteins[1], but transcriptome study revealed that about 70% region of DNA is transcribed to RNA. In addition, widespread transcription of non-coding RNAs with important functions throughout the genome have been recently discovered[2].

Transfer RNA (tRNA) is one of the most abundant and popular non-coding RNAs, which works as an adaptor molecule linking the language of nucleotide to the language of amino acid in the translation step. Canonical tRNA forms cloverleaf secondary structure and L-shaped tertiary structure as shown in Figure 2. Aminoacyl-tRNA sythetase (ARS) charge corresponding amino acid onto the 3′-end of specific tRNA to produce aminoacyl-tRNA (aa-tRNA) and aa-tRNA is recruited into ribosome by translation factors (Figure 1). Three consecutive nucleotides called as anticodon of tRNA forms base pairs with the sequential triplet codons of mRNA in the ribosome, and corresponding amino acid is incorporated into nascent peptide chain (Figure 1 and 2). The correspondence between codons and amino acids are defined as genetic code (Figure 3). This genetic code is well conserved in the most of all organisms (some exceptions have been found[3]) and accurate translation of mRNA to protein is essential for living organisms.

CCA sequence at the 3′-end are universally conserved in all three domains, prokaryote, eukaryote and archaea. During the translation reaction, CCA-3′ end interacts with ribosome and translation factors, and it is important for efficient translation. In the aminoacylation step, ARSs strictly recognize the body sequences of substrate tRNAs to charge specific amino acids onto specific tRNAs. In addition, because most ARSs also interact with the universally conserved CCA-3′ end of tRNAs[4,5], it is difficult to prepare the various aa-tRNAs bearing mutations in the CCA-3′ end by ARSs[6-8]. To prepare such aa-tRNAs, so-called “chemical acylation” method where chemically synthesized aminoacyl-nucleotide was ligated to tRNA lacking the 3′-end by T4 RNA ligase[9,10] may be usable. However, in this method, it is laborious work that various aminoacyl-nuleotides are prepared by chemical synthesis. If the easy method to prepare various aa-tRNA bearing mutation in the CCA-3′ end, it is usable for analyzing the role of CCA-3′ end during the translation and engineering the translation machinery.

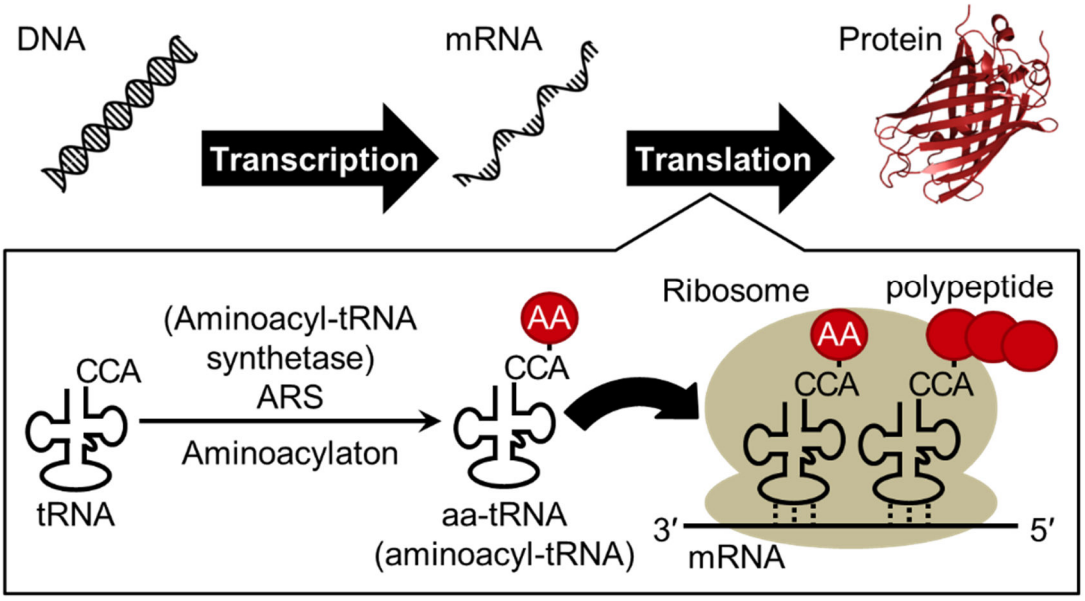

Figure 1 | Schematic illustration of central dogma. "AA" in the red circle indicate amino acid. An image of protein was prepared from crystal structure of GFP (green fluorescent protein) (PDB ID: 1GFL)[11].

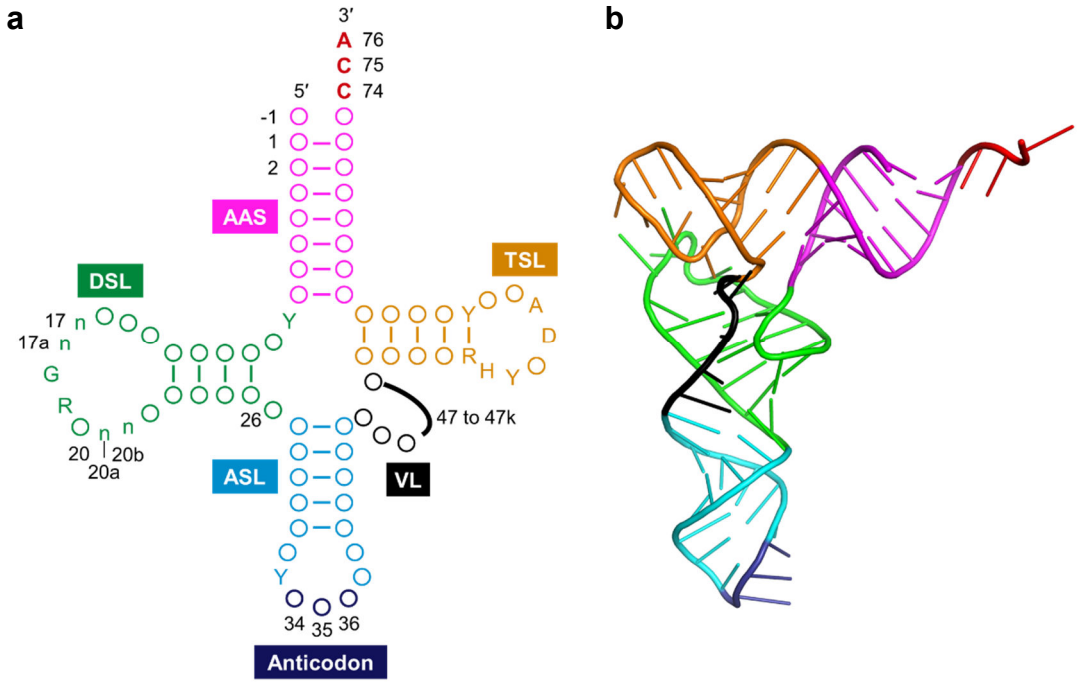

Figure 2 | Structure of tRNA. (**a**) Clover leaf structure of tRNA. Conserved bases are described, circles represent non-conserved bases, and numbers indicate the nucleotide position. Positions 34 to 36 (dark blue) form corresponds to the anticodon and position 74 to 76 (red) are universally conserved as CCA. The sequences composing the intron and the extra bases presenting in the variable loop (position noted 47 to 47k) are shown as black line. AAS (magenta), the amino acid-accepting stem; DSL (green), the dihydrouridine stem and loop; ASL (cyan), the anticodon stem and loop; VL (black); the variable loop, TSL (orange); the thymidine stem and loop, Y; pyrimidine, R; purine, H; not G, D; not C. The n bases at position 17, 17a, 20a and 20b are optional bases not present in all tRNAs. Position -1 bases are found in all cytoplasmic mature $tRNA^{His}_{GUG}$ from the three biological domains. This figure was adapted from the previous paper[12]. (**b**) L-shaped tertiary structure of tRNA. Each color indicates the same region described in Figure 2a. This image was prepared from crystal structure of *Saccharomyces cerevisiae* $tRNA^{Phe}$ (PDB ID: 1EHZ)[13].

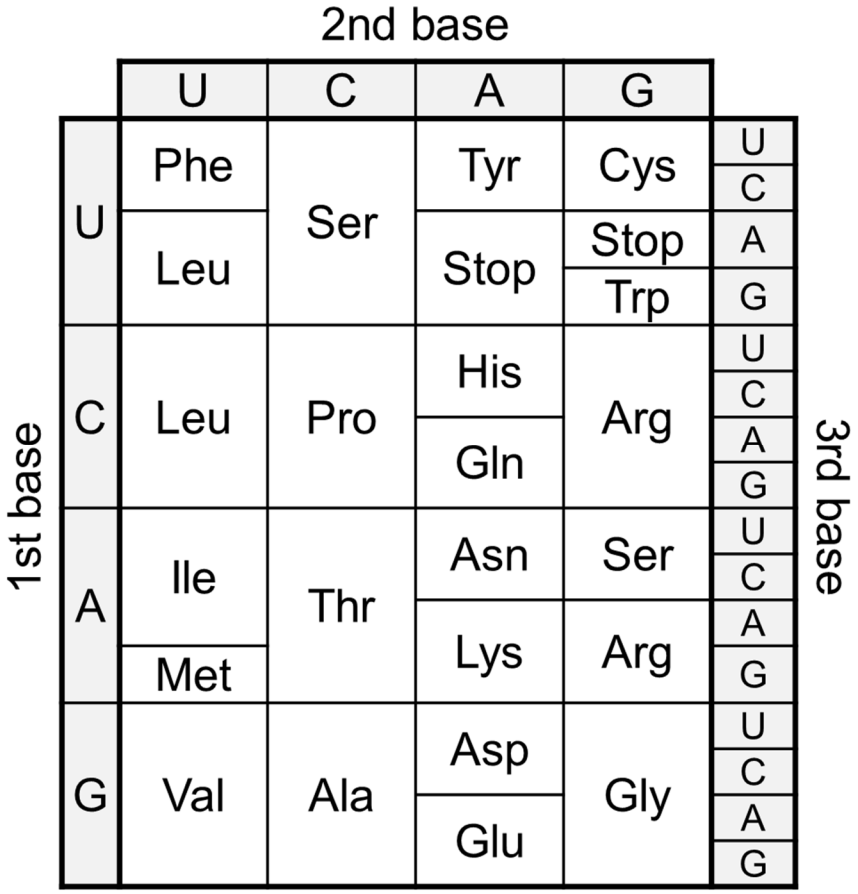

Figure 3 | Standard genetic code. All 64 codons were assigned to corresponding amino acids or termination signal (termed as "Stop" in this figure).

1.2. Other non-coding RNAs than tRNAs

As I mentioned above, tRNA is one of the most abundant and popular non-coding RNA (ncRNA). In addition to tRNA, ribosomal RNAs (rRNAs) are also known as popular ncRNA, which constitute a ribosome. There are three kinds of rRNAs (5S, 16S and 23S rRNAs) in prokaryote, and four kinds of rRNAs (5S, 5.8S, 18S and 28S rRNA) in eukaryote. Peptidyl transfer reaction, which makes a peptide bond, is catalyzed by 23S rRNA or 28S rRNA. These tRNA and rRNA are much abundant and account for about 95% of all RNAs in mammalian cells[14].

The various other ncRNAs than tRNA and rRNA have been identified mainly in mammalian[15]. These ncRNAs are roughly classified into two groups; long ncRNA (longer than 200 nt) and small ncRNA (smaller than 200 nt)[16,17]. Unlike small ncRNA such as microRNA, function of long ncRNA has been less studied. However, recent studies revealed the broad functional repertoire including roles in high-order chromosomal dynamics, telomere biology and subcellular structural organization[18,19].

The class of small ncRNA includes various RNAs such as snRNAs (small nuclear RNAs), snoRNAs (small nucleolar RNAs) and miRNAs (micro RNAs)[20]. snRNA makes complex with protein to form snRNP (ribonucleoprotein) and consists of spliceosome involving the reaction of mRNA splicing. snoRNA generally ranges from 60 to 300 nucleotides in length and guide the site-specific modification of nucleotide in target RNA[21]. In the past few years, about 22 nt short ncRNA (miRNA) has been widely studied. miRNA is loaded into Ago protein to form a RISC (RNA induced silencing complex)[22]. A RISC binds to target mRNA via base pairs between miRNA and 3′-UTR (untranslated region) of mRNA to inhibit the translation and suppress the expression of the target gene. To date more than 1,800 miRNAs in human are registered in databases (miRbase[23], http://www.mirbase.org/) and are estimated to regulate the expression of more than 60% of genes to control a wide range of biological processes including proliferation, differentiation, aptosis and development[15].

As mentioned above, various small ncRNAs have been identified and the function of those RNAs have been studied. However, it is not easy to discover the small ncRNAs with very low abundance because tRNAs are too much abundant in small RNA fraction (less than 200 nt). To overcome this problem, easy method to remove tRNAs from small RNA fraction is required.

1.3. Aminoacylation ribozyme "flexizyme"

In order to achieve these requirements described in former sections; (i) to easily prepare the various aa-tRNAs bearing mutations in the CCA-3′ end, and (ii) to remove endogenous tRNAs from small RNA fraction, I focused attention on the unique characteristics of flexizymes which is *in vitro* selected aminoacylation ribozyme[24,25]. First, I briefly summarize how to develop the artificial ribozyme.

A ribozyme which is an RNA having catalytic activity which was first discovered in nature in 1980s[26,27]. These discoveries indicated RNA can store information like DNA and catalyze chemical reaction like proteins to strongly support the "RNA world hypothesis" that the origin of life may have relied on RNA[28]. However, the chemical reactions catalyzed by natural ribozymes are limited to the cleavage or ligation of the RNA phosphodiester backbone except for peptidyl transfer by ribosome[29]. To expand the variety of chemical reactions catalyzed by RNA, *in vitro* selection of ribozyme have been developed. SELEX (Systematic Evolution of Ligands by EXponential enrichment, or *in vitro* selection) is a method based on molecular evolution engineering to identify the active molecules having specific activities from pools of various compounds[30,31]. It was first reported from two groups in 1990 and they discovered RNA aptamers binding to specific target molecules from random RNA library. In addition to discovery of aptamers, the molecules having catalytic activities such as ribozyme have also been identified[32]. A general scheme of SELEX of active RNAs was shown in Figure 4. DNA library containing T7 promoter sequence, 5′-constant sequence, random sequences and 3′-constant sequence is prepared by PCR. These constant sequences are used for following reverse transcription and PCR reaction. The DNA templates are transcribed to RNAs by T7 RNA polymerase *in vitro*. From the RNA library, portion of RNAs are selected based on their ability to carry out a specific function (*e.g.* binding affinity and catalytic activity). Selected RNAs are converted to cDNAs by reverse transcription. These cDNAs are amplified by PCR and the same procedure is repeated. After the several times repeats of this process, the resulting RNA fraction will be enriched with functional molecules with the desired activity. By SELEX, many artificial ribozymes have been discovered[33] (*e.g.* RNA polymerization[33-35], alcohol oxidation[36], Diels-Alder reaction[37] and aminoacylation[38-42]). Among them, aminoacylation ribozymes were important because these are candidates as key molecules which link the ancient RNA world with modern protein world.

In nature, ARS catalyze aminoacylation reaction. Aminoacylation by ARS involves two steps; (i) activation of carboxyl group amino acids by adenylation using ATP

to yield aminoacyl-adenosine monophosphate (aminoacyl-AMP), and (ii) condensation of aminoacyl-AMP and tRNA to yield aminoacyl-tRNA (aa-tRNA). Although no ribozyme catalyzing these both reactions have been discovered in nature, some artificial ribozymes catalyzing (i) or (ii) step were identified by *in vitro* selection. The KK13 ribozyme catalyze the activation of amino acid by using the 5′-triphosphate of RNA instead of ATP[43]. The #29 ribozyme family catalyzes *cis*-acyl transfer reaction using aminoacyl-AMP to produce both aminoacyl- and peptidyl-tRNAs[38,44,45], and the C3 ribozyme family consisting of only three essential nucleotides catalyze both *cis*- and *trans*-acyl transfer reaction using aminoacyl-AMP[42,46]. Flexizyme is also *trans*-acyl-transfer ribozyme using activated amino acids, which was developed by SELEX method and was engineered to aminoacylate various activated amino acids. There are three types of flexizymes, dFx (dinitrobenzyl flexizyme)[41], eFx (enhanced flexizyme)[41] and aFx (amino flexizyme)[47] (Figure 5). dFx charges wide variety of amino acid whose carboxyl group is activated with 3,5-dinitrobenzyl ester (DBE), eFx charges aromatic amino acids activated with cyanomethyl ester (CME) or nonaromatic amino acids with 4-chlorobenzyl thioester (CBT), and aFx charges substrates activated with benzyl thioester group bearing a protonated primary amine (ABT).These flexizymes can charge wide variety of amino acids including noncanonical amino acids (ncAAs) such as *N*-methyl-amino acids[48-50], *N*-alkyl-glycines[51], cyclic *N*-alkyl amino acids[52], *N*-acyl-amino acids[53], exotic peptides[54], α-hydroxy acids[49,55], and D-amino acids[56] onto tRNAs.

In addition to the characteristic to charge diverse kinds of amino acids, flexizymes have the flexibility to aminoacylate the substrate tRNA bearing any body sequence because flexizymes recognize the CCA-3′ end by two consecutive base pairs with GGU-3′ of flexizymes[40,57,58] (Figure 5 and 6). The crystal structure of flexizyme fused with substrate minihelix RNA docked onto the tRNAs indicated the no interaction between flexizyme and body sequence of substrate tRNA[57] (Figure 6). This is because, fleixizyme can aminoacylate other RNAs different from tRNAs even if those have single strand CCA-3′ end. In addition, in the experiment using prototype of flexizymes “pre-24”, these two consecutive base pairs are important for aminoacylation and this base pairs can be substituted with certain base pairs[40,58].

In summary, flexizymes have following unique characteristics; (i) substrate RNA is recognized by two consecutive base pairs between 3′-end of substrate RNA and 3′-end of flexizyme, (ii) these base pairs can be substituted with other base pairs and (iii) various activated amino acids can be used as substrates including both canonical and noncanonical amino acids. These features are different from natural protein ARSs and

other aminoacylation ribozymes. From these characteristics of flexizymes, I considered that flexizymse enable to achieve the requirements described early in this section. In other words, flexizymes enable to label all endogenous tRNAs bearing CCA-3′ end with ncAAs to be removed, and to aminoacylate CCA-3′ mutated tRNAs using compensatory mutated flexizymes.

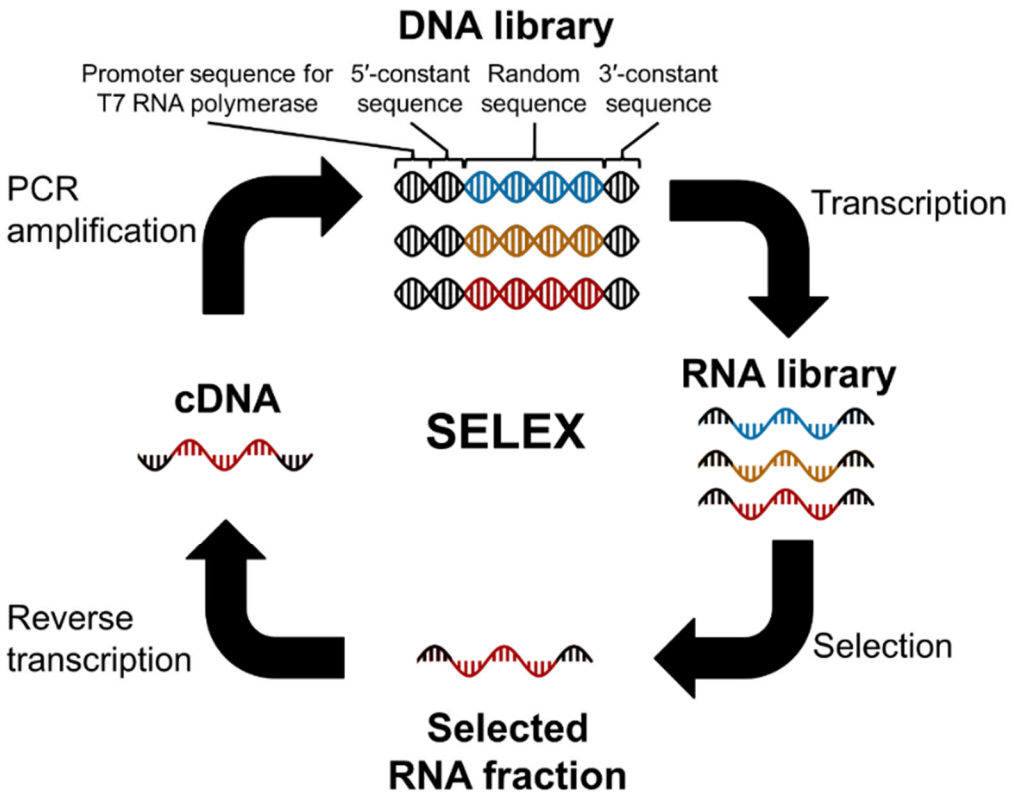

Figure 4 | General scheme of SELEX (*in vitro* selection) of functional RNAs. Black sequences are constant regions including the promoter sequence for T7 RNA polymerase and the 5′/3′-constant sequences necessary for reverse transcription and PCR amplification. Colored (blue, orange and red) sequences indicate the random region. This figure is adapted from the previous review[25].

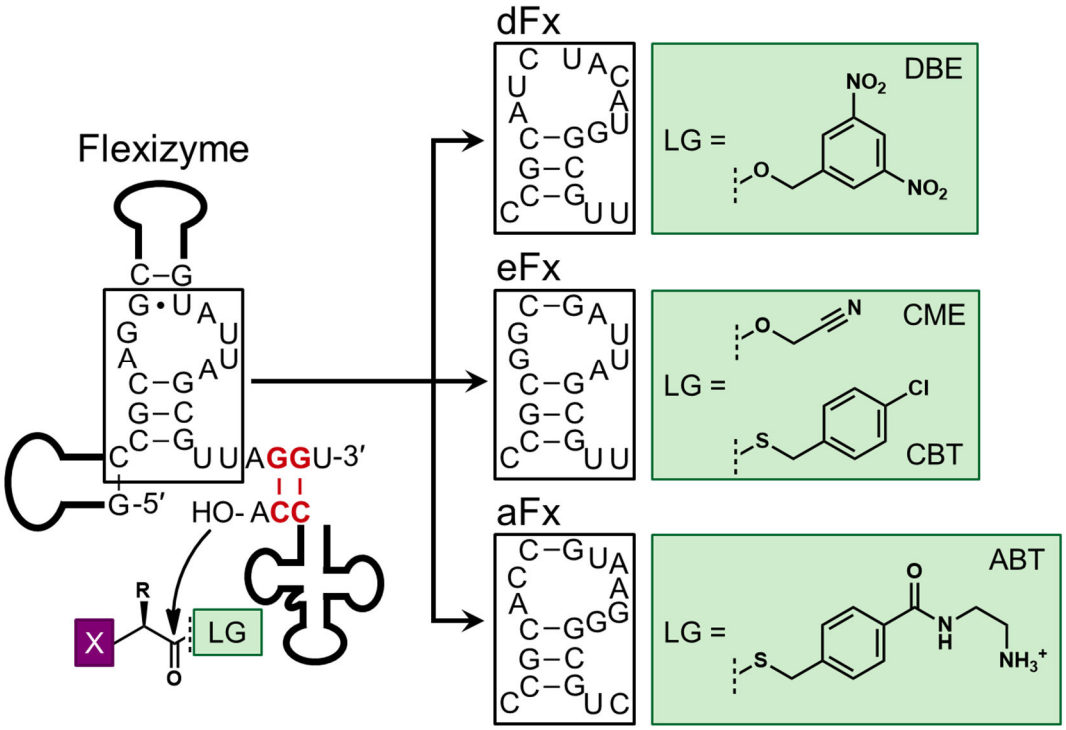

Figure 5 | Schematic illustration of aminoacylation by three derivatives of flexizymes and acyl-donor substrates with cognate leaving group. This figure is adapted from previous review[25]. Lines denote the Watson-Crick base pairs and bullets denote wobble base pairs. R represents amino acid side chains including noncanonical ones. X represent a-functional groups, such as amino, *N*-alkylamino, and hydroxy groups. LG; leaving group, dFx; dinitrobenzyl flexizyme, eFx; enhanced flexizyme, aFx; amino flexizyme, DBE; dinitrobenzyl ester, CME; cyanomethyl ester, CBT; *p*-chlorobenzylthioester, ABT; amino-derivatized benzyl thioester.

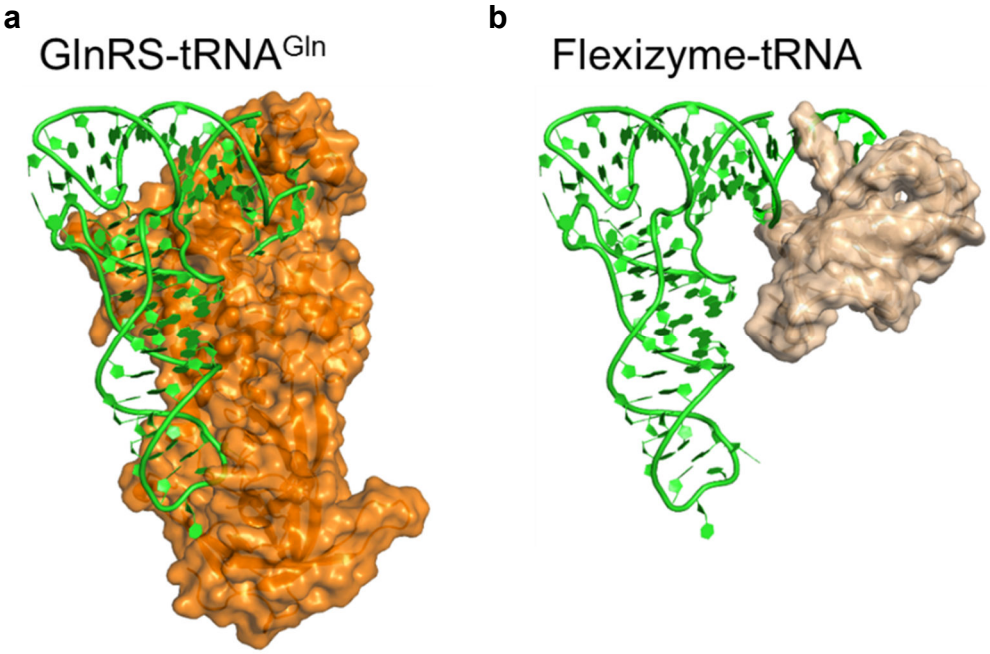

Figure 6 | Structures of protein and RNA aminoacyl-tRNA synthetases. (**a**) Crystal structure of *Escherichia coli* glutamiyl-tRNA synthetase (GlnRS; orange) and tRNA^{Gln} complex (PDB ID: 1GTS)[59]. (**b**) Crystal structure of flexizyme (wheat color) docked onto *E. coli* tRNA^{Gln} by superimposing the minihelix on the ASL of the tRNA (PDB ID: 3CUL)[57].

1.4. Discovery of human microRNA precursor binding to folic acid (a brief introduction of chapter 2)

Micro RNA (miRNA) is about 22 nt single strand RNA, which is one of the small ncRNAs[60]. miRNAs are loaded into an Argonaute protein to form the RNA-induced silencing complex (RISC) and RISC binds to target mRNA via base paring with miRNA to regulate the expression of target gene. miRNAs are transcribed from genome as part of a long primary transcripts (pri-miRNAs) and then pri-miRNAs are cleaved to produce precursor miRNAs (pre-miRNAs) which form hairpin structure. Pre-miRNAs are exported to the cytoplasm and they are cleaved to mature miRNA. Biogenesis and functions of miRNAs is regulated by proteins and RNAs in many steps. Recently, regulation of miRNA functions by binding with small molecules were discovered in human[61]. This discovery suggests the existence of RNA aptamer elements binding to small molecules with specific biochemical activities like in various human miRNAs.

In chapter 2, I will report the discovery of human microRNA precursor binding to metabolite and analyses of the function of this RNA (Figure 7). In my master course tRNA and rRNA-depleted human small ncRNA library was constructed by aminoacylation of tRNA with biotin-conjugated amino acids using flexizymes. Then, SELEX to folic acid was performed to obtain three kinds of RNAs. In doctor course I measured the binding affinity of these RNAs and discovered that precursor microRNA 125a (hsa-pre-miR125a) bound to folic acid. Then, further mutation study revealed the essential motif of hsa-pre-miR125a for binding to folic acid.

1.5. Development of orthogonal translation machinery (a brief introduction of chapter 3)

The CCA-3′ sequence of tRNA is conserved among all organisms and is important for aminoacylation and translation. In bacteria, the CCA-3′ end makes Watson-Crick base pairs with bases of 23S rRNA in the peptidyl transferase center (PTC) in the classical state[62] and translocation[63] during translation. These base pairs are important for translation activity, and the compensatory mutations in these base pairs are tolerated during peptidyl transfer reaction[64] and translocation[63]. However, it is yet empirically unknown whether such mutations accommodates translation in its entirety.

In chapter 3, I will report development of orthogonal translation machinery[65]. First, I developed method to easily aminoacylate various amino acids onto tRNAs bearing mutation (or mutations) in CCA-3′ end using compensatory mutated flexizymes. Then

the translation activity of the PTC-mutated ribosomes and tRNAs were measured. Consequently, I developed the mutant ribosome and tRNA pair which had comparable translation activity and orthogonality to the wild-type ribosome and tRNA pair. Finally two different peptides was expressed simultaneously in one pot from a single mRNA template using the mutant and wild-type ribosome–tRNA pair.

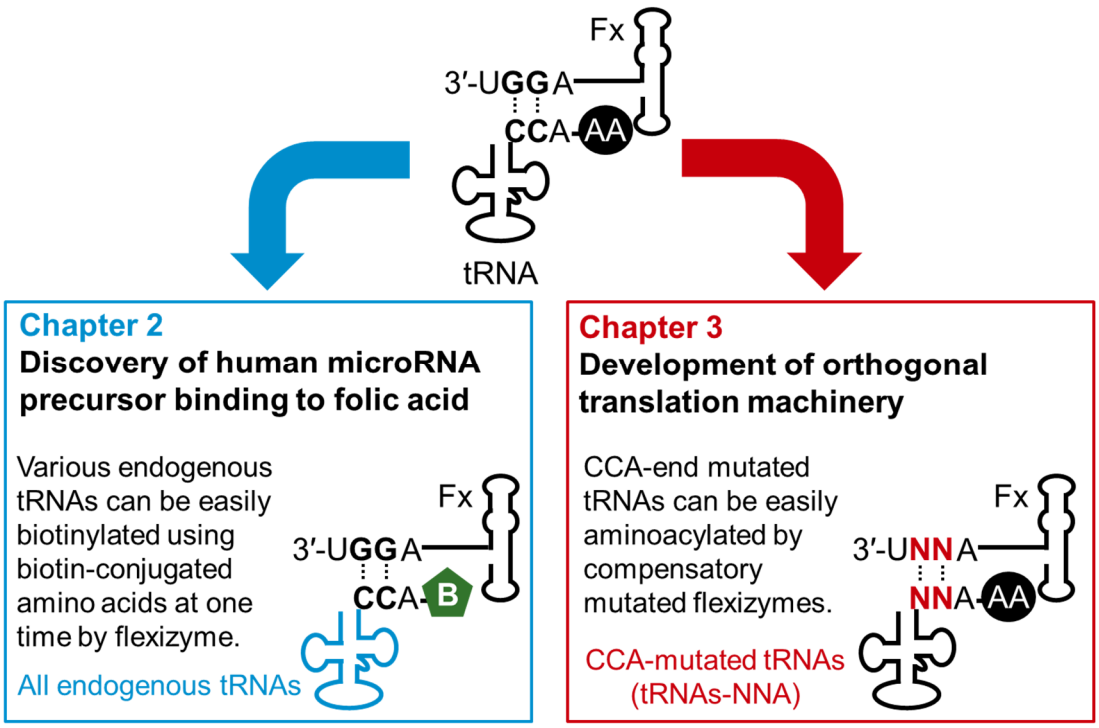

Figure 7 | Overview of the thesis. Discovery of human microRNA precursor binding to folic acid in chapter 2, and development of orthogonal translation machinery in chapter 3.

Chapter 2

Discovery of human microRNA precursor binding to folic acid by small RNA transcriptomic SELEX

2.1. Introduction

Various small non-coding RNAs (ncRNAs) have been discovered recent years. Among them, the functions and biogenesis of microRNA (miRNA) has been extensively studied. miRNAs are about 22 nt RNAs pairing to mRNAs to regulate the translation. Canonical pathway of miRNA processing in mammalian was shown in Figure 8[66]. A miRNA is initially transcribed as part of a long primary transcript, primary microRNA (pri-miRNA) by RNA polymerase II or III[67-69]. In mammalian, pri-miRNAs are cleaved by Drosha-DGCR8 microprocessor complex in nucleus to produce about 60-80 nucleotide precursor microRNA (pre-miRNA) forming hairpin structure[70-74]. The pre-miRNA is then exported to the cytoplasm by Exportin-5-RanGTP[75,76]. In the cytoplasm pre-miRNA is cleaved to about 22 nt mature miRNA duplex by the RNase III enzyme Dicer[77] in complex with the double stranded RNA binding protein TRBP[78]. One strand of the duplex is loaded into an Argonaute protein to form the RNA-induced silencing complex (RISC) and the other strand is discarded[79-81]. RISC binds to target mRNA via base paring with miRNA, whereas Argonaute protein functions as effectors by recruiting factors that induce translational repression, mRNA deadenylation and mRNA decay[82].

Biogenesis and functions of miRNAs is regulated by proteins and RNAs in many steps including their transcription, their processing by Drosha and Dicer, their loading onto Argonaute proteins and miRNA turn over, and their dysregulation which is associated with many human diseases including cancer[60,83]. Recently it was reported that direct binding of specific polyphenols (epigallocatechin gallate and resveratrol) to hsa-miR-33a-5p and hsa-miR-122-5p ("hsa" indicates *Homo sapiens*, "5p" indicates mature miRNA derived from 5′ end of pre-miRNA) and modulate divergently their levels in hepatic cells (Figure 9)[61]. This is the first report about posttranscriptional regulation of miRNAs by binding of natural small molecules. In addition to these miRNAs, the RNA elements in human mRNA binding to ATP or GTP have been discovered by genomic SELEX (Figure 10)[84,85]. From these discoveries, I considered that more RNA aptamer elements binding to metabolites with specific biochemical activities exist in other human small ncRNAs. In this chapter 2, I aimed to discover human small ncRNA including microRNAs and these precursors which bind to metabolite.

As many as 145 miRNAs, involved in the control of a variety of carcinogenesis mechanisms, were modulated by natural agents, including vitamins, oligoelements, polyphenols, isoflavones, indoles, isothiocyanates, phospholipids, saponins, anthraquinones and polyunsaturated fatty acids[86]. Although the detailed mechanisms how to modulate expression of miRNAs by these molecules are unknown, there is a possibility

that some of these molecules directly bind to miRNAs or those precursors to regulate the functions. Among these natural agents, I focused attention on folic acid (vitamin B9), which is one of the water soluble vitamins. The term "folic acid" refers to the synthetic form of the vitamin whereas "folate" refers to the natural forms, such as those present in food[87]. Folate includes oxidized, reduced, polyglutamylated, methylated and formylated forms of folic acid[88]. Deficiency of dietary folate has been linked to developmental anomalies[89] as well as increased risk for a number of cancers[90]. In addition, folic acid deficient culture medium change the expression level of many human miRNAs and folic acid supplementation in the culture medium restored miRNA levels, which indicate that modulation of miRNAs by folate is reversible[91]. From these results, I chose folic acid as a candidate molecule binding to miRNA.

In order to discover the certain RNAs interacting with small molecules or proteins, pull down method, computational prediction[92] and genomic SELEX[93] have been widely used. These approaches have been very successful, but they have some limitations. By classical pull down method, although RNAs expressing *in vivo* can be discovered, it is difficult to detect RNAs with low abundance *in vivo*. Computational predictions rely primarily on conservation and structural stability as signal for an active molecule, thus constraining the range of possible predictions[94,95]. Genomic SELEX is one of the SELEX method, where libraries are derived from genomic DNA. This methodology enable to identify new RNA aptamer elements even if the RNA is low abundant. However post-transcriptionally processed RNA sequence (*e.g.* splicing, nontemplated addition of nucleotides[96] and circular permutation[97]) cannot be obtained by genomic SELEX because the library derived from genome. In addition, obtained sequence may not be transcribed from genome *in vivo*. To overcome these problems, I decided to perform SELEX using library prepared from natural transcript, which is called as cDNA-SELEX[98,99] or transcriptomic SELEX[100]. This method enable to discover the small molecule-binding RNAs transcribed *in vivo* even if the level of expression is very low.

In the small RNA fraction (shorter than 200 nt), tRNAs and rRNAs are much abundant[14]. Therefore, it is necessary to remove tRNAs and rRNAs from small RNAs in order to construct cDNA library of high quality, which enable every potential small ncRNAs binding to metabolite equal candidacy for selection. Although several kits to remove the rRNAs using hybridization probes or ribonuclease are commercially available, the reported methods to deplete tRNAs require that each sequences are individually identified and preparation of about 30 different probes[101,102]. In my master course, tRNAs in the small RNA fraction were labeled with biotin-conjugated amino acids using

flexizymes and were depleted. From this small RNA library, cDNA library for SELEX was constructed and then selection to folic acid was performed to obtain three kinds of RNA sequences. To my knowledge, this is the first example of transcriptomic SELEX using small RNA library and this method of SELEX was named as small RNA transcriptomic SELEX.

In this chapter, the study during my master course will be briefly summarized and I will report binding analyses of RNAs obtained from small RNA transcriptomic SELEX performed during my Ph.D. course.

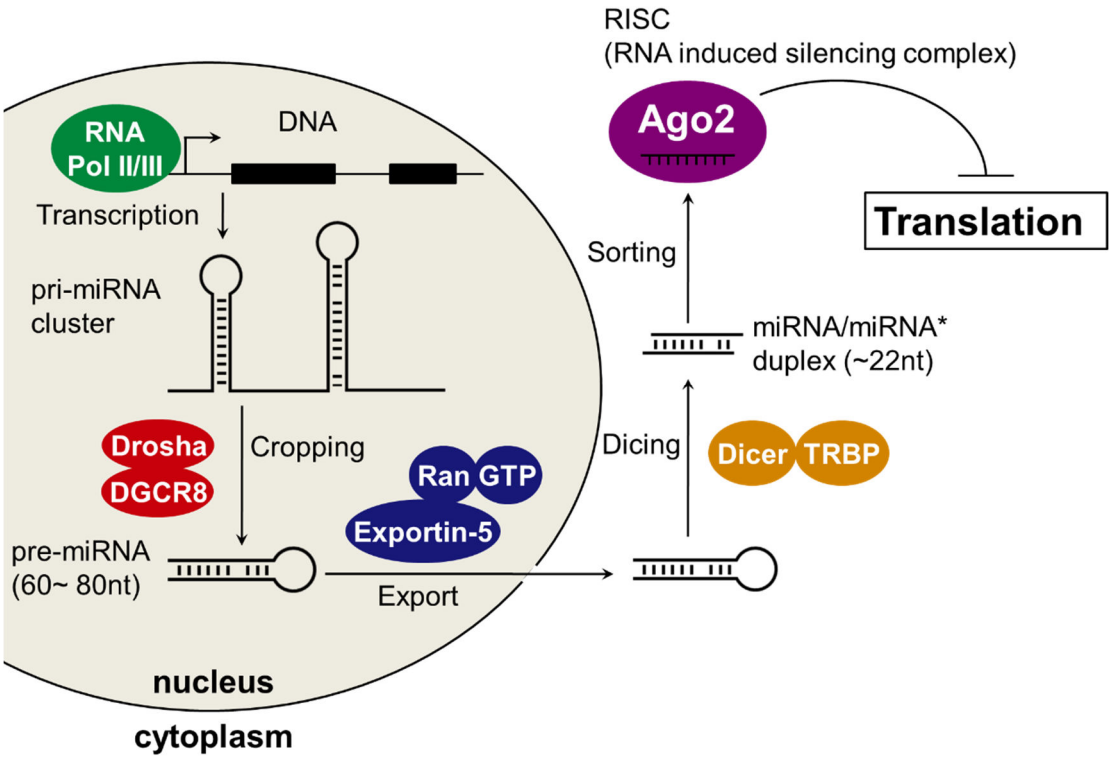

Figure 8 | Canonical pathway of microRNA processing. A miRNA is initially transcribed as part of a long pri-miRNA by RNA polymerase II or III. In mammalian, pri-miRNAs are cleaved by Drosha-DGCR8 microprocessor complex in nucleus to produce about 60-80 nucleotide pre-miRNA forming hairpin structure. The pre-miRNA is then exported to the cytoplasm by Exportin-5-Ran-GTP. In the cytoplasm pre-miRNA is cleaved to about 22 nt mature miRNA duplex by Dicer in complex with TRBP. One strand of the duplex is loaded into an Argonaute protein (Ago2) to form RISC. RISC binds to target mRNA via base paring with miRNA, whereas Ago2 functions as effectors by recruiting factors that induce translational repression.

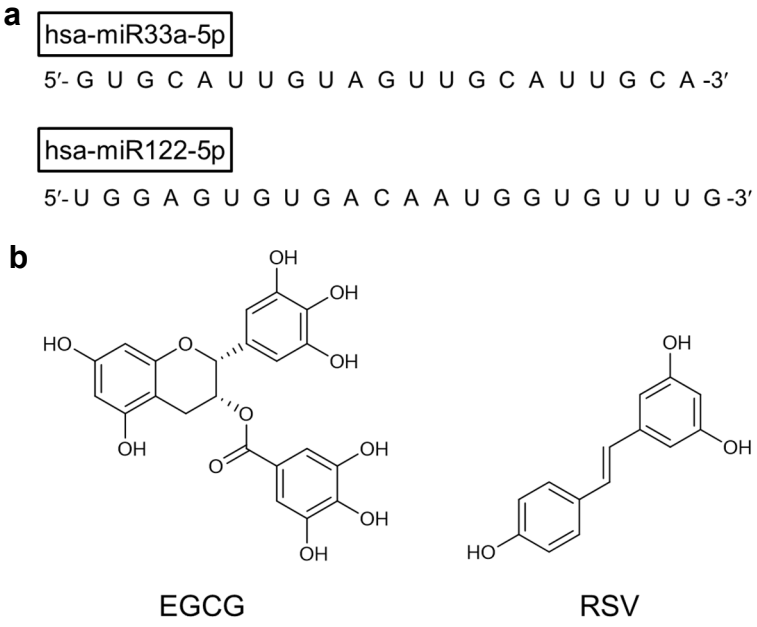

Figure 9 | Human miRNAs directly binding to polyphenols. (**a**) Sequences of hsa-miR33a-5p and hsa-miR122-5p which directly bind to specific polyphenols. (**b**) Chemical structures of epigallocatechin gallate (EGCG) and resveratrol (RSV).

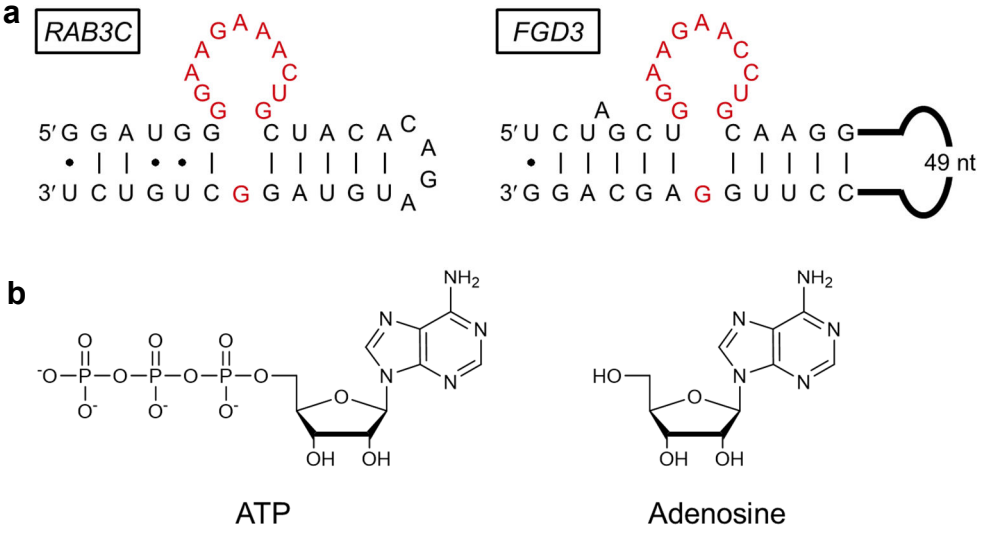

Figure 10 | Adenosine aptamers discovered in human transcripts. (**a**) Secondary structures of the regions containing adenosine aptamers of human *RAB3C* gene and human *FGD3* intron[85]. Red bases are essential motifs binding to adenosine. (**b**) Chemical structures of adenosine 5′-triphosphate and adenosine.

2.2. Results and discussions

2.2.1. Summary of master course study

In my master course, tRNA-depleted small RNA library was constructed and some RNAs binding to folic acid was discovered using small RNA transcriptomic SELEX. First, I briefly summarize the study during my master course.

Because the expression level of miRNAs and precursors are different among the cells, eight human cell lines (BeWo, HuH-7, VMRC-RCW, NEC14, HuO-3N1, RAJI, MCF7 and HeLaS3 cells) were chosen based on the expression levels of miRNAs to include various miRNAs into the library (Table 1)[103]. After extracting small RNAs from these cell lines, rRNAs were removed by hybridizing with complementary nucleotide probe and tRNAs were depleted using flexiyzme. Flexizymes specifically aminoacylate tRNAs via base pairs with CCA-end of tRNA using α-*N*-biotinyl-phenylalanine-CME (biotin-Phe-CME) and then biotin conjugated aa-tRNAs were captured by streptavidin magnetic beads (Figure 5 and 11). Consequently tRNA/rRNA-depleted small RNA library was constructed.

Next, tRNA/rRNA-depleted small RNA library was converted to cDNA library. RNA library was ligated with 3′- and 5′-adpators containing promoter sequence for T7 RNA polymerase (Figure 12a). Then, RNA library ligated with adaptors was reverse transcribed to cDNA and amplified by PCR (Figure 12a).

Folic acid was immobilized onto the magnetic beads via ester bond, whose surface is covered by hydroxyl group (Figure 12b). Small RNA transcriptomic SELEX using the folic acid-immobilized beads and the library described above was performed four rounds, and three RNA sequences (hsa-pre-miR125a, FA1 and FA2) were enriched (Table 2). FA1 is antisense sequence to mitochondrial mRNA of ND1 (NADH dehydrogenase 1) bearing additional poly-A/C sequence at the 3′ end, which is not coded in genome. FA2 is 5′ fragment of $tRNA^{Gly}_{GCC}$, whose 3′ fragment was discovered as small RNA (tRF3006)[104].

Because 3′- and 5′-adpators were ligated with naturally transcribed RNA sequences for reverse transcription and PCR during SELEX, it was possible that these adaptors are essential for binding to folic acid. The secondary structures predicted using mfold software[105] are shown in Figure 13. In the case of FA1 and FA2, ligation of adaptors dramatically changed the secondary structure (referred as FA1+adp and FA2+adp). On the other hand, the structure of hsa-pre-miR125a was retained when adpators were ligated

(named as hsa-pre-miR125a+adp). Therefore, it was necessary to investigate whether these RNAs can bind to folic acid without adaptor sequences.

Table 1 | Human miRNA precursor profile[103] and cell lines used in chapter 2

Organ systems and cell types	Numbers of miRNA precursors	Corresponding cell lines used in chapter 2
hsa_Osteosarcoma-U2Os-uninduced	168	HuO-3N1
hsa_Placenta	152	BeWo
hsa_renal_carcinoma-DH1-diff-3d	159	VMRC-RCW
hsa_Teratocarcinoma-NT2-norm	111	NEC14
hsa_Hepatoma-PLC	139	HuH-7
hsa_Breast-adenocarcinoma-MCF7	100	MCF7
hsa_Cervix-HeLa-IFNa	120	HeLaS3
hsa_Burkitt-patient3	128	RAJI

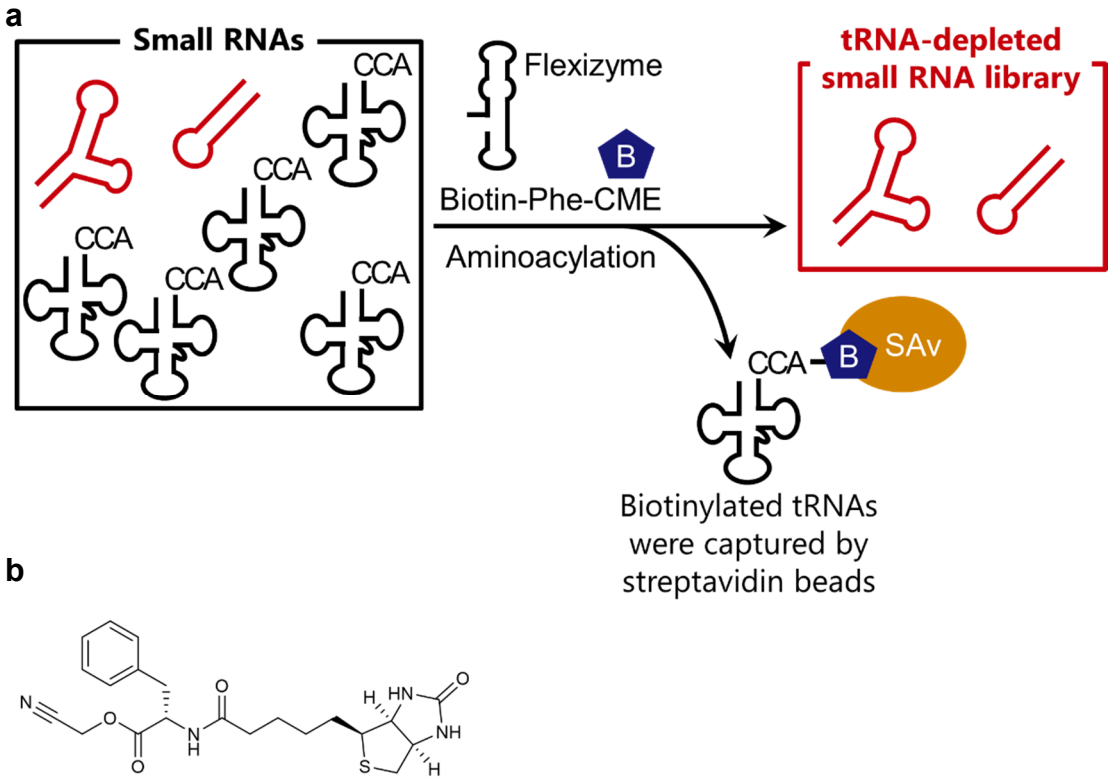

Figure 11 | Construction of a human tRNA-depleted small RNA library. (**a**) Schematic illustration of depletion of tRNAs. Small RNAs shorter than about 200 nt extracted from human cell lines contains large amount of tRNAs (black) and small amount of other small RNAs (red). Flexizyme aminoacylated tRNA with biotin-Phe-CME (deep blue pentagon), and biotinylated tRNAs were captured by streptavidin beads (orange circle). (**b**) Chemical structure of biotin-Phe-CME.

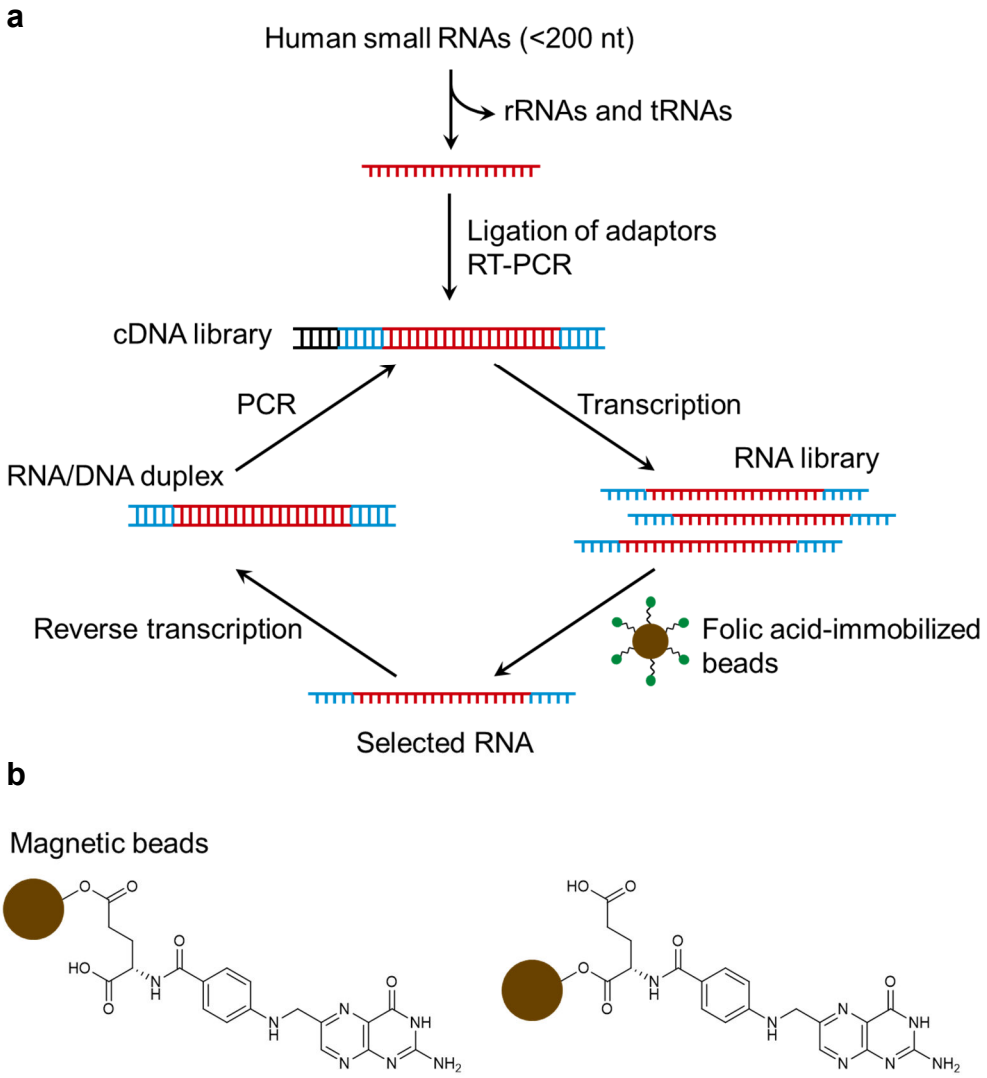

Figure 12 | Overview of small RNA transcriptomic SELEX to folic acid. (**a**) Ribosomal RNAs (rRNAs) and tRNAs were removed from human small RNAs using Ribominus^{TM} Eykaryote kit and flexiyzyme. Both 5′- and 3′-adaptors were ligated with RNA library and converted to cDNA library by RT-PCR. Standard SELEX was performed using this cDNA library and folic acid immobilized on magnetic beads. Red lines indicate sequences of natural transcript, cyan lines indicate adaptor sequences and black lines indicate T7 promoter sequences. (**b**) Structure of folic acid immobilized on the magnetic beads (brown circle). Folic acid was immobilized on the magnetic beads whose surface is hydroxyl group via ester bond. Because folic acid has two carboxyl group, two structures are possible.

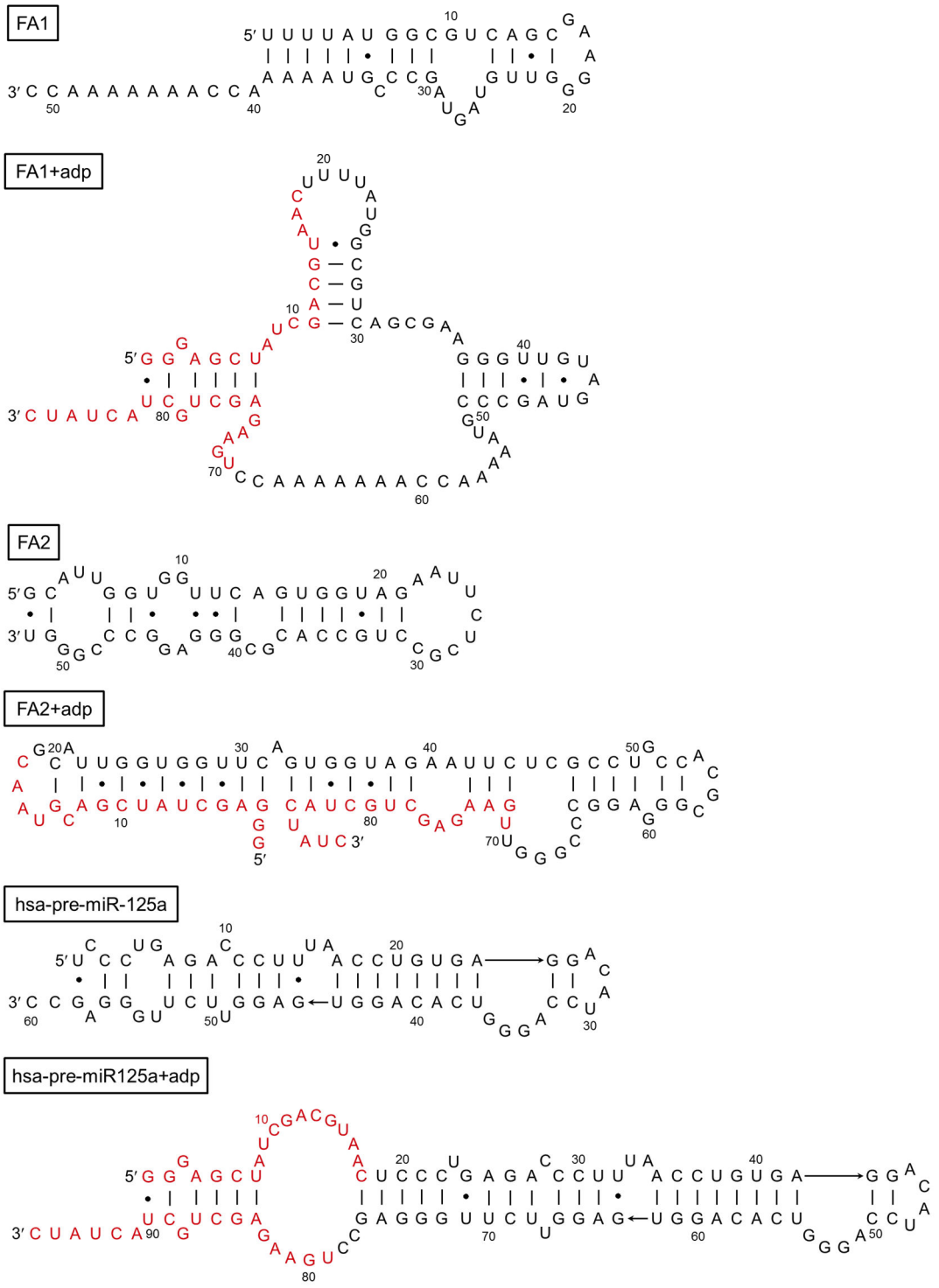

Figure 13 | Predicted secondary structure of RNAs obtained from SELEX to folic acid. Secondary structures were predicted using mfold software[105]. Red bases indicated adaptor sequences and numbers indicate the nucleotide positions. Lines denote the Watson-Crick base pairs and bullets denote wobble base pairs.

Table 2 | RNAs obtained from SELEX.

Sequence (5′ to 3′)	Remarks	frequency
TTTTATGGCGTCAGCGAAGGGTTGTAGTAGCCCGTAAAAACCAAA AAAACC	FA1 (Antisense of MT-ND1 mRNA)	20/40
GCATTGGTGGTTCAGTGGTAGAATTCTCGCCTGCCACGCGGGAGGC CCGGGT	FA2 (5′ fragment of $tRNA^{Gly}_{GCC}$)	7/40
TCCCTGAGACCCTTTAACCTGTGAGGACATCCAGGGTCACAGGTGA GGTTCTTGGGAGCC	pre-miR125a	3/40
TCCTCTTTAGTATAGTGGTGAGTATCCCCGCCTGTCACGCGGGAGA CCGG	5′ fragment of $tRNA^{Asp}_{GTC}$	1/40
CCTGTCACGCGGGAGACCGGGGTTCGATTCCCCGACGGGGAGCCA	3′ fragment of $tRNA^{Asp}_{GTC}$	1/40
TCACTGACTGTCTTGGGAGGAGGGGCTGGGTGTGGCACACAGTGA	chromosome 22 antisense of MYH9 gene	1/40
GGGAGGGTGCCTGGAGGAGTGGAGGGATTGGATTTACACCCTCTT A	chromosome 2	1/40
CAGCCGCAAGGGAGGCTGGGAAGTACAGTCATGCCTTCAGGTAGC CGTGAGCCCTG	chromosome 1	1/40
GAAAAAGTCATGGAGGCCATGGGGTTGGCTTGAAACCAGCTTTGG GGGGTTCGATTCCTTCCTTTTTTTGCCA	$tRNA^{Ser}_{UGA}$	1/40
TACGGACTCACTATAGGGAGCTATCGACGTAACGCTGGGGATTGT GGGTTCGTCCCATCTGGGTCGCCA	Similar to $tRNA^{Arg}_{CCG}$	1/40
TTTTGGGGTTTGGCAAAAACCAAAAAAACCAACCCAAACCCACCA AAAACCAA	No match in the databases	1/40
TTGGCTGAGGATGCGGAGGGGAGGAGGCTGAGTA	chromosome 11 intron of TAF6L gene	1/40
TGAGGATGGTGGTCAAGGGACCCCTATCACACCACCACCACCAA	Mitochondrial genome	1/40

2.2.2. Kd determination and mutation study using Bio-Layer Interferometry

During my doctor course, binding interaction between folic acid and RNAs obtained from SELEX were investigated. Three RNAs ligated with adaptors were prepared by *in vitro* transcription and three RNAs without adpators which were chemically synthesized were purchased from Gene design Inc. Dissociation constant (K_D) values were measured using Bio-layer interferometory (BLI). In order to immobilize folic acid onto the biosensors, ethylenediamine was immobilized onto AR2G (Amine Reactive Second-Generation) biosensors whose surface bears carboxyl group. Then, folic acid was conjugated with this sensor via amide bond. Six to eight different concentrations of RNA were used for steady-state analysis to determine the K_D values.

The results of binding analysis about FA1 and FA2 were shown in Table 3. FA1 bound to folic acid only when adaptors were ligated (K_D value was 3.8 μM). This indicates that ligation of adaptors is essential for FA1 to bind to folic acid, which is consistent with the difference of predicted structure between FA1 and FA1+adp (Figure 13). On the other hand, FA2 did not bind to folic acid with or without adaptors. This result indicates that it was false positive that FA2 sequence was obtained by SELEX to folic acid. During SELEX, cDNAs are amplified by PCR. However, each cDNA is not uniformly amplified at the same efficiency by PCR, which means that the sequences whose PCR efficiency are higher could be easily amplified and therefore such sequences are obtained from final round as false positive[106].

The results of binding analysis about hsa-pre-miR125a were shown in Table 4. In contrast to FA1 and FA2, hsa-pre-miR125a can bind to folic acid with or without adaptors, and both hsa-pre-miR-125a and hsa-pre-miR125a+adp showed the almost the same binding affinity (K_D values were 2.8 μM and 1.9 μM each). These results indicate that adaptors do not affect the binding affinity of hsa-pre-miR125a to folic acid.

To identify the essential motif of hsa-pre-miR125a to bind to folic acid, mutants were prepared by *in vitro* transcription and K_D values of those mutants were measured. Because 5′ end base of the wild-type hsa-pre-miR125a is not G, the efficiency of *in vitro* transcription by T7 RNA polymerase was too low. Therefore, U1G/G58C mutant hsa-pre-miR125a whose predicted secondary structure was the same as that of wild-type was prepared, and U1G/G58C mutant showed the almost the same K_D value as that of wild-type (2.6 μM) (Figure 14 and Table 4). Therefore, I introduced the same mutations (U1G/G58C) into other mutants except for m1 mutant (Figure 14 and Table 4). First, as a negative control hsa-pre-miR21 U1G mutant was prepared. The binding affinity of hsa-

pre-miR21 U1G was too low to determine the K_D value. Next, the some regions of pre-miR125a U1G/G58C mutant were replaced from pre-miR21 U1G mutant to prepare m1 and m3 mutant, and the region of pre-miR125a which is same as hsa-pre-miR21 was substituted with other bases to prepare m2 mutant (Figure 14a). Mutant m1 and m2 had comparable K_D value but the binding affinity of m3 was too low to be determined (Table 4). In addition, Mutations at three base pairs next to the terminal loop did not change the binding affinity (mutants m4, m5 and m6 in Figure 14a). These results indicate that terminal loop region of hsa-pre-miR125a is important for binding to folic acid. Therefore binding affinity of several mutants whose bases of terminal loop region were substituted with other bases were measured (Figure 14a). Mutants m10 and A33U showed the almost the same binding affinity (K_D = 4.2 μM and 3 μM each in Table 4) as that of wild-type, but the K_D values of other mutants (m7, m8, m9, G34A, G35A and G36A) increased about ten to twenty times more than that of wild-type (Figure 14a and Table 4).

For further analysis, several point mutants, deletion mutants and insertion mutants were prepared and divided into four groups (Figure 14b). Each mutant in group 1 have a point mutation in the terminal loop to disrupt a structure of terminal loop. Mutants G25A, G26C and C32G showed decreased binding affinity but C31G had the comparable binding affinity (K_D = 6.0 μM in Figure 14b and Table 4), which indicate that G24, G26 and C32 are important but the predicted structure of terminal loop is not essential for binding. Next, the other bases than important bases for binding observed in mutation study of group 1 were deleted or a base was inserted (mutants in group 2 in Figure 14b). All these mutants except for ΔA27/C28/A29/U30/C31 mutant showed the comparable K_D values, which indicates that minimal length of terminal loop necessary for binding is 8 nt. Both mutants ΔA33 and G34.5 (group 3 in Figure 14b) had weak binding affinity (K_D = 11 μM and 48 μM each in Table 4). This result and the fact that A33U showed the comparable affinity (Table 4) indicate that one nucleotide between C32 and G34 is important but the variety of bases is not. Deletion of a base at the next to the terminal loop (ΔA24 and ΔU37 in group 4) and two mispairs next to the terminal loop (U37A/C38G in group 4) abolished the binding affinity but one mispair at the same position (U37A in group 4) did not show the decrease of binding affinity.

From mutation study described above, the essential motif for binding to folic acid was shown in Figure 15. This motif was not observed in other RNAs obtained by SELEX (Table 2). By searching the structure of human miRNA precursors registered in the database of microRNA as of 8 December 2014 (miRbase[23], http://www.mirbase.org/), this motif was discovered only in miR125a precursors.

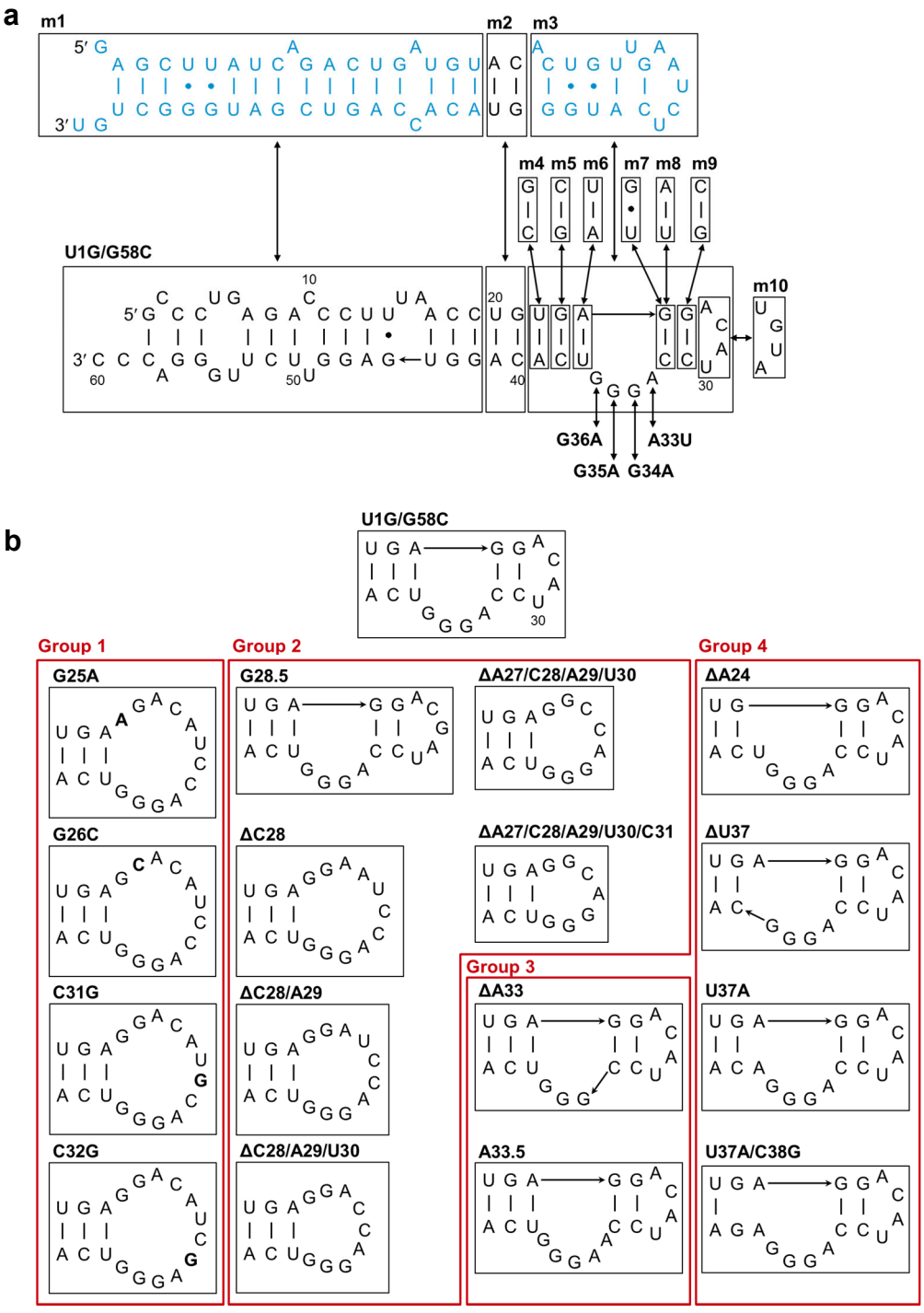

Figure 14 | Mutants of hsa-pre-miR125a. Lines denote the Watson-Crick base pairs and bullets denote wobble base pairs. Secondary structures were predicted using mfold software[105]. (**a**) Roughly mutated hsa-pre-miR125a mutants. Cyan bases denote corresponding regions to hsa-pre-miR21 used as negative control. (**b**) Loop region mutants were divided into four groups. Bold bases in group 1 are mutated bases.

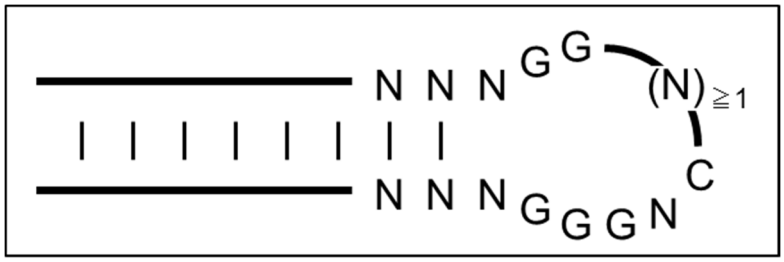

Figure 15 | Essential motif for binding to folic acid. N means A/U/G/C. Black bold lines indicate RNA and lines between bases denote base pairs.

Table 3 | Dissociation constant of RNAs obtained from SELEX. ND indicates "not determined" and R^2 value is derived from curve fitting in the steady-state analysis.

RNA	K_D [μM]	R^2
FA1+Adp	3.8	0.98
FA2+Adp	ND	
FA1	ND	
FA2	ND	

Table 4 | Dissociation constant of hsa-pre-miR125a mutants. ND indicates "not determined".

RNA	Mutation	K_D [μM]	R^2
hsa-pre-miR21	U1G	ND	
hsa-pre-miR125a	WT	2.8	0.98
	Adp+	1.9	0.97
	U1G/G58C	2.6	0.99
hsa-pre-miR125a U1G/G58C	m1	4.6	0.98
	m2	6.2	0.99
	m3	ND	
	m4	2.1	0.98
	m5	3.2	1.00
	m6	3.2	0.99
	m7	32	0.98
	m8	62	0.96
	m9	48	0.95
	m10	4.2	0.88

RNA	Mutation	K_D [μM]	R^2
	A33U	3.0	0.96
	G34A	38	0.94
	G35A	ND.	
	G36A	35	0.99
	G25A	73	0.99
	G26C	62	0.99
	C31G	6.0	0.99
	C32G	11	0.99
	G28.5	5.9	0.98
hsa-pre-miR125a	ΔC28	1.8	0.96
U1G/G58C	ΔC28/A29	1.4	0.94
	ΔC28/A29/U30	1.2	0.93
	ΔA27/C28/A29/U30	5.2	0.97
	ΔA27/C28/A29/U30/C31	12	0.98
	ΔA33	11	1.00
	A33.5	48	0.93
	ΔA24	ND	
	ΔU37	ND	
	U37A	2.1	0.95
	U37A/C38G	ND	

2.3. Conclusion

In this chapter, novel interaction between hsa-pre-miR125a and folic acid was identified using small RNA transcriptomic SELEX. Because the abundance of miRNA precursors including both pri- and pre-miRNAs are too low, the methods to discover the interaction of these RNAs and small molecules have been limited. This small RNA transcriptomic SELEX has great potential to identify the interaction of low abundant small RNAs. In addition, new RNA motif for binding to folic acid was discovered, which is difficult to be discovered by computational analysis.

However, there are still some problems in this method. The first problem is the influence of the ligation of adaptor sequences. Because these adaptors may dramatically change the structure of natural transcript, the potential RNA aptamer elements whose structure are changed by adaptors cannot be identified by this method. To overcome this problem, it is necessary to cleavage the adaptors after PCR or transcription and re-ligate these before RT-PCR (reverse transcription-PCR), or perform multiple selections using different adaptor sets. Second problem is diversity of transcripts extracted from cells. The expression levels of RNAs are different among the cells and conditions of culture. To include all RNAs transcribed in various cells, it is necessary to extract RNAs from various cells or organs including both cancer cells and normal cells.

Analysis of binding affinity of hsa-pre-miR125a revealed the essential motif in the terminal loop region for binding to folic acid. Because this loop region is common in both pri- and pre-miR125a, folic acid binds to both precursors. In addition, this motif is conserved in mammalian (Figure 16). It was reported the terminal loop region controls miRNA processing by Drosha and Dicer[107]. In addition, several regions of pri-miRNAs are important for processing by Dorhsa, apical stem and terminal loop elements of pri-miR125a have significant contributions on processing by Drosha[108]. These reports suggested the potential of folic acid binding to terminal loop of miR125a precursors to regulate the processing. Although it was reported that folic acid deficient culture medium changed the expression level of many human miRNAs, miR-125a was not included in these miRNAs[91]. In this report, expression level of miRNAs were quantified after 6-day folate deficient treatment. They also suggested that reduced dietary folate can reduce levels of *S*-adenosyl methionine required for cellular methylation reactions resulting the epigenetic alterations at the DNA or histone code levels[91]. Therefore, increase or decrease of folic acid may regulate the processing of miR125a in a short time. Nonetheless, further studies are necessary to elucidate whether the interaction between folic acid and miR125a precursors regulate the processing.

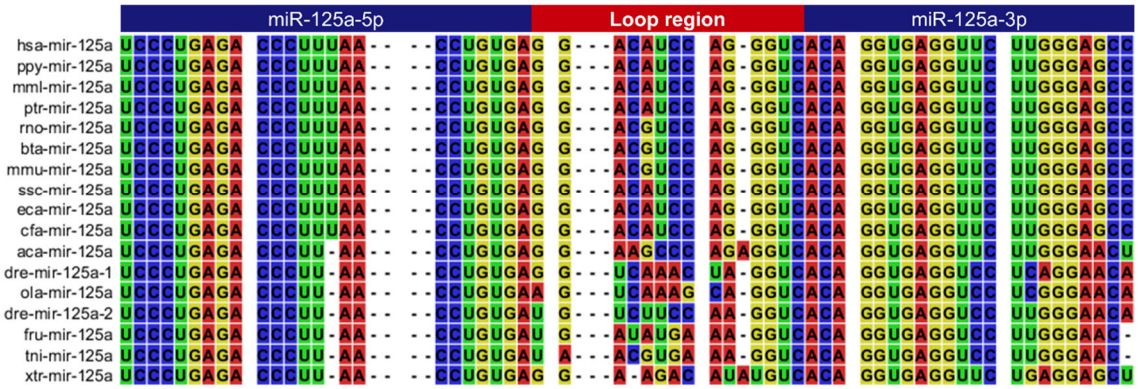

Figure 16 | The sequences of pre-miR125a of various species. These sequences were from miRbase[23] (http://www.mirbase.org/). The abbreviations indicate as follows, Hsa; *Homo sapiens,* ppy; *Pongo pygmaeus*, mml; *Macaca mulatta*, ptr; *Pan troglodytes*, rno; *Rattus norvegicus*, bta; *Bos taurus*, mmu; *Mus musculus*, ssc; *Sus scrofa*, eca; *Equus caballus*, cfa; *Canis familiaris*, aca; *Anolis carolinensis*, dre; *Danio rerio*, ola; *Oryzias latipes*, fru; *Fugu rubripes*, tni; *Tetraodon nigroviridis*, xtr; *Xenopus tropicalis.*

2.4. Materials and methods

Chemical synthesis of α-*N*-biotinyl-phenylalanine cyanomethyl ester (Biotin-Phe-CME)

Biotin-Phe-CME was synthesized as previously described[40].

Preparation of RNAs

The Fx3+N1 and RNAs for BLI analysis were prepared using run-off *in vitro* transcription with T7 RNA polymerase and purified by denaturing PAGE (8M Urea, 1×TBE, %T = 8 and %C = 5). The primers for preparing transcription templates and RNAs are shown in Table 5 and 6. All primers were purchased from Eurofins Genomics K.K. (Japan). FA1, FA2 and hsa-pre-miR125a RNAs were purchased from Gene Design Inc. (Japan). The concentrations of RNAs were determined by absorbance at 260 nm.

Preparation of dial-Fx3+N1

In order to avoid the amplification of Fx3+N1 during the construction of small RNA library, 3′-end of Fx3+N1 purified by denaturing PAGE was oxidized to dialdehyde group (named as dial-Fx3+N1) under the following conditions; 20 μM Fx3+N1 was incubated with 30 mM $NaIO_4$ on ice for 20 min and then 2% (w/v) $LiClO_4$ in acetone was added. The mixture was centrifuged at 15,000× *g* (room temperature) for 10 min to precipitate dial-Fx3+N1. The pellet was rinsed with acetone and centrifuged at 15,000× *g* (room temperature) for 3 min two times, and dissolved in water. The concentrations of RNAs were determined by absorbance at 260 nm.

Cell cultures for construction of small RNA library

BeWo cells (RIKEN Cell Bank, code RCB1644) were grown in Ham's F-12 medium (17458-65, Nacalai Tesque Inc.) supplemented with 100 unit/mL penicillin, 100 μg /mL streptomycin (15140-122, Thermo Fisher Scientific Inc.) and 10% fetal bovine serum (10437-010, Thermo Fisher Scientific Inc.) in an incubator with 5% CO_2 at 37°C. HuH-7, VMRC-RCW, NEC14, HuO-3N1 and RAJI Cells (RIKEN Cell Bank, code RCB1942, RCB1963, RCB0490, RCB2104, RCB1647) were grown in RPMI 1640 medium (30264-85, Nacalai Tesque Inc.) supplemented with 100 unit/mL penicillin, 100 μg /mL streptomycin (15140-122, Thermo Fisher Scientific Inc.) and 10% fetal bovine

serum (10437-010, Thermo Fisher Scientific Inc.) in an incubator with 5% CO_2 at 37°C. MCF7 cells (RIKEN Cell Bank, code RCB1940) were grown in MEM (21442-25, Nacalai Tesque Inc.) supplemented with 100 unit/mL penicillin, 100 μg /mL streptomycin (15140-122, Thermo Fisher Scientific Inc.), 0.1 mM non-essential amino acids (M71455, Sigma-Aldrich Co. LLC.), 1 mM sodium pyruvate (S8636, Sigma-Aldrich Co. LLC.) and 10% fetal bovine serum (10437-010, Thermo Fisher Scientific Inc.) in an incubator with 5% CO_2 at 37°C. HeLa S3 cells (JCRB Cell Bank, JCRB0713) were grown in DMEM (08458-45, Nacalai Tesque Inc.) supplemented with 100 unit/mL penicillin, 100 μg /mL streptomycin (15140-122, Thermo Fisher Scientific Inc.) and 10% fetal bovine serum (10437-010, Thermo Fisher Scientific Inc.) in an incubator with 5% CO_2 at 37°C.

Construction of the rRNA/tRNA-depleted human small RNA library

Once cells arrived to about 20% of confluence, the media were removed and ISOGEN II (311-07361, NIPPON GENE) was added to dissolve cells. Small RNAs (10-200 nt) were isolated according to the attached protocol. The concentrations were determined by absorbance at 260 nm.

Small RNAs extracted from each cell lines were mixed with an equal ratio and 10 μg small RNAs were treated with Ribominus^{TM} Eykaryote kit (A10837-08, Thermo Fisher Scientific Inc.) according to the attached protocol. Ribosomal RNA-depleted small RNAs were separated to 10-60 nt fraction and 60-200 nt fraction by denaturing PAGE (8M Urea, 1×TBE, %T = 8 and %C = 5). tRNAs in the 60-200 nt fraction of small RNAs were removed under the following conditions; A mixture of 48 μL of 73.3 ng/ μL 60-200 nt fraction of small RNAs and 33.3 μM dial-Fx3+N1 in 166.7 mM HEPES-KOH (pH 7.5) was heated at 95°C for 2 min and cooled to room temperature over 5 min. Sixteen μL of 3 M $MgCl_2$ was added, and the mixture was transferred to an ice bath, 16 μL of 25 mM Biotin-Phe-CME in DMSO was added, and the mixture was incubated on ice for 2 h. After 2 h incubation, 8 μL of 3 M NaCl and 220 μL of ethanol were added, and then precipitated. To capture biotin-Phe-tRNAs, the RNA was dissolved with 120 μL SAv binding buffer (5 mM HEPES-KOH (pH 7.5), 500 μM EDTA, 1 M NaCl and 0.1% tween-20) and then mixed with 1.2 mg Dynabeads M-280 Streptavidin (11205D, Thermo Fisher Scientific Inc.) which were pre-washed with SAv wash buffer (5 mM Tris-HCl (pH 7.5), 500 μM EDTA and 1M NaCl) three times before use, SAv solution A (0.1M NaOH and 50mM NaCl) two times and 0.1 M NaCl one time. After 15 min incubation using rotary shaker at room temperature, supernatant was recovered and precipitated with ethanol. For the second aminoacylation, the RNA was dissolved in 32 μL of 125 mM HEPES-KOH

(pH 7.5) with 0.75 M $MgCl_2$ and incubated at room temperature for 5 min. The mixture was transferred to an ice bath, 8 μL of 25 mM Biotin-Phe-CME was added. After 2 h incubation on ice, 48 μL of 500 mM NaCl and 220 μL of ethanol were added. RNAs were precipitated and dissolved in 80 μL SAv binding buffer. To capture biotin-Phe-tRNAs, the same procedure described above was performed using 800 μg beads. Supernatant was precipitated with ethanol and dissolved in 16 μL of 125 mM HEPES-KOH (pH 7.5) with 0.75 M $MgCl_2$ and incubated at room temperature for 5 min. The mixture was transferred to an ice bath, 4 μL of 25 mM Biotin-Phe-CME was added. After 2 h incubation on ice, 68 μL of 500 mM NaCl and 220 μL of ethanol were added. RNAs were precipitated and dissolved in 40 μL SAv binding buffer. To capture biotin-Phe-tRNAs, the same procedure described above was performed using 400 μg beads. Supernatant was precipitated with ethanol and dissolved in water. The concentrations were determined by absorbance at 260 nm.

A dial-Fx3+N1 in the tRNA-depleted small RNA fraction was removed by denaturing PAGE (8M Urea, 1×TBE, %T = 8 and %C = 5). All amount of rRNA/tRNA-depleted small RNA fraction and 10-60 nt fraction were mixed and used for preparation of cDNA library.

Construction of the cDNA library from rRNA/tRNA-depleted human small RNA library

The 3′-adaptor (purchased from Integrated DNA Technologies) and 5′-adaptor (purchased from Japan Bio Service) were purified by denaturing PAGE (8M Urea, 1×TBE, %T = 20 and %C = 5) before use. For ligation of 3′-adaptor, 6.67 μL mixture of 70 ng/μL rRNA/tRNA-depleted human small RNA library, 15 μM 3′-adaptor, 25 % (v/v) PEG8000, 40 U/ μL T4 RNA ligase 2 truncated (M0242S, New England BioLabs) and 1×reaction buffer attached with T4 RNA ligase 2 truncated was incubated at 16°C for 12 h. After the incubation, RNAs were extracted with phenol-chloroform mixture and precipitated with ethanol. Non-reacted 3′-adaptor was removed by denaturing PAGE (8M Urea, 1×TBE, %T = 8 and %C = 5). RNA library ligated with 3′-adaptor was treated with $NaIO_4$ according to protocol described above, in order to avoid the ligation of contaminated RNA library not ligated with 3′-adaptor in the next step. For ligation of 5′-adaptor, 10 μL mixture of all amount of RNA library ligated with 3′-adaptor, 20 μM 5′-adaptor, 15% (v/v) PEG8000, 30 U/ μL T4 RNA ligase 1 (M0204S, New England BioLabs) and 1×reaction buffer attached with T4 RNA ligase 1 was incubated at 37°C for 30 min and then at 16°C for 12 h. After the incubation, RNAs were extracted with

phenol-chloroform mixture and precipitated with ethanol. Non-reacted 5′-adaptor was removed by denaturing PAGE (8M Urea, 1×TBE, %T = 8 and %C = 5). RNA library ligated with 3′/5′-adaptors was converted to cDNA by reverse transcription and PCR using primers sele-1 and sele-2 and PrimeScript^{TM} One Step RT-PCR Kit Ver.2 (RR055A, TaKaRa), and cDNA was purified by Native-PAGE (1×TBE, %T = 8 and %C = 5).

Immobilization of folic acid onto the magnetic beads

Magnetic beads whose surface was modified with hydroxyl group (FG-beads, TAS8848 N1120, Tamagawa Seiki) were used for immobilization of folic acid. FG-beads were suspended in *N,N*-dimethylformamide (DMF, 10344-00, KANTO CHEMICAL) at 5 μg/μL and washed with DMF three times. FG-beads (5 μg/μL) were incubated at 25°C for 24 h with 10 mM folic acid (16221-91, Nacalai Tesque Inc.) in DMF in the presence of 10 mM EDC (1-ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride, 15022-86, Nacalai Tesque Inc.) and 10% (v/v) triethylamine (202-02641, WAKO). Thereafter, the folic acid-immobilized beads were washed with DMF three times and suspended in DMF at 10 μg/μL.

SELEX

cDNA library was transcribed to RNA library by run-off *in vitro* transcription with T7 RNA polymerase and purified by denaturing PAGE (8M Urea, 1×TBE, %T = 8 and %C = 5). A 10 μL of 510 ng/μL RNA library and 1×selection buffer (50 mM HEPES-KOH (pH 7.5), 300 mM KCl, 5 mM $MgCl_2$ and 0.05% tween-20) was mixed with of FG-beads whose surface is modified with hydroxyl group which pre-washed by 1×selection buffer three times before use, and this mixture was incubated using rotary shaker at 4°C for 30 min (referred to as “pre-clear” process). After the incubation, the supernatant was added to folic acid-immobilized beads which were washed by 1×selection buffer three times before use, and this mixture was incubated using rotary shaker at 4°C for 1 h. After the incubation, the supernatant was discarded and the beads were washed with 100 μL of 1×selection buffer (referred to as “wash” process). The beads were suspended in 1 mM folic acid in 1×selection buffer and incubated using rotary shaker at 4°C for 30 min to elute the RNAs binding to folic acid-immobilized beads. The supernatant was recovered and this elution cycle repeated three times. All of the supernatants were mixed and precipitated with ethanol. The pellet was dissolved in 10 μL of water.

To quantify the amount of recovered RNAs from beads, quantitative RT-PCR

was performed as follows; 0.5% of recovered RNAs were mixed with 10 μL of RT mix (500 μM dNTP, 5 μM primer sele-2, 8 U/μL M-MLV Reverse Transcriptase (M1701, Promega K.K.) and 1×reaction buffer attached with reverse transcriptase), and this mixture was incubated at 50°C for 1 h. After the incubation, 1 μL of reverse transcription reaction solution was mixed with 19 μL of qPCR mix (5 μM primer sele-1, 5 μM primer sele-2 10 mM Tris-HCl (pH 9.0), 50 mM KCl, 0.1% Triton-X100, 2.5 mM $MgCl_2$, 0.25 mM dNTP, 1/1,000,000 SYBR® Green I (50512, LONZA), Taq DNA polymerase) and qPCR was performed using LightCycler®1.5 (Roche Diagnostics K.K.) at 94°C (20°C/sec) for 1 min, followed by 35 cycles of 94°C for 10 sec (20°C/sec), 51°C for 10 sec (20°C/sec) and 72°C for 30 sec (0.5°C/sec) to measure C_p values.

The all amount of recovered RNAs not used for qPCR was mixed with RT mix and this mixture was incubated at 50°C for 1 h. After the incubation, all of the reverse transcription reaction solution was mixed with 200 μL of qPCR mix without SYBR® Green I and DNA was amplified by PCR by C_p+2 cycle of 94°C for 10 sec, 51°C for 10 sec and 72°C for 30 sec. DNA was extracted with phenol-chloroform mixture and precipitated with ethanol, and then used for next round of selection.

For the first round of selection, pre-clear process was not performed, wash process was performed one time and 200 μg beads were used. For the second round of selection, pre-clear process was performed one time, wash process was performed one time and 100 μg beads were used. For the third round of selection, pre-clear process was performed one time, wash process was repeated three times and 100 μg beads were used. For the fourth round of selection, pre-clear process was performed one time, wash process was repeated three times and 20 μg beads were used.

Sequencing

The cDNAs of tRNA-depleted human small RNA library and RNAs after SELEX were TA 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. Inserted cDNAs in the vectors were amplified by colony-PCR using primers seq-1 and seq-2, and sequenced by FASMAC Co., Ltd. (Japan).

To annotate the sequenced RNAs, BLAST search[109] using NCBI human genomic plus transcript database (http://blast.ncbi.nlm.nih.gov/Blast.cgi), fRNA database[110] (http://www.ncrna.org/frnadb/), genomic tRNA database[111] (http://gtrnadb.ucsc.edu/), snoRNABase[112] (https://www-snorna.biotoul.fr/) and miRbase[23]

(http://www.mirbase.org/) was performed.

Biolayer interferometry (BLI)

Binding of RNAs to folic acid was measured using the Octet RED96 biolayer interferometry (BLI) instrument (ForteBio, Inc.) and all measurement were conducted at 30°C and agitation rate is 1,000 rpm.

Folic acid was immobilized onto the amine reactive second-generation (AR2G) biosensor from ForteBio as follows; Each AR2G sensors were prewet for more than 10 min in water prior to use. Each AR2G sensors was equilibrated in water for 3 min, then activated with EDC/NHS solution (200 mM EDC and 50 mM *N*-hydroxysuccinimide (NHS, A00013, Watanabe Chemical Industries) in water) for 10 min. The activated sensors was modified with amine group by incubating with EDA solution (100 mM ethylenediamine (15020-22, Nacalai Tesque Inc.), 10 mM boric acid buffer (pH 8.5) and 1M NaCl) for 150 sec, then quenched with 1 M ethanolamine (E6133, Sigma-Aldrich Co. LLC., pH 8.5 adjusted by KOH) for 3 min. The amine group-immobilized sensors were equilibrated with 250 mM phosphate buffer (pH 7.0) for 3 min and then folic acid was immobilized by incubating with FA/EDC/NHS/Phos solution (20 mM folic acid, 200 mM EDC, 50 mM NHS and 250 mM phosphate buffer (pH 7.0)) for 20 min. The folic acid-immobilized sensors were incubated with water for 3 min. Reference sensors for the binding analysis were prepared following the same protocols described above using EDC/NHS/Phos solution (200 mM EDC, 50 mM NHS and 250 mM phosphate buffer (pH 7.0)) instead of FA/EDC/NHS/Phos solution.

Binding of RNAs to folic acid was measured as follows; folic acid-immobilized sensors or reference sensors were prewet for more than 10 min in Octet buffer (50 mM HEPES-KOH (pH 7.5), 300 mM KCl, 50 mM $MgCl_2$, 0.05% tween-20) prior to use. The sensors were equilibrated in Octet buffer for 3 min, then incubated in Octet buffer for 1 min to generate baseline. As an association step, the sensors were incubated in RNA dissolved in Octet buffer. After an association step, the sensors were incubated in Octet buffer as a dissociation step. The sensors were regenerated by incubating in R buffer (20 mM EDTA and 0.05% tween-20) for 10 sec 5 times. The ForteBio Data Analysis Octet software was used to perform steady-state analysis. For each RNAs, six to eight concentrations were used, from 200 nM to 240 μM (depending on the affinity of each RNAs).

Table 5 | Primers and adaptors used in chapter 2. N means A/T/G/C, italic types indicate ribonucleotides, ddC indicates dideoxyribocytidine.

Name	Sequence (5′ to 3′)	Remarks
primer Fx3-1	TAATACGACTCACTATAGGATCGAAAGATTTCCGCAGGCC	transcription of Fx3+N1
primer Fx3-2	NACCTAACGCCAATACCCTTTCGGGCCTGCGGAAATCTTT	transcription of Fx3+N1
primer Fx3-3	GGCGTAATACGACTCACTATAGG	transcription of Fx3+N1
primer Fx3-4	NACCTAACGCCAATACCCTT	transcription of Fx3+N1
primer seq-1	GCCTCTTCGCTATTACGCCAGC	sequencing
primer seq-2	TGTTGTGTGGAATTGTGAGCGG	sequencing
3′-adaptor	*A*pp-TGAAGAGCTGCTACTATC-ddC	Construction of the cDNA library
5′-adaptor	*GGCGUAAUACGACUCACUAUAGGGAGCUAUCGACGUAAC*	Construction of the cDNA library
primer sele-1	GGCGTAATACGACTCACTATAGGGAGCTATCGACGTAAC	Construction of the cDNA library and selection
primer sele-2	GATAGTAGCAGCTCTTCA	Construction of the cDNA library and selection
primer FA1+Adp-1	GGGAGCTATCGACGTAACTTTTATGGCGTCAGCGAAGGGTTG	Transcription of FA1+Adp
primer FA1+Adp-2	GGTTTTTTTGGTTTTTACGGGCTACTACAACCCTTCGCTGACGCC	Transcription of FA1+Adp
primer FA1+Adp-3	GATAGTAGCAGCTCTTCAGGTTTTTTTGGTTTTTACGGGCTAC	Transcription of FA1+Adp
primer FA2+Adp-1	GGGAGCTATCGACGTAACGCATTGGTGGTTCAGTGGTAGAATTC	Transcription of FA2+Adp
primer FA2+Adp-2	ACCCGGGCCTCCCGCGTGGCAGGCGAGAATTCTACCACTGAACCACC	Transcription of FA2+Adp
primer FA2+Adp-3	GATAGTAGCAGCTCTTCAACCCGGGCCTCCCGC	Transcription of FA2+Adp

Name	Sequence (5′ to 3′)	Remarks
Primer hsa-pre-miR125a+Adp-1	GGGAGCTATCGACGTAACTCCCTGAGACCCTTTAACCTG	Transcription of hsa-pre-miR125a+Adp
Primer hsa-pre-miR125a+Adp-2	GGCTCCCAAGAACCTCACCTGTGACCCTGGATGTCCTCAC	Transcription of hsa-pre-miR125a+Adp
Primer hsa-pre-miR125a+Adp-3	GATAGTAGCAGCTCTTCAGGCTCCCAAGAACCTCACC	Transcription of hsa-pre-miR125a+Adp
primer T7+Adp	GGCGTAATACGACTCACTATAGGGAGCTATCGACGTAAC	Transcription of FA1/FA2/hsa-pre-miR125a+Adp
primer T7	GGCGTAATACGACTCACTATAG	Transcription of pre-miRNAs
pre-miRNA template 1	ACAGCCCATCGACTGGTGTTGCCATGAGATTCAACAGTCAACATCAGTCTGATAAGCTCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 2	GGGCTCCAAGAACCTCACCTGTGACCCTGGATGTCCTCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 3	ACAGCCCATCGACTGGTGTTGTGACCCTGGATGTCCTCACAACATCAGTCTGATAAGCTCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 4	GGGCTCCAAGAACCTCACCACTGACCCTGGATGTCCTCAGTGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 5	GGGTCCCAAGAACCTCACCTGCCATGAGATTCAACAGTCAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 6	GGGCTCCAAGAACCTCACCTGGGACCCTGGATGTCCTCCCAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 7	GGGCTCCAAGAACCTCACCTGTCACCCTGGATGTCCTGACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 8	GGGCTCCAAGAACCTCACCTGTGTCCCTGGATGTCCACACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 9	GGGCTCCAAGAACCTCACCTGTGACCCTAGATGTCCTCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 10	GGGCTCCAAGAACCTCACCTGTGACCCTAGATGTCTTCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs

Name	Sequence (5′ to 3′)	Remarks
pre-miRNA template 11	GGGCTCCAAGAACCTCACCTGTGACCCTGCATGTGCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 12	GGGCTCCAAGAACCTCACCTGTGACCCTGGTACACCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 13	GGGCTCCAAGAACCTCACCTGTGACCCAGGATGTCCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 14	GGGCTCCAAGAACCTCACCTGTGACCTTGGATGTCCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 15	GGGCTCCAAGAACCTCACCTGTGACTCTGGATGTCCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 16	GGGCTCCAAGAACCTCACCTGTGATCCTGGATGTCCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 17	GGGCTCCAAGAACCTCACCTGTGACCCTGGATGTCTTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 18	GGGCTCCAAGAACCTCACCTGTGACCCTGGATGTGCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 19	GGGCTCCAAGAACCTCACCTGTGACCCTGCATGTCCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 20	GGGCTCCAAGAACCTCACCTGTGACCCTCGATGTCCTCACAGGTTA AAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 21	GGGCTCCAAGAACCTCACCTGTGACCCTGGATCGTCCTCACAGGTT AAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 22	GGGCTCCAAGAACCTCACCTGTGACCCTGGATTCCTCACAGGTTAA AGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 23	GGGCTCCAAGAACCTCACCTGTGACCCTGGATCCTCACAGGTTAAA GGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 24	GGGCTCCAAGAACCTCACCTGTGACCCTGGTCCTCACAGGTTAAAG GGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 25	GGGCTCCAAGAACCTCACCTGTGACCCTGGCCTCACAGGTTAAAGG GTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 26	GGGCTCCAAGAACCTCACCTGTGACCCTGCCTCACAGGTTAAAGGG TCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 27	GGGCTCCAAGAACCTCACCTGTGACCCGGATGTCCTCACAGGTTAA AGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs

Name	Sequence (5′ to 3′)	Remarks
pre-miRNA template 28	GGGCTCCAAGAACCTCACCTGTGACCCTTGGATGTCCTCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 29	GGGCTCCAAGAACCTCACCTGTGACCCTGGATGTCCCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 30	GGGCTCCAAGAACCTCACCTGTGCCCTGGATGTCCTCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 31	GGGCTCCAAGAACCTCACCTGTGTCCCTGGATGTCCTCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs
pre-miRNA template 32	GGGCTCCAAGAACCTCACCTGTCTCCCTGGATGTCCTCACAGGTTAAAGGGTCTCAGGGCTATAGTGAGTCGTATTACGCC	Transcription of pre-miRNAs

Table 6 | RNAs used in chapter 2. N means A/U/G/C, italic types indicate ribonucleotides.

Name	Sequence (5′ to 3′)
Fx3+N1	*GGAUCGAAAGAUUUCCGCAGGCCCGAAAGGGUAUUGGCGUUAGGUN*
FA1+adp	*GGGAGCUAUCGACGUAACUUUUAUGGCGUCAGCGAAGGGUUGUAGUAGCCCGUAAAAACCAAAAAAACCUGAAGAGCUGCUACUAUC*
FA2+adp	*GGGAGCUAUCGACGUAACGCAUUGGUGGUUCAGUGGUAGAAUUCUCGCCUGCCACGCGGGAGGCCCGGGUUGAAGAGCUGCUACUAUC*
FA1	*UUUUAUGGCGUCAGCGAAGGGUUGUAGUAGCCCGUAAAAACCAAAAAAACC*
FA2	*GCAUUGGUGGUUCAGUGGUAGAAUUCUCGCCUGCCACGCGGGAGGCCCGGGU*
hsa-pre-miR21GU1G	*GAGCUUAUCAGACUGAUGUUGACUGUUGAAUCUCAUGGCAACACCAGUCGAUGGGCUGU*
hsa-pre-miR125a	*UCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGGGUCACAGGUGAGGUUCUUGGGAGCC*
hsa-pre-miR125a+adp	*GGGAGCUAUCGACGUAACUCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGGGUCACAGGUGAGGUUCUUGGGAGCCUGAAGAGCUGCUACUAUC*
hsa-pre-miR125aU1G/G58C	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
m1	*GAGCUUAUCAGACUGAUGUUGUGAGGACAUCCAGGGUCACAACACCAGUCGAUGGGCUGU*
m2	*GCCCUGAGACCCUUUAACCACUGAGGACAUCCAGGGUCAGUGGUGAGGUUCUUGGAGCCC*
m3	*GCCCUGAGACCCUUUAACCUGACUGUUGAAUCUCAUGGCAGGUGAGGUUCUUGGGACCC*
m4	*GCCCUGAGACCCUUUAACCUGGGAGGACAUCCAGGGUCCCAGGUGAGGUUCUUGGAGCCC*

Name	Sequence (5′ to 3′)
m5	*GCCCUGAGACCCUUUAACCUGUCAGGACAUCCAGGGUGACAGGUGAGGUUCUUGGAGCCC*
m6	*GCCCUGAGACCCUUUAACCUGUGUGGACAUCCAGGGACACAGGUGAGGUUCUUGGAGCCC*
m7	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCUAGGGUCACAGGUGAGGUUCUUGGAGCCC*
m8	*GCCCUGAGACCCUUUAACCUGUGAAGACAUCUAGGGUCACAGGUGAGGUUCUUGGAGCCC*
m9	*GCCCUGAGACCCUUUAACCUGUGAGCACAUGCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
m10	*GCCCUGAGACCCUUUAACCUGUGAGGUGUACCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
A33U	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCUGGGUCACAGGUGAGGUUCUUGGAGCCC*
G34A	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAAGGUCACAGGUGAGGUUCUUGGAGCCC*
G35A	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGAGUCACAGGUGAGGUUCUUGGAGCCC*
G36A	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGGAUCACAGGUGAGGUUCUUGGAGCCC*
G25A	*GCCCUGAGACCCUUUAACCUGUGAAGACAUCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
G26C	*GCCCUGAGACCCUUUAACCUGUGAGCACAUCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
C31G	*GCCCUGAGACCCUUUAACCUGUGAGGACAUGCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
C32G	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCGAGGGUCACAGGUGAGGUUCUUGGAGCCC*
G28.5	*GCCCUGAGACCCUUUAACCUGUGAGGACGAUCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔC28	*GCCCUGAGACCCUUUAACCUGUGAGGAAUCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔC28/A29	*GCCCUGAGACCCUUUAACCUGUGAGGAUCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔC28/A29/U30	*GCCCUGAGACCCUUUAACCUGUGAGGACCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔA27/C28/A29/U30	*GCCCUGAGACCCUUUAACCUGUGAGGCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔA27/C28/A29/U30/C31	*GCCCUGAGACCCUUUAACCUGUGAGGCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔA33	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCGGGUCACAGGUGAGGUUCUUGGAGCCC*
A33.5	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔA24	*GCCCUGAGACCCUUUAACCUGUGGGACAUCCAGGGUCACAGGUGAGGUUCUUGGAGCCC*
ΔU37	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGGGCACAGGUGAGGUUCUUGGAGCCC*
U37A	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGGGACACAGGUGAGGUUCUUGGAGCCC*
U37A/C38G	*GCCCUGAGACCCUUUAACCUGUGAGGACAUCCAGGGAGACAGGUGAGGUUCUUGGAGCCC*

Chapter 3

Orthogonal ribosome-tRNAs pair by engineering of peptidyl transferase center

3.1. Introduction

Assignment of 20 canonical amino acids to codons, is achieved by specific acylation of tRNA with cognate amino acid catalyzed by aminoacyl-tRNA synthetase (ARS). Since each tRNA has an anticodon that pairs with the codon, the codons on mRNA can be decoded by the cognate aminoacyl-tRNAs (aa-tRNAs) according to the genetic code; and thus ribosome is able to catalyze the formation of peptide bond along the mRNA template, to yield a polypeptide with the encoded sequence. While 20 canonical amino acids are used in the natural translation system, it has been demonstrated that more than hundreds of different noncanonical amino acids (ncAAs) can be incorporated into nascent polypeptide chain by engineering of the genetic code[113]. Assignment of ncAAs in the genetic code has been achieved by two methodologies: genetic code expansion[114,115] and genetic code reprogramming[116]. The former method generally assigns a ncAA (or multiple ncAAs) to a nonsense codon(s) such as stop codons and an artificially programmed quadruplet codon(s). Genetic code expansion has been applied for both *in vivo* and *in vitro* expression of proteins with ncAAs.

When the genetic code expansion is performed *in vivo*, availability of ncAA-tRNA generally relies on an exogenously introduced ARS specifically paired with an orthogonal tRNA. The orthogonal tRNA should not be aminoacylated by endogenous ARSs. At the same time, the exogenous ARS should be engineered to specifically charge the ncAA onto the orthogonal tRNA, but not onto endogenous tRNAs. Several orthogonal ARS–tRNA pairs have been successfully developed[113], and thus far demonstrated expression of proteins containing ncAAs. However, the choices of usable ncAAs are yet limited to a few subgroups, because the majority examples of the engineered ARSs are based on *Methanococcus jannaschii* TyrRS[115], *Methanosarcina barkeri* PylRS[117,118], or *Methanosarcina mazei* PylRS[119,120]; therefore, the usable ncAAs for these mutant enzymes are limited to Phe or Lys analogs.

In addition to orthogonal ARS–tRNA pair, orthogonal ribosome–mRNA pair has been developed[121]. In this system engineered ribosome containing orthogonal 16S rRNA which does not recognizes endogenous mRNA but recognizes engineered mRNA bearing mutated Shine-Dalgarno (SD) sequence, was used[121]. Decoding center of this orthogonal 16S rRNAs have been further evolved for efficient amber codon decoding[122] and quadruplet codon[123]. These orthogonal ribosome–mRNA pairs enabled to express specific protein containing ncAAs *in vivo*[124]. However, the numbers of usable ncAAs and codons

were still limited.

On the other hand, in the genetic code reprograming, multiple sense codons can be simultaneously reassigned with various ncAAs. Genetic code reprogramming has been dominantly utilized in *in vitro* reconstituted translation system[125] to produce short polypeptides containing diverse exotic ncAAs such as D-amino acids[56] and *N*-methyl-amino acids[48-50]. Although in the genetic code reprogramming method, multiple ncAAs can be incorporated, some canonical amino acids cannot be used simultaneously. It is because the canonical amino acids originally assigned in natural genetic code need to be removed to suppress the incorporation of canonical amino acids instead of ncAAs[126]. To overcome this problem in genetic code reprogramming, I conceived the new concept of orthogonal translation system, which is orthogonal ribosome–tRNA pair. This engineered translation system means that wild-type ribosome only use canonical aa-tRNAs and the engineered ribosome only use ncAA-tRNAs as substrates. In this system, it is not necessary to deplete canonical amino acids and to change the original genetic code in order to incorporate ncAAs.

To develop the orthogonal ribosome–tRNA pair, the interaction between ribosome and tRNA needs to be engineered. During the elongation cycle of bacterial translation, where amino acids were sequentially added to the growing peptide chain, CCA-end of tRNA interacts with 23S rRNA of 30S subunit. The elongation cycle is divided to three steps[127,128] (Figure 18). First step is accommodation of aa-tRNA into the A-site of ribosome. aa-tRNA is delivered to the A-site by EF-Tu•GTP in the A/T state where the anticodon forms base pairs with codon in the A-site of 30S and the acceptor end is bound to EF-Tu (the first letter represents the binding site on the 30S and the second letter the site on the 50S). Then GTP is hydrolyzed to GDP to release EF-Tu•GDP and the accommodation of the acceptor end of the aa-tRNA into the A-site of the 50S (A/A state, the classical state). Second is peptide bond formation between aa-tRNA in A/A state and peptidyl-tRNA in P/P state, which is catalyzed in the peptidyl transferase center (PTC). After the peptide bond formation, peptidyl-tRNA in A/A state is translocated to P/P state and deacylated tRNA in P/P state is translocated to E/E state facilitated by EF-G with associated GTP hydrolysis, which is the third step.

In the PTC of classical state of *E. coli* ribosome complex, critical Watson-Crick base pairs occur between the universally conserved 3′ end of tRNAs (C74 and C75) and 23S rRNA G2251 and G2252 at the P site as well as G2553 at the A site, which were revealed by biochemical experiments and crystal structural analyses[62,129-136] (Figure 17 and 19a). Using an analogue of aa-tRNA fragment (C75 mutant puromycin derivatives of

the form NPm) as an A-site substrate (Figure 19b and c), peptidyl transfer activity of wild-type and G2553N mutant ribosome 50S subunits was measured by Kim and Green[64]. They reported that the wild-type ribosome preferred the wild-type substrate (CPm) by about 2- to 5-fold relative to the other substrates (APm, GPm and UPm) (Figure 19b). Furthermore, G2553C ribosome preferred the compensatory mutated substrate (GPm) by at least 20-fold relative to the other three substrates (Figure 19c). These results indicated that G2253C ribosome-GPm pair has weak orthogonality to wild-type ribosome-CPm pair in peptidyl transfer reaction.

The first step of translocation is movement of acceptor end of peptidyl-tRNA in A/A state into A/P state and acceptor end of deacylated tRNA in P/P state into P/E state, which is called as hybrid state[127] (Figure 18 and 20). This first step is promoted by EF-G•GTP but it can occur in the absence of GTP[137]. Then EF-G promotes translocation of peptidyl-tRNA from the A/P state to the P/P state and deacylated tRNA from the P/E state to the E/E state to be released from the ribosome. To my knowledge, there is no direct evidence of the interaction between 23S rRNA and CCA-end of tRNAs during the hybrid state. However, Dorner *et. al.* indicated the potential pairs between C74/C75 of peptidyl-tRNA in the A/P state and G2252/G2251 of 23S rRNA in the P-site[63] (Figure 20). In this report, they measured EF-G translocation rate with the pairs of wild-type and mutant ribosomes and tRNAs (Table 7 and 8). When C74G tRNA is in A-site of wild-type ribosome, translocation rate reduced (Table 7). When potential pairing between tRNA in the A/P state and 23S rRNA in the P-site was restored (the pair of C74G tRNA and G2252C ribosome), translocation rate was restored (Table 7). Similar result was observed about the interaction between C75 of tRNA in the A/P state and G2251 of 23S rRNA (Table 7). These results indicates that the pair of G2251C or G2252C ribosome and C75G or C74G tRNA have weak orthogonality to the pair of wild-type ribosome and tRNA in translocation.

On the other hand, disruption of the interaction in the A/A state (the pairs of wild-type tRNA and G2253C/U/A ribosomes) increased the rate of translocation (Table 8) and restoring the base pair in the A/A state (the pair of C75G tRNA and G2253C ribosome) diminished the translocation rate. These results indicate that ribosome and tRNAs favored the stable state forming base pairs.

These studies described above showed that engineering the base pairs between G2251/G2252/G2553 of ribosome and C74/C75 of tRNAs in PTC had potential to develop the orthogonal ribosome–tRNA pairs. However, it was unknown whether a mutation (or mutations) of 23S rRNA and CCA-3′ end of tRNAs in the PTC affect the

whole translation reaction to produce polypeptides using aa-tRNAs as substrates in an mRNA-dependent manner, including other steps.

In this chapter 3, I measured the translation activity of the PTC-mutated ribosome–tRNA pairs using an *E. coli* custom-made *in vitro* translation system that was integrated with the flexizyme technology, referred to as the FIT (Flexible *In vitro* Translation) system[126]. In natural, tRNAs are aminoacylated by protein aminoacyl-tRNA synthetases (ARSs) and ARSs specifically recognize the sequence of cognate tRNAs. Because most ARSs also interact with the universally conserved CCA-3′ end of tRNAs[4,5], ARSs are not suitable for preparation of aminoacyl-tRNA (aa-tRNA) bearing mutation in the CCA-3′ end[6-8]. In order to overcome this problem, flexizymes (eFx and dFx) bearing compensatory mutated 3′ ends were used because flexizymes recognize CCA-3′ end of substrate RNA via base pairs as described in chapter 1. Finally, the pair of mutant ribosome and tRNAs which acted orthogonally and used only the cognate genetic code consisting of ncAAs[65].

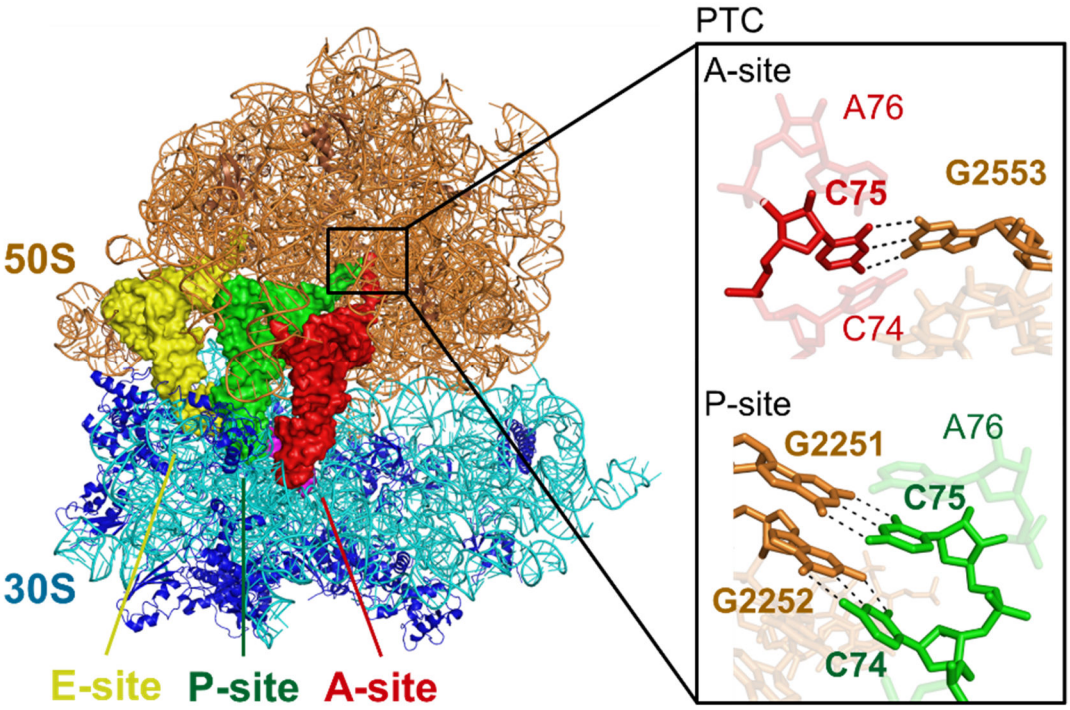

Figure 17 | Crystal structure of classical state of ribosome-tRNAs-mRNA complex. The interaction between 23S rRNA and CCA-end of tRNAs at the peptidyl transferase center (PTC) is shown in black rectangle. Black dashed lines indicate hydrogen bonds. This structure was reported in previous paper[62] (PDB ID: 2WDK/2WDL).

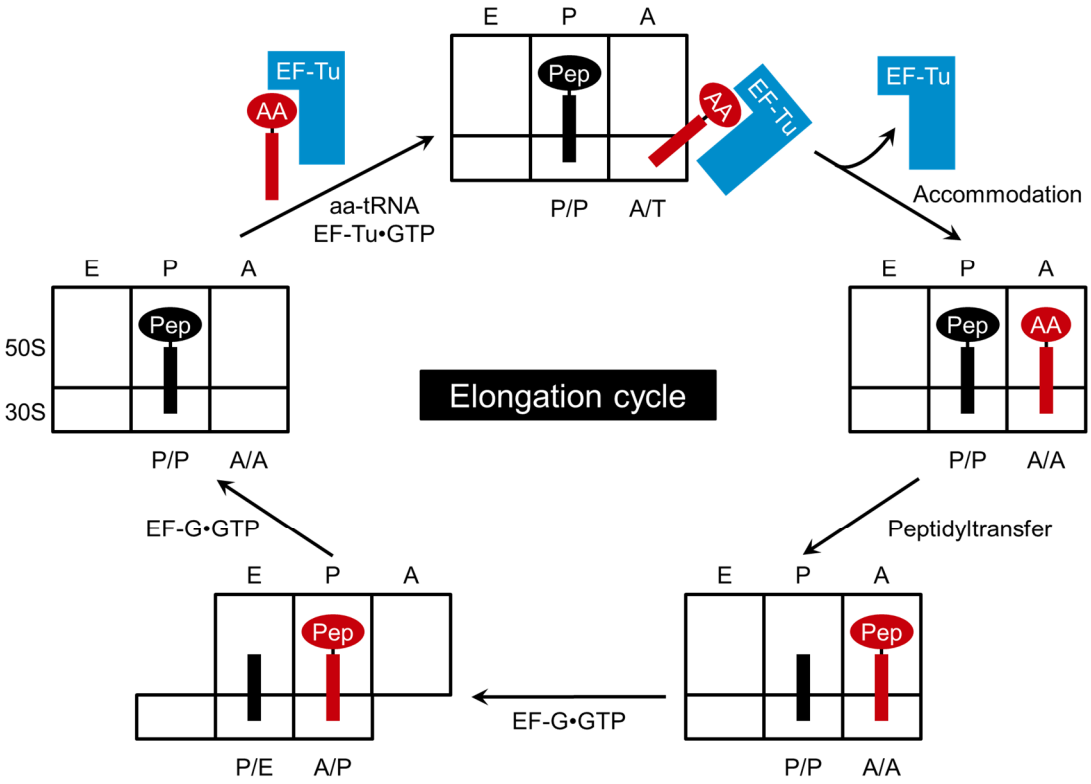

Figure 18 | Schematic illustration of elongation cycle of bacterial translation. The ribosomes are described as black rectangles and divided into 30S and 50S subunits and partitioned into the A-, P- and E-sites. The biding states of the tRNAs as previously defined[127] are indicated at the bottom of each sites. "Pep" means peptide and "AA" means amino acid. Bold lines indicate tRNAs.

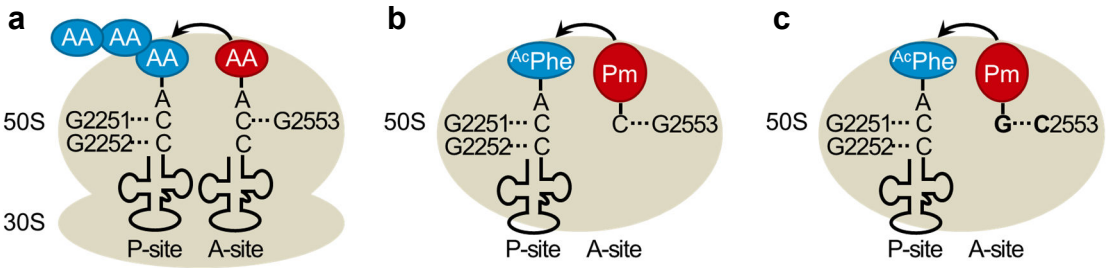

Figure 19 | Peptidyl transfer reaction using G2553N mutant ribosome and C75 mutant puromycin derivatives[64]. The ribosomes are described as wheat ellipse and divided into 50S and 30S subunits. (**a**) Schematic illustration of natural peptidyl transfer between peptidyl-tRNA at P-site and aa-tRNA at A-site. (**b**) Schematic illustration of peptidyl transfer between *N*-acetyl-phenylalanyl-tRNA (^{Ac}Phe-tRNA) at P-site and C75 puromycin derivatives shown as CPm at A-site using reconstituted wild-type 50S subunit. (**c**) Schematic illustration of peptidyl transfer between *N*-acetyl-phenylalanyl-tRNA (^{Ac}Phe-tRNA) at P-site and C75G puromycin derivatives shown as GPm at A-site using reconstituted G2553C mutant 50S subunit.

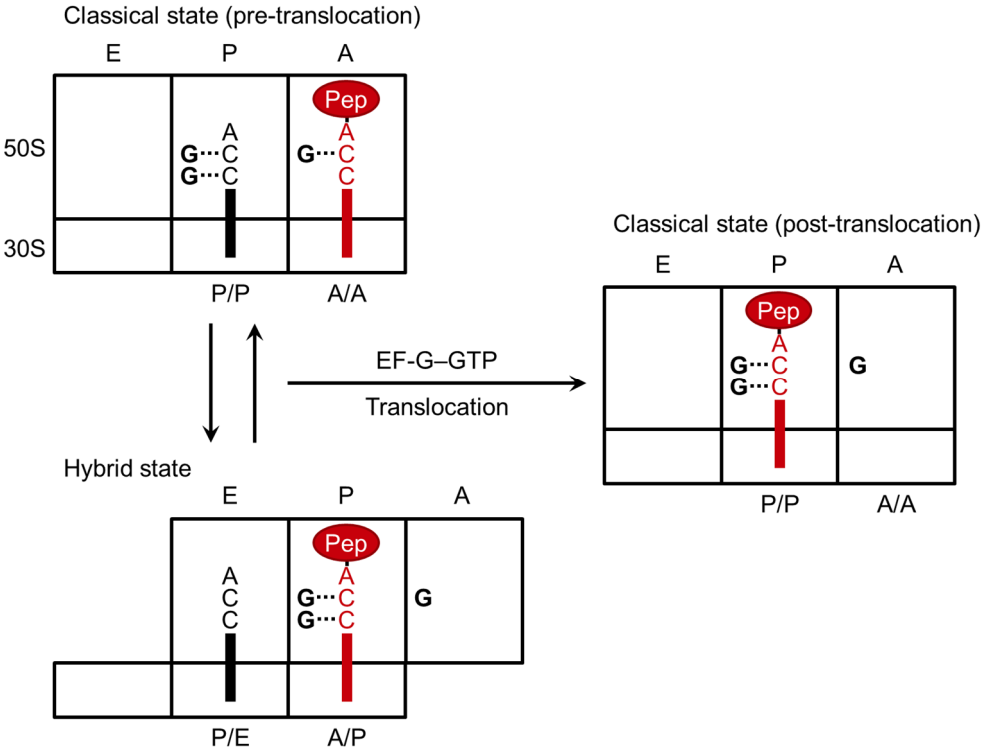

Figure 20 | Schematic illustration of classical and hybrid state of tRNA-ribosome complex and translocation induced by EF-G–GTP. The ribosomes are described as black rectangles and divided into 30S and 50S subunits and partitioned into the A-, P- and E-sites. The biding states of the tRNAs as previously defined[127] are indicated at the bottom of each sites. "Pep" in red circle means peptidyl-tRNA. Bold lines indicate tRNAs and bold Gs indicate G2251, G2252 and G2553 of 23S rRNA interacting with CCA-end of tRNAs. This figure was adapted from previous report[63,138].

Table 7 | EF-G translocation of WT, G2252C and G2251C mutant ribosomes. This table was adapted from previous report[63].

		Translocation rate (s^{-1})				Translocation rate (s^{-1})	
P-site tRNA	A-site tRNA	WT	G2252C	P-site tRNA	A-site tRNA	WT	G2251C
WT	WT	3.8	1.5	WT	WT	3.8	2.1
WT	C74G	1.5	5.0	WT	C75G	1.4	9.3

Table 8 | EF-G translocation with WT and G2553 mutant ribosomes. ND indicates "not determined". This table was adapted from previous report[63].

		Translocation rate (s^{-1})			
P-site tRNA	A-site tRNA	WT	G2553C	G2553U	G2553A
WT	WT	3.8	10.7	12.2	12.4
WT	C75G	1.4	1.1	ND	ND

3.2. Results and discussions

The study in this chapter 3 was published as a paper in the scientific journal and the figures in this chapter are reproduced or adapted from the published paper[65].

3.2.1. Aminoacylation of CCA-mutated tRNAs

To confirm that CCA-3′ end mutated tRNAs (denoted as tRNAs-NNA) can be aminoacylated by compensatory mutated flexizymes, I arbitrarily chose lysine and tyrosine as amino acids substrates (Figure 21), and $tRNA^{AsnE2}_{CUU}$ and $tRNA^{AsnE2}_{GUA}$ (Figure 22) bearing corresponding anticodons. Three mutant tRNAs derived from $tRNA^{AsnE2}_{CUU}$ and $tRNA^{AsnE2}_{GUA}$ bearing single or double C to G mutation (or mutations) (C74G, C75G or C74G/C75G) and the compensatory single or double G to C mutated flexizymes derived from dFx and eFx (Figure 5) were prepared by *in vitro* transcription. The aminoacylation efficiency were measured by acid-PAGE, which indicates the cognate mutant dFxs and eFxs can aminoacylate the tRNAs-NNA in the presence of Lys-DBE and Tyr-CME (Figure 23). On the other hand, wild-type flexizymes could not aminoacylate non-cognate tRNA mutants (Figure 23). In the case of C75U tRNA mutant, aminoacylation efficiency of mutant flexizymes bearing GAU-3′ end, which form two Watson-Crick base pairs with tRNAs, was lower than that of wild-type bearing GGU-3′ end, which form one Watson-Crick base pair and one G•U wobble base pair (Figure 24). This result indicates that G•U wobble base pair is more suitable for aminoacylation by flexizymes than A-U Watson-Crick base pairs. Mutant flexizymes could also charge several other amino acids (Figure 21) onto the corresponding mutant tRNAs for further experiments (Figure 25).

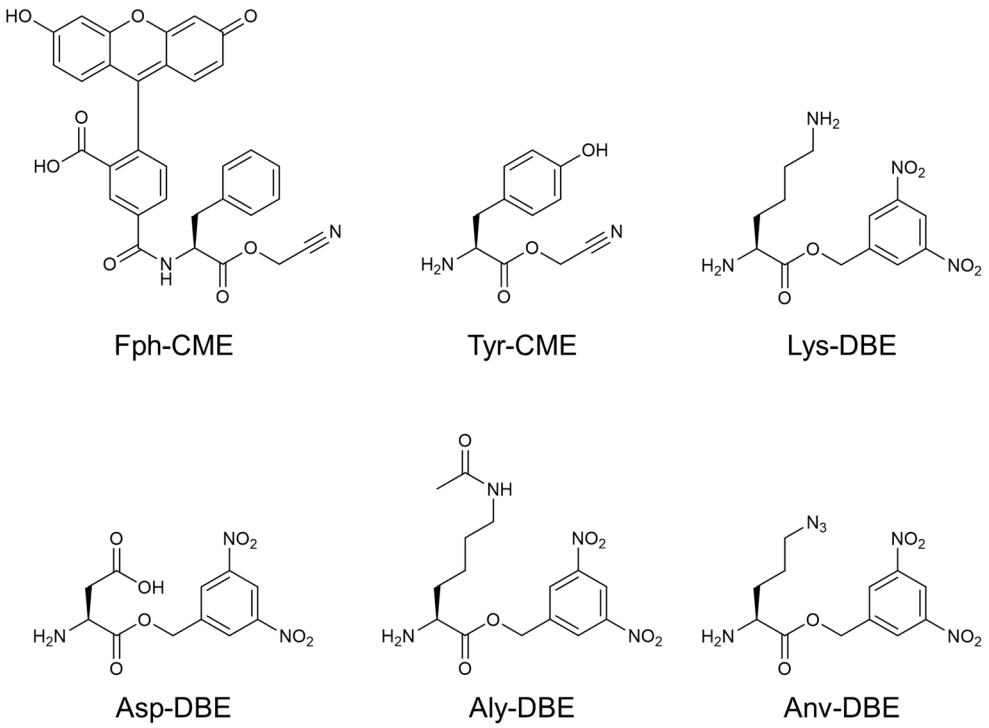

Figure 21 | Amino acid substrates used in the chapter 3. The abbreviations of Fph, Aly, and Anv represent *N*-(5-FAM)-L-phenylalanine, L-acetyllysine, and L-azidonorvaline, respectively. Carboxyl groups were activated using cyanomethyl ester (CME) or 3,5-dinitrobenzyl ester (DBE).

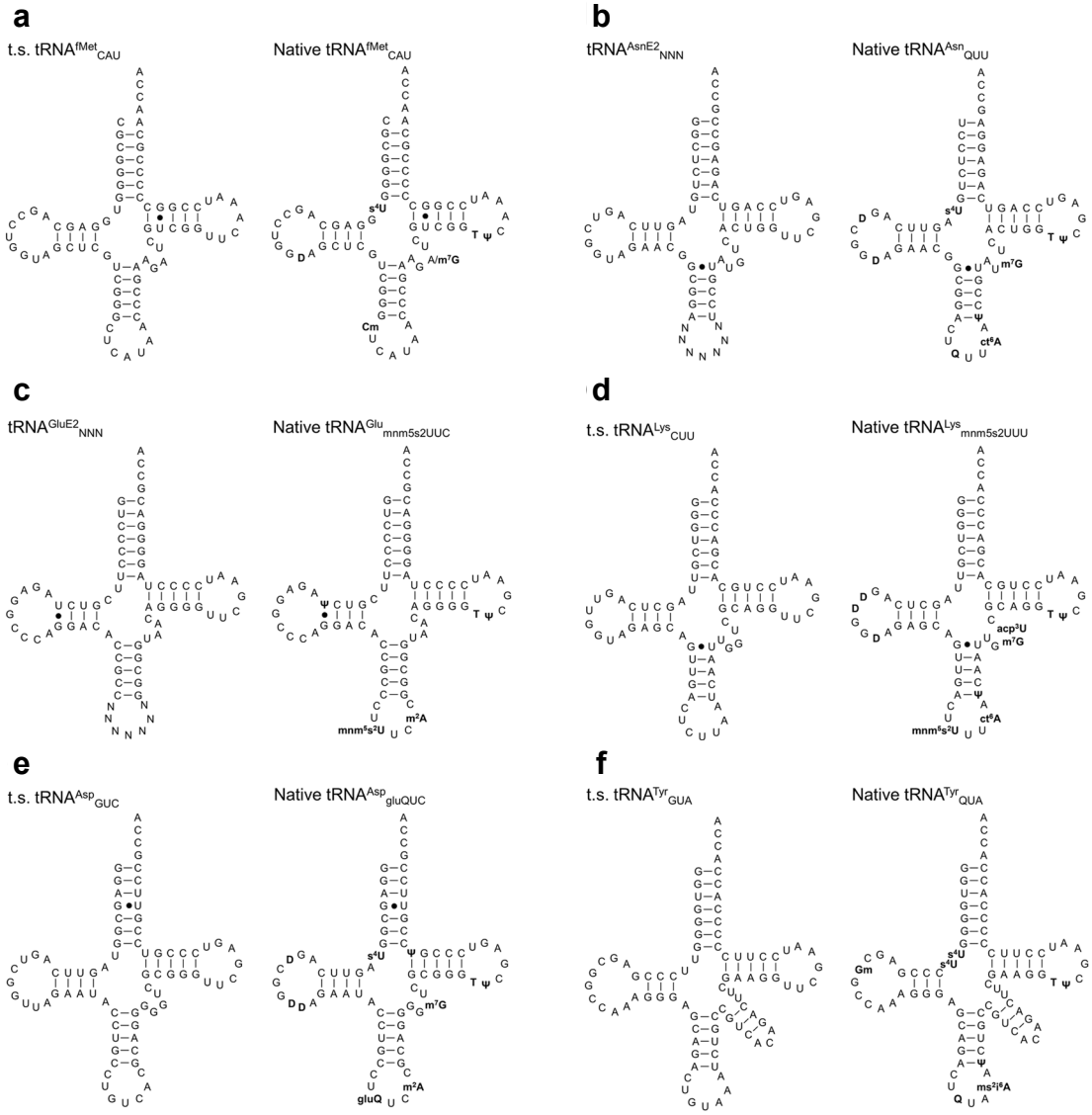

Figure 22 | Secondary structures of transcribed (t.s.) tRNAs compared with native tRNAs of *E. coli*. The figure presents the secondary structures of $tRNA^{fMet}_{CAU}$ (**a**), $tRNA^{AsnE2}_{NNN}$ (**b**), $tRNA^{GluE2}_{NNN}$ (**c**), $tRNA^{Lys}_{CUU}$ (**d**), $tRNA^{Asp}_{GUC}$ (**e**), and $tRNA^{Tyr}_{GUA}$ (**f**). s^4U; 4-thiouridine, D; dihydrouridine, Cm; 2′-O-methylcytidine, m^7G; 7-methylguanosine, ψ; pseudouridine, Q; queuosine, ct^6A: cyclic N^6-threonylcarbamoyladenosine, mnm^5s^2U; 5-methylaminomethyl-2-thiouridine, m^2A; 2-methyladenosine, acp^3U; 3-(3-amino-3-carboxypropyl)uridine, T; ribothymidine, gluQ; glutamyl-queuosine, ms^2i^6A; 2-methylthio-N^6-isopentenyladenosine. The sequences are based on the Modomics database[139] (http://modomics.genesilico.pl/).

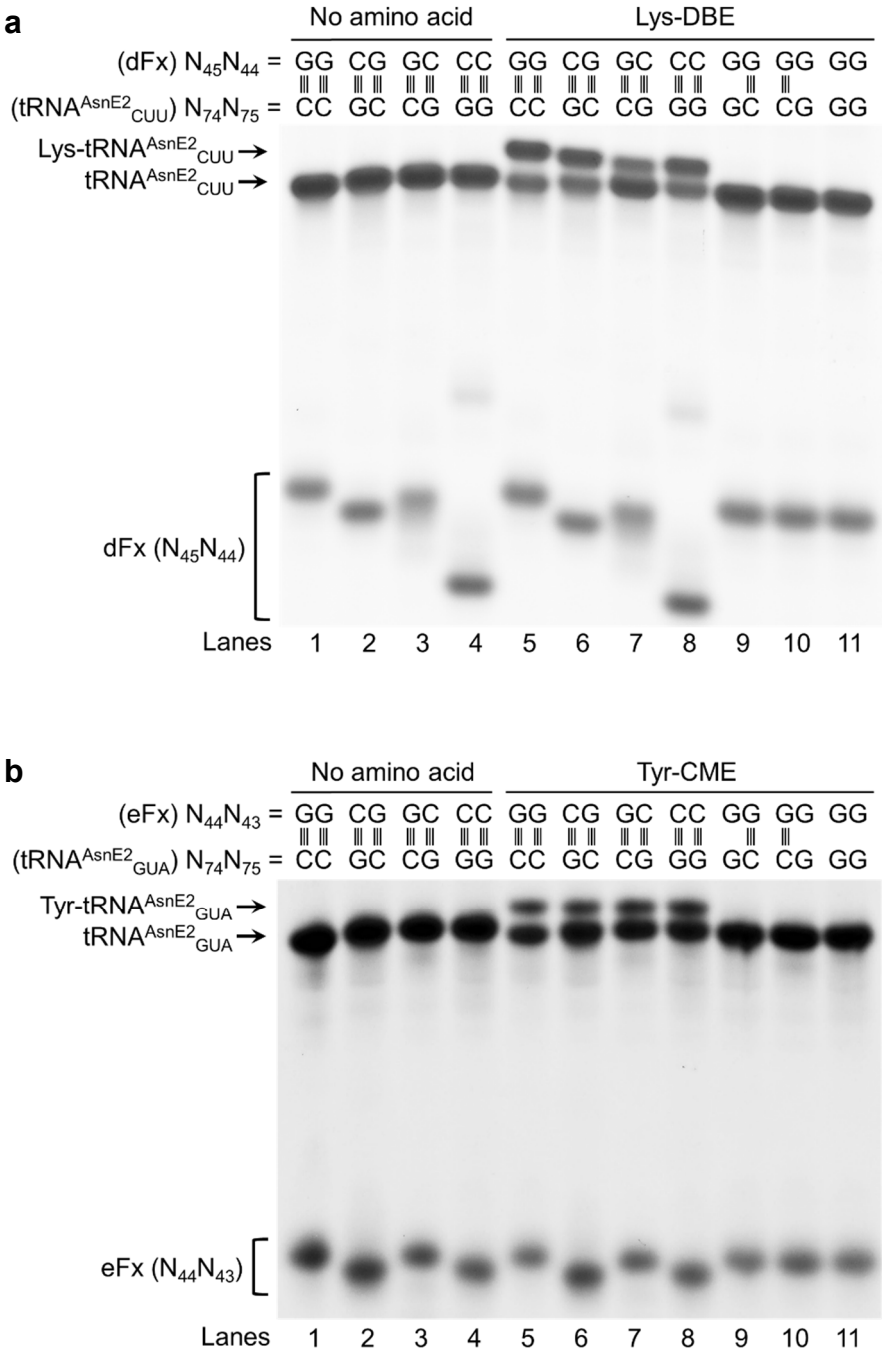

Figure 23 | A model study of aminoacylation of tRNAs-NNA derived from tRNA^{AsnE2}. tRNAs were acylated by dFx in the presence of Lys-DBE (**a**), and eFx in the presence of Tyr-CME (**b**). Acid-PAGE separated the bands of aa-tRNA (upper) and uncharged tRNA (lower). The data were generated from a sample of the end product of aminoacylation reaction. Note that the mobility of flexizymes was different depending on the 3′-terminal mutations.

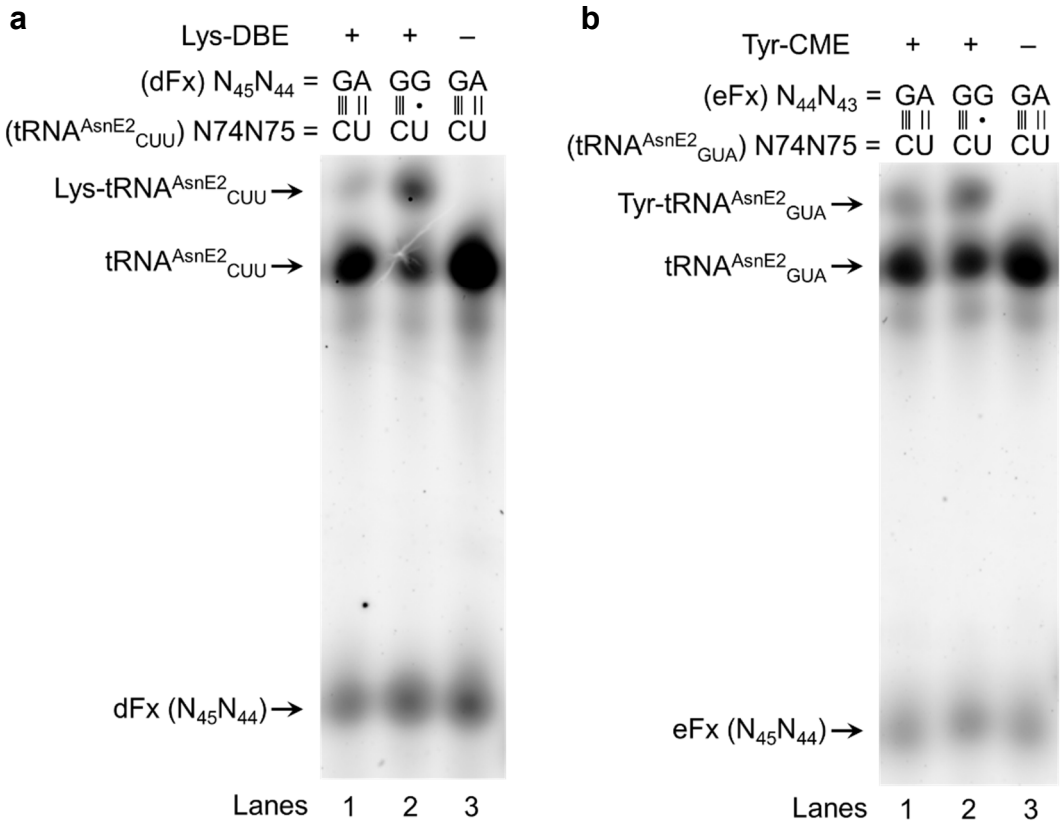

Figure 24 | Aminoacylation of tRNA-CUA derived from tRNA^{AsnE2}. tRNAs were acylated by dFx in the presence of Lys-DBE (**a**), and eFx in the presence of Tyr-CME (**b**). Acid-PAGE separated the bands of aa-tRNA (upper) and uncharged tRNA (lower). The data were generated from a sample of the end product of aminoacylation reaction.

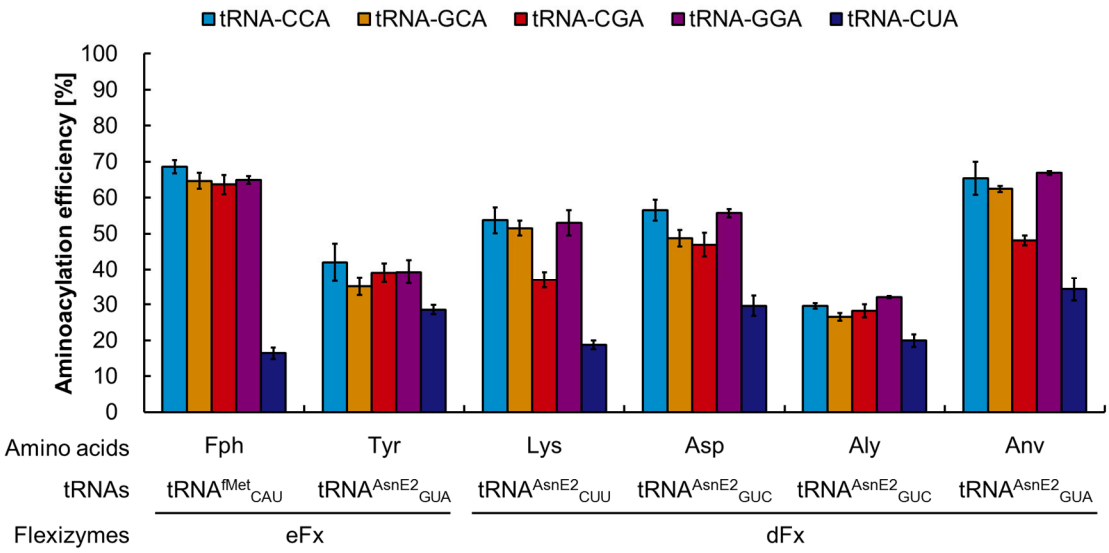

Figure 25 | Efficiency of flexizymes-catalyzed aminoacylation with canonical (Tyr, Lys, and Asp) and non-canonical (Fph, Aly, and Anv) amino acids. The tRNAs-CCA, -GCA, -CGA, -GGA and -CUA were acylated by the compensatory flexizymes, dFx/eFx-GGU, -GCU, -CGU, -CCU and -AGU. The data represent the average of three independent reactions. The error bars represent the standard deviation. The data were generated from a sample of the end product of aminoacylation reaction.

3.2.2. Preparation of tagged ribosomes

Mutant ribosomes bearing a mutation (or mutations) at PTC were prepared using an MS2-tag affinity purification method[140]. Professor Rachel Green in Johns Hopkins University kindly gifted me a plasmid p278 MS2 (Amp^R) encoding 23S rRNA inserted with MS2-tag sequence in helix 98 of 23S rRNA (Figure 26a), pcI^{857} (Kan^R) encoding a temperature-sensitive mutant of the lambda repressor protein cI and a plasmid pMAL-c2g (Amp^R) encoding His_6-tagged MS2 coat protein fused with maltose binding protein (His_6-MS2-MBP).

The plasmids p278 MS2 bearing mutations at PTC were constructed by site-directed mutagenesis. First, the plasmid pMAL-c2g was used to transform *E. coli* and His_6-MS2-MBP was expressed to be purified by FPLC using Ni-NTA affinity column. Next, the plasmids pcI^{857} and p278 MS2 were used to transform *E. coli*. Because the mutant ribosomes may be toxic to *E. coli*, this transformed *E. coli* was pre-cultured at 30°C where mutant ribosomes were not expressed and the temperature was increased to 42°C to express the mutant ribosomes. His_6-MS2-MBP was immobilized onto the MBP column by FPLC and the expressed tagged-ribosomes (wild-type, G2251C, G2252C, G2553C, G2251C/G2553C, G2251C/G2252C/G2553C and G2251A/G2553A) were purified by MS2 affinity purification using FPLC (Figure 26b). Contamination of wild-type untagged-ribosome into the purified tagged-ribosomes were quantified by primer extension method[140]. This result showed the negligible contamination rate less than 3% (Figure 27).

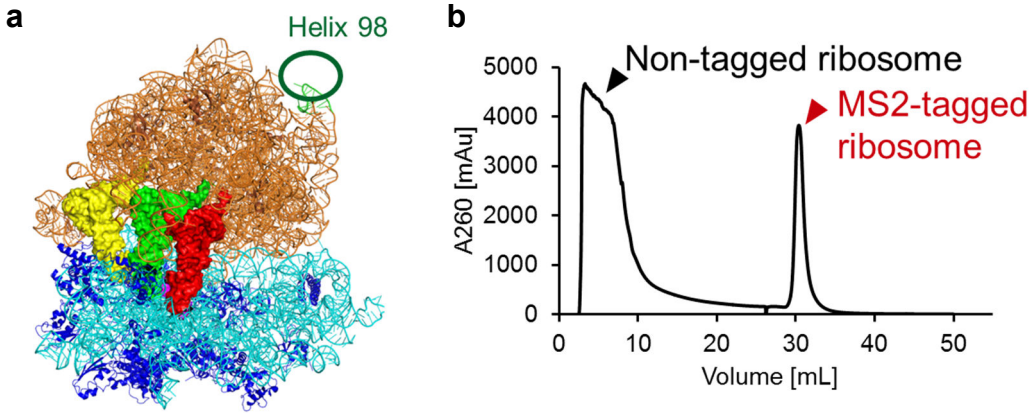

Figure 26 | Schematic illustration of the insertion sites for the MS2 tag and purification of tagged ribosome. (**a**) Helix 98 in 23S rRNA which is the insertion sites for the MS2 tag was indicated as green circle. The image of structure was produced from the previous report [62] (PDB ID: 2WDK/2WDL). (**b**) Representative results of FPLC purification of MS2-tagged ribosome.

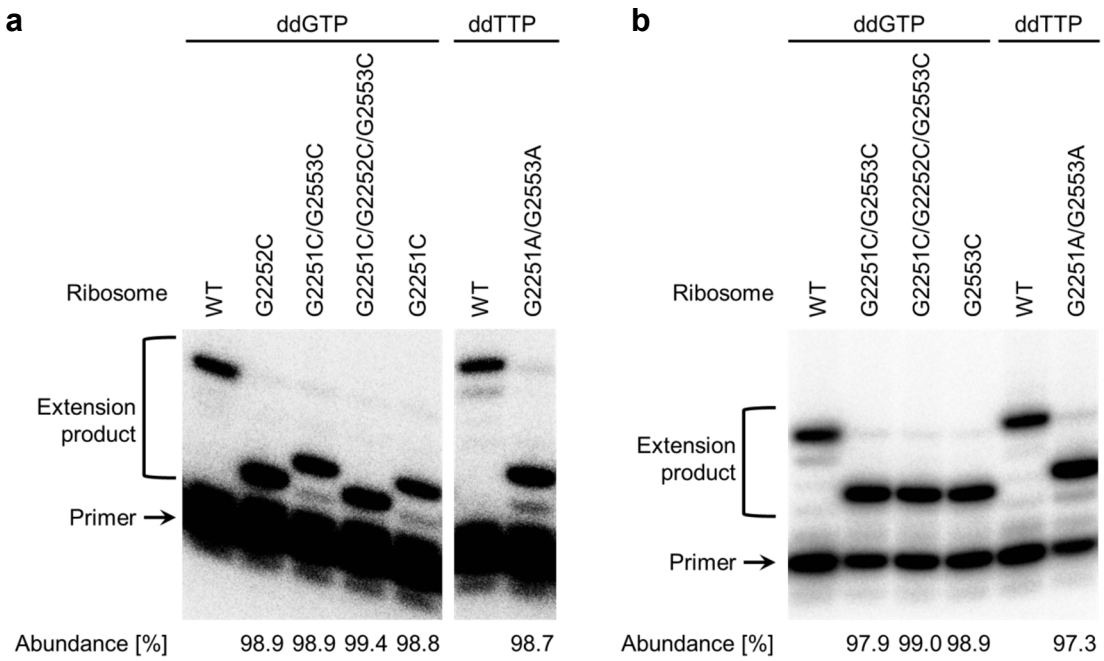

Figure 27 | Primer extension analysis of rRNAs extracted from MS2-tagged ribosomes purified by FPLC. (**a**) A primer that is complementary to the bases of 2254–2273 in 23S rRNA was used for the detection of mutations of 2251 and 2252. (**b**) A primer that is complementary to the bases of 2556–2575 in 23S rRNA was used for the detection of mutation of 2553. MS2-taggd ribosomes bearing mutations produced different ddGTP or ddTTP stops from those produced by WT ribosome. The abundance of tagged ribosomes relative to untagged wild-type in each population was shown below the gel. The calculated values of abundance were derived from a single experimental data set.

3.2.3. Translation activity of wild-type/mutant ribosome-tRNA pairs

An assay system to detect peptides translated by tagged-ribosomes and aa-tRNAs-NNA was established. Because *E. coli* ARSs do not uniformly charge amino acids onto tRNAs-NNA[6-8], a conventional methods using radioisotope-labeling amino acids is unsuitable. Instead, fluorescently-labeled amino acid (*N*-(5-FAM)-L-phenylalanine, Fph, Figure 21) was aminoacylated onto initiator $tRNA^{fMet}_{CAU}$ and incorporated at N-terminus of peptide by FIT system[126] to reassign the initiation codon from fMet to Fph. Lysine and tyrosine were charged onto $tRNA^{AsnE2}_{CUU}$-CCA and $tRNA^{AsnE2}_{GUA}$-CCA by cognate dFx and eFx, respectively. As a model template, a DNA template encoding the heptapeptide-1 (Fph-Lys-Tyr-Lys-Lys-Tyr-Lys) was prepared and the expression of heptapeptide-1 in the FIT system, whose components were described in the section of materials and methods, was confirmed. Tricine-SDS-PAGE analysis of the translated product enabled to visualize a fluorescent band (Figure 28a) and quantify an expressed peptide by comparison with a band generated by a known concentration of Fph (Figure 28b), and the peak corresponding to heptatpeptide-1 was confirmed by MALDI-TOF-MS (Figure 28c). The production of heptapeptide-1 plateaued at 30 min, with a final concentration of approximately 1.0 μM. We also examined expression of heptapeptide-1 using Lys-$tRNA^{GluE2}_{CUU}$-CCA and Tyr-$tRNA^{GluE2}_{GUA}$-CCA under the same condition, and the amount of expressed peptide increased to 3.5 μM (Figure 29a and b). The increase of expression rate could be attributed to the higher binding affinity of aa-$tRNAs^{GluE2}$-CCA compared to aa-$tRNA^{AsnE2}$-CCA (Figure 30). Therefore, we decided to use $tRNA^{GluE2}$ for further translation experiments.

Next, I investigated whether tRNAs-NNA can be used by the wild-type ribosome for mRNA-dependent translation. Four mutant aa-tRNAs bearing GCA-3′, CGA-3′ and GGA-3′ were prepared by cognate dFxs and eFxs (Figure 25) and tested these expressions. Surprisingly tRNA-GCA could be used by wild-type ribosome to express heptapeptide-1 even only one base pair forms at P-site and the final concentration of the peptide was 2.7 μM (Figure 29). On the other hand, tRNAs-CGA and -GGA were inactive with wild-type ribosome (Figure 29). These results indicate that tRNAs-CGA and tRNAs-GGA but not tRNAs-GCA are orthogonal to the wild-type tRNAs-CCA (Figure 31).

Then, I measured the translation activity of mutant ribosomes and tRNAs. As predicted from the results with the pair of wild-type ribosome and non-cognate tRNAs-GCA, the G2252C ribosome was also active with both cognate tRNAs-GCA and non-cognate tRNAs-CCA, but not with tRNAs-CGA and tRNAs-GGA (Figure 29a). The expression rate of cognate pair of G2252C ribosome and tRNAs-GCA was almost the

same as that of wild-type pair, and that of non-cognate pair of G2252C ribosome and tRNAs-CCA was also the same as that of non-cognate pair of wild-type ribosome and tRNAs-GCA (Figure 29b). These both cognate and non-cognate pairs yielded heptapeptide-1 as major products (Figure 29c). These results indicated that mispair between G2252 and G74 or C2252 and C74 could be tolerated during the translation reaction.

The double mutant ribosome G2251C/G2553C ribosome was sufficiently active with only cognate tRNAs-CGA to yield heptapeptide-1 (Figure 29). Time-course analysis indicated that expression plateaued at a final concentration of 1.7 μM. This suggests that the cognate pair of G2251C/G2553C ribosome and tRNAs-CGA functions orthogonally to the wild-type pair (Figure 31). However, this mutant pair has a slower translation rate than the wild-type pair, resulting in a lower yield of peptide due to the hydrolysis of the aa-tRNAs. Faint bands observed when using tRNAs-CCA and tRNAs-GCA could be ascribed to a peptide originating from the background translation by little amount (< 3%) of wild-type ribosome contamination to the respective mutant ribosome (Figure 27). When tRNAs-GGA was used, a very faint band was detected because mispair between G2252 and G74 was tolerated as described above. However, these bands were too faint to reliably quantify the intensity for the time-course experiment below and determine the product identity by MALDI-TOF analysis.

The triple mutant ribosome (G2251C/G2252C/G2553C) was active with only cognate tRNAs-GGA to yield heptapeptide-1 like the double mutant ribosome (Figure 29). However, the translation rate dramatically decreased and the final concentration of heptapeptide-1 was 0.4 μM, which was reflected as a tiny peak of heptapaptide-1 in MALDI-TOF-MS spectrum (Figure 29c).

I also tested the translational activity of other ribosome mutants (G2251A/G2553A, G2251C and G2553C) and tRNA mutant (tRNA-CUA) (Figure 32). Like G2251C/G2553 ribosome and G2251C/G2252C/G2553C ribosome, G2251A/G2553A ribosome was active with only cognate tRNA-CUA, but the band was too faint to reliably quantify the intensity for the time-course experiment and determine the product identity by MALDI-TOF analysis. A very faint band observed when using tRNAs-CCA could be due to the contamination of wild-type ribosome (Figure 27). These results indicate that A-U pair between 23S rRNA and tRNA is not appropriate for translation. In addition, wild-type ribosome was inactive with tRNAs-CUA, which suggests that G•U wobble pair between ribosome and tRNA at the PTC is not tolerated. In the case of G2251C and G2553C ribosomes, translation activity was not restored when

using tRNA-CCA or tRNA-CGA. This result and the fact that the pair of G2251C/G2553C ribosome and tRNA-CGA had comparable translation activity to wild-type pair, combination of two base pairs, G2553-C75 at the A-site and G2251-C75 at the P-site, is essential for translation.

These decrease of translation rate of cognate pair of mutant ribosome and tRNAs indicates that, despite the fact that PT activity and translocation activity of mutant ribosomes bearing mutations at PTC was restored by the compensatory mutations into CCA end of tRNAs[63,64], these compensatory mutations did not fully restore the whole translation activity. This further suggests that these bases have crucial roles in not only the PT reaction and translocation but also other steps of translation, potentially involving interactions with other rRNAs or protein factors. It was confirmed that the binding affinity of aa-tRNAs-NNA to form ternary complexes (EF-TU•GTP•aa-tRNA) were almost the same as that of wild-type, which was coincident with the previous report[130,141]. In the accommodation step, C75 of tRNA pack between EF-Tu residue 219 and the flipped base of A55 of 16S rRNA, which is 99.5% conserved in all species of known sequence[142]. It is possible that these interaction is critical for efficient translation reaction. Nonetheless, further studies are necessary to elucidate why compensatory mutations at CCA-end could not fully restore the translation activity of mutant ribosomes.

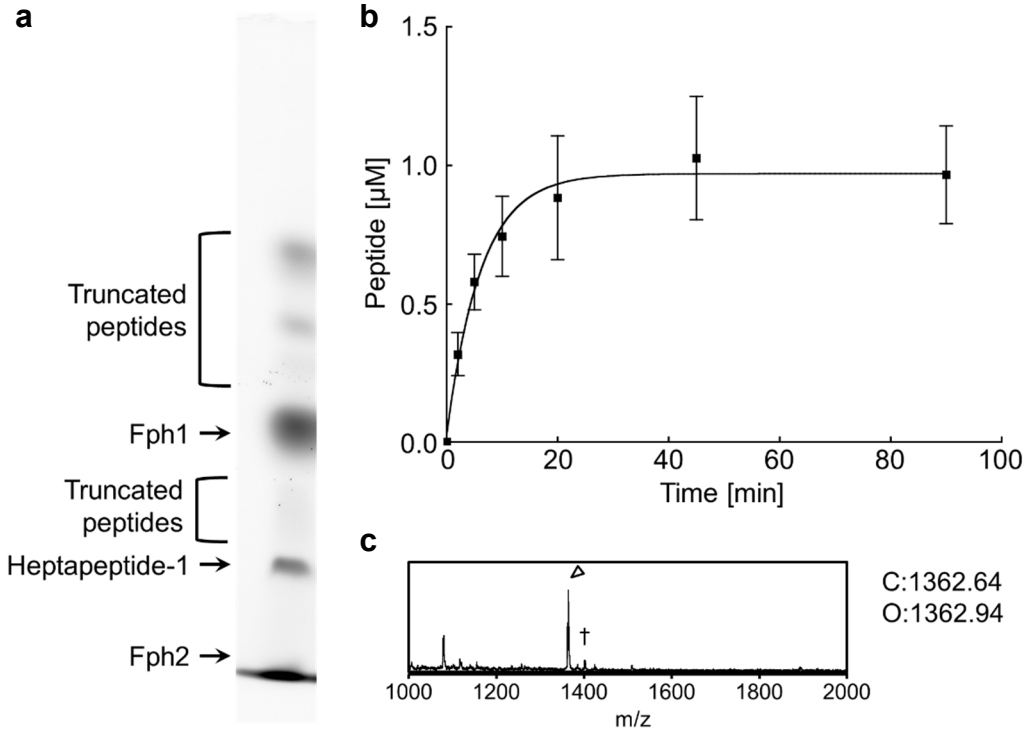

Figure 28 | Translation activity of pairs of wild-type ribosome–tRNA using tRNAs^{AsnE2} as elongator tRNAs. (**a**) Tricine-SDS-PAGE analysis of heptapeptide-1 (FphLysTyrLysLysTyrLys) synthesized using the wild-type pairs of ribosome–tRNAs. tRNA$^{fMet}_{CAU}$ bearing the CCA-3′ end were used as initiator tRNAs, and tRNAs$^{AsnE2}_{NNN}$ (NNN denotes CUU and GUA designating Lys and Tyr, respectively) bearing the CCA-3′ end were used as elongator tRNAs. This image showed the result of the end of the translation reaction. Two bands were observed for Fph (indicated by Fph1 and Fph2), presumably due to the presence of Fph, Fph-CME and Fph-tRNA$^{fMet}_{CAU}$. The bands above and below of Fph1 showed truncated peptides generated by peptidyl-tRNA drop-off[143] (**b**) Time-course analysis of heptapeptide-1 production. The data represent the average of three independent reactions and error bars represent the standard deviation of the individual measurements. (**c**) MALDI-TOF MS analysis of heptapeptide-1. C and O denote calculated and observed mass values, respectively. † indicates a potassium adduct of heptapeptide-1. Other minor peaks were also present in the non-templated translation product likely originating from the translation system. The data were generated from a sample of the end product of translation reaction.

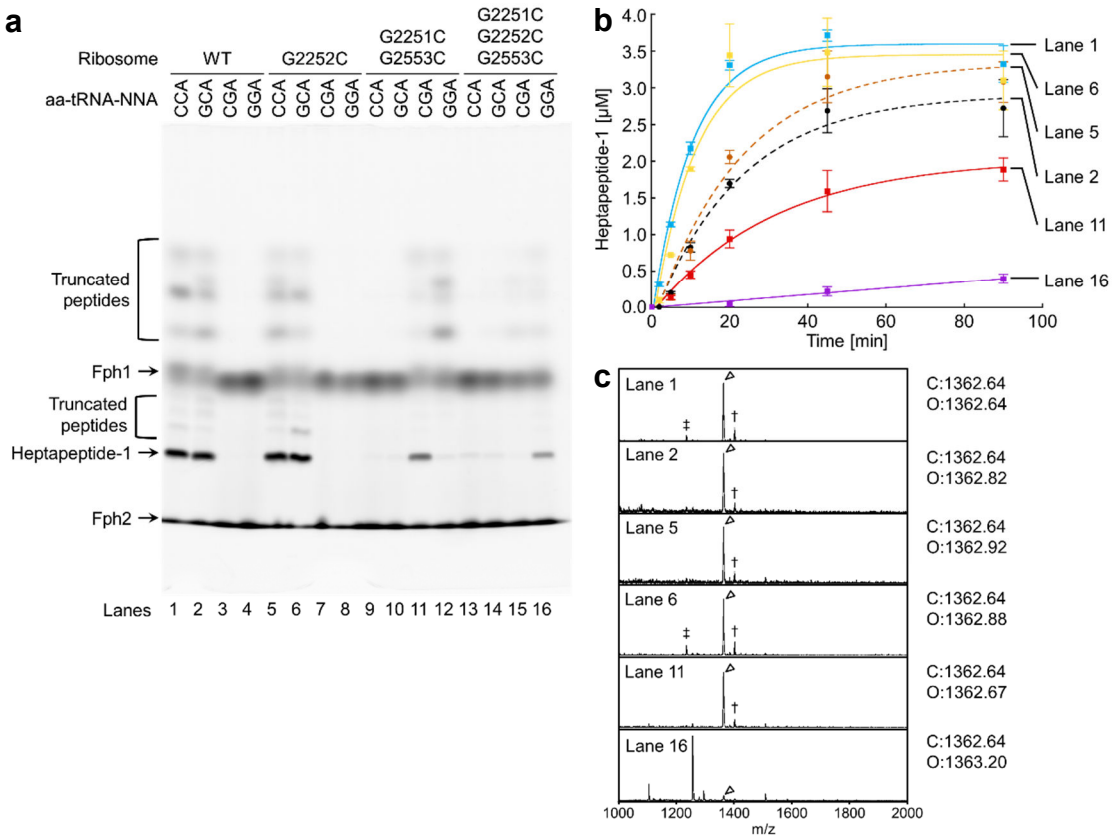

Figure 29 | Translational activity of pairs of ribosome and tRNA mutants. (**a**) Tricine-SDS-PAGE analysis of heptapeptide-1 (Fph-Lys-Tyr-Lys-Lys-Tyr-Lys) synthesized using the designated pairs of ribosome–tRNA mutants. tRNAs$^{fMet}_{CAU}$ bearing the respective NNA-3′ end were used as initiator tRNAs, and tRNAs$^{GluE2}_{NNN}$ (NNN denotes CUU and GUA designating Lys and Tyr, respectively) bearing the respective NNA-3′ end were used as elongator tRNAs. This image showed the result of the end of the translation reaction. (**b**) Time-course analysis of heptapeptide-1 production by active pairs of ribosomes and tRNAs. I investigated four cognate pairs of the wild-type ribosome–tRNAs-CCA (cyan solid line), G2252C-ribosome–tRNAs-GCA (yellow solid), G2251C/G2553C-ribosome–tRNAs-CGA (red solid), and G2251C/G2252C/G2553C-ribosome–tRNAs-GGA (purple solid), and two non-cognate pairs of the wild-type ribosome–tRNAs-GCA (black dashed) and G2252C-ribosome–tRNAs-CCA (brown dashed). The data represent the average of three independent reactions. The error bars represent the standard deviation. (**c**) MALDI-TOF MS analysis of heptapeptide-1 synthesized using cognate and non-cognate ribosome–tRNA pairs. Lane numbers are those of the tricine-SDS-PAGE gel described in Figure 29. C and

O denote calculated and observed mass values, respectively. † indicates a potassium adduct of heptapeptide-1. ‡ indicates the molecular mass value consistent with FphLysTyrLysLysTyr, which was presumably produced by peptidyl-tRNA drop-off[143]. The identity of major peak in lane 16 is unknown; however, since this peaks as well as other minor peaks were also present in the sample of G2251C/G225C/G2553C-ribosome alone, it could be originated from a contaminant of the G2251C/G225C/G2553C-ribosome preparation. The data were generated from a sample of the end product of translation reaction.

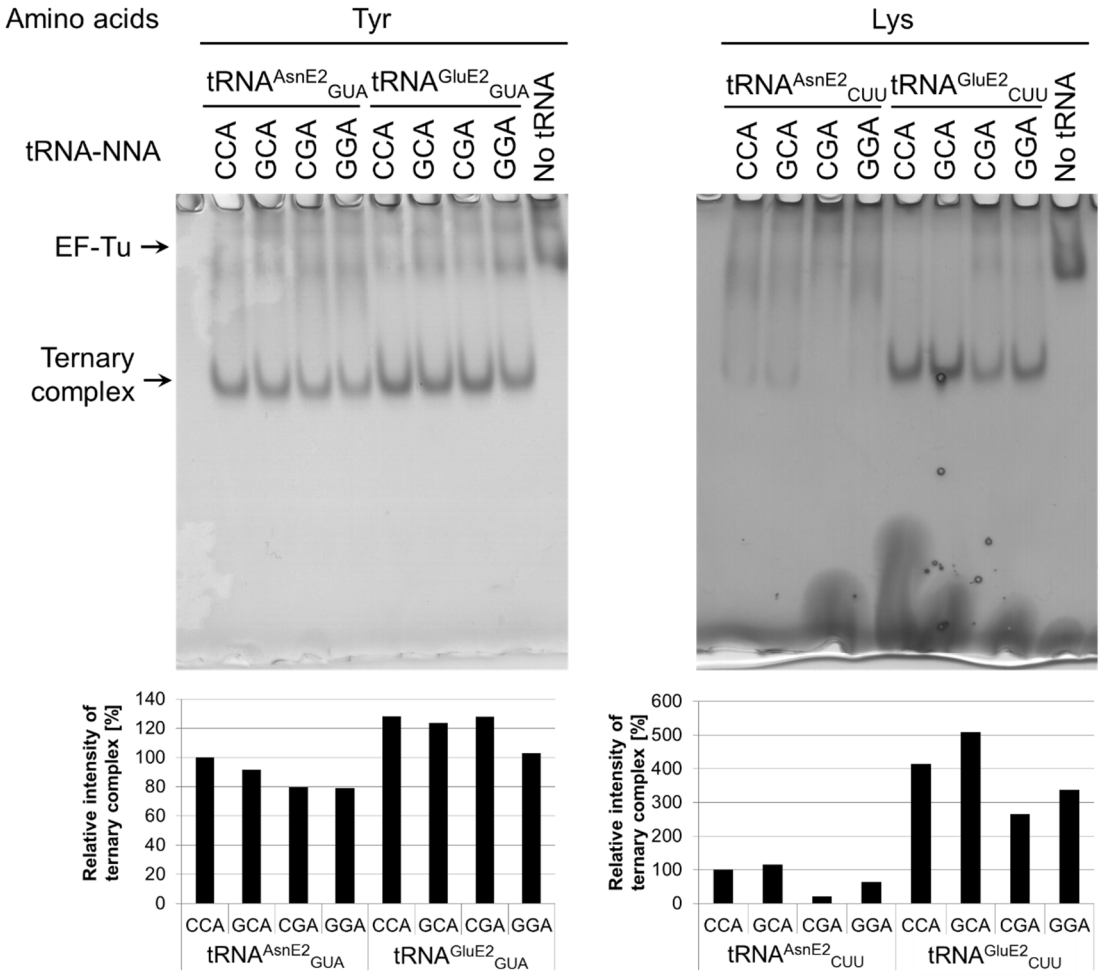

Figure 30 | Gel-shift assay for ternary complex of elongation factor Tu (EF-Tu), guanosine triphosphate (GTP) and aa-tRNAs-CCA, GCA, CGA and GGA. Tyrosine was charged onto $tRNA^{AsnE2}_{GUA}$ and $tRNA^{GluE2}_{GUA}$, and lysine was charged onto $tRNA^{AsnE2}_{CUU}$ and $tRNA^{GluE2}_{CUU}$. Band intensities of the ternary complexes were normalized relative to EF-Tu, GTP and aa-tRNAs-CCA complexes. The values of the relative intensity of the ternary complex (y-axis) were derived from a single experimental set.

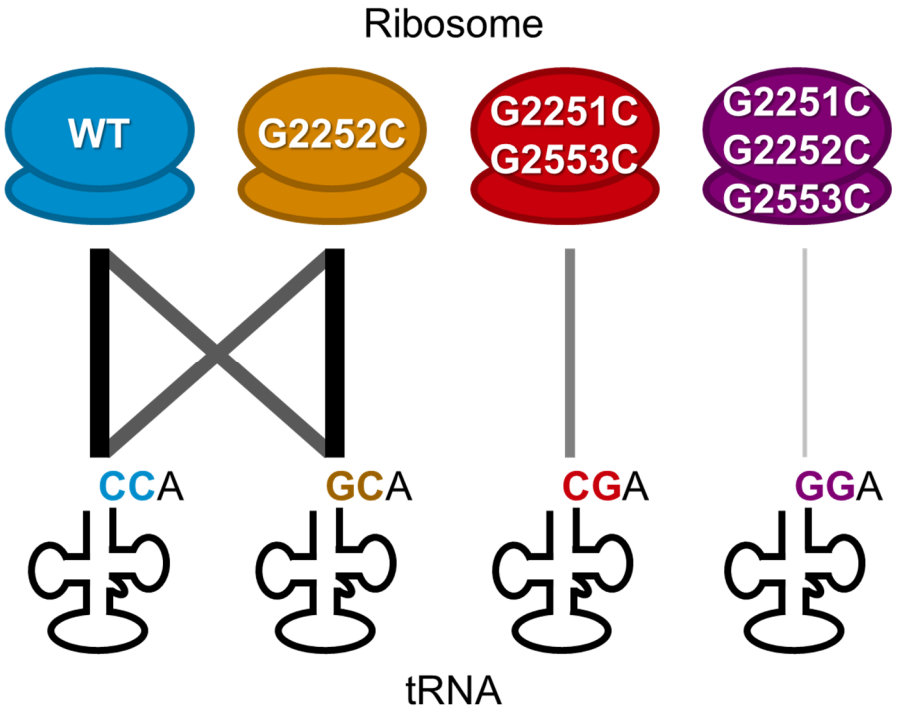

Figure 31 | Illustration of compatibility and orthogonality of the ribosome–tRNAs mutant pairs. Line thickness indicates the compatibility of translational activity between each ribosome–tRNA pair.

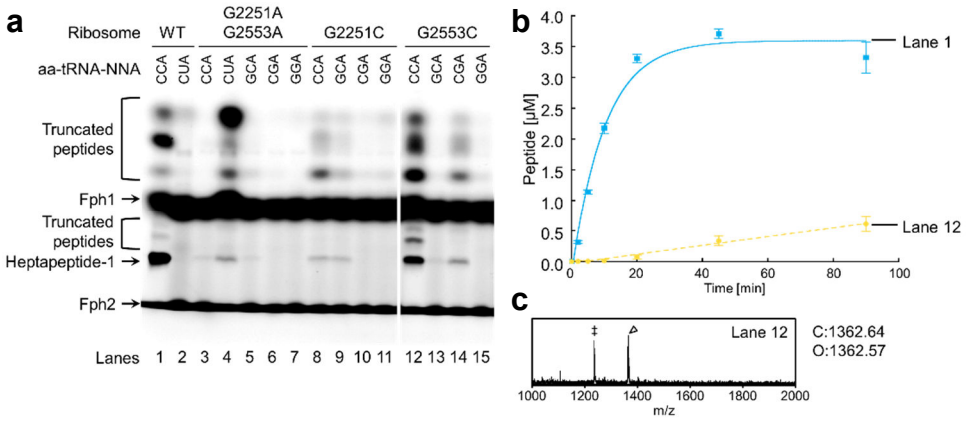

Figure 32 | Translational activity of pairs of ribosome–tRNA mutants bearing CCA, GCA, CGA, GGA and CUA-3′ ends. tRNAs^{fMet} were used as initiator tRNAs and tRNAs^{GluE2} as elongator tRNAs. (**a**) The entire image of the tricine-SDS-PAGE analysis of a heptapeptide-1 (FphLysTyrLysLysTyrLys) synthesized using pairs of ribosome–tRNA mutants. This image showed the result of the end of the translation reaction. The bands above and below of Fph1 showed truncated peptides generated by peptidyl-tRNA drop-off[143], which were also observed for the wild-type control of lane 1. The faint bands in lanes 3, 8, 9 and 13 could be ascribed to a peptide originating from the background translation by little amount (< 3%) of wild-type ribosome contamination to the respective mutant ribosome (Figure 27). (**b**) Time-course analysis of heptapeptide-1 production by active pairs of ribosomes and tRNAs. Lane numbers are those of the tricine-SDS-PAGE gel described in Figure 32a, in which the solid line in cyan indicates the reaction of a cognate pairs of the wild-type ribosome–tRNAs-CCA (the identical result shown in Figure 29b) while the dashed line in yellow shows that of a non-cognate pair G2253C-ribosome–tRNAs-CCA. The data represent the mean value of each sample (n = 3) and error bars show the standard deviations obtained from independent measurements. (**c**) MALDI-TOF MS analysis of heptapeptide-1 synthesized using a non-cognate pair G2253C-ribosome–tRNAs-CCA. Lane numbers are those of the Tricine-SDS-PAGE gel described in Figure 32a. C and O denote calculated and observed mass values, respectively. ‡ indicates the molecular mass value consistent with FphLysTyrLysLysTyr, which could be one of the bands corresponding to the truncated peptides appeared in the tricine-SDS-PAGE. The data shown in Figure 32a and c were generated from a sample of the end product of translation reaction.

3.2.4. Simultaneous expression of two different peptides from single mRNA template

Because the pair of G2251C/G2553C ribosome and tRNAs-CGA had comparable translation activity and orthogonality to the wild-type ribosome and tRNA-CCA pair, I examined whether two distinct peptides could be expressed from a single mRNA template under artificially programmed genetic code in one pot. Four amino acids were assigned to the respective genetic codes as follows; For the wild-type ribosome–tRNAs pair, a genetic code (referred to as WT-code) assigned Fph (AUG, initiation), lysine (AAG), tyrosine (UAC) and aspartic acid (GAC), and for the G2251C/G2553C ribosome–tRNAs-CGA pair, an orthogonal genetic code (referred to as OR-code) assigned Fph (AUG, initiation), lysine (AAG), L-azidonorvaline (Anv; UAC) and L-acetyllysine (Aly; GAC) (Figure 33). The tRNAs-CCA and tRNAs-CGA were aminoacylated with respective amino acids using wild-type or compensatory mutated flexizymes. The DNA template encoding heptapeptide-2 according to the WT-code and heptapeptide-3 according to the OR-code was added to the FIT system, which comprised the combination of the wild-type ribosomes. G2251C/G2553C ribosomes and/or the above aa-tRNAs (Figure 34a).

In the presence of wild-type ribosome and wild-type aa-tRNAs-CCA, heptapeptide-2 was expressed according to WT-code (Figure 34 b-d). When aa-tRNAs-CGA was added into wild-type pair of ribosome and aa-tRNAs-CCA, only heptapeptide-2 was observed in tiricine-SDS-PAGE and MALDI-TOF-MS spectrum without any hybrid products (for example, Fph-Lys-Anv-Asp-Lys-Tyr-Aly, $[M+H]^+$ = 1313.51) generated from potential cross-reading (or readings) of codons in non-cognate genetic code (Figure 34 b and c). This indicates that aa-tRNAs-CGA was not used as translation substrates by wild-type ribosome. Similarly, the mutant pair of G2251C/G2553C ribosome and aa-tRNAs-CGA or this pair with the wild-type aa-tRNAs-CCA yielded only heptapeptide-2 (Figure 34 b-d). The decrease of peptide expression was observed in the presence of one ribosome with both cognate and non-cognate aa-tRNAs (Figure 34 c and d). This decreased expression of peptide could not be restored by increasing EF-Tu concentrations from 10 μM to 20 μM, which rules out EF-Tu sequestration by the non-cognate aa-tRNAs as an explanation for this effect (Figure 35). From this result and the fact that the affinity of aa-tRNAs-CCA to EF-Tu was almost the same as that of aa-tRNAs-CGA (Figure 30), it is possible that ternary complex including non-cognate aa-tRNAs compete with that including cognate aa-tRNAs in accommodation step. Nonetheless, further studies are necessary to restore this decrease of expression rate.

In the presence of both wild-type and mutant pair of ribosome and aa-tRNAs, the desired both heptapeptide-2 and -3 were expressed from a single DNA template in one pot according to the WT- and OR-codes (Figure 34). In the absence of DNA template, no peptide was observed (Figure 34c), which is consistent with the observation that translation proceeded in an mRNA-dependent manner. Moreover, the same experiment using different DNA template and aa-tRNAs produced two different heptapeptides from one DNA template (Figure 36 and 37).

In these experiments, two distinct peptides can be expressed from one DNA template without any hybrid products. These results indicate that these coexisting translational machineries acted orthogonally and used only the cognate genetic codes.

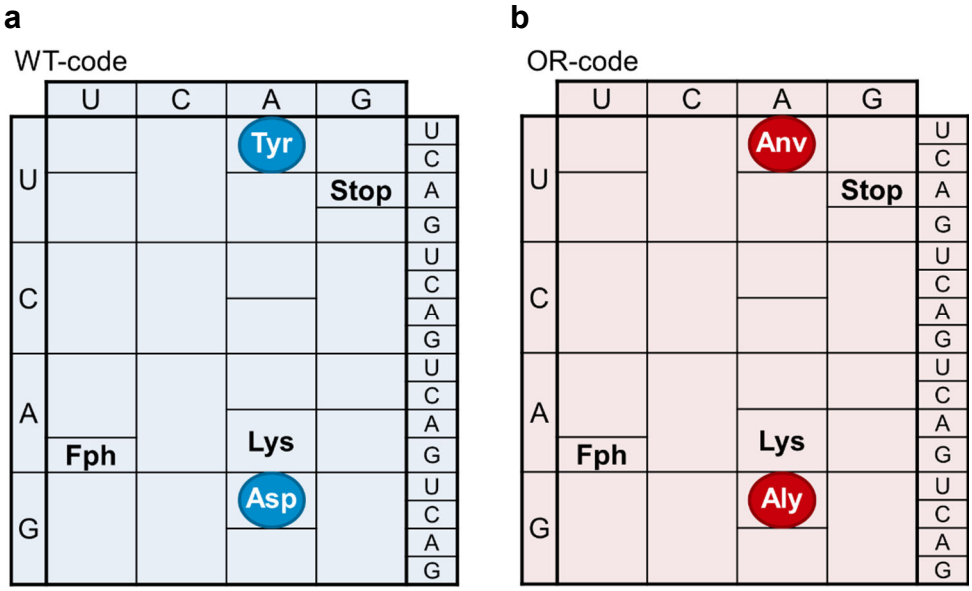

Figure 33 | Two genetic codes programmed for simultaneous expression of two distinct peptides from a single mRNA sequence. (a) WT-code. This code comprises the wild-type ribosome–tRNAs-CCA pair. (b) OR-code. This code comprises the G2251C/G2553C-ribosome–tRNAs-CGA pair.

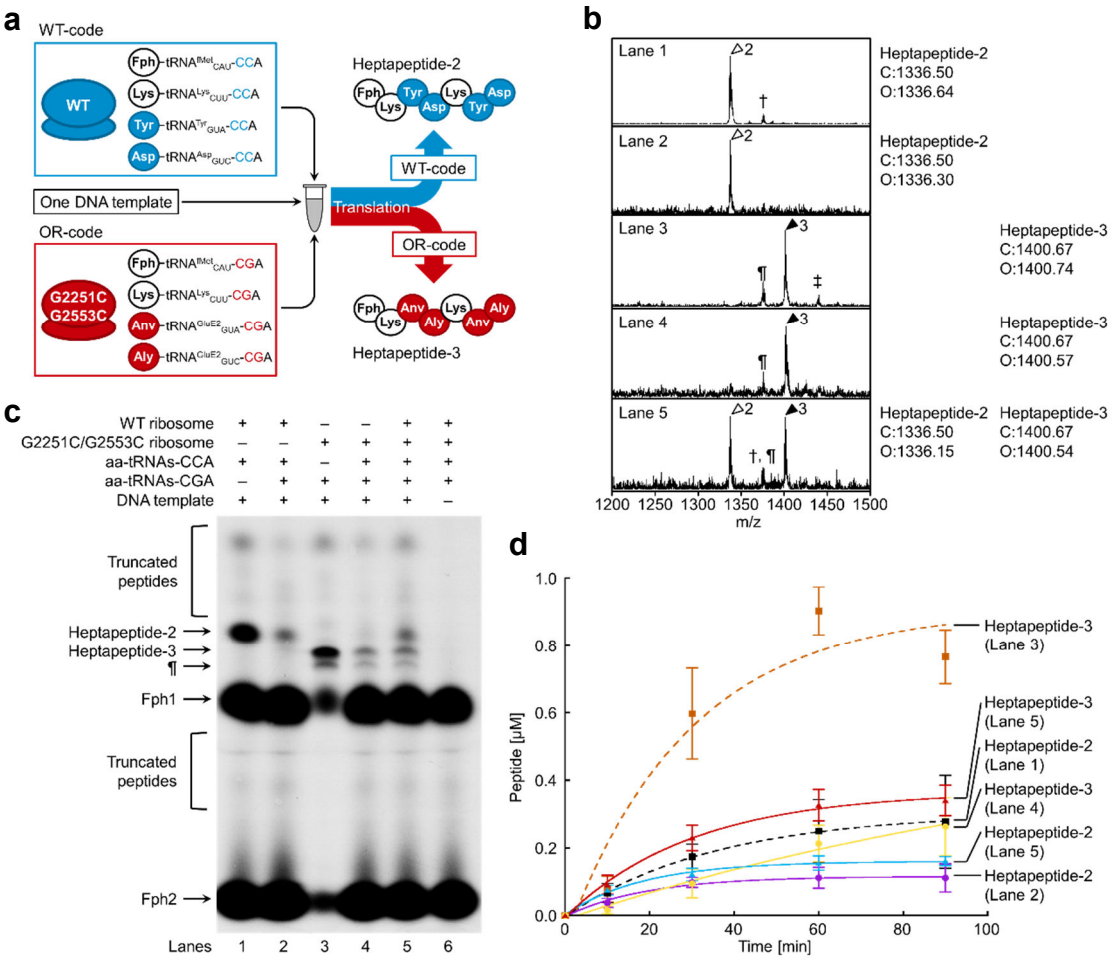

Figure 34 | Simultaneous expression of two distinct peptides from a single mRNA sequence under two artificially programmed genetic codes. (**a**) Schematic illustration of simultaneous expression. The wild-type ribosome–tRNAs-CCA and G2251C/G2553C-ribosome–tRNA-CGA pairs generated heptapeptide-2 and heptapeptide-3, respectively, according to the WT- and OR-codes. (**b**) MALDI-TOF MS analysis of translational products. Calculated (C) and observed mass (O) values are shown in the right panel of the spectra. † and ‡ denote a potassium adduct of heptapeptide-2 and heptapeptide-3, respectively. ¶ denotes heptapeptide-3 whose azide group was reduced to a primary amino group presumably by thiols present in the FIT system[144]. Since other minor unidentified peaks were present in the non-templated translation product, they are likely present in the translation system. The data shown in Figure 34b and c were generated from a sample of the end product of translation reaction. (**c**) Tricine-SDS-PAGE analysis of the respective heptapeptides. (**d**) Time-course analysis of simultaneous

expression of heptapeptides. Heptapeptide-2 expression is seen in lane 1 (black dashed line), lane 2 (purple) and lane 5 (cyan, simultaneous expression); heptapeptide-3 expression is seen in lane 3 (brown dashed line) lane 4 (yellow) and lane 5 (red, simultaneous expression). Lane numbers are those described in Figure 34c. The data represent the average of three independent reactions and error bars represent the standard deviation.

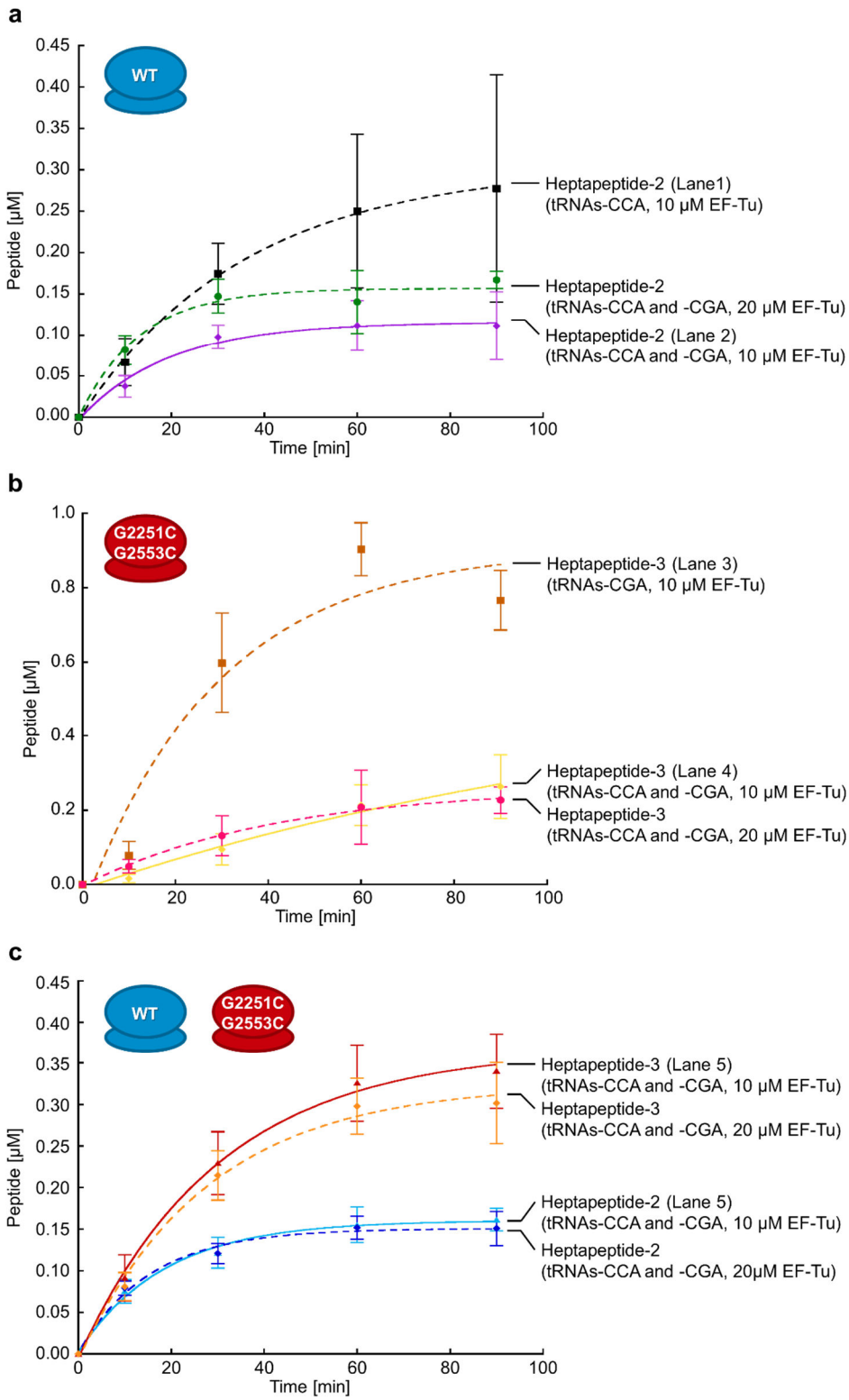

Figure 35 | Time-course analysis of heptapeptide production with an elevated concentration of EF-Tu. The concentration of EF-Tu was increased to 20 μM from 10 μM which was the condition in the experiments in Figure

34. (**a**) Production of heptapeptide-2 when wildype ribosome coexisted with cognate aa-tRNAs-CCA and noncognate aa-tRNA-CGA with 20 μM EF-Tu/Ts were shown as green dashed line. (**b**) Production of heptapeptide-3 when orthogonal ribosome coexisted with its noncognate and cognate aa-tRNAs with 20 μM EF-Tu/Ts were shown as pink dashed line. (**c**) Production of heptapeptide-2 and -3 when both ribosomes with all aa-tRNAs coexisted with 20 μM EF-Tu/Ts were shown as blue and orange dashed lines. The other lines represent the same result as those shown in Figure 34d. Lane numbers are those described in Figure 34c. The data represent the average of three independent reactions, and the error bars represent the standard deviation of the individual measurements.

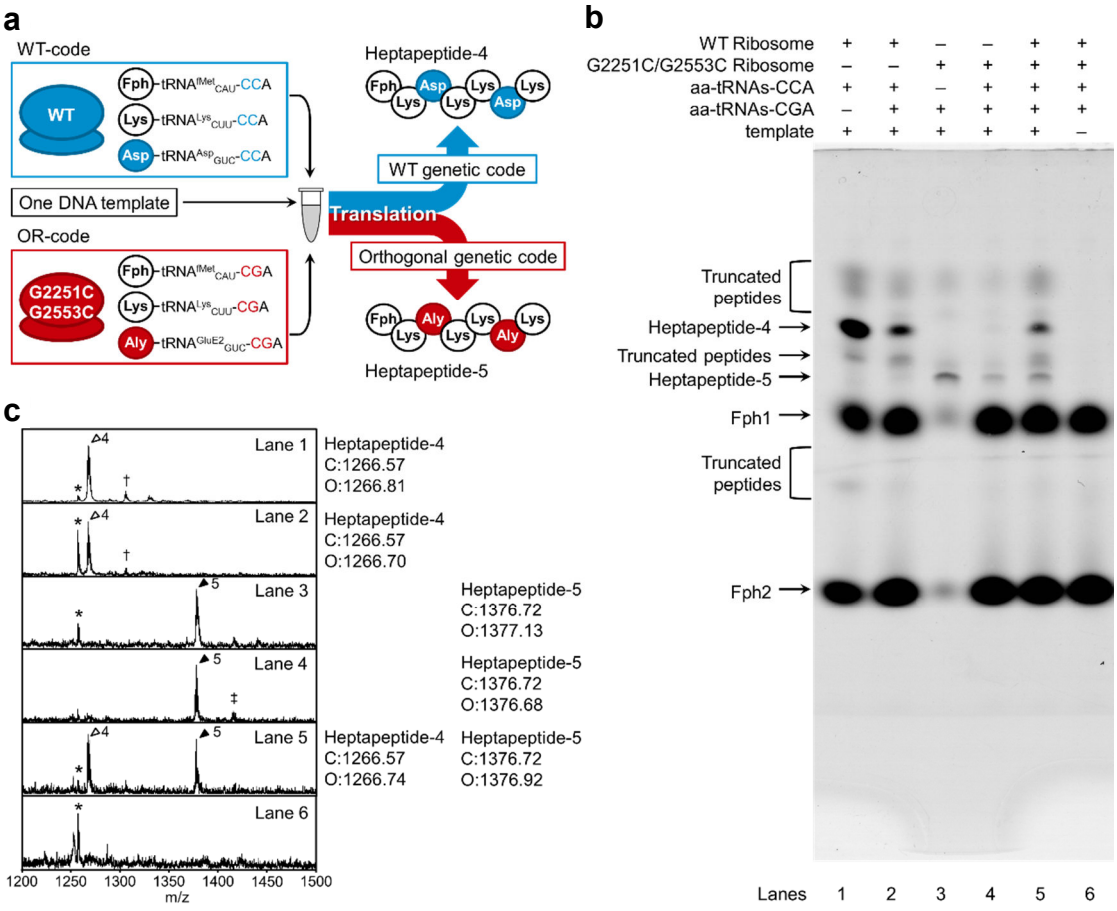

Figure 36 | Simultaneous expression of two heptapeptides-4 and -5 from a single mRNA template according to the two artificially programmed genetic codes. (**a**) Peptide translation from a single mRNA sequence in a single reaction mixture. The pair of wild-type ribosome–tRNAs-CCA and G2251C/G2553C-ribosome–tRNA-CGA generated Fph-Lys-Asp-Lys-Lys-Asp-Lys (heptapeptide-4) and Fph-Lys-Aly-Lys-Lys-Aly-Lys (heptapeptide-5), respectively, according to their respective genetic codes. (**b**) Tricine-SDS-PAGE analysis of the respective heptapeptides, 4 and 5. Other faint based are truncated peptides generated by peptidyl-tRNA drop-off[143]. (**c**) MALDI-TOF MS of the peptides isolated by gel filtration. Calculated (C) and observed mass (O) values are shown in the right panel of the spectra. † indicates the potassium adduct of heptapeptide-4, ‡ indicates the potassium adduct of heptapeptide-5, *species also present in the "no template" translation product and were likely derived from the translation mixture. The data shown in Figure 36b and c were generated from a sample of the end product of translation reaction.

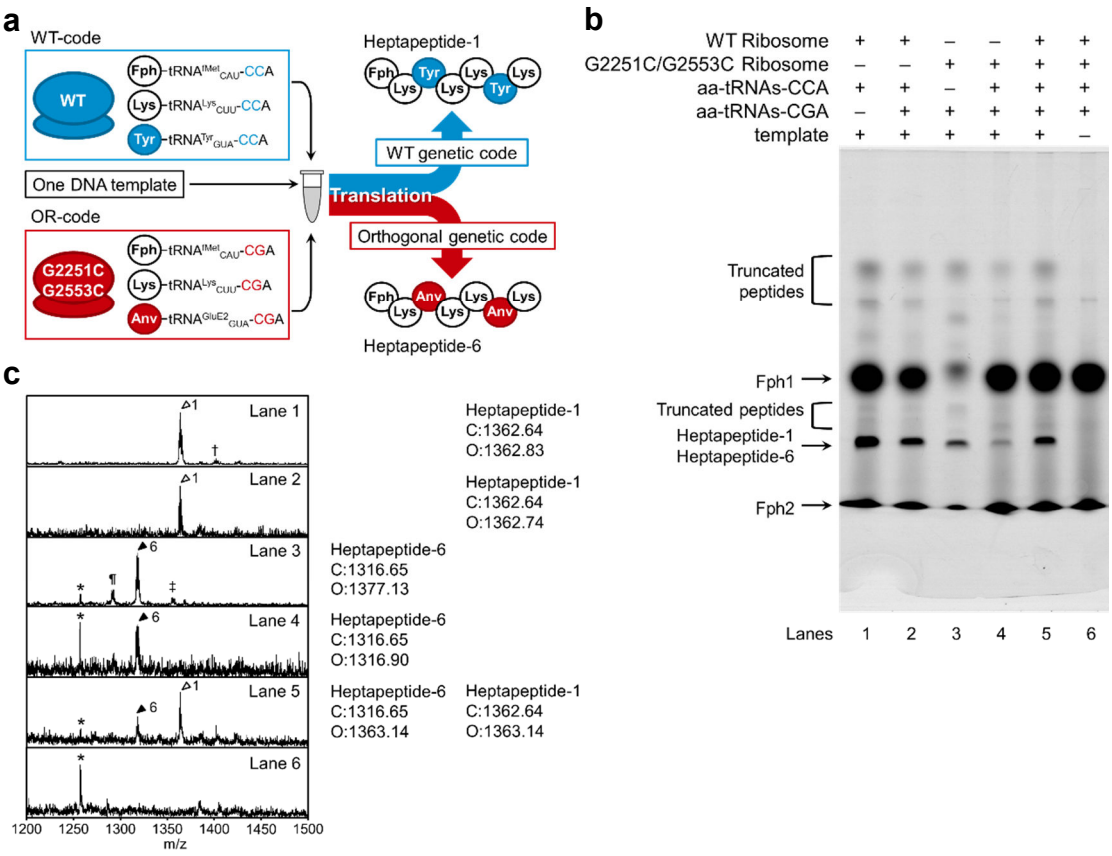

Figure 37 | Simultaneous expression of two heptapeptides-1 and -6 from a single mRNA template according to the two artificially programmed genetic codes. (**a**) Peptide translation from a single mRNA sequence in a single reaction mixture. The pair of wild-type ribosome–tRNAs-CCA and G2251C/G2553C-ribosome–tRNA-CGA generated Fph-Lys-Asp-Lys-Lys-Asp-Lys (heptapeptide-1) and Fph-Lys-Anv-Lys-Lys-Anv-Lys (heptapeptide-6), respectively, according to their respective genetic codes. (**b**) Tricine-SDS-PAGE analysis. Heptapeptides-1 and -6 could were observed at the same position. Other faint based are truncated peptides generated by peptidyl-tRNA drop-off[143]. (**c**) MALDI-TOF MS of the peptides isolated by gel filtration. Calculated (C) and observed mass (O) values are shown in the right panel of the spectra. † indicates the potassium adduct of heptapeptide-1, ‡ indicates the potassium adduct of heptapeptide-2, ¶ indicates heptapeptide-6 whose azide group was reduced to a primary amino group presumably by thiols present in the FIT system[144], *species also present in the "no template" translation product and were likely derived from the translation mixture. The data shown in Figure 37b and c were generated from a sample of the end product of translation reaction.

3.3. Conclusion

In this chapter, I developed the method to prepare aa-tRNAs-NNA using compensatory mutated flexizymes. Then, I measured the translation activity of PTC-mutated ribosome–tRNA pair and developed the orthogonal ribosome–tRNAs pair.

First, CCA-3′ end mutated tRNAs were aminoacylated with various amino acids by compensatory mutated flexizymes. In addition to C74N/C75N mutants, A76N mutant tRNA also can be aminoacylated by prototype of flexizymes "pre-24"[40,58], which indicates that also compensatory mutated flexizymes (eFx, dFx and aFx) can aminoacylate A76N mutant tRNAs. From other groups, an efficient method to prepare homogeneous aa-tRNAs using flexizyme was recently reported[145]. These results indicate that wild-type and compensatory mutated flexizymes have great potential to charge various amino acids onto any RNA substrates bearing NNN-3′ end.

Next, translation activity of several mutant ribosome–tRNAs pairs were measured. To my knowledge, this is the first report about the whole translation activity of CCA-3′ end mutated tRNAs and compensatory mutated ribosome although peptidyl transfer[64] and translocation[63] of those were already reported. Compensatory mutation of 23S rRNA in the PTC restored the whole translation activity, which is consistent with the results of peptidyl transfer[64] and translocation[63]. However, compensatory mutations did not fully restore the activity of some mutant ribosome–tRNAs pairs. To fully restore the translation activity, more engineering of translation machinery including other positions of rRNAs and translation factors will be required.

Finally, orthogonal mutant ribosome–tRNAs functioning independently to wild-type ribosome–tRNAs according to artificially designed genetic code. Combination of wild-type and orthogonal ribosome–tRNAs pair can produce two different peptides from one mRNA. Unfortunately, the efficiency of translation was decreased when both cognate and non-cognate tRNA were present with ribosomes. To increase the translation efficiency, the more detailed interaction of CCA-3′ end with ribosomes and translation factors should be investigated and more engineering of translation machinery will be required.

The study in this chapter 3 established a novel method for genetic code reprogramming, and also demonstrated the importance of interactions between the rRNA and the tRNAs in translation. These results open an entrance to a new opportunity of *in vitro* synthetic biology involving the engineering of the genetic codes and translation machineries.

3.4. Materials and methods

The experiments in this chapter were conducted mainly following the methods in the previous report[65].

Chemical synthesis of L-tyrosine cyanomethyl ester (Tyr-CME), L-aspartic acid 3,5-dinitrobenzyl ester (Asp-DBE), L-lysine 3,5-dinitrobenzyl ester (Lys-DBE), L-acetyllysine 3,5-dinitrobenzyl ester (Aly-DBE), and L-azidonorvaline 3,5-dinitrobenzyl ester (Anv-DBE)

Tyr-CME, Asp-DBE, Lys-DBE, Aly-DBE and Anv-DBE were synthesized as previously described[41,144].

Chemical synthesis of *N*-(5-FAM)-L-phenylalanine-cyanomethyl ester (Fph-CME)

L-phenylalanine-cyanomethyl ester (Phe-CME) was synthesized as previously described[40]. Triethylamine (22 μL, 160 μmol) was added to a mixture of Phe-CME (14.4 mg, 60 μmol) and 5-carboxyfluorescein succinimidyl ester (Thermo Fisher Scientific Inc., 19 mg, 40 μmol) in 500 μL of *N,N*-dimethylformamide and then the mixture was stirred at room temperature for 13.5 h. After the reaction, ethyl acetate was added, and the solution was treated with 1 M HCl (3 mL × 3) and saturated NaCl (3 mL × 1). The organic layer was dried over Na_2SO_4, and concentrated under reduced pressure. The crude residue was dissolved in methanol and injected into a high-pressure liquid chromatography (HPLC) system equipped with a Cadenza 5CD-18 reverse-phase chromatography column (Code 5CD0Q6). The column was equilibrated with 20% (v/v) acetonitrile in H_2O containing 0.1% (v/v) TFA, and eluted with a 2.2%/min gradient of 99.9% (v/v) acetonitrile in 0.1% (v/v) TFA to 90% (v/v) acetonitrile. The HPLC profile was monitored by measuring the absorbance at 494 nm. The fractions of Fph-CME were lyophilized. The concentration was determined by absorbance at 494 nm in 500 mM Tris-HCl (pH 8.0).

Fph-CME: ^{1}H NMR (300 MHz, DMSO-d_6, δ): 10.16 (br, 2H), 9.36 (d, J = 7.5 Hz, 1H), 8.45 (d, J = 0.6 Hz, 1H), 8.19 (dd, J = 8.1, 1.5 Hz, 1H), 7.39 (d, J = 8.1 Hz, 1H), 7.36–7.20 (m, 5H), 6.68 (d, J = 2.1 Hz, 2H), 6.61–6.52 (m, 4H), 5.05 (s, 1H), 4.87–4.80 (m, 1H), 3.33–3.14 (m, 2H); ^{13}C NMR (125 MHz, DMSO-d_6, δ): 171.0, 168.5, 165.4, 160.0, 155.6, 152.2, 137.6, 135.4, 135.3, 129.7, 129.5, 128.8, 127.1, 127.0, 124.9, 123.9, 116.2, 113.1, 109.3, 102.7, 83.8, 54.5, 50.0, 36.2.

The high-resolution mass spectrum (HRMS) was determined using flow injection

(direct electrospray ionization (ESI)-MS) in positive mode (Thermo Exactive). Calculated [$(M+H)^+$] *m/z* for $C_{32}H_{23}N_2O_8^+$ = 563.1449, found = 563.1445.

In vitro transcription

The tRNAs, eFxs, and dFxs were prepared using run-off in vitro transcription with T7 RNA polymerase. 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 polymerase[146]. The primers for preparing transcription templates are shown in Table 9. All non-methylated primers were purchased from Eurofins Genomics K.K. (Japan) and methylated primers were purchased from Gene Design Inc. (Japan).

Expression and purification of maltose binding protein (MBP) -MS2-His_6

The plasmid pMAL-c2g (Amp^R) encoding His_6-tagged MS2 coat protein (a gift from R. Green, Johns Hopkins University) was used to transform *E. coli* BL21-(DE3) in LB medium with 100 μg/mL carbenicillin. A single colony isolated by streaking was used to inoculate a 30 mL overnight culture in LB supplemented with 100 μg/mL carbenicillin, and 10 mL of this culture was used to inoculate one liter of the same medium. The culture was grown to an OD_{600} of 0.7 at 37°C. The expression of MBP-MS2-His_6 was induced using 1 mM isopropyl β-D-1-thiogalactopyranoside, and the cells were cultured for another 4 h at 37°C. The culture was then centrifuged for 10 min at 4,000× *g*, the pelleted cells were resuspended in 30 mL of lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, 6 mM β-mercaptoethanol (β-ME), 1 mM phenylmethanesulfonyl fluoride, adjusted to pH 8.0 with NaOH), and were then sonicated. The lysate was clarified twice by centrifugation for 20 min at 15,000× *g* (4°C) and filtered through a Millex-LH 0.45 μm filter unit (Merck Millipore). MBP-MS2-His_6 was purified using a fast protein liquid chromatography (FPLC) system (AKTA Avant, GE Healthcare) with a His-Trap HP column (GE Healthcare). The column was washed with NiNTA buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10mM imidazole, 6 mM β-ME, adjusted to pH 8.0 with NaOH) and the protein was eluted with NiNTA buffer containing 250 mM imidazole. Protein-containing fractions were pooled and concentrated using a membrane filter (Amicon Ultra15, 3,000 MWCO, Merck Millipore). The concentrated protein was dialyzed against 20 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20% glycerol, and 1 mM dithiothreitol (DTT). The protein concentration was determined by absorbance at 280 nm, assuming an extinction coefficient of 84,800 cm^{-1} M^{-1} (calculated on: http://web.expasy.org/protparam/).

Expression and purification of tagged ribosomes

The plasmid pcI^{857} (Kan^{R}) encoding a temperature-sensitive mutant of the lambda repressor protein cI (a gift from R. Green, Johns Hopkins University)[140] was used to transform *E. coli* DH5α in LB medium containing 50 μg/mL kanamycin. Any 23S rRNA mutation of interest was introduced into the plasmid p278 MS2 encoding 23S rRNA tagged with the MS2 stem-loop (a gift from R. Green, Johns Hopkins University) by site directed mutagenesis using the Quickchange Lightning Site-Directed Mutagenesis kit (Agilent Technologies). The primers for mutagenesis are shown in Table 9. The mutations were confirmed by sequencing. The mutated plasmid p278 MS2 was used to transform DH5α with pcI^{857} in LB medium containing 100 μg/mL ampicillin and 50 μg/mL kanamycin. A colony isolated by streaking was used to inoculate a 50 mL overnight culture grown in LB supplemented with 100 μg/mL ampicillin and 50 μg/mL kanamycin at 30°C. One liter of this medium was inoculated with 20 mL of an overnight culture and grown to an OD_{600} of 0.7–0.9 at 42°C. The cultures were centrifuged for 10 min at 4,000× *g* and the pelleted cells were resuspended in 5.5 mL ribosome buffer A (20 mM Tris-HCl, pH 7.5, 100 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, and 6 mM β-ME). Lysozyme (Nacalai Tesque Inc.) was added (final concentration, 0.1 mg/mL) and incubated at 4°C for 30 min. This solution was disrupted twice using a cell disruption vessel (Parr Instrument Company). The lysate was clarified twice by centrifugation for 15 min at 15,000× *g* (4°C), layered onto 10 mL of ribosome buffer D (20 mM Tris-HCl, pH 7.5, 500 mM NH_4Cl, 10 mM $MgCl_2$, 0.5 mM EDTA, 1.1 M sucrose) in a 25PC tube (Hitachi), and centrifuged at 104,000× *g* in an S50A rotor (Hitachi) for 18–20 h at 4°C. Ribosome buffer A was used to rinse and then dissolve the pellet, which was then stored at −80°C.

Tagged ribosomes were purified using FPLC (AKTA Avant, GE Healthcare). First, 2 mg MBP-MS2-His_6 was loaded on an MBP Trap HP column (GE Healthcare), and the bound protein was washed with ribosome binding buffer (20 mM Tris-HCl, pH 7.5, 100 mM NH_4Cl, and 10 mM $MgCl_2$). Crude ribosomes were loaded onto the column and washed with 25 mL of ribosome-binding buffer. Tagged ribosomes bound to MBP-MS2-His_6 were eluted with 20 mL of ribosome-elution buffer (20 mM Tris-HCl, pH 7.5, 100 mM NH_4Cl, 10 mM $MgCl_2$, and 10 mM maltose). Purified tagged ribosomes were concentrated to 10–20 μM, and the buffer was exchanged with RE buffer (20 mM HEPES-KOH, pH 7.6, 10 mM $MgCl_2$, 50 mM KCl, 1 mM DTT) using membrane filtration (Amicon Ultra15, 100,000 MWCO, Merck Millipore). The concentration of 70S ribosomes was determined by absorbance at 260 nm using the conversion factor 1 A_{260} = 23 pmol.

Aminoacylation by flexizymes

Aminoacylation reactions of Tyr-CME, Asp-DBE, Lys-DBE, Aly-DBE, and Anv-DBE were performed as previously described[41]. The reaction times were changed to 30 min (Tyr-CME), 9 hours (Asp-DBE), and 3 hours (Lys-DBE, Aly-DBE and Anv-DBE). Aminoacylation of Fph-CME was performed as follows: A mixture of 3 μL of 41.7 μM $tRNA^{fMet}_{CAU}$ and 41.7 μM eFx in 83.3 mM HEPES-KOH (pH 7.5) was heated at 95°C for 2 min and cooled to room temperature over 5 min. One microliter of 3 M $MgCl_2$ was added, and the mixture was transferred to an ice bath, 1 μL of 5 mM Fph-CME was added, and the mixture was incubated on ice for 16 h. After the reaction, aminoacyl-tRNA (aa-tRNA) was precipitated with ethanol as previously described[41].

Analysis of acylation

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 10 V cm^{-1}) for 21 h. The gels were stained with ethidium bromide and analyzed using an FLA-5100 fluorescent image analyzer (Fujifilm) or PharosFX molecular imager (BIO-RAD). Aminoacylation efficiency was calculated according to the band intensity of aa-tRNA (A) and free tRNA (T) and is presented as (A)/[(A) + (T)]. Because of the fluorescence of Fph-tRNA, the efficiency for Fph-CME was calculated using the expression $1-[(T_{aa+})(F_{aa-})]/[(T_{aa-})(F_{aa+})]$ according to the band intensity of flexizyme in the lane without amino acids (F_{aa-}), flexizyme in the lane with amino acids (F_{aa+}), free tRNA in the lane without amino acids (T_{aa-}), and free tRNA in the lane with amino acids (T_{aa+}). Values reported are the average of three independent reactions, and error bars represent the standard deviation.

DNA templates for translation

The DNA templates for translation were prepared using the polymerase chain reaction (PCR), and the amplicons were extracted with phenol-chloroform mixture and precipitated with ethanol. The DNA templates were purified using 12% native-PAGE, and the concentrations were determined by absorbance at 260 nm. The primers for preparing DNA templates for translation are shown in Table 9.

***In vitro* translation using the FIT system**

The FIT system comprised a mixture of all desired components for translation[126]. The tagged-ribosomes and aa-tRNAs were added as necessary in each experiment. The composition of the FIT system was as follows: 50 mM HEPES-KOH (pH 7.6), 12 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, 0.6 μM MTF, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 0.26 μM EF-G, 10 μM EF-Tu, 10 μM EF-Ts, 0.25 μM RF1, 0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 0.1 μM T7 RNA polymerase, 4 μg/mL creatine kinase, 3 μg/mL myokinase, 0.1 μM pyrophosphatase, 0.1 μM nucleotide-diphosphatase kinase, and 400 nM DNA template. The concentration of EF-Tu and EF-Ts only in the experiment shown in Figure 35 was increased to 20μM. Translation reactions were started by adding aa-tRNAs at 37°C after a 5 min incubation. Translation reactions were terminated by adding an equal volume of 2×Tricine-SDS-PAGE loading buffer (900 mM Tris-HCl (pH 8.45), 8% (w/v) SDS, 30% (v/v) glycerol).

In the experiment shown in Figure 29 and 32, the concentration of ribosomes, Fph-tRNA$^{fMet}_{CAU,}$ Lys-tRNA$^{GluE2}_{CUU}$, and Tyr-tRNA$^{GluE2}_{GUA}$ were 2 μM, 5 μM, 20 μM, 10 μM, respectively. The sequence of the DNA template was 5′-GGCGT AATAC GACTC ACTAT AGGGC TTTAA TAAGG AGAAA AACAT GAAGT ACAAG AAGTA CAAGT GAGCT TCG-3′.

In the experiment shown in Figure 34 and 35, the concentration of the wild-type ribosomes, G2251C/G2553C ribosomes, Fph-tRNA$^{fMet}_{CAU}$-CCA, Fph-tRNA$^{fMet}_{CAU}$-CGA, Lys-tRNA$^{Lys}_{CUU}$-CCA, Lys-tRNA$^{Lys}_{CUU}$-CGA, Asp-tRNA$^{Asp}_{GUC}$-CCA, Aly-tRNA$^{GluE2}_{GUC}$-CGA, Tyr-tRNA$^{Tyr}_{GUA}$-CCA, and Anv-tRNA$^{GluE2}_{GUA}$-CGA were 0.1 μM, 2 μM, 5 μM, 1.5 μM, 10 μM, 3 μM, 10 μM, 3 μM, 10 μM and 3 μM, respectively. The sequence of the DNA template was 5′-GGCGT AATAC GACTC ACTAT AGGGC TTTAA TAAGG AGAAA AACAT GAAGT ACGAC AAGTA CGACT GAGCT TCG-3′.

In the experiment shown in Figure 28, the concentration of wild-type ribosomes, Fph-tRNA$^{fMet}_{CAU,}$ Lys-tRNA$^{AsnE2}_{CUU}$, and Tyr-tRNA$^{AsnE2}_{GUA}$ were 2 μM, 5 μM, 20 μM, 10 μM, respectively. The sequence of the DNA template was 5′-GGCGT AATAC GACTC ACTAT AGGGC TTTAA TAAGG AGAAA AACAT GAAGT ACAAG AAGTA CAAGT GAGCT TCG-3′.

In the experiment shown in Figure 36, the concentration of the wild-type ribosomes, G2251C/G2553C ribosomes, Fph-tRNA$^{fMet}_{CAU}$-CCA, Fph-tRNA$^{fMet}_{CAU}$-

CGA, Lys-tRNA$^{Lys}_{CUU}$-CCA, Lys-tRNA$^{Lys}_{CUU}$-CGA, Asp-tRNA$^{Asp}_{GUC}$-CCA, and Aly-tRNA$^{GluE2}_{GUC}$-CGA were 0.3 μM, 2 μM, 5 μM, 1.5 μM, 20 μM, 6 μM, 10 μM, 3 μM, respectively. The sequence of the DNA template was 5′-GGCGT AATAC GACTC ACTAT AGGGC TTTAA TAAGG AGAAA AACAT GAAGG ACAAG AAGGA CAAGT GAGCT TCG-3′.

In the experiment shown in Figure 37, the concentration of the wild-type ribosomes, G2251C/G2553C ribosomes, Fph-tRNA$^{fMet}_{CAU}$-CCA, Fph-tRNA$^{fMet}_{CAU}$-CGA, Lys-tRNA$^{Lys}_{CUU}$-CCA, Lys-tRNA$^{Lys}_{CUU}$-CGA, Tyr-tRNA$^{Tyr}_{GUA}$-CCA, and Anv-tRNA$^{AsnE2}_{GUA}$-CGA were 0.3 μM, 2 μM, 5 μM, 1.5 μM, 20 μM, 6 μM, 10 μM, 3 μM, respectively. The sequence of the DNA template was 5′-GGCGT AATAC GACTC ACTAT AGGGC TTTAA TAAGG AGAAA AACAT GAAGT ACAAG AAGTA CAAGT GAGCT TCG-3′.

The aa-tRNAs were prepared by aminoacylation of tRNAs using cognate flexizymes. The concentration of aa-tRNA was calculated from acylation efficiency shown in Figure 25.

MALDI-TOF analysis

After incubation, the *in vitro* translation mixtures were desalted using a *C*18-tip (Nikkyo Technos) and analyzed using MALDI-TOF mass in linear positive modes. All MALDI-TOF analysis was performed using an Autoflex II (Bruker Daltonics) or ultrafleXtreme (Bruker Daltonics) with external calibration (Peptide Calibration Standard II, Bruker Daltonics).

Tricine-SDS-PAGE analysis of translation products

The quantity of Fph incorporated into peptides was determined using Tricine-SDS-PAGE as previously described[126]. The concentration of acrylamide in separation gel was 12% for the experiment shown in Figure 28, 29 and 32, 19% for Figure 34 and 20% for Figure 36. The fluorescence of the peptides was determined using the FLA-5100 (Fujifilim) or PharosFX molecular imager (BIO-RAD), and the data were fit to an exponential curve using KaleidaGraph (Synergy Software). The amounts of peptides were quantified according to that of a fluorescent band generated by a known concentration (0 μM to 4 μM) of Fph. Values reported are the average of three independent reactions, and error bars represent the standard deviation.

Quantification of the abundance of tagged ribosome by primer extension

Primer extension analysis was performed mainly following the previous report[140]. Primer extension using a primer that is complementary to the bases 2254-2273 in 23S rRNA produced a dideoxyguanosine stop at +2 for tagged G2252C , +3 for tagged G2251C, or +6 for untagged wild-type ribosomes, or a dideoxythymidine stop at +3 for tagged G2251A or +7 for untagged wild-type ribosomes. Primer extension using a primer that is complementary to the bases 2556-2575 in 23S rRNA produced a dideoxyguanosine stop at +3 for tagged G2553C or +5 for untagged wild-type ribosomes, or a dideoxythymidine stop at +3 for tagged G2553A or +9 for untagged wild-type ribosomes. Ribosomal RNAs were extracted from purified ribosomes by phenol-chloroform extraction and ethanol precipitation. Primer extension was carried out as follows: A mixture of 7 μL of 57.1 nM rRNA, 171.4 nM [^{32}P] 5′-end labeled primer, 7.14 mM each dATP/dCTP/dTTP/ddGTP or dATP/dGTP/dCTP/ddTTP was heated at 65°C for 5 min and cooled on ice 1 min. Two microliter of 5× FS buffer (Thermo Fisher Scientific Inc.), 0.5 μL of 0.1 M DTT and 0.5 μL of SuperScript™ III (Thermo Fisher Scientific Inc.) were added, and the mixture were incubated at 55°C for 1 hour. Primer extension products were resolved on a 15% denaturing polyacrylamide gel (8M urea, 1×TBE) and the intensities of the bands were quantified using FLA-5100 (Fujifilim).

Gel-shift analysis of ternary complex of EF-Tu, GTP and aa-tRNA

Binding of EF-Tu to aa-tRNAs was investigated by following previous report[147]. First, 10 μM EF-Tu was preincubated with 1 mM GTP at 37°C for 15 min in 5 μL total volume containing 70 mM HEPES-KOH (pH 7.6), 52 mM NH_4OAc, 8 mM $Mg(OAc)_2$, 30 mM KCl, 0.8 mM DTT, 1.6 μM GDP, 6% glycerol, 10 mM phosphoenolpyruvate, and 0.08 U/μL pyruvate kinase. After aminoacylation reaction by flexizyme, the ethanol precipitated mixtures which contain flexizymes, aa-tRNAs and tRNAs were resuspended in 6 mM KOAc at 8.3 μM final concentration of aa-tRNA. The concentrations of aa-tRNAs were calculated from acylaton efficiency shown in Figure 25. This 3 μL aa-tRNA solution and 2 μL of ternary complex buffer, containing 150 mM HEPES-KOH (pH 7.6), 195 mM NH_4OAc, and 30 mM $Mg(OAc)_2$ were added to the preincubated EF-Tu solution. The mixture was incubated at 37°C for 10 min. Native-PAGE was performed using 8% polyacrylamide gels at 4°C in a running buffer containing 50 mM Tris-HCl (pH 6.8), 65 mM NH_4OAc, and 10 mM $Mg(OAc)_2$. Gels were stained by SimplyBlue™ SafeStain (Thermo Fisher Scientific Inc.). Intensities of bands were analyzed by Multi Gauge (Fujifilim).

Table 9 | Primers used in the chapter 3.

Sequence (5′ to 3′)	used for
GGCGTAATACGACTCACTATAG	dFx, eFx, tRNA, DNA template
GTAATACGACTCACTATAGGATCGAAAGATTTCCGC	dFx-NNU and eFx-NNU
AACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCC	dFx-NNU
AACGCTAATCCCCTTTCGGGGCCGCGGAAATCTTTCGATCC	eFx-NNU
AC(M)CTAACGCCATGTACCCTTTCGGG	dFx-GGU
AC(M)CTAACGCTAATCCCCTTTCGGG	eFx-GGU
AG(M)CTAACGCCATGTACCCTTTCGGG	dFx-GCU
AG(M)CTAACGCTAATCCCCTTTCGGG	eFx-GCU
AC(M)GTAACGCCATGTACCCTTTCGGG	dFx-CGU
AC(M)GTAACGCTAATCCCCTTTCGGG	eFx-CGU
AG(M)GTAACGCCATGTACCCTTTCGGG	dFx-GGU
AG(M)GTAACGCTAATCCCCTTTCGGG	eFx-GGU
AC(M)TTAACGCCATGTACCCTTTCGGG	dFx-AGU
AC(M)TTAACGCTAATCCCCTTTCGGG	eFx-AGU
GTAATACGACTCACTATACGCGGGGTGGAGCAGCCTGGTAGCTCGTCGG	$tRNA^{fMet}_{CAU}$-NNA
GAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCT	$tRNA^{fMet}_{CAU}$-NNA
TTGCGGGGGCCGGATTTGAACCGACGATCTTCGGG	$tRNA^{fMet}_{CAU}$-NNA
GGCGTAATACGACTCACTATAC	$tRNA^{fMet}_{CAU}$-NNA
TG(M)GTTGCGGGGGCCGGATTTG	$tRNA^{fMet}_{CAU}$-CCA
TG(M)CTTGCGGGGGCCGGATTTG	$tRNA^{fMet}_{CAU}$ -GCA
TC(M)GTTGCGGGGGCCGGATTTG	$tRNA^{fMet}_{CAU}$ -CGA
TC(M)CTTGCGGGGGCCGGATTTG	$tRNA^{fMet}_{CAU}$ -GGA
TA(M)GTTGCGGGGGCCGGATTTG	$tRNA^{fMet}_{CAU}$ -CUA
GGCGTAATACGACTCACTATAGTCCCCTTCGTCTAGA	$tRNA^{GluE2}_{NNN}$-NNA
CGTCCCCTAGGGGATTCGAACCCCTGTTACCGCC	$tRNA^{GluE2}_{NNN}$-NNA
TATAGTCCCCTTCGTCTAGAGGCCCAGGACACCGCCCTCT	$tRNA^{GluE2}_{CUU}$-NNA
GAACCCCTGTTACCGCCTTAAGAGGGCGGTGTCCTGG	$tRNA^{GluE2}_{CUU}$-NNA
TATAGTCCCCTTCGTCTAGAGGCCCAGGACACCGCCCTGT	$tRNA^{GluE2}_{GUA}$-NNA
GAACCCCTGTTACCGCCTTTACAGGGCGGTGTCCTGG	$tRNA^{GluE2}_{GUA}$-NNA
TATAGTCCCCTTCGTCTAGAGGCCCAGGACACCGCCTTGT	$tRNA^{GluE2}_{GUC}$-NNA
GAACCCCTGTTACCGCCTTGACAAGGCGGTGTCCTGG	$tRNA^{GluE2}_{GUC}$-NNA
TG(M)GCGTCCCCTAGGGGATTC	$tRNA^{GluE2}_{NNN}$-CCA
TG(M)CCGTCCCCTAGGGGATTC	$tRNA^{GluE2}_{NNN}$-GCA

Sequence (5′ to 3′)	used for
TC(M)GCGTCCCCTAGGGGATTC	$tRNA^{GluE2}_{NNN}$-CGA
TC(M)CCGTCCCCTAGGGGATTC	$tRNA^{GluE2}_{NNN}$-GGA
TA(M)GCGTCCCCTAGGGGATTC	$tRNA^{GluE2}_{NNN}$-CUA
TATAGGGTCGTTAGCTCAGTTGGTAGAGCAGTTGACTCTTAATC	$tRNA^{Lys}_{CUU}$-NNA
GGCGTAATACGACTCACTATAGGGTCGTTAGCTCAGTTG	$tRNA^{Lys}_{CUU}$-NNA
TTCGAACCTGCGACCAATTGATTAAGAGTCAACTGCTCTACC	$tRNA^{Lys}_{CUU}$-NNA
TGGGTCGTGCAGGATTCGAACCTGCGACCAATTGATT	$tRNA^{Lys}_{CUU}$-NNA
TG(M)GTGGGTCGTGCAGGATTCG	$tRNA^{Lys}_{CUU}$-CCA
TC(M)GTGGGTCGTGCAGGATTCG	$tRNA^{Lys}_{CUU}$-CGA
TATAGGAGCGGTAGTTCAGTCGGTTAGAATACCTGCCTGT	$tRNA^{Asp}_{GUC}$-NNA
GGCGTAATACGACTCACTATAGGAGCGGTAGTTCAGTC	$tRNA^{Asp}_{GUC}$-NNA
GAACCCGCGACCCCCTGCGTGACAGGCAGGTATTCTAACCG	$tRNA^{Asp}_{GUC}$-NNA
CGGAACGGACGGGACTCGAACCCGCGACCCCC	$tRNA^{Asp}_{GUC}$-NNA
TG(M)GCGGAACGGACGGGACT	$tRNA^{Asp}_{GUC}$-CCA
AGGTGGGGTTCCCGAGCGGCCAAAGGGAGCAGACTGTAAATCT	$tRNA^{Tyr}_{GUA}$-NNA
GGCGTAATACGACTCACTATAGGTGGGGTTCCCGAG	$tRNA^{Tyr}_{GUA}$-NNA
GAACCTTCGAAGTCTGTGACGGCAGATTTACAGTCTGCTCCCT	$tRNA^{Tyr}_{GUA}$-NNA
TGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGA	$tRNA^{Tyr}_{GUA}$-NNA
TG(M)GTGGTGGGGGAAGGATTCG	$tRNA^{Tyr}_{GUA}$-CCA
GTAATACGACTCACTATAGGCTCTGTAGTTCAGTCGGTAGAACGGCGGA	$tRNA^{AsnE2}_{NNN}$-NNA
CGGCTCTGACTGGACTCGAACCAGTGACATACGGA	$tRNA^{AsnE2}_{NNN}$-NNA
GAACCAGTGACATACGGATTAAGAGTCCGCCGTTCTACCGACT	$tRNA^{AsnE2}_{CUU}$-NNA
GAACCAGTGACATACGGATTTACAGTCCGCCGTTCTACCGACT	$tRNA^{AsnE2}_{GUA}$-NNA
GAACCAGTGACATACGGATTGACAGTCCGCCGTTCTACCGACT	$tRNA^{AsnE2}_{GUC}$-NNA
TG(M)GCGGCTCTGACTGGACTC	$tRNA^{AsnE2}_{NNN}$-CCA
TG(M)CCGGCTCTGACTGGACTC	$tRNA^{AsnE2}_{NNN}$-GCA
TC(M)GCGGCTCTGACTGGACTC	$tRNA^{AsnE2}_{NNN}$-CGA
TC(M)CCGGCTCTGACTGGACTC	$tRNA^{AsnE2}_{NNN}$-GGA
TA(M)GCGGCTCTGACTGGACTC	$tRNA^{AsnE2}_{NNN}$-CUA
ATACGACTCACTATAGGGCTTTAATAAGGAGAAAAACATG	DNA template
GGCGTAATACGACTCACTATAGGGCTTT	DNA template
GTACTTCTTGTACTTCATGTTTTTCTCCTTATTAAAGCC	DNA template (Heptapeptide-1)
GTACTTGTCGTACTTCATGTTTTTCTCCTTATTAAAGCC	DNA template (Heptapeptide-2, 3)
GTCCTTCTTGTCCTTCATGTTTTTCTCCTTATTAAAGCC	DNA template (Heptapeptide-4, 5)

Sequence (5′ to 3′)	used for
CGAAGCTCACTTGTACTTCTTGTACTTCATGTTTTTC	DNA template (Heptapeptide-1)
CGAAGCTCAGTCGTACTTGTCGTACTTCATGTTTTTC	DNA template (Heptapeptide-2, 3)
CGAAGCTCACTTGTCCTTCTTGTCCTTCATGTTTTTC	DNA template (Heptapeptide-4, 5)
GGTGGGTAGTTTGACTGCGGCGGTCTCCTCCTAAAGAG	Quick Change (G2251C)
CTCTTTAGGAGGAGACCGCCGCAGTCAAACTACCCACC	Quick Change (G2251C)
GGTGGGTAGTTTGACTGGCGCGGTCTCCTCCTAAAGAG	Quick Change (G2252C)
CTCTTTAGGAGGAGACCGCGCCAGTCAAACTACCCACC	Quick Change (G2252C)
GGTGGGTAGTTTGACTGCCGCGGTCTCCTCCTAAAGAG	Quick Change (G2251C/G2252C)
CTCTTTAGGAGGAGACCGCGGCAGTCAAACTACCCACC	Quick Change (G2251C/G2252C)
GGTGGGTAGTTTGACTGAGGCGGTCTCCTCCTAAAGAG	Quick Change (G2251A)
CTCTTTAGGAGGAGACCGCCTCAGTCAAACTACCCACC	Quick Change (G2251A)
GTCCCAAGGGTATGGCTCTTCGCCATTTAAAGTGGTAC	Quick Change (G2553C)
GTACCACTTTAAATGGCGAAGAGCCATACCCTTGGGAC	Quick Change (G2553C)
GTCCCAAGGGTATGGCTATTCGCCATTTAAAGTGGTAC	Quick Change (G2553A)
GTACCACTTTAAATGGCGAATAGCCATACCCTTGGGAC	Quick Change (G2553A)
TACTCTTTAGGAGGAGACCG	Primer extension (2251, 2252)
GCGTACCACTTTAAATGGCG	Primer extension (2553)
CCACCCTTTAATGTTTGATGTTC	Sequencing (2251, 2252)
CAGTCAAGCTGGCTTATGC	Sequencing (2251, 2252)
CATATCGACGGCGGTGTTTG	Sequencing (2553)
GGCAGATAGGGACCGAAC	Sequencing (2553)

Table 10 | RNAs prepared in the chapter 3. N means A/U/G/C, italic types indicate ribonucleotides.

Name	Sequence (5′ to 3′)
eFx-NNU	*GGAUCGAAAGAUUUCCGCGGCCCCGAAAGGGGAUUAGCGUUANNU*
dFx-NNU	*GGAUCGAAAGAUUUCCGCAUCCCCGAAAGGGUACAUGGCGUUANNU*
$tRNA^{fMet}_{CAU}$-NNA	*CGCGGGGUGGAGCAGCCUGGUAGCUCGUCGGGCUCAUAACCCGAAGAUCGUCGGUUCAAAU CCGGCCCCCGCAANNA*
$tRNA^{EnAsn}_{CUU}$-NNA	*GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACUCUUAAUCCGUAUGUCACUGGUUCGAGU CCAGUCAGAGCCGNNA*
$tRNA^{EnAsn}_{GUA}$-NNA	*GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACUGUAAAUCCGUAUGUCACUGGUUCGAGU CCAGUCAGAGCCGNNA*
$tRNA^{EnAsn}_{GUC}$-NNA	*GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUUGUCAAUCCGUAUGUCACUGGUUCGAGU CCAGUCAGAGCCGNNA*
$tRNA^{EnGlu}_{CUU}$-NNA	*GUCCCCUUCGUCUAGAGGCCCAGGACACCGCCCUCUUAAGGCGGUAACAGGGGUUCGAAUC CCCUAGGGGACGNNA*
$tRNA^{EnGlu}_{GUA}$-NNA	*GUCCCCUUCGUCUAGAGGCCCAGGACACCGCCCUGUAAAGGCGGUAACAGGGGUUCGAAUC CCCUAGGGGACGNNA*
$tRNA^{EnGlu}_{GUC}$-NNA	*GUCCCCUUCGUCUAGAGGCCCAGGACACCGCCUUGACAAGGCGGUAACAGGGGUUCGAAUC CCCUAGGGGACGNNA*
$tRNA^{Lys}_{CUU}$-NNA	*GGGUCGUUAGCUCAGUUGGUAGAGCAGUUGACUCUUAAUCAAUUGGUCGCAGGUUCGAAU CCUGCACGACCCANNA*
$tRNA^{Tyr}_{GUA}$-NNA	*GGUGGGGUUCCCGAGCGGCCAAAGGGAGCAGACUGUAAAUCUGCCGUCACAGACUUCGAAG GUUCGAAUCCUUCCCCCACCANNA*
$tRNA^{Asp}_{GUC}$-NNA	*GGAGCGGUAGUUCAGUCGGUUAGAAUACCUGCCUGUCACGCAGGGGGUCGCGGGUUCGAG UCCCGUCCGUUCCGNNA*

Chapter 4

General conclusion

In natural, each ARS specifically recognizes each tRNA and amino acid, and tRNAs are specifically aminoacylated with cognate amino acids by cognate ARSs. This rigorously-defined correspondence of tRNAs to amino acids defines genetic code, which is essential for accurate translation of mRNAs to proteins in living organism. However, this specific aminoacylation reaction is not favored in the case of application of aminoacylation to other studies or engineering of translation machinery, because it is difficult to arbitrarily aminoacylate RNAs with various amino acids. Flexizymes are *in vitro* selected aminoacylation ribozymes which have two unique characteristics; (i) using various substrates including both natural amino acids and ncAAs. (ii) recognizing substrate RNA via two consecutive base pairs, which can be substituted with other base pairs[65]. These characteristics enable to flexibly aminoacylate RNAs bearing various 3′ ends with various amino acids including both canonical and noncanonical amino acids.

In the study described in chapter 2, novel interaction between folic acid and hsa-pre-miR125a was discovered using flexizymes. This is the first example of interaction between metabolite and microRNA precursors, which suggests that more RNA aptamer elements exist in microRNA precursors. The same procedure will be usable to discover other small RNAs interacting with small molecules. MicroRNAs have been extensively studied as therapeutic targets and some small molecule inhibitors have been developed targeting microRNA precursors[148,149]. Because miR125a-5p is antioncogene, folic acid have a potential as a small molecule drug which binds to miR125a precursors and regulates the processing of miR-125a. In addition, folic acid derivatives will be also drug candidates targeting miR-125a precursors. Recently, macrocyclic peptides armed with a mechanism-based warhead inhibiting human deacetylase SIRT2 was developed[150]. If folic acid moiety could be incorporated into peptide as warhead, it is possible to develop peptide drug targeting miR-125a precursors. As described above, the discovery of new interaction between miRNAs and metabolite have not only scientific importance but also essential information to develop drugs targeting miRNAs.

Flexizymes specifically aminoacylated tRNAs in many small RNAs. This result demonstrated that aminoacylation by flexizymes is usable technology to easily label and deplete endogenous tRNAs with much abundance, and to discover new small ncRNAs whose length is similar to tRNAs. In addition, this technology can be used for tRNAs of any species even if the sequences were unknown because flexizymes only recognizes CCA-3′ end which is universally conserved in all organisms.

On the other hand, this methodology can be applied not only for depleting but also for recovering specific RNAs. Recently several small ncRNAs bearing CCA-3′ end

such as mascRNA[151], MEN β tRNA-like small RNA[152] and Y RNA[153] are discovered, which are produced by cleavage by RNaseP and CCA addition by CCA-adding enzyme. It has been suggested that CCA addition is related to stability of RNAs[154]. Labeling CCA-3′ end of RNAs by flexizymes will be usable to discover such RNAs bearing CCA-3′ end.

In the study described in chapter 3, compensatory mutated flexizymes aminoacylated tRNAs bearing mutation at the 3′ end. In addition to C74N/C75N mutants, A76N mutant tRNA also can be aminoacylated by prototype of flexizymes "pre-24"[40,58], which suggests that also compensatory mutated flexizymes (eFx, dFx and aFx) can aminoacylate A76N mutant tRNAs. These results indicate that wild-type and compensatory mutated flexizymes have great potential to charge various amino acids onto any tRNAs bearing NNN-3′ end. Therefore A76 of tRNA is also candidate to be mutated for engineering of peptidyl transferase center to develop more active orthogonal translation machinery. However, further engineering of ribosome and translation factors will be necessary because A76 residue is important for formation of aa-tRNA•EF-Tu•GTP ternary complex[141] and translocation of peptidyl-tRNA from A-site to P-site[155].

In the study in chapter 3, only *in vitro* translation activity and orthogonality was demonstrated. If this orthogonal translation machinery will work *in vivo*, it will be possible that engineered proteins containing ncAAs are expressed without affecting the endogenous translation machinery. In order to achieve this, there are problems to be solved. First is the low efficiency of translation by orthogonal ribosome and tRNA pair. In this study short seven polypeptide was expressed but it is necessary to increase translation efficiency in order to produce full-length protein. Second is the misincorporation of aa-tRNAs into non-cognate ribosomes. In the case of FIT system, the amount of translation components including ARS, factors, amino acids, tRNAs and ribosomes can be changed. However, it is very difficult to express the appropriate amount of translation components *in vivo* to avoid the misincorporation. Further engineering of translation components as described above will be necessary to solve these problems. Finally, it is necessary that orthogonal ribosome-tRNA pair translates the specific mRNA encoding target protein in order not to affect the endogenous translation of other proteins. To achieve this, we need to construct orthogonal ribosome-tRNA-mRNA. Orthogonal ribosome-mRNA pair was already reported, but 16S rRNA was engineered in the study[121-123]. Therefore, orthogonal 50S subunit -30S subunit pair will be needed to construct orthogonal ribosome-tRNA-mRNA. As described above, there are many problems to express the orthogonal translation machinery *in vivo*, but these will be solved in future.

From the results in both chapter 2 and 3, flexizymes can aminoacylate any RNAs bearing not only CCA-3′ end but also other 3′ end sequences with various amino acids bearing. In addition, 3′ end of flexizymes will be further engineered, flexizymes will be able to recognize longer 3′ end sequences. Therefore, flexizymes have great potential to deplete or enrich RNAs bearing specific 3′ end sequences.

Until now, flexizymes has been mainly applied for preparing aa-tRNA bearing ncAAs used for genetic code reprogramming to produce non-standard peptides[25]. However, as discussed above, aminoacylation at 3′ end of RNA with various amino acids by flexizymes has a great potential to be applied for RNA studies including analysis of translation mechanisms, engineering of translation, discoveries of novel small ncRNAs. The discovery of interaction between folic acid and hsa-pre-miR125a proposes new regulation mechanism of miRNA biogenesis and the development of orthogonal translation system opens an entrance to a new opportunity of *in vitro* synthetic biology.

References

1 Collins, F. S., Lander, E. S., Rogers, J., Waterston, R. H. & Conso, I. H. G. S. Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-945 (2004).

2 Cheng, J. *et al.* Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* **308**, 1149-1154 (2005).

3 Ambrogelly, A., Palioura, S. & Söll, D. Natural expansion of the genetic code. *Nat. Chem. Biol.* **3**, 29-35 (2006).

4 Cavarelli, J. & Moras, D. Recognition of tRNAs by aminoacyl-tRNA synthetases. *FASEB J.* **7**, 79-86 (1993).

5 Ruff, M. *et al.* Class II aminoacyl transfer RNA synthetases: crystal structure of yeast aspartyl-tRNA synthetase complexed with tRNA(Asp). *Science* **252**, 1682-1689 (1991).

6 Schulman, L. H. & Pelka, H. Structural requirements for aminoacylation of Escherichia coli formylmethionine transfer RNA. *Biochemistry* **16**, 4256-4265 (1977).

7 Liu, M. & Horowitz, J. Functional transfer RNAs with modifications in the 3'-CCA end: differential effects on aminoacylation and polypeptide synthesis. *Proc. Natl. Acad. Sci. U. S. A.* **91**, 10389-10393 (1994).

8 Zhou, X. L. *et al.* Role of tRNA amino acid-accepting end in aminoacylation and its quality control. *Nucleic Acids Res.* **39**, 8857-8868 (2011).

9 Hecht, S. M., Alford, B. L., Kuroda, Y. & Kitano, S. "Chemical aminoacylation" of tRNA's. *J. Biol. Chem.* **253**, 4517-4520 (1978).

10 Lodder, M., Wang, B. X. & Hecht, S. M. The N-pentenoyl protecting group for aminoacyl-tRNAs. *Methods* **36**, 245-251 (2005).

11 Yang, F., Moss, L. G. & Phillips, G. N. The molecular structure of green fluorescent protein. *Nat. Biotechnol.* **14**, 1246-1251 (1996).

12 Marck, C. & Grosjean, H. tRNomics: Analysis of tRNA genes from 50 genomes of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **8**, 1189-1232 (2002).

13 Shi, H. J. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 angstrom resolution: A classic structure revisited. *RNA* **6**, 1091-1105 (2000).

14 Lindberg, J. & Lundeberg, J. The plasticity of the mammalian transcriptome. *Genomics* **95**, 1-6 (2010).

15 Esteller, M. Non-coding RNAs in human disease. *Nat. Rev. Genet.* **12**, 861-874 (2011).

16 Kapranov, P. *et al.* RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484-1488 (2007).

17 Huang, Y. *et al.* Molecular functions of small regulatory noncoding RNA. *Biochem (Mosc)* **78**, 221-230 (2013).

18 Mercer, T. R., Dinger, M. E. & Mattick, J. S. Long non-coding RNAs: insights into functions. *Nat. Rev. Genet.* **10**, 155-159 (2009).

19 Amaral, P. P. & Mattick, J. S. Noncoding RNA in development. *Mamm. Genome* **19**, 454-492 (2008).

20 Matera, A. G., Terns, R. M. & Terns, M. P. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat. Rev. Mol. Cell. Biol.* **8**, 209-220 (2007).

21 Mattick, J. S. & Makunin, I. V. Non-coding RNA. *Hum. Mol. Genet.* **15**, R17-R29 (2006).

22 Snead, N. M. & Rossi, J. J. Biogenesis and function of endogenous and exogenous siRNAs. *Wires RNA* **1**, 117-131 (2010).

23 Kozomara, A. & Griffiths-Jones, S. miRBase: annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **42**, D68-D73 (2014).

24 Suga, H., Hayashi, G. & Terasaka, N. The RNA origin of transfer RNA aminoacylation and beyond. *Phil. Trans. R. Soc. B* **366**, 2959-2964 (2011).

25 Terasaka, N. & Suga, H. Flexizymes-facilitated genetic code reprogramming leading to the discovery of drug-like peptides. *Chem. Lett.* **43**, 11-19 (2014).

26 Kruger, K. *et al.* Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-157 (1982).

27 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).

28 Walter, G. Origin of life: The RNA world. *Nature* **319** (1986).

29 Doudna, J. & Cech, T. The chemical repertoire of natural ribozymes. *Nature* **418**, 222-228 (2002).

30 Ellington, A. & Szostak, J. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).

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 Robertson, D. & Joyce, G. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467-468 (1990).

33 Johnston, W., Unrau, P., Lawrence, M., Glasner, M. & Bartel, D. RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319-1325 (2001).

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

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

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

37 Agresti, J. J., Kelly, B. T., Jaschke, 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).

38 Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643-647 (1995).

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

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

41 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).

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

43 Kumar, R. & Yarus, M. RNA-catalyzed amino acid activation. *Biochemistry* **40**, 6998-7004 (2001).

44 Illangasekare, M., Kovalchuke, O. & Yarus, M. Essential structures of a self-aminoacylating RNA. *J. Mol. Biol.* **274**, 519-529 (1997).

45 Illangasekare, M. & Yarus, M. A tiny RNA that catalyzes both aminoacyl-RNA and peptidyl-RNA synthesis. *RNA* **5**, 1482-1489 (1999).

46 Yarus, M. The meaning of a minuscule ribozyme. *Phil. Trans. R. Soc. B* **366**, 2902-2909 (2011).

47 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).

48 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).

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

50 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).

51 Kawakami, T., Murakami, H. & Suga, H. Ribosomal synthesis of polypeptoids and peptoid-peptide hybrids. *J. Am. Chem. Soc.* **130**, 16861-16863 (2008).

52 Kawakami, T., Ishizawa, T. & Murakami, H. Extensive Reprogramming of the Genetic Code for Genetically Encoded Synthesis of Highly N-Alkylated Polycyclic Peptidomimetics. *J. Am. Chem. Soc.* **135**, 12297-12304 (2013).

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

54 Goto, Y. & Suga, H. Translation initiation with initiator tRNA charged with exotic peptides. *J. Am. Chem. Soc.* **131**, 5040-5041 (2009).

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

56 Goto, Y., Murakami, H. & Suga, H. Initiating translation with D-amino acids. *RNA* **14**, 1390-1398 (2008).

57 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).

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

59 Perona, J. J., Rould, M. A. & Steitz, T. A. Structural basis for transfer RNA aminoacylation by Escherichia Coli glutaminyl-tRNA synthetase. *Biochemistry* **32**, 8758-8771 (1993).

60 Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15**, 509-524 (2014).

61 Baselga-Escudero, L. *et al.* Resveratrol and EGCG bind directly and distinctively to miR-33a and miR-122 and modulate divergently their levels in hepatic cells. *Nucleic Acids Res.* **42**, 882-892 (2014).

62 Voorhees, R., Weixlbaumer, A., Loakes, D., Kelley, A. & Ramakrishnan, V. Insights into substrate stabilization from snapshots of the peptidyl transferase center of the intact 70S ribosome. *Nat. Struct. Mol. Biol.* **16**, 528-533 (2009).

63 Dorner, S., Brunelle, J., Sharma, D. & Green, R. The hybrid state of tRNA binding is an authentic translation elongation intermediate. *Nat. Struct. Mol. Biol.* **13**, 234-241 (2006).

64 Kim, D. F. & Green, R. Base-pairing between 23S rRNA and tRNA in the ribosomal A site. *Mol. Cell* **4**, 859-864 (1999).

65 Terasaka, N., Hayashi, G., Katoh, T. & Suga, H. An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. *Nat. Chem. Biol.* **10**, 555-557 (2014).

66 Winter, J., Jung, S., Keller, S., Gregory, R. I. & Diederichs, S. Many roads to maturity: microRNA biogenesis pathways and their regulation. *Nat. Cell Biol.* **11**, 228-234 (2009).

67 Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051-4060 (2004).

68 Cai, X. Z., Hagedorn, C. H. & Cullen, B. R. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957-1966 (2004).

69 Borchert, G. M., Lanier, W. & Davidson, B. L. RNA polymerase III transcribes human microRNAs. *Nat. Struct. Mol. Biol.* **13**, 1097-1101 (2006).

70 Lee, Y. *et al.* The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**, 415-419 (2003).

71 Denli, A. M., Tops, B. B. J., Plasterk, R. H. A., Ketting, R. F. & Hannon, G. J. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**, 231-235 (2004).

72 Gregory, R. I. *et al.* The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**, 235-240 (2004).

73 Han, J. J. *et al.* The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev.* **18**, 3016-3027 (2004).

74 Landthaler, M., Yalcin, A. & Tuschl, T. The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. *Curr. Biol.* **14**, 2162-2167 (2004).

75 Okada, C. *et al.* A High-Resolution Structure of the Pre-microRNA Nuclear Export Machinery. *Science* **326**, 1275-1279 (2009).

76 Yi, R., Qin, Y., Macara, I. G. & Cullen, B. R. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev.* **17**, 3011-3016 (2003).

77 Hutvagner, G. *et al.* A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**, 834-838 (2001).

78 Haase, A. D. *et al.* TRBP, a regulator of cellular PKR and HIV-1 virus expression, interacts with Dicer and functions in RNA silencing. *EMBO Rep.* **6**, 961-967 (2005).

79 Khvorova, A., Reynolds, A. & Jayasena, S. D. Functional siRNAs and miRNAs exhibit strand bias *Cell* **115**, 209-216 (2003).

80 Schwarz, D. S. *et al.* Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**, 199-208 (2003).

81 Liu, J. D. *et al.* Argonaute2 is the catalytic engine of mammalian RNAi. *Science* **305**, 1437-1441 (2004).

82 Huntzinger, E. & Izaurralde, E. Gene silencing by microRNAs: contributions of translational repression and mRNA decay. *Nat. Rev. Genet.* **12**, 99-110 (2011).

83 Hansen, T. B. *et al.* Natural RNA circles function as efficient microRNA sponges. *Nature* **495**, 384-388 (2013).

84 Curtis, E. & Liu, D. Discovery of Widespread GTP-Binding Motifs in Genomic DNA and RNA. *Chem. Biol.* **20**, 521-532 (2013).

85 Vu, M. M. *et al.* Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chem. Biol.* **19**, 1247-1254 (2012).

86 Izzotti, A., Cartiglia, C., Steele, V. E. & De Flora, S. MicroRNAs as targets for dietary and pharmacological inhibitors of mutagenesis and carcinogenesis. *Mutat. Res.* **751**, 287-303 (2012).

87 Melse-Boonstra, A., West, C. E., Katan, M. B., Kok, F. J. & Verhoef, P. Bioavailability of heptaglutamyl relative to monoglutamyl folic acid in healthy adults. *Am. J. Clin. Nutr.* **79**, 424-429 (2004).

88 Selhub, J. Determination of Tissue Folate Composition by Affinity-Chromatography Followed by High-Pressure Ion-Pair Liquid-Chromatography. *Anal. Biochem.* **182**, 84-93 (1989).

89 Mulinare, J., Cordero, J. F., Erickson, J. D. & Berry, R. J. Periconceptional use of multivitamins and the occurrence of neural tube defects. *Jama-J Am Med Assoc* **260**, 3141-3145 (1988).

90 Jiang, R. *et al.* Joint association of alcohol and folate intake with risk of major chronic disease in women. *Am. J. Epidemiol.* **158**, 760-771 (2003).

91 Marsit, C. J., Eddy, K. & Kelsey, K. T. MicroRNA responses to cellular stress. *Cancer Res.* **66**, 10843-10848 (2006).

92 Barrick, J. E. *et al.* New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6421-6426 (2004).

93 Singer, B., Shtatland, T., Brown, D. & Gold, L. Libraries for genomic SELEX. *Nucleic Acids Res.* **25**, 781-786 (1997).

94 Zimmermann, B., Bilusic, I., Lorenz, C. & Schroeder, R. Genomic SELEX: A discovery tool for genomic aptamers. *Methods* **52**, 125-132 (2010).

95 Dinger, M. E., Pang, K. C., Mercer, T. R. & Mattick, J. S. Differentiating protein-coding and noncoding RNA: challenges and ambiguities. *PLoS Comput. Biol.* **4** (2008).

96 Koppers-Lalic, D. *et al.* Nontemplated nucleotide additions distinguish the small RNA composition in cells from exosomes. *Cell reports* **8**, 1649-1658 (2014).

97 Pan, T. & Uhlenbeck, O. C. Circularly permuted DNA, RNA and proteins - a review. *Gene* **125**, 111-114 (1993).

98 Chen, L., Yun, S. W., Seto, J., Liu, W. & Toth, M. The fragile X mental retardation protein binds and regulates a novel class of mRNAs containing U rich target sequences. *Neuroscience* **120**, 1005-1017 (2003).

99 Dobbelstein, M. & Shenk, T. In vitro selection of RNA ligands for the ribosomal L22 protein associated with Epstein-Barr virus-expressed RNA by using randomized and cDNA-derived RNA libraries. *J. Virol.* **69**, 8027-8034 (1995).

100 Fujimoto, Y., Nakamura, Y. & Ohuchi, S. HEXIM1-binding elements on mRNAs identified through transcriptomic SELEX and computational screening. *Biochimie* **94**, 1900-1909 (2012).

101 Yuan, G., Klambt, C., Bachellerie, J. P., Brosius, J. & Huttenhofer, A. RNomics in Drosophila melanogaster: identification of 66 candidates for novel non-messenger RNAs. *Nucleic Acids Res.* **31**, 2495-2507 (2003).

102 Liu, J. M. *et al.* Experimental discovery of sRNAs in Vibrio cholerae by direct cloning, 5S/tRNA depletion and parallel sequencing. *Nucleic Acids Res.* **37**, e46 (2009).

103 Landgraf, P. *et al.* A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**, 1401-1414 (2007).

104 Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev.* **23**, 2639-2649 (2009).

105 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406-3415 (2003).

106 Tsuji, S. *et al.* Effective isolation of RNA aptamer through suppression of PCR bias. *Biochem. Biophys. Res. Commun.* **386**, 223-226 (2009).

107 Zhang, X. X. & Zeng, Y. The terminal loop region controls microRNA processing by Drosha and Dicer. *Nucleic Acids Res.* **38**, 7689-7697 (2010).

108 Auyeung, V. C., Ulitsky, I., McGeary, S. E. & Bartel, D. P. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* **152**, 844-858 (2013).

109 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

110 Mituyama, T. *et al.* The Functional RNA Database 3.0: databases to support mining and annotation of functional RNAs. *Nucleic Acids Res.* **37**, D89-92 (2009).

111 Chan, P. P. & Lowe, T. M. GtRNAdb: a database of transfer RNA genes detected in genomic sequence. *Nucleic Acids Res.* **37**, D93-D97 (2009).

112 Lestrade, L. & Weber, M. J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res.* **34**, D158-D162 (2006).

113 Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413-444 (2010).

114 Magliery, T. J., Anderson, J. C. & Schultz, P. G. Expanding the genetic code: Selection of efficient suppressors of four-base codons and identification of "shifty" four-base codons with a library approach in Escherichia coli. *J. Mol. Biol.* **307**, 755-769 (2001).

115 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001).

116 Forster, A. *et al.* Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 6353-6357 (2003).

117 Polycarpo, C. R. *et al.* Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580**, 6695-6700 (2006).

118 Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Genetically encoding N(ε)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **4**, 232-234 (2008).

119 Mukai, T. *et al.* Adding L-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochem. Biophys. Res. Commun.* **371**, 818-822 (2008).

120 Yanagisawa, T. *et al.* Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(ε)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem. Biol.* **15**, 1187-1197 (2008).

121 Rackham, O. & Chin, J. W. A network of orthogonal ribosome•mRNA pairs. *Nat. Chem. Biol.* **1**, 159-166 (2005).

122 Wang, K. H., Neumann, H., Peak-Chew, S. Y. & Chin, J. W. Evolved orthogonal ribosomes enhance the efficiency of synthetic genetic code expansion. *Nat. Biotechnol.* **25**, 770-777 (2007).

123 Neumann, H., Wang, K., Davis, L., Garcia-Alai, M. & Chin, J. W. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. *Nature* **464**, 441-444 (2010).

124 Wang, K. H. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat Chem* **6**, 393-403 (2014).

125 Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751-755 (2001).

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

127 Moazed, D. & Noller, H. F. Intermediate States in the Movement of Transfer-RNA in the Ribosome. *Nature* **342**, 142-148 (1989).

128 Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234-1242 (2009).

129 Bashan, A. *et al.* Structural basis of the ribosomal machinery for peptide bond formation, translocation, and nascent chain progression. *Mol. Cell* **11**, 91-102 (2003).

130 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).

131 Steitz, T. A. A structural understanding of the dynamic ribosome machine. *Nat. Rev. Mol. Cell. Biol.* **9**, 242-253 (2008).

132 Moazed, D. & Noller, H. F. Sites of interaction of the CCA end of peptidyl-tRNA with 23S rRNA. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 3725-3728 (1991).

133 Samaha, R. R., Green, R. & Noller, H. F. A base pair between tRNA and 23S rRNA in the peptidyl transferase centre of the ribosome. *Nature* **377**, 309-314 (1995).

134 Lescoute, A. & Westhof, E. The interaction networks of structured RNAs. *Nucleic Acids Res.* **34**, 6587-6604 (2006).

135 Green, R., Switzer, C. & Noller, H. F. Ribosome-catalyzed peptide-bond formation with an A-site substrate covalently linked to 23S ribosomal RNA. *Science* **280**, 286-289 (1998).

136 Polikanov, Y. S., Steitz, T. A. & Innis, C. A. A proton wire to couple aminoacyl-tRNA accommodation and peptide-bond formation on the ribosome. *Nat. Struct. Mol. Biol.* **21**, 787-793 (2014).

137 Spiegel, P. C., Ermolenko, D. N. & Noller, H. F. Elongation factor G stabilizes the hybrid-state conformation of the 70S ribosome. *RNA* **13**, 1473-1482 (2007).

138 Cornish, P. V., Ermolenko, D. N., Noller, H. F. & Ha, T. Spontaneous intersubunit rotation in single ribosomes. *Mol. Cell* **30**, 578-588 (2008).

139 Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Res.* **41**, D262-267 (2013).

140 Youngman, E. M. & Green, R. Affinity purification of in vivo-assembled ribosomes for in vitro biochemical analysis. *Methods* **36**, 305-312 (2005).

141 Liu, J., Liu, M. & Horowitz, J. Recognition of the universally conserved 3'-CCA end of tRNA by elongation factor EF-Tu. *RNA* **4**, 639-646 (1998).

142 Schmeing, T. M. *et al.* The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* **326**, 688-694 (2009).

143 Kang, T. & Suga, H. Translation of a histone H3 tail as a model system for studying peptidyl-tRNA drop-off. *FEBS Lett.* **585**, 2269-2274 (2011).

144 Sako, Y., Morimoto, J., Murakami, H. & Suga, H. Ribosomal synthesis of bicyclic peptides via two orthogonal inter-side-chain reactions. *J. Am. Chem. Soc.* **130**, 7232-7234 (2008).

145 Zhang, J. W. & Ferre-D'Amare, 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).

146 Kao, C., Zheng, M. & Rüdisser, S. A simple and efficient method to reduce nontemplated nucleotide addition at the 3' terminus of RNAs transcribed by T7 RNA polymerase. *RNA* **5**, 1268-1272 (1999).

147 Doi, Y., Ohtsuki, T., Shimizu, Y., Ueda, T. & Sisido, M. Elongation factor Tu mutants expand amino acid tolerance of protein biosynthesis system. *J. Am. Chem. Soc.* **129**, 14458-14462 (2007).

148 Bose, D. *et al.* The tuberculosis drug streptomycin as a potential cancer therapeutic: inhibition of miR-21 function by directly targeting its precursor. *Angew. Chem. Int. Ed. Engl.* **51**, 1019-1023 (2012).

149 Velagapudi, S. P., Gallo, S. M. & Disney, M. D. Sequence-based design of bioactive small molecules that target precursor microRNAs. *Nat. Chem. Biol.* **10**, 291-297 (2014).

150 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).

151 Wilusz, J. E., Freier, S. M. & Spector, D. L. 3' end processing of a long nuclear-retained noncoding RNA yields a tRNA-like cytoplasmic RNA. *Cell* **135**, 919-932 (2008).

152 Sunwoo, H. *et al.* MEN ε/β nuclear-retained non-coding RNAs are up-regulated upon muscle differentiation and are essential components of paraspeckles. *Genome Res.* **19**, 347-359 (2009).

153 Chen, X. G., Sim, S., Wurtmann, E. J., Feke, A. & Wolin, S. L. Bacterial noncoding Y RNAs are widespread and mimic tRNAs. *RNA* **20**, 1715-1724 (2014).

154 Wilusz, J. E., Whipple, J. M., Phizicky, E. M. & Sharp, P. A. tRNAs marked with CCACCA are targeted for degradation. *Science* **334**, 817-821 (2011).

155 Virumae, K., Saarma, U., Horowitz, J. & Remme, J. Functional importance of the 3'-terminal adenosine of tRNA in ribosomal translation. *J. Biol. Chem.* **277**, 24128-24134 (2002).

List of accomplishments

Papers

1. Terasaka N., Hayashi G., Katoh H. & Suga H. An engineered ribosome-tRNAs pair functions orthogonally to the wild-type under an artificially programmed genetic code. *Nat. Chem. Biol.*, **10**, 555–557. (2014)
2. Terasaka N. & Suga H. Flexizymes-facilitated genetic code reprogramming leading to the discovery of drug-like peptides. *Chem. Lett.*, **43**, 11-19 (2014).

Oral presentations

1. Terasaka N., Hayashi G., Katoh H. & Suga H., An orthogonal ribosome-tRNAs pair by the engineering of peptidyl transferase center. 25th tRNA conference, Kyllini, GREECE, Sep. 2014

2. 寺坂尚紘, 林剛介, 加藤敬行, 菅裕明, 天然リボソーム・tRNA ペアと直交性を持つ変異リボソーム・tRNA ペアの構築, 日本ケミカルバイオロジー学会第 9 回年会, 大阪府, 2014 年 6 月

3. Terasaka N., Hayashi G., Katoh H. & Suga H., Mutant ribosome•tRNA pair towards orthogonal genetic code. 9th International Symposium on Aminoacyl-tRNA Synthetases, Hakone, JAPAN, Oct. 2013

4. 寺坂尚紘, 林剛介, 加藤敬行, 菅裕明, 変異リボソーム・tRNA ペアを用いた直交性遺伝暗号の創成, 第 15 回日本 RNA 学会年会, 愛媛県, 2013 年 7 月,

5. Terasaka N. & Suga H., The translational activity of a CCA-mutated tRNA-ribosome mutant pair. Annual Symposium on Academic English for Chemistry, Tokyo, JAPAN, Feb. 2013

6. 寺坂尚紘, 林剛介, 加藤敬行, 菅裕明, 変異リボソーム・tRNA ペアを用いた直交性遺伝暗号の創成, RNA フロンティアミーティング 2012, 熊本県, 2012 年 9 月

Poster presentations

1. Terasaka N., Hayashi G., Katoh H. & Suga H., An orthogonal ribosome-tRNA pair via engineering of the peptidyl transferase center. The 3rd Japan-Swiss Chemical Biology Symposium, Bern, SWITZERLAND, Oct. 2014

2. 寺坂尚紘，二井一樹，加藤敬行，菅裕明, In vitro セレクションを用いて小分子結合性 small non-coding RNA を探索する，第 16 回生命化学研究会，静岡県, 2014 年 1 月

3. Terasaka N., Futai K., Katoh T., Suga H., Discovery of human small non-coding RNAs binding to small molecules by SELEX-NT. RiboClub Annual Meeting 2013, Quebec, CANADA, Sep. 2013

4. Terasaka N., Hayashi G., Katoh H. & Suga H., Translational activity of CCA-mutated tRNA-ribosome mutant pair. XXIV tRNA conference, Olmue, CHILE, Dec. 2012

5. 寺坂尚紘，二井一樹，加藤敬行，後藤佑樹，菅裕明, In vitro セレクションによる生体内小分子に結合するヒト small non-coding RNA の探索，第 24 回高遠シンポジウム，長野県, 2012 年 8 月

6. 寺坂尚紘，二井一樹，加藤敬行，後藤佑樹，菅裕明, In vitro セレクションを用いた生体内小分子に結合するヒト small non-coding RNA の探索，第 14 回日本 RNA 学会年会，宮城県, 2012 年 7 月

7. 寺坂尚紘，二井一樹，加藤敬行，菅裕明，生体内小分子に結合するヒト small non-coding RNA の探索，新規素材探索第 11 回セミナー，神奈川県, 2012 年 6 月

Awards and fellowships

1. Sidney Altman Endowment Travel Award in the 25th tRNA conference (Sep. 2014).
2. Best presenter award in the 15th annual meeting of the RNA Society of Japan (Jul. 2013).
3. Travel Grant for Attending the 63rd Lindau Nobel Laureate Meeting from the Japan Society of the Promotion of Science (Jul. 2013).
4. Research Fellowship for young scientists, the Japan Society of the Promotion of Science (2012-2014).

Acknowledgement

I would like to express my sincere gratitude to Professor Hiroaki Suga for his kind guidance, valuable suggestions, and encouragement throughout this work. I am deeply grateful to Associate Professor Naokazu Kano, Assistant professor Takayuki Katoh, Assistant Professor Yuki Goto and Project Assistant Professor Toby Passioura for their helpful advice and discussion. I am greatly indebted to Dr. Kazuki Futai and Dr. Gosuke Hayashi for teaching me about experimental things and discussing with me about the researches. I also thank to all members of the laboratory.

I thank to Japan Society for the Promotion of Science (JSPS) for a grant and financial support.

Finally, I would like to thank my parents and brother for invaluable assistance and cheering words.

Naohiro Terasaka