学位論文

Development of tRNA-engineered *in vitro* translation systems to synthesize peptides with expanded repertoire of nonproteinogenic amino acids

(多種の非タンパク質性アミノ酸を基質として利用する
tRNA 改変型ペプチド翻訳合成系の開発)

平成 28 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

岩根　由彦

Abstract

Proteins and peptides, which are collectively called as polypeptides, perform a vast array of biological functions and attract a great interest in biological and pharmaceutical sciences. In living organisms, polypeptides are synthesized by ribosomal translation reaction in an order specified by mRNA sequence composed of nucleotide triplets named codons. The relationships between codons and amino acids are referred to as the genetic code. Among 64 codons, 61 codons designate elongation of the nascent polypeptides with specific proteinogenic amino acids, synthesizing amino acid polymer with 20 proteinogenic building blocks. Beyond this nature's limitation, artificial engineering of translation machinery has allowed for the ribosomal synthesis of polypeptides containing nonproteinogenic amino acids (npAAs), which have created innovative and practical methodologies. For example, the synthesis of polypeptides containing optical probes facilitates structural biology by NMR and X-ray crystallography, and also it facilitates the development of chemically modified "nonstandard" peptide drugs with improved pharmacological properties. In particular, our laboratory has developed a selection technology, named RaPID (Random nonstandard peptides integrated discovery) system, to discover peptide binders to drug target proteins from a sequence-randomized peptide library, and successfully developed 'nonstandard' peptide drugs with noncanonical scaffolds that mimic the characteristics of naturally occurring bioactive peptides.

In spite of the great potential applicability, the current engineered translation systems still have methodological limitations that diminish its scope of application: (1) Accurate synthesis of a polypeptide containing more than two distinct npAAs can be achieved only when several proteinogenic amino acids are excluded from *in vitro* translation systems to create vacant codons that are assignable for npAAs, which leads to the decrease in building block repertoire. (2) Even when such an engineered translation system is used, it is still impossible to efficiently synthesize polypeptides containing many kinds of *N*-methyl amino acids (^{Me}AAs) due to an unspecified translational disorder. In this context, there is a need to develop *in vitro* translation systems enabling (1) expansion of the building block repertoire by utilizing multiple npAAs without the need to exclude any of 20 proteinogenic ones and (2) the synthesis of nonstandard peptides containing a variety of ^{Me}AAs. As the Ph.D. degree research, I developed these two kinds of engineered translation systems.

Chapter 1 is the general introduction that describes the relating biological knowledge and the previous methods to synthesize npAA-containing polypeptides in translation reaction. This chapter also explains the merit of peptide-based drugs along with a selection methodology to discover novel peptide binders to drug targets. The limitations of conventional engineered translation systems are described at the end.

Chapter 2 describes the first research topic entitled 'Expanding the amino acid repertoire of ribosomal polypeptide synthesis via artificial division of codon boxes'. In native translation machinery, the repertoire of amino acids available for ribosomal polypeptide synthesis is restricted by the genetic code, where 61 sense codons redundantly code for 20 proteinogenic amino acids. In this study, I have developed a method, named 'artificial division of codon boxes', to reduce the genetic code redundancy and create vacant codons without the need to sacrifice any of 20 proteinogenic amino acids. The reassignment of npAAs to these vacant codons allowed for the expansion of the amino acid repertoire from the standard 20. The proof of concept of this novel translation system has been

achieved by the accurate synthesis of various model peptides such as a 32-mer linear peptide composed of 23 amino acid repertoire and a 14-mer macrocyclic *N*-methyl-peptide drug that inhibits E6AP protein.

Chapter 3 describes the second research topic entitled 'Ribosomal synthesis of highly *N*-methylated peptides by tuning affinities between EF-Tu and *N*-methyl-aminoacyl-tRNAs'. For the selection of peptide-based drugs, it is ideal to synthesize a peptide library containing multiple ^{Me}AAs because backbone *N*-methyl modification improves the peptide's cell-membrane permeability and stability against enzymatic decomposition. However, the synthesis of such highly *N*-methylated peptides is difficult because the ribosomal incorporation of multiple ^{Me}AAs drastically reduces the synthetic efficiency and accuracy due to an unspecified mechanism. In this study, I revealed the problem that ^{Me}AA-tRNAs did not bind to a carrier protein named EF-Tu, and therefore ^{Me}AA-tRNAs could not be delivered into ribosome efficiently. To overcome this problem, I developed a method to reinforce the weak affinities, which enabled all of the tested ^{Me}AA-tRNAs to bind to EF-Tu with appropriate strengths. The resultant optimal translation system exhibited the improved ^{Me}AA incorporation efficiency and accuracy and successfully expressed highly *N*-methylated peptides containing up to nine distinct ^{Me}AAs.

Chapter 4 is the general conclusion of the entire thesis. In this study, I have developed the engineered translation systems that allow for the synthesis of nonstandard polypeptides composed of expanded repertoire of proteinogenic and nonproteinogenic amino acids. In particular, the integration of these methods with peptide selection technologies, such as the RaPID system, would enable us to express a library of nonstandard macrocyclic peptides with highly modified scaffolds and to discover nonstandard peptide drugs that possess improved binding potencies and pharmacokinetic properties in future.

Table of contents

Abstract 1

Table of contents 3

Abbreviation list 5

Chapter 1. General introduction 6

1.1. Translation reaction: ribosomal polypeptide synthesis according to the genetic code 7

1.2. Structure and functions of tRNAs 12

1.3. Methods to synthesize nonproteinogenic aminoacyl-tRNAs 16

1.4. Methods to assign npAAs in the genetic code 20

1.5. Methods to improve the synthetic efficiency of polypeptides containing suboptimal npAA substrates 23

1.6. Development of nonstandard macrocyclic peptide drugs 27

1.7. Limitations of the conventional engineered translation systems 33

Chapter 2. Expanding the amino acid repertoire of ribosomal polypeptide synthesis via artificial division of codon boxes 34

2.1. Introduction 35

2.2. Results and discussion 40

2.2.1. Accurate decoding by *in vitro* tRNA transcripts 40

2.2.2. Demonstration of artificial division of a single codon box 45

2.2.3. Assignment of two different npAAs in the Arg CGN codon box 49

2.2.4. Artificial division of multiple codon boxes 51

2.2.5. Expression of macrocyclic *N*-methyl-peptide CM_{11}-1 57

2.3. Conclusion 59

2.4. Supplementary results 61

2.4.1. Preparation of tRNAs that do not have a guanosine at the 5'-terminus ($tRNA^{Gln}_{CUG}$, $tRNA^{Pro}_{CGG}$, $tRNA^{Pro}_{GGG}$, and $tRNA^{Trp}_{GCA}$) 61

2.4.2. Concentration optimization of three tRNA transcripts that cannot be acylated by AARSs efficiently 62

2.4.3. Optimizations of translation conditions to achieve artificial division of a single codon box 65

2.5. Methods 71

2.6. Supplementary tables 74

Chapter 3. Ribosomal synthesis of highly *N*-methylated peptides by tuning affinities between EF-Tu and *N*-methyl-aminoacyl-tRNAs 82

3.1. Introduction 84

3.2. Results and discussion 87

3.2.1. Experimental proof that many ^{Me}AA-tRNAs cannot bind to EF-Tu with sufficient strengths 87

3.2.2. Development of a method to reinforce the weak EF-Tu affinities of ^{Me}AA-tRNAs 89

3.2.3. Expression of highly *N*-methylated peptide via EF-Tu affinity tuning for ^{Me}AA-tRNAs 95

3.3. Conclusion 98

3.4. Supplementary results 100

3.4.1. Development of a novel method to quantify affinities between EF-Tu and aminoacyl-tRNAs 100

3.4.2. Optimization of translation condition to express peptides containing multiple ^{Me}AAs 103

3.5. Methods 108

5.6 Supplementary tables 113

Chapter 4. General conclusion 120

References 123

List of achievements 136

Acknowledgement 137

Abbreviation list

A-site	Aminoacyl-tRNA site
AARS	Aminoacyl-tRNA Synthetase
^{Ac}K	*ε-N*-acetyllysine
Cit	Citrulline
ClAc^{D}W	*N*-chloroacetyl-D-tryptophan
CME	Cyanomethyl ester
DBE	3,5-dinitrobenzyl ester
dFx	Dinitro-Flexizyme
E-site	Exit site
E. coli	*Escherichia coli*
EF-Tu	Elongation factor thermo unstable
eFx	Enhanced Flexizyme
FIT	Flexible *In vitro* Translation
FIT-nt	FIT system composed of native tRNAs
FIT-XXt	FIT system composed of XX kinds of *in vitro* tRNA transcripts (XX denotes number)
FLAG	A peptide tag sequence: DYKDDDDK
Flexizyme	Flexible tRNA-acylation ribozyme
fMet	Formylmethionine
IF	Initiation factor
^{Iodo}F	4-iodophenylalanine
MALDI-TOF-MS	Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry
^{Me}A	*N*-methylalanine
^{Me}AA	*N*-methyl-amino acid
^{Me}AA-tRNA	*N*-methyl-aminoacyl-tRNA
^{Me}D	*N*-methylaspartic acid
^{Me}F	*N*-methylphenylalanine
^{Me}G	*N*-methylglycine
^{Me}L	*N*-methylleucine
^{Me}M	*N*-methylmethionine
^{Me}Nl	*N*-methylnorleucine
^{Me}Nv	*N*-methylnorvaline
^{Me}S	*N*-methylserine
^{Me}T	*N*-methylthreonine
^{Me}V	*N*-methylvaline
^{Me}Y	*N*-methyltyrosine
^{Me}Ym	*N*-methyl-p-methoxyphenylalanine
mRNA	Messenger RNA
npAA	Nonproteinogenic amino acids
npAA-tRNA	Nonproteinogenic aminoacyl-tRNA
P-site	Peptidyl-tRNA site
pAA	Proteinogenic amino acid
pAA-tRNA	Proteinogenic aminoacyl-tRNA
PAGE	Polyacrylamide gel electrophoresis
PTC	Peptidyl transfer center
RaPID	Random nonstandard Peptides Integrated Discovery
RF	Release factor
tRNA	Transfer RNA

Chapter 1

General introduction

1.1. Translation reaction: ribosomal polypeptide synthesis according to the genetic code

Translation reaction is an essential biological process where a ribosome synthesizes proteins and peptides, which are collectively called as polypeptides, according to genetic information encoded in mRNA sequences (Fig. 1a). mRNAs consist of sequential nucleotide triplets named codons. Among the 64 codons, 61 'sense' codons designate the nascent polypeptide elongation with the specific amino acid according to the genetic code, whereas the other three 'stop' codons designate termination of polypeptide synthesis (Fig. 1b). Translation is composed of three events: (1) initiation, (2) elongation, and (3) termination (Fig. 1c). The initiation event construct an initiation complex composed of ribosome, mRNA, and formylmethionyl-tRNA (fMet-tRNA^{fMet}). In the elongation event, a nascent polypeptide is elongated with 20 kinds of proteinogenic amino acids in the order specified by the mRNA sequence. During the termination event, the full-length nascent polypeptide is released from ribosome and then the ribosome complex is disassembled. In this section, the molecular mechanisms of these events are described based on *E. coli*'s system.

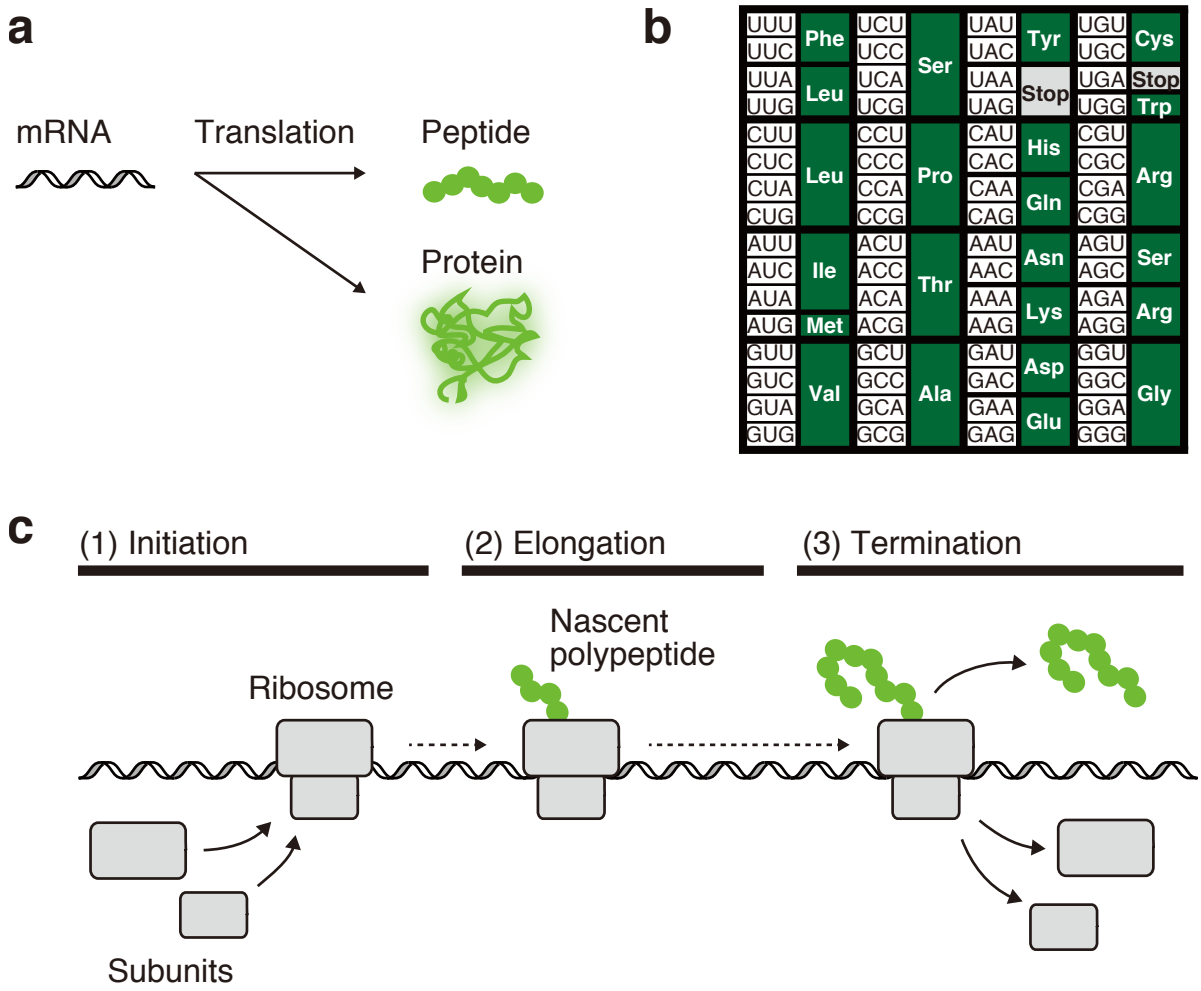

Figure 1. Ribosomal synthesis of polypeptides according to the genetic code. a, Polypeptide synthesis by translation reaction. **b**, The genetic code that designates the specific amino acid in response to each codon. **c**, Schematic illustration of three events involved in translation reaction.

1.1.1. Initiation

At the first step, an mRNA interacts with a small ribosomal subunit by forming base pairs between the mRNA Shine-Dalgarno (SD) sequence and the rRNA anti-Shine-Dalgarno (ASD) sequence (Fig. 2). Presence of AU-rich enhancer element in mRNA can further stabilize the complex via interaction with the small subunit[1]. Then three initiation factors (IF1, IF2, IF3) and fMet-tRNA^{fMet} bind to the complex[2]. The fMet-tRNA^{fMet} forms codon-anticodon base pairs with the initiation AUG codon, located 8–10 nucleotides downstream of the SD sequence[3]. As IF2 can selectively recognize fMet-tRNA^{fMet} by its unique structural features such as the formylmethionine moiety and the presence of C1-A72 mismatch[4], undesired tRNA species such as non-formylated Met-tRNA^{fMet} and elongator aminoacyl-tRNAs are not involved in this process. Subsequently, a large ribosomal subunit binds to the complex, and IFs dissociate from the ribosome complex using the energy of GTP hydrolysis on IF2[5]. The fMet-tRNA^{fMet} is located in a space inside of the ribosome called P-site (peptidyl-tRNA site). The resultant initiation complex composed of 70S ribosome, mRNA, and fMet-tRNA^{fMet} acts as the starting point for the next elongation event.

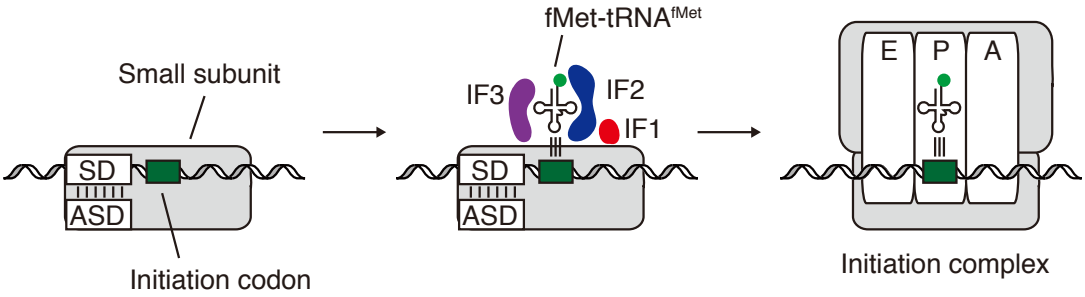

Figure 2. Initiation event. SD: Shine-Dalgarno sequence, ASD: anti-Shine-Dalgarno sequence.

1.1.2. Elongation

During elongation event, tRNAs work as adaptor molecules that bring specific amino acids to the corresponding codons present in ribosomal A-site (aminoacyl-tRNA site) (Fig. 3a). Each of 20 proteinogenic amino acids is charged to the corresponding tRNAs by specific aminoacyl-tRNA synthetase (AARS) [6,7]. For example, Phe is charged to a tRNA specific to Phe, which is represented as tRNA^{Phe}, by an AARS specific to Phe (PheRS: phenylalanyl-tRNA synthetase). The resultant aminoacyl-tRNAs then bind to GTP-bound EF-Tu (elongation factor thermo unstable), forming a complex named ternary complex[8]. EF-Tu binds to the L12 ribosomal protein and thereby positions the bound aminoacyl-tRNA in proximity of ribosome[9,10]. Subsequently, the aminoacyl-tRNA is inserted into ribosome and its anticodon nucleotides forms base pairs with the codon nucleotides present in A-site (Fig. 3b)[11,12]. When all of the nucleotide triplet match[13], GTP hydrolysis on EF-Tu is induced via a conformational change involving tRNA, ribosome, and EF-Tu[12,14-18]. The GTP hydrolysis leads to an extensive structural change of EF-Tu from GTP- to GDP-bound form[19], which release the bound aminoacyl-tRNA into A-site. After the accommodation, the peptidyl group of the peptidyl-tRNA in P-site (or fMet group in the first elongation event) and the aminoacyl group of the aminoacyl-tRNA in A-site are positioned in the peptidyl transferase center (PTC) via base pair interactions between ribosomal RNA (rRNA) and the two tRNAs[20]. The PTC then catalyzes the peptide

bond formation, elongating the nascent polypeptide chain by one amino acid[21,22]. The elongated nascent polypeptide goes through ribosome exit tunnel and eventually gets out into cytosol[23]. After the peptidyl transfer reaction, the acceptor stem of uncharged tRNA in P-site moves toward E-site (exit site), causing a ribosomal structural change. To this state of ribosome, a GTP-bound EF-G (elongation factor G) binds and push the two tRNAs from P- and A-sites to E- and P-sites, respectively using the energy of GTP hydrolysis[24,25]. The mRNA is simultaneously shifted to the same direction by one codon. The vacant A-site can start another cycle of elongation event.

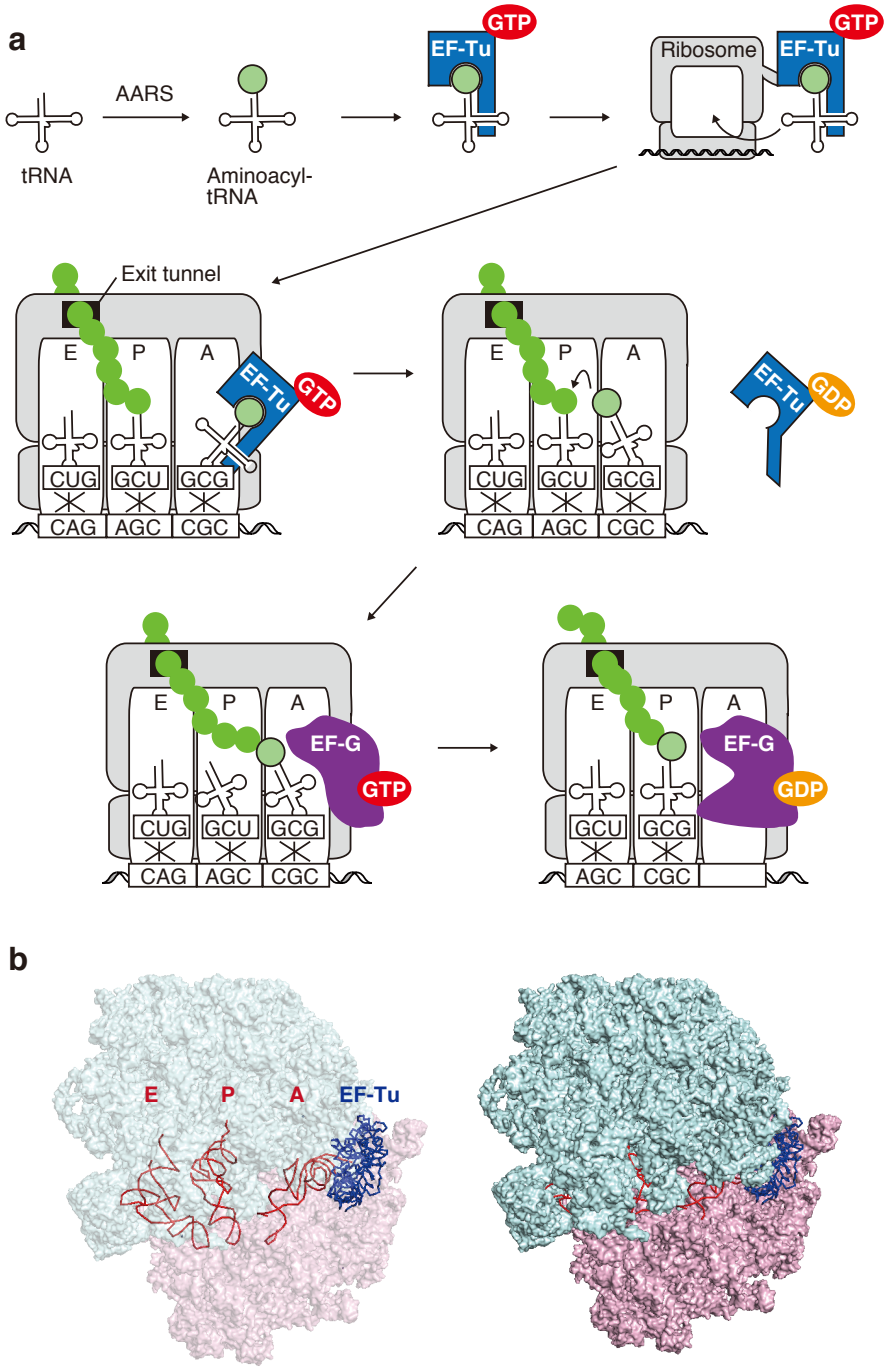

Figure 3. Elongation event. a, Processes that elongate the nascent polypeptide with one amino acid according to the genetic code of the A-site codon. **b**, Structure of a ribosome with a ternary complex (PDB ID: 2XQD/2XQE)[18]. The transparent ribosome structure is represented in the left image to show the presence of three tRNAs inside of the ribosome.

1.1.3. Termination

Among 64 codons, three stop codons (UAG, UAA, and UGA) designate termination of the polypeptide synthesis (Fig. 4). Termination reaction is catalyzed either by release factor 1 or 2 (RF1 and RF2). RF1 recognizes UAG and UAA stop codons, whereas RF2 recognizes UGA and UAA stop codons[26]. The recognition is achieved by direct binding of domain 2 of the RFs to the stop codons present in A-site, where PVT and SPF motifs of RF1 and RF2, respectively, contribute to the recognition[27,28]. When RFs recognize the corresponding stop codons, their GGQ motif in domain 3 directly hydrolyzes the ester bond of peptidyl-tRNA in P-site[27,28], which releases the nascent polypeptide from ribosome. Release factor 3 (RF3) and ribosome recycling factor (RRF) subsequently disassemble the ribosome complex to recycle the components[29-31].

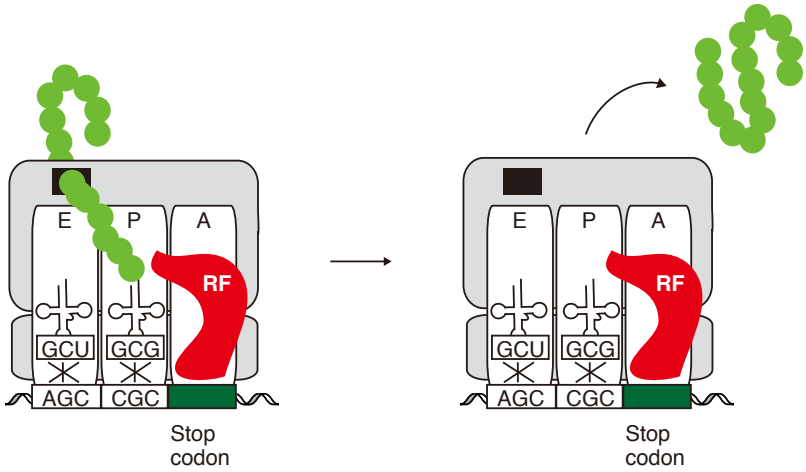

Figure 4. Termination event. RF1 recognize UAA and UAG stop codons, whereas RF2 recognize UAA and UGA stop codons. The ribosome complex is subsequently disassembled by RF3 and RRF.

1.1.4. *In vitro* reconstruction of translation system

Shimizu *et al.* have achieved *in vitro* reconstruction of translation system (so-called PURE (protein synthesis using recombinant elements) system)[32]. The system composed of individually purified *E. coli* factors such as ribosome, 20 proteinogenic amino acids, native tRNA mixture, 20 endogenous AARSs, initiation/elongation/termination factors, an RNA polymerase derived from T7 phage, NTPs, and enzymes for energy regeneration (Fig. 5). The reconstructed translation system has four major advantages compared to *in vivo* translation: (1) Concentrations of individual components (such as ions, proteins, RNAs) can be readily adjusted. It also enables us to add any desired chemical additives and biomolecules. (2) The reconstructed system minimizes the risk of mRNA and polypeptide degradation by cellular RNases and proteinases, respectively. (3) Purification and analyses of polypeptide products can be simplified. (4) Any undesired components can be omitted from the translation system, which facilitates the artificial engineering of the genetic code as described in the following section 1.4.3. For example, when RF1 is omitted, the corresponding UAG codon is no more recognized as the stop codon. The resultant vacant codon can be used for the reassignment of a nonproteinogenic amino acid.

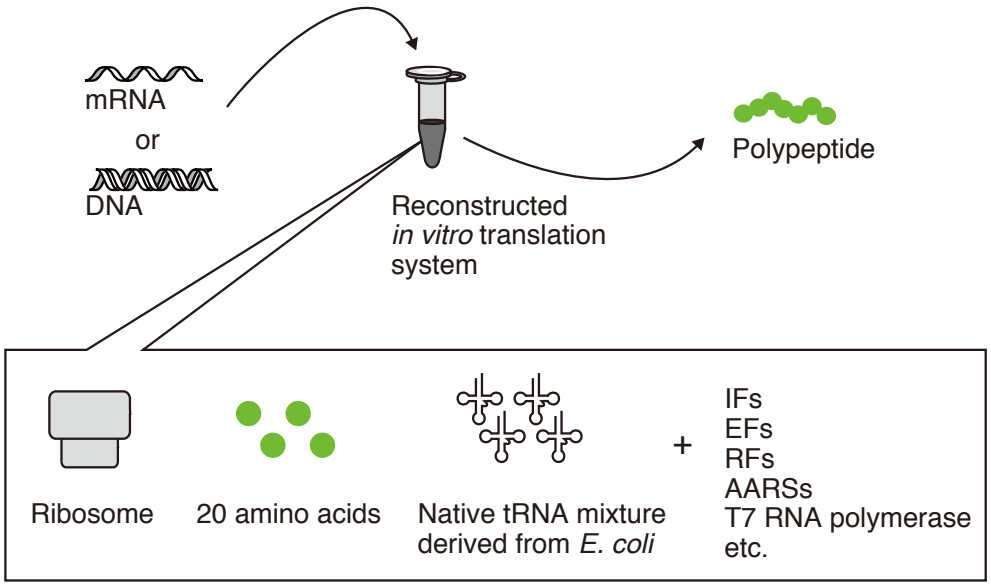

Figure 5. A reconstructed *in vitro* translation system.

1.2. Structure and functions of tRNAs

E. coli has around 45 distinct tRNAs[33], and individual tRNAs bring specific amino acids to the corresponding codons during translation reaction. The individual tRNA structures have evolved so that (1) they are accurately recognized by specific AARS(s), (2) they are efficiently delivered into ribosome, and (3) they accurately decode the A-site codons. This section describes how the tRNA structure confers these characteristics.

1.2.1. tRNA structure

tRNAs exhibit cloverleaf-shaped secondary structures by means of intra-molecular base pairing (Fig. 6a). The consensus structure conserved in all tRNAs comprises (1) acceptor stem, (2) D arm, (3) anticodon arm, (4) variable loop, (5) T arm, and (6) 3′-terminal NCCA end (N =U, C, A, or G). The secondary structure is further folded to form L-shaped tertiary structure[34], where D loop and T loop interacts each other, and two continuous helixes are observed: one helix is formed by continuous acceptor and T stems and the other is formed by continuous anticodon and D stems (Fig. 6b). In living organisms, tRNA transcripts composed of U, C, A, and G are post-transcriptionally modified by enzymes. As the result, native 'matured' tRNAs include 3–13 post-transcriptionally modified nucleosides that have important function as explained below (Fig. 6c).

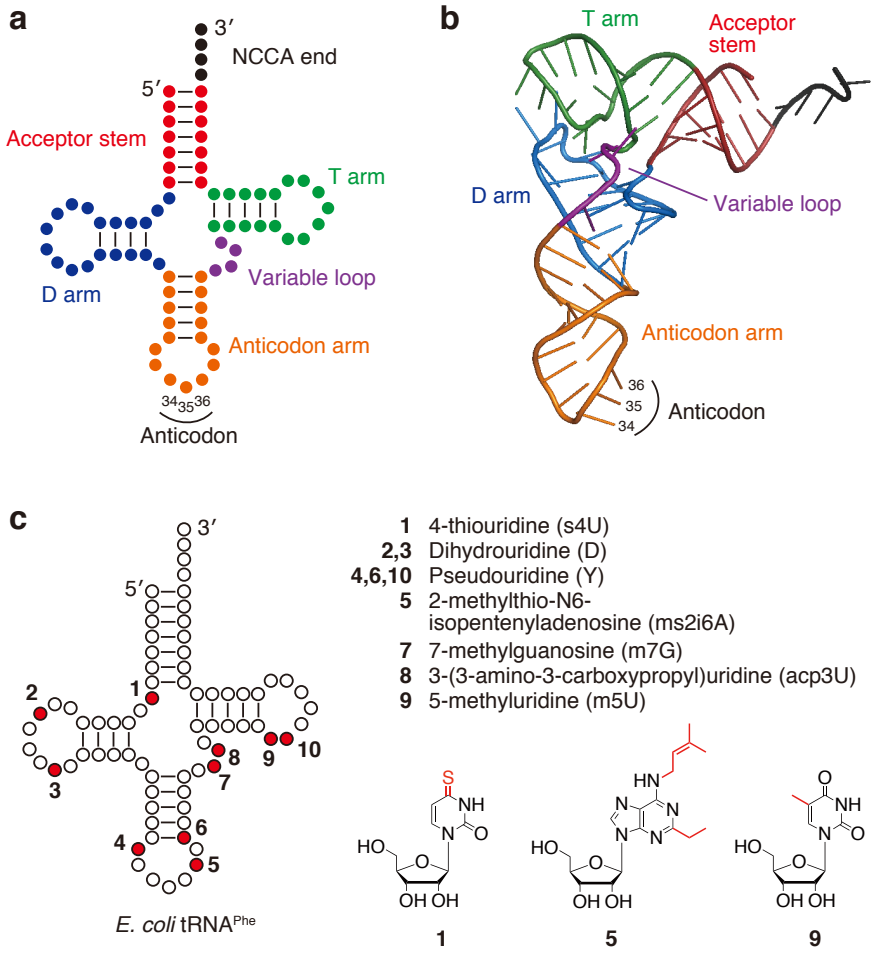

Figure 6. tRNA structure and its modifications. a, Typical secondary structure of a tRNA. **b**, Typical tertiary structure of a tRNA (PDB ID: 6TNA)[34]. **c**, Examples of nucleoside modifications.

1.2.2. Aminoacylation of tRNAs by AARSs

An AARS charge specific amino acid to specific tRNAs by forming an ester bond between the 3′-hydroxy group of the 3′-terminal adenosine and the carboxyl group of amino acid (Fig. 7). To ensure the correct aminoacylation, each AARS must discriminate the specific tRNA from the others. AARSs most frequently recognize the identity of acceptor stem, anticodon loop, and 3′-terminal NCCA sequences as discrimination sites, whereas D arm, the anticodon stem, and variable loop are used less frequently for the discrimination[6,7]. Some AARSs also recognize the anticodon post-transcriptional modifications[6,7,35]: LysRS and GluRS recognize mnm^{5}s^{2}U$_{34}$ (the first nucleotide of anticodon) of tRNA^{Lys} and tRNA^{Glu}, respectively, whereas IleRS recognizes k^{2}C$_{34}$ and t^{6}A$_{37}$ (the 3′-neighbour of anticodon) of tRNA^{Ile}. When the tRNAs lack their post-transcriptional modifications, the k_{cat}/K_M values drastically decrease to around 1%[6,7,35-38].

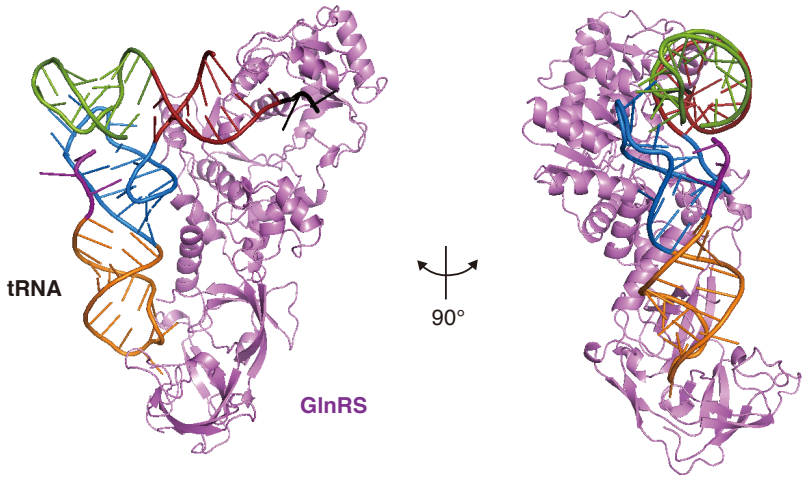

Figure 7. Cocrystal structure of GlnRS and tRNA^{Gln} (PDB ID: 1GTS)[39]. The tRNA domains are colored as Figure 6.

1.2.3. EF-Tu-mediated delivery of aminoacyl-tRNA into ribosome

During translation reaction, aminoacyl-tRNAs are delivered into ribosome by EF-Tu: EF-Tu binds to an aminoacyl-tRNA and then insert it into A-site using the energy of GTP hydrolysis (Fig. 8a)[11,12,40]. EF-Tu binds to an aminoacyl-tRNA at two distinct sites: (1) amino acid site and (2) T-stem site[8,41,42] (Fig. 8b). When only the former interaction at the amino acid site is considered, the binding affinities are drastically different depending on the amino acid speices[43] (Fig. 8c). In order to compensate the differences, nature has elegantly evolved the T-stem structures so that the affinities between EF-Tu and each of 20 aminoacyl-tRNAs become equivalent[44]. Due to this equivalent affinities, EF-Tu can deliver all of 20 proteinogenic aminoacyl-tRNAs into ribosome with similar efficiencies[45]. Interestingly, EF-Tu does not bind to the undesired aminoacyl-tRNA species such as Met-tRNA^{fMet}, which should be used for the initiation event rather than elongation event, and Sec-tRNA^{Sec} (Sec: selenocysteine), which should be delivered into ribosome by specific carrier protein SelB in response to SECIS (selenocysteine insertion sequence) in mRNA[46,47]. These facts suggest that the equivalent binding affinities between EF-Tu and the canonical aminoacyl-tRNAs also works as a quality control mechanism to reject undesired substrates[48]. The complete depletion of post-transcriptional modifications does not affect the affinities between the aminoacyl-tRNAs and EF-Tu[45,49], possibly because the direct binding regions (*i.e.* the acceptor stem and T-stem) contain few nucleoside modifications.

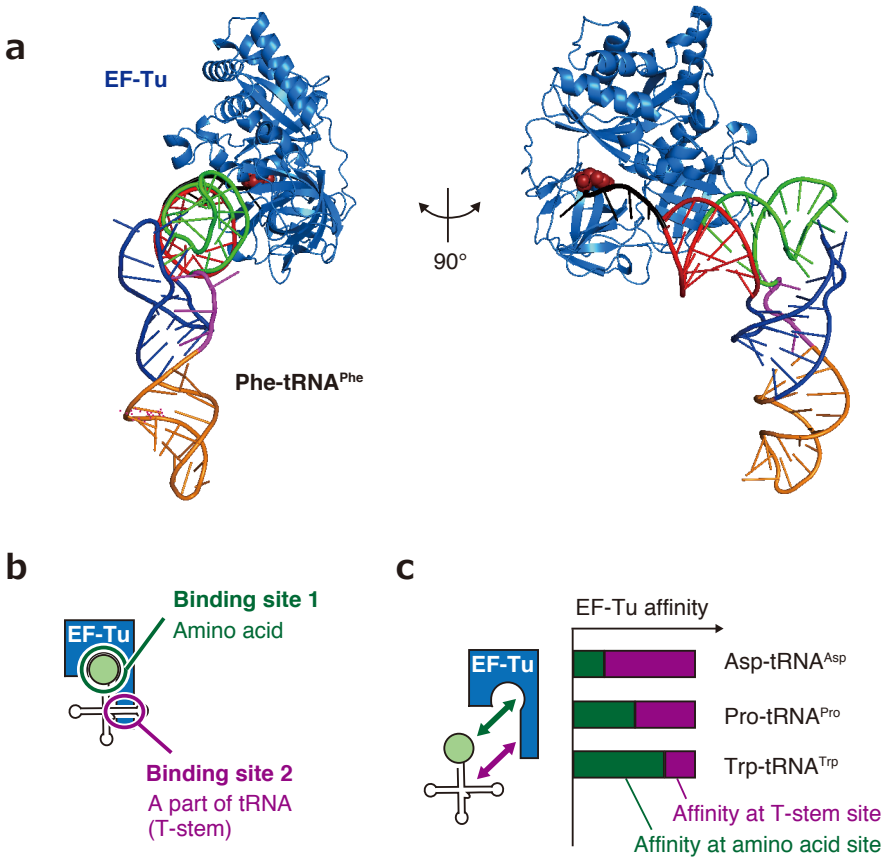

Figure 8. Complex formation of EF-Tu and aminoacyl-tRNA. a, Cocrystal structure of a ternary complex (PDB ID: 1TTT)[8]. The charged Phe is represented by red spheres. **b**, Two binding sites between EF-Tu and aminoacyl-tRNA. **c**, Compensatory relationship between affinities at the two binding sites, which contributes to the equivalent affinity between EF-Tu and each of 20 proteinogenic aminoacyl-tRNAs.

1.2.4. Decoding of the A-site codon

During decoding of the A-site codon, the anticodon loop interacts with some rRNA nucleosides as well as the codon so that a fixed anti-parallel duplex between codon and anticodon is formed[13,15,50]. Due to the fixed conformation, the first and second base pairs (N_1-N_{36} and N_2-N_{35}) are restricted to only Watson-Crick type base pairs (*i.e.* U-A, G-C, A-U, and C-G). In contrast, the base interaction between the third codon nucleotide and the first anticodon nucleotide (N_3-N_{34}) tolerates extended geometries including the Wobble base pairs such as G-U and U-G[51]. Furthermore, the 34th tRNA nucleotide exhibits a variety of nucleoside modifications, which leads to a complex variety of N_3-N_{34} base pairs (Fig. 9)[52-54]. As the modified anticodon often recognizes multiple codons via the 'extended' Wobble base pairs, all of 61 sense codons can be recognized by fewer numbers of tRNAs (around 45 tRNAs in case of *E. coli*)[55]. The 37th tRNA nucleotide is also often possess post-transcriptionally modifications[54,56]. The modifications at the 34th and 37th nucleotides synergistically help to pre-order the anticodon loop conformation so that it can fit to the decoding center with subtle structural change[57,58]. The modifications also contribute to the reading frame maintenance[35]. Surprisingly, it has been reported that reconstituted *in vitro* translation systems composed of several *in vitro* transcribed tRNAs that lack nucleoside modifications can synthesize polypeptides[59-63]. Therefore, the tRNA modifications are not essential for translation reaction.

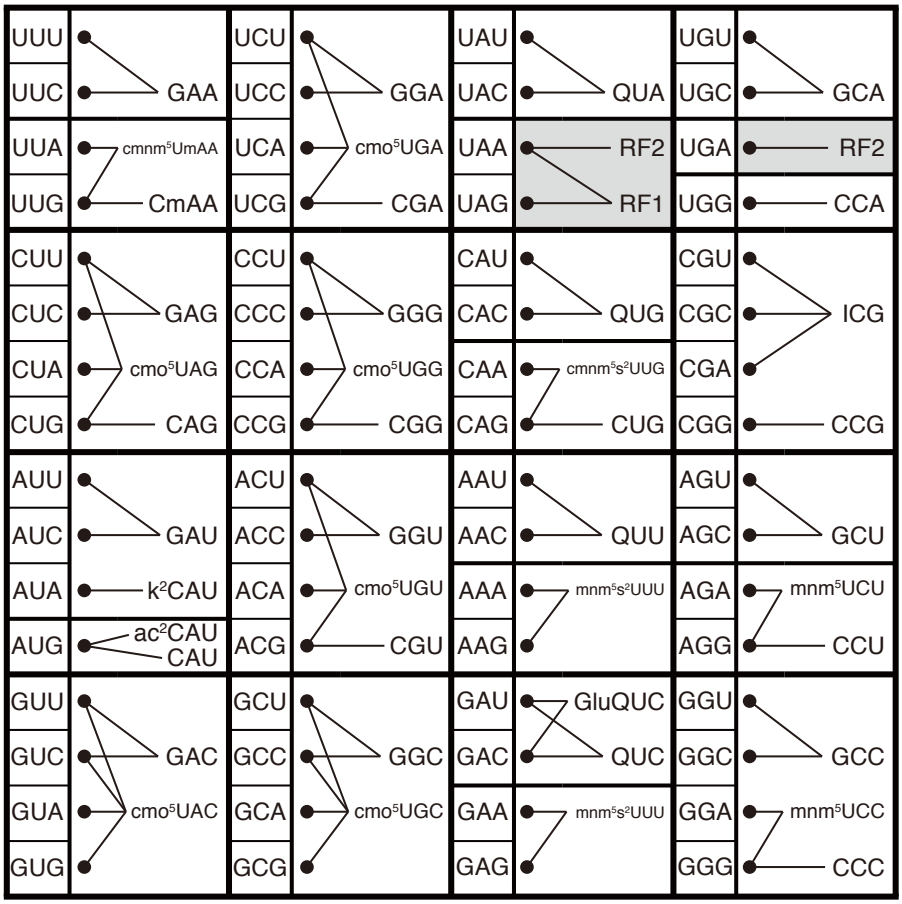

Figure 9. Redundant decoding rule via expanded Wobble base pairing. The relationships between anticodons and their corresponding codons are represented by lines. As illustrated, the 61 sense codons are decoded by around 42 anticodons via expanded Wobble base pairing. This decoding rule is based on a previous report[55].

1.3. Methods to synthesize nonproteinogenic aminoacyl-tRNAs

In native translation system, a fixed variety of 20 proteinogenic amino acids (pAAs) are generally used as the building blocks. Beyond this nature's limitation, artificial engineering of translation machinery has allowed for the ribosomal synthesis of polypeptides containing nonproteinogenic amino acids (npAAs), which have been applied widely to chemical biology and pharmacology. For example, the synthesis of proteins containing site-specific post-translational modifications, crosslinkers, optical probes, or controllable protective groups has facilitated the studies of protein structures and functions[64-66]. Also, it also facilitates the development of chemically modified "nonstandard" peptide drugs with improved pharmacological properties[67]. The incorporation of npAA into a nascent polypeptide is achieved in two steps: (1) synthesis of nonproteinogenic aminoacyl-tRNA (npAA-tRNA) and (2) engineering of the genetic code so that the code of some codons is reprogrammed to designate npAA incorporation (as described in the next section 1.4). The methods to synthesize npAA-tRNAs include (i) chemical modification of pAA-tRNAs[68,69], (ii) npAA 'mischarge' by endogenous AARSs[70,71], (iii) chemical synthesis of npAA-dinucleotides followed by its enzymatic ligation to tRNAs[72,73], (iv) npAA charge by engineered AARSs[74,75], (v) npAA charge by an artificial aminoacylation ribozyme 'flexizymes'[76,77]. Among these methods, this section focuses on two methods: (iv) npAA charge by engineered AARSs, which is the most general method applicable to both *in vivo* and *in vitro* translation, and (v) npAA charge by flexizymes, which is a versatile method applicable to *in vitro* translation.

1.3.1. npAA-tRNA synthesis by engineered AARSs

Living organisms typically use 20 orthogonal pairs of tRNA-AARS. If a novel exogenous AARS-tRNA pair specific to npAA could be introduced into a living organism, an npAA-tRNA can be generated in addition to the original set of 20 pAA-tRNAs (Fig. 10). The exogenous AARS-tRNA pair must fulfill two requirements. First, the exogenous AARS should selectively charge specific npAA to the exogenous tRNA, and it should not recognize any of 20 pAAs and endogenous tRNAs. Second, the exogenous tRNA should be recognized only by the exogenous AARS, not by any of endogenous AARSs. Such an 'orthogonal' AARS-tRNA pair specific to npAA has been established generally by three-step engineerings[74,75,78]: (1) an exogenous archaeal (or eukaryotic) AARS-tRNA pair is introduced into bacterial *E. coli*. Two dominantly-used examples are TyrRS-tRNA pair from *Methanocaldococcus jannaschii* and pyrrolysine (Pyl)-specific PylRS-tRNA pair from *Methanosarcina barkeri*. (2) As the exogenous tRNA is not completely unreactive to *E. coli*'s endogenous AARSs, several tRNA nucleotides are mutated to avoid the acylation with pAA. (3) The exogenous AARS is also mutated so that it specifically recognizes the npAA as substrates using molecular evolution approaches. It should be noted that the scope of applicable npAAs is limited to the analogs of original AARS substrate (*i.e.* Tyr analogs in case of TyrRS-tRNA pair from *Methanocaldococcus jannaschii* and Pyl analogs in case of PylRS-tRNA pair from *Methanosarcina barkeri*).

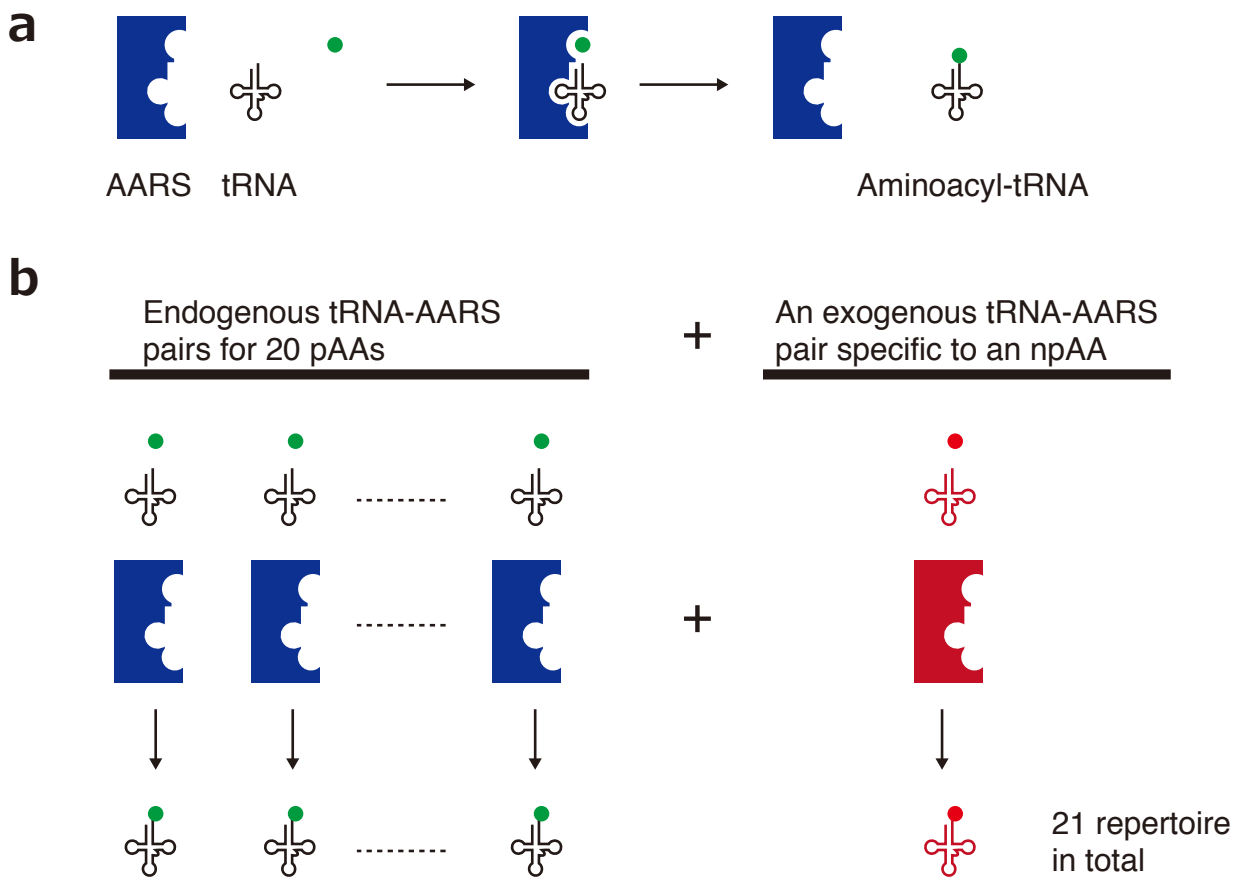

Figure 10. Expansion of the amino acid repertoire by introduction of an orthogonal tRNA-AARS pair specific to npAA. a, Schematic illustration of a canonical pAA-tRNA synthesis by an endogenous AARS. **b**, Addition of an npAA as a substrate by introduction of an exogenous tRNA-AARS pair specific to an npAA.

1.3.2. npAA-tRNA synthesis by flexizymes

Flexizymes (flexible tRNA-acylation ribozymes) are artificially developed ribozymes that allow for synthesis of npAA-tRNA using a chemically activated npAA as substrate[76,77,79] (Fig. 11a). The substrate tolerance of flexizymes is versatile: (1) As flexizymes recognize only the common 3′-terminal CCA end of a tRNA (Fig. 11b), it can charge any desired tRNAs regardless of the body sequence. (2) Flexizymes can charge virtually any desired npAA without any structural restrictions. It has been demonstrated that flexizymes are compatible with over 100 different npAAs including *N*-methyl-amino acids, *N*-acyl-amino acid, α-hydroxy-acids, D-amino acids, β-amino acids, oligopeptides, and a vast array of npAAs with unnatural side-chains[80]. Currently, three types of flexizymes are utilized (Fig. 11a)[76,81]: (1) dFx (dinitro-flexizyme) charges npAAs activated with a 3,5-dinitrobenzyl ester (DBE) regardless of the side-chain structure. (2) eFx (enhanced flexizyme) charges npAAs with aromatic side-chain activated with a cyanomethyl ester (CME). As the acylation efficiency of eFx is generally higher than dFx, it is preferable to use eFx for npAAs that possess aromatic side-chains. (3) aFx (amino flexizyme) charge npAAs activated with a 4-[(2-aminoethyl)carbomyl]benzyl thioester (ABT) regardless of the side-chain structure. aFx is especially useful for water-insoluble npAA as the ABT activation improves its water solubility. Accordingly, the best flexizyme can be chosen based on the npAA side-chain and solubility.

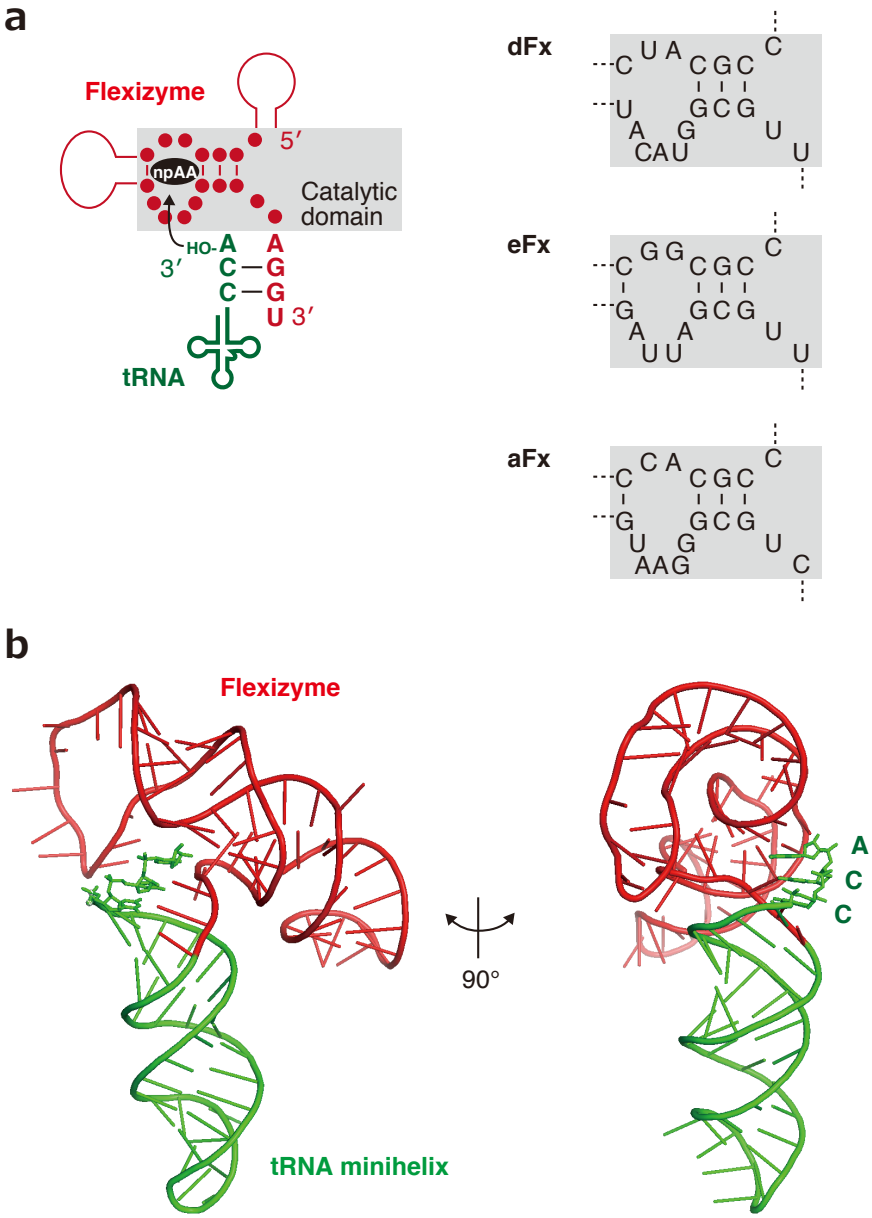

Figure 11. Synthesis of npAA-tRNA by flexizymes. a, Schematic illustration of npAA-tRNA synthesis by flexizymes. The catalytic domain sequences of dFx, eFx, and aFx are represented. **b**, Crystal structure of the flexizyme-tRNA minihelix fusion (PDB ID: 3CUL)[79]. The minihelix mimics the structure of acceptor stem and D-arm of a tRNA.

1.4. Methods to assign npAAs in the genetic code

In the genetic code, all of 64 codons are used in native translation system: 61 sense codons designate specific pAA incorporation whereas 3 stop codons designate termination of nascent polypeptide synthesis. Therefore, there is no remaining codon (*i.e.* vacant codon) that can be used specifically for npAA incorporation. To overcome this problem, researchers have developed three methods to engineer the genetic code: (1) stop codon suppression, (2) four-base codon suppression, and (3) genetic code reprogramming.

1.4.1. Stop codon suppression

In the stop codon suppression method[74,82,83], npAA is incorporated in response to a stop codon by means of npAA-tRNA that possesses the corresponding anticodon (Fig. 12). This stop codon-based method has a drawback that the desired npAA incorporation is competed with the termination reaction by a release factor(s). Recently the competition issue was overcome by the deletion of release factor one, which improved the synthetic efficiency and accuracy of full-length polypeptide containing one or two kinds of npAAs[84-87]. However, the number of assignable npAA by this method is limited up to two because only two of the three stop codons can be used for npAA incorporation at a time.

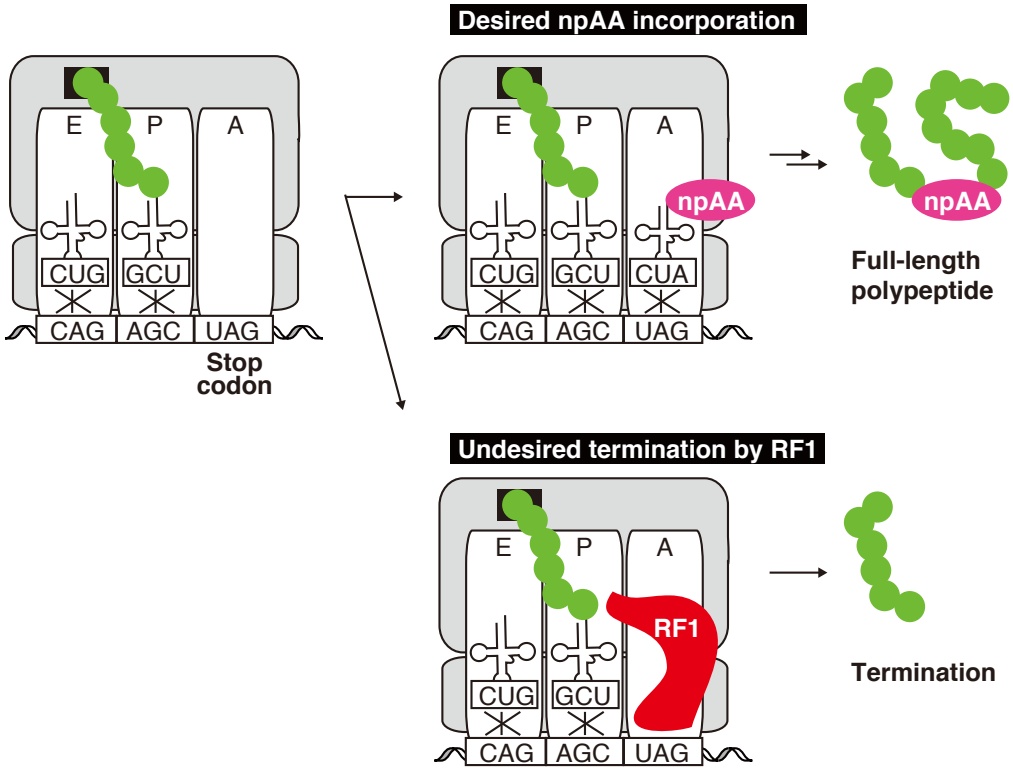

Figure 12. Incorporation of an npAA by stop codon suppression method. The desired incorporation of npAA in response to UAG stop codon leads to synthesis of full-length polypeptide. This reaction occurs under competition with polypeptide termination by means of RF1.

1.4.2. Four-base codon suppression

In the four-base codon suppression method[88-90], npAA is incorporated in response to a four-base codon by npAA-tRNA that possesses the corresponding four-base anticodon (Fig. 13). For example, an AGGU codon, designed based on a rare AGG codon, is decoded by npAA-tRNA$_{ACCU}$ under the competition with endogenous Arg-tRNA$_{CCU}$ and Arg-tRNA$_{mnm5UCU}$ (mnm^{5}U: 5-methylaminomethyluridine). The decoding of the four-base codon by npAA-tRNA$_{ACCU}$ causes a programmed frameshift that yields the full-length polypeptide, whereas undesired decoding by endogenous Arg-tRNAs results in termination of translation by a downstream stop codon. As the general problem of this method, the competition with endogenous aminoacyl-tRNAs leads to decrease in translation efficiency and accuracy. Due to this problem, it is difficult to incorporate multiple distinct npAAs simultaneously, and the best case thus far is incorporation of three npAAs with a substantially decrease in the expression level[91].

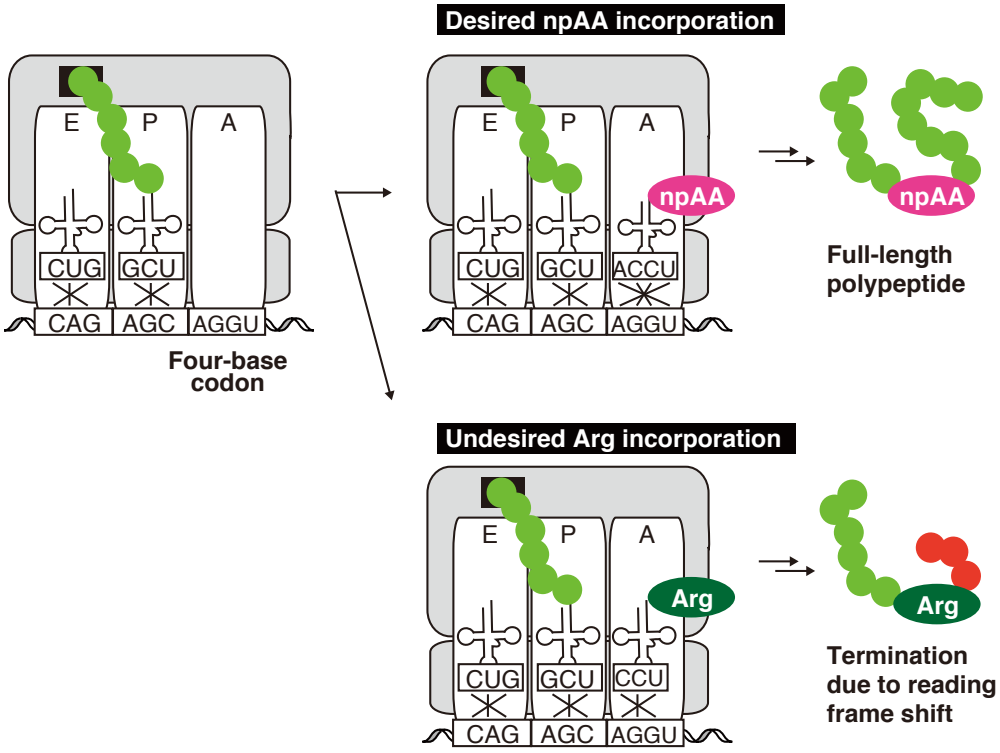

Figure 13. Incorporation of an npAA by four-base codon suppression method. While the decoding of the four-base codon by npAA-tRNA$_{ACCU}$ causes a programmed frameshift, undesired decoding by endogenous Arg-tRNAs results in termination of translation by a downstream stop codon.

1.4.3. Genetic code reprogramming method

In contrast to the two methods described above, the genetic code reprogramming method[59,60,92] allows for simultaneous use of multiple distinct npAAs by assigning them to vacant codons that are created by excluding some pAAs and the corresponding AARSs from a reconstituted cell-free translation system[32] (Fig. 14). For example, the omission of Phe makes the corresponding UUU/UUC codons vacant, which can be used for the npAA assignment by means of npAA-tRNA$_{GAA}$. However, this method has shortcomings that (1) some of 20 pAAs become unavailable for polypeptide synthesis and (2) the variety of building blocks is still restricted to 20 (or less than 20 in practice).

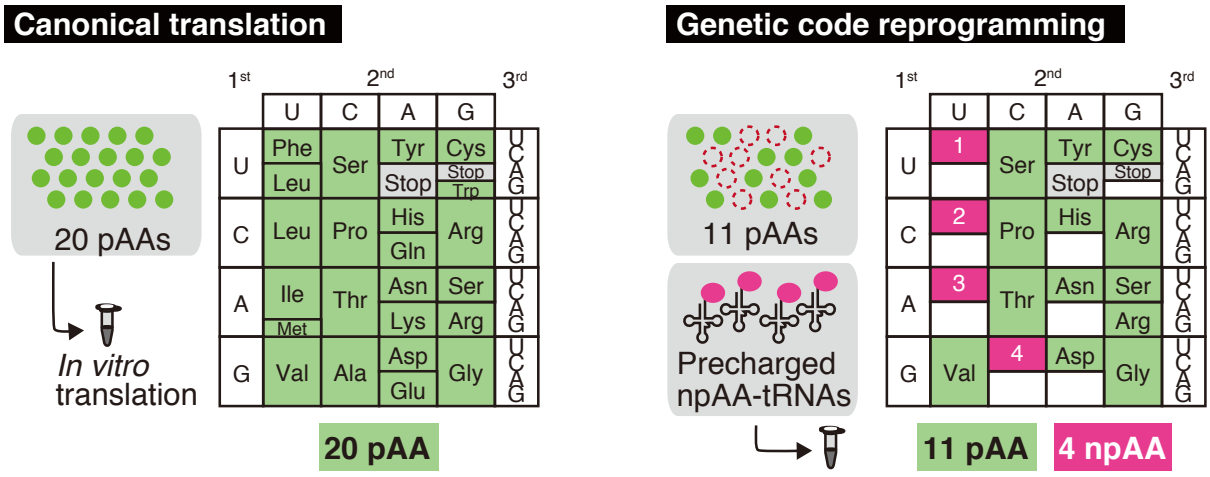

Figure 14. Encoding npAAs by genetic code reprogramming method. In contrast to a canonical translation system containing 20 pAAs, the genetic code reprogramming system contains reduced repertoire of pAAs along with precharged npAA-tRNAs. The reprogrammed genetic code achieved in a previous research is represented as an example[93].

1.5. Methods to improve the synthetic efficiency of polypeptides containing suboptimal npAA substrates

Considering that the native translation system has evolved to utilize 20 pAAs, it is a surprising fact that ribosome exhibits a vast substrate tolerance for hundreds of different npAAs[64,94]. As expected, the synthetic efficiency of npAA-containing polypeptides is generally less than canonical polypeptides, and some npAAs, especially for npAAs with huge and/or charged side-chains, cannot be incorporated at all. There are four possible obstacles for the npAA incorporation (Fig. 15). First, the npAA-tRNA might not bind to EF-Tu, which leads to inefficient delivery of npAA-tRNA into ribosome. Second, the decoding of noncanonical codons such as stop codons and four-base codons might be inefficient, which might hinder the accommodation of npAA-tRNA. Third, the peptidyl transfer reaction might be hindered when an npAA is involved, which could lead to inefficient peptide chain elongation. Fourth, the npAA-containing nascent polypeptide might not go through the exit tunnel, which would result in arrest or complete stop of translation reaction. To overcome these problems, researchers have further engineered translation systems including (1) EF-Tu mutation, (2) use of another tRNA for npAA delivery, (3) ribosome evolution, and (4) npAA engineering. These contrivances have successfully improved the synthetic efficiencies of npAA-containing polypeptides and expanded the scope of available npAAs.

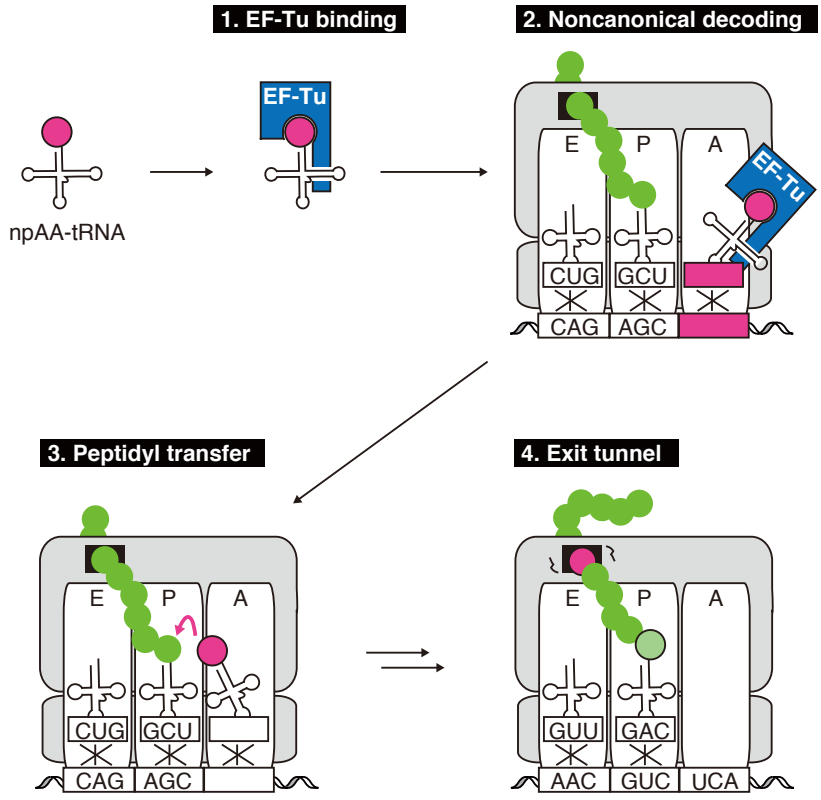

Figure 15. Four possible obstacles for the synthesis of npAA-containing polypeptides.

1.5.1. EF-Tu mutation

When an npAA-tRNA binds to EF-Tu, the npAA moiety is accommodated into a pocket composed of EF-Tu residues (Fig. 16a). In some cases, steric or electrostatic repulsion between npAA and the pocket residues hinders their complex formation. In such a case, expansion of the pocket space could allow for the binding of npAA-tRNA. For example, E215A or D216A mutations has successfully achieved EF-Tu-mediated delivery and subsequent polypeptide synthesis containing npAA with bulky side-chains such as L-1-naphtylalanine, L-2-pyrenylalanine, and fluorescent ε-*N*-Bodipy576/589-lysine[49,95] (Fig. 16b). In another example, a polypeptide containing O-phosphoserine (Sep) was successfully synthesized by mutating the pocket residues so that the electrostatic repulsion could be diminished (as the result of molecular evolution approach)[96,97].

1.5.2. Use of another tRNA for npAA delivery

Along with the above EF-Tu mutation strategy, the insufficient EF-Tu affinity could be reinforced by using another tRNA that binds to EF-Tu more strongly. For example, the use of $tRNA^{Ala}$ and $tRNA^{Glu}$ instead of $tRNA^{Phe}$ and $tRNA^{Asn}$, respectively, enabled npAA-tRNAs to bind to EF-Tu and improved their incorporation efficiencies[61,98]. It should be noted that this method using the body sequences of native tRNAs, such as $tRNA^{Ala}$ and $tRNA^{Glu}$, is applicable only when the corresponding pAAs are excluded from a reconstituted translation system, which strictly limits its application scope. In another example, the molecular evolution of tRNA body sequence was conducted[99], in which a sequence-randomized tRNA library was prepared and the tRNAs that can deliver the npAA efficiently can be selectively recovered. As the result, the incorporation efficiencies of *p*-benzoylphenylalanine and o-nitrobonzyltyrosine were successfully improved by 20-fold although the mechanism of improvement has not been elucidated.

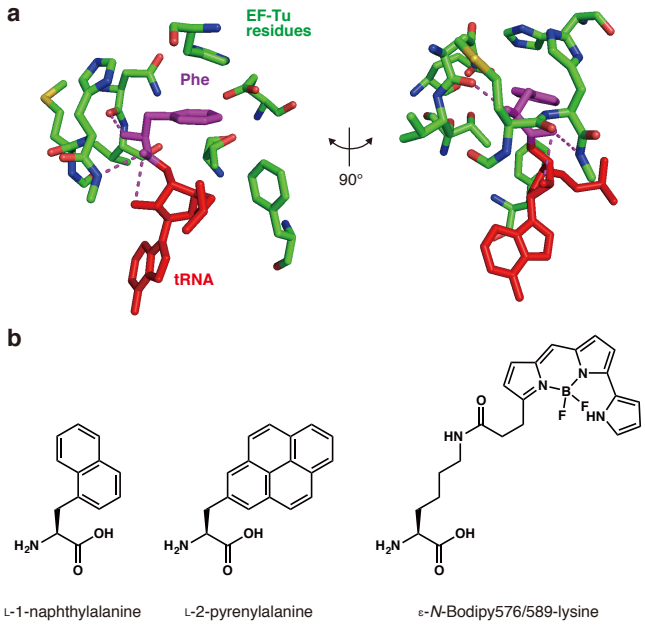

Figure 16. Improvement of npAA incorporation efficiencies by EF-Tu mutation. a, Structure of the EF-Tu pocket accommodating charged phenylalanine of Phe-$tRNA^{Phe}$ (PDB ID: 1TTT)[8]. Some of these EF-Tu residues are mutated so that it can accommodate bulky side-chains of npAAs. **b**, Examples of npAAs that could be incorporated into nascent polypeptides by EF-Tu mutation strategy.

1.5.3. Ribosome evolution

In order to improve synthetic efficiency of npAA-containing polypeptide, researchers have engineered rRNA sequences that consist either the decoding center or the peptidyl transferase center via molecular evolution strategy. The engineering of decoding center includes the development of mutated ribosomes named ribo-X and ribo-Q that can efficiently incorporate npAAs in response to stop and four-base codon, respectively[83,90] (Fig. 17a). The engineering allowed for incorporation of two distinct npAA in a protein and intramolecular Click reaction between the two npAAs (*p*-azido-L-phenylalanine and *N6*-[(2-propynyloxy)carbonyl]-L-lysine)[90]. Whereas, the engineering of peptidyl transferase center, has successfully improved incorporation efficiencies of D-amino acids and β-amino acids[100,101] (Fig. 17b).

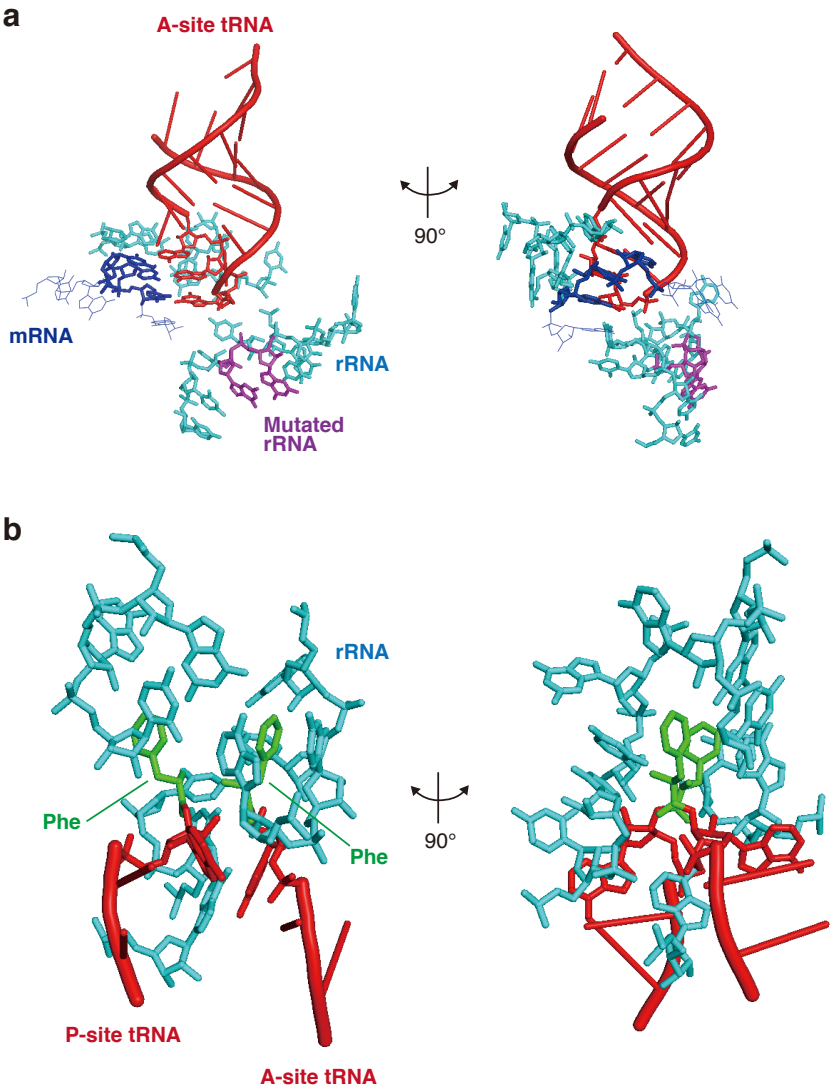

Figure 17. Ribosome evolution for efficient incorporation of npAAs. a, Decoding center of a ribosome (PDB ID: 2J00, 2J01)[102]. The codon and anticodon nucleosides are represented as stick image, whereas the rRNA nucleosides mutated as a result of ribosome evolution are represented in purple. **b**, Peptidyl transferase center of a ribosome (PDB ID: 2WDK, 2WDL)[20]. Some of these rRNAs are mutated to improve the synthetic efficiency of polypeptides containing npAAs.

1.5.4. npAA engineering

Another approach to expand the scope of applicable npAAs is to chemically or enzymatically convert the npAA structure after translation reaction. For example, the synthesis of polypeptides containing charged *N*-alkyl-amino acids has been achieve by post-translational chemical conversion from the charge-masked npAAs to charged side-chains[103]. For example, azide and ester groups can be converted to amino and carboxylic acid groups, respectively (Fig. 18a and b). This method is practical because the direct incorporation of the charged *N*-alkyl-amino acids is incapable[104]. Similarly, polypeptides containing npAAs with either alkyne or azide groups allow for post-translational side-chain modification by Click reaction, which can extensively expand the scope of modification structures[66,105] (Fig. 18c).

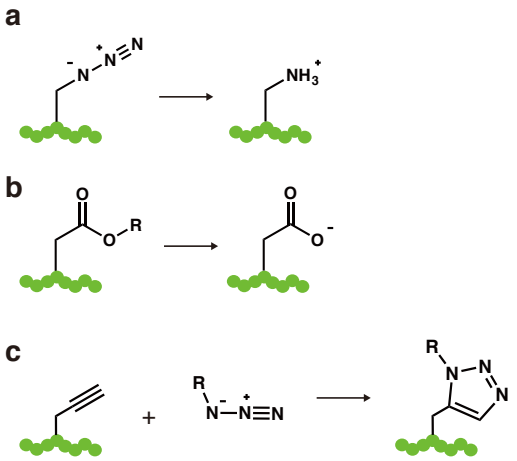

Figure 18. Post-translational conversion of npAA structure. a,b, Post-translational formation of amino group (**a**) and carboxylic acid group (**b**). **c**, Post-translational side-chain modification by Click chemistry.

1.6. Development of nonstandard macrocyclic peptide drugs

The use of npAAs has enabled us to create a variety of innovative biotechnologies, among of which our laboratory focuses on development of selection technologies to discover chemically modified peptide-based drugs with improved pharmacological properties by the use of npAAs.

1.6.1. Small-molecule drugs and protein-based drugs

In drug development studies, small molecules and proteins are dominantly used as the molecular bases[106,107]. These two classes of molecules exhibit the opposite advantages and disadvantages in respect of binding properties and membrane permeability. Regarding binding properties, small molecules can bind only to pockets or clefts of target proteins in general (Fig. 19a), which limit the range of targetable proteins: only ~10% of proteins are estimated to be targetable by small molecules[108,109]. In contrast, protein-based drugs can bind to surface of target proteins with large area (Fig. 19a), which improves their binding affinity, selectivity, and the range of targetable proteins[110]. In the aspect of membrane permeability, most small molecules that satisfy the Lipinski's rule of five can penetrate cell membrane passively[111], and therefore they are orally available and can target both intracellular and extracellular proteins (Fig. 19b). In contrast, protein-based drugs are not membrane permeable, and therefore they need to be taken by injection and can target only extracellular proteins (Fig. 19b).

1.6.2. Noncanonical structures observed in natural bioactive peptides

Peptide-based molecules could combine the large-surface binding of protein-based drugs and the membrane permeability of small molecules. In fact, living organisms synthesize such bioactive peptides[67,112,113]. Fig. 19c represents some examples of clinically available peptide-based drugs that have been developed from naturally occurring bioactive peptides. Notably, these practical peptide drugs possess common nonproteinogenic features: macrocyclic structure, backbone *N*-methylation, D-configuration, and nonproteinogenic side-chains. These structural characteristics confer structural rigidity, high target-binding affinity, membrane permeability, and resistance to enzymatic degradation, which lead to improved pharmacokinetics properties[114-124].

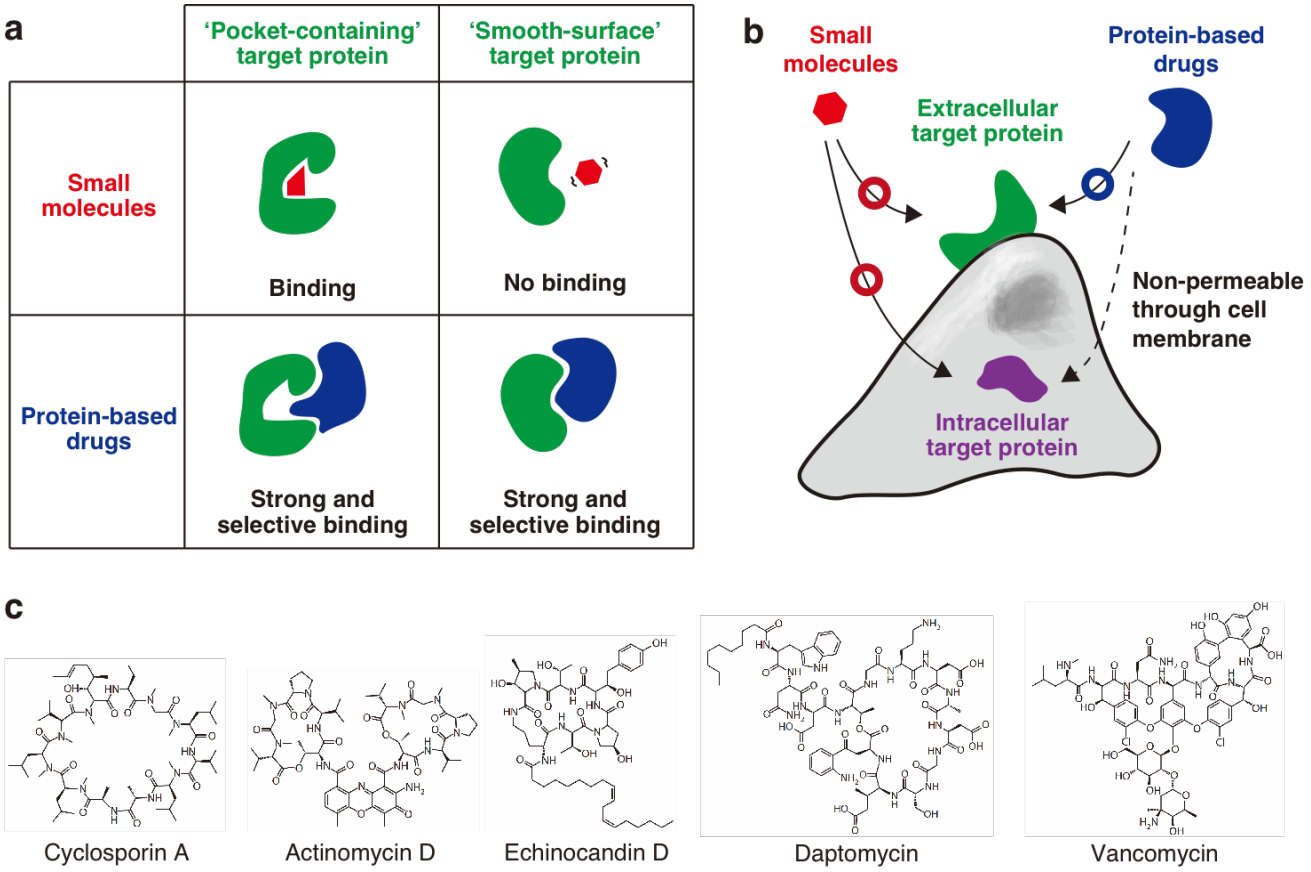

Figure 20. Three classes of drug molecules. **a**, Comparison between small molecules and protein-based drugs in respect of target-binding property. **b**, Comparison between small molecules and protein-based drugs in respect of cell-membrane permeability. **c**, Examples of clinically available peptide-based drugs.

1.6.3. Selection strategies to discover novel peptide drugs

Peptide-based drug scaffolds are attractive because peptides can bind to target proteins with large surface area, which achieves high binding affinity and selectivity. Also, small peptides have a potential to penetrate cell membrane passively, which is impossible for larger protein-based drugs. Due to such appealing properties of peptide-based drugs, researchers have developed screening technologies, such as mRNA display[125,126], in which target-binding peptides can be selectively recovered from a mass population of sequence-randomized peptide library. In predominant cases, peptide libraries are prepared by ribosomal synthesis because this method can yield a huge diversity of sequence-randomized peptides that overwhelms that of chemical peptide synthesis (Typically 10^{12}, 10^{9}, and 10^{6} diversity in case of *in vitro* translation, phage-based expression, and chemical synthesis, respectively).

Although potent peptide binders have been successfully discovered by the screening methods, it is still challenging to develop practical peptide drugs with sufficient pharmacokinetic properties: First, canonical peptides composed of 20 pAAs are instable against enzymatic degradation and cannot retain their structures in human blood or organs. Second, canonical peptides are not cell-membrane permeable in general due to its rich hydrogen bonding property, and therefore it is difficult to develop orally available peptide drugs or peptide drugs that target intracellular proteins. In this context, it is preferable to develop peptide drugs with chemical modifications that improve their pharmacokinetic properties.

To overcome these problems, Suga's group has developed the engineered translation system, named FIT (Flexible *In vitro* Translation) system, to ribosomally synthesize 'nonstandard' peptides with noncanonical scaffolds that mimic the characteristics of naturally occurring bioactive peptides[77,92,104,127-129]. The FIT system relies on a reconstituted *E. coli* cell-free translation system[32] supplemented with npAA-tRNAs prepared by flexizyme-catalyzed aminoacylation. Owing to the flexizyme's versatility, the FIT system has successfully expresses nonstandard peptides containing *N*-methyl-amino acids, D-amino acids, unnatural side-chains, and macrocyclic structure. Furthermore, they combined the FIT system with the mRNA display method and have discovered nonstandard peptide-based drugs with potent binding properties[93,130-136]. The selection system named RaPID (Random nonstandard peptides integrated discovery), in brief, consists of the following manipulations (Fig. 20a). (1) A sequence-randomized DNA library is prepared by primer extension and PCR. (2) An mRNA library is synthesized by *in vitro* transcription of the DNA library. (3) An npAA-containing peptide library is synthesized by means of the FIT system. The synthesized peptides are covalently linked to their mRNA templates via puromycin linker. The peptide can be macrocyclized by spontaneous reaction between the *N*-terminal chloroacetyl group and downstream cysteine thiol (Fig. 20b). (4) The target-binding peptides are recovered using a target protein immobilized to a solid phase. (5) The mRNAs covalently attached to the recovered peptides are amplified by reverse transcription, PCR, and *in vitro* transcription. This cycle of RaPID selection is repeated typically 4–6 times, and the sequences of target-binding peptides are identified by DNA sequencing. To date, RaPID screening technology has yielded nonstandard peptides that can bind to over 20 proteins with high affinities (low nM K_d).

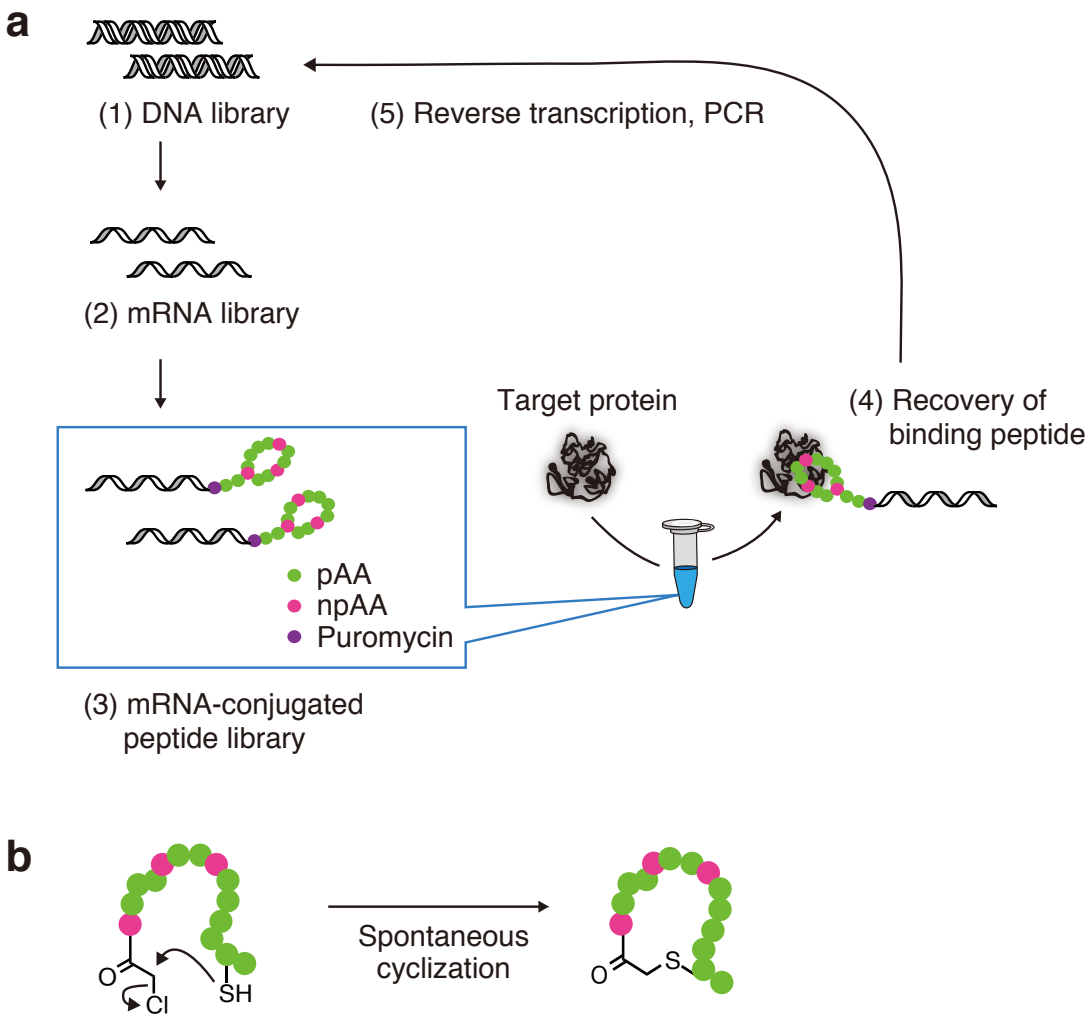

Figure 21. Schematic illustration of peptide selection by the RaPID system. a, Manipulations of the RaPID system. **b**, Formation of the macrocyclic structure by spontaneous reaction between the *N*-terminal chloroacetyl group and downstream cysteine thiol.

1.6.4. Examples of macrocyclic nonstandard peptides discovered by the RaPID system

Here I describe three examples of peptide drugs discovered by the RaPID system. The first example is macrocyclic peptide inhibitors of a deacetylase named SIRT2[131,132] (Fig. 21a). These peptides contain ε-*N*-trifluoroacetyllysine as a mechanism-based warhead: the noncanonical side-chain mimicking the canonical substrate (*i.e.* acetyllysine) directly binds to the catalytic pocket of SIRT2 and inhibits its activity, as demonstrated by X-ray crystallography of the complex. Interestingly, the peptide has a rigid structure via multiple intra-molecular hydrogen bonds and exhibits several nM of K_d and IC_{50} with 10-fold selectivity against SIRT1 and SIRT3.

The second example is macrocyclic peptides that work as co-crystallization probe of a MATE multidrug transporter derived from *P. furiosus* (PfMATE)[133,135] (Fig. 21b). Also in this case, the discovered peptides possessed multiple intra-molecular hydrogen binds, which facilitated their stable settlement inside the cleft of PfMATE. The peptides successfully improved the resolution by fixing the PfMATE conformation, and also they gave insights into the mechanism of multidrug transporter.

The third example is macrocyclic peptide-based PPI (protein-protein interaction) inducer[136] (Fig. 21c–e). In living organisms, HGF (hepatocyte growth factor) transmits its signal by dimerizing PTKs (receptor tyrosine kinases) named Met. Mimicking this mechanism, Ito *et al.* developed Met-binding macrocyclic peptides and synthesized the homodimer of peptides with covalent linkers. This dimer-macrocycle could dimerize two Met receptors and successfully activated their downstream signal transduction.

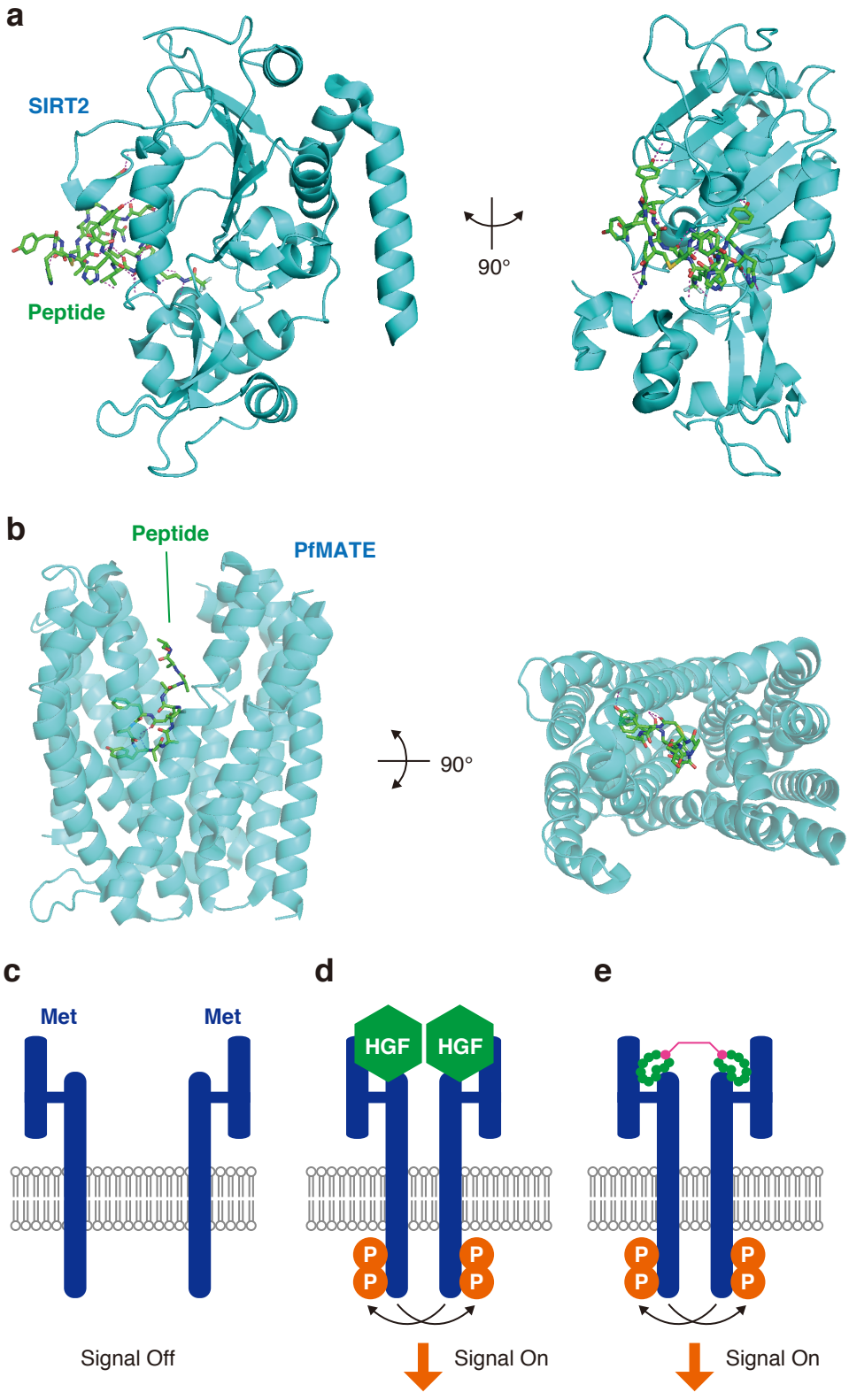

Figure 22. Examples of bioactive nonstandard peptides discovered by the RaPID system. a, Cocrystal structure of SIRT2 and its inhibitor macrocyclic peptide (S2iL5) (PDB ID: 4L3O)[132]. **b**, Cocrystal structure of PfMATE and a macrocyclic peptide (MaD5) that works as a co-crystallization probe (PDB ID: 3VVR)[133]. **c–e**, Receptor activation by homodimer macrocyclic peptides with a covalent linker. The inactive Met receptor (**c**) is activated by HGF-mediated dimerization (**d**). The peptide-based dimer-macrocycle can mimic this activation mechanism and transmit the downstream signaling (**e**). 'P' represents the phosphorylation of tyrosine residues. The phosphorylated receptor can interact with other intra-cellular proteins, which transmit the downstream signaling.

1.7. Limitations of the conventional engineered translation systems.

Although the use of engineered translation systems including the FIT system enabled us to synthesize nonstandard peptides containing a couple of npAAs, the conventional methods still have limitations in repertoire of building blocks (*i.e.* pAAs and npAAs) that can be used simultaneously[93]. Therefore, it is still challenging to synthesize highly modified nonstandard peptides like cyclosporine that contains multiple distinct npAAs. Accordingly, we cannot discover peptide drugs with sufficient pharmacokinetic characters such as cell-membrane permeability and proteolytic resistance in most cases.

More specifically, the conventional engineered translation systems have methodological limitations as follows: (1) Accurate synthesis of a polypeptide containing more than two npAAs can be achieved only when several proteinogenic amino acids are excluded from *in vitro* translation systems to create vacant codons assignable for npAAs, resulting in the decreased building block repertoire[93]. (2) Even when such an engineered translation system is used, it is impossible to efficiently synthesize polypeptides containing many kinds of *N*-methyl amino acids (^{Me}AAs) due to an unspecified translational disorder. In this context, there is a need to develop *in vitro* translation systems enabling (1) expansion of the building block repertoire by utilizing multiple npAAs in addition to the 20 proteinogenic ones and (2) the synthesis of nonstandard peptides containing a variety of ^{Me}AAs. As the Ph.D. degree research, I studied to develop these two kinds of engineered translation systems.

Chapter 2

Expanding the amino acid repertoire of ribosomal polypeptide synthesis via artificial division of codon boxes

2.1. Introduction

Ribosomal synthesis of the small peptides composed of both proteinogenic and nonproteinogenic amino acids (pAAs and npAAs, respectively) has revolutionized drug discovery and chemical biology[67,137-140]. For example, thioether-macrocyclic peptides, including *N*-methyl-peptides, that strongly and selectively bind to target proteins have been discovered[93,130,131,133] by the RaPID (Random nonstandard peptides integrated discovery) system that combined an mRNA display method[125,126] with an *in vitro* translation method, named FIT (flexible *in vitro* translation) system[77,79,92,104,127-129]. The FIT system relies on a reconstituted *E. coli* cell-free translation system[32] supplemented with the *E. coli* total tRNA mixture and synthetic tRNA molecules charged with npAAs by flexizyme-catalyzed aminoacylation[92]. As all the 64 codons in the genetic code are assigned to 20 kinds of pAAs and three stop signals, vacant codons for npAAs have been prepared by omitting a few pAAs and their specific aminoacyl-tRNA synthetases (AARSs) from the reaction mixtures[59,60,77]. Therefore, there has been a contradiction that increasing the number of npAAs assigned decreases the number of pAAs assigned from the standard 20.

One of alternative approaches to expand the amino acid repertoire is the stop codon suppression method. In this method, a stop codon, such as UAG amber codon, designates the incorporation of npAA by means of npAA-tRNA$_{CUA}$[74,82]. However, the number of assignable npAAs in this method should still be limited to two because only two of the three stop codons can be overwritten at a time. A method of programmed frameshift suppression using four-base codons is another alternative to or can be combined with the stop codon suppression method for the incorporation of a few kinds of npAAs[88-91]. Although the method of programmed frameshift suppression conceptually enables us to encode more than two npAAs, the best and single case thus far reported is that three npAAs are incorporated simultaneously with a substantial decrease in the expression level because of the competitive misincorporation of pAAs by certain endogenous pAA-tRNAs[91]. Collectively, the methods of genetic code expansion mentioned above have a severe limitation on the number of assignable npAAs. Accordingly, there is a need to reconstitute the translation system *in vitro* to overcome this limitation as well as the contradiction mentioned above.

To encode multiple npAAs without sacrificing any of the 20 pAAs, I aimed to reprogram multiple sense codons. Although as many as 61 codons can code for specific amino acids, only 20 pAAs are generally assigned to these codons (Fig. 22a–c). This means that many codons are redundantly used for the assignment of the same amino acids. In *E. coli* and most other organisms, fewer than 45 types of tRNAs are used[33], and therefore some tRNA species decode multiple cognate codons via 'expanded' wobble base pairing mechanisms[35,51,141] (Fig. 22d). For example, in case of the Val codon box, two species of $tRNA^{Val}_{NAC}$ are responsible for decoding the GUU, GUC, GUA, and GUG codons (Fig. 23a, Val). The $tRNA^{Val}$ with GAC anticodon decodes two Val codons (GUU and GUC); the $tRNA^{Val}$ with cmo^5UAC (cmo^5U: 5-oxyacetic acid) anticodon decodes all four Val codons. In the case of arginine $tRNA^{Arg}_{NCG}$, the $tRNA^{Arg}$ with ICG (I: inosine) anticodon decodes three Arg codons (Fig. 23a, Arg); the $tRNA^{Arg}$ with CCG anticodon decodes a single codon, CGG. Evolution of the translation system has made the sophisticated decoding mechanism and it has achieved the high accuracy in translation reaction. An intriguing question is whether it is capable to reduce the redundancy of the genetic code and create vacant codons that can be used for the reassignment of npAAs.

Here, I report a conceptually new method to encode multiple distinct npAAs and the full set of pAAs by reducing the redundancy of the genetic code, referred to as the 'artificial division of codon boxes'. To achieve this, I first supplemented a native tRNA-free *in vitro* translation system with 32 *in vitro* transcripts of $tRNA_{SNN}$ (S = G or C) (FIT-32t system). These 32 tRNA transcripts can be charged with 20 pAAs by endogenous AARSs and independently decode the corresponding 31 NNS elongation codons as well as the AUG initiation codon, which covers the assignment of 20 pAAs (Fig. 23b). As 19 of these $tRNA_{SNN}$ molecules redundantly assign eight pAAs to the corresponding NNS codons, it is capable to omit some of the redundant $tRNA_{SNN}$'s from the FIT-32t system and replace them with the corresponding $tRNA^{AsnE2}_{SNN}$'s, which have a body sequence orthogonal to endogenous AARSs[77,93,127]. When these $tRNA^{AsnE2}_{SNN}$'s are pre-charged with different npAAs (Fig. 22e) by flexizymes, the corresponding codons can be reassigned to npAAs without abandoning any of 20 pAAs (Fig. 23c and 24). I have utilized this methodology to assign three distinct npAAs to the vacant codons and demonstrated the expression of a 32-mer peptide composed of 23 building blocks (three npAAs and the 20 pAAs) as well as a macrocyclic *N*-methyl-peptide that inhibits a ubiquitin ligase E6AP[93]. Our results supported the idea that this new method can replace the conventional method for constructing macrocyclic *N*-methyl-peptide libraries for the discovery of bioactive ligands against drug targets. This is the first report on the extensive artificial division of codon boxes and the first step towards making the vacant codons for up to nearly a dozen of npAAs.

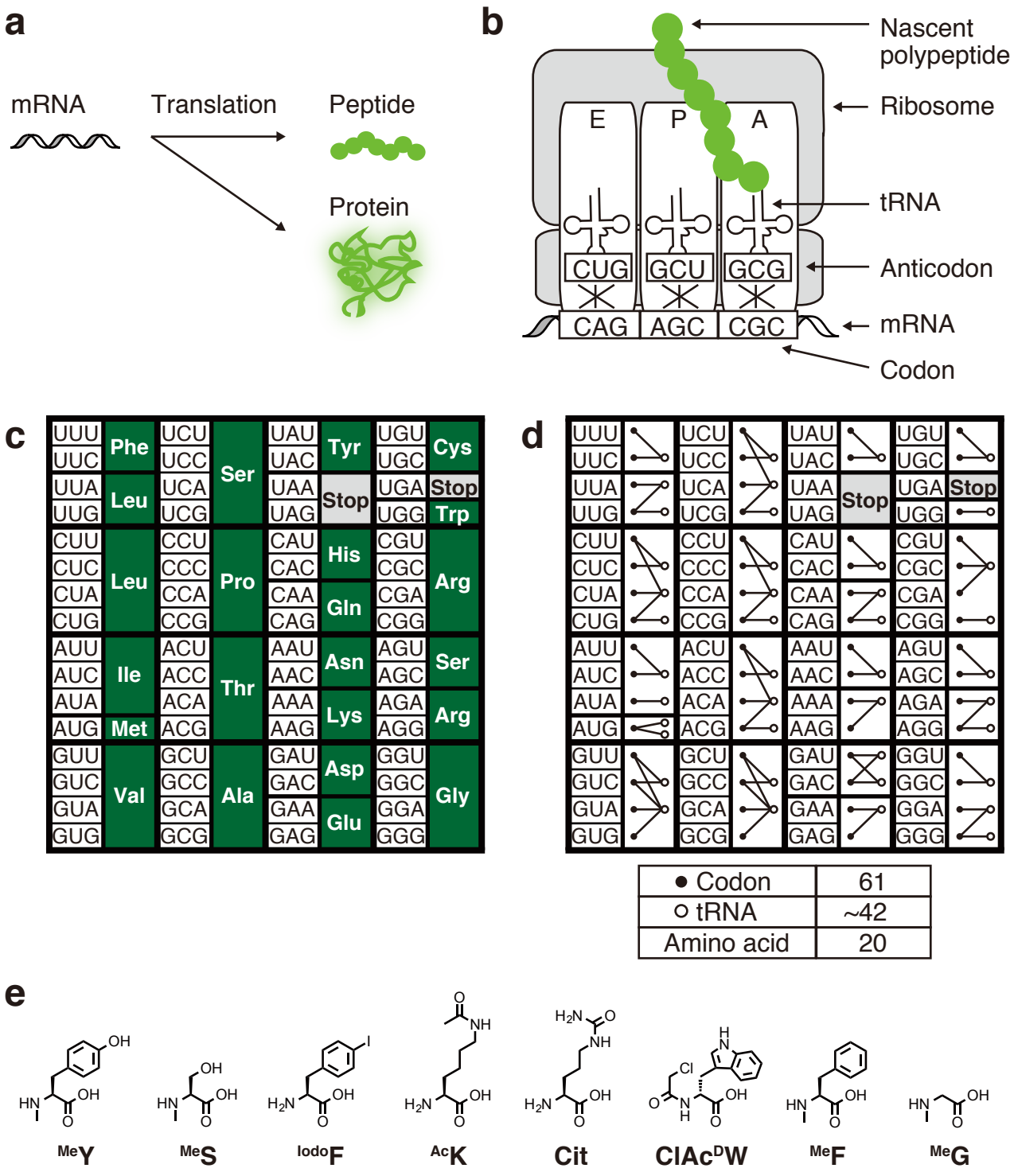

Figure 22. Redundant decoding observed in native translation reaction. a, Synthesis of polypeptides via translation of mRNAs. **b**, Schematic illustration of ribosomal polypeptide synthesis where tRNA brings the specific amino acid in response to the corresponding codon. The anticodon sequence of a tRNA decodes the corresponding codon by forming base pairs. **c**, The genetic code. 61 sense codons code for 20 proteinogenic amino acids redundantly. **d**, The redundant and complex decoding by native tRNAs via expanded Wobble base pairing. As illustrated, most native tRNAs decode multiple codons. Filled and empty circles indicate codons and tRNAs, respectively. The corresponding anticodons are represented in Fig. 9. **e**, npAAs used in this study. ^{Me}Y: *N*-methyltyrosine, ^{Me}S: *N*-methylserine, ^{Iodo}F: p-iodophenylalanine, ^{Ac}K: ε-*N*-acetyllysine, Cit: citrulline, ClAc^{D}W: *N*-chloroacetyl-D-tryptophan, ^{Me}F: *N*-methylphenylalanine, ^{Me}G: *N*-methylglycine.

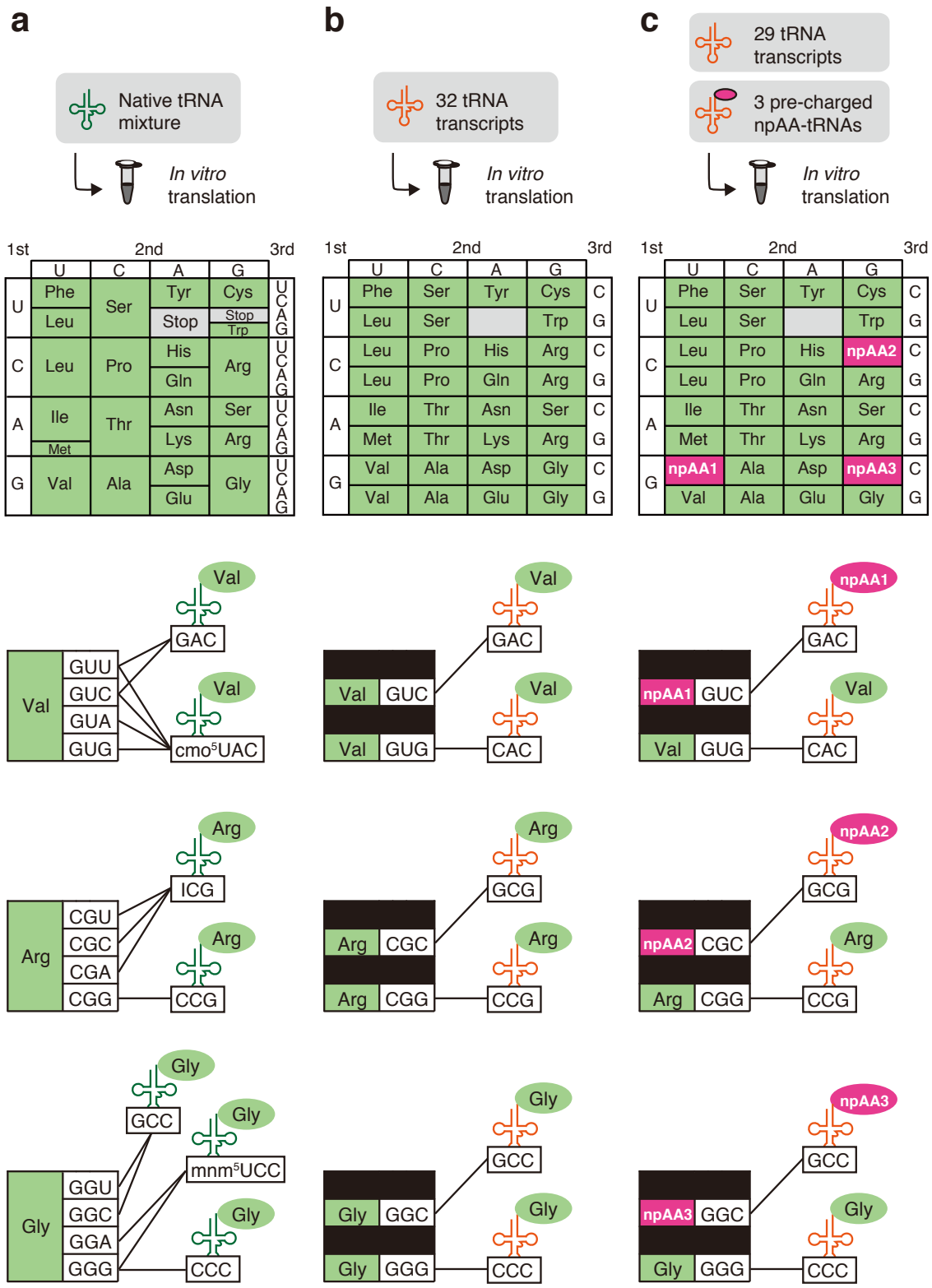

Figure 23. Schematic illustration of artificial division of codon boxes. a, The genetic code of a native translation system. The Val, Arg and Gly family codon boxes are decoded by several native tRNAs via Watson-Crick or wobble base pairings as illustrated. **b**, A genetic code reprogrammed in the FIT-32t system. In this system, native tRNAs are replaced with 32 *in vitro* tRNA transcripts that decode 31 NNS sense codons along with the AUG initiation codon. **c**, A reprogrammed genetic code with reduced redundancy that enables the assignment of three distinct npAAs in addition to 20 pAAs. The Val, Arg, and Gly codon boxes are divided artificially by replacing the redundant $tRNA^{Val}_{GAC}$, $tRNA^{Arg}_{GCG}$, and $tRNA^{Gly}_{GCC}$ with three npAA-$tRNA^{EnAsn}_{GNN}$'s prepared by flexizyme-mediated aminoacylation. The UAG stop codon is vacant in this study because RF1 is omitted in the FIT system. The UAA and UGA codons, however, designate termination by means of RF2.

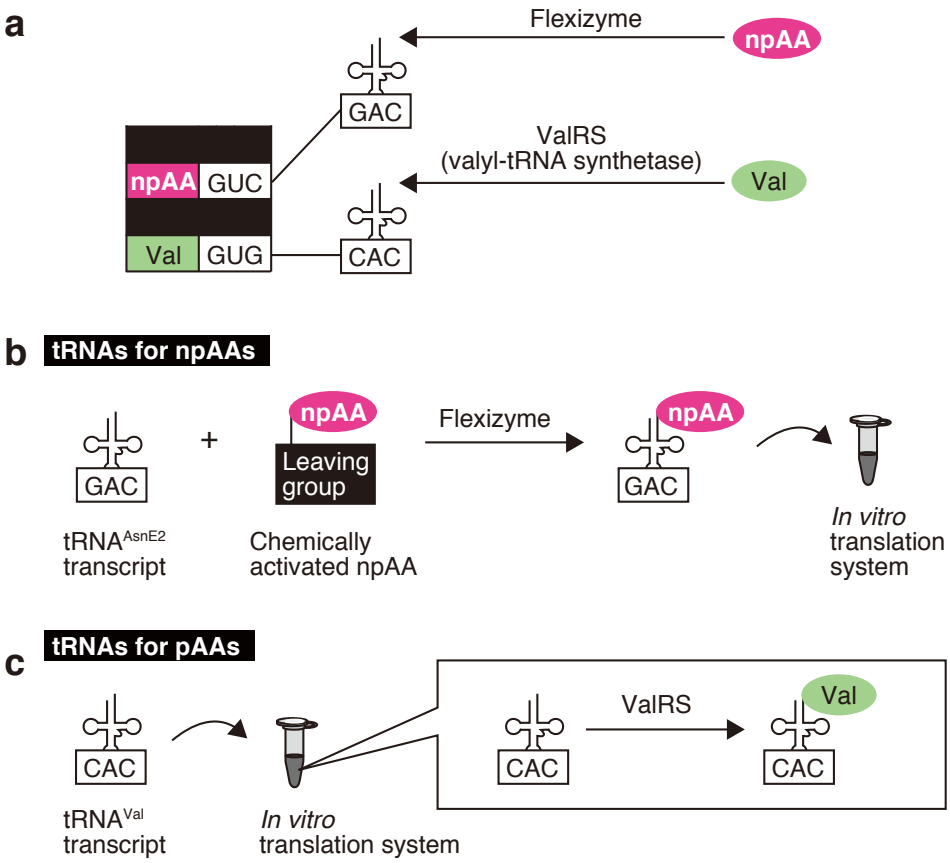

Figure 24. Methods to charge npAA and pAA to tRNAs. a–c, Artificial division of Val codon box. $tRNA^{AsnE2}_{GAC}$ is charged with npAA by means of flexizymes, and the resultant pre-charged npAA-tRNA is added into the translation mixture (**b**), whereas $tRNA^{Val}_{CAC}$ is charged with Val by ValRS *in situ* during translation reaction (**c**).

2.2. Results and discussion

2.2.1. Accurate decoding by *in vitro* tRNA transcripts

To achieve our goal of 'artificial division of codon boxes', I first tested whether the respective 32 *in vitro* $tRNA_{SNN}$ transcripts could decode the corresponding codons and incorporate the specific pAAs with sufficient efficiency and accuracy. This examination was required because the use of tRNA transcripts lacking post-transcriptional nucleoside modifications could lead to inefficiency or inaccuracy in acylation and/or decoding steps (Fig. 25). The respective $tRNA_{SNN}$'s with representative naturally occurring body sequences were prepared by *in vitro* transcription (Supplementary results 2.4.1). Among these 32 $tRNA_{SNN}$'s, $tRNA^{Ini}_{CAU}$ transcript decodes the initiation AUG codon whereas the other 31 $tRNA_{SNN}$ transcripts, including $tRNA^{Met}_{CAU}$, decode the NNS elongation codons except for the UAG stop codon.

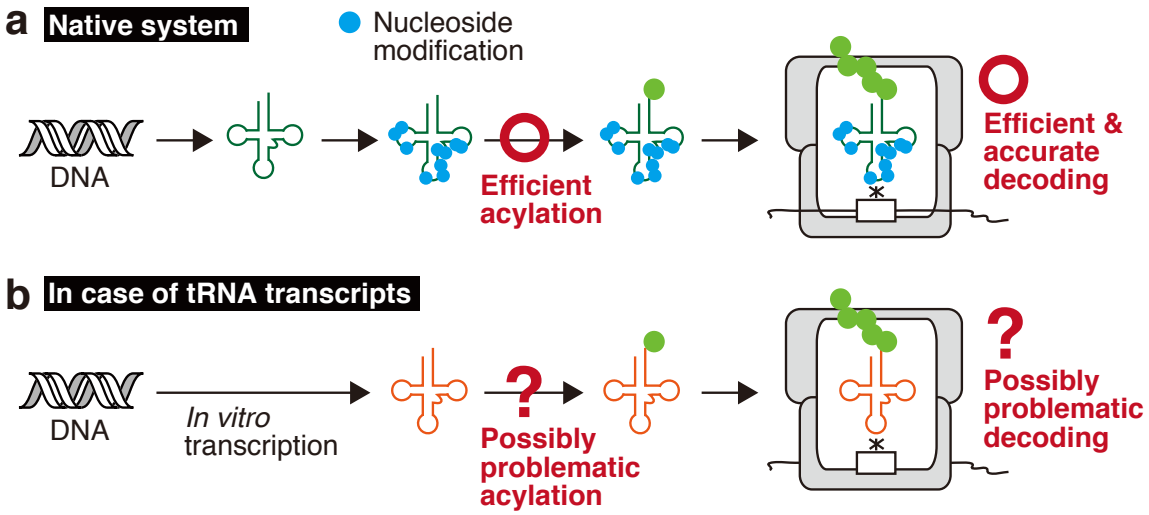

Figure 25. Possible obstacles due to the use of tRNA transcripts that lack nucleoside modifications. a, The correct activities of native tRNAs containing post-transcriptional nucleoside modifications. These tRNA modifications contribute to the efficient and accurate reaction during acylation and decoding steps. **b**, Possible problems of tRNA transcripts. As *in vitro* transcribed tRNAs do not contain any nucleoside modifications, their activities during acylation and decoding might become inefficient or inaccurate.

The activities of 32 tRNA transcripts were examined by expression of a series of short model peptides (Fig. 26a). The mRNA templates were designed to determine the peptide expression level by means of tricine-SDS-PAGE (autoradiographic detection of [^{14}C]-Asp in the C-terminal FLAG peptide region), and accuracy of the decoding was evaluated by MALDI-TOF-MS (Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry) of translation products. To simplify the analysis, I prepared a 'minimal' FIT system that contained only four kinds of tRNA transcripts ($tRNA^{Ini}_{CAU}$, $tRNA^{Tyr}_{GUA}$, $tRNA^{Asp}_{GUC}$, and $tRNA^{Lys}_{CUU}$) (Fig. 26b). 29 templates form the $mRNA1_{NNS*}$ set (Fig. 26a) were translated in the minimal FIT system supplemented with the corresponding tRNA transcript in appropriate concentration (Supplementary results 2.4.2) (Fig. 26b). In nearly all cases a single band that corresponded to the desired peptide was observed with similar expression levels, with exception in the Cys lane where multiple bands appeared due to disulfide bond formation with thiol-containing molecules. The activities of $tRNA^{Glu}_{CUC}$ and $tRNA^{Gly}_{CCC}$ were confirmed using $mRNA2_{NNS*}$ templates instead of $mRNA1_{NNS*}$ (Fig. 26a,d). The mRNA sequences were changed because $mRNA1_{NNS*}$'s sequence-dependent poor expression was observed even in the presence of *E. coli* native tRNA mixture, which could be solved by using $mRNA2_{NNS*}$ as templates (Fig. 26e).

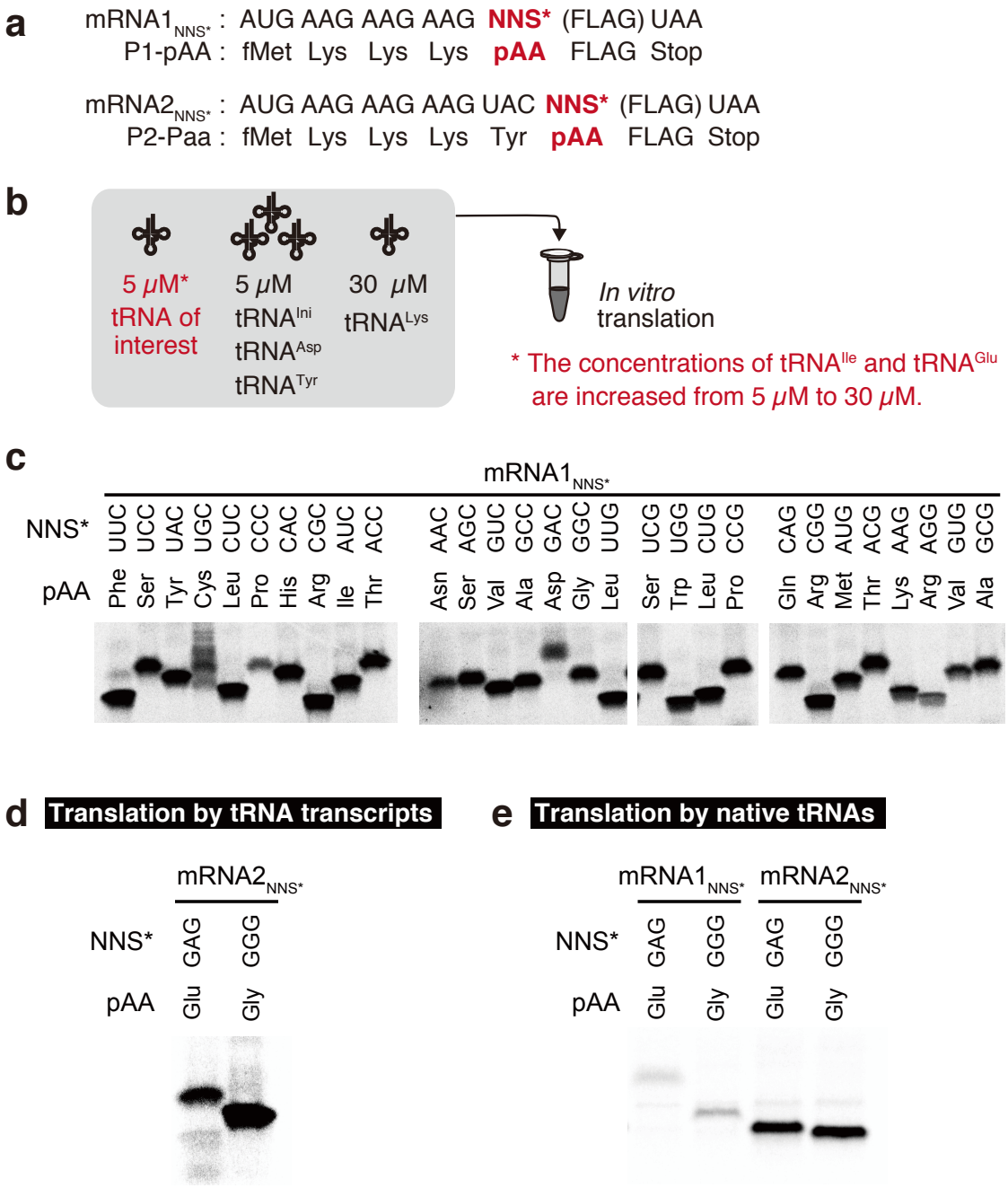

Figure 26. Decoding by the minimal FIT system composed of *in vitro* tRNA transcripts. a, Sequences of the mRNA1$_{NNS*}$ and mRNA2$_{NNS*}$ sets, and their resultant P1-pAA and P2-pAA peptides. '(FLAG)' represents an mRNA sequence that codes for the FLAG peptide (DYKDDDDK). **b**, The minimal FIT systems used in this experiment. mRNA1$_{NNS*}$'s and mRNA2$_{NNS*}$'s containing an NNS* (S = G or C) codon at the fifth and sixth positions, respectively, were translated in the presence of *in vitro* transcripts of tRNA$^{Ini}_{CAU}$, tRNA$^{Tyr}_{GUA}$, tRNA$^{Asp}_{GUC}$, tRNA$^{Lys}_{CUU}$, and tRNAs that decode the relevant NNS* codon. **c,d,** Confirmation of the activities of *in vitro* tRNA transcripts by means of the tricine-SDS-PAGE analyses of P1(**c**) and P2(**d**) peptides synthesized by the minimal FIT systems. Multiple bands are generally observed when a peptide containing a free Cys residue is analyzed by tricine-SDS-PAGE, which result from disulfide bond formation with thiol-containing molecules. **e**, Optimization of mRNA sequences containing GAG or GGG codon. In this experiment, mRNAs were translated in a FIT system containing *E. coli* total tRNA.

Next, I designed two additional mRNA templates, mRNA3 and mRNA4 that fully covered all 31 elongator NNS codons, and expressed them in a FIT system that contained the 32 $tRNA_{SNN}$ transcripts at their optimal concentrations, referred to as the FIT-32t system (Fig. 27). MALDI-TOF-MS of the peptide products expressed from the respective mRNAs (Fig. 27b,c) indicated the correct molecular weights of the desired peptides. Furthermore, a model peptide containing all of 20 pAAs was expressed in the FIT-32t system (Fig. 28a). The expression level was 29% compared to the native tRNA control (Fig. 28b), and the accurate synthesis was confirmed by MALDI-TOF-MS of the peptide products (Fig. 28c). Based on these results, I concluded that the 32 *in vitro* $tRNA_{SNN}$ transcripts are able to decode cognate codons accurately in the FIT-32t system.

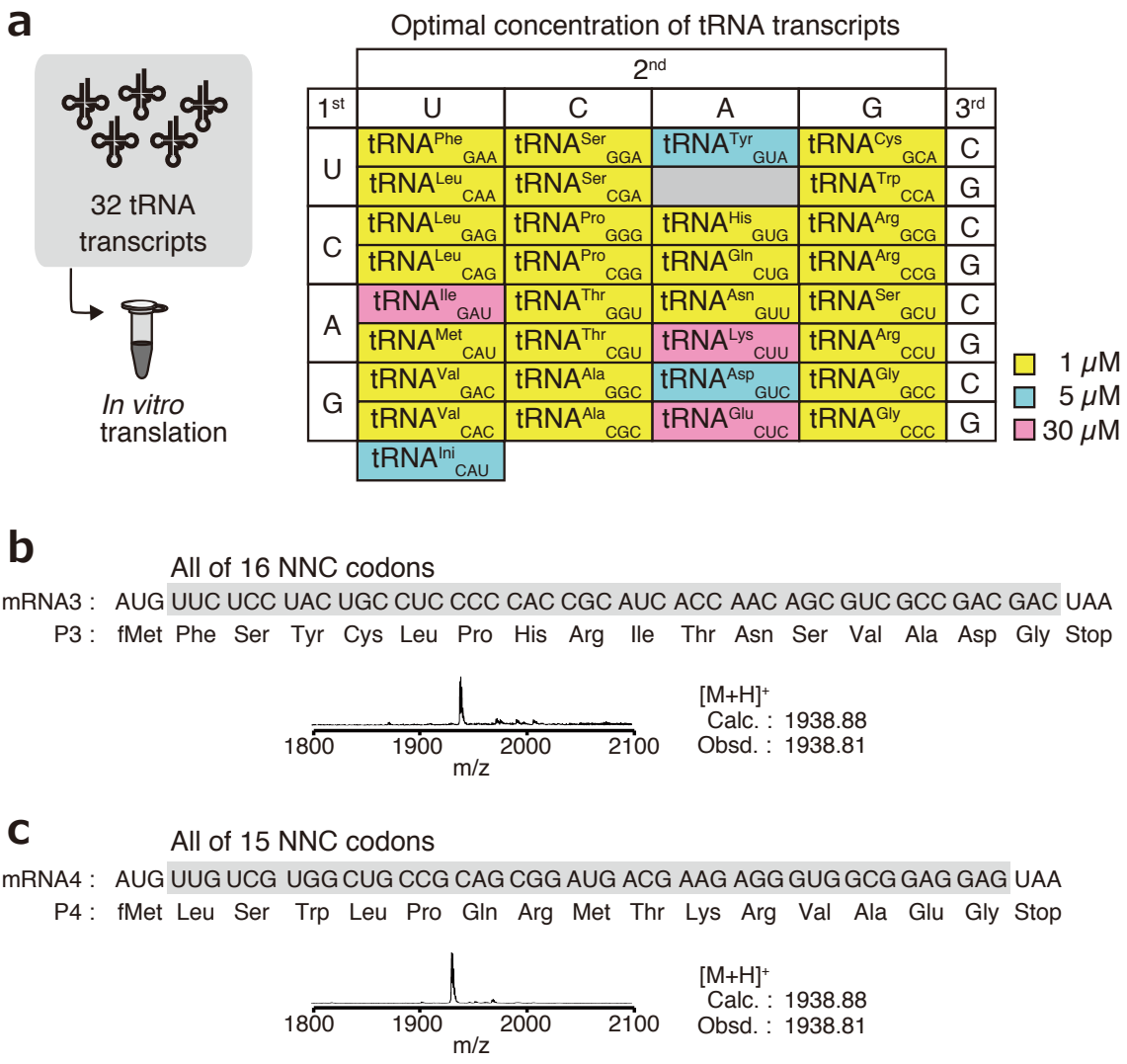

Figure 27. Accurate decoding by 32 tRNA transcripts in the FIT-32t system. a, The FIT-32t system containing optimal concentrations of 32 tRNA transcripts. **b,c**, Accurate decoding of every 31 NNS elongator codons. Peptide products expressed in the FIT-32t system from mRNA3 containing 16 NNC sense codons (**d**) and mRNA4 containing 15 NNG sense codons (**e**) were analyzed by MALDI-TOF-MS.

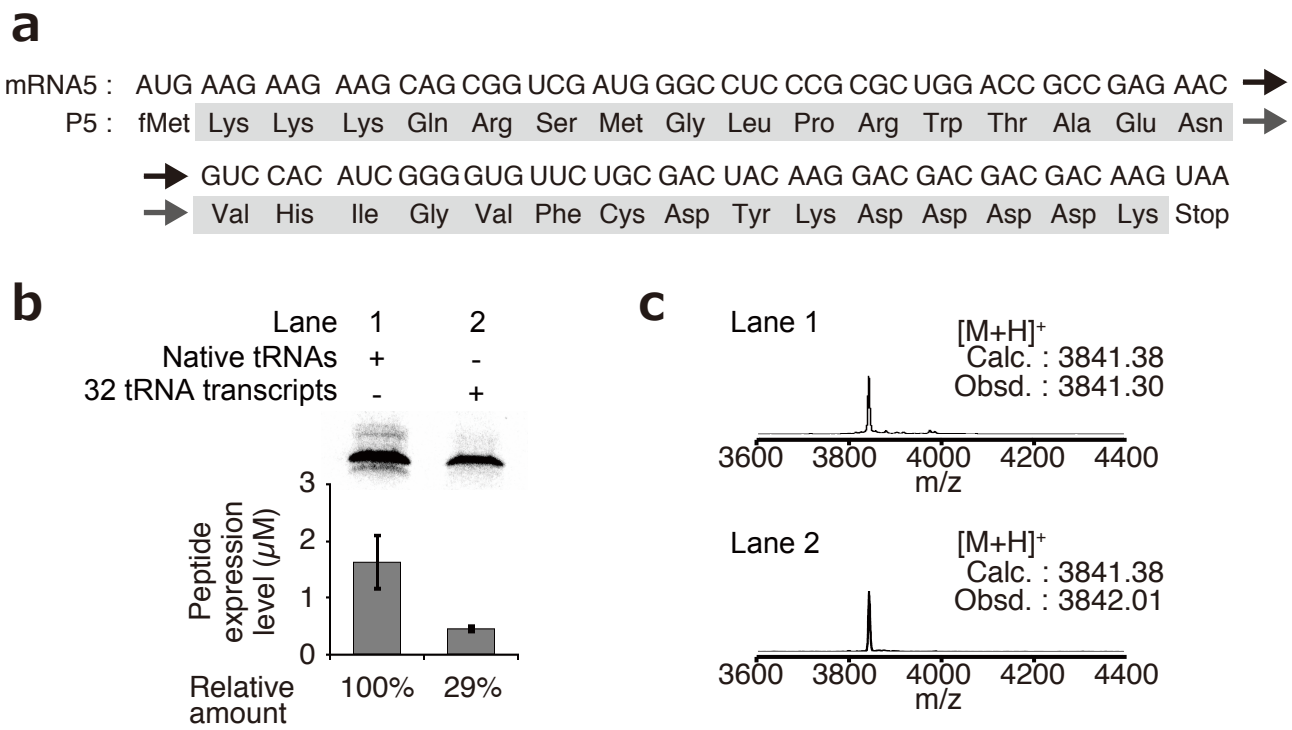

Figure 28. Demonstration of the polypeptide synthesis with 20 pAA repertoire in the FIT-32t system. a, Sequences of the mRNA5 and its resultant P5 peptide. Arrows indicate the sequence connections. **b,c**, Expression of a model peptide containing all 20 pAAs in the FIT-32t system. The peptide products were analyzed by tricine-SDS-PAGE (**b**) and MALDI-TOF-MS (**c**). Native tRNAs: 1.5 mg/mL of *E. coli* total tRNA. A part of these data is reused in Fig. 35.

2.2.2. Demonstration of artificial division of a single codon box

To realize the concept of artificial codon-box division, we attempted to divide the Val GUN codon box under the optimized translation condition (see Supplementary results 2.4.3 for the optimization experiments). We used $mRNA1_{NNG}$ and $mRNA1_{NNC}$ templates and assigned Val and *N*-methyltyrosine (^{Me}Y) to GUG and GUC codons, respectively (Fig. 29a). The assignment was achieved by the addition of either $tRNA^{Val}_{CAC}$, which is *in situ* charged with Val by ValRS, or an orthogonal $tRNA^{AsnE2}_{GAC}$ pre-charged with ^{Me}Y by enhanced flexizyme (eFx) catalysis[92]. As a positive control, $mRNA1_{GUG}$ was translated in the FIT system that consisted of native tRNAs, referred to as FIT-nt, and the expression of P1-Val peptide was observed in tricine-SDS-PAGE (Fig. 29b, lane 1). When the minimal FIT system was used, only a negligible faint band with the same migration as P1-Val was detected (Fig. 29b, lane 2), presumably because of an unavoidable trace amount of contamination by native $tRNA^{Val}$ in the FIT system. The addition of $tRNA^{Val}_{CAC}$ dramatically increased the expression level with a comparable efficiency to that of the FIT-nt system (Fig. 29b, lanes 1 versus 3). Further inclusion of ^{Me}Y-$tRNA^{AsnE2}_{GAC}$ did not disturb the expression of P1-Val (Fig. 29b, lanes 3 versus 4), and the accurate synthesis of correct peptide was confirmed by MALDI-TOF-MS of the peptide product after purification by anti-FLAG M2 agarose (Fig. 29c). The above results indicated that ^{Me}Y-$tRNA^{AsnE2}_{GAC}$ does not cross-read the GUG codon.

I next examined the expression of P1-^{Me}Y from $mRNA1_{GUC}$ in the presence of ^{Me}Y-$tRNA^{AsnE2}_{GAC}$ and/or $tRNA^{Val}_{CAC}$. This template produced P1-Val in the FIT-nt system (Fig. 29b, lane 5). However, the minimal FIT system alone produced a negligible faint band with the same migration as P1-Val and another band at a slower migrated position (Fig. 29b, lane 6). These side-product bands turned out to be P1-Val and P1-Asp, as indicated by MALDI-TOF-MS (Fig. 29c). Expression of P1-Asp was probably caused by the misreading of GUC by Asp-$tRNA^{Asp}_{GUC}$, which was forced by the lack of the corresponding tRNA that decoded GUC codon. In fact, these side-product bands were suppressed by the addition of ^{Me}Y-$tRNA^{AsnE2}_{GAC}$, and the desired P1-^{Me}Y was expressed with a comparable efficiency to that of the FIT-nt system (Fig. 29b, lanes 5 versus 7). Expression of P1-^{Me}Y was not disturbed by the addition of $tRNA^{Val}_{CAC}$ (Fig. 29b, lanes 7 versus 8), as confirmed by MALDI-TOF-MS (Fig. 29c). Taking all the data together, the Val codon box was successfully divided to assign Val and ^{Me}Y to GUG and GUC codons, respectively.

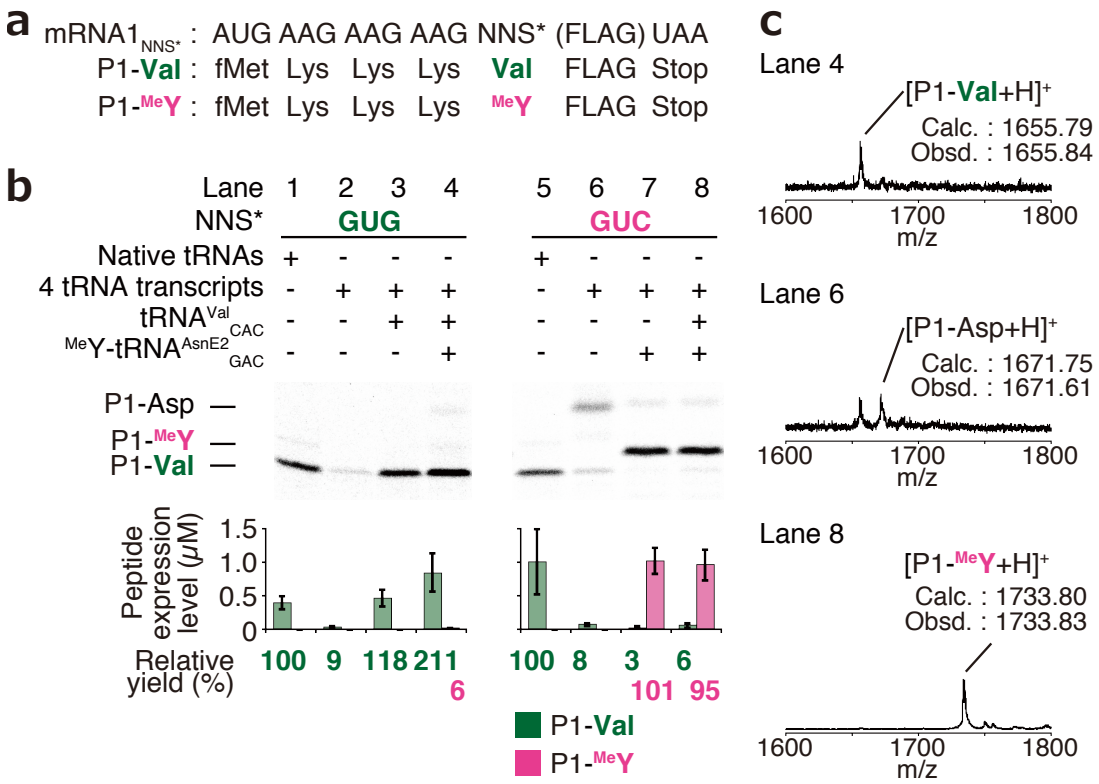

Figure 29. Artificial division of the Val GUN codon box. a, Sequence of the mRNA1$_{NNS*}$ set and their resultant P1 peptides. GUG and GUC codons present at the fifth position of mRNA1$_{NNS*}$ were decoded by Val-tRNA$^{Val}_{CAC}$ and ^{Me}Y-tRNA$^{AsnE2}_{GAC}$, respectively. **b**,**c**, Accurate and efficient decoding of GUG and GUC codons to Val and ^{Me}Y, respectively. Peptide products expressed in the minimal FIT systems were analyzed by tricine-SDS-PAGE (**b**) and MALDI-TOF-MS (**c**). Native tRNAs: 1.5 mg/mL of *E. coli* total tRNA; 4 *in vitro* tRNA transcripts: 5 μM each tRNA$^{Ini}_{CAU}$, tRNA$^{Asp}_{GUC}$, and tRNA$^{Tyr}_{GUA}$, and 30 μM tRNA$^{Lys}_{CUU}$; tRNA$^{Val}_{CAC}$: 5 μM tRNA$^{Val}_{CAC}$; ^{Me}Y-tRNA$^{AsnE2}_{GAC}$: 100 μM tRNA$^{AsnE2}_{GAC}$ pre-charged with ^{Me}Y; Error bar: standard deviations.

I also performed the division of two additional codon boxes, those of Arg and Gly. In the Arg CGN codon box, CGG and CGC were assigned to Arg and *N*-methylserine (^{Me}S) by the addition of tRNA$^{Arg}_{CCG}$ transcript and ^{Me}S-tRNA$^{AsnE2}_{CGC}$ prepared by dinitro-flexizyme (dFx)[92] catalysis, respectively (Fig. 30a). Similarly, in the Gly GGN codon box, GGG and GGC were assigned to Gly and 4-iodophenylalanine (^{Iodo}F) by the addition of tRNA$^{Gly}_{CCC}$ transcript and ^{Iodo}F-tRNA$^{AsnE2}_{GCC}$ prepared by eFx catalysis, respectively (Fig. 30b). In both experiments, the desired peptides, P1-Arg and P1-^{Me}S (Fig. 30a) as well as P2-Gly and P2-^{Iodo}F (Fig. 30b), were expressed from the respective mRNA templates in a similar manner to the Val codon box's result. Taken together, I have demonstrated the concept of artificial codon-box division that assign a single npAA to NNC codons and the corresponding NNG codon remains assigned with the cognate pAA.

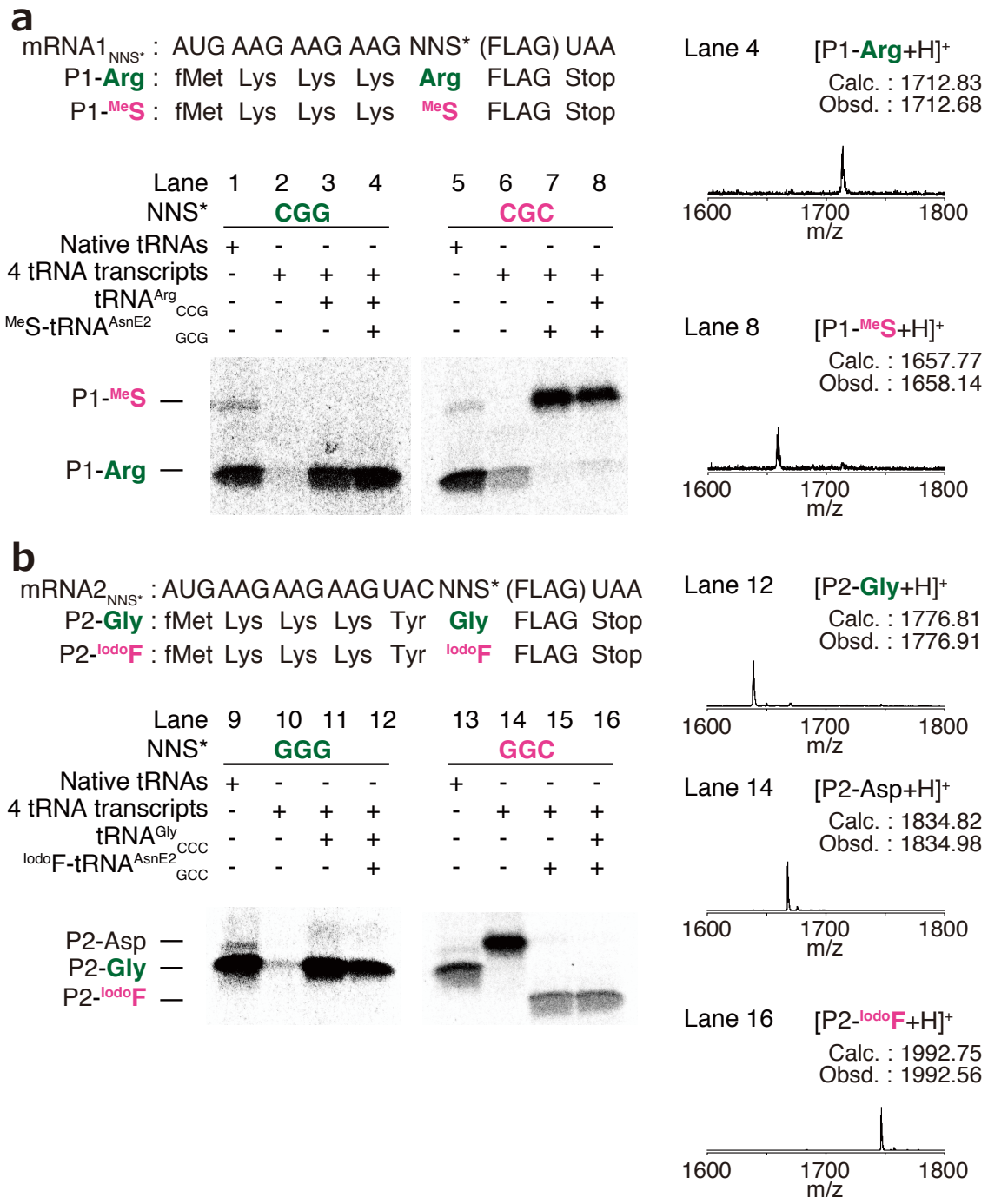

Figure 30. Artificial division of Arg and Gly codon boxes. Accurate and efficient decoding of the artificially divided Arg codon box (**a**) and Gly codon box (**b**). Peptide products expressed in the minimal FIT systems were analyzed by tricine-SDS-PAGE and MALDI-TOF-MS. Native tRNAs: 1.5 mg/mL of *E. coli* total tRNA; 4 *in vitro* tRNA transcripts: 5 μM each $tRNA^{Ini}_{CAU}$, $tRNA^{Asp}_{GUC}$, $tRNA^{Tyr}_{GUA}$, and 30 μM $tRNA^{Lys}_{CUU}$; $tRNA^{Arg}_{CCG}$: 5 μM $tRNA^{Arg}_{CCG}$; ^{Me}S-$tRNA^{AsnE2}_{GCG}$: 100 μM $tRNA^{AsnE2}_{GCG}$ pre-charged with ^{Me}S; $tRNA^{Gly}_{CCC}$: 5 μM $tRNA^{Gly}_{CCC}$; ^{Iodo}F-$tRNA^{AsnE2}_{GCC}$: 100 μM $tRNA^{AsnE2}_{GCC}$ pre-charged with ^{Iodo}F.

2.2.3. Assignment of two different npAAs in the Arg CGN codon box

In the AGN codon box, AGA and AGG codons code for Arg, whereas AGU and AGC codons code for Ser (Fig. 23a). In addition to these codons, Arg is assigned to the CGN codon box, whereas Ser is assigned to the UCN codon box. Thus, these redundant codon boxes can be divided artificially to reassign two different npAAs. Here I chose the Arg CGN codon box to demonstrate this. The mRNA6 template that contained three different Arg codons, (AGG, CGC, and CGG) was translated in two FIT systems as follows (Fig. 31a). As a control for the ordinary translation, the FIT-32t system decoded the above codons with three consecutive Arg's (Fig. 31b). The other FIT system contained only 30 *in vitro* tRNA transcripts (FIT-30t), which excluded $tRNA^{Arg}_{GCG}$ and $tRNA^{Arg}_{CCG}$, which resulted in that the corresponding CGC and CGG codons became vacant whereas only the AGG codon coded for Arg by means of $tRNA^{Arg}_{CCU}$. To this FIT-30t, I added $tRNA^{AsnE2}_{GCG}$ and $tRNA^{AsnE2}_{CCG}$ pre-charged with ^{Me}S and ε-*N*-acetyllysine (^{Ac}K), respectively; thereby, ^{Me}S and ^{Ac}K were assigned to CGC and CGG codons, respectively (Fig. 31c).

Expression of mRNA6 in the control FIT-32t expressed the P6-Arg/Arg/Arg peptide that contained three Arg residues (Fig. 31b). The FIT-30t with the two npAA-tRNAs yielded a single product with a mass value consistent with the desired P6-Arg/^{Me}S/^{Ac}K peptide (Fig. 31c). Thus, the Arg CGC and CGG codons were reassigned to two different npAAs. This demonstrated that the redundant Arg codon box can be utilized for the assignment of two different npAAs with the AGG codon coding for Arg.

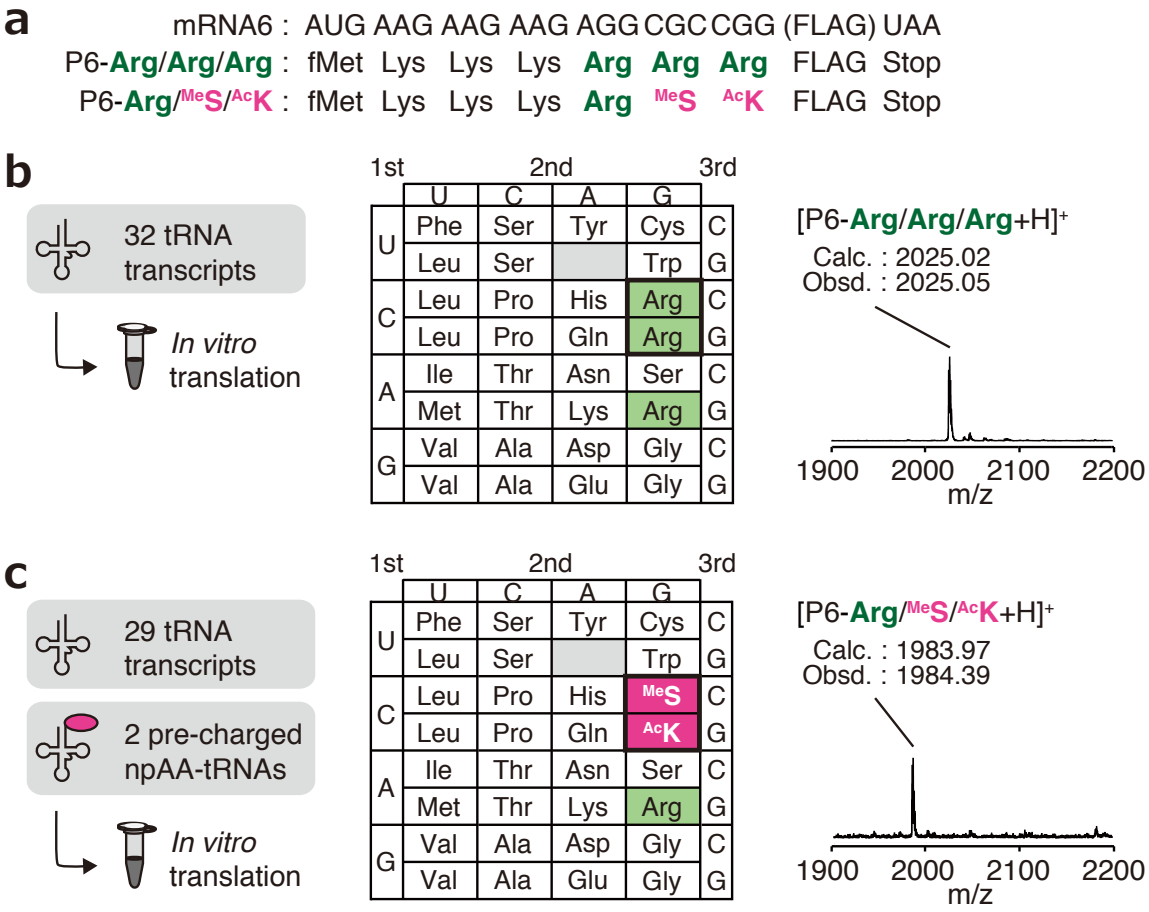

Figure 31. Assignments of two different npAAs in Arg CGN codon box. a, Sequences of mRNA6 and the resultant P6 peptides. **b**, A control experiment in which the three different Arg codons (AGG, CGC, and CGG) were decoded to three consecutive Arg in the FIT-32t system. The peptide product expressed in FIT-32t system was analyzed by MALDI-TOF-MS. **c**, Assignment of two different npAAs in the Arg CGN codon box. The peptide product expressed in the FIT-30t system (excluding $tRNA^{Arg}_{GCG}$ and $tRNA^{Arg}_{CCG}$) supplied with $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ and $^{Ac}K\text{-}tRNA^{AsnE2}_{CCG}$ was analyzed by MALDI-TOF-MS.

2.2.4. Artificial division of multiple codon boxes

I next attempted the simultaneous division of multiple codon boxes to achieve assignments of multiple npAAs. The mRNA7, which contained two Arg CGS codons (CGC and CGG) and two Val GUS codons (GUC and GUG), was designed to express four different peptides according to the genetic codes reprogrammed by the four FIT systems (Fig. 32a,b). First, the FIT-32t system should decode the above codons with Arg and Val. Second, the FIT-31t system, in which $tRNA^{Arg}_{GCG}$ was replaced with $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$, would decode CGC codon to ^{Me}S. Third, the FIT-31t system, in which $tRNA^{Val}_{GAC}$ was replaced with $^{Me}Y\text{-}tRNA^{AsnE2}_{GAC}$, would decode GUC codon to ^{Me}Y. Fourth, the FIT-30t system, in which both $tRNA^{Arg}_{GCG}$ and $tRNA^{Val}_{GAC}$ were replaced with $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ and $^{Me}Y\text{-}tRNA^{AsnE2}_{GAC}$, would at once decode both CGC and GUC codons to ^{Me}S and ^{Me}Y, respectively. Translation of mRNA7 in these FIT systems each yielded a single major product (P7-Arg/Val, P7-^{Me}S/Val, P7-Arg/^{Me}Y, and P7-^{Me}S/^{Me}Y) accurately (Fig. 32c–f). Clearly, the artificial division of multiple codon boxes can be achieved by the designated FIT systems. The translation efficiency determined by the expression level of P7-Arg/Val in the FIT-32t system was nearly twice that of the FIT-nt system (Fig. 32g, lanes 1 versus 2). The introduction of $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ or $^{Me}Y\text{-}tRNA^{AsnE2}_{GAC}$ dropped the expression level of P7-^{Me}S/Val or P7-Arg/^{Me}Y twofold (Fig. 32g, lanes 3 and 4). The introduction of both npAA-tRNAs yielded P7-^{Me}S/^{Me}Y with an approximately twofold reduction of the single suppression (Fig. 32g, lane 5).

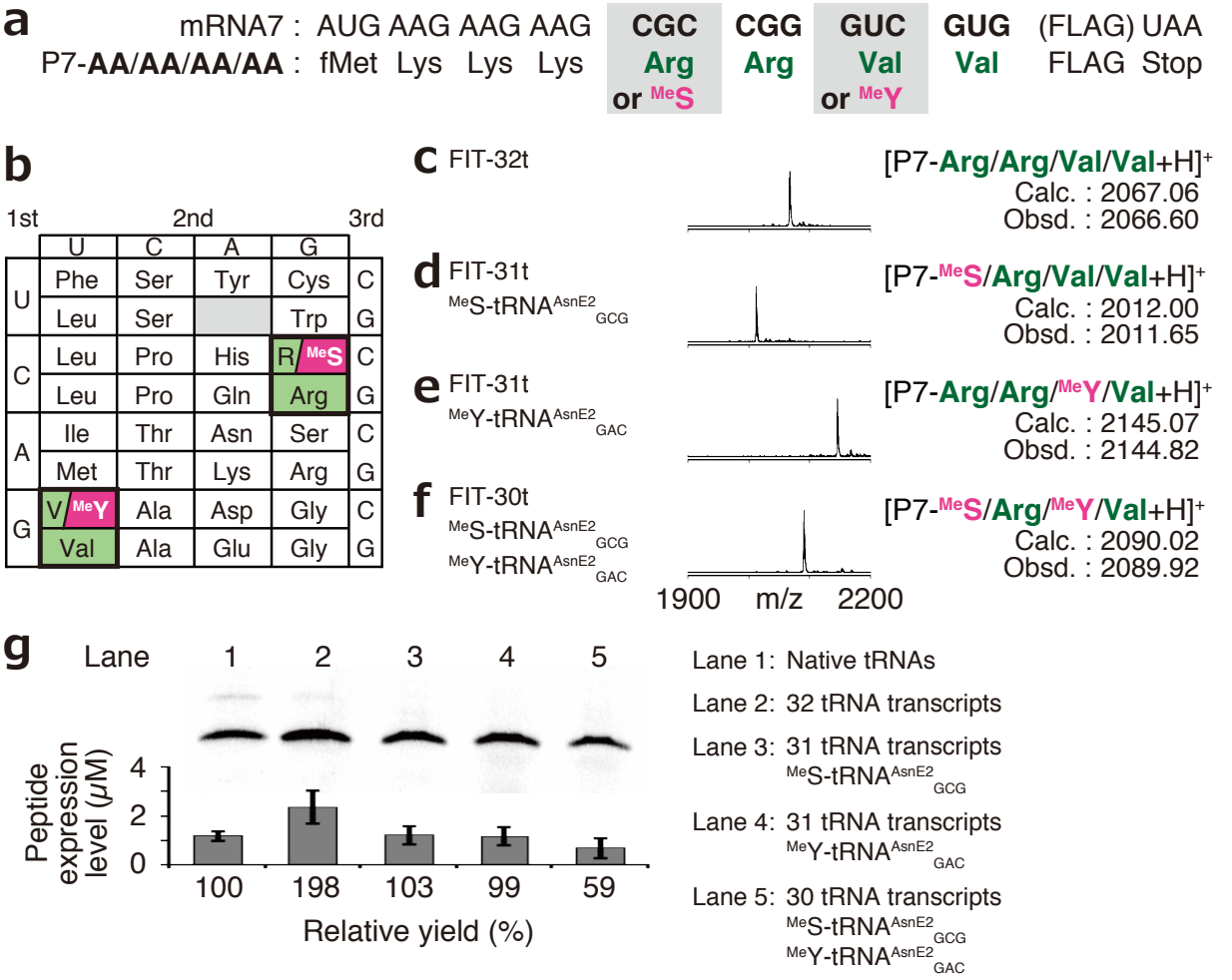

Figure 32. Simultaneous division of Arg and Val codon boxes. a, Sequences of mRNA7 and its resultant P7 peptides. **b**, A reprogrammed genetic code containing artificially divided Arg and Val codon boxes. **c**–**f**, MALDI-TOF-MS of the peptide products expressed in the FIT-32t system (**c**), the FIT-31t system (excluding $tRNA^{Arg}_{GCG}$) supplied with $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ (**d**), the FIT-31t system (excluding $tRNA^{Val}_{GAC}$) supplied with $^{Me}Y\text{-}tRNA^{AsnE2}_{GAC}$ (**e**), and the FIT-30t system (excluding $tRNA^{Arg}_{GCG}$ and $tRNA^{Val}_{GAC}$) supplied with $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ and $^{Me}Y\text{-}tRNA^{AsnE2}_{GAC}$ (**f**). **g**, Quantification of the peptide expression levels. Native tRNAs: 1.5 mg/mL of *E. coli* total tRNA; Error bar: standard deviation.

Similarly, I performed expression of mRNA8, which contained two Gly GGS codons (GGC and GGG) and two Arg CGS codons (CGC and CGG), in the similar FIT systems (excluding $tRNA^{Gly}_{GCC}$ and/or $tRNA^{Arg}_{GCG}$) supplied with $^{Iodo}F\text{-}tRNA^{AsnE2}_{GCC}$ and/or $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ (Fig. 33). Again, a single major product that corresponded to the expected peptide was observed in each case, which demonstrated that the simultaneous division of the Gly GGN and Arg CGN codon boxes can also be achieved.

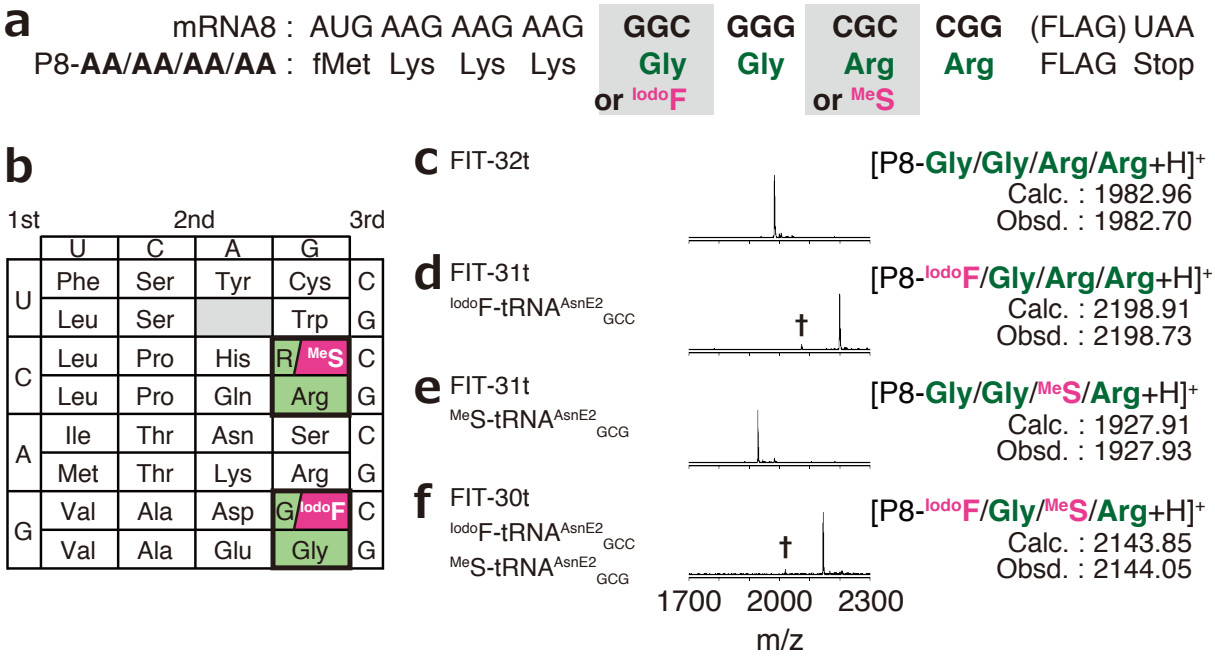

Figure 33. Simultaneous division of Gly and Arg codon boxes. a, Sequences of mRNA8 and its resultant P8 peptides. **b**, A reprogrammed genetic code containing artificially divided Gly and Arg codon boxes. **c**–**f**, MALDI-TOF-MS of peptide products expressed in FIT-32t system (**c**), FIT-31t system (excluding $tRNA^{Gly}_{GCC}$) supplied with $^{Iodo}F\text{-}tRNA^{AsnE2}_{GCC}$ (**d**), FIT-31t system (excluding $tRNA^{Arg}_{GCG}$) supplied with $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ (**e**), and FIT-30t system (excluding $tRNA^{Gly}_{GCC}$ and $tRNA^{Arg}_{GCG}$) supplied with $^{Iodo}F\text{-}tRNA^{AsnE2}_{GCC}$ and $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$ (**f**). †Peak corresponds to the peptide containing Phe in place of ^{Iodo}F, which originates from a trace amount of contaminant in the ^{Iodo}F substrate.

Next, I performed the simultaneous division of three codon boxes (Fig. 34a,b). When the mRNA9 was translated in the FIT-32t system, the corresponding P9-Gly/Gly/Arg/Arg/Val/Val peptide was expressed as a single major product (Fig. 34c). When the same template was translated in a FIT-29t system (excluding $tRNA^{Gly}_{GCC}$, $tRNA^{Arg}_{GCG}$, and $tRNA^{Val}_{GAC}$) supplied with ^{Iodo}F-$tRNA^{AsnE2}_{GCC}$, ^{Me}S-$tRNA^{AsnE2}_{GCG}$, and ^{Me}Y-$tRNA^{AsnE2}_{GAC}$, P9-^{Iodo}F/Gly/^{Me}S/Arg/^{Me}Y/Gly was expressed as the only major product (Fig. 34d). This triple suppression with three npAAs yielded 0.28 μM P9-^{Iodo}F/Gly/^{Me}S/Arg/^{Me}Y/Gly with approximately fivefold less than the wild-type P9-Gly/Gly/Arg/Arg/Val/Val expression in the FIT-nt system (Fig. 34e). This range of expression level should be sufficient for the application to *in vitro* selection technologies, such as the RaPID system.

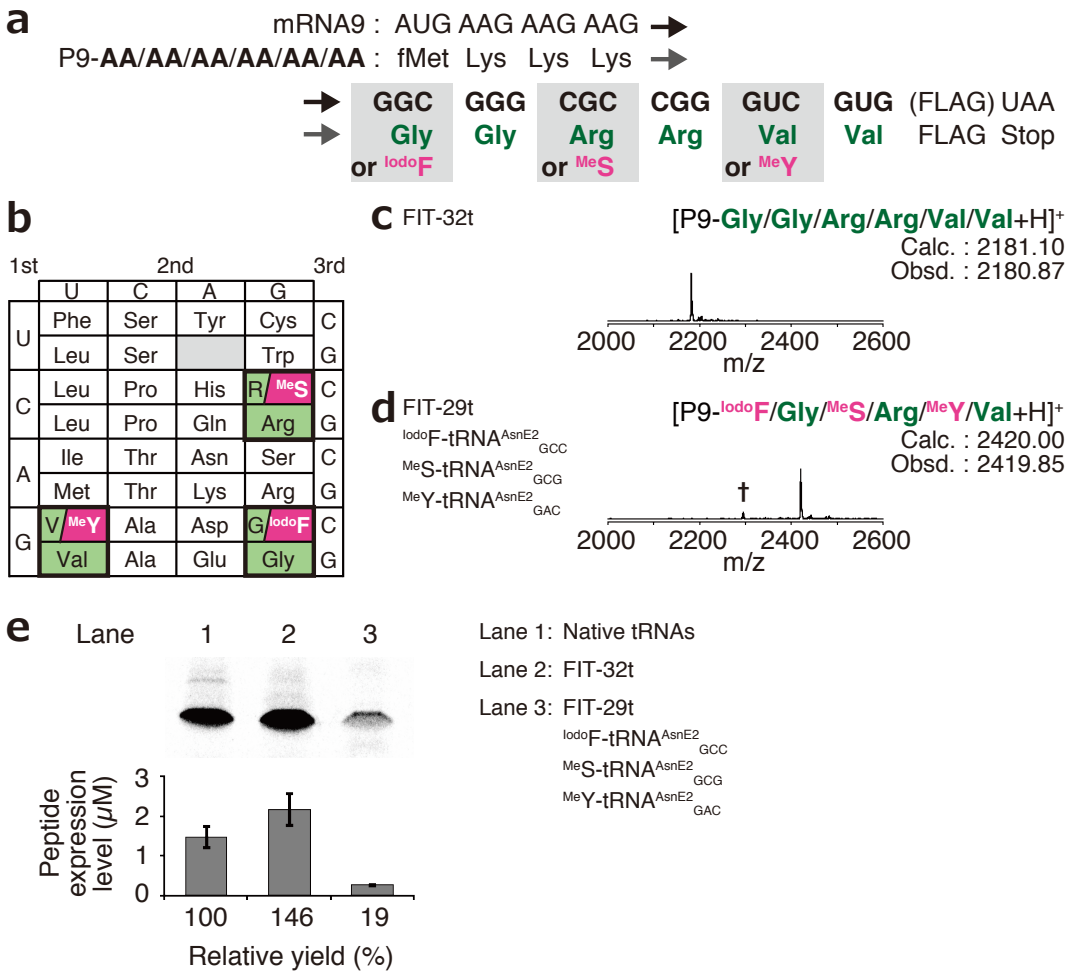

Figure 34. Simultaneous division of three codon boxes. a, Sequences of mRNA9 and its resultant P9 peptides. **b**, A reprogrammed genetic code with three artificially divided codon boxes. **c,d**, MALDI-TOF-MS of the peptide products expressed in the FIT-32t system (**c**) and the FIT-29t system (excluding $tRNA^{Gly}_{GCC}$, $tRNA^{Arg}_{GCG}$, and $tRNA^{Val}_{GAC}$) supplied with ^{Iodo}F-$tRNA^{AsnE2}_{GCC}$, ^{Me}S-$tRNA^{AsnE2}_{GCG}$, and ^{Me}Y-$tRNA^{AsnE2}_{GAC}$ (**d**). †Peak corresponds to the peptide that contains Phe in place of ^{Iodo}F, which originates from a trace amount of contaminant in the ^{Iodo}F substrate. **e**, Quantification of the peptide expression levels. Native tRNAs: 1.5 mg/mL of *E. coli* total tRNA; Error bar: standard deviation.

To validate whether the incorporation of npAAs along with 20 pAAs could be achieved in a single peptide molecule, I designed mRNA10 that coded for not only three npAAs (^{Ac}K, ^{Iodo}F, and citrulline (Cit)) along with all the 20 pAAs (Fig. 35a,b; note that Tyr and Asp are present in the FLAG peptide sequence (DYKDDDDK)). When mRNA10 was translated in the FIT-32t system, P10-Gly^{9}/Arg^{12}/Val^{18} was expressed in an approximately 30% expression level relative to that of the FIT-nt system (Fig. 35c,e, lanes 1 versus 2). The translation of mRNA10 in the FIT-29t supplied with ^{Ac}K-tRNA$^{AsnE2}_{GCC}$, ^{Iodo}F-tRNA$^{AsnE2}_{GCG}$, and Cit-tRNA$^{AsnE2}_{GAC}$ yielded the desired 32-mer P10-^{Ac}K^{9}/^{Iodo}F^{12}/Cit^{18} peptide containing the expected repertoire of 23 amino acids (Fig. 35d,e, lane 3). Although the expression level of P10-^{Ac}K^{9}/^{Iodo}F^{12}/Cit^{18} was approximately 15% of the P10-Gly^{9}/Arg^{12}/Val^{18} expressed in the FIT-nt system, it is clearly detectable in both tricine-SDS-PAGE and MALDI-TOF-MS as a single major product.

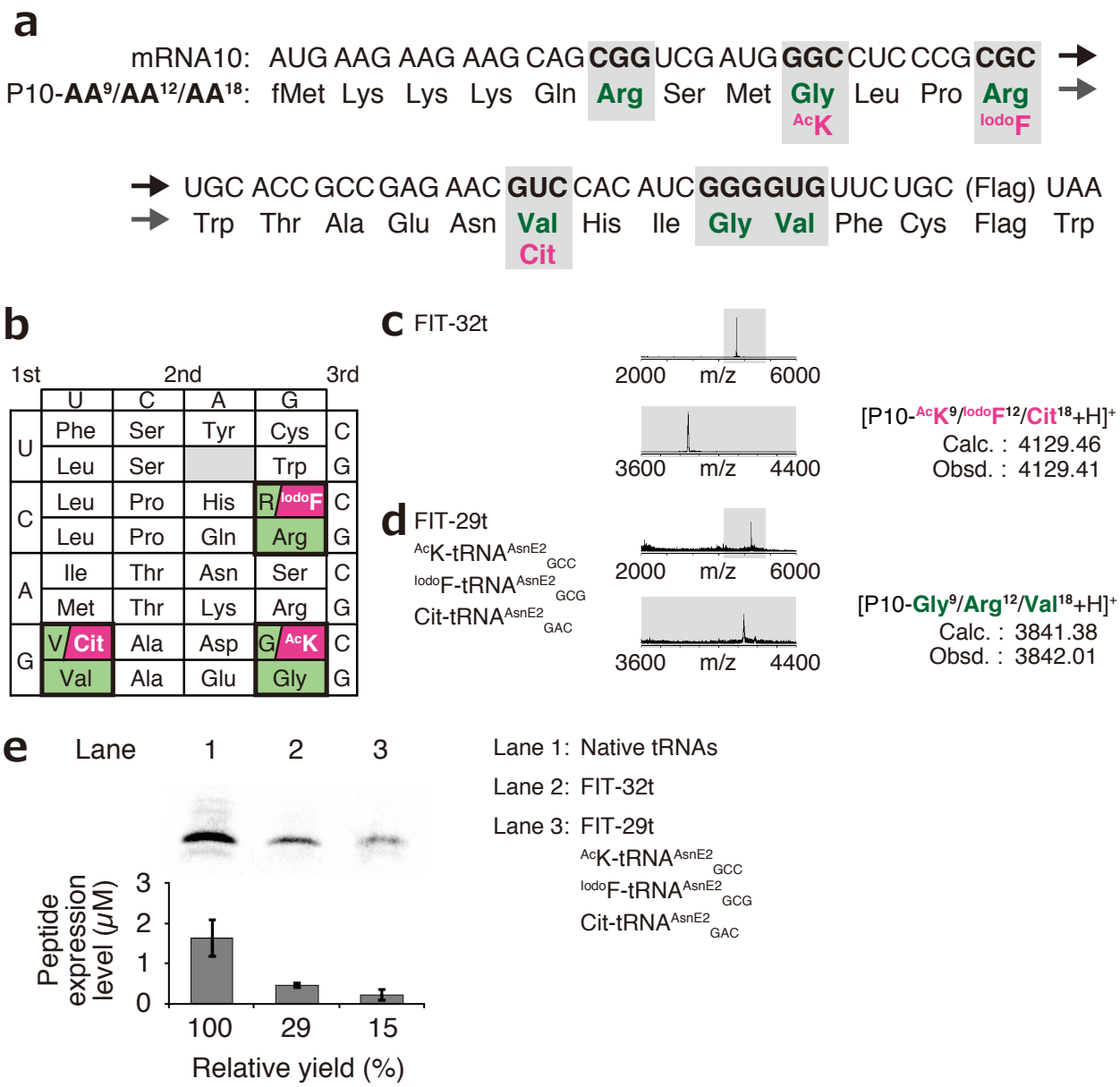

Figure 35. Accurate use of the expanded repertoire of amino acids (20 pAAs and 3 npAAs) via artificial division of three codon boxes. a, Sequences of mRNA10 and its resultant P10 peptide. **b**, A reprogrammed genetic code with three artificially divided codon boxes. **c,d**, MALDI-TOF-MS of the peptide products expressed in the FIT-32t system (**c**) and FIT-29t system (excluding $tRNA^{Gly}_{GCC}$, $tRNA^{Arg}_{GCG}$, and $tRNA^{Val}_{GAC}$) supplied with $^{Ac}K\text{-}tRNA^{AsnE2}_{GCC}$, $^{Iodo}F\text{-}tRNA^{AsnE2}_{GCG}$, and $Cit\text{-}tRNA^{AsnE2}_{GAC}$ (**d**). The calculated and observed m/z values in this figure are average mass values. **e**, Quantification of the peptide expression levels. Native tRNAs: 1.5 mg/mL of *E. coli* total tRNA; Error bar: standard deviation.

2.2.5. Expression of macrocyclic *N*-methyl-peptide CM_{11}-1

As a concluding experiment to demonstrate artificial codon-box division, I expressed a macrocyclic *N*-methyl-peptide, CM_{11}-1, that comprised of 14 amino acids (nine proteinogenic, five nonproteinogenic). This 'natural product-like' macrocycle was developed by the RaPID selection targeting E6AP, which belongs to a family of ubiquitin ligases, HECT E3, and plays a crucial role in promoting the degradation of tumor-associating proteins such as p63 during tumor development[93]. When this macrocycle, which contains ^{Me}S, *N*-methylphenylalanine (^{Me}F) and *N*-methylglycine (^{Me}G), was originally selected, they utilized a reprogrammed genetic code in which *N*-chloroacetyl-D-tryptophan (ClAc^{D}W), ^{Me}S, ^{Me}F, ^{Me}G, and ^{Me}A (*N*-methylalanine) were assigned to the initiator AUG codon and the elongator UUU, CUU, AUU, and GCU codons, respectively. The corresponding Met, Phe, Leu, Ile, and Ala were removed from the genetic code. Moreover, because they utilized an NNU codon library, the genetic code also lacked Gln, Lys, Glu, and Trp. Thus, in the previous experiment, a total of nine pAAs was omitted from the construction of the macrocyclic *N*-methyl-peptide library. Here I aimed to realize the expression of CM_{11}-1 with maintaining as many pAAs as possible.

I designed a reprogrammed genetic code wherein the initiator *N*-formylmethionine was replaced with ClAc^{D}W (Fig. 36a), and ^{Me}F, ^{Me}S, and ^{Me}G were assigned to the GUC, CGC, and GGC codons, respectively (Fig. 36b). It should be noted that all the elongator amino acids except for Met were kept in the genetic code The corresponding mRNA11 was designed based on the above genetic code, and expressed in the FIT-28t system that contained ClAc^{D}W-tRNA$^{Ini}_{CAU}$, ^{Me}F-tRNA$^{AsnE2}_{GAC}$, ^{Me}S-tRNA$^{AsnE2}_{GCG}$, and ^{Me}G-tRNA$^{AsnE2}_{GCC}$. The MALDI-TOF-MS analysis of the peptide product revealed that the desired CM_{11}-1 peptide was correctly translated from the mRNA11 under the reprogrammed genetic code generated by our methodology for the artificial division of codon boxes (Fig. 36c,d).

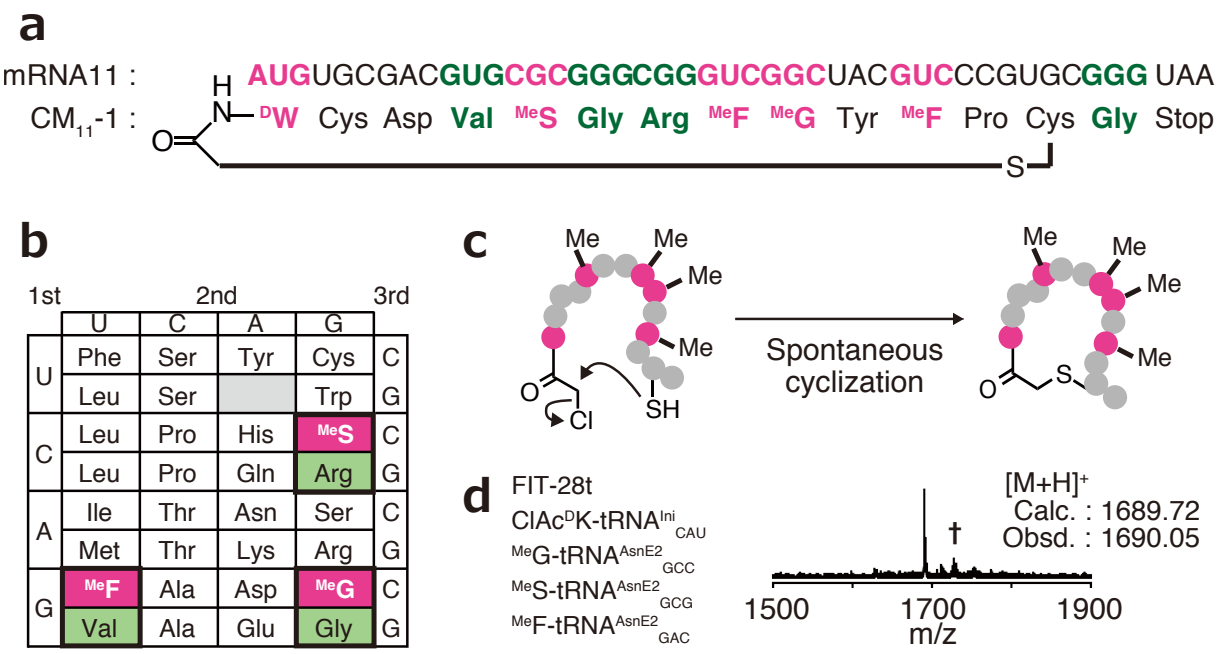

Figure 36. Artificial division of three codon boxes to express a nonstandard macrocyclic *N*-methyl-peptide. a, Sequences of mRNA11 and its resultant peptide 'CM_{11}-1'. **b**, A reprogrammed genetic code with three artificially divided codon boxes. **c**, Scheme of the post-translational macrocyclization. **d**, MALDI-TOF-MS of the peptide products expressed in the FIT-28t system (excluding $tRNA^{Ini}_{CAU}$, $tRNA^{Gly}_{GCC}$, $tRNA^{Arg}_{GCG}$, and $tRNA^{Val}_{GAC}$) supplied with $ClAc^{D}W\text{-}tRNA^{Ini}_{CAU}$, $^{Me}G\text{-}tRNA^{AsnE2}_{GCC}$, $^{Me}S\text{-}tRNA^{AsnE2}_{GCG}$, and $^{Me}F\text{-}tRNA^{AsnE2}_{GAC}$. †Peak corresponds to the linear peptide that contains the unreacted *N*-terminal chloroacetyl group.

2.3. Conclusion

Here I report a novel methodology to create vacant codons by reducing the redundancy of codon assignment based on the concept of 'artificial division of codon boxes'. To realize this concept, I have prepared a native tRNA-free FIT system and supplied it with 32 *in vitro* transcribed $tRNA_{SNN}$'s (S = G or C), which covered the assignment of the 20 pAAs. In principle, this system can create 11 vacant codons, which can then be reassigned with npAAs without the need to abandon any of 20 pAAs. In this study, up to three codon boxes were simultaneously divided and three different npAAs were assigned therein. To demonstrate this concept, I expressed various model peptides, which included a macrocyclic *N*-methyl-peptide inhibitor of E6AP[93]. Notably, the reprogrammed codons were accurately decoded to the desired npAAs and the misincorporation of either cognate or non-cognate pAAs did not occur, which demonstrated the maintenance of translation accuracy. The present proof-of-concept study on the artificial division of codon boxes opens a new opportunity for genetic code reprogramming[59,60]. In particular, when this technology is coupled with a peptide selection technology, such as the RaPID system[93,130,131,133], it enables us to express libraries of nonstandard macrocyclic peptides that are composed of 23 or more pAA and npAA building blocks to discover bioactive ligands against drug targets.

The expansion of amino acid repertoire has three practical benefits: (1) Whereas the canonical translation often redundantly expresses the same peptide from different mRNA sequences, the engineered translation system developed in this study can synthesize more diverse peptides from the same mRNA library using the expanded building block repertoire (Fig. 37a). Accordingly, the diversity of a peptide library can be increased by fourfold (from ~1×10^{12} to ~4×10^{12}) in case of a typical RaPID selection, which is expected to improve the binding properties of the discovered peptides. (2) Expansion of the building block allows for the construction of binding interface between peptide drug and target protein with expanded structural and chemical diversity (Fig. 37b). The precisely optimized surface structure is expected to improve the binding affinity and target selectivity. (3) In theory, the artificial codon-box division method allows for the synthesis of chemically modified proteins composed of 20 pAAs along with multiple chemical modifications (Fig. 37c).

Recent advances in the methods of genetic code expansion[74,82,88-91,96,142] and genetic code reprogramming[59,60,92,128,129] have yielded many applications for protein engineering and macrocyclic peptide ligand discovery. Importantly, the concept of the artificial division of codon boxes demonstrated in this work introduces a new methodology to *in vitro* synthetic biology, and I envision that this methodology is, in principle, also applicable to cellular synthetic biology.

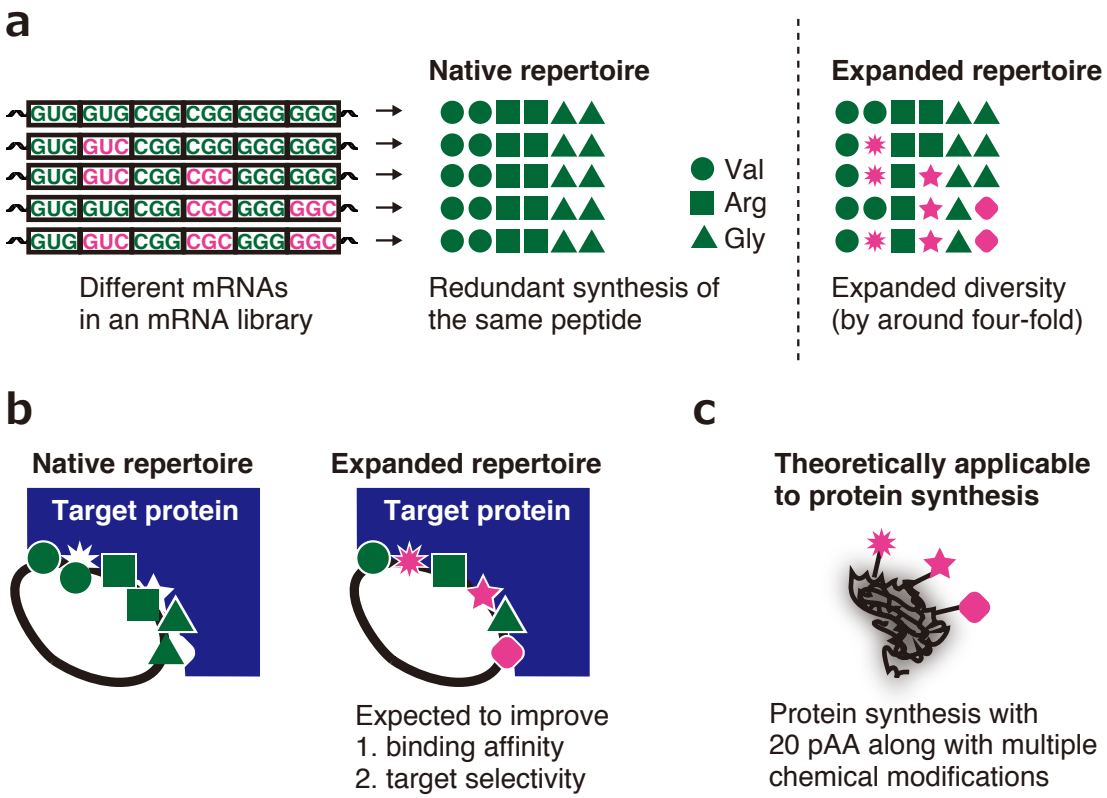

Figure 37. Significance of the expansion in building block repertoire. a, Increase in the diversity of a peptide library, which is expected to improve the binding properties of the discovered peptide ligands. **b**, Construction of binding interface with expanded structural and chemical repertoire. **c**, Synthesis of protein with 20 pAAs along with multiple chemical modifications.

2.4. Supplementary results

2.4.1. Preparation of tRNAs that do not have a guanosine at the 5'-terminus ($tRNA^{Gln}_{CUG}$, $tRNA^{Pro}_{CGG}$, $tRNA^{Pro}_{GGG}$, and $tRNA^{Trp}_{GCA}$)

As tRNAs that possess a 5′-terminal non-guanine nucleotide ($tRNA^{Gln}_{CUG}$, $tRNA^{Pro}_{CGG}$, $tRNA^{Pro}_{GGG}$, and $tRNA^{Trp}_{GCA}$) cannot be efficiently synthesized by T7 RNA polymerase[143], such tRNAs were transcribed with a 5′-terminal leader sequence (5′-GGAACGCGCGACUCUAAU-3′) (Fig. 38). Subsequently, the leader peptides were removed by RNase P reaction (Fig. 38b) and the tRNA transcripts were purified by denaturing PAGE (Fig. 38c).

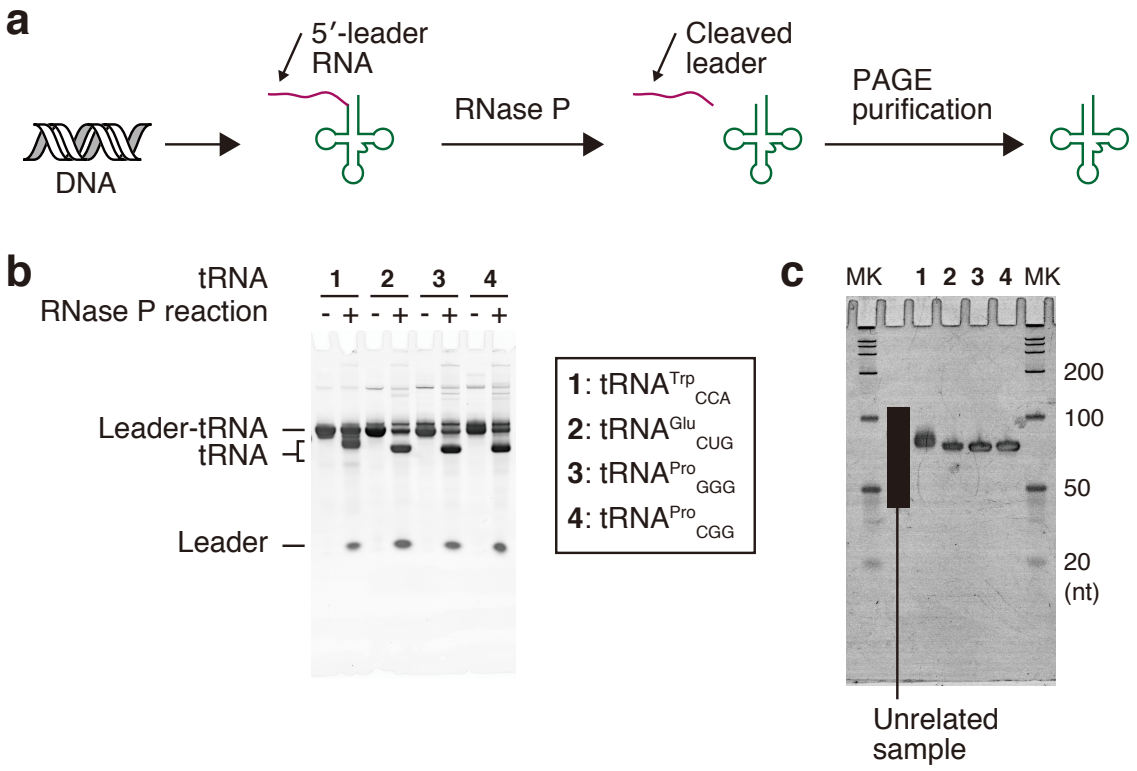

Figure 38. Preparation of tRNA transcripts that do not have a guanosine at the 5′-terminus. a, Method to prepare the four tRNA transcripts that possess a non-guanine nucleotide at the 5′-terminus. **b**, Confirmation of the leader-RNA cleavage by RNase P reaction. The RNA samples were analyzed by denaturing PAGE followed by ethidium bromide staining. **c**, Purity confirmation of the desired tRNA transcripts after PAGE purification.

2.4.2. Concentration optimization of three tRNA transcripts that cannot be acylated by AARSs efficiently.

It is known that the lack of the enzymatic nucleoside modifications in $tRNA^{Lys}$, $tRNA^{Ile}$, and $tRNA^{Glu}$ decreases their k_{cat}/K_M values for aminoacylation by their AARS, which results in less pAA-tRNA formation[6,7,35-38]. Accordingly, I first tried to compensate the inefficient acylation by increasing the concentrations of the respective $tRNA_{SNN}$. The mRNA1 was translated in the minimal FIT system in the presence of various concentrations of $tRNA^{Lys}_{CUU}$ transcripts and the expression levels of peptide products were evaluated by autoradiography after tricine-SDS-PAGE (Fig. 39a,b). The translation efficiency was improved by elevating the $tRNA^{Lys}_{CUU}$ concentration up to 30 μM and remained the same beyond 30 μM (Fig. 39c). MALDI-TOF-MS analysis of the peptide product expressed with 30 μM $tRNA^{Lys}_{CUU}$ revealed that the correct peptide product containing five Lys was synthesized accurately (Fig. 39d). When the expression levels were compared to a control condition with native tRNA mixture, they were improved from 12% to 54% by increasing $tRNA^{Lys}_{CUU}$ concentration from 5 to 30 μM, respectively (Fig. 39e). The expression level under the optimized condition was comparable to another control that contained the optimal concentration (5 μM) of native isolated $tRNA^{Lys}_{mnm5s2UUU}$. Based on these results, I concluded that the inefficiency problem derived from $tRNA^{Lys}$ transcript was successfully overcome by its concentration optimization. It should be mentioned that only the native isolated $tRNA^{Lys}_{mnm5s2UUU}$ was available as a kind gift from Prof. Suzuki's laboratory, therefore other native tRNAs could not be used for the reconstruction.

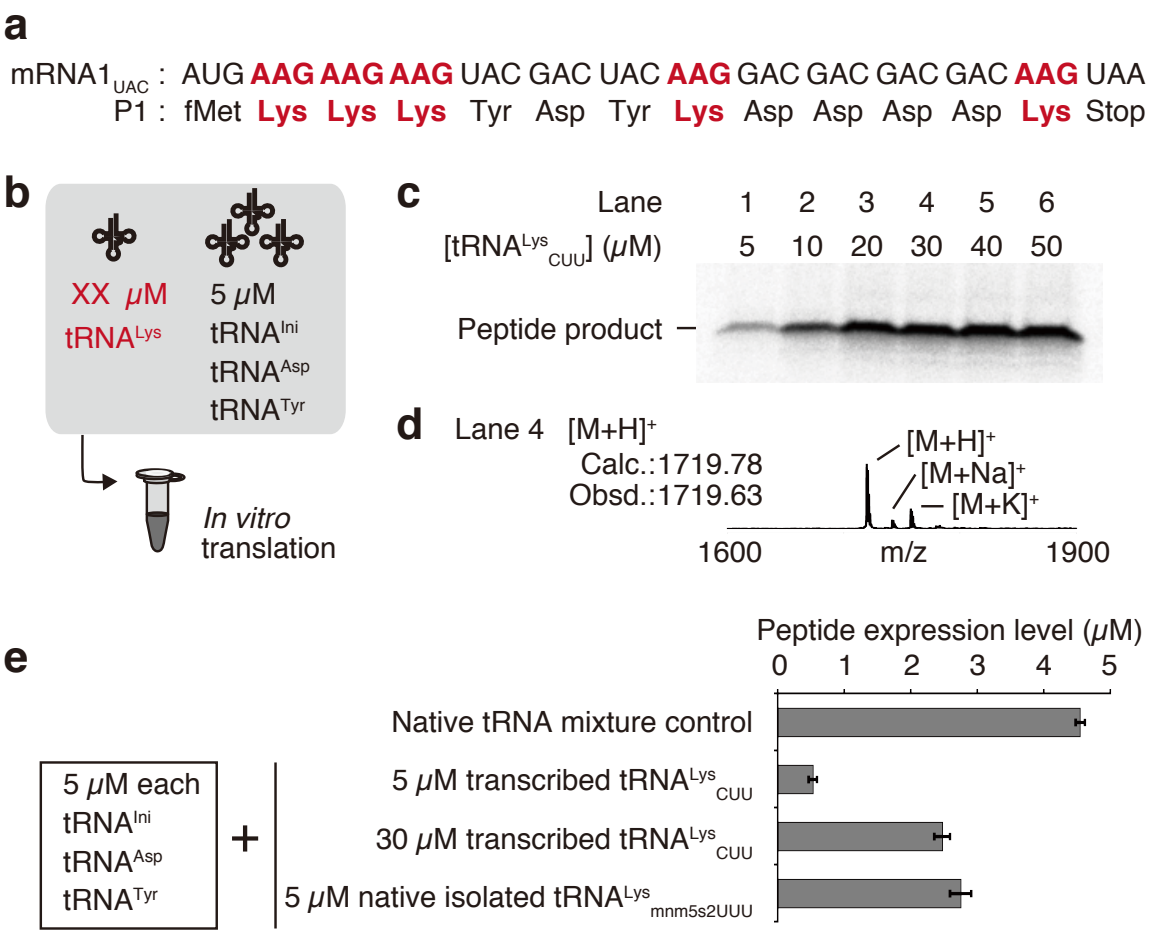

Figure 39. Improvement of the peptide expression level by optimization of $tRNA^{Lys}_{CUU}$ concentration. a, Sequences of the mRNA and its resultant P1 peptide. **b**, The minimal FIT system that contains 5 μM each $tRNA^{Ini}_{CAU}$, $tRNA^{Asp}_{GUC}$, $tRNA^{Tyr}_{GUA}$, and various concentrations of $tRNA^{Lys}_{CUU}$. **c,d**, Optimization of $tRNA^{Lys}_{CUU}$ concentration. The peptide products were analyzed by tricine-SDS-PAGE (**c**) and MALDI-TOF-MS (**d**). **e**, Confirmation of the alleviated inefficiency problem derived from the use of transcribed $tRNA^{Lys}_{CUU}$. Error bar: standard deviation.

Next, the concentrations of $tRNA^{Ile}_{GAU}$ and $tRNA^{Glu}_{CUC}$ were optimized. In the minimal FIT system supplied with a fixed 30 μM $tRNA^{Lys}_{CUU}$ concentration, an increase in the $tRNA^{Ile}_{GAU}$ concentration to 30 μM improved the expression yield (Fig. 40a). At 30 μM of $tRNA^{Ile}_{GAU}$, the correct product was expressed as confirmed by its MALDI-TOF-MS analysis. Similarly, the elevation of $tRNA^{Glu}_{CUC}$ concentration not only increased the expression level, but also suppressed an unknown side-product band (Fig. 40b). At 30 μM of $tRNA^{Glu}_{CUC}$, the expression of the correct product was confirmed by MALDI-TOF-MS, which indicated that the GAG codon was decoded accurately by $tRNA^{Glu}_{CUC}$.

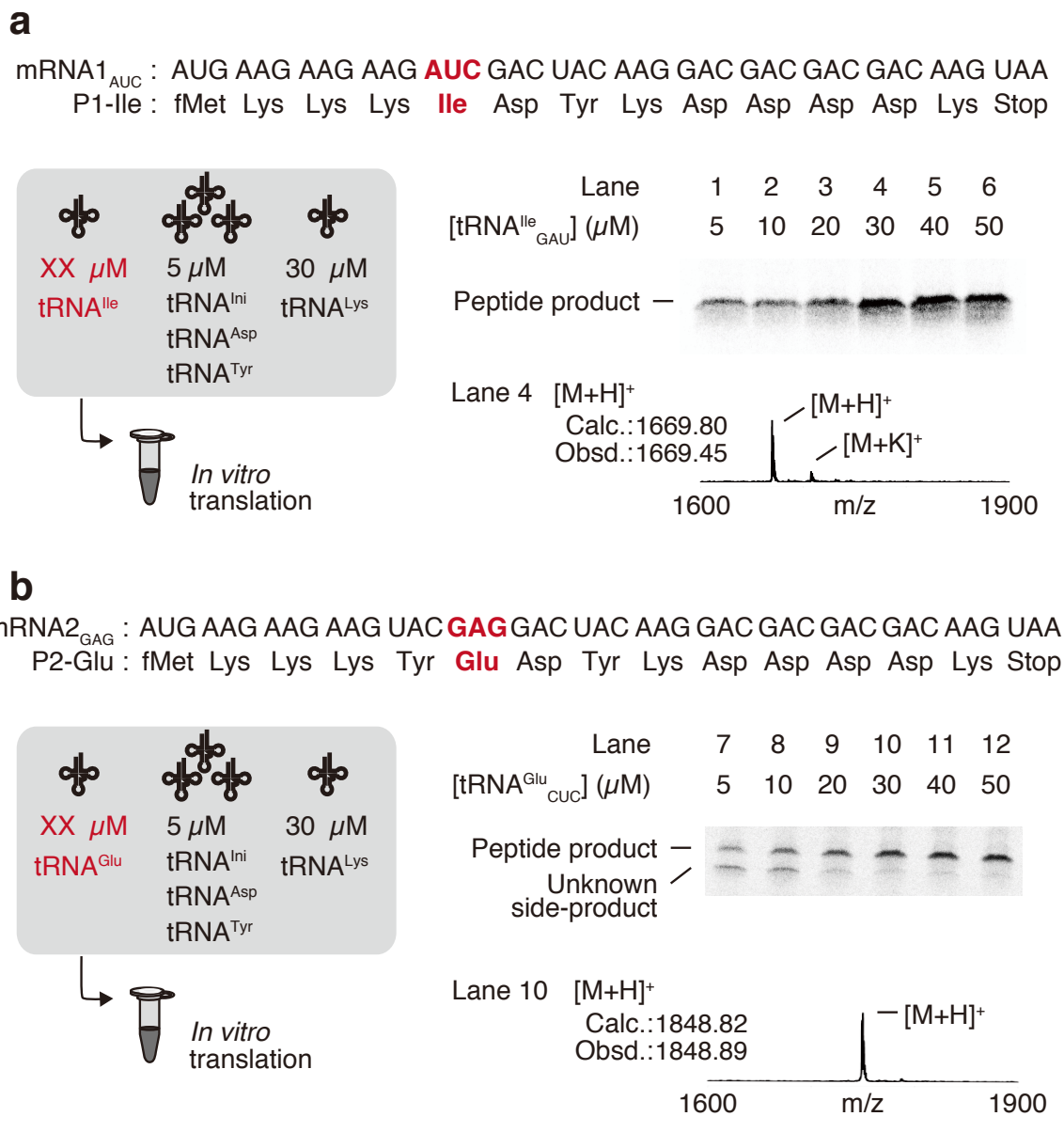

Figure 40. Improvement of the peptide expression levels by optimization of $tRNA^{Ile}_{GAU}$ and $tRNA^{Glu}_{CUC}$ concentrations. Optimization of $tRNA^{Ile}_{GAU}$ (**a**) and $tRNA^{Glu}_{CUC}$ (**b**) concentrations. Efficient and accurate translation was confirmed by tricine-SDS-PAGE and MALDI-TOF-MS of the peptide products.

2.4.3. Optimizations of translation conditions to achieve artificial division of a single codon box

To realize the concept of 'artificial codon-box division', I attempted to divide the arginine CGN codon box, and tried to assign Arg and *N*-serine (^{Me}S) to CGG and CGC codons, respectively, by the addition of $tRNA^{Arg}_{CCG}$, which would be *in situ* charged with Arg by ArgRS, and/or an orthogonal $tRNA^{AsnE2}_{GCG}$ pre-charged with ^{Me}S by dFx catalysis. In this system, the $mRNA1_{CGG}$ and $mRNA1_{CGC}$ templates should be translated to P1-Arg and P1-^{Me}S peptides, possessing Arg and ^{Me}S at the fifth position, respectively (Fig. 41a).

As a positive control, $mRNA1_{CGG}$ was translated in the FIT-nt system composed of native tRNAs, the desired P1-Arg, was observed in tricine-SDS-PAGE analysis (Fig. 41b, lane 1). When the minimal FIT system was used, only a negligible faint band with the same migration as P1-Arg was detected (Fig. 41b, lane 2), presumably due to an unavoidable trace amount of contamination of native $tRNA^{Arg}$ in the FIT system. Addition of $tRNA^{Arg}_{CCG}$ dramatically increased the expression level with a comparable efficiency to the FIT-nt system (Fig. 41b, lanes 1 versus 3), and further inclusion of ^{Me}S-$tRNA^{AsnE2}_{GCG}$ did not disturb the expression of P1-Arg (Fig. 41b, lanes 3 versus 4). The above results indicated that ^{Me}Y-$tRNA^{AsnE2}_{GAC}$ does not cross-read the GUG codon.

I next examined the expression of P1-^{Me}S from $mRNA1_{CGC}$ in the presence of ^{Me}S-$tRNA^{AsnE2}_{GCG}$ and/or $tRNA^{Arg}_{CCG}$. This template produced P1-Arg in the FIT-nt system (Fig. 41b, lane 5). On the other hand, the minimal FIT system alone produced a band with the same migration as P1-Val supposedly due to the contamination of an unavoidable trace amount of contamination of native $tRNA^{Arg}$ in the FIT system (Fig. 41b, lane 6). When ^{Me}S-$tRNA^{AsnE2}_{GCG}$ was added to the minimal FIT system, P1-^{Me}Y was observed at a slower migrated position, whereas the undesired expression of P1-Arg was still observed (Fig. 41b, lane 7). The expression level of the undesired P1-Arg was further increased when $tRNA^{Arg}_{CCG}$ was added (Fig. 41b, lane 7 versus 8), suggested that the CGC codon was mis-decoded by Arg-$tRNA^{Arg}_{CCG}$ although there is one base mismatch at the wobble position (C_3-C_{34}) (Fig. 41c). I supposed that the cause of misincorporation of Arg might be attributed to the problem that the concentration of pre-charged ^{Me}S-$tRNA^{AsnE2}_{GCG}$ gradually decrease as the reaction continues, whereas the concentration of Arg-$tRNA^{Arg}_{CCG}$ should be constant as the $tRNA^{Arg}_{CCG}$ are continuously charged with Arg by ArgRS *in situ* (Fig. 41d).

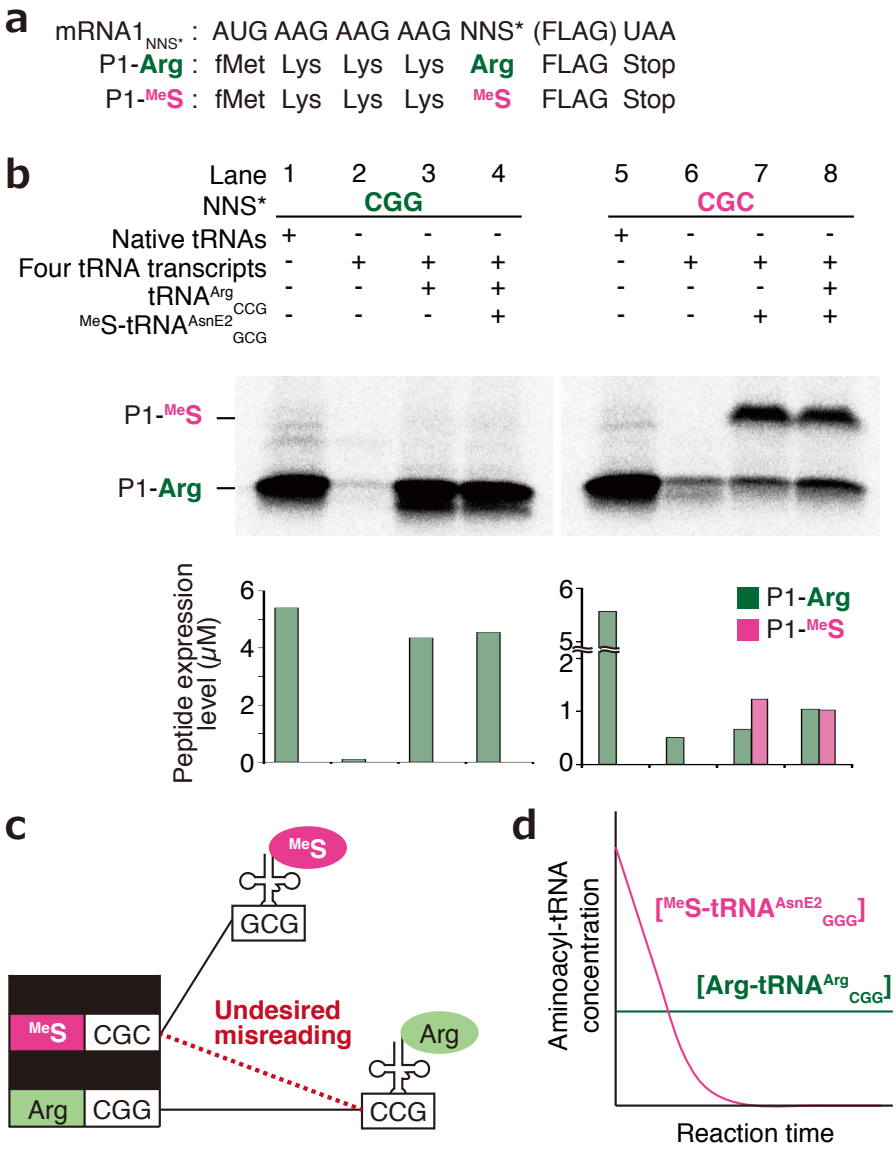

Figure 41. Inaccurate translation of artificially divided Arg codon box observed in the unoptimized condition. a, Sequences of mRNA1$_{NNS*}$ and its resultant P1 peptides. CGG and CGC codons present at the fifth position of mRNA1$_{NNS*}$ were decoded by Arg-tRNA$^{Arg}_{CCG}$ and ^{Me}S-tRNA$^{AsnE2}_{GCG}$, respectively. **b**, Verification of the decoding of CGG and CGC codons, present in the Arg codon box, to Arg and ^{Me}S, respectively. Native tRNAs: 1.5 mg/mL *E. coli* total tRNA; 4 *in vitro* tRNA transcripts: 5 μM each tRNA$^{Ini}_{CAU}$, tRNA$^{Asp}_{GUC}$, tRNA$^{Tyr}_{GUA}$, and 30 μM tRNA$^{Lys}_{CUU}$; tRNA$^{Arg}_{CCG}$: 5 μM tRNA$^{Arg}_{CCG}$; ^{Me}S-tRNA$^{AsnE2}_{GCG}$: 50 μM tRNA$^{AsnE2}_{GCG}$ pre-charged with ^{Me}S. **c**, The undesired observation that the CGC codon was misread by Arg- tRNA$^{Arg}_{CCG}$. **d**, A possible cause of the misreading problem: a gradual decrease of ^{Me}S-tRNA$^{AsnE2}_{GCG}$ concentration depending on reaction time.

In order to avoid the unbalanced the aminoacyl-tRNA concentrations, the reaction time was decreased from 30 min to 10 min (Fig. 42). As expected, the relative expression of P1-^{Me}S to P1-Arg was improved by reducing reaction time (Fig. 42a,b, lane 5–8), but the undesired P1-Arg was still expressed as confirmed by MALDI-TOF-MS of the peptide products (Fig. 42c, lane 8). Therefore, further optimization of translation condition was required to achieve artificial division of a codon box with maintained translation accuracy and efficiency.

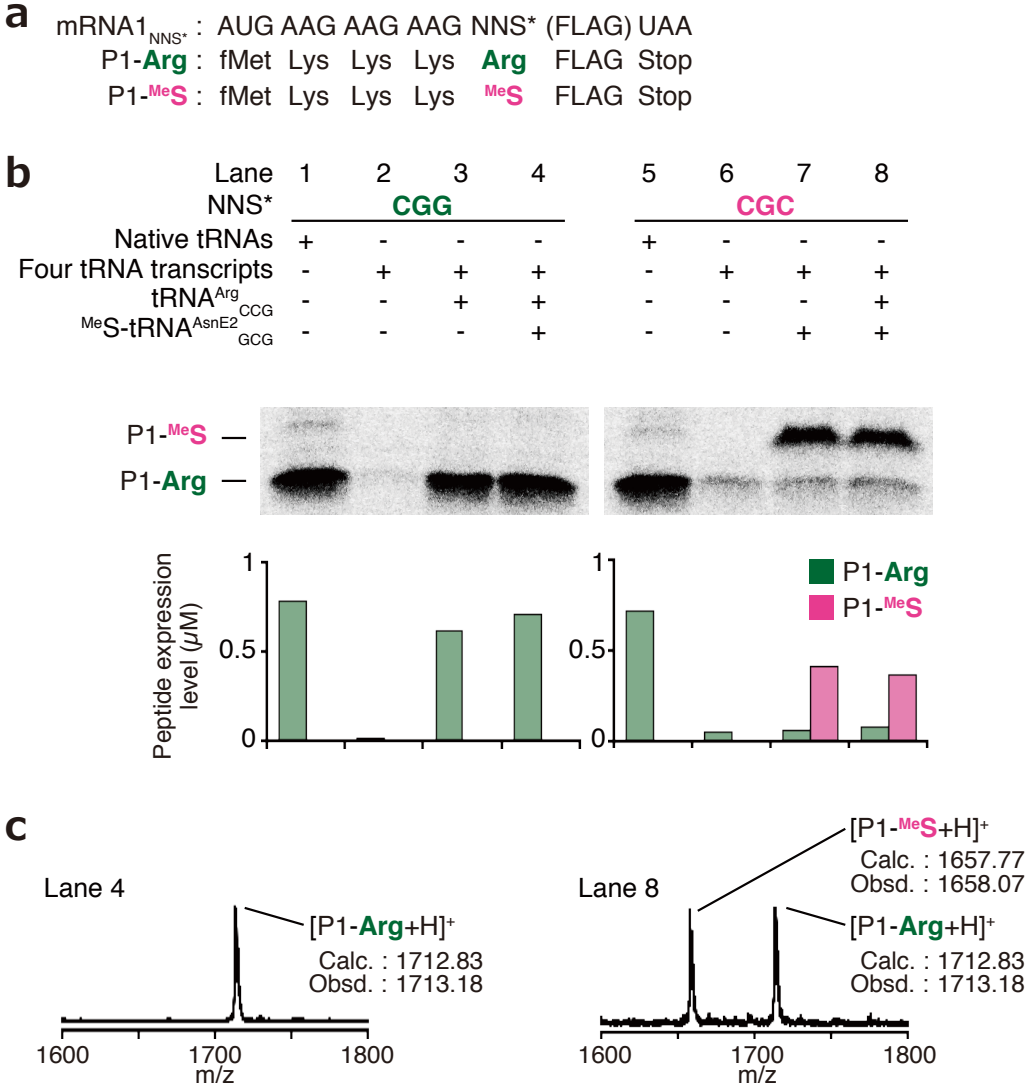

Figure 42. Alleviated inaccuracy problem by reducing the reaction time from 30 to 10 min. a, Sequences of mRNA1$_{NNS*}$ and its resultant P1 peptides. CGG and CGC codons present at the fifth position of mRNA1$_{NNS*}$ were decoded by Arg-tRNA$^{Arg}_{CCG}$ and ^{Me}S-tRNA$^{AsnE2}_{GCG}$, respectively. **b,c**, Verification of the decoding of CGG and CGC codons, present in the Arg codon box, to Arg and ^{Me}S, respectively. The peptide products were analyzed by tricine-SDS-PAGE (**b**) and MALDI-TOF-MS (**c**). Native tRNAs: 1.5 mg/mL *E. coli* total tRNA; 4 *in vitro* tRNA transcripts: 5 μM each tRNA$^{Ini}_{CAU}$, tRNA$^{Asp}_{GUC}$, tRNA$^{Tyr}_{GUA}$, and 30 μM tRNA$^{Lys}_{CUU}$; tRNA$^{Arg}_{CCG}$: 5 μM tRNA$^{Arg}_{CCG}$; ^{Me}S-tRNA$^{AsnE2}_{GCG}$: 50 μM tRNA$^{AsnE2}_{GCG}$ pre-charged with ^{Me}S.

Regarding the inefficiency of P1-^{Me}S expression, I hypothesized the cause could be attributed to either of (1) inefficient delivery of ^{Me}S-tRNA$^{AsnE2}_{GCG}$ by EF-Tu and (2) slow accommodation and peptidyl transfer reaction (Fig. 43a). If either of these processes was hindered by the use of ^{Me}S, the Arg misincorporation could occur with slow kinetics (Fig. 43a). Therefore, I attempted to overcome the problems as follows. First, the problem of inefficient delivery of ^{Me}S-tRNA$^{AsnE2}_{GCG}$ by EF-Tu could be alleviated by increasing the concentration of ^{Me}S-tRNA$^{AsnE2}_{GCG}$. Second, the problem of slow accommodation and peptidyl transfer involving ^{Me}S could be overcome by optimizing Mg^{2+} concentration, as it has been reported that the increase in Mg^{2+} concentration dramatically increases these steps[16,144] (Fig. 43b).

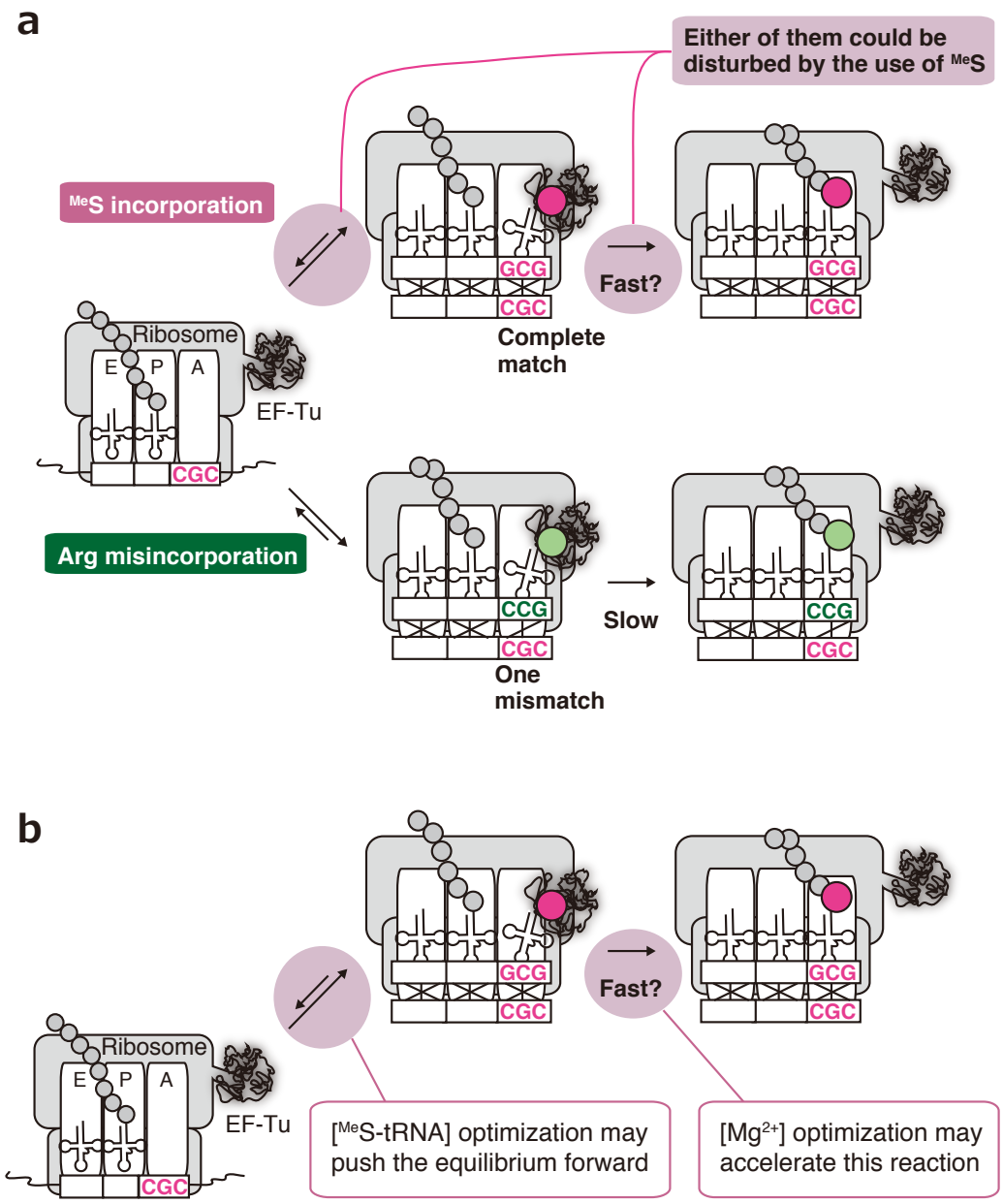

Figure 43. Hypothesized causes of the observed inaccurate translation and methods to overcome the problem. a, The competition between the desired ^{Me}S incorporation and undesired Arg misincorporation. If the desired ^{Me}S incorporation were inefficient, the undesired Arg misincorporation would occur with slow kinetics. **b**, Two approaches to accelerate the ^{Me}S incorporation.

Based on this hypothesis, the concentration of ^{Me}S-tRNA$^{AsnE2}_{GCG}$ and $Mg(OAc)_2$ were optimized. The mRNA1$_{CGC}$ that code for P1-^{Me}S peptide was translated in the minimal FIT system that contained both ^{Me}S-tRNA$^{AsnE2}_{GCG}$ and tRNA$^{Arg}_{CCG}$ (Fig. 44a). When the concentration of $Mg(OAc)_2$ was increased, the relative expression levels of desired P1-^{Me}S to undesired P1-Arg was improved from 1.8 to 15.9 (Fig. 44c, lanes 1–4). When the concentration of tRNA$^{AsnE2}_{GCG}$ pre-charged with ^{Me}S was increased from 50 μM to 100 μM, the relative expression levels of P1-^{Me}S was increased by nearly twofold, and the relative expression levels of P1-^{Me}S to P1-Arg was improved further (Fig. 44b, lanes 5–8). Among these conditions, the condition of 13 mM $Mg(OAc)_2$ and 100 μM tRNA$^{AsnE2}_{GCG}$ pre-charged with ^{Me}S was chosen as the optimal condition.

The accurate decoding of the artificially divided Arg codon box was examined by translation of mRNA7 containing both CGC and CGG codons in the Arg CGN codon box (Fig. 44c). The MALDI-TOF-MS of the peptide product demonstrated that the correct peptide containing ^{Me}S and Arg was accurately synthesized.

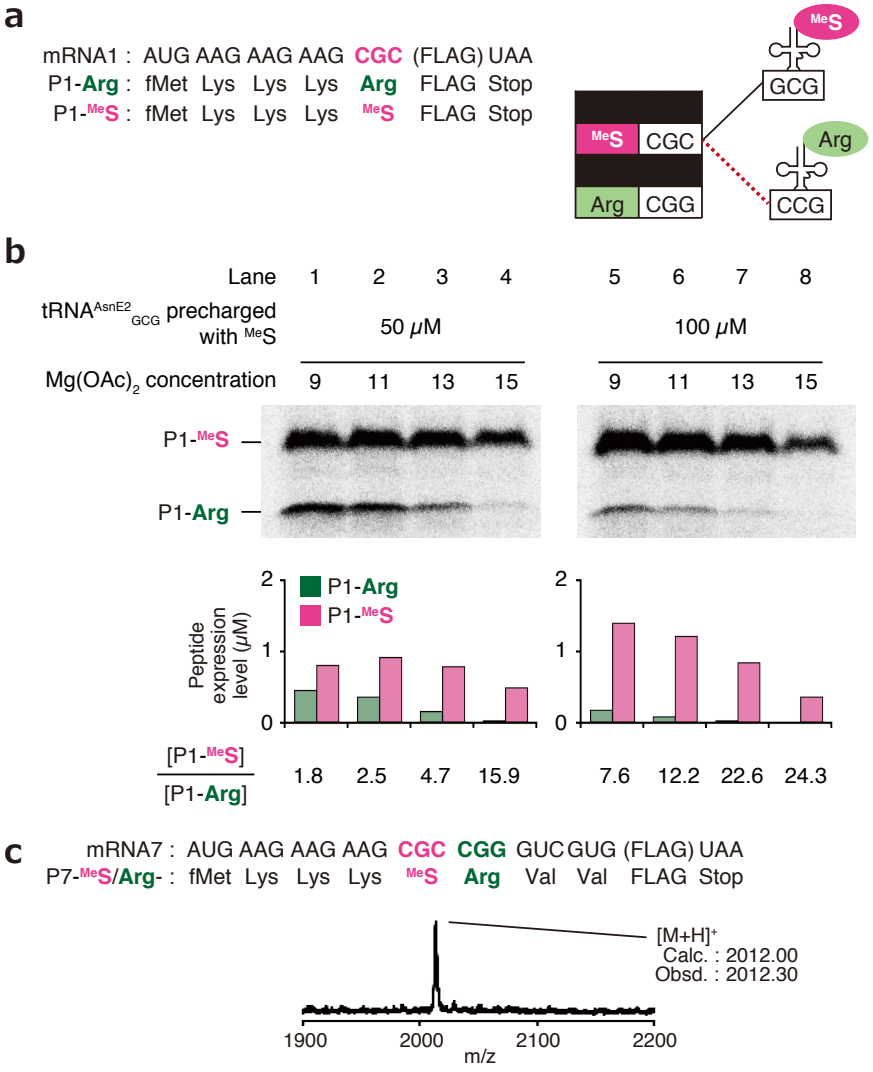

Figure 44. Improvement of the translation efficiency and accuracy of artificially divided Arg codon box. **a**, Sequences of mRNA1$_{CGC}$ and its resultant P1 peptides. **b,** Concentration optimization of tRNA$^{AsnE2}_{GCG}$ pre-charged with ^{Me}S and $Mg(OAc)_2$. The peptide products were analyzed by tricine-SDS-PAGE. **c,** Accurate decoding of the CGC and CGG codons demonstrated by MALDI-TOF-MS of the peptide product.

Next, the synthetic efficiency under the optimized condition was compared to the original condition that had been utilized in our laboratory. For this experiment, the Val GUN codon box was divided, and mRNA1$_{GUC}$ coding P1-^{Me}Y was translated in the minimal FIT system that contained both ^{Me}Y-tRNA$^{AsnE2}_{GAC}$ and tRNA$^{Val}_{CAC}$ (Fig. 45a). In case of the original condition, the expression of desired P1-^{Me}Y reached plateau at 8 min, and afterwards the undesired P1-Val appeared and its expression level reached comparable level to P1-^{Me}Y at 30 min (Fig. 45b). Under the optimized condition, in contrast, the expression of P1-^{Me}Y was sustained until 20 min and the undesired expression of P1-Val was successfully suppressed as confirmed by both tricine-SDS-PAGE and MALDI-TOF-MS analyses of the peptide products (Fig. 45c).

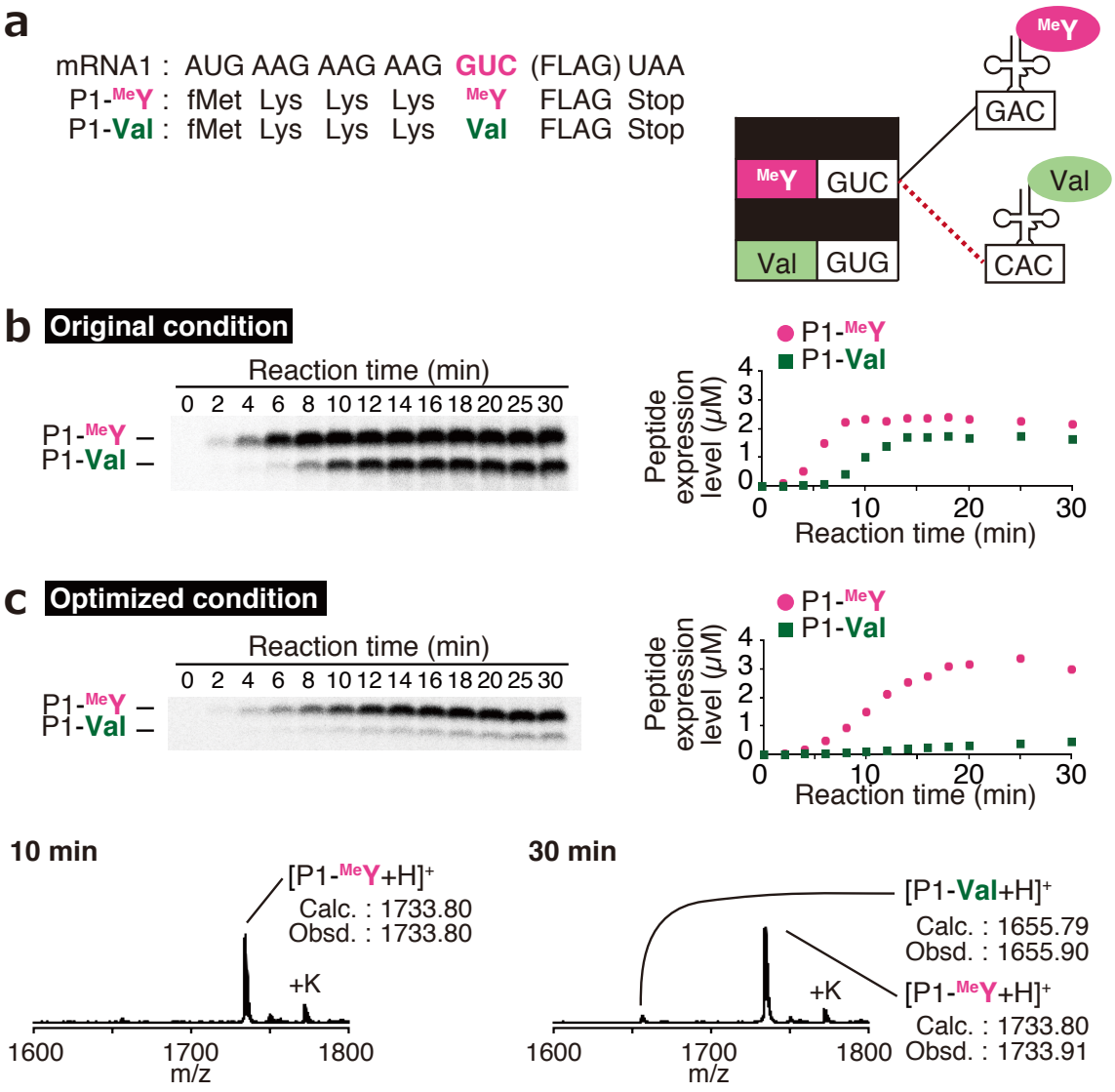

Figure 45. Examination of the improved efficiency and accuracy by the optimization. a, Sequences of mRNA1$_{GUC}$ and its resultant P1 peptides. **b**, Inaccurate translation of the GUC codon under the original condition. **c**, Accurate translational under the optimized condition. The peptide products were analyzed by tricine-SDS-PAGE and MALDI-TOF-MS.

2.5. Methods

2.5.1. Preparation of tRNA transcripts, flexizymes, and mDNAs

All the oligonucleotides listed in Supplementary Table 1 were purchased from Eurofins, Japan. The sequences of tRNA transcripts prepared in this study are listed in Supplementary Table 2. DNA templates that coded for tRNAs were prepared by primer extension followed by two-step PCR as follows: Appropriate forward and reverse primers (1 μM each, see Supplementary Table 3 for the primers used) were mixed in the PCR mixture (10 mM Tris·HCl (pH 9.0), 50 mM KCl, 2.5 mM $MgCl_2$, 0.25 mM each dNTPs, 0.1% (v/v) Triton X-100, and 45 nM *Taq* DNA polymerase). Primer extension was conducted in a 10-μL reaction mixture by denaturing (95 °C for 1 min) followed by 5 cycles of annealing (50 °C for 1 min) and extending (72 °C for 1 min). The resulting reaction mixture was 10-fold diluted with the PCR mixture, and the extension product was amplified using the specific forward and reverse primers (0.5 μM each, see Supplementary Table 3 for the primers used). PCR was conducted in a 200-μL reaction mix by 5 cycles of denaturing (95 °C for 40 sec), annealing (50 °C for 40 sec), and extending (72 °C for 40 sec). The resulting PCR mixture was 200-fold diluted with the PCR mixture, and the 1st PCR product was amplified using the designated forward and reverse primers (0.5 μM each, see Supplementary Table 3 for the primers used). The 2nd PCR was conducted in a 200-μL reaction mix by 12 cycles of denaturing (95 °C for 40 sec), annealing (50 °C for 40 sec), and extending (72 °C for 40 sec). Amplification of the final PCR product was confirmed by 3% agarose gel electrophoresis and ethidium bromide staining. The resulting DNA was purified by phenol/chloroform extraction and ethanol precipitation, and then dissolved in 20 μL of water.

Transcription reaction was conducted by incubating the 200-μL transcription mixture (40 mM Tris·HCl (pH 8.0), 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 22.5 mM $MgCl_2$, 3.75 mM each NTPs, 5 mM GMP, 22.5 mM KOH, 10% (v/v) DNA template prepared above, and 120 nM T7 RNA polymerase) at 37 °C overnight. The transcription mixture was mixed with $MnCl_2$ (100 mM, 4 μL) and RQ1 RNase-Free DNase (1 U/μL, 1μL, Promega), and incubated at 37 °C for 30 min. The resultant tRNA transcript was precipitated by isopropanol and dissolved in water. The tRNA transcript was purified by 8% denaturing PAGE and ethanol precipitation, and then dissolved in 10 μL water. For the preparation of tRNAs without a 5'-terminal guanine nucleotide ($tRNA^{Gln}_{CUG}$, $tRNA^{Pro}_{CGG}$, $tRNA^{Pro}_{GGG}$, and $tRNA^{Trp}_{GCA}$), tRNAs with a 5'-terminal leader sequence (5'-GGAACGCGCGACUCUAAU-3') were transcribed, and the leader sequences were removed by RNase P reaction as follows: the tRNA transcript with the 5'-leader was prepared by the 200 μL-scale *in vitro* transcription and DNase reaction as described above, and it was mixed with $MgCl_2$ (3 M, 1.34 μL) and 1 μL of RNase P solution (40 mM Tris·HCl (pH 8.0), 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 50 μM C5 protein, and 50 μM M1 RNA), and incubated at 37 °C overnight. The tRNA transcript was purified by 8% denaturing PAGE as described above.

Flexizymes (dFx and eFx) were prepared as previously described[77,92] using primers O-1, O-138, O-139, and O-141 for dFx, and O-1, O-138, O-140, and O-142 for eFx.

Double stranded DNA templates that coded for mRNAs (mDNAs, see Supplementary Table 4 for the mRNA sequences) were prepared by primer extension and PCR (see Supplementary Table 5 for the primers used). The resulting mDNAs were purified by phenol/chloroform extraction and ethanol precipitation, and then dissolved in 10

μL of water. The concentration was measured by 8% native PAGE and ethidium bromide staining using 100 bp Quick-Load DNA Ladders (New England BioLabs) as reference.

2.5.2. Synthesis of npAA-tRNAs

Amino acids activated with an appropriate ester groups (^{Me}Y-CME, ^{Iodo}F-CME, ClAc^{D}W-CME, ^{Me}F-CME, ^{Me}S-DBE, ^{Ac}K-DBE, ^{Me}G-DBE, and Cit-DBE) were synthesized by the procedure reported previously[92,104]. Aminoacyl-tRNAs for 5 μL translation reactions were prepared as follows: 12 μL HEPES-KOH buffer (pH 7.5, 83 mM) that contained 42 μM tRNA and 42 μM flexizyme (eFx for CME-activated amino acids and dFx for DBE-activated ones) was heated at 95 °C for 2 min and cooled to 25 °C over 5 min. $MgCl_2$ (3 M, 4 μL) was added and the mixture was incubated at 25 °C for 5 min. The acylation reaction was initiated by the addition of activated amino acid substrate in DMSO (25 mM, 4 μL) and the mixture was incubated on ice for appropriate time (1 h for ClAc^{D}W-CME; 2 h for ^{Iodo}F-CME, ^{Ac}K-DBE, ^{Me}G-CME, and Cit-DBE; 6 h for ^{Me}S-DBE and ^{Me}F-CME; and 10 h for ^{Me}Y-CME). After acylation, the reaction was stopped by the addition of 80 μL of 0.3 M sodium acetate (pH 5.2), and the RNA was precipitated by ethanol precipitation. The pellet was rinsed twice with 70% ethanol that contained 0.1 M sodium acetate (pH 5.2), and once with 70% ethanol. The resulting npAA-tRNA was dissolved in 1 mM sodium acetate (pH 5.2) just before adding it to the translation solution.

2.5.3. *In vitro* translation

The reconstituted *in vitro* translation system[32] used in this study contained all the necessary components for translation except for tRNAs and RF1. Concentrations of translation components have been optimized in our previous studies[77] as follows: 50 mM HEPES-KOH (pH 7.6), 100 mM KOAc, 2 mM GTP, 2 mM ATP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 9–15 mM $Mg(OAc)_2$ (see Supplementary Table 6 for $Mg(OAc)_2$ concentration in each experiment), 2 mM spermidine, 1 mM DTT, 100 μM 10-formyl-5,6,7,8-tetrahydrofolic acid, 1.2 μM ribosome, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 10 μM EF-Tu, 10 μM EF-Ts, 0.26 μM EF-G, 0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 0.6 μM MTF, 4 μg/mL creatine kinase, 3 μg/mL myokinase, 0.1 μM pyrophosphatase, 0.1 μM nucleotide diphosphate kinase, 0.1 μM T7 RNA polymerase, 0.73 μM AlaRS, 0.03 μM ArgRS, 0.38 μM AsnRS, 0.13 μM AspRS, 0.02 μM CysRS, 0.06 μM GlnRS, 0.23 μM GluRS, 0.09 μM GlyRS, 0.02 μM HisRS, 0.4 μ IleRS, 0.04 μM LeuRS, 1.1 μM LysRS, 0.03 μM MetRS, 0.68 μM PheRS, 0.16 μM ProRS, 0.04 μM SerRS, 0.09 μM ThrRS, 0.03μM TrpRS 0.02 μM TyrRS, 0.02 μM ValRS, 200 μM each 20 pAAs, 0.25 μM mDNA, appropriate concentrations of tRNA transcripts (see Supplementary Table 7 for their concentrations in each experiment) and/or 50–100 μM each tRNA^{AsnE2}s pre-charged with npAAs (see Supplementary Table 6 for their concentration in each experiment) unless noted otherwise in the figure legends. For control experiments where 1.5 mg/mL native *E. coli* tRNA mixture (Roche) was used, the concentrations of $Mg(OAc)_2$ and LysRS were adjusted to 12 mM and 0.11 μM, respectively, being the optimal concentration.

The translation products were analyzed by MALDI-TOF-MS and/or autoradiography after tricine-SDS-PAGE. For MALDI-TOF-MS analysis, the translation reaction mixture (5 μL) was incubated at 37 °C for 0–30 min (see

Supplementary Table 6 for the incubation time in each experiment). After the reaction, the mixture was diluted with 5 μL of FLAG purification buffer (100 mM Tris-HCl (pH 7.6), 300 mM NaCl). The expressed peptide was immobilized on anti-FLAG M2 agarose beads (Sigma) by incubating at 25 °C for 1 hour. After washing the beads with 25 μL of wash buffer (50 mM Tris-HCl (pH 7.6), 150 mM NaCl), the immobilized peptides were eluted with 15 μL of 0.2% TFA. The FLAG purification procedure was omitted in some experiments (Supplementary Table 6). After purification, the peptide was desalted with SPE C-tip (Nikkyo Technos) and eluted with 1 μL of 80% acetonitrile, 0.5% acetic acid solution 50% saturated with the matrix (R)-cyano-4-hydroxycinnamic acid (Bruker Daltonics). MALDI-TOF-MS measurement was conducted using an autoflex II TOF/TOF (Bruker Daltonics) or an ultrafleXtreme (Bruker Daltonics) under linear/positive mode and externally calibrated with peptide calibration standard II (Bruker Daltonics) and/or protein calibration standard I (Bruker Daltonics).

For autoradiography analysis, the translation reaction (2.5 μL) was conducted in the presence of 50 μM [^{14}C]-Asp instead of 200 μM cold Asp. The translation product was analyzed by 15% tricine-SDS-PAGE and autoradiography using an FLA-5100 (Fujifilm Life Science) or a Typhoon FLA 7000 (GE Healthcare) without FLAG purification. The amount of peptide products was quantified based on the band intensities of known amounts of [^{14}C]-Asp as calibration standards.

For translation of mRNA10 and mRNA11 (Fig. 35 and 36), several translation conditions were changed. First, the translation solution contained 1 mM TCEP (tris(2-carboxyethyl)phosphine) in order to prevent disulfide bond formation at the cysteine residue. Second, the concentration of HEPES-KOH (pH 7.6) was increased from 50 mM to 100 mM to prevent pH change caused by 1 mM TCEP. Third, the volume of translation solution to synthesize P10-^{Ac}K^{9}/^{Iodo}F^{12}/Cit^{18} (Fig. 6d lane 3) and CM$_{11}$-1 (Fig. 6h) was increased to 20 μL in order to improve the S/N ratio of MALDI-TOF-MS. For mRNA11 translation, methionine was omitted from the translation solution for the reprogramming of the AUG initiation codon. After translation, the reaction mixture (20 μL) was mixed with wash buffer (20 μL, 50 mM Tris-HCl (pH 7.6), 150 mM NaCl) and incubated at 25˚C for 60 min for macrocyclization.

2.6. Supplementary Tables

Supplementary Table 1. List of oligonucleotides used in this study.

Name	Sequence
O-1	5'-GGCGTAATACGACTCACTATAG-3'
O-2	5'-GTAATACGACTCACTATAGGAACGCGCGACTCTAAT-3'
O-3	5'-GTAATACGACTCACTATAGGCGGGGTGGAGCAGCCTGGTAGCTCGTCGG-3'
O-4	5'-GAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCT-3'
O-5	5'-TGGTTGCGGGGGCCGGATTTGAACCGACGATCTTCGGG-3'
O-6	5'-TGGTTGCGGGGGCCGGATTT-3'
O-7	5'-GTAATACGACTCACTATAGGGGCTATAGCTCAGCTGGGAGAGCGCCTGC-3'
O-8	5'-GAACCGCAGACCTCCTGCGTGCGAAGCAGGCGCTCTCCCAGCT-3'
O-9	5'-TGGTGGAGCTATGCGGGATCGAACCGCAGACCTCCTGC-3'
O-10	5'-TGGTGGAGCTATGCGGGATC-3'
O-11	5'-GTAATACGACTCACTATAGGGGCTATAGCTCAGCTGGGAGAGCGCCTGC-3'
O-12	5'-GAACCGCAGACCTCCTGCATGCCATGCAGGCGCTCTCCCAGCT-3'
O-13	5'-TGGTGGAGCTATGCGGGATCGAACCGCAGACCTCCTGC-3'
O-14	5'-TGGTGGAGCTATGCGGGATC-3'
O-15	5'-GTAATACGACTCACTATAGCGCCCGTAGCTCAGCTGGATAGAGCGCTGC-3'
O-16	5'-CGAACCTGAGACCTCTGCCTCCGGAGGGCAGCGCTCTATCCAGCT-3'
O-17	5'-TGGCGCGCCCGACAGGATTCGAACCTGAGACCTCTGC-3'
O-18	5'-TGGCGCGCCCGACAGGATTC-3'
O-19	5'-GTAATACGACTCACTATAGCGCCCGTAGCTCAGCTGGATAGAGCGCTGC-3'
O-20	5'-CGAACCTGAGACCTCTGCCTTAGGAGGGCAGCGCTCTATCCAGCT-3'
O-21	5'-TGGCGCGCCCGACAGGATTCGAACCTGAGACCTCTGC-3'
O-22	5'-TGGCGCGCCCGACAGGATTC-3'
O-23	5'-GTAATACGACTCACTATAGCGCCCGTAGCTCAGCTGGATAGAGCGCTGC-3'
O-24	5'-CGAACCTGAGACCTCTGCCTTCGCAGGGCAGCGCTCTATCCAGCT-3'
O-25	5'-TGGCGCGCCCGACAGGATTCGAACCTGAGACCTCTGC-3'
O-26	5'-TGGCGCGCCCGACAGGATTC-3'
O-27	5'-GTAATACGACTCACTATAGCCTCTGTAGTTCAGTCGGTAGAACGGCGGA-3'
O-28	5'-GAACCAGTGACATACGGATTAACAGTCCGCCGTTCTACCGACT-3'
O-29	5'-TGGCGCCTCTGACTGGACTCGAACCAGTGACATACGGA-3'
O-30	5'-TGGCGCCTCTGACTGGACTC-3'
O-31	5'-GTAATACGACTCACTATAGGAGCGGTAGTTCAGTCGGTTAGAATACCTG-3'
O-32	5'-GAACCCGCGACCCCCTGCGTGACAGGCAGGTATTCTAACCGACT-3'
O-33	5'-TGGCGGAACGGACGGGACTCGAACCCGCGACCCCCTGC-3'
O-34	5'-TGGCGGAACGGACGGGACTC-3'
O-35	5'-GTAATACGACTCACTATAGGCGCGTTAACAAAGCGGTTATGTA-3'
O-36	5'-CGCGAACCGGACTAGACGGATTTGCAGTCCGCTACATAACCGCTT-3'
O-37	5'-TGGAGGCGCGTTCCGGAGTCGAACCGGACTAGACGGA-3'
O-38	5'-TGGAGGCGCGTTCCGGAGTC-3'
O-39	5'-GGAACGCGCGACTCTAATTGGGGTATCGCCAAGCGGTAAGGCACCGGA-3'
O-40	5'-GAACCTCGGAATGCCGGAATCAGAGTCCGGTGCCTTACCGCTT-3'
O-41	5'-TGGCTGGGGTACGAGGATTCGAACCTCGGAATGCCGGA-3'
O-42	5'-TGGCTGGGGTACGAGGATTC-3'
O-43	5'-GTAATACGACTCACTATAGTCCCCTTCGTCTAGAGGCCCAGGACACCGC-3'
O-44	5'-CGAACCCCTGTTACCGCCGTGAGAGGGCGGTGTCCTGGGCCTCT-3'
O-45	5'-TGGCGTCCCCTAGGGGATTCGAACCCCTGTTACCGCC-3'
O-46	5'-TGGCGTCCCCTAGGGGATTC-3'
O-47	5'-GTAATACGACTCACTATAGCGGGCGTAGTTCAATGGTAGAACGAGAGC-3'
O-48	5'-CGAACCCTCGTATAGAGCTTGGGAAGCTCTCGTTCTACCATTG-3'
O-49	5'-TGGAGCGGGCGAAGGGAATCGAACCCTCGTATAGAGC-3'
O-50	5'-TGGAGCGGGCGAAGGGAATC-3'
O-51	5'-GTAATACGACTCACTATAGCGGGCGTAGTTCAATGGTAGAACGAGAGC-3'
O-52	5'-CGAACCCTCGTATAGAGCTTGGCAAGCTCTCGTTCTACCATTG-3'
O-53	5'-TGGAGCGGGCGAAGGGAATCGAACCCTCGTATAGAGC-3'
O-54	5'-TGGAGCGGGCGAAGGGAATC-3'
O-55	5'-GTAATACGACTCACTATAGGTGGCTATAGCTCAGTTGGTAGAGCCCTGG-3'
O-56	5'-GAACCCACGACAACTGGAATCACAATCCAGGGCTCTACCAACTG-3'
O-57	5'-TGGGGTGGCTAATGGGATTCGAACCCACGACAACTGGA-3'
O-58	5'-TGGGGTGGCTAATGGGATTC-3'
O-59	5'-GTAATACGACTCACTATAGGGCTTGTAGCTCAGGTGGTTAGAGCGCACC-3'
O-60	5'-GAACCACCGACCTCACCCTTATCAGGGGTGCGCTCTAACCACCT-3'
O-61	5'-TGGTGGGCCTGAGTGGACTTGAACCACCGACCTCACCC-3'
O-62	5'-TGGTGGGCCTGAGTGGACTT-3'
O-63	5'-GTAATACGACTCACTATAGCGAAGGTGGCGGAATTGGTAGACGCGCTAG-3'
O-64	5'-GTCCGTAAGGACACTAACTTTTGAAGCTAGCGCGTCTACCAATT-3'
O-65	5'-TGGTGCGAGGGGGGGGACTTGAACCCCCACGTCCGTAAGGACACTAAC-3'

Supplementary Table 1. (Continued)

Name	Sequence
O-66	5'-TGGTGCGAGGGGGGGGACTT-3'
O-67	5'-GTAATACGACTCACTATAGCGAAGGTGGCGGAATTGGTAGACGCGCTAG-3'
O-68	5'-GTCCGTAAGGACACTAACACCTGAAGCTAGCGCGTCTACCAATT-3'
O-69	5'-TGGTGCGAGGGGGGGGGACTTGAACCCCCACGTCCGTAAGGACACTAAC-3'
O-70	5'-TGGTGCGAGGGGGGGGGACTT-3'
O-71	5'-GTAATACGACTCACTATAGCGAAGGTGGCGGAATTGGTAGACGCGCTAG-3'
O-72	5'-GTCCGTAAGGACACTAACACCTCAAGCTAGCGCGTCTACCAATT-3'
O-73	5'-TGGTGCGAGGGGGGGGGACTTGAACCCCCACGTCCGTAAGGACACTAAC-3'
O-74	5'-TGGTGCGAGGGGGGGGGACTT-3'
O-75	5'-GTAATACGACTCACTATAGGGTCGTTAGCTCAGTTGGTAGAGCAGTTGA-3'
O-76	5'-GAACCTGCGACCAATTGATTAAGAGTCAACTGCTCTACCAACT-3'
O-77	5'-TGGTGGGTCGTGCAGGATTCGAACCTGCGACCAATTGA-3'
O-78	5'-TGGTGGGTCGTGCAGGATTC-3'
O-79	5'-GTAATACGACTCACTATAGGCTACGTAGCTCAGTTGGTTAGAGCACATC-3'
O-80	5'-GAACCTGTGACCCCATCATTATGAGTGATGTGCTCTAACCAACT-3'
O-81	5'-TGGTGGCTACGACGGGATTCGAACCTGTGACCCCATCA-3'
O-82	5'-TGGTGGCTACGACGGGATTC-3'
O-83	5'-GTAATACGACTCACTATAGCCCGGATAGCTCAGTCGGTAGAGCAGGGGA-3'
O-84	5'-GAACCAAGGACACGGGGATTTTCAGTCCCCTGCTCTACCGACT-3'
O-85	5'-TGGTGCCCGGACTCGGAATCGAACCAAGGACACGGGGA-3'
O-86	5'-TGGTGCCCGGACTCGGAATC-3'
O-87	5'-GGAACGCGCGACTCTAATCGGTGATTGGCGCAGCCTGGTAGCGCACTTC-3'
O-88	5'-GAACCTCCGACCCCTTCGTCCCGAACGAAGTGCGCTACCAGGCT-3'
O-89	5'-TGGTCGGTGATAGAGGATTCGAACCTCCGACCCCTTCG-3'
O-90	5'-TGGTCGGTGATAGAGGATTC-3'
O-91	5'-GGAACGCGCGACTCTAATCGGTGATTGGCGCAGCCTGGTAGCGCACTTC-3'
O-92	5'-GAACCTCCGACCCCTTCGACCCCATCGAAGTGCGCTACCAGGCT-3'
O-93	5'-TGGTCGGTGATAGAGGATTCGAACCTCCGACCCCTTCG-3'
O-94	5'-TGGTCGGTGATAGAGGATTC-3'
O-95	5'-GTAATACGACTCACTATAGGAGAGATGCCGGAGCGGCTGAACGGACCGG-3'
O-96	5'-AGAGTTGCCCCTACTCCGGTTTTCGAGACCGGTCCGTTCAGCCGCT-3'
O-97	5'-TGGCGGAGAGAGGGGGATTTGAACCCCCGGTAGAGTTGCCCCTACTCCG-3'
O-98	5'-TGGCGGAGAGAGGGGGATTT-3'
O-99	5'-GTAATACGACTCACTATAGGAGAGATGCCGGAGCGGCTGAACGGACCGG-3'
O-100	5'-AGAGTTGCCCCTACTCCGGTTTAGCAGACCGGTCCGTTCAGCCGCT-3'
O-101	5'-TGGCGGAGAGAGGGGGATTTGAACCCCCGGTAGAGTTGCCCCTACTCCG-3'
O-102	5'-TGGCGGAGAGAGGGGGATTT-3'
O-103	5'-GTAATACGACTCACTATAGGAGAGATGCCGGAGCGGCTGAACGGACCGG-3'
O-104	5'-AGAGTTGCCCCTACTCCGGTTTTCCAGACCGGTCCGTTCAGCCGCT-3'
O-105	5'-TGGCGGAGAGAGGGGGATTTGAACCCCCGGTAGAGTTGCCCCTACTCCG-3'
O-106	5'-TGGCGGAGAGAGGGGGATTT-3'
O-107	5'-GTAATACGACTCACTATAGCCGATATAGCTCAGTTGGTAGAGCAGCGCA-3'
O-108	5'-GAACCTACGACCTTCGCATTACGAATGCGCTGCTCTACCAACT-3'
O-109	5'-TGGTGCCGATAATAGGAGTCGAACCTACGACCTTCGCA-3'
O-110	5'-TGGTGCCGATAATAGGAGTC-3'
O-111	5'-GTAATACGACTCACTATAGCCGATATAGCTCAGTTGGTAGAGCAGCGCA-3'
O-112	5'-GAACCTACGACCTTCGCACTACCAATGCGCTGCTCTACCAACT-3'
O-113	5'-TGGTGCCGATAATAGGAGTCGAACCTACGACCTTCGCA-3'
O-114	5'-TGGTGCCGATAATAGGAGTC-3'
O-115	5'-GGAACGCGCGACTCTAATAGGGGCGTAGTTCAATTGGTAGAGCACCGGT-3'
O-116	5'-GAACTCCCAACACCCGGTTTTGGAGACCGGTGCTCTACCAATT-3'
O-117	5'-TGGCAGGGGCGGAGAGACTCGAACTCCCAACACCCGGT-3'
O-118	5'-TGGCAGGGGCGGAGAGACTC-3'
O-119	5'-GTAATACGACTCACTATAGGTGGGGTTCCCGAGCGGCCAAAGGGAGCAG-3'
O-120	5'-GAAGTCTGTGACGGCAGATTTACAGTCTGCTCCCTTTGGCCGCT-3'
O-121	5'-TGGTGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGA-3'
O-122	5'-TGGTGGTGGGGGAAGGATTC-3'
O-123	5'-GTAATACGACTCACTATAGGGTGATTAGCTCAGCTGGGAGAGCACCTCC-3'
O-124	5'-GAACCGCCGACCCCCTCCATGTCAAGGAGGTGCTCTCCCAGCT-3'
O-125	5'-TGGTGGGTGATGACGGGATCGAACCGCCGACCCCCTCC-3'
O-126	5'-TGGTGGGTGATGACGGGATC-3'
O-127	5'-GTAATACGACTCACTATAGGGTGATTAGCTCAGCTGGGAGAGCACCTCC-3'
O-128	5'-GAACCGCCGACCCCCTCCGTGTGAGGGAGGTGCTCTCCCAGCT-3'
O-129	5'-TGGTGGGTGATGACGGGATCGAACCGCCGACCCCCTCC-3'
O-130	5'-TGGTGGGTGATGACGGGATC-3'

Supplementary Table 1. (Continued)

Name	Sequence
O-131	5'-GTAATACGACTCACTATAGGCTCTGTAGTTCAGTCGGTAGAACGGCGGA-3'
O-132	5'-GAACCAGTGACATACGGAATGTCAATCCGCCGTTCTACCGACT-3'
O-133	5'-TGGCGGCTCTGACTGGACTCGAACCAGTGACATACGGA-3'
O-134	5'-TGGCGGCTCTGACTGGACTC-3'
O-135	5'-GAACCAGTGACATACGGATTGGCAATCCGCCGTTCTACCGACT-3'
O-136	5'-GAACCAGTGACATACGGATCCGCAGTCCGCCGTTCTACCGACT-3'
O-137	5'-GAACCAGTGACATACGGATCCGGAGTCCGCCGTTCTACCGACT-3'
O-138	5'-GTAATACGACTCACTATAGGATCGAAAGATTTCCGC-3'
O-139	5'-ACCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCC-3'
O-140	5'-ACCTAACGCTAATCCCCTTTCGGGGCCGCGGAAATCTTTCGATCC-3'
O-141	5'-ACCTAACGCCATGTACCCT-3'
O-142	5'-ACCTAACGCTAATCCCCT-3'
O-143	5'-TAATACGACTCACTATAGGGCTTTAATAAGGAGAAAAACATGAAGAAGA-3'
O-144	5'-GTCGTCGTCCTTGTAGTCGAACTTCTTCTTCATGTTTTTCTC-3'
O-145	5'-CGAAGCTTACTTGTCGTCGTCGTCCTTGTAGTC-3'
O-146	5'-GTCGTCGTCCTTGTAGTCGGACTTCTTCTTCATGTTTTTCTC-3'
O-147	5'-GTCGTCGTCCTTGTAGTCGTACTTCTTCTTCATGTTTTTCTC-3'
O-148	5'-GTCGTCGTCCTTGTAGTCGCACTTCTTCTTCATGTTTTTCTC-3'
O-149	5'-GTCGTCGTCCTTGTAGTCGAGCTTCTTCTTCATGTTTTTCTC-3'
O-150	5'-GTCGTCGTCCTTGTAGTCGGGCTTCTTCTTCATGTTTTTCTC-3'
O-151	5'-GTCGTCGTCCTTGTAGTCGTGCTTCTTCTTCATGTTTTTCTC-3'
O-152	5'-GTCGTCGTCCTTGTAGTCGCGCTTCTTCTTCATGTTTTTCTC-3'
O-153	5'-GTCGTCGTCCTTGTAGTCGATCTTCTTCTTCATGTTTTTCTC-3'
O-154	5'-GTCGTCGTCCTTGTAGTCGGTCTTCTTCTTCATGTTTTTCTC-3'
O-155	5'-GTCGTCGTCCTTGTAGTCGTTCTTCTTCTTCATGTTTTTCTC-3'
O-156	5'-GTCGTCGTCCTTGTAGTCGCTCTTCTTCTTCATGTTTTTCTC-3'
O-157	5'-GTCGTCGTCCTTGTAGTCGACCTTCTTCTTCATGTTTTTCTC-3'
O-158	5'-GTCGTCGTCCTTGTAGTCGGCCTTCTTCTTCATGTTTTTCTC-3'
O-159	5'-GTCGTCGTCCTTGTAGTCGTCCTTCTTCTTCATGTTTTTCTC-3'
O-160	5'-GTCGTCGTCCTTGTAGTCGCCCTTCTTCTTCATGTTTTTCTC-3'
O-161	5'-GTCGTCGTCCTTGTAGTCCAACTTCTTCTTCATGTTTTTCTC-3'
O-162	5'-GTCGTCGTCCTTGTAGTCCGACTTCTTCTTCATGTTTTTCTC-3'
O-163	5'-GTCGTCGTCCTTGTAGTCCCACTTCTTCTTCATGTTTTTCTC-3'
O-164	5'-GTCGTCGTCCTTGTAGTCCAGCTTCTTCTTCATGTTTTTCTC-3'
O-165	5'-GTCGTCGTCCTTGTAGTCCGGCTTCTTCTTCATGTTTTTCTC-3'
O-166	5'-GTCGTCGTCCTTGTAGTCCTGCTTCTTCTTCATGTTTTTCTC-3'
O-167	5'-GTCGTCGTCCTTGTAGTCCCGCTTCTTCTTCATGTTTTTCTC-3'
O-168	5'-GTCGTCGTCCTTGTAGTCCATCTTCTTCTTCATGTTTTTCTC-3'
O-169	5'-GTCGTCGTCCTTGTAGTCCGTCTTCTTCTTCATGTTTTTCTC-3'
O-170	5'-GTCGTCGTCCTTGTAGTCCTTCTTCTTCTTCATGTTTTTCTC-3'
O-171	5'-GTCGTCGTCCTTGTAGTCCCTCTTCTTCTTCATGTTTTTCTC-3'
O-172	5'-GTCGTCGTCCTTGTAGTCCACCTTCTTCTTCATGTTTTTCTC-3'
O-173	5'-GTCGTCGTCCTTGTAGTCCGCCTTCTTCTTCATGTTTTTCTC-3'
O-174	5'-GTCGTCGTCCTTGTAGTCCTCCTTCTTCTTCATGTTTTTCTC-3'
O-175	5'-GTCGTCGTCCTTGTAGTCCCCCTTCTTCTTCATGTTTTTCTC-3'
O-176	5'-GTCGTCGTCCTTGTAGTCCTCGTACTTCTTCTTCATGTTTTTCTC-3'
O-177	5'-GTCGTCGTCCTTGTAGTCCCCGTACTTCTTCTTCATGTTTTTCTC-3'
O-178	5'-GTCGTCGTCCTTGTAGTCGCCGTACTTCTTCTTCATGTTTTTCTC-3'
O-179	5'-TAATACGACTCACTATAGGGCTTTAATAAGGAGAAAAACATGTTCTCCTACTG-3'
O-180	5'-TTAGCCGTCGGCGACGCTGTTGGTGATGCGGTGGGGGAGGCAGTAGGAGAACATGTTTTTCTC-3'
O-181	5'-CGAAGCTTAGCCGTCGGCGAC-3'
O-182	5'-TAATACGACTCACTATAGGGCTTTAATAAGGAGAAAAACATGTTGTCGTGG-3'
O-183	5'-TTACCCCTCCGCCACCCTCTTCGTCATCCGCTGCGGCAGCCACGACAACATGTTTTTCTC-3'
O-184	5'-CGAAGCTTACCCCTCCGCCAC-3'
O-185	5'-GAAAAACATGAAGAAGAAGCAGCGGTCGATGGGCCTCCCGCGCTGGACCGCCGA-3'
O-186	5'-TCGTCGTCCTTGTAGTCGCAGAACACCCCGATGTGGACGTTCTCGGCGGTCCAGCGC-3'
O-187	5'-TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGAAGAAGAAG-3'
O-188	5'-CGAAGCTTACTTGTCGTC-3'
O-189	5'-GTCGTCGTCCTTGTAGTCCCGGCGCCTCTTCTTCTTCATGTTTTTCTC-3'
O-190	5'-TAATACGACTCACTATAGGGCTTTAATAAGGAGAAAAACATG-3'
O-191	5'-GTCGTCGTCCTTGTAGTCCACGACCCGGCGCTTCTTCTTCATGTTTTTCTCCTTATTAAAGCCC-3'
O-192	5'-GTCGTCGTCCTTGTAGTCCCGGCGCCCGCCCTTCTTCTTCATGTTTTTCTCCTTATTAAAGCCC-3'
O-193	5'-GTCGTCGTCCTTGTAGTCCACGACCCGGCGCCCGCCCTTCTTCTTCATGTTTTTCTCCTTATTAAAGCCC-3'
O-194	5'-TAATACGACTCACTATAGGGCTTTAATAAGGAGAAAAACATGTGCGACGTGCGCGGGCG-3'
O-195	5'-CGAAGCTTACCCGCACGGGACGTAGCCGACCCGCCCGCGCACGTCG-3'
O-196	5'-CGAAGCTTACCCGCACG-3'

Supplementary Table 2. List of *in vitro* tRNA transcripts prepared in this study.

tRNA	Anticodon	tRNA sequence
Ini	CAU	5'-GGCGGGGUGGAGCAGCCUGGUAGCUCGUCGGGCUCAUAACCCGAAGAUCGUCGGUUCAAAUCCGGCCCCCGCAACCA-3'
Ala	CGC	5'-GGGGCUAUAGCUCAGCUGGGAGAGCGCCUGCUUCGCACGCAGGAGGUCUGCGGUUCGAUCCCGCAUAGCUCCACCA-3'
Ala	GGC	5'-GGGGCUAUAGCUCAGCUGGGAGAGCGCCUGCAUGGCAUGCAGGAGGUCUGCGGUUCGAUCCCGCAUAGCUCCACCA-3'
Arg	CCG	5'-GCGCCCGUAGCUCAGCUGGAUAGAGCGCUGCCCUCCGGAGGCAGAGGUCUCAGGUUCGAAUCCUGUCGGGCGCGCCA-3'
Arg	CCU	5'-GCGCCCGUAGCUCAGCUGGAUAGAGCGCUGCCCUCCUAAGGCAGAGGUCUCAGGUUCGAAUCCUGUCGGGCGCGCCA-3'
Arg	GCG	5'-GCGCCCGUAGCUCAGCUGGAUAGAGCGCUGCCCUGCGAAGGCAGAGGUCUCAGGUUCGAAUCCUGUCGGGCGCGCCA-3'
Asn	GUU	5'-GCCUCUGUAGUUCAGUCGGUAGAACGGCGGACUGUUAAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGGCGCCA-3'
Asp	GUC	5'-GGAGCGGUAGUUCAGUCGGUUAGAAUACCUGCCUGUCACGCAGGGGGUCGCGGGUUCGAGUCCCGUCCGUUCCGCCA-3'
Cys	GCA	5'-GGCGCGUUAACAAAGCGGUUAUGUAGCGGACUGCAAAUCCGUCUAGUCCGGUUCGACUCCGGAACGCGCCUCCA-3'
Gln	CUG	5'-UGGGGUAUCGCCAAGCGGUAAGGCACCGGACUCUGAUUCCGGCAUUCCGAGGUUCGAAUCCUCGUACCCCAGCCA-3'
Glu	CUC	5'-GUCCCCUUCGUCUAGAGGCCCAGGACACCGCCCUCUCACGGCGGUAACAGGGGUUCGAAUCCCCUAGGGGACGCCA-3'
Gly	CCC	5'-GCGGGCGUAGUUCAAUGGUAGAACGAGAGCUUCCCAAGCUCUAUACGAGGGUUCGAUUCCCUUCGCCCGCUCCA-3'
Gly	GCC	5'-GCGGGCGUAGUUCAAUGGUAGAACGAGAGCUUGCCAAGCUCUAUACGAGGGUUCGAUUCCCUUCGCCCGCUCCA-3'
His	GUG	5'-GGGGCUAUAGCUCAGUUGGUAGAGCCCUGGAUUGUGAUUCCAGUUGUCGUGGGUUCGAAUCCCAUUAGCCACCCCA-3'
Ile	GAU	5'-GGGCUUGUAGCUCAGGUGGUUAGAGCGCACCCCUGAUAAGGGUGAGGUCGGUGGUUCAAGUCCACUCAGGCCCACCA-3'
Leu	CAA	5'-GCGAAGGUGGCGGAAUUGGUAGACGCGCUAGCUUCAAGUGUUAGUGUCCUUACGGACGUGGGGGUUCAAGUCCCCCCCCUCGCACCA-3'
Leu	CAG	5'-GCGAAGGUGGCGGAAUUGGUAGACGCGCUAGCUUCAGGUGUUAGUGUCCUUACGGACGUGGGGGUUCAAGUCCCCCCCCUCGCACCA-3'
Leu	GAG	5'-GCGAAGGUGGCGGAAUUGGUAGACGCGCUAGCUUGAGGUGUUAGUGUCCUUACGGACGUGGGGGUUCAAGUCCCCCCCCUCGCACCA-3'
Lys	CUU	5'-GGGUCGUUAGCUCAGUUGGUAGAGCAGUUGACUCUUAAUCAAUUGGUCGCAGGUUCGAAUCCUGCACGACCCACCA-3'
MeU	CAU	5'-GGCUACGUAGCUCAGUUGGUUAGAGCACAUCACUCAUAAUGAUGGGGUCACAGGUUCGAAUCCCGUCGUAGCCACCA-3'
Phe	GAA	5'-GCCCGGAUAGCUCAGUCGGUAGAGCAGGGGACUGAAAAUCCCCGUGUCCUUGGUUCGAUUCCGAGUCCGGGCACCA-3'
Pro	CGG	5'-CGGUGAUUGGCGCAGCCUGGUAGCGCACUUCGUUCGGGACGAAGGGGUCGGAGGUUCGAAUCCUCUAUCACCGACCA-3'
Pro	GGG	5'-CGGUGAUUGGCGCAGCCUGGUAGCGCACUUCGUUGGGGACGAAGGGGUCGGAGGUUCGAAUCCUCUAUCACCGACCA-3'
Ser	CGA	5'-GGAGAGAUGCCGGAGCGGCUGAACGGACCGGUCUCGAAAACCGGAGUAGGGGCAACUCUACCGGGGGUUCAAAUCCCCCUCUCUCCGCCA-3'
Ser	GCU	5'-GGAGAGAUGCCGGAGCGGCUGAACGGACCGGUCUGCUAAACCGGAGUAGGGGCAACUCUACCGGGGGUUCAAAUCCCCCUCUCUCCGCCA-3'
Ser	GGA	5'-GGAGAGAUGCCGGAGCGGCUGAACGGACCGGUCUGGAAAACCGGAGUAGGGGCAACUCUACCGGGGGUUCAAAUCCCCCUCUCUCCGCCA-3'
Uhr	CGU	5'-GCCGAUAUAGCUCAGUUGGUAGAGCAGCGCAUUCGUAAUGCGAAGGUCGUAGGUUCGACUCCUAUUAUCGGCACCA-3'
Uhr	GGU	5'-GCCGAUAUAGCUCAGUUGGUAGAGCAGCGCAUUGGUAAUGCGAAGGUCGUAGGUUCGACUCCUAUUAUCGGCACCA-3'
Urp	CCA	5'-AGGGGCGUAGUUCAAUUGGUAGAGCACCGGUCUCCAAAACCGGGUGUUGGGAGUUCGAGUCUCUCCGCCCCUGCCA-3'
Uyr	GUA	5'-GGUGGGGUUCCCGAGCGGCCAAAGGGAGCAGACUGUAAAUCUGCCGUCACAGACUUCGAAGGUUCGAAUCCUUCCCCCACCACCA-3'
Val	GAC	5'-GGGUGAUUAGCUCAGCUGGGAGAGCACCUCCCUGACACGGAGGGGGUCGGCGGUUCGAUCCCGUCAUCACCCACCA-3'
Val	CAC	5'-GGGUGAUUAGCUCAGCUGGGAGAGCACCUCCCUCACACGGAGGGGGUCGGCGGUUCGAUCCCGUCAUCACCCACCA-3'
AsnE2	GAC	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUUGACAUUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	GCC	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUUGCCAAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	GCG	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACUGCGGAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	CCG	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACUCCGGAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'

Supplementary Table 3. List of oligonucleotides used for the preparation of DNA templates coding tRNAs.

tRNA	Anticodon	Extension		1st PCR		2nd PCR	
		Forward primer	Reverse primer	Forward primer	Reverse primer	Forward primer	Reverse primer
Ini	CAU	O-3	O-4	O-1	O-5	O-1	O-6
Ala	CGC	O-7	O-8	O-1	O-9	O-1	O-10
Ala	GGC	O-11	O-12	O-1	O-13	O-1	O-14
Arg	CCG	O-15	O-16	O-1	O-17	O-1	O-18
Arg	CCU	O-19	O-20	O-1	O-21	O-1	O-22
Arg	GCG	O-23	O-24	O-1	O-25	O-1	O-26
Asn	GUU	O-27	O-28	O-1	O-29	O-1	O-30
Asp	GUC	O-31	O-32	O-1	O-33	O-1	O-34
Cys	GCA	O-35	O-36	O-1	O-37	O-1	O-38
Gln	CUG	O-39	O-40	O-2	O-41	O-2	O-42
Glu	CUC	O-43	O-44	O-1	O-45	O-1	O-46
Gly	CCC	O-47	O-48	O-1	O-49	O-1	O-50
Gly	GCC	O-51	O-52	O-1	O-53	O-1	O-54
His	GUG	O-55	O-56	O-1	O-57	O-1	O-58
Ile	GAU	O-59	O-60	O-1	O-61	O-1	O-62
Leu	CAA	O-63	O-64	O-1	O-65	O-1	O-66
Leu	CAG	O-67	O-68	O-1	O-69	O-1	O-70
Leu	GAG	O-71	O-72	O-1	O-73	O-1	O-74
Lys	CUU	O-75	O-76	O-1	O-77	O-1	O-78
Met	CAU	O-79	O-80	O-1	O-81	O-1	O-82
Phe	GAA	O-83	O-84	O-1	O-85	O-1	O-86
Pro	CGG	O-87	O-88	O-2	O-89	O-2	O-90
Pro	GGG	O-91	O-92	O-2	O-93	O-2	O-94
Ser	CGA	O-95	O-96	O-1	O-97	O-1	O-98
Ser	GCU	O-99	O-100	O-1	O-101	O-1	O-102
Ser	GGA	O-103	O-104	O-1	O-105	O-1	O-106
Thr	CGU	O-107	O-108	O-1	O-109	O-1	O-110
Thr	GGU	O-111	O-112	O-1	O-113	O-1	O-114
Trp	CCA	O-115	O-116	O-2	O-117	O-2	O-118
Tyr	GUA	O-119	O-120	O-1	O-121	O-1	O-122
Val	GAC	O-123	O-124	O-1	O-125	O-1	O-126
Val	CAC	O-127	O-128	O-1	O-129	O-1	O-130
AsnE2	GAC	O-131	O-132	O-1	O-133	O-1	O-134
AsnE2	GCC	O-131	O-135	O-1	O-133	O-1	O-134
AsnE2	GCG	O-131	O-136	O-1	O-133	O-1	O-134
AsnE2	CCG	O-131	O-137	O-1	O-133	O-1	O-134

Supplementary Table 4. List of mRNA sequences used in this study.

Name	Corresponding mDNA	mRNA sequence
$mRNA1_{UUC}$	$mDNA1_{UUC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUUCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{UCC}$	$mDNA1_{UCC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUCCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{UAC}$	$mDNA1_{UAC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUACGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{UGC}$	$mDNA1_{UGC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUGCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{CUC}$	$mDNA1_{CUC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCUCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{CCC}$	$mDNA1_{CCC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCCCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{CAC}$	$mDNA1_{CAC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCACGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{CGC}$	$mDNA1_{CGC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCGCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{AUC}$	$mDNA1_{AUC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGAUCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{ACC}$	$mDNA1_{ACC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGACCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{AAC}$	$mDNA1_{AAC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGAACGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{AGC}$	$mDNA1_{AGC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGAGCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GUC}$	$mDNA1_{GUC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGUCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GCC}$	$mDNA1_{GCC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGCCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GAC}$	$mDNA1_{GAC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGACGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GGC}$	$mDNA1_{GGC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGGCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{UUG}$	$mDNA1_{UUG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUUGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{UCG}$	$mDNA1_{UCG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUCGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{UGG}$	$mDNA1_{UGG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUGGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{CUG}$	$mDNA1_{CUG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCUGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{CCG}$	$mDNA1_{CCG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCCGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mDNA1_{CAG}$	$mDNA1_{CAG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCAGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{CGG}$	$mDNA1_{CGG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCGGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{AUG}$	$mDNA1_{AUG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGAUGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{ACG}$	$mDNA1_{ACG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGACGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{AAG}$	$mDNA1_{AAG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGAAGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{AGG}$	$mDNA1_{AGG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGAGGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GUG}$	$mDNA1_{GUG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGUGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GCG}$	$mDNA1_{GCG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGCGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GAG}$	$mDNA1_{GAG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGAGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA1_{GGG}$	$mDNA1_{GGG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGGGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA2_{GAG}$	$mDNA2_{GAG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUACGAGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA2_{GGG}$	$mDNA2_{GGG}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUACGGGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
$mRNA2_{GGC}$	$mDNA2_{GGC}$	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGUACGGCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA3	mDNA3	5'-GGGCUUUAAUAAGGAGAAAAACAUGUUCUCCUACUGCCUCCCCCACCGCAUCACCAACAGCGUCGCCGACGGCUAAGCUUCG-3'
mRNA4	mDNA4	5'-GGGCUUUAAUAAGGAGAAAAACAUGUUGUCGUGGCUGCCGCAGCGGAUGACGAAGAGGGUGGCGGAGGGGUAAGCUUCG-3'
mRNA5	mDNA5	5'-GGGUUAACUUUAACAAGGAGAAAAACAUGAAGAAGAAGCAGCGGUCGAUGGGCCUCCCGCGCUGG ACCGCCGAGAACGUCCACAUCGGGGUGUUCUGCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA6	mDNA6	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGAGGCGCCGGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA7	mDNA7	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCGCCGGGUCGUGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA8	mDNA8	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGGCGGGCGCCGGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA9	mDNA9	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGGCGGGCGCCGGGUCGUGGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA10	mDNA10	5'-GGGUUAACUUUAACAAGGAGAAAAACAUGAAGAAGAAGCAGCGGUCGAUGGGCCUCCCGCGCUGG ACCGCCGAGAACGUCCACAUCGGGGUGUUCUGCGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA11	mDNA11	5'-GGGCUUUAAUAAGGAGAAAAACAUGUGCGACGUGCGCGGGCGGGUCGGCUACGUCCCGUGCGGGUAAGCU-3'

Supplementary Table 5. List of oligonucleotides used for the preparation of mDNAs.

Name	Corresponding mRNA	Extension		1st PCR		2nd PCR	
		Forward primer	Reverse primer	Forward primer	Reverse primer	Forward primer	Reverse primer
$mDNA1_{UUC}$	$mRNA1_{UUC}$	O-143	O-144	O-1	O-145	-	-
$mDNA1_{UCC}$	$mRNA1_{UCC}$	O-143	O-146	O-1	O-145	-	-
$mDNA1_{UAC}$	$mRNA1_{UAC}$	O-143	O-147	O-1	O-145	-	-
$mDNA1_{UGC}$	$mRNA1_{UGC}$	O-143	O-148	O-1	O-145	-	-
$mDNA1_{CUC}$	$mRNA1_{CUC}$	O-143	O-149	O-1	O-145	-	-
$mDNA1_{CCC}$	$mRNA1_{CCC}$	O-143	O-150	O-1	O-145	-	-
$mDNA1_{CAC}$	$mRNA1_{CAC}$	O-143	O-151	O-1	O-145	-	-
$mDNA1_{CGC}$	$mRNA1_{CGC}$	O-143	O-152	O-1	O-145	-	-
$mDNA1_{AUC}$	$mRNA1_{AUC}$	O-143	O-153	O-1	O-145	-	-
$mDNA1_{ACC}$	$mRNA1_{ACC}$	O-143	O-154	O-1	O-145	-	-
$mDNA1_{AAC}$	$mRNA1_{AAC}$	O-143	O-155	O-1	O-145	-	-
$mDNA1_{AGC}$	$mRNA1_{AGC}$	O-143	O-156	O-1	O-145	-	-
$mDNA1_{GUC}$	$mRNA1_{GUC}$	O-143	O-157	O-1	O-145	-	-
$mDNA1_{GCC}$	$mRNA1_{GCC}$	O-143	O-158	O-1	O-145	-	-
$mDNA1_{GAC}$	$mRNA1_{GAC}$	O-143	O-159	O-1	O-145	-	-
$mDNA1_{GGC}$	$mRNA1_{GGC}$	O-143	O-160	O-1	O-145	-	-
$mDNA1_{UUG}$	$mRNA1_{UUG}$	O-143	O-161	O-1	O-145	-	-
$mDNA1_{UCG}$	$mRNA1_{UCG}$	O-143	O-162	O-1	O-145	-	-
$mDNA1_{UGG}$	$mRNA1_{UGG}$	O-143	O-163	O-1	O-145	-	-
$mDNA1_{CUG}$	$mRNA1_{CUG}$	O-143	O-164	O-1	O-145	-	-
$mDNA1_{CCG}$	$mRNA1_{CCG}$	O-143	O-165	O-1	O-145	-	-
$mDNA1_{CAG}$	$mRNA1_{CAG}$	O-143	O-166	O-1	O-145	-	-
$mDNA1_{CGG}$	$mRNA1_{CGG}$	O-143	O-167	O-1	O-145	-	-
$mDNA1_{AUG}$	$mRNA1_{AUG}$	O-143	O-168	O-1	O-145	-	-
$mDNA1_{ACG}$	$mRNA1_{ACG}$	O-143	O-169	O-1	O-145	-	-
$mDNA1_{AAG}$	$mRNA1_{AAG}$	O-143	O-170	O-1	O-145	-	-
$mDNA1_{AGG}$	$mRNA1_{AGG}$	O-143	O-171	O-1	O-145	-	-
$mDNA1_{GUG}$	$mRNA1_{GUG}$	O-143	O-172	O-1	O-145	-	-
$mDNA1_{GCG}$	$mRNA1_{GCG}$	O-143	O-173	O-1	O-145	-	-
$mDNA1_{GAG}$	$mRNA1_{GAG}$	O-143	O-174	O-1	O-145	-	-
$mDNA1_{GGG}$	$mRNA1_{GGG}$	O-143	O-175	O-1	O-145	-	-
$mDNA2_{GAG}$	$mRNA2_{GAG}$	O-143	O-176	O-1	O-145	-	-
$mDNA2_{GGG}$	$mRNA2_{GGG}$	O-143	O-177	O-1	O-145	-	-
$mDNA2_{GGC}$	$mRNA2_{GGC}$	O-143	O-178	O-1	O-145	-	-
mDNA3	mRNA3	O-179	O-180	O-1	O-181	-	-
mDNA4	mRNA4	O-182	O-183	O-1	O-184	-	-
mDNA5	mRNA5	O-185	O-186	O-1	O-187	O-1	O-188
mDNA6	mRNA6	O-143	O-189	O-1	O-145	-	-
mDNA7	mRNA7	O-190	O-191	O-1	O-145	-	-
mDNA8	mRNA8	O-190	O-192	O-1	O-145	-	-
mDNA9	mRNA9	O-190	O-193	O-191	O-145	-	-
mDNA10	mRNA10	O-185	O-186	O-1	O-187	O-1	O-188
mDNA11	mRNA11	O-194	O-195	O-1	O-196	-	-

Supplementary Table 6. Conditions for translation reactions conducted in this study.

Figure	Each pre-charged $tRNA^{AsnE2}$ (μM)	$Mg(OAc)_2$ (mM)	Reaction time (min)	Anti-flag purification for MALDI-TOF-MS
26	-	13	30	No MS
27	-	13	30	-
28	-	13	10	+
23 - 36	100	13	10	+
39, 40	-	9	30	-
41	50	9	30	No MS
42	50	9	10	-
44	50 or 100	9 - 15	10	-
45c	50	9	0 - 30	No MS
45d	100	13	0 - 30	-

Supplementary Table 7. Concentrations of *in vitro* tRNA transcripts used for the experiments. Asterisk denotes that the tRNA transcript was added only when the $mRNA1_{NNS*}$ and $mRNA2_{NNS*}$ contained the codon corresponding to each tRNA.

tRNA	Anticodon	Concentration of tRNA transcript (μM)							
		Fig. 26	Fig. 27, 28	Fig. 29	Fig. 30a	Fig. 30b	Fig. 31	Fig. 32	Fig. 33
Lys	CUU	30	30	30	30	30	30	30	30
Ile	GAU	30*	30	0	0	0	30	30	30
Glu	CUC	30*	30	0	0	0	30	30	30
Ini	CAU	5	5	5	5	5	5	5	5
Asp	GUC	5	5	5	5	5	5	5	5
Tyr	GUA	5	5	5	5	5	5	5	5
Ala	CGC	5*	1	0	0	0	1	1	1
Ala	GGC	5*	1	0	0	0	1	1	1
Arg	CCG	5*	1	0	0 or 5	0	0 or 1	1	1
Arg	GCG	5*	1	0	0	0	0 or 1	0 or 1	0 or 1
Arg	CCU	5*	1	0	0	0	1	1	1
Asn	GUU	5*	1	0	0	0	1	1	1
Cys	GCA	5*	1	0	0	0	1	1	1
Gln	CUG	5*	1	0	0	0	1	1	1
Gly	CCC	5*	1	0	0	0 or 5	1	1	1
Gly	GCC	5*	1	0	0	0	1	1	0 or 1
His	GUG	5*	1	0	0	0	1	1	1
Leu	CAA	5*	1	0	0	0	1	1	1
Leu	CAG	5*	1	0	0	0	1	1	1
Leu	GAG	5*	1	0	0	0	1	1	1
Met	CAU	5*	1	0	0	0	1	1	1
Phe	GAA	5*	1	0	0	0	1	1	1
Pro	CGG	5*	1	0	0	0	1	1	1
Pro	GGG	5*	1	0	0	0	1	1	1
Ser	CGA	5*	1	0	0	0	1	1	1
Ser	GCU	5*	1	0	0	0	1	1	1
Ser	GGA	5*	1	0	0	0	1	1	1
Thr	CGU	5*	1	0	0	0	1	1	1
Thr	GGU	5*	1	0	0	0	1	1	1
Trp	CCA	5*	1	0	0	0	1	1	1
Val	CAC	5*	1	0 or 5	0	0	1	1	1
Val	GAC	5*	1	0	0	0	1	0 or 1	1

tRNA	Anticodon	Concentration of tRNA transcript (μM)							
		Fig. 34, 35	Fig. 36	Fig. 39	Fig. 40a	Fig. 40b	Fig. 41, 42	Fig. 44	Fig. 45
Lys	CUU	30	30	5-50	30	30	30	30	30
Ile	GAU	30	30	0	5-50	0	0	30	0
Glu	CUC	30	30	0	0	5-50	0	30	0
Ini	CAU	5	0	5	5	5	5	5	5
Asp	GUC	5	5	5	5	5	5	5	5
Tyr	GUA	5	5	5	5	5	5	5	5
Ala	CGC	1	1	0	0	0	0	1	0
Ala	GGC	1	1	0	0	0	0	1	0
Arg	CCG	1	1	0	0	0	0 or 5	5	0
Arg	GCG	0 or 1	0	0	0	0	0	0	0
Arg	CCU	1	1	0	0	0	0	1	0
Asn	GUU	1	1	0	0	0	0	1	0
Cys	GCA	1	1	0	0	0	0	1	0
Gln	CUG	1	1	0	0	0	0	1	0
Gly	CCC	1	1	0	0	0	0	1	0
Gly	GCC	0 or 1	0	0	0	0	0	1	0
His	GUG	1	1	0	0	0	0	1	0
Leu	CAA	1	1	0	0	0	0	1	0
Leu	CAG	1	1	0	0	0	0	1	0
Leu	GAG	1	1	0	0	0	0	1	0
Met	CAU	1	1	0	0	0	0	1	0
Phe	GAA	1	1	0	0	0	0	1	0
Pro	CGG	1	1	0	0	0	0	1	0
Pro	GGG	1	1	0	0	0	0	1	0
Ser	CGA	1	1	0	0	0	0	1	0
Ser	GCU	1	1	0	0	0	0	1	0
Ser	GGA	1	1	0	0	0	0	1	0
Thr	CGU	1	1	0	0	0	0	1	0
Thr	GGU	1	1	0	0	0	0	1	0
Trp	CCA	1	1	0	0	0	0	1	0
Val	CAC	1	1	0	0	0	0	5	5
Val	GAC	0 or 1	0	0	0	0	0	0	0

Supplementary Table 8. Numerical data of quantification experiments.

Fig. 28	Lane	Amount (μM)		S.D.	N
	1	1.65	±	0.46	4
	2	0.48	±	0.05	4

Fig. 29	Lane	Peptide-Val				Peptide-^{Me}Y			
		Amount (μM)		S.D.	N	Amount (μM)		S.D.	N
	1	0.40	±	0.10	4				
	2	0.04	±	0.02	4				
	3	0.48	±	0.12	4				
	4	0.85	±	0.28	4	0.02	±	0.01	4
	5	1.02	±	0.48	4				
	6	0.08	±	0.03	4				
	7	0.03	±	0.02	4	1.03	±	0.20	6
	8	0.06	±	0.03	4	0.97	±	0.23	6

Fig. 32	Lane	Amount (μM)		S.D.	N
	1	1.19	±	0.18	4
	2	2.36	±	0.68	4
	3	1.23	±	0.38	4
	4	1.18	±	0.37	4
	5	0.71	±	0.40	4

Fig. 34	Lane	Amount (μM)		S.D.	N
	1	1.49	±	0.27	4
	2	2.18	±	0.40	4
	3	0.28	±	0.00	4

Fig. 35	Lane	Amount (μM)		S.D.	N
	1	1.65	±	0.46	4
	2	0.48	±	0.05	4
	3	0.24	±	0.12	4

Fig. 39e	Lane	Amount (μM)		S.D.	N
	1	4.55	±	0.07	3
	2	0.54	±	0.06	3
	3	2.48	±	0.12	3
	4	2.76	±	0.16	3

Fig. 41	Lane	Peptide-Arg		Peptide-^{Me}S	
		Amount (μM)	N	Amount (μM)	N
	1	5.42	1		
	2	0.11	1		
	3	4.37	1		
	4	4.56	1		
	5	5.58	1		
	6	0.51	1		
	7	0.67	1	1.24	1
	8	1.05	1	1.04	1

Fig. 42	Lane	Peptide-Arg		Peptide-^{Me}S	
		Amount (μM)	N	Amount (μM)	N
	1	0.79	1		
	2	0.01	1		
	3	0.62	1		
	4	0.71	1		
	5	0.72	1		
	6	0.06	1		
	7	0.07	1	0.42	1
	8	0.08	1	0.37	1

Fig. 44	Lane	Peptide-Arg		Peptide-^{Me}S	
		Amount (μM)	N	Amount (μM)	N
	1	0.46	1	0.82	1
	2	0.36	1	0.92	1
	3	0.17	1	0.80	1
	4	0.03	1	0.51	1
	5	0.18	1	1.41	1
	6	0.10	1	1.23	1
	7	0.04	1	0.86	1
	8	0.02	1	0.38	1

Fig. 45c	Time (min)	Peptide-Val		Peptide-^{Me}Y	
		Amount (μM)	N	Amount (μM)	N
	0	0.00	1	0.00	1
	2	0.01	1	0.09	1
	4	0.03	1	0.52	1
	6	0.08	1	1.50	1
	8	0.41	1	2.22	1
	10	1.00	1	2.32	1
	12	1.39	1	2.27	1
	14	1.71	1	2.38	1
	16	1.71	1	2.35	1
	18	1.76	1	2.40	1
	20	1.67	1	2.32	1
	25	1.75	1	2.28	1
	30	1.65	1	2.14	1

Fig. 45d	Time (min)	Peptide-Val		Peptide-^{Me}Y	
		Amount (μM)	N	Amount (μM)	N
	0	0.00	1	0.00	1
	2	0.01	1	0.04	1
	4	0.02	1	0.19	1
	6	0.04	1	0.47	1
	8	0.07	1	0.94	1
	10	0.10	1	1.50	1
	12	0.15	1	2.11	1
	14	0.20	1	2.54	1
	16	0.24	1	2.73	1
	18	0.29	1	3.11	1
	20	0.30	1	3.18	1
	25	0.40	1	3.37	1
	30	0.45	1	2.99	1

Chapter 3

Ribosomal synthesis of highly *N*-methylated peptides by tuning affinities between EF-Tu and *N*-methyl-aminoacyl-tRNAs

3.1. Introduction

In drug development studies, small molecules and proteins are dominantly used as the molecular bases[106]. Recently, protein-based drugs, especially antibodies, have become one of the most successful classes of therapeutics owing to their high affinity and selectivity to target proteins with large surface interaction (*i.e.* surface recognition property), which contributed to less side-effects and higher success rate in clinical trial compared to small-molecule drugs[107,110]. In spite of their effectiveness, the use of protein-based drugs has a major drawback that proteins cannot penetrate through cell membranes, and therefore they are not orally available (*i.e.* injection required) and they cannot target intracellular proteins that comprise most part of human proteins. In order to discover an ideal drug that possess both surface recognition property and cell-membrane permeability, small peptide molecules attract great attention as a new drug scaffold[67,112]. In fact, such drugs have been found in naturally occurring bioactive peptides[113]. For example, an immunosuppressant cyclosporine, a macrocyclic *N*-methylated peptide, exhibits membrane permeability and acceptable stability in human body, and therefore it can effectively interact with intracellular target proteins named cyclophilins[145-148]. Notably, bioactive peptides often possess multiple backbone *N*-methyl modifications, and it has been revealed that such *N*-methylations confer increased membrane permeability and protease resistance on peptide molecules[114-124].

To discover novel peptide-based drugs, researchers have developed several methods to discover peptides that bind to target proteins from a sequence-randomized peptide library, which include phage-based expression method and mRNA display method[125,126]. Our laboratory has combined the mRNA display method with an engineered *in vitro* translation system named FIT (flexible *in vitro* translation) system[77,92,104,127-129], and successfully developed a novel method, named RaPID (Random nonstandard peptides integrated discovery) system, to discover bioactive peptides with nonstandard scaffold such as *N*-methyl modification and macrocyclic structure[93,130-136]. The system relies on a reconstituted *E. coli* cell-free translation system[32] supplemented with synthetic tRNAs pre-charged with nonproteinogenic amino acids (npAAs) by catalysis of flexizymes[92]. Although they have discovered potent binders with a few *N*-methyl modifications, the desired peptides that exhibit membrane permeability and stability against enzymatic degradation could not be found in most cases.

To further improve such pharmacokinetic properties, it is ideal to develop a nonstandard peptide library that comprises a rich variety of distinct *N*-methyl-amino acids (^{Me}AAs). However, there is a methodological obstacle that ^{Me}AAs are suboptimal substrates for the translational machinery, and therefore premature termination of translation can occur or incorrect amino acids can be misincorporated in place of the desired ^{Me}AA, resulting in reduced incorporation efficiency and accuracy of *N*-methylated peptide synthesis[69,104,149]. A previous study reported that, among 19 ^{Me}AAs, only five ^{Me}AAs exhibited efficient incorporation efficiencies (>80% to a proteinogenic amino acid (pAA) control); six ^{Me}AAs were moderate (10–80%); the other eight ^{Me}AAs were poor and undetectable (<10%)[104]. Although several contrivances to improve the incorporation efficiency of a single ^{Me}AA into a peptide have been reported[69,150,151], it is still challenging to incorporate multiple distinct ^{Me}AAs into the same peptide. The best case in literature so far is the synthesis of a peptide containing only three kinds of ^{Me}AAs[69,93]. In this study, I aimed to develop a novel method that allows for the simultaneous incorporation of multiple distinct ^{Me}AAs, which

would enable us to discover highly *N*-methylated peptide drugs with improved membrane permeability and stability against enzymatic decomposition.

To achieve this goal, I first hypothesized the cause of the inefficient ^{Me}AA incorporation as follows. Ribosome recruits aminoacyl-tRNAs by means of EF-Tu and accommodates them into the A-site space in ribosome (Fig. 46a)[40]. EF-Tu binds to an aminoacyl-tRNA at two distinct sites: (1) amino acid site and (2) T-stem site (Fig. 46b)[8,41,42]. When only the former interaction at the amino acid site is considered, the binding affinities are drastically different depending on the amino acid species (Fig. 46c)[43]. In order to compensate the different affinities, nature has elegantly evolved the T-stem structures so that the affinities between EF-Tu and each of 20 aminoacyl-tRNAs become equivalent[44]. Due to this equivalent affinities, EF-Tu can deliver all of 20 proteinogenic aminoacyl-tRNAs (pAA-tRNAs) into ribosome with similar efficiencies[45] (Fig. 46d). In contrast, if the affinity between EF-Tu and an *N*-methylaminoacyl-tRNA (^{Me}AA-tRNA) were weak, the ^{Me}AA-tRNA could not bind to EF-Tu under the competition with pAA-tRNAs, which would lead to decrease in the synthetic efficiency of the ^{Me}AA-containing peptide. In other words, if the EF-Tu affinity of ^{Me}AA-tRNA could be tuned to the mean EF-Tu affinity of pAA-tRNAs, then the efficiency could be improved and a variety of ^{Me}AAs could be utilized simultaneously.

To prove this concept, in this study, I first experimentally confirmed that the affinities of most ^{Me}AA-tRNAs were less than the mean value of pAA-tRNAs, which supported the idea that ^{Me}AA-tRNAs could not bind to EF-Tu with sufficient strengths (Fig. 46e). Next, I developed a method to tune the weak EF-Tu affinity of ^{Me}AA-tRNAs to the value of pAA-tRNAs (Fig. 46e). The FIT system containing EF-Tu affinity-tuned ^{Me}AA-tRNAs exhibited improved synthetic efficiency and accuracy of *N*-methylated peptides. The proof of concept was demonstrated by the expression of a 32-mer model peptide composed of 9 different ^{Me}AAs as well as 14 kinds of pAAs. These results supported the idea that this new method enables us to synthesize a highly *N*-methylated peptide library, which is expected to facilitate the discovery of practical nonstandard peptide drugs with improved pharmacokinetic properties in future.

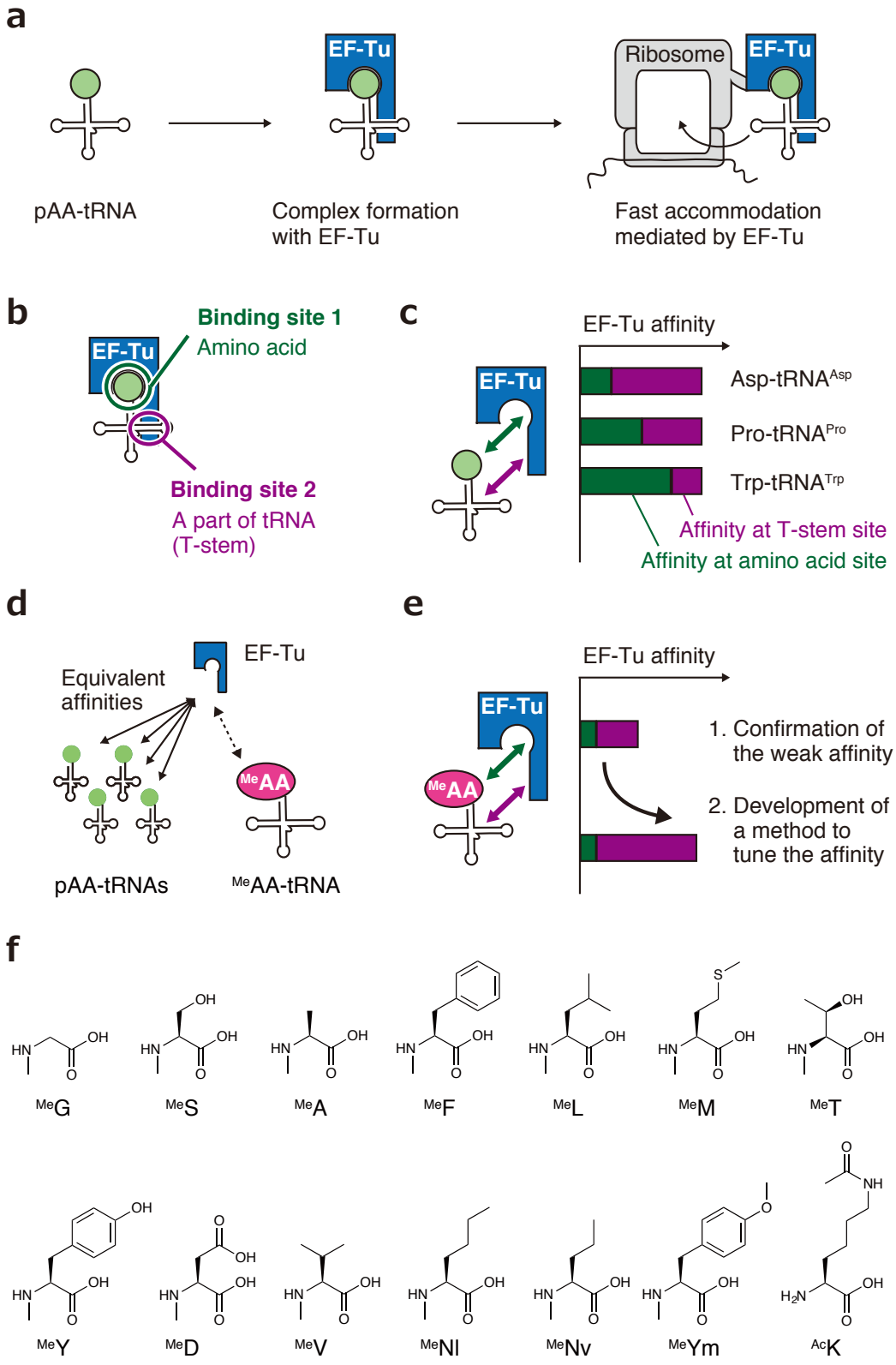

Figure 46. Schematic illustration of the interaction between EF-Tu and aminoacyl-tRNAs. a, EF-Tu-mediated delivery of pAA-tRNA into ribosome. **b**, Two binding sites between EF-Tu and pAA-tRNA. **c**, Compensatory relationship between affinities at the two binding sites, achieving equivalent EF-Tu affinities for 20 pAA-tRNAs. **d**, Hypothesized obstacle for ^{Me}AA-tRNA: the insufficient EF-Tu affinity under the competition with pAA-tRNAs. **e**, Schematic representation of this study. **f**, npAAs used in this study. ^{Me}G: *N*-methylglycine, ^{Me}S: *N*-methylserine, ^{Me}A: *N*-methylalanine, ^{Me}F: *N*-methylphenylalanine, ^{Me}L: *N*-methylleucine, ^{Me}M: *N*-methylmethionine, ^{Me}T: *N*-methylthreonine, ^{Me}Y: *N*-methyltyrosine, ^{Me}D: *N*-methylaspartic acid, ^{Me}V: *N*-methylvaline, ^{Me}Nl: *N*-methylnorleucine, ^{Me}Nv: *N*-methylnorvaline, ^{Me}Ym: *N*-methyl-p-methoxyphenylalanine, ^{Ac}K: ε-*N*-acetyllysine.

3.2. Results and Discussion

3.2.1. Experimental proof that many ^{Me}AA-tRNAs cannot bind to EF-Tu with sufficient strengths

First, I compared the EF-Tu affinities of canonical pAA-tRNAs and ^{Me}AA-tRNAs. As the control, three canonical pAA-tRNAs (Phe-tRNA^{Phe}, Tyr-tRNA^{Tyr}, and Ser-tRNA^{Ser}) were prepared by flexizyme-catalyzed aminoacylation, and their EF-Tu affinities were quantified by the method developed in this study (see Supplementary results 3.4.1 for the method development). As previously reported, they exhibited similar affinities ranging from -8 to -9.4 kcal/mol (Fig. 47a). It should be noted that the observed affinity values were slightly less than those previously reported (from -9.5 to 10.5 kcal/mol)[45]. The cause of difference might be attributed to the difference in buffer composition or the presence of His6 tag at the C-terminus of EF-Tu used in this study. In contrast to canonical pAA-tRNAs, many ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s exhibited undetectable level of weak EF-Tu affinities with the exception of ^{Me}G-tRNA$^{AsnE2}_{GAC}$, ^{Me}S-tRNA$^{AsnE2}_{GA}$, and ^{Me}A-tRNA$^{AsnE2}_{GAC}$ (Fig. 47b). Considering that the α-amino group of the charged amino acid is accommodated in a pocket composed of EF-Tu residues (Fig. 47c), the *N*-methyl modification is supposed to cause steric hindrance between ^{Me}AA and the pocket, which would decrease the EF-Tu affinities. The three exceptions of ^{Me}G-tRNA$^{AsnE2}_{GAC}$, ^{Me}S-tRNA$^{AsnE2}_{GA}$, and ^{Me}A-tRNA$^{AsnE2}_{GAC}$ might alleviated the steric hindrance by positioning the side-chain (-H, -CH_3, and -OH groups, respectively) into the pocket for the amino group. Next, six kinds of ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s were used for the ribosomal synthesis of a model peptide containing a single ^{Me}AA (Fig. 47d). The mRNA template was designed to determine the translation accuracy by MALDI-TOF-MS (Matrix-Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry) of translation products (Fig. 47d). The genetic code of GUC codon was reprogrammed by excluding the corresponding Val from the *in vitro* translation system and instead supplementing one of the ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s. When the ^{Me}G-tRNA$^{AsnE2}_{GAC}$ and ^{Me}S-tRNA$^{AsnE2}_{GA}$, which exhibited strong EF-Tu affinities, were used, the correct peptide was observed as a single major peak in MALDI-TOF-MS. On the other hand, when the ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s, which exhibited weak EF-Tu affinities, were used, an undesired side-product containing either Leu or Ile in place of the ^{Me}AA was detected in MALDI-TOF-MS along with the correct peptide product (Fig. 47d). As an exception, a single major peak was observed for ^{Me}F although ^{Me}F-tRNA$^{AsnE2}_{GAC}$ did not exhibit detectable level of EF-Tu affinity. A supplementary experiment revealed that the misincorporated amino acid was Ile, not Leu, as the side-product was observed only when Ile was present in translation mixture (Fig. 47e). Similar weak EF-Tu affinity and translation inaccuracy were observed for another type of npAA, ε-*N*-acetyllysine (^{Ac}K), which was examined to confirm the generality of this study (Fig. 47b,d). Collectively, above experiments demonstrated that the EF-Tu affinities of many ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s were much weaker than the mean value of canonical pAA-tRNAs and the weak EF-Tu affinity often causes the translational inaccuracy during ribosomal polypeptide synthesis.

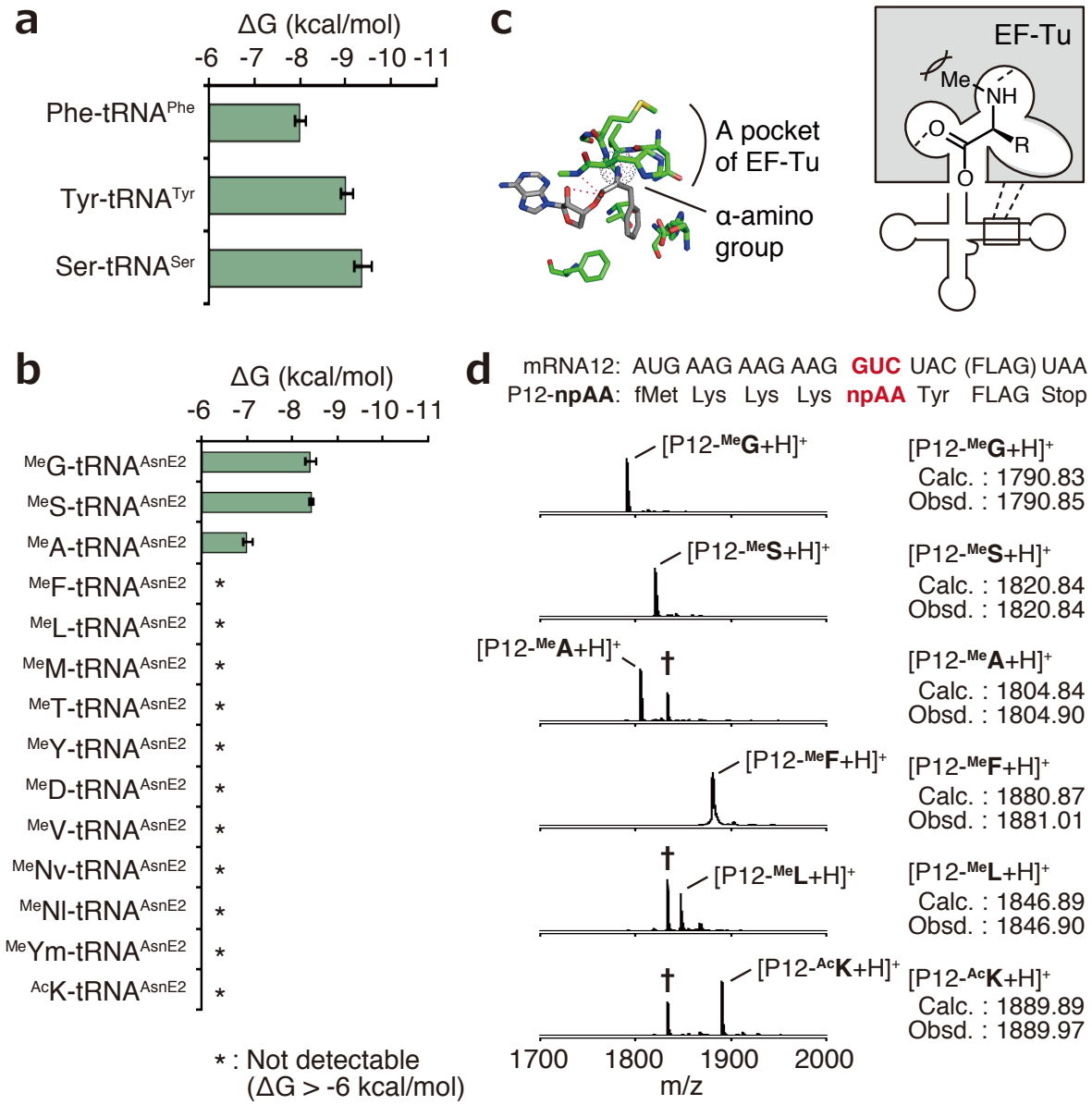

Figure 47. Weak EF-Tu affinities of ^{Me}AA-tRNAs and correlate inaccurate synthesis of ^{Me}AA-containing peptides. a, EF-Tu affinities of canonical pAA-tRNAs. Error bar: standard error. **b**, Weak EF-Tu affinities of many ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s along with ^{Ac}K-tRNA$^{AsnE2}_{GAC}$. **c**, Supposed cause of the weakened EF-Tu affinities of ^{Me}AA-tRNAs. This image was made from the cocrystal structure of EF-Tu and Phe-tRNA^{Phe} (PDB ID: 1TTT)[8] **d**, Examination of the translation accuracy of peptides containing each of ^{Me}AAs or ^{Ac}K. In each experiment, the mRNA12 containing GUC codon was translated in the FIT system containing npAA-tRNA$^{AsnE2}_{GAC}$ of interest. †Peak corresponds to the side-product containing Ile instead of the npAA. **e**, Confirmation of the side-product observed in **d**.

3.2.2. Development of a method to reinforce the weak EF-Tu affinities of ^{Me}AA-tRNAs.

Discovery of the fact that many ^{Me}AA-tRNAs cannot bind to EF-Tu with sufficient strengths encouraged me to develop a novel method to reinforce the weak EF-affinity of ^{Me}AA-tRNAs. As it had been reported that the binding affinity between EF-Tu and Phe-tRNA^{Phe} changes depending on its T-stem sequence[42], I expected that the EF-Tu affinity of an ^{Me}AA-tRNA could also be changed by using different T-stem sequences. Based on this concept, three different T-stem sequences were selected from literature (Fig. 48)[42]. According to their results, the EF-Tu affinities of these four ^{Me}AA-tRNA are supposed to increase as the T-stem number increases. Note that the original T-stem was renamed as T-stem No.2.

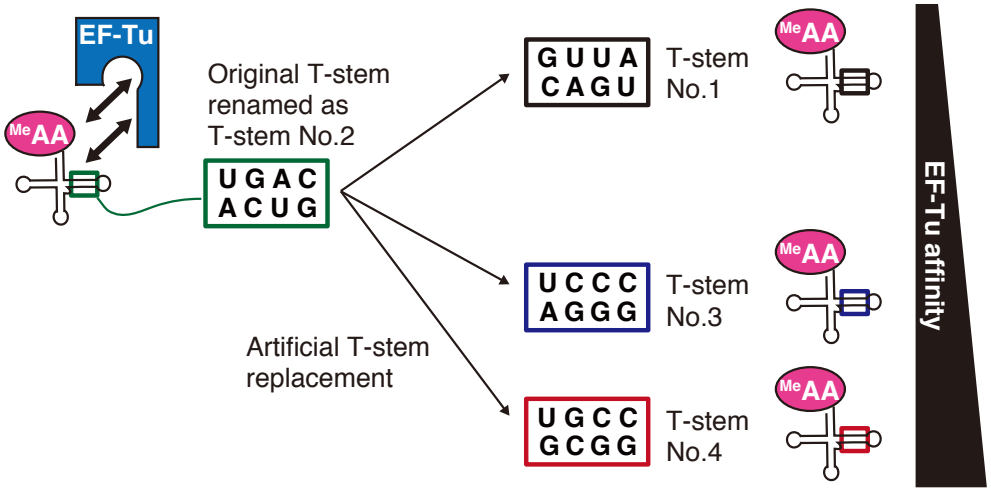

Figure 48. Strategy to reinforce the weak EF-Tu affinity of ^{Me}AA-tRNA. The T-stem sequence of ^{Me}AA-tRNA was replaced with three kinds of different sequences that possess different EF-Tu affinities. The original T-stem was renamed as T-stem No.2, and newly prepared T-stems are named as T-stem No.1, 3, and 4. Their EF-Tu affinities are supposed to increase as the number increases.

The feasibility of affinity reinforcement by means of T-stem replacement was confirmed using Phe-tRNA$^{AsnE2}_{GAC}$ and ^{Me}F-tRNA$^{AsnE2}_{GAC}$. When the T-stem sequence of Phe-tRNA$^{AsnE2}_{GAC}$ was replaced with the other T-stem sequences, their EF-Tu affinities were changed in the expected order (Fig. 49a). Similar change in EF-Tu affinity was observed also in case of ^{Me}F-tRNA$^{AsnE2}_{GAC}$ (Fig. 49b). The control experiment using uncharged tRNA$^{AsnE2}_{GAC}$ revealed that these tRNAs did not bind to EF-Tu unless they were charged with amino acid (Fig. 49c).

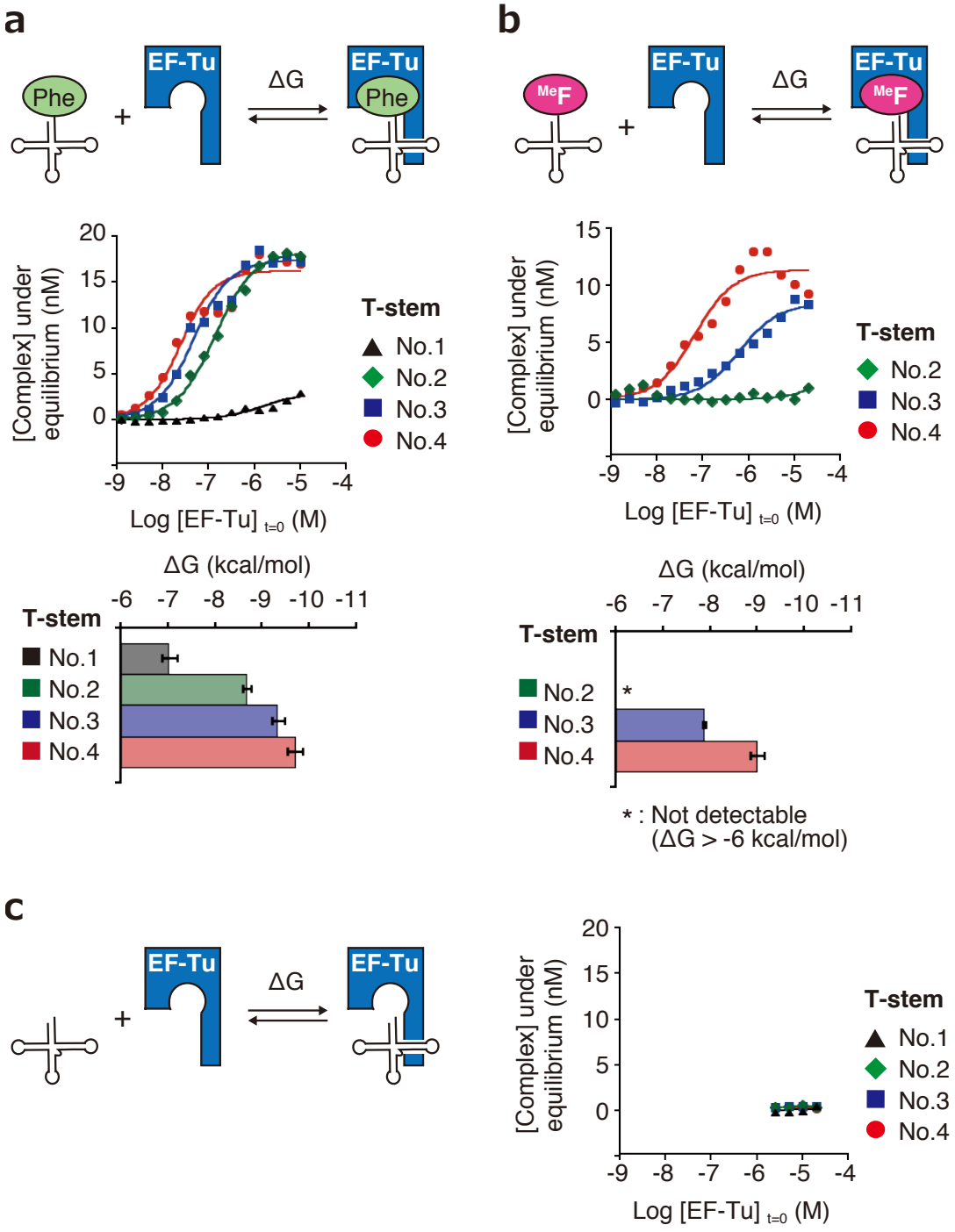

Figure 49. Reinforcement of the EF-Tu affinity of aminoacyl-tRNAs by the T-stem replacement strategy. **a,b**, Quantification of the EF-Tu affinities of Phe-tRNA$^{AsnE2}_{GAC}$ (**a**) and ^{Me}F-tRNA$^{AsnE2}_{GAC}$ (**b**) with different T-stem sequences. Asterisk indicates that the affinity is undetectable level (ΔG > -6 kcal/mol). Error bar: standard error. **c**, Control experiment where uncharged tRNA was used.

The effect of the EF-Tu affinity reinforcement on translational accuracy and efficiency was examined using the above Phe-tRNA$^{AsnE2}_{GAC}$ and ^{Me}F-tRNA$^{AsnE2}_{GAC}$ derivatives. The mRNA template was designed to determine the peptide expression level by means of tricine-SDS-PAGE (autoradiographic detection of [^{14}C]-Asp in the C-terminal FLAG peptide region), and accuracy of the decoding was evaluated by MALDI-TOF-MS of translation products (Fig. 50a,b). When the weak T-stem No.1 was used, a side-product containing Ile in place of Phe or ^{Me}F was observed in MALDI-TOF-MS, whereas tRNAs with relatively stronger T-stems yielded only the correct peptide (Fig. 50a,b). When the translational efficiency of the Phe-containing peptide was quantified using the FIT system that did not contain Ile, the synthetic rate was decreased as the EF-Tu affinity increased (Fig. 50c). The cause of decrease in synthetic rate could be attributed to following hypotheses: (1) Phe-tRNA$^{AsnE2}_{GAC}$ with stronger T-stems might dominantly bind to EF-Tu, which would hinder the binding of canonical pAA-tRNAs and lead to decrease in the synthetic rate of full-length peptide or (2) the too strong interaction between EF-Tu and the replaced T-stem might hinder the fast accommodation of aminoacyl-tRNA into ribosome as observed in previous researches[49,152]. Further investigation is required to identify the exact reason. In case of ^{Me}F-tRNA$^{AsnE2}_{GAC}$, the use of T-stem No.2 and No.3 exhibited the best synthetic rate compared to T-stem No.1 or 4 (Fig. 50d). Collectively, above results demonstrated that the EF-Tu affinity of aminoacyl-tRNAs can be dramatically changed by the T-stem replacement strategy and the translational accuracy and efficiency can be improved by selecting the best T-stem from the four candidates.

The effect of T-stem replacement on EF-Tu affinity and translational accuracy was examined also for other four ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s along with ^{Ac}K-tRNA$^{AsnE2}_{GAC}$ (Fig. 51). In all cases, the reinforcement of EF-Tu affinity was achieved, and accordingly the translational accuracy was improved. Notably, in case of ^{Me}L and ^{Ac}K incorporation, the side-product was observed as the major peak when the original T-stem No.2 was used, whereas the undesired side-product was successfully suppressed by replacing the T-stem with No.4 sequence. The EF-Tu affinity reinforcement was confirmed for another eight kinds of ^{Me}AA-tRNA$^{AsnE2}_{GAC}$'s as well (Fig. 52).

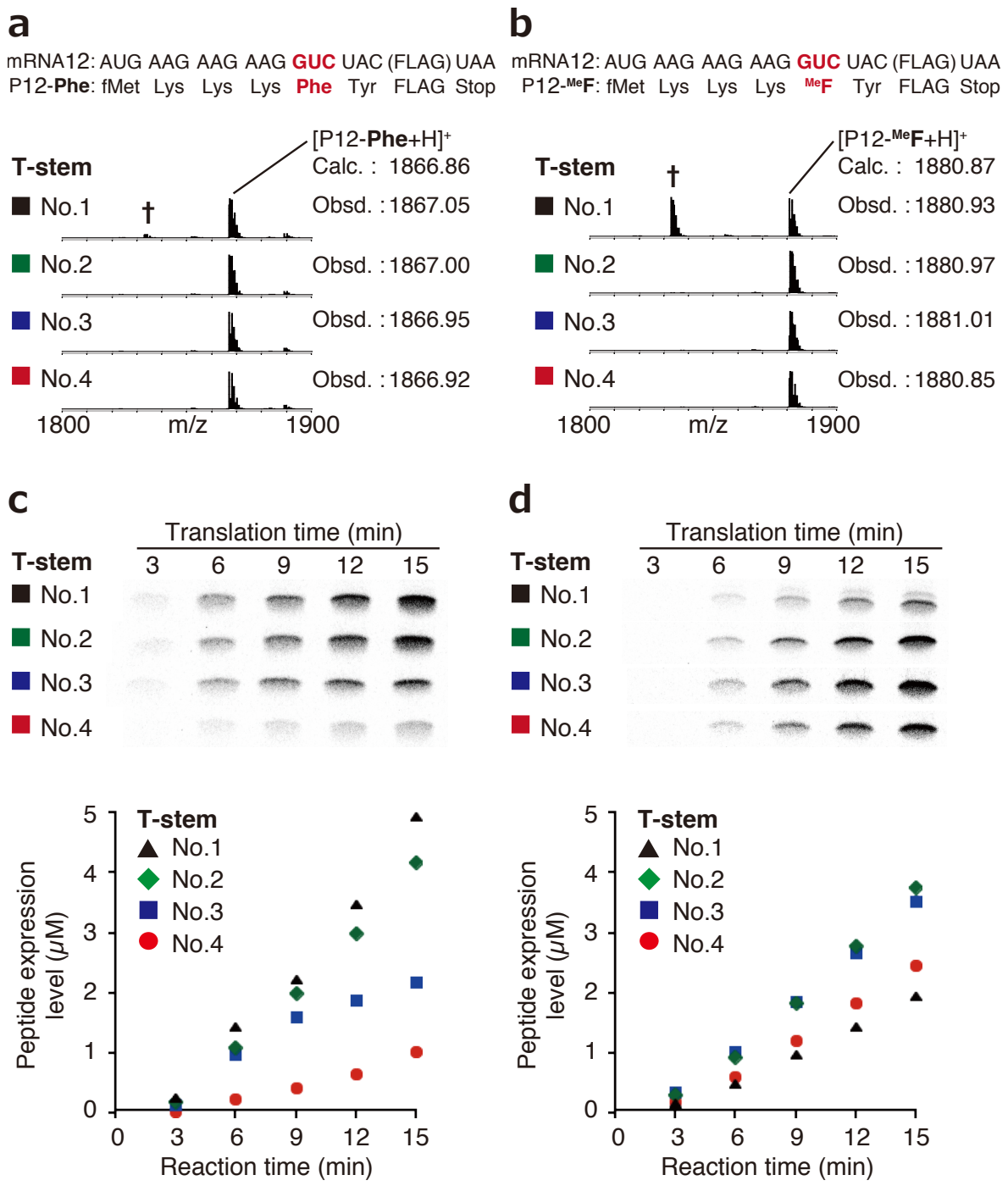

Figure 50. Improvement in translation accuracy and efficiency achieved by selecting the appropriate T-stem. a,b, MALDI-TOF-MS of peptide products after 15 min translation reaction. The GUC codon was decoded by Phe-tRNA$^{AsnE2}_{GAC}$ (**a**) and ^{Me}F-tRNA$^{AsnE2}_{GAC}$ (**b**) that possessed different T-stem sequences. †Peak corresponds to the side-product containing Ile in place of Phe (**a**) and ^{Me}F (**b**). **c,d**, Time course of the peptide expression levels in the FIT system containing Phe-tRNA$^{AsnE2}_{GAC}$ (**c**) and ^{Me}F-tRNA$^{AsnE2}_{GAC}$ (**d**).

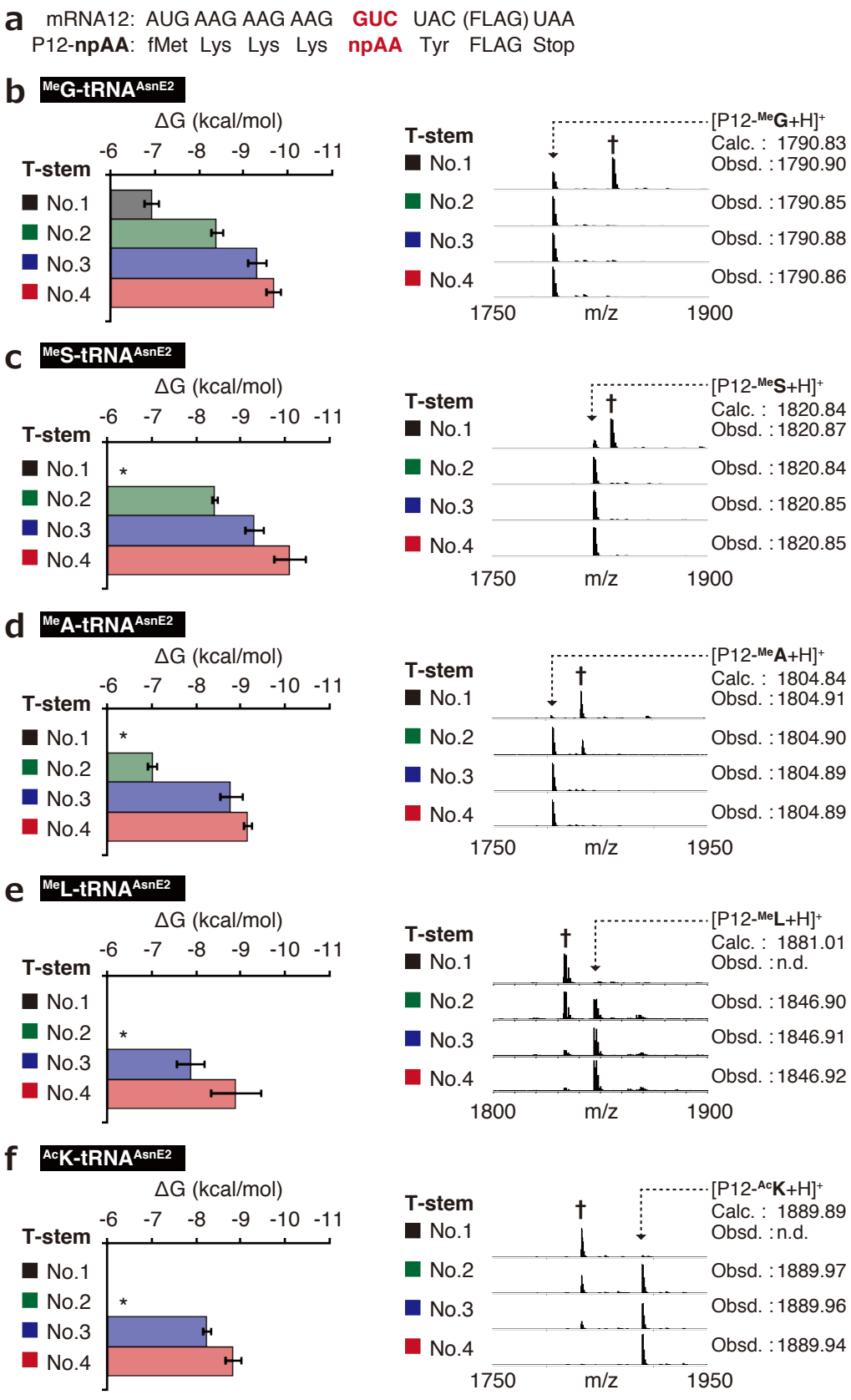

Figure 51. Improved translation accuracy by reinforcing EF-Tu affinities of ^{Me}AA-tRNAs and ^{Ac}K-tRNA. a, Sequences of mRNA12 and its resultant P12-npAA peptide. The GUC codon was decoded by npAA-tRNA$^{AsnE2}_{GAC}$ of interest. **b**–**f**, EF-Tu affinities and their effects on translational accuracy examined for ^{Me}G-tRNA$^{AsnE2}_{GAC}$ (**b**), ^{Me}S-tRNA$^{AsnE2}_{GAC}$ (**c**), ^{Me}A-tRNA$^{AsnE2}_{GAC}$ (**d**), ^{Me}L-tRNA$^{AsnE2}_{GAC}$ (**e**), and ^{Ac}K-tRNA$^{AsnE2}_{GAC}$ (**f**) that possessed different T-stem sequences. Asterisk indicates that the affinity is undetectable level (ΔG > -6 kcal/mol). Error bar: standard error. †Peak corresponds to the side-product containing Ile in place of npAA.

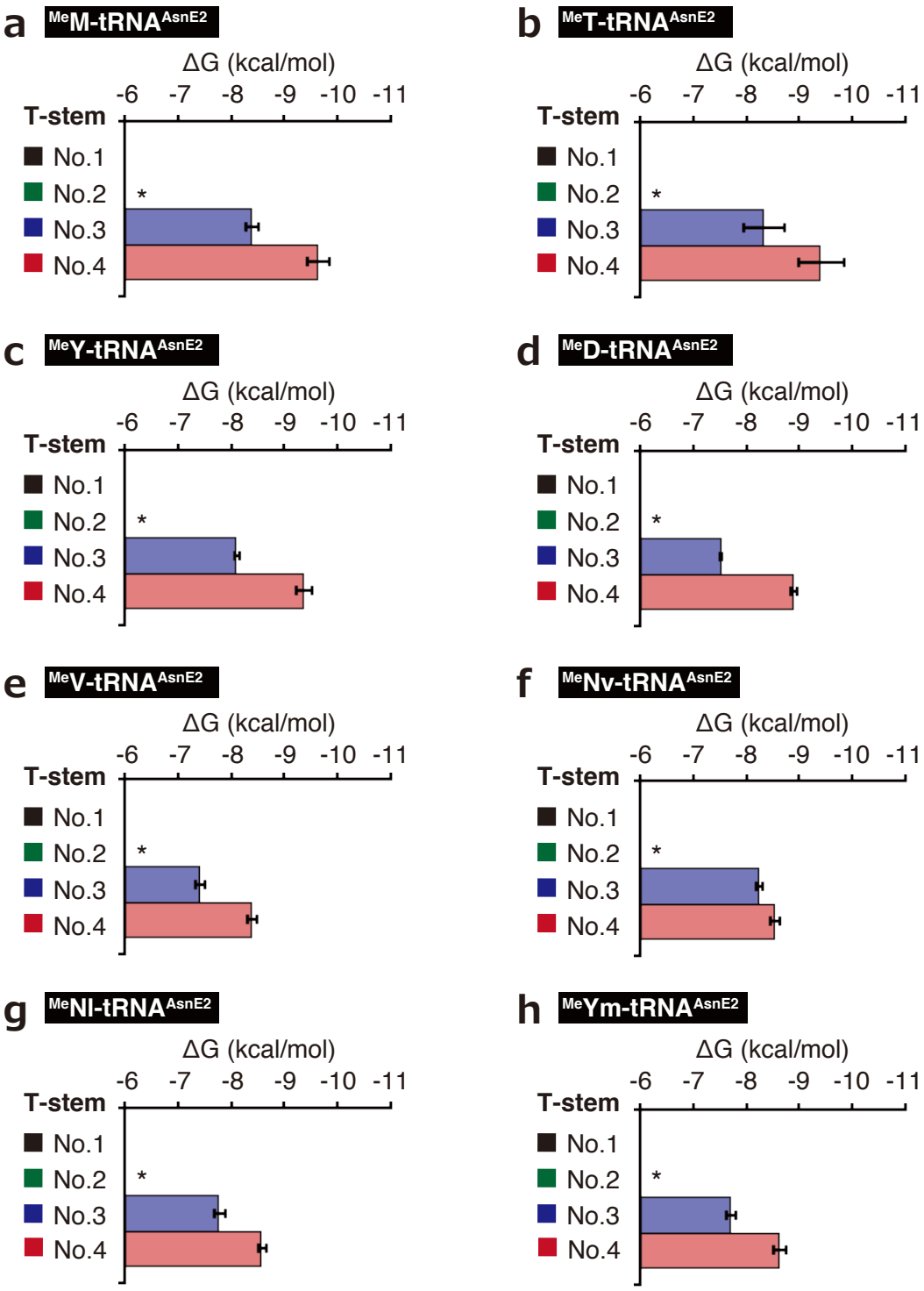

Figure 52. EF-Tu affinity reinforcement for other ^{Me}AA-tRNAs. a–h, EF-Tu affinities of ^{Me}M-tRNA$^{AsnE2}_{GAC}$ (**a**), ^{Me}T-tRNA$^{AsnE2}_{GAC}$ (**b**), ^{Me}Y-tRNA$^{AsnE2}_{GAC}$ (**c**), ^{Me}D-tRNA$^{AsnE2}_{GAC}$ (**d**), ^{Me}V-tRNA$^{AsnE2}_{GAC}$ (**e**), ^{Me}Nv-tRNA$^{AsnE2}_{GAC}$ (**f**), ^{Me}Nl-tRNA$^{AsnE2}_{GAC}$ (**g**), ^{Me}Ym-tRNA$^{AsnE2}_{GAC}$ (**h**) containing different T-stem sequences. Asterisk indicates that the affinity is undetectable level (ΔG > -6 kcal/mol). Error bar: standard error.

3.2.3. Expression of highly *N*-methylated peptide via EF-Tu affinity tuning for ^{Me}AA-tRNAs.

In order to demonstrate the improved *N*-methyl-peptide synthesis by the EF-Tu affinity tuning strategy, I next tried to express highly *N*-methylated peptide containing six distinct ^{Me}AAs. To achieve this, I first optimized the translation conditions (see Supplementary results 3.4.2 for the optimization experiments), and as the result, two contrivances have been developed. First, the EF-Tu affinities of ^{Me}AA-tRNAs were successfully tuned to the range observed for pAA-tRNAs by selecting the appropriate T-stem sequence for individual ^{Me}AA-tRNAs (Fig. 53a). Under this condition, EF-Tu can bind to all of pAA-tRNAs and ^{Me}AA-tRNAs with equivalent affinities. Second, the concentration of each ^{Me}AA-tRNA was adjusted based on flexizyme-catalyzed acylation efficiencies, and the concentration was adjusted to 10 μM; whereas, in conventional method, 50 μM each ^{Me}AA-tRNAs were added without considering acylation efficiencies (Fig. 53b). In this experiment, a model peptide composed of 6 ^{Me}AAs along with 14 pAAs was expressed in the FIT system in which five pAAs (Phe, Leu, Ile, Val, and Ala) were excluded and six ^{Me}AA-tRNAs (^{Me}Y-tRNA$^{AsnE2TS3}_{GAA}$, ^{Me}A-tRNA$^{AsnE2TS3}_{GAU}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GGC}$, ^{Me}V-tRNA$^{AsnE2TS1}_{GAC}$, ^{Me}G-tRNA$^{AsnE2TS5}_{CAG}$, and ^{Me}S-tRNA$^{AsnE2TS5}_{GAG}$) were supplemented instead (Fig. 53 c,d). When the translation was conducted by the conventional system[93], the products were a mixture of the desired peptide and various side-products containing pAAs in place of ^{Me}AAs, and the expression level of the mixture was 0.7 μM in total (Fig. 53e,f). On the other hand, the optimized system developed in this study yielded only the desired peptide with improved expression level of 1.8 μM. Clearly, this result demonstrated that the cause of translational disorder can be attributed to the insufficient EF-Tu affinities of ^{Me}AA-tRNAs, and the affinity tuning strategy can dramatically improve the synthetic efficiency and accuracy of *N*-methyl peptides.

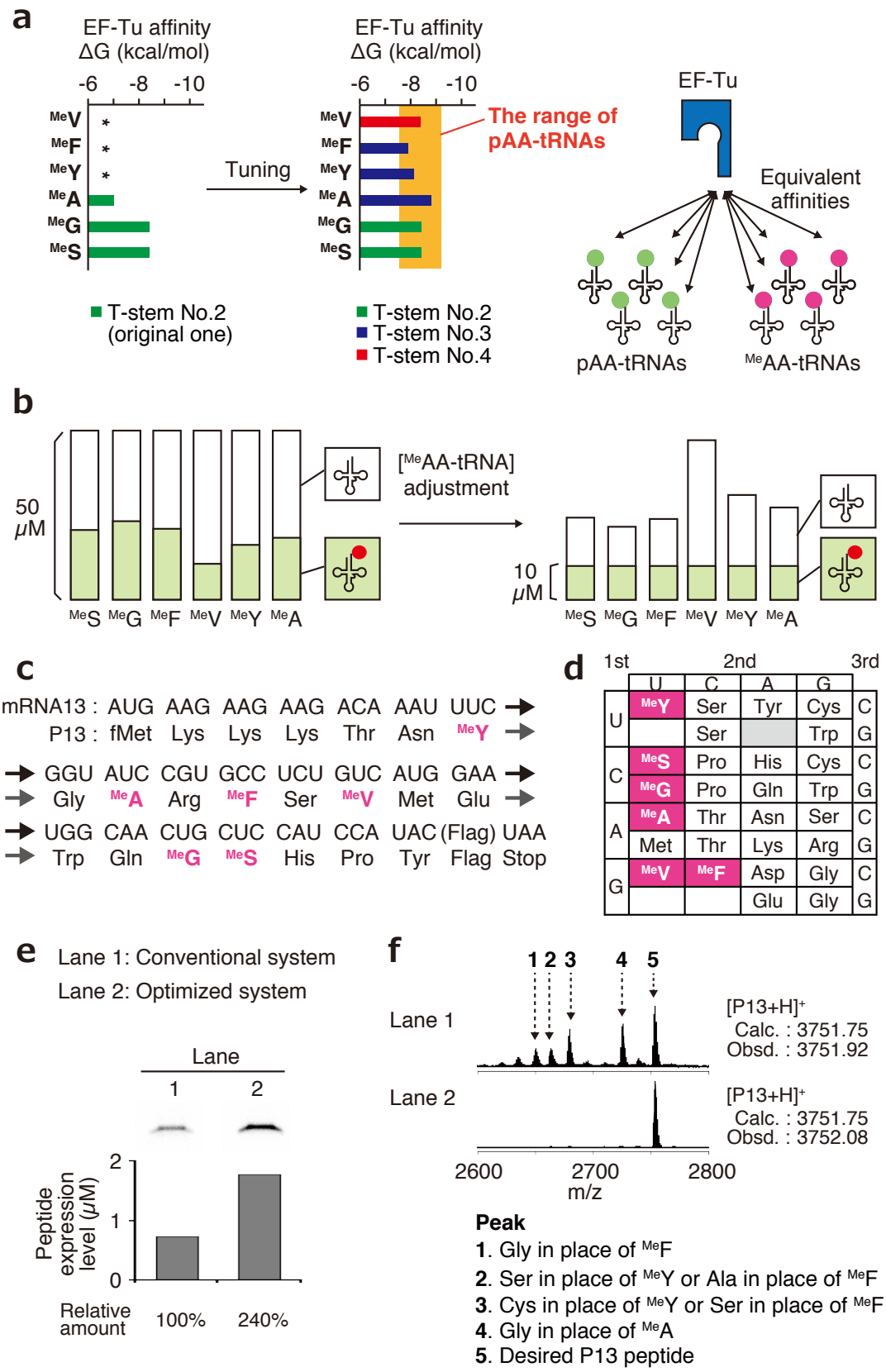

Figure 53. Accurate and efficient synthesis of a model peptide containing six distinct ^{Me}AAs under the optimized condition. a, Summary of the EF-Tu affinity tuning. **b**, The adjustment of each ^{Me}AA-tRNA concentration based on the flexizyme-catalyzed acylation efficiencies. **c**, Sequences of mRNA13 and its resultant P13 peptide. Arrows indicate the sequence connections. **d**, A reprogrammed genetic that codes for 6 ^{Me}AAs along with 15 pAAs. **e,f**, Comparison between the conventional system and the optimized system based on the analysis of the peptide products by tricine-SDS-PAGE (**e**) and MALDI-TOF-MS (**f**).

Next, the variety of ^{Me}AAs was further expanded to nine, and a 32-mer model peptide composed of nine distinct ^{Me}AAs along with 14 pAAs was expressed under the optimized condition (Fig. 54 a–c). In this experiment, three kinds of ^{Me}AAs with unnatural side-chains were added to the repertoire. Compared to a control translation using 20 pAAs, the P14 peptide with nine *N*-methyl modifications was expressed in 34% efficiency to the control with maintenance of translational accuracy as demonstrated by MALDI-TOF-MS (Fig. 54 d,e), which demonstrated that the EF-Tu affinity tuning strategy dramatically expands the variety of simultaneously usable ^{Me}AAs and enables us to express highly *N*-methylated peptides.

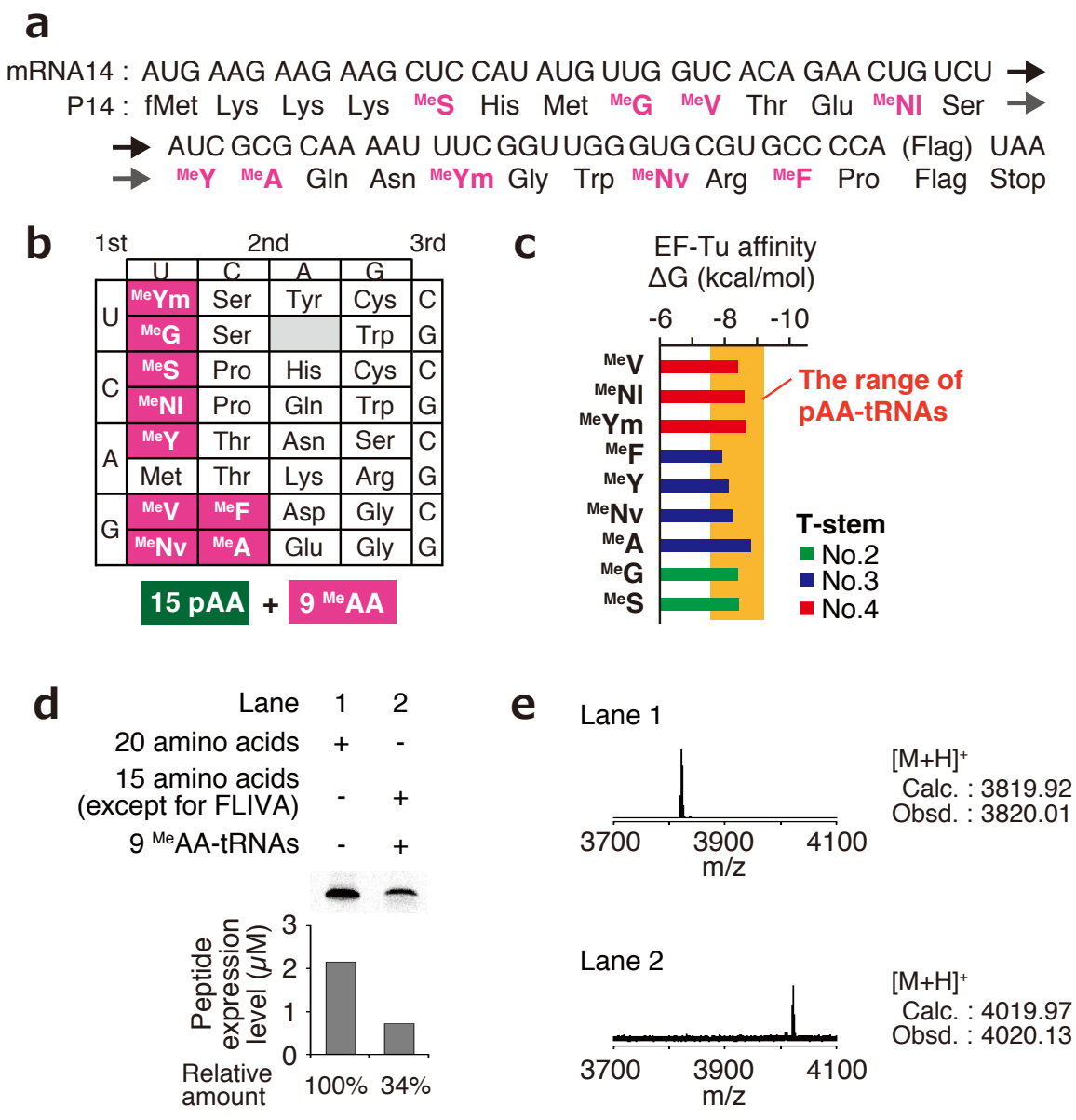

Figure 54. Expression of a highly *N*-methylated peptide containing nine distinct ^{Me}AAs along with 15 pAAs. a, Sequences of mRNA14 and its resultant P14 peptide. The CUC, UUG, GUC, CUG, AUC, GCG, UUC, GUG, and GCC codons were translated by ^{Me}S-tRNA$^{AsnE2TS5}_{GAG}$, ^{Me}G-tRNA$^{AsnE2TS5}_{CAA}$, ^{Me}V-tRNA$^{AsnE2TS1}_{GAC}$, ^{Me}Nl-tRNA$^{AsnE2TS1}_{CAG}$, ^{Me}Y-tRNA$^{AsnE2TS3}_{GAU}$, ^{Me}A-tRNA$^{AsnE2TS3}_{CGC}$, ^{Me}Ym-tRNA$^{AsnE2TS1}_{GAA}$, ^{Me}Nv-tRNA$^{AsnE2TS3}_{CAC}$, and ^{Me}F-tRNA$^{AsnE2TS3}_{GGC}$, respectively. **b**, A reprogrammed genetic code that codes for nine ^{Me}AAs as well as 15 pAAs. **c**, Summary of the EF-Tu affinity tuning. **d,e**, Expression of the P14 peptide containing nine different ^{Me}AAs. The mRNA was translated in either the FIT system that contained 20 pAAs (as control) or the FIT system that contained 15 pAAs (excluding Phe, Leu, Ile, Val, and Ala) and the 9 ^{Me}AA-tRNAs. The peptide products were analyzed by tricine-SDS-PAGE (**d**) and MALDI-TOF-MS (**e**).

3.3. Conclusion

Here I report a novel methodology that achieves the ribosomal synthesis of nonstandard peptides containing a rich variety of ^{Me}AAs. First, this study quantitatively demonstrated that the binding affinity between EF-Tu and many ^{Me}AA-tRNAs were weak (Fig. 47), which supported my hypothesis that the cause of inefficient ^{Me}AA incorporation could be attributed to the impaired EF-Tu-mediated delivery of ^{Me}AA-tRNAs into ribosome. This study also demonstrated that the weak EF-Tu affinities of ^{Me}AA-tRNAs can be reinforced by the T-stem replacement strategy (Fig. 48, 52, 54, and 55). Notably, when the EF-Tu binding affinity was adjusted to too strong values, the synthetic efficiency of the corresponding peptide was decreased (Fig. 50 and 59), which was consistent with previous reports where pAAs and fluorescent npAA were examined[49,152].

Tuning of the affinities between EF-Tu and each of ^{Me}AA-tRNAs successfully improved synthetic efficiency and accuracy of a model peptide containing six distinct ^{Me}AAs (Fig. 53). Moreover, the developed translation system also achieved the synthesis of a highly *N*-methylated peptide containing as many as nine distinct ^{Me}AAs (Fig. 54), which has expanded the variety of simultaneously usable ^{Me}AAs from three[69,93] to nine. Notably, the nine ^{Me}AAs included ^{Me}V that has been reported as an incompatible substrate for ribosomal polypeptide synthesis[104]. Notably, the variety of 24 building blocks (15 pAAs and nine ^{Me}AAs) achieved in this study has overcome the previous limitation of 23 variety (20 pAAs and three npAAs)[91,153]. These facts clearly demonstrated the improved synthetic efficiency and accuracy of highly *N*-methylated peptides.

As the cause of inefficiency observed for npAA incorporation, three problems could be supposed: (1) impaired EF-Tu-mediated delivery of the corresponding npAA-tRNAs, (2) slow peptidyl transfer reaction involving the npAA, and (3) hindered ejection of nascent polypeptides containing npAA through the exit tunnel. Previous researches rarely identified which process was the actual cause of inefficiency, and therefore it is possible that many npAAs could not be incorporated efficiently just because the corresponding npAA-tRNA could not bind to EF-Tu with sufficient strengths. This study has discovered an impressive and practical fact that the synthetic efficiency and accuracy of *N*-methylated peptides can be improved dramatically only by tuning the insufficient EF-Tu affinities of ^{Me}AA-tRNAs. This fact implies that the same strategy can facilitate efficient incorporation of a rich variety of npAAs as well as ^{Me}AAs. Such an example was demonstrated for ^{Ac}K in this study (Fig. 51).

The development of EF-Tu affinity tuning method has another potential benefit that it enables us to directly examine the peptidyl transfer reaction without the need to consider the possible EF-Tu-binding problem, which would facilitate the further engineering of peptidyl transfer machinery. It is also possible that the method contributes to biological studies regarding the mechanisms of ribosomal peptidyl transfer reaction and substrate discrimination (*i.e.* quality control of translation).

Although the conventional RaPID system have discovered macrocyclic *N*-methyl-peptide binders to drug target proteins, the variety of ^{Me}AAs found therein has been limited up to three[93]. As a result, most discovered nonstandard peptide drugs did not exhibit the desired pharmacokinetic properties such as cell-membrane permeability and stability against enzymatic decomposition. The present engineered translation system developed in this study would enable us to synthesize a highly *N*-methylated peptide library with sufficient accuracy and efficiency. The

integration of such a peptide library with peptide screening technologies is expected to facilitate the discovery of practical nonstandard peptide drugs with excellent pharmacological properties in future.

3.4. Supplementary results

3.4.1. Development of a novel method to quantify affinities between EF-Tu and aminoacyl-tRNAs

The conventional method to quantify the affinities between EF-Tu and aminoacyl-tRNAs consists of following steps (Fig. 55a)[41,154]: (1) a pAA-tRNA of interest is synthesized by the corresponding AARS using radiolabeled pAA. (2) The pAA-tRNA is mixed with EF-Tu. (3) After reaching equilibrium, RNase A is added, which decomposes only unbound pAA-tRNA within 30 seconds, whereas the pAA-tRNA bound on EF-Tu is protected. (4) The addition of 10% TCA quenches the RNase A reaction and precipitates the ternary complex. (5) Detection of the radioactivity from the precipitate allows for the quantification of the complex under the equilibrium state. As radiolabeled ^{Me}AA substrates are not available, I had to modify this quantification method so that tRNA was radiolabeled instead of amino acid. Two approaches to synthesize radioactive tRNA were tested. In the first approach, the 5′-terminal phosphate group was radiolabeled by dephosphorylation followed by phosphorylation in the presence of [γ-^{32}P]-ATP (Fig. 55b). The radioactive tRNA can be charged with ^{Me}AAs by flexizymes, and the resultant ^{Me}AA-tRNA would allow for the quantification of their binding affinities. In the other approach, the 3′-terminal phosphate group was radiolabeled by ligation of a tRNA lacking 3′-terminal adenosine with [α-^{32}P]-ATP (Fig. 55c).

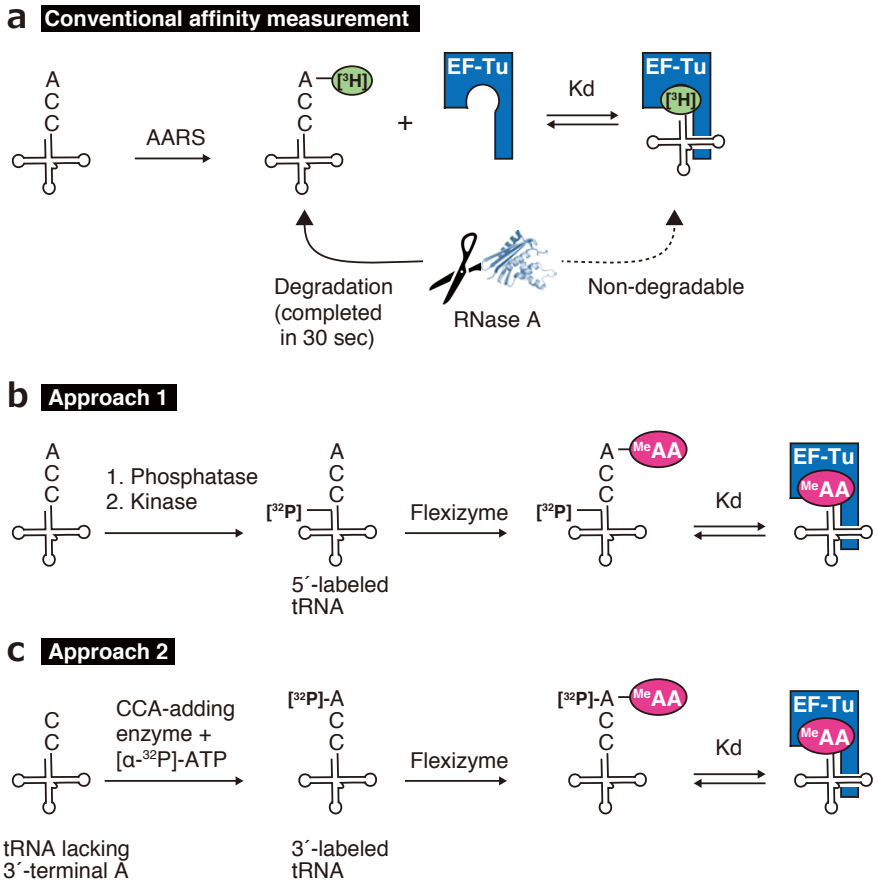

Figure 55. Method to quantify the affinity between EF-Tu and aminoacyl-tRNA. a, Conventional affinity measurement using radiolabeled amino acid. **b**, The first approach to radiolabel tRNA. The 5′-terminal phosphate group was radiolabeled by dephosphorylation followed by phosphorylation in the presence of [γ-^{32}P]-ATP. The radioactive tRNA can be charged with ^{Me}AA by flexizymes and the binding affinity could be measured. **c**, The second approach to radiolabel tRNA. A tRNA lacking 3′-terminal adenosine was ligated with [α-^{32}P]-ATP by means of CCA-adding enzyme.

Next, I examined whether the two radiolabeled tRNAs could be rapidly decomposed by RNase A within 30 seconds. When the 5′-labeled $tRNA^{Phe}$ was decomposed by various concentration of RNase A in the absence of EF-Tu, around 10% intensity of remaining radioactivity was observed even when 2909 ng/μL of RNase A was present (Fig. 56a). In contrast, the 3′-labeled $tRNA^{Phe}$ was completely decomposed by lower concentration of RNase A (Fig. 56b), which conferred sufficient S/N ratio for the complex quantification in the following experiments. The quantification of $tRNA^{Phe}$ without RNase A decomposition exhibited the linear qualitative relationship between the radioactivity of the precipitate and the input tRNA amount ranging from 2 to 2000 fmol (*i.e.* 0.2 to 200 nM) (Fig. 56c). Based on these results, the 3′-labeling was used in the following experiments.

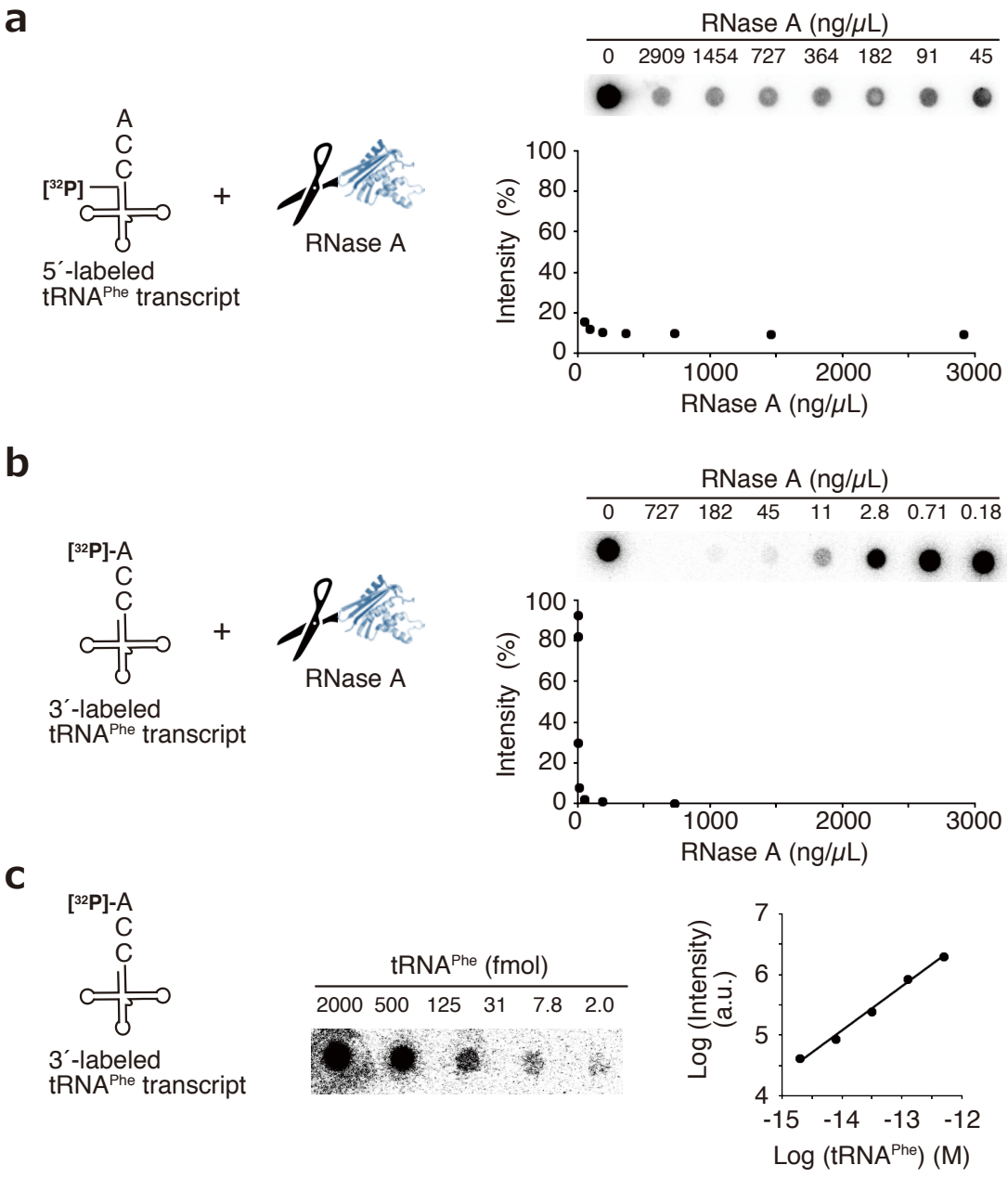

Figure 56. Comparison of the two radiolabeling methods. a, Incomplete degradation of 5′-labeled tRNA by RNase A. The tRNA sample decomposed by RNase A was quantified by 10% TCA precipitation, filtration, and autoradiography. **b**, Complete degradation of 3′-labeled tRNA by RNase A. Note that the range of RNase A concentration is different from **a**. **c**, Quantification of calibration standards. In this experiment, different amounts of 3′-labeled tRNAs were quantified without RNase A reaction.

Next, I attempted to quantify the EF-Tu affinity of Phe-tRNA$^{AsnE2}_{GAC}$. The acylation efficiency by flexizyme (eFx) catalysis[92] was determined to be 53.2% by acid denaturing PAGE and ethidium bromide staining (Fig. 57b). The radioactivity measurement of known concentrations of Phe-tRNA$^{AsnE2}_{GAC}$ exhibited the linear qualitative relationship between the radioactivity and the input tRNA amount, which allowed for the quantification of the tRNA$^{AsnE2}_{GAC}$ amount based on the radioactivity (Fig. 57c). 20 μM pre-charged Phe-tRNA$^{AsnE2}_{GAC}$ was mixed with various concentrations of EF-Tu and the complex concentrations under equilibrium were quantified (Fig. 57d). The plots of the complex concentration to the initial concentration of EF-Tu correlated well with the theoretical curve of the binding equation, and allowed for the quantifications of the dissociation constant and the binding affinity value. Collectively, I successfully developed the method to quantify the binding affinity between EF-Tu and aminoacyl-tRNA using 3′-radiolabeled tRNA.

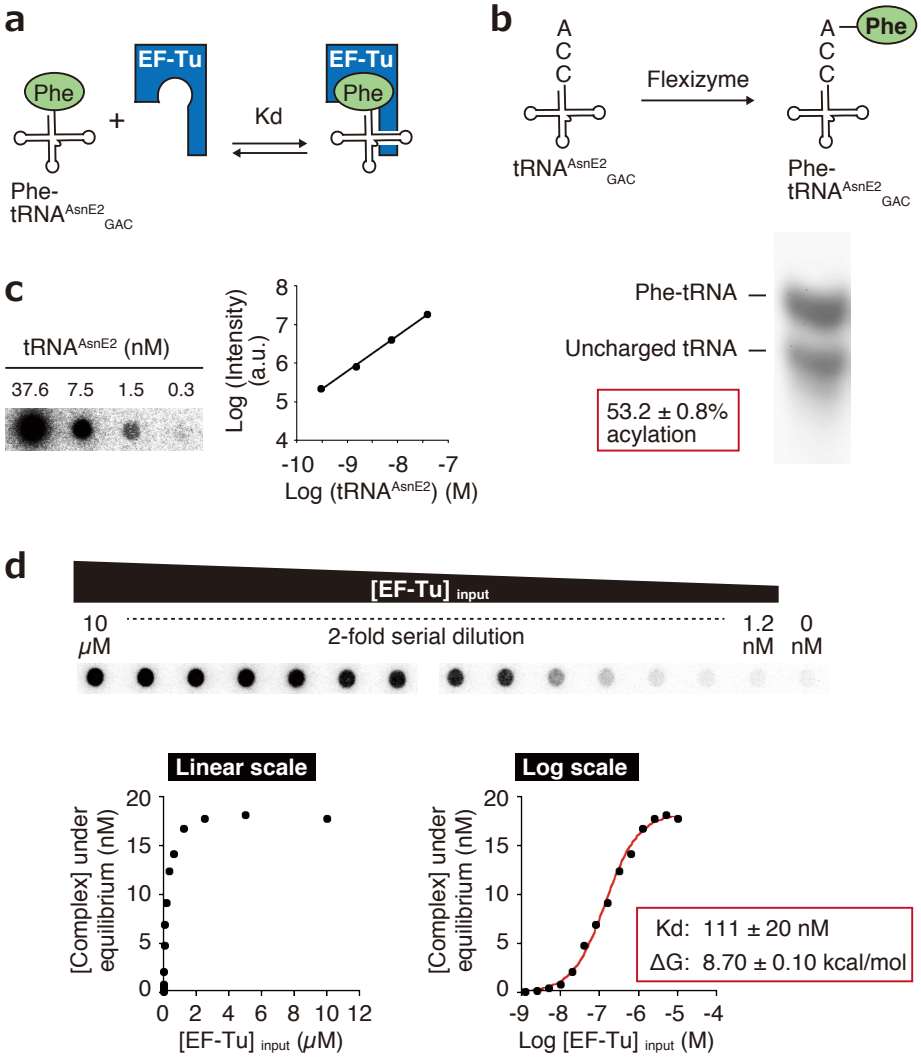

Figure 57. Affinity measurement between EF-Tu and Phe-tRNA$^{AsnE2}_{GAC}$. a, Scheme of the binding equilibrium of interest. **b**, Measurement of the flexizyme-catalyzed acylation efficiency. The acylation product was analyzed by acid denaturing PAGE and ethidium bromide staining. **c**, Quantification of calibration standards of known concentrations of Phe-tRNA$^{AsnE2}_{GAC}$. Their radioactivities were measured without RNase A reaction. **d**, Quantification of the affinity between EF-Tu and Phe-tRNA$^{AsnE2}_{GAC}$. The complex concentrations were plotted against the concentration of EF-Tu. The plot is represented in both linear and logarithmic scales. The fitting of theoretical curve allowed for quantifications of the Kd value as well as ΔG value. Error: standard error.

3.4.2. Optimization of translation condition to express peptides containing multiple ^{Me}AAs.

As it had been known that the Mg^{2+} concentration dramatically affects the translational accuracy and efficiency[16,144], I tried to optimize the concentration to improve the synthetic efficiency and accuracy of *N*-methyl-peptides. Although a fixed 12 mM Mg^{2+} condition has been utilized in our laboratory, I noticed that the pre-charged ^{Me}AA-tRNAs added to translation mixture intake the free Mg^{2+} ions into their structures as counter cations, and accordingly the concentration of free Mg^{2+} decreases (Fig. 58a). Therefore, the concentration of Mg^{2+} should be optimized depending on the amount of tRNA. In order to examine the relationship between the optimal Mg^{2+} concentration and the tRNA amount, mRNA15 was translated to P15-3^{Me}F peptide containing three consecutive ^{Me}F's in the FIT system in which two pAAs (Leu and Val) were excluded and three pre-charged ^{Me}F-tRNAs (^{Me}F-tRNA$^{AsnE2TS3}_{GAG}$, ^{Me}F-tRNA$^{AsnE2TS3}_{CAG}$, and ^{Me}F-tRNA$^{AsnE2TS3}_{GAC}$) and/or various concentrations of non-functional tRNA (Fig. 58a,b) were supplemented instead. As the non-functional tRNA, a tRNA$^{AsnE2}_{CUA}$ containing a T-stem that has extremely low EF-Tu affinity was used. Note that the corresponding UAG codon is a stop codon that was not present in the mRNA15 sequence, and therefore the non-functional tRNA is supposed not to interact either with EF-Tu or ribosome. When the non-functional tRNA was not added, the peptide expression level decreased as the Mg^{2+} concentration increased. In contrast, when 200 and 400 μM non-functional tRNAs were added, the optimal Mg^{2+} concentration shifted to 14 and 18 mM, respectively. MALDI-TOF-MS of the optimal conditions revealed that the correct P15-3^{Me}F peptide was accurately synthesized (Fig. 58d). These results demonstrated that the Mg^{2+} concentration should be increased depending on the amount of tRNAs present in translation mixture.

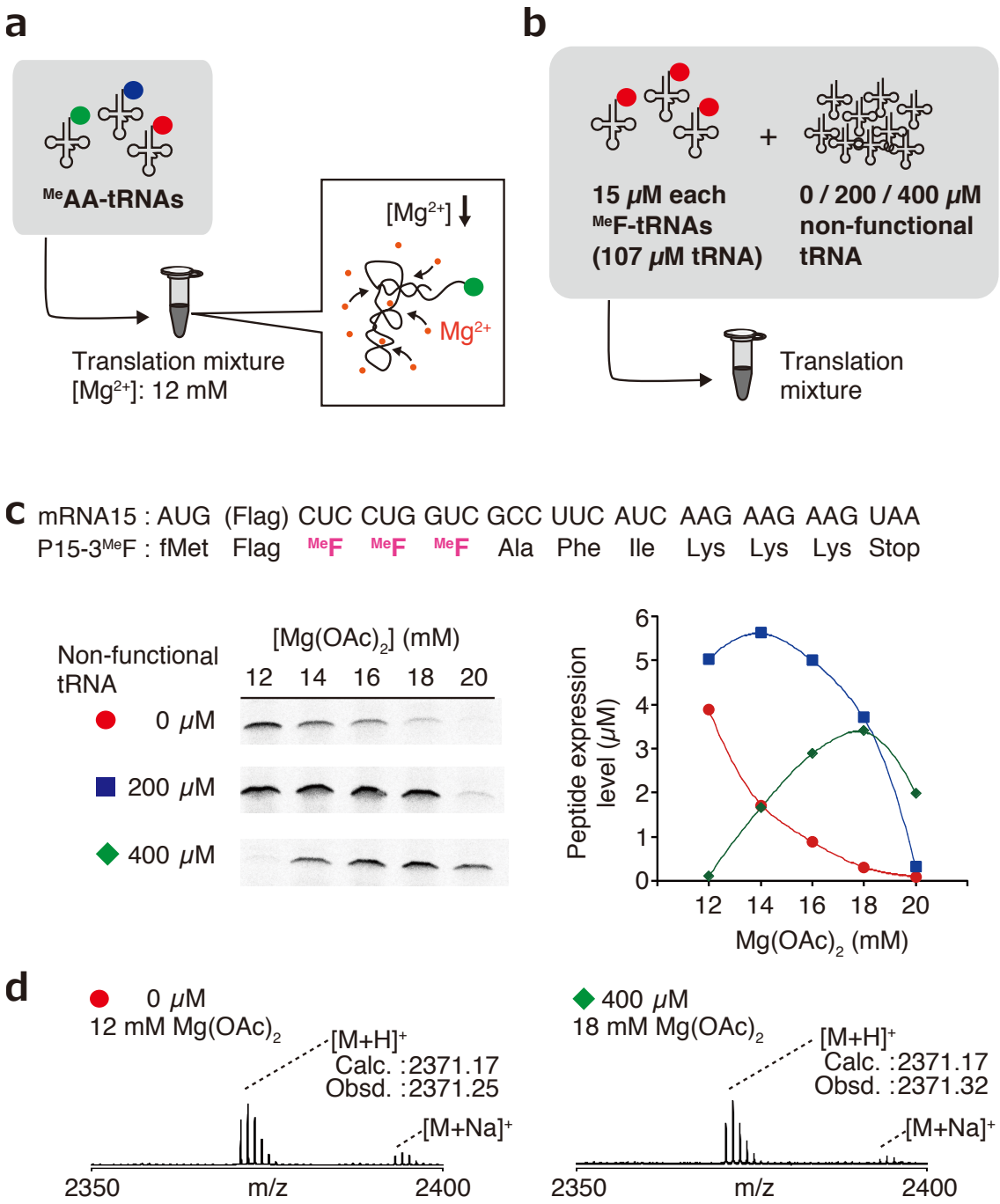

Figure 58. Improved translational efficiency by optimization of Mg^{2+} concentration. a, Hypothesized cause of the translational inaccuracy observed when a mass amount of pre-charged npAA-tRNAs are added. As the free Mg^{2+} ions chelate to the pre-charged npAA-tRNAs, the concentration of free Mg^{2+} must decrease as the amount of npAA-tRNA increases. The decrease in free Mg^{2+} concentration would decrease the translation efficiency. **b**, A model experiment to examine the relationship between the optimal Mg^{2+} concentration and the tRNA amount. **c,d**, Quantification of the peptide expression levels in the presence of different concentrations of $[Mg(OAc)_2]$ and non-functional tRNA. The GUC, CUG, and GUC codons were decoded by $^{Me}F\text{-}tRNA^{AsnE2TS3}_{GAG}$, $^{Me}F\text{-}tRNA^{AsnE2TS3}_{CAG}$, and $^{Me}F\text{-}tRNA^{AsnE2TS3}_{GAC}$, respectively. The peptide products were analyzed by tricine-SDS-PAGE (**c**) and MALDI-TOF-MS (**d**).

As higher amount of tRNA decreased the peptide expression level (Fig. 58c), I next tried to minimize the concentration of pre-charged ^{Me}AA-tRNAs. In this experiment, the mRNA15 was translated to a model peptide containing six consecutive ^{Me}Fs in the FIT system in which five pAAs (Phe, Leu, Ile, Val, and Ala) were excluded and six ^{Me}F-tRNAs (with GAG, CAG, GAC, GGC, GAA, and GAU anticodons; all T-stem No.3) were supplemented instead (Fig. 59a). By decreasing the concentration of each ^{Me}F-tRNA from 40 to 20 μM, the synthetic rate was improved as expected (Fig. 59b). When the concentration of each ^{Me}F-tRNA were further decreased from 20 μM, the synthetic rate did not change, and in all cases the accurate synthesis of the correct peptide was observed in MALDI-TOF-MS (Fig. 59b,c). Among them, the 10 μM condition was used in the following experiment.

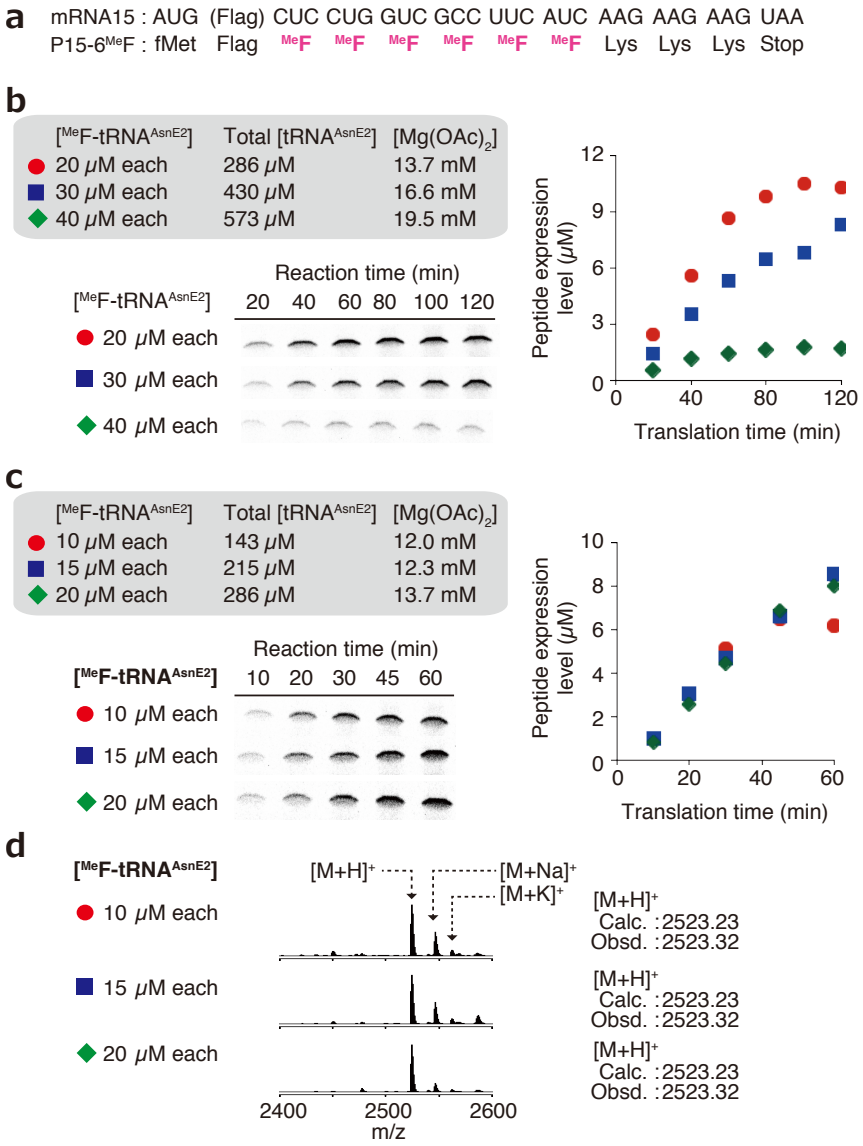

Figure 59. Improved translational efficiency by optimization of ^{Me}F-tRNA concentration. a, Sequences of mRNA15 and its resultant P15-6^{Me}F peptide. The GUC, CUG, GUC, GCC, UUC, and AUC codons were translated by ^{Me}F-tRNA$^{AsnE2TS3}_{GAG}$, ^{Me}F-tRNA$^{AsnE2TS3}_{CAG}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GAC}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GGC}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GAA}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GAU}$, respectively. **b–d**, Optimization of ^{Me}F-tRNA concentration. The optimal $Mg(OAc)_2$ concentrations were estimated according to the amount of tRNA^{AsnE2}s based on the result of Figure 58. The peptide products were analyzed by tricine-SDS-PAGE (**b**,**c**) and MALDI-TOF-MS after 45 min translation reaction (**d**).

I tried to further improve the translational efficiency by optimizing EF-Tu concentration, IF2 concentration, pH, and Mg^{2+} concentration (Fig. 60). Unfortunately, none of them significantly improved the efficiency.

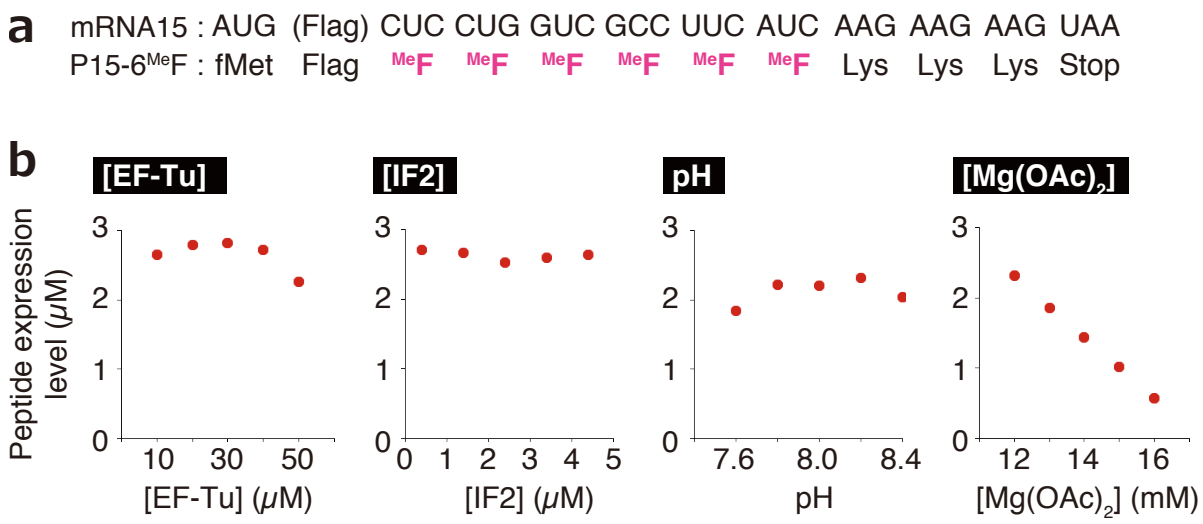

Figure 60. Attempt to further improve translation efficiency by titrating other factors. a, Sequences of mRNA15 and its resultant P15-6^{Me}F peptide. The GUC, CUG, GUC, GCC, UUC, and AUC codons were translated by ^{Me}F-tRNA$^{AsnE2TS3}_{GAG}$, ^{Me}F-tRNA$^{AsnE2TS3}_{CAG}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GAC}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GGC}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GAA}$, ^{Me}F-tRNA$^{AsnE2TS3}_{GAU}$, respectively. **b**, Titration of EF-Tu, IF2, and $Mg(OAc)_2$ concentrations and pH. In every figure, the leftmost point is the original condition.

Although the EF-Tu affinity of ^{Me}AA-tRNA was successfully reinforced by the T-stem mutation strategy, the appropriate value to which the affinity should be tuned was still unclear. In order to examine the appropriate EF-Tu affinity for ^{Me}AA incorporation, the model peptide containing six consecutive ^{Me}F's was translated in the FIT system that contained six ^{Me}F-tRNAs that possessed different T-stem sequences (Fig. 61a). As the result, the use of T-stem No.3 maximized the efficiency with sufficient translational accuracy demonstrated by MALDI-TOF-MS (Fig. 61b,c) These results revealed that the EF-Tu affinity of ^{Me}AA-tRNA should be adjusted to around -8 kcal/mol rather than -9 kcal/mol (Fig. 61d).

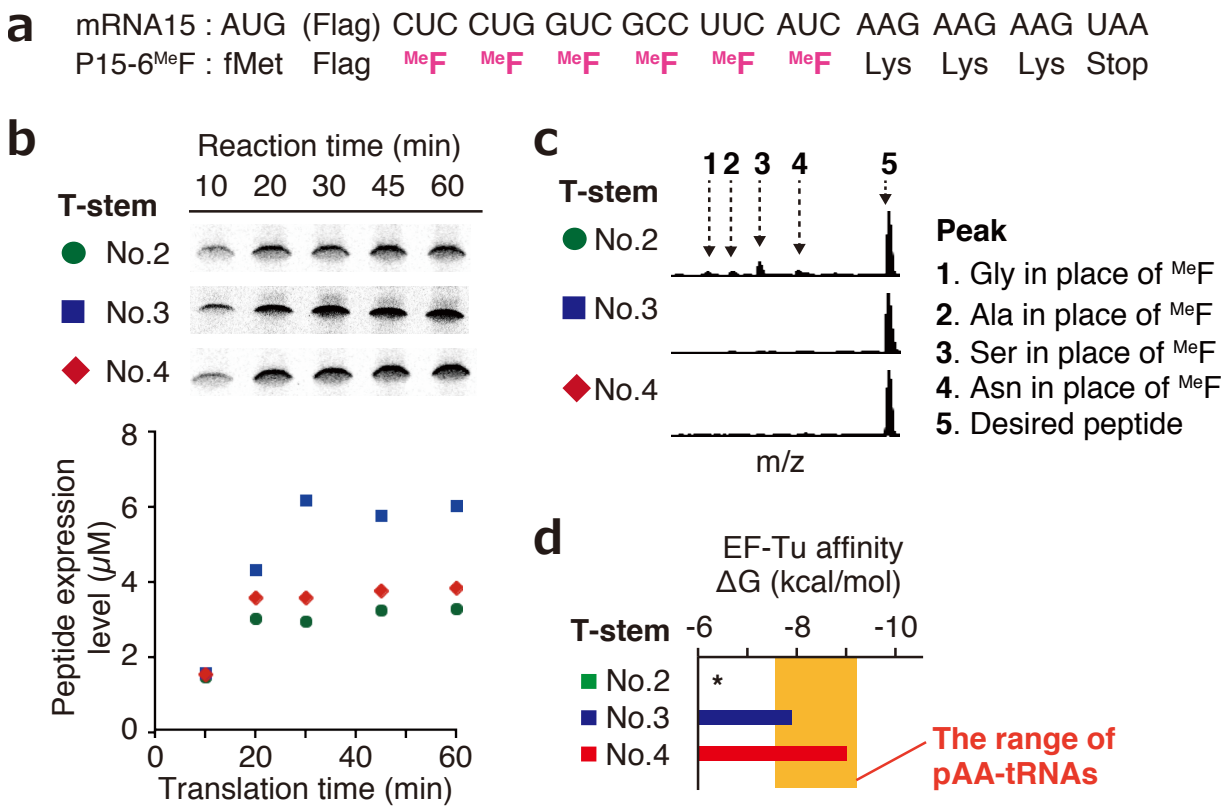

Figure 61. Determination of the appropriate affinities between EF-Tu and ^{Me}AA-tRNAs. a, Sequences of mRNA15 and its resultant P15-6^{Me}F peptide. The GUC, CUG, GUC, GCC, UUC, and AUC codons were translated by ^{Me}F-tRNA$^{AsnE2}_{GAG}$, ^{Me}F-tRNA$^{AsnE2}_{CAG}$, ^{Me}F-tRNA$^{AsnE2}_{GAC}$, ^{Me}F-tRNA$^{AsnE2}_{GGC}$, ^{Me}F-tRNA$^{AsnE2}_{GAA}$, ^{Me}F-tRNA$^{AsnE2}_{GAU}$, respectively, that possessed different T-stem sequences. **b,c**, Comparison among different T-stem sequences by the analysis of the peptide products by tricine-SDS-PAGE (**b**) and MALDI-TOF-MS after 45 min translation reaction (**c**). **d**, The EF-Tu affinities of ^{Me}F-tRNA$^{AsnE2}_{GAC}$ that possessed different T-stem sequences.

3.5. Methods

3.5.1. Preparation of tRNAs, flexizymes, mDNAs, and mRNAs

All oligonucleotides listed in Supplementary Table 9 were purchased from Eurofins, Japan. The sequences of tRNA transcripts prepared in this study were listed in Supplementary Table 10. Double stranded DNA templates that coded for tRNAs were prepared by primer extension followed by PCR as follows: Appropriate forward and reverse primers (1 μM each, see Supplementary Table 11 for the primers used) were mixed in 100 μL of the PCR mixture (10 mM Tris-HCl (pH 9.0), 50 mM KCl, 2.5 mM $MgCl_2$, 0.25 mM each dNTPs, 0.1% (v/v) Triton X-100, and 45 nM *Taq* DNA polymerase). Primer extension was conducted by denaturing (95 °C for 1 min) followed by 5 cycles of annealing (50 °C for 1 min) and extending (72 °C for 1 min). 1 μL of the resulting reaction mixture was 200-fold diluted with 199 μL of the PCR mixture containing 0.5 μM each appropriate forward and reverse primers (0.5 μM each, see Supplementary Table 11 for the primers used), and PCR was conducted by 12 cycles of denaturing (95 °C for 40 sec), annealing (50 °C for 40 sec), and extending (72 °C for 40 sec). Amplification of the PCR product was confirmed by 3% agarose gel electrophoresis and ethidium bromide staining. The resulting DNA was purified by phenol/chloroform extraction and ethanol precipitation, and then dissolved in 20 μL of water. The concentrations of tRNA was measured by A_{260} of the 10-fold diluted solution.

Transcription reaction was conducted by incubating the 200-μL transcription mixture (40 mM Tris·HCl (pH 8.0), 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 22.5 mM $MgCl_2$, 3.75 mM each NTPs, 5 mM GMP, 22.5 mM KOH, 10% (v/v) DNA template prepared above, and 120 nM T7 RNA polymerase) at 37 °C overnight. The transcription mixture was mixed with $MnCl_2$ (100 mM, 4 μL) and RQ1 RNase-Free DNase (1 U/μL, 1μL, Promega), and incubated at 37 °C for 30 min. The resultant tRNA transcript was precipitated by isopropanol and dissolved in water. The tRNA transcript was purified by 8% denaturing PAGE and ethanol precipitation, and then dissolved in 10 μL water.

Flexizymes (dFx and eFx) were prepared as previously described[77,92].

Double stranded DNA templates that coded for mRNAs (mDNAs, see Supplementary Table 12 for the mRNA sequences) were prepared by primer extension and PCR (see Supplementary Table 13 for the primers used). The resulting mDNAs were purified by phenol/chloroform extraction and ethanol precipitation, and then dissolved in 10 μL of water. The concentration was measured by 8% native PAGE and ethidium bromide staining with 100 bp Quick-Load DNA Ladders (New England BioLabs) as reference.

mRNAs were prepared by *in vitro* transcription as follows: the 200-μL transcription mixture (40 mM Tris·HCl (pH 8.0), 1 mM spermidine, 0.01% (v/v) Triton X-100, 10 mM DTT, 30 mM $MgCl_2$, 5 mM each NTPs, 30 mM KOH, 10% (v/v) DNA template prepared above, and 120 nM T7 RNA polymerase) was incubated at 37 °C overnight. The transcription mixture was mixed with $MnCl_2$ (100 mM, 4 μL) and RQ1 RNase-Free DNase (1 U/μL, 1μL, Promega), and incubated at 37 °C for 30 min. The resultant mRNA was precipitated by isopropanol and dissolved in water. The tRNA transcript was purified by 8% denaturing PAGE and ethanol precipitation, and then dissolved in 10 μL water. The concentrations of mRNA was measured by A_{260} of the 10-fold diluted solution.

3.5.2. Synthesis of aminoacyl-tRNAs by flexizymes

Amino acids activated with an appropriate ester group (Phe-CME, Tyr-CME, Ser-DBE, ^{Me}G-DBE, ^{Me}S-DBE, ^{Me}A-DBE, ^{Me}F-CME, ^{Me}L-DBE, ^{Me}M-DBE, ^{Me}T-DBE, ^{Me}Y-CME, ^{Me}D-DBE, ^{Me}V-DBE, ^{Me}Nl-DBE, ^{Me}Ym-CME, ^{Me}Nv-DBE, and ^{Ac}K-DBE; CME: cyanomethyl ester; DBE:3,5-dinitrobenzyl ester) were synthesized by the procedure reported previously[92,104,128].

Aminoacyl-tRNAs were prepared by the following procedure: 12 μL HEPES-KOH buffer (pH 7.5, 83 mM) containing 42 μM tRNA and 42 μM flexizyme (eFx for CME-activated amino acids and dFx for DBE-activated ones) was heated at 95 °C for 2 min and cooled to 25 °C over 5 min. $MgCl_2$ (3 M, 4 μL) was added and the mixture was incubated at 25 °C for 5 min. The reaction was initiated by addition of activated amino acid substrate in DMSO (25 mM, 4 μL) and incubated on ice for acylation (incubation time: 2 h for Phe-CME, Tyr-CME, ^{Me}G-DBE, ^{Me}A-DBE, and ^{Ac}K-DBE; 6 h for Ser-DBE, ^{Me}S-DBE, ^{Me}F-CME, ^{Me}L-DBE, ^{Me}M-DBE, and ^{Me}Ym-CME; and 10 h for ^{Me}Y-CME; 24h for ^{Me}T-DBE, ^{Me}D-DBE, ^{Me}V-DBE, ^{Me}Nl-DBE, and ^{Me}Nv-DBE). After acylation, the reaction was stopped by addition of 80 μL of 0.3 M sodium acetate (pH 5.2), and the RNA was precipitated by ethanol precipitation. The pellet was rinsed twice with 70% ethanol containing 0.1 M sodium acetate (pH 5.2), and once with 70% ethanol. The resulting aminoacyl-tRNA was dissolved in 1 mM sodium acetate (pH 5.2) just before adding to translation solution. The sample of non-functional $tRNA^{AsnE2}_{CUA}$ (Fig. 58) was prepared by the mock reaction where DMSO was added instead of activated amino acid substrate solution.

3.5.3. Measurement of the efficiency of flexizyme-catalyzed aminoacylation

The pellet of ethanol-precipitated 50 pmol tRNA pre-charged with pAA or npAA was dissolved in 0.52 μL of 10 mM sodium acetate (pH 5.2) and mixed with 5.0 μL of acid PAGE loading buffer (83% (v/v) formamide, 150 mM sodium acetate (pH 5.2), and 10 mM EDTA). The solution was loaded on an acid denaturing polyacrylamide gel (12% (w/v) acrylamide/bisacrylamide (19:1), 8M urea, and 50 mM sodium acetate (pH 5.2)) and electrophoresis was conducted at 300 V (approximately 10 V/cm) for 20 h. The gel was stained with ethidium bromide and analyzed using a Typhoon FLA 7000 (GE Healthcare). Aminoacylation efficiency was calculated based on the band intensities of aminoacyl-tRNA (A) and free tRNA (T) and is presented as (A)/[(A)+(T)].

3.5.4. Radiolabeling of either 5′- or 3′-terminus of tRNA

The 5′-terminus of a tRNA was radiolabeled as follows: the 4-μL dephosphorylation reaction mixture (1 μM tRNA, 50 mM Bis-Tris-Propane-HCl (pH 6.0), 1 mM $MgCl_2$, 0.1 mM $ZnCl_2$, and 0.5 U antarctic phosphatase (New England BioLabs)) was incubated at 37 °C for 30 min, and then incubated at 70 °C for 5 min. The solution (4 μL) was mixed with 9.1 μL of H_2O, 2.0 μL of 10× buffer (700 mM Tris-HCl (pH 7.6), 100 mM $MgCl_2$, 50 mM DTT), 2.9 μL of 1.7 μM [γ-^{32}P]-ATP, and 2.0 μL of 10 U/μL T4 polynucleotide kinase (New England BioLabs). The resulting 20-μL kinase reaction mixture was incubated at 37 °C for 60 min. The ^{32}P-labeled tRNA (4 pmol) was mixed with non-labeled tRNA (121 pmol). The tRNA was purified by phenol/chloroform extraction and ethanol

precipitation, and then dissolved in 10 μL of water. The concentrations of tRNA was measured by A_{260} of the 50-fold diluted solution.

The 3′-terminus of a tRNA was radiolabeled as follows: 7.5 μL of 16.7 μM tRNA lacking 3′-terminal adenosine was incubated at 80 °C for 3 min, and then put on ice for 10 min. The 25-μL CCA-adding reaction mixture (120 mM Gly-NaOH (pH 9.0), 75 mM $MgCl_2$, 30 mM DTT, 5 μM tRNA lacking 3′-terminal adenosine prepared above, 10 mM non-radiolabeled ATP, 0.67 μM [α-^{32}P]-ATP, and 200 nM CCA-adding enzyme) was incubated at 37 °C for 20 min. The tRNA was purified by phenol/chloroform extraction, Micro Bio-Spin 30 column (Bio-Rad), and ethanol precipitation, and then dissolved in 5 μL of water. The concentrations of tRNA was measured by A_{260} of the 100-fold diluted solution.

3.5.5. Quantification of the affinity between EF-Tu and aminoacyl-tRNA

The aminoacyl-tRNA was prepared as described above (section 3.4.2) except that the acylation reaction was conducted using the mixture of 3′-radiolabeled tRNA (10%) and non-radiolabeled tRNA (90%). The resulting pellet of aminoacyl-tRNA was dissolved in 3.2 μL of 10 mM sodium acetate (pH 5.2). The concentration of aminoacyl-tRNA was adjusted as follows: 0.7 μL of the aminoacyl-tRNA solution was 100-fold diluted with 69.3 μL of 10 mM sodium acetate (pH 5.2) and the concentrations of tRNA and flexizyme were measured by A_{260} of the diluted solution. The concentration of aminoacyl-tRNA was calculated based on the tRNA concentration and the flexizyme-catalyzed acylation efficiency (Supplementary Table 14). The concentration of aminoacyl-tRNA was adjusted to 2.0 μM by adding the appropriate volume of 10 mM sodium acetate (pH 5.2). Just prior to use, the concentration of aminoacyl-tRNA was adjusted to 100 nM by mixing the 2.0 μM aminoacyl-tRNA solution with buffer A (50 mM HEPES-KOH (pH7.6), 100 mM KOAc, 12 mM $Mg(OAc)_2$, 1 mM GTP, 1 mM DTT, 20 mM creatine phosphate, 2 mM spermidine, 3 mM phosphoenol pyruvate, and 0.1 μg/μL pyruvate kinase from rabbit muscle (Sigma)).

EF-Tu stored in its GDP-bound form was converted to GTP-bound form by incubating the EF-Tu solution in buffer A at 37 °C for 30 min. 12 μL each of samples containing different concentrations of EF-Tu, typically ranging from 1.5 nM to 25 μM, were prepared by twofold serial dilution on ice. 9.6 μL of the EF-Tu solution was mixed with 2.4 μL of 100 nM aminoacyl-tRNA and incubated on ice for 20 min. Under the equilibrium binding condition, 10 μL each solution was mixed with 1 μL of 1 mg/mL RNase A (Sigma) on ice. After 20 sec, the reaction was quenched by addition of 50 μL of 10% (v/v) trichloroacetic acid (TCA) containing 0.1 mg/mL unfractionated yeast tRNA on ice. The precipitate was filtrated using 0.45 μM pore-size nitrocellulose membrane assembled in Bio-Dot microfiltration apparatus (Bio-Rad), and washed by six times of 200 μl each 5% (v/v) TCA. The membrane was soaked in 95% (v/v) ethanol for 5 min, and the dried membrane was analyzed by autoradiography using a Typhoon FLA 7000 (GE Healthcare). To correct for the background signal derived from any aminoacyl-tRNA that may remain after the 20 sec treatment with RNase A, a no EF-Tu control was analyzed in parallel and its radioactivity was subtracted from the experimental data. The resultant radioactivity was converted to the concentration of ternary complex using the conversion factor determined by the calibration aminoacyl-tRNA samples in buffer A containing 20, 4, 0.8, and 0.16

nM tRNA. Equilibrium dissociation constants were determined by fitting the binding data to the following equation **3** using KaleidaGraph program (Hulinks).

The K_D equation is defined as follows, where aatRNA, EFTu, and aatRNA•EFTu denote aminoacyl-tRNA, EF-Tu, and their complex, respectively.

$$K_D = \frac{[\text{aatRNA}]_{eq} \times [\text{EFTu}]_{eq}}{[\text{aatRNA•EFTu}]_{eq}} - (\mathbf{1})$$

$$K_D = \frac{([\text{aatRNA}]_{input} - [\text{aatRNA•EFTu}]_{eq}) \times ([\text{EFTu}]_{input} - [\text{aatRNA•EFTu}]_{eq})}{[\text{aatRNA•EFTu}]_{eq}} - (\mathbf{2})$$

$$[\text{aatRNA} \bullet \text{EFTu}]_{eq} = \frac{\left([\text{aatRNA}]_{input} + [\text{EFTu}]_{input} + K_D - \sqrt{([\text{aatRNA}]_{input} + [\text{EFTu}]_{input} + K_D)^2 - 4 \times [\text{aatRNA}]_{input} \times [\text{EFTu}]_{input}}\right)}{2} - (\mathbf{3})$$

As the concentration of aminoacyl-tRNA varied slightly in every experiment due to its multistep preparation, both K_D and the aminoacyl-tRNA concentration were set as variables in the fitting analysis.

3.5.6. *In vitro* translation

The reconstituted cell-free translation system[32] contained all necessary components for translation except for RF1. Concentrations of translation components were optimized in our previous studies[77] as follows: 50 mM HEPES-KOH (pH 7.6), 100 mM KOAc, 2 mM GTP, 2 mM ATP, 1 mM CTP, 1 mM UTP, 20 mM creatine phosphate, 12–20 mM $Mg(OAc)_2$, 2 mM spermidine, 1 mM DTT, 100 μM 10-formyl-5,6,7,8-tetrahydrofolic acid, 1.2 μM ribosome, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 10 μM EF-Tu, 10 μM EF-Ts, 0.26 μM EF-G, 0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 0.6 μM MTF, 4 μg/mL creatine kinase, 3 μg/mL myokinase, 0.1 μM pyrophosphatase, 0.1 μM nucleotide diphosphate kinase, 0.1 μM T7 RNA polymerase, 0.73 μM AlaRS, 0.03 μM ArgRS, 0.38 μM AsnRS, 0.13 μM AspRS, 0.02 μM CysRS, 0.06 μM GlnRS, 0.23 μM GluRS, 0.09 μM GlyRS, 0.02 μM HisRS, 0.4 μ IleRS, 0.04 μM LeuRS, 0.11 μM LysRS, 0.03 μM MetRS, 0.68 μM PheRS, 0.16 μM ProRS, 0.04 μM SerRS, 0.09 μM ThrRS, 0.03μM TrpRS 0.02 μM TyrRS, 0.02 μM ValRS, 1.5 mg/mL native *E. coli* tRNA mixture (Roche), 200 μM each 3–20 pAAs (see Supplementary Table 16 for the amino acids added in each experiment), 10–50 μM each $tRNA^{AsnE2}$s pre-charged with npAAs, and either 0.25 μM mDNA 6.0 μM mRNA unless noted otherwise in the figure legends (see Supplementary Table 15 for the detail condition in each experiment).

The translation products were analyzed by MALDI-TOF-MS and/or autoradiography after tricine-SDS-PAGE. For MALDI-TOF-MS analysis, 1.0–10 μL of the translation reaction mixture was incubated at 37 °C for 3–120 min (see Supplementary Table 15 for the volume and reaction time in each experiment). After the reaction, the mixture was diluted with the same volume of FLAG-purification buffer (100 mM Tris-HCl (pH 7.6), 300 mM NaCl). The expressed peptide was immobilized on anti-FLAG M2 agarose beads (Sigma) by incubating at 25 °C for 1 hour. After washing the beads with 25 μL of wash buffer (50 mM Tris-HCl (pH 7.6), 150 mM NaCl), the immobilized peptides were eluted with 15 μL of 0.2% TFA. Following purification, the peptide was desalted with SPE C-tip (Nikkyo

Technos) and eluted with 1 μL of 80% acetonitrile, 0.5% acetic acid solution 50% saturated with the matrix (R)-cyano-4-hydroxycinnamic acid (Bruker Daltonics). MALDI-TOF-MS measurements were performed using a ultrafleXtreme (Bruker Daltonics) under reflect/positive mode and externally calibrated with peptide calibration standard II (Bruker Daltonics) and/or protein calibration standard I (Bruker Daltonics). For autoradiography analysis, the translation reaction was performed in the presence of 50 μM [^{14}C]-Asp instead of 200 μM cold Asp. The translation product was analyzed by 15% tricine-SDS-PAGE and autoradiography using a Typhoon FLA 7000 (GE Healthcare) without FLAG purification. The amount of peptide products was quantified based on the relative band intensity of the peptide product to the sum intensities present in the lane (*i.e.* the sum intensities of unreacted free Asp and peptide products).

For translation of mRNA13, 14 and 15, (Fig. 53, 54, 58, 59, 60, and 61), the method to prepare ^{Me}AA-tRNAs was changed as follows. First, the ethanol precipitate after the flexizyme-catalyzed acylation was dissolved in 300 μL of 0.3 M sodium acetate (pH 5.2), and precipitated by the second ethanol precipitation. This precipitation was conducted to wash out the co-precipitated Mg salt. The pellet was rinsed twice with 70% ethanol containing 0.1 M sodium acetate (pH 5.2), and once with 70% ethanol. The resulting aminoacyl-tRNA was dissolved in 1 mM sodium acetate (pH 5.2) just before adding to translation solution. The concentration of aminoacyl-tRNA was adjusted as follows: 0.25 μL of the aminoacyl-tRNA solution was 5000-fold diluted with 1250 μL of 1 mM sodium acetate (pH 5.2) and the concentrations of tRNA and flexizyme were measured by A_{260} of the diluted solution. The concentration of aminoacyl-tRNA was calculated based on the tRNA concentration and the flexizyme-catalyzed acylation efficiencies (Supplementary Table 14).

2.5. Supplementary Tables

Supplementary Table 9. List of oligonucleotides used in this study. G(Me): 2′-O-methylguanosine.

Name	Sequence
O-197	5'-TAATACGACTCACTATAGCCCGGATAGCTCAGTCGGTAGAGCAGGGGACTGAAAATCCCCGTGTCC-3'
O-198	5'-GTGCCCGGACTCGGAATCGAACCAAGGACACGGGGATTTTCA-3'
O-199	5'-GGCGTAATACGACTCACTATAG-3'
O-200	5'-GG(Me)TGCCCGGACTCGG-3'
O-201	5'-TG(M)GTGCCCGGACTCGG-3'
O-202	5'-GGCGTAATACGACTCACTATAGGTGGGGTTCCCGAGCGGCCAAAGGGAGCAGACTGTAAATCTGCCG-3'
O-203	5'-GGTGGTGGGGGAAGGATTCGAACCTTCGAAGTCTGTGACGGCAGATTTACAGTCTG-3'
O-204	5'-GG(Me)TGGTGGGGGAAGGAT-3'
O-205	5'-TG(M)GTGGTGGGGGAAGGAT-3'
O-206	5'-GGCGTAATACGACTCACTATAGGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTCGAAAACCGGAG-3'
O-207	5'-GGCGGAGAGAGGGGGATTTGAACCCCCGGTAGAGTTGCCCCTACTCCGGTTTTCGAGACC-3'
O-208	5'-GG(Me)CGGAGAGAGGGGGA-3'
O-209	5'-TG(M)GCGGAGAGAGGGGGA-3'
O-210	5'-GTAATACGACTCACTATAGGCTCTGTAGTTCAGTCGGTAGAACGGCGG-3'
O-211	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGATTGTCAGTCCGCCGTTCTACCG-3'
O-212	5'-GG(Me)CGGCTCTGCAATG-3'
O-213	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGATTGTCAGTCCGCCGTTCTACCG-3'
O-214	5'-GG(Me)CGGCTCTGACTGG-3'
O-215	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGATTGTCAGTCCGCCGTTCTACCG-3'
O-216	5'-GG(Me)CGGCTCTGAGGGG-3'
O-217	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGATTGTCAGTCCGCCGTTCTACCG-3'
O-218	5'-GG(Me)CGGCTCTGACGGG-3'
O-219	5'-TG(Me)GCGGCTCTGCAATG-3'
O-220	5'-TG(Me)GCGGCTCTGACTGG-3'
O-221	5'-TG(Me)GCGGCTCTGAGGGG-3'
O-222	5'-TG(Me)GCGGCTCTGACGGG-3'
O-223	5'-GTAATACGACTCACTATAGGCTCTGTAGTTCAGTCGGAAGAACGGCGG-3'
O-224	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGATTTTCAGTCCGCCGTTCTTCCG-3'
O-225	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGATTTTCAGTCCGCCGTTCTTCCG-3'
O-226	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGATTTTCAGTCCGCCGTTCTTCCG-3'
O-227	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGATTTTCAGTCCGCCGTTCTACCG-3'
O-228	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGATTTTGAGTCCGCCGTTCTACCG-3'
O-229	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGATTTTGAGTCCGCCGTTCTACCG-3'
O-230	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGATTTTGAGTCCGCCGTTCTACCG-3'
O-231	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGATTTTGAGTCCGCCGTTCTACCG-3'
O-232	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGAACCTCAATCCGCCGTTCTACCG-3'
O-233	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGAACCTCAATCCGCCGTTCTACCG-3'
O-234	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGAACCTCAATCCGCCGTTCTACCG-3'
O-235	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGAACCTCAATCCGCCGTTCTACCG-3'
O-236	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGAACCTGAATCCGCCGTTCTACCG-3'
O-237	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGAACCTGAATCCGCCGTTCTACCG-3'
O-238	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGAACCTGAATCCGCCGTTCTACCG-3'
O-239	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGAACCTGAATCCGCCGTTCTACCG-3'
O-240	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGATTATCAGTCCGCCGTTCTACCG-3'
O-241	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGATTATCAGTCCGCCGTTCTACCG-3'
O-242	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGATTATCAGTCCGCCGTTCTACCG-3'
O-243	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGATTATCAGTCCGCCGTTCTACCG-3'
O-244	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGAGTGTGAGTCCGCCGTTCTACCG-3'
O-245	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGAGTGTGAGTCCGCCGTTCTACCG-3'
O-246	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGAGTGTGAGTCCGCCGTTCTACCG-3'
O-247	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGAGTGTGAGTCCGCCGTTCTACCG-3'
O-248	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGACTGCCAGTCCGCCGTTCTACCG-3'
O-249	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGACTGCCAGTCCGCCGTTCTACCG-3'
O-250	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGACTGCCAGTCCGCCGTTCTACCG-3'
O-251	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGACTGCCAGTCCGCCGTTCTACCG-3'
O-252	5'-GCGGCTCTGCAATGACTCGAACACTGGACATACGGATTGCGAATCCGCCGTTCTACCG-3'
O-253	5'-GCGGCTCTGACTGGACTCGAACCAGTGACATACGGATTGCGAATCCGCCGTTCTACCG-3'
O-254	5'-GCGGCTCTGAGGGGACTCGAACCCCTGACATACGGATTGCGAATCCGCCGTTCTACCG-3'
O-255	5'-GCGGCTCTGACGGGACTCGAACCCGCGACATACGGATTGCGAATCCGCCGTTCTACCG-3'
O-256	5'-GCGGCTCTGCAATGACTCGAACACCGGACATACGGATTGTCAGTCCGCCGTTCTACCG-3'
O-257	5'-GCGGCTCTGCAATGACTCGAACACCGGACATACGGATTTAGAGTCCGCCGTTCTACCG-3'
O-258	5'-TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGAAGAAGAAG-3'
O-259	5'-CGAAGCTTACTTGTCGTCGTCGTCCTTGTAGTCGTAGACCTTCTTCTTCATGTTTTTCTC-3'
O-260	5'-CGAAGCTTACTTGTCGTC-3'
O-261	5'-GAGAAAAACATGAAGAAGAAGACAAATTTCGGTATCCGTGCCTCTGTCATGGA-3'
O-262	5'-GTCGTCCTTGTAGTCGTATGGATGGAGCAGTTGCCATTCCATGACAGAGGCACG-3'
O-263	5'-CGAAGCTTACTTGTCGTCGTCGTCCTTGTAGTC-3'
O-264	5'-GAGAAAAACATGAAGAAGAAGCTCCATATGTTGGTCACAGAACTGTCTATCGCGCAAAA-3'
O-265	5'-GTCGTCCTTGTAGTCTGGGGCACGCACCCAACCGAAATTTTGCGCGATAGACAGT-3'
O-266	5'-TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGGACTACAAGGACGACGACGACAAG-3'
O-267	5'-CGAAGCTTACTTCTTCTTGATGAAGGCGACCAGGAGCTTGTCGTCGTCGTCCT-3'
O-268	5'-CGAAGCTTACTTCTTCTT-3'

Supplementary Table 10. List of *in vitro* tRNA transcripts prepared in this study.

tRNA	Anticodon	T-stem	tRNA sequence
Phe without 3'-A	GAA	-	5'-GCCCGGAUAGCUCAGUCGGUAGAGCAGGGGACUGAAAAUCCCCGUGUCCUUGGUUCGAUUCCGAGUCCGGGCACC-3'
Tyr without 3'-A	GUA	-	5'-GGUGGGGUUCCCGAGCGGCCAAAGGGAGCAGACUGUAAAAUCUGCCGUCACAGACUUCGAAGGUUCGAAUCCUUCCCCCACCACC-3'
Ser without 3'-A	CGA	-	5'-GGAGAGAUGCCGGAGCGGCUGAACGGACCGGUCUCGAAAACCGGAGUAGGGGCAACUCUACCGGGGGUUCAAAUCCCCCUCUCUCCGCC-3'
AsnE2 without 3'-A	GAC	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCC-3'
AsnE2 without 3'-A	GAC	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCC-3'
AsnE2 without 3'-A	GAC	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCC-3'
AsnE2 without 3'-A	GAC	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCC-3'
AsnE2	GAC	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	GAC	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	GAC	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	GAC	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	GAA	No.1	5'-GGCUCUGUAGUUCAGUCGGAAGAACGGCGGACU GAA AAUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	GAA	No.2	5'-GGCUCUGUAGUUCAGUCGGAAGAACGGCGGACU GAA AAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	GAA	No.3	5'-GGCUCUGUAGUUCAGUCGGAAGAACGGCGGACU GAA AAUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	GAA	No.4	5'-GGCUCUGUAGUUCAGUCGGAAGAACGGCGGACU GAA AAUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	CAA	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAA AAUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	CAA	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAA AAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	CAA	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAA AAUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	CAA	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAA AAUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	GAG	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAG GUUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	GAG	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAG GUUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	GAG	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAG GUUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	GAG	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAG GUUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	CAG	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CAG GUUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	CAG	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CAG GUUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	CAG	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CAG GUUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	CAG	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CAG GUUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	GAU	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAU AAUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	GAU	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAU AAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	GAU	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAU AAUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	GAU	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAU AAUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	CAC	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAC ACUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	CAC	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAC ACUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	CAC	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAC ACUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	CAC	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU CAC ACUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	GGC	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU GGC AGUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	GGC	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU GGC AGUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	GGC	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU GGC AGUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	GGC	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACU GGC AGUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	CGC	No.1	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CGC AAUCCGUAUGUCCAGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
AsnE2	CGC	No.2	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CGC AAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA-3'
AsnE2	CGC	No.3	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CGC AAUCCGUAUGUCAGGGGUUCGAGUCCCCUCAGAGCCGCCA-3'
AsnE2	CGC	No.4	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU CGC AAUCCGUAUGUCGCGGGUUCGAGUCCCGUCAGAGCCGCCA-3'
AsnE2	GAC	No.0	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCCGGUGUUCGAGUCAUUGCAGAGCCGCCA-3'
Non-functional	CUA	No.0	5'-GGCUCUGUAGUUCAGUCGGUAGAACGGCGGAUU GAC AUUCCGUAUGUCCGGUGUUCGAGUCAUUGCAGAGCCGCCA-3'

Supplementary Table 11. List of oligonucleotides used for the preparation of DNA templates coding tRNAs.

tRNA	Anticodon	T-stem	Extension		PCR	
			Forward primer	Reverse primer	Forward primer	Reverse primer
Phe without 3'-A	GAA	Native	O-197	O-198	O-199	O-200
Phe	GAA	Native	O-197	O-198	O-199	O-201
Tyr without 3'-A	GUA	Native	O-202	O-203	O-199	O-204
Tyr	GUA	Native	O-202	O-203	O-199	O-205
Ser without 3'-A	CGA	Native	O-206	O-207	O-199	O-208
Ser	CGA	Native	O-206	O-207	O-199	O-209
AsnE2 without 3'-A	GAC	No.1	O-210	O-211	O-199	O-212
AsnE2 without 3'-A	GAC	No.2	O-210	O-213	O-199	O-214
AsnE2 without 3'-A	GAC	No.3	O-210	O-215	O-199	O-216
AsnE2 without 3'-A	GAC	No.4	O-210	O-217	O-199	O-218
AsnE2	GAC	No.1	O-210	O-211	O-199	O-219
AsnE2	GAC	No.2	O-210	O-210	O-199	O-220
AsnE2	GAC	No.3	O-210	O-212	O-199	O-221
AsnE2	GAC	No.4	O-210	O-214	O-199	O-222
AsnE2	GAA	No.1	O-223	O-224	O-199	O-219
AsnE2	GAA	No.2	O-223	O-225	O-199	O-220
AsnE2	GAA	No.3	O-223	O-226	O-199	O-221
AsnE2	GAA	No.4	O-223	O-227	O-199	O-222
AsnE2	CAA	No.1	O-210	O-228	O-199	O-219
AsnE2	CAA	No.2	O-210	O-229	O-199	O-220
AsnE2	CAA	No.3	O-210	O-230	O-199	O-221
AsnE2	CAA	No.4	O-210	O-231	O-199	O-222
AsnE2	GAG	No.1	O-210	O-232	O-199	O-219
AsnE2	GAG	No.2	O-210	O-233	O-199	O-220
AsnE2	GAG	No.3	O-210	O-234	O-199	O-221
AsnE2	GAG	No.4	O-210	O-235	O-199	O-222
AsnE2	CAG	No.1	O-210	O-236	O-199	O-219
AsnE2	CAG	No.2	O-210	O-237	O-199	O-220
AsnE2	CAG	No.3	O-210	O-238	O-199	O-221
AsnE2	CAG	No.4	O-210	O-239	O-199	O-222
AsnE2	GAU	No.1	O-210	O-240	O-199	O-219
AsnE2	GAU	No.2	O-210	O-241	O-199	O-220
AsnE2	GAU	No.3	O-210	O-242	O-199	O-221
AsnE2	GAU	No.4	O-210	O-243	O-199	O-222
AsnE2	CAC	No.1	O-210	O-244	O-199	O-219
AsnE2	CAC	No.2	O-210	O-245	O-199	O-220
AsnE2	CAC	No.3	O-210	O-246	O-199	O-221
AsnE2	CAC	No.4	O-210	O-247	O-199	O-222
AsnE2	GGC	No.1	O-210	O-248	O-199	O-219
AsnE2	GGC	No.2	O-210	O-249	O-199	O-220
AsnE2	GGC	No.3	O-210	O-250	O-199	O-221
AsnE2	GGC	No.4	O-210	O-251	O-199	O-222
AsnE2	CGC	No.1	O-210	O-252	O-199	O-219
AsnE2	CGC	No.2	O-210	O-253	O-199	O-220
AsnE2	CGC	No.3	O-210	O-254	O-199	O-221
AsnE2	CGC	No.4	O-210	O-255	O-199	O-222
AsnE2	GAC	No.0	O-210	O-256	O-199	O-219
Non-functional	CUA	No.0	O-210	O-257	O-199	O-219

Supplementary Table 12. List of mRNA sequences used in this study.

Name	Corresponding mDNA	mRNA sequence
mRNA12	mDNA12	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGGUCUACGACUACAAGGACGACGACGACAA GUAAGCUUCG-3'
mRNA13	mDNA13	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGACAAAUUUCGGUAUCCGUGCCUCUGUCAU GGAAUGGCAACUGCUCCAUCCAUACGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA14	mDNA14	5'-GGGCUUUAAUAAGGAGAAAAACAUGAAGAAGAAGCUCCAUAUGUUGGUCACAGAACUGUCUAU CGCGCAAAAUUUCGGUUGGGUGCGUGCCCCAGACUACAAGGACGACGACGACAAGUAAGCUUCG-3'
mRNA15	mDNA15	5'-GGGCUUUAAUAAGGAGAAAAACAUGGACUACAAGGACGACGACGACAAGCUCCUGGUCGCCUU CAUCAAGAAGAAGUAAGCUUCG-3'

Supplementary Table 13. List of oligonucleotides used for the preparation of mDNAs.

Name	Corresponding mRNA	Extension		1st PCR		2nd PCR	
		Forward primer	Reverse primer	Forward primer	Reverse primer	Forward primer	Reverse primer
mRNA12	mDNA12	O-258	O-259	O-199	O-260	-	-
mRNA13	mDNA13	O-261	O-262	O-258	O-263	O-199	O-260
mRNA14	mDNA14	O-264	O-265	O-258	O-263	O-199	O-260
mRNA15	mDNA15	O-266	O-267	O-199	O-268	-	-

Supplementary Table 14. Efficiencies of flexizymes-catalyzed acylation. As flexizymes recognize only the 3′-terminal CCA consensus sequence[79], the acylation efficiencies were similar independent of the tRNA specie. Therefore, the mean acylation values for each amino acid were used for the [^{Me}AA-tRNA] adjustment in experiments of Fig. 53, 54, 58–61. *The acylation efficiencies for Tyr and Ser were cited from a previous report[92].

Amino acid	tRNA	Anticodon	T-stem	Acylation (%)		S.D.	N
Phe	Phe	GAA	Native	52.7	±	3.7	3
	AsnE2	GAC	No.0	55.0	±	1.0	3
			No.1	51.8	±	0.9	3
			No.2	53.2	±	0.8	3
			No.3	53.9	±	0.3	3
			No.4	49.4	±	0.3	3
		GAA	No.2	51.7	±	1.2	3
		GAU	No.2	57.5	±	4.3	3
		CAU	No.2	53.1	±	1.6	3
			Mean	53.1	±	2.8	27
Tyr*	Asn	CUA	Native	34.4	±	6.4	
Ser*	Asn	CUA	Native	37.5	±	4.9	
^{Me}G	AsnE2	GAC	No.0	44.0	±	0.9	3
			No.1	46.2	±	0.4	3
			No.2	47.3	±	0.2	3
			No.3	44.6	±	3.2	3
			No.4	49.7	±	0.1	3
			Mean	46.4	±	2.4	15
^{Me}S	AsnE2	GAC	No.0	39.2	±	4.7	3
			No.1	39.9	±	2.9	3
			No.2	43.0	±	1.9	3
			No.3	39.8	±	1.7	3
			No.4	44.0	±	1.2	3
			Mean	41.2	±	3.1	15
^{Me}A	AsnE2	GAC	No.0	33.0	±	1.0	3
			No.1	38.4	±	1.4	3
			No.2	40.8	±	1.4	3
			No.3	36.5	±	1.9	3
			No.4	34.2	±	1.6	3
			Mean	36.6	±	3.2	15
^{Me}F	AsnE2	GAC	No.0	43.5	±	1.3	3
			No.1	39.8	±	0.9	3
			No.2	34.8	±	0.8	3
			No.3	47.5	±	3.2	3
			No.4	44.1	±	2.4	3
			Mean	41.9	±	4.8	15
^{Me}L	AsnE2	GAC	No.0	45.6	±	6.1	3
			No.1	46.5	±	0.6	3
			No.2	47.4	±	2.5	3
			No.3	44.9	±	2.7	3
			No.4	47.6	±	3.2	3
			Mean	46.4	±	3.2	15
^{Me} M	AsnE2	GAC	No.0	49.1	±	0.4	3
			No.1	51.8	±	0.4	3
			No.2	54.6	±	0.8	3
			No.3	52.0	±	1.5	3
			No.4	51.6	±	1.6	3
			Mean	51.8	±	2.0	15

Supplementary Table 14. (Continued)

Amino acid	tRNA	Anticodon	T-stem	Acylation (%)		S.D.	N
			No.0	41.2	±	0.3	3
			No.1	45.6	±	0.5	3
^{Me}T	AsnE2	GAC	No.2	41.3	±	0.7	3
			No.3	43.3	±	0.4	3
			No.4	43.7	±	1.0	3
			Mean	43.0	±	1.8	15
			No.0	32.3	±	0.8	3
			No.1	35.0	±	0.8	3
^{Me}Y	AsnE2	GAC	No.2	31.0	±	0.8	3
			No.3	33.2	±	0.8	3
			No.4	30.2	±	2.9	3
			Mean	32.4	±	2.1	15
			No.0	38.8	±	2.2	3
			No.1	40.6	±	4.3	3
^{Me}D	AsnE2	GAC	No.2	36.5	±	1.7	3
			No.3	36.5	±	3.2	3
			No.4	32.0	±	2.3	3
			Mean	36.9	±	3.8	15
			No.0	22.1	±	0.2	3
			No.1	23.1	±	0.3	3
^{Me}V	AsnE2	GAC	No.2	22.2	±	2.2	3
			No.3	17.8	±	2.2	3
			No.4	21.3	±	1.0	3
			Mean	21.3	±	2.3	15
			No.0	42.2	±	2.5	3
			No.1	47.3	±	1.8	3
^{Me}Nl	AsnE2	GAC	No.2	44.0	±	3.3	3
			No.3	43.0	±	1.0	3
			No.4	43.5	±	1.0	3
			Mean	44.0	±	2.5	15
			No.0	32.3	±	1.2	3
			No.1	32.3	±	1.8	3
^{Me}Nv	AsnE2	GAC	No.2	29.2	±	1.3	3
			No.3	29.3	±	0.4	3
			No.4	32.1	±	0.3	3
			Mean	31.1	±	1.8	15
			No.0	39.0	±	0.5	3
			No.1	36.7	±	1.0	3
^{Me}Ym	AsnE2	GAC	No.2	37.1	±	0.5	3
			No.3	36.6	±	0.9	3
			No.4	31.7	±	2.7	3
			Mean	36.2	±	2.8	15
			No.0	24.5	±	1.2	3
			No.1	33.7	±	2.4	3
^{Ac}K	AsnE2	GAC	No.2	29.4	±	1.1	3
			No.3	30.4	±	1.6	3
			No.4	32.1	±	0.8	3
			Mean	30.0	±	3.5	15

Supplementary Table 15. Conditions for translation reactions conducted in this study. Asterisk denotes that the concentrations of each ^{Me}AA-tRNA were adjusted to 10 μM based on the acylation efficiencies as described in section 3.5.6.

Figure	Each pre-charged $tRNA^{AsnE2}$ (μM)	$Mg(OAc)_2$ (mM)	Reaction time (min)	Reaction volume (μL)	Peptide template added into the translation mixture
47d	15	12	15	5.0 (^{Me}F, ^{Me}L) and 2.5 (^{Me}G, ^{Me}S, ^{Me}A, ^{Ac}K)	mDNA
47e	-	12	30	2.5	mDNA
50a,b	15	12	15	5.0	mDNA
50c,d	15	12	3 - 15	1.5	mDNA
51	15	12	15	5.0 (^{Me}F, ^{Me}L) and 2.5 (^{Me}G, ^{Me}S, ^{Me}A, ^{Ac}K)	mDNA
53e	50 (conventional) 10* (optimized)	12	20	2.5	mRNA
53f	50 (conventional) 10* (optimized)	12	20	5.0	mRNA
54d	10*	15	20	1.5	mRNA
54e	10*	15	20	10.0	mRNA
58c	15	12 - 20	15	1.0	mDNA
58d	15	12, 18	15	2.5	mDNA
59b,c	10* - 40*	12.0 - 19.5	10 - 120	1.0	mDNA
59d	10* - 20*	12.0 - 13.7	45	6.0	mDNA
60	10*	12.0	15	1.0	mDNA
61b	10*	12.0	10 - 60	1.0	mDNA
61c	10*	12.0	45	6.0	mDNA

Supplementary Table 16. The proteinogenic amino acids added in each experiment.

Amino acid	Fig. 47d	Fig. 47e	Fig. 50a,b	Fig. 50c,d	Fig. 51	Fig. 53,54	Fig. 58	Fig. 59-61
Ala	+	Not added	+	+	+	Not added	+	Not added
Arg	+	Not added	+	+	+	+	+	+
Asn	+	Not added	+	+	+	+	+	+
Asp	+	Not added	+	+	+	+	+	+
Cys	+	Not added	+	+	+	+	+	+
Gln	+	Not added	+	+	+	+	+	+
Glu	+	Not added	+	+	+	+	+	+
Gly	+	Not added	+	+	+	+	+	+
His	+	Not added	+	+	+	+	+	+
Ile	+	+ / -	+	Not added	+	Not added	+	Not added
Leu	+	+ / -	+	+	+	Not added	Not added	Not added
Lys	+	+	+	+	+	+	+	+
Met	+	+	+	+	+	+	+	+
Phe	+	Not added	+	+	+	Not added	+	Not added
Pro	+	Not added	+	+	+	+	+	+
Ser	+	Not added	+	+	+	+	+	+
Thr	+	Not added	+	+	+	+	+	+
Trp	+	Not added	+	+	+	+	+	+
Tyr	+	+	+	+	+	+	+	+
Val	Not added	+ / -	Not added	Not added	Not added	Not added	Not added	Not added

Chapter 4

General conclusion

Development of engineered translation systems such as the FIT system enabled us to synthesize nonstandard polypeptides containing a couple of npAAs[77,92,104,127-129]. In particular, the peptide screening technologies including the RaPID system have created an innovative field to discover potent peptide binders to drug target proteins, and successfully developed nonstandard peptide drugs with noncanonical scaffolds that mimic the characteristics of naturally occurring bioactive peptides[93,130-136]. In spite of the supposed practicability, the conventional methods still have a major limitations in the repertoire of building blocks (*i.e.* pAAs and npAAs) that can be used simultaneously[93]. Therefore, it was still challenging to synthesize highly modified nonstandard peptides like cyclosporine A containing multiple distinct npAAs and to develop peptide drugs with sufficient pharmacokinetic properties such as cell-membrane permeability and proteolytic resistance. As the Ph.D. degree research, I studied to develop two kinds of engineered translation systems enabling (1) expansion of the building block repertoire by utilizing multiple npAAs in addition to the 20 proteinogenic ones and (2) the synthesis of nonstandard peptides containing a variety of ^{Me}AAs.

In the study described in chapter 2, I have developed a novel method to expand the amino acid repertoire of ribosomal polypeptide synthesis via artificial division of codon boxes[153]. To realize this concept, I constructed a native tRNA-free FIT system that was supplied with *in vitro* transcripts of $tRNA_{SNN}$ (S = G or C), covering the assignment of the 20 pAAs to 31 NNS codons. In principle, this system can create 11 vacant codons without abandoning codons coding the pAAs, which can then be reassigned with npAAs. In this study, up to 3 codon boxes were divided simultaneously and reassigned with three distinct npAAs without abandoning any of 20 pAAs. The developed translation system was able to express various model peptides, including a macrocyclic *N*-methyl-peptide inhibitor of E6AP[93]. It should be noted that misreading of the reassigned codons with either cognate or noncognate pAAs did not occur, and thus the accuracy of translation was maintained. The proof-of-concept study on the artificial division of codon boxes demonstrated in this work opens a new opportunity for genetic code reprogramming. In particular, when this technology is coupled with mRNA display methods[125,126], such as the RaPID system, it enables us to express libraries of nonstandard macrocyclic peptides that are composed of 23 or more pAA and npAA building blocks to discover bioactive ligands against drug targets.

In the study described in chapter 3, I have developed a novel method to express nonstandard peptides containing a variety of ^{Me}AAs. To realize this, I first quantitatively demonstrated that ^{Me}AA-tRNAs cannot be delivered efficiently into ribosome by means of EF-Tu. Next, I developed a strategy to tune the weak EF-Tu affinity of ^{Me}AA-tRNAs to the mean value of proteinogenic aminoacyl-tRNAs by artificial replacement of their T-stem sequences. Tuning of the affinities between EF-Tu and each of multiple distinct ^{Me}AA-tRNAs successfully improved synthetic efficiency and accuracy of a model peptide containing six distinct ^{Me}AAs. Moreover, the developed translation system also achieved the synthesis of a highly *N*-methylated peptide containing as many as nine distinct ^{Me}AAs, which demonstrated the dramatic improvement in the incorporation efficiency and accuracy. Notably, the variety of 24 building blocks (15 pAAs and 9 ^{Me}AAs) achieved in this study has overcome the previous limitation of 23 variety (20 pAAs and 3 npAAs) [91,153]. The present engineered translation system would enable us to synthesize a highly *N*-methylated peptide library with sufficient accuracy and efficiency.

In this study, I have developed the engineered translation systems that allow for the synthesis of nonstandard polypeptides composed of expanded repertoire of pAAs and npAAs. These technologies will open a new opportunity for the synthesis of polypeptides containing a rich variety of npAAs. In particular, the integration of these methods with the RaPID system would enable us to express libraries of nonstandard macrocyclic peptides with highly modified scaffolds and to rapidly discover nonstandard peptide drugs. The technological improvement is expected to facilitate the discovery of practical nonstandard peptide drugs with improved binding potencies and pharmacokinetic properties in future.

References

1 Komarova, A. V., Tchufistova, L. S., Dreyfus, M. & Boni, I. V. AU-rich sequences within 5' untranslated leaders enhance translation and stabilize mRNA in Escherichia coli. *J. Bacteriol.* **187**, 1344-1349, doi:10.1128/JB.187.4.1344-1349.2005 (2005).

2 Milon, P. *et al.* The ribosome-bound initiation factor 2 recruits initiator tRNA to the 30S initiation complex. *EMBO Rep* **11**, 312-316, doi:10.1038/embor.2010.12 (2010).

3 Gold, L. Posttranscriptional regulatory mechanisms in Escherichia coli. *Annu. Rev. Biochem.* **57**, 199-233, doi:10.1146/annurev.bi.57.070188.001215 (1988).

4 Simonetti, A. *et al.* Structure of the 30S translation initiation complex. *Nature* **455**, 416-420, doi:10.1038/nature07192 (2008).

5 Milon, P., Konevega, A. L., Gualerzi, C. O. & Rodnina, M. V. Kinetic checkpoint at a late step in translation initiation. *Mol. Cell* **30**, 712-720, doi:10.1016/j.molcel.2008.04.014 (2008).

6 Vasil'eva, I. A. & Moor, N. A. Interaction of aminoacyl-tRNA synthetases with tRNA: general principles and distinguishing characteristics of the high-molecular-weight substrate recognition. *Biochemistry. Biokhimiia* **72**, 247-263 (2007).

7 Giege, R., Sissler, M. & Florentz, C. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Res.* **26**, 5017-5035 (1998).

8 Nissen, P. *et al.* Crystal structure of the ternary complex of Phe-tRNAPhe, EF-Tu, and a GTP analog. *Science* **270**, 1464-1472 (1995).

9 Diaconu, M. *et al.* Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell* **121**, 991-1004, doi:10.1016/j.cell.2005.04.015 (2005).

10 Kothe, U., Wieden, H. J., Mohr, D. & Rodnina, M. V. Interaction of helix D of elongation factor Tu with helices 4 and 5 of protein L7/12 on the ribosome. *J. Mol. Biol.* **336**, 1011-1021, doi:10.1016/j.jmb.2003.12.080 (2004).

11 Blanchard, S. C., Gonzalez, R. L., Kim, H. D., Chu, S. & Puglisi, J. D. tRNA selection and kinetic proofreading in translation. *Nat. Struct. Mol. Biol.* **11**, 1008-1014, doi:10.1038/nsmb831 (2004).

12 Schmeing, T. M. *et al.* The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* **326**, 688-694, doi:10.1126/science.1179700 (2009).

13 Ogle, J. M. *et al.* Recognition of cognate transfer RNA by the 30S ribosomal subunit. *Science* **292**, 897-902, doi:10.1126/science.1060612 (2001).

14 Rodnina, M. V., Fricke, R. & Wintermeyer, W. Transient conformational states of aminoacyl-tRNA during ribosome binding catalyzed by elongation factor Tu. *Biochemistry* **33**, 12267-12275 (1994).

15 Ogle, J. M., Murphy, F. V., Tarry, M. J. & Ramakrishnan, V. Selection of tRNA by the ribosome requires a transition from an open to a closed form. *Cell* **111**, 721-732 (2002).

16 Gromadski, K. B. & Rodnina, M. V. Kinetic determinants of high-fidelity tRNA discrimination on the ribosome. *Mol. Cell* **13**, 191-200 (2004).

17 Cochella, L. & Green, R. An active role for tRNA in decoding beyond codon:anticodon pairing. *Science* **308**, 1178-1180, doi:10.1126/science.1111408 (2005).

18 Voorhees, R. M., Schmeing, T. M., Kelley, A. C. & Ramakrishnan, V. The mechanism for activation of GTP hydrolysis on the ribosome. *Science* **330**, 835-838, doi:10.1126/science.1194460 (2010).

19 Berchtold, H. *et al.* Crystal structure of active elongation factor Tu reveals major domain rearrangements. *Nature* **365**, 126-132, doi:10.1038/365126a0 (1993).

20 Voorhees, R. M., Weixlbaumer, A., Loakes, D., Kelley, A. C. & 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, doi:10.1038/nsmb.1577 (2009).

21 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).

22 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, doi:10.1038/nsmb.2871 (2014).

23 Bhushan, S. *et al.* alpha-Helical nascent polypeptide chains visualized within distinct regions of the ribosomal exit tunnel. *Nat. Struct. Mol. Biol.* **17**, 313-317, doi:10.1038/nsmb.1756 (2010).

24 Yamamoto, H. *et al.* EF-G and EF4: translocation and back-translocation on the bacterial ribosome. *Nat. Rev. Microbiol.* **12**, 89-100, doi:10.1038/nrmicro3176 (2014).

25 Gao, Y. G. *et al.* The structure of the ribosome with elongation factor G

trapped in the posttranslocational state. *Science* **326**, 694-699, doi:10.1126/science.1179709 (2009).

26 Scolnick, E., Tompkins, R., Caskey, T. & Nirenberg, M. Release factors differing in specificity for terminator codons. *Proceedings of the National Academy of Sciences of the United States of America* **61**, 768-774 (1968).

27 Laurberg, M. *et al.* Structural basis for translation termination on the 70S ribosome. *Nature* **454**, 852-857, doi:10.1038/nature07115 (2008).

28 Weixlbaumer, A. *et al.* Insights into translational termination from the structure of RF2 bound to the ribosome. *Science* **322**, 953-956, doi:10.1126/science.1164840 (2008).

29 Grentzmann, G. *et al.* Release factor RF-3 GTPase activity acts in disassembly of the ribosome termination complex. *RNA* **4**, 973-983 (1998).

30 Hirashima, A. & Kaji, A. Role of elongation factor G and a protein factor on the release of ribosomes from messenger ribonucleic acid. *J. Biol. Chem.* **248**, 7580-7587 (1973).

31 Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981-984, doi:10.1126/science.1202692 (2011).

32 Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751-755, doi:10.1038/90802 (2001).

33 Komine, Y., Adachi, T., Inokuchi, H. & Ozeki, H. Genomic organization and physical mapping of the transfer RNA genes in *Escherichia coli* K12. *J. Mol. Biol.* **212**, 579-598, doi:10.1016/0022-2836(90)90224-A (1990).

34 Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S. H. Crystal structure of yeast phenylalanine transfer RNA. I. Crystallographic refinement. *J. Mol. Biol.* **123**, 607-630 (1978).

35 Gustilo, E. M., Vendeix, F. A. & Agris, P. F. tRNA's modifications bring order to gene expression. *Curr. Opin. Microbiol.* **11**, 134-140, doi:10.1016/j.mib.2008.02.003 (2008).

36 Tamura, K., Himeno, H., Asahara, H., Hasegawa, T. & Shimizu, M. *In vitro* study of *E.coli* $tRNA^{Arg}$ and $tRNA^{Lys}$ identity elements. *Nucleic Acids Res.* **20**, 2335-2339 (1992).

37 Nureki, O. *et al.* Molecular recognition of the identity-determinant set of isoleucine transfer RNA from *Escherichia coli*. *J. Mol. Biol.* **236**, 710-724, doi:10.1006/jmbi.1994.1184 (1994).

38 Sylvers, L. A., Rogers, K. C., Shimizu, M., Ohtsuka, E. & Soll, D. A

2-thiouridine derivative in $tRNA^{Glu}$ is a positive determinant for aminoacylation by *Escherichia coli* glutamyl-tRNA synthetase. *Biochemistry* **32**, 3836-3841 (1993).

39 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).

40 Semenkov, Y. P., Rodnina, M. V. & Wintermeyer, W. Energetic contribution of tRNA hybrid state formation to translocation catalysis on the ribosome. *Nat. Struct. Biol.* **7**, 1027-1031, doi:10.1038/80938 (2000).

41 Sanderson, L. E. & Uhlenbeck, O. C. Directed mutagenesis identifies amino acid residues involved in elongation factor Tu binding to yeast Phe-tRNAPhe. *J. Mol. Biol.* **368**, 119-130, doi:10.1016/j.jmb.2007.01.075 (2007).

42 Schrader, J. M., Chapman, S. J. & Uhlenbeck, O. C. Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis. *J. Mol. Biol.* **386**, 1255-1264 (2009).

43 Dale, T., Sanderson, L. E. & Uhlenbeck, O. C. The affinity of elongation factor Tu for an aminoacyl-tRNA is modulated by the esterified amino acid. *Biochemistry* **43**, 6159-6166, doi:10.1021/bi036290o (2004).

44 Asahara, H. & Uhlenbeck, O. C. The tRNA specificity of Thermus thermophilus EF-Tu. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3499-3504, doi:10.1073/pnas.052028599 (2002).

45 LaRiviere, F. J., Wolfson, A. D. & Uhlenbeck, O. C. Uniform binding of aminoacyl-tRNAs to elongation factor Tu by thermodynamic compensation. *Science* **294**, 165-168, doi:10.1126/science.1064242 (2001).

46 Louie, A. & Jurnak, F. Kinetic studies of Escherichia coli elongation factor Tu-guanosine 5'-triphosphate-aminoacyl-tRNA complexes. *Biochemistry* **24**, 6433-6439 (1985).

47 Rudinger, J., Hillenbrandt, R., Sprinzl, M. & Giege, R. Antideterminants present in minihelix(Sec) hinder its recognition by prokaryotic elongation factor Tu. *EMBO J.* **15**, 650-657 (1996).

48 Dale, T. & Uhlenbeck, O. C. Amino acid specificity in translation. *Trends Biochem. Sci.* **30**, 659-665, doi:10.1016/j.tibs.2005.10.006 (2005).

49 Mittelstaet, J., Konevega, A. L. & Rodnina, M. V. A kinetic safety gate controlling the delivery of unnatural amino acids to the ribosome. *J. Am. Chem. Soc.* **135**, 17031-17038, doi:10.1021/ja407511q (2013).

50 Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M. & Yusupova, G. A new

understanding of the decoding principle on the ribosome. *Nature* **484**, 256-259, doi:10.1038/nature10913 (2012).

51 Crick, F. H. Codon—anticodon pairing: the wobble hypothesis. *J. Mol. Biol.* **19**, 548-555 (1966).

52 Agris, P. F. Wobble position modified nucleosides evolved to select transfer RNA codon recognition: a modified-wobble hypothesis. *Biochimie* **73**, 1345-1349 (1991).

53 Rozov, A., Demeshkina, N., Westhof, E., Yusupov, M. & Yusupova, G. New Structural Insights into Translational Miscoding. *Trends Biochem. Sci.* **41**, 798-814, doi:10.1016/j.tibs.2016.06.001 (2016).

54 Grosjean, H. & Westhof, E. An integrated, structure- and energy-based view of the genetic code. *Nucleic Acids Res.* **44**, 8020-8040, doi:10.1093/nar/gkw608 (2016).

55 Nasvall, S. J., Chen, P. & Bjork, G. R. The wobble hypothesis revisited: Uridine-5-oxyacetic acid is critical for reading of G-ending codons. *Rna-a Publication of the Rna Society* **13**, 2151-2164, doi:10.1261/rna.731007 (2007).

56 Yarus, M. Translational efficiency of transfer RNA's: uses of an extended anticodon. *Science* **218**, 646-652 (1982).

57 Sundaram, M., Durant, P. C. & Davis, D. R. Hypermodified nucleosides in the anticodon of tRNALys stabilize a canonical U-turn structure. *Biochemistry* **39**, 12575-12584 (2000).

58 Agris, P. F., Vendeix, F. A. & Graham, W. D. tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* **366**, 1-13, doi:10.1016/j.jmb.2006.11.046 (2007).

59 Forster, A. C. *et al.* Programming peptidomimetic syntheses by translating genetic codes designed *de novo*. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 6353-6357, doi:10.1073/pnas.1132122100 (2003).

60 Josephson, K., Hartman, M. C. & Szostak, J. W. Ribosomal synthesis of unnatural peptides. *J. Am. Chem. Soc.* **127**, 11727-11735, doi:10.1021/ja0515809 (2005).

61 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, doi:10.1038/nchembio.1549 (2014).

62 Shepotinovskaya, I. & Uhlenbeck, O. C. tRNA residues evolved to promote translational accuracy. *RNA* **19**, 510-516, doi:10.1261/rna.036038.112 (2013).

63 Cui, Z., Stein, V., Tnimov, Z., Mureev, S. & Alexandrov, K. Semisynthetic tRNA complement mediates *in vitro* protein synthesis. *J. Am. Chem. Soc.* **137**, 4404-4413, doi:10.1021/ja5131963 (2015).

64 Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413-444, doi:10.1146/annurev.biochem.052308.105824 (2010).

65 Davis, L. & Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.* **13**, 168-182, doi:10.1038/nrm3286 (2012).

66 Chin, J. W. Expanding and reprogramming the genetic code of cells and animals. *Annu. Rev. Biochem.* **83**, 379-408, doi:10.1146/annurev-biochem-060713-035737 (2014).

67 Passioura, T., Katoh, T., Goto, Y. & Suga, H. Selection-based discovery of druglike macrocyclic peptides. *Annu. Rev. Biochem.* **83**, 727-752, doi:10.1146/annurev-biochem-060713-035456 (2014).

68 Merryman, C. & Green, R. Transformation of aminoacyl tRNAs for the in vitro selection of "drug-like" molecules. *Chem. Biol.* **11**, 575-582, doi:10.1016/j.chembiol.2004.03.009 (2004).

69 Subtelny, A. O., Hartman, M. C. & Szostak, J. W. Ribosomal synthesis of N-methyl peptides. *J. Am. Chem. Soc.* **130**, 6131-6136, doi:10.1021/ja710016v (2008).

70 Hartman, M. C., Josephson, K. & Szostak, J. W. Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4356-4361, doi:10.1073/pnas.0509219103 (2006).

71 Hartman, M. C., Josephson, K., Lin, C. W. & Szostak, J. W. An expanded set of amino acid analogs for the ribosomal translation of unnatural peptides. *PLoS One* **2**, e972, doi:10.1371/journal.pone.0000972 (2007).

72 Heckler, T. G. *et al.* T4 RNA ligase mediated preparation of novel "chemically misacylated" tRNAPheS. *Biochemistry* **23**, 1468-1473 (1984).

73 Robertson, S. A., Ellman, J. A. & Schultz, P. G. A General and Efficient Route for Chemical Aminoacylation of Transfer-Rnas. *J. Am. Chem. Soc.* **113**, 2722-2729, doi:DOI 10.1021/ja00007a055 (1991).

74 Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of *Escherichia coli*. *Science* **292**, 498-500, doi:10.1126/science.1060077 (2001).

75 Hoesl, M. G. & Budisa, N. Recent advances in genetic code engineering in

Escherichia coli. *Curr. Opin. Biotechnol.* **23**, 751-757, doi:10.1016/j.copbio.2011.12.027 (2012).

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

77 Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat Protoc* **6**, 779-790, doi:10.1038/nprot.2011.331 (2011).

78 Polycarpo, C. R. *et al.* Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *FEBS Lett.* **580**, 6695-6700, doi:10.1016/j.febslet.2006.11.028 (2006).

79 Xiao, H., Murakami, H., Suga, H. & Ferre-D'Amare, A. R. Structural basis of specific tRNA aminoacylation by a small *in vitro* selected ribozyme. *Nature* **454**, 358-361, doi:10.1038/nature07033 (2008).

80 Rogers, J. M. & Suga, H. Discovering functional, non-proteinogenic amino acid containing, peptides using genetic code reprogramming. *Org Biomol Chem* **13**, 9353-9363, doi:10.1039/c5ob01336d (2015).

81 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, doi:10.1016/j.bmcl.2009.03.114 (2009).

82 Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C. & Schultz, P. G. A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **244**, 182-188 (1989).

83 Wang, K., 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, doi:10.1038/nbt1314 (2007).

84 Mukai, T. *et al.* Codon reassignment in the *Escherichia coli* genetic code. *Nucleic Acids Res.* **38**, 8188-8195, doi:10.1093/nar/gkq707 (2010).

85 Mukai, T. *et al.* Genetic-code evolution for protein synthesis with non-natural amino acids. *Biochem. Biophys. Res. Commun.* **411**, 757-761, doi:10.1016/j.bbrc.2011.07.020 (2011).

86 Johnson, D. B. F. *et al.* RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. *Nat. Chem. Biol.* **7**, 779-786, doi:10.1038/nchembio.657 (2011).

87 Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357-360, doi:10.1126/science.1241459 (2013).

88 Hohsaka, T., Ashizuka, Y., Murakami, H. and Sisido, M. Incorporation of

nonnatural amino acids into streptavidin trhrough *in vitro* frame-shift suppression. *J. Am. Chem. Soc.* **118**, 9778-9779 (1996).

89 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, doi:10.1006/jmbi.2001.4518 (2001).

90 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, doi:10.1038/nature08817 (2010).

91 Ohtsuki, T., Manabe, T. & Sisido, M. Multiple incorporation of non-natural amino acids into a single protein using tRNAs with non-standard structures. *FEBS Lett.* **579**, 6769-6774, doi:10.1016/j.febslet.2005.11.010 (2005).

92 Murakami, H., Ohta, A., Ashigai, H. & Suga, H. A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nat. Methods* **3**, 357-359, doi:10.1038/nmeth877 (2006).

93 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, doi:10.1016/j.chembiol.2011.09.013 (2011).

94 Ellman, J., Mendel, D., Anthony-Cahill, S., Noren, C. J. & Schultz, P. G. Biosynthetic method for introducing unnatural amino acids site-specifically into proteins. *Methods Enzymol.* **202**, 301-336 (1991).

95 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, doi:10.1021/ja075557u (2007).

96 Park, H. S. *et al.* Expanding the genetic code of *Escherichia coli* with phosphoserine. *Science* **333**, 1151-1154, doi:10.1126/science.1207203 (2011).

97 Lee, S. *et al.* A facile strategy for selective incorporation of phosphoserine into histones. *Angew. Chem. Int. Ed. Engl.* **52**, 5771-5775, doi:10.1002/anie.201300531 (2013).

98 Forster, A. C. Low modularity of aminoacyl-tRNA substrates in polymerization by the ribosome. *Nucleic Acids Res.* **37**, 3747-3755, doi:10.1093/nar/gkp240 (2009).

99 Guo, J., Melancon, C. E., 3rd, Lee, H. S., Groff, D. & Schultz, P. G. Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids. *Angew. Chem. Int. Ed. Engl.* **48**,

9148-9151, doi:10.1002/anie.200904035 (2009).

100 Dedkova, L. M., Fahmi, N. E., Golovine, S. Y. & Hecht, S. M. Enhanced D-amino acid incorporation into protein by modified ribosomes. *J. Am. Chem. Soc.* **125**, 6616-6617, doi:10.1021/ja035141q (2003).

101 Maini, R. *et al.* Protein Synthesis with Ribosomes Selected for the Incorporation of beta-Amino Acids. *Biochemistry* **54**, 3694-3706, doi:10.1021/acs.biochem.5b00389 (2015).

102 Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935-1942, doi:10.1126/science.1131127 (2006).

103 Kawakami, T., Sasaki, T., Reid, P. C. & Murakami, H. Incorporation of electrically charged N-alkyl amino acids into ribosomally synthesized peptides via post-translational conversion. *Chem Sci* **5**, 887-893, doi:10.1039/c3sc52744a (2014).

104 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, doi:10.1016/j.chembiol.2007.12.008 (2008).

105 Elliott, T. S. *et al.* Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat. Biotechnol.* **32**, 465-472, doi:10.1038/nbt.2860 (2014).

106 Verdine, G. L. & Walensky, L. D. The challenge of drugging undruggable targets in cancer: lessons learned from targeting BCL-2 family members. *Clin. Cancer. Res.* **13**, 7264-7270, doi:10.1158/1078-0432.CCR-07-2184 (2007).

107 Swinney, D. C. & Anthony, J. How were new medicines discovered? *Nature Reviews Drug Discovery* **10**, 507-519, doi:10.1038/nrd3480 (2011).

108 Hopkins, A. L. & Groom, C. R. The druggable genome. *Nat. Rev. Drug Discov.* **1**, 727-730, doi:10.1038/nrd892 (2002).

109 Russ, A. P. & Lampel, S. The druggable genome: an update. *Drug Discov. Today* **10**, 1607-1610, doi:10.1016/S1359-6446(05)03666-4 (2005).

110 Reichert, J. M., Rosensweig, C. J., Faden, L. B. & Dewitz, M. C. Monoclonal antibody successes in the clinic. *Nat. Biotechnol.* **23**, 1073-1078, doi:10.1038/nbt0905-1073 (2005).

111 Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Methods* **44**, 235-249 (2000).

112 Driggers, E. M., Hale, S. P., Lee, J. & Terrett, N. K. The exploration of macrocycles for drug discovery--an underexploited structural class. *Nat. Rev. Drug Discov.* **7**, 608-624, doi:10.1038/nrd2590 (2008).

113 Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nat. Rev. Drug Discov.* **4**, 206-220, doi:10.1038/nrd1657 (2005).

114 Conradi, R. A., Hilgers, A. R., Ho, N. F. & Burton, P. S. The influence of peptide structure on transport across Caco-2 cells. II. Peptide bond modification which results in improved permeability. *Pharm. Res.* **9**, 435-439 (1992).

115 Haviv, F. *et al.* Effect of N-methyl substitution of the peptide bonds in luteinizing hormone-releasing hormone agonists. *J. Med. Chem.* **36**, 363-369 (1993).

116 Chikhale, E. G., Ng, K. Y., Burton, P. S. & Borchardt, R. T. Hydrogen bonding potential as a determinant of the in vitro and in situ blood-brain barrier permeability of peptides. *Pharm. Res.* **11**, 412-419 (1994).

117 Miller, S. M. *et al.* Comparison of the Proteolytic Susceptibilities of Homologous L-Amino-Acid, D-Amino-Acid, and N-Substituted Glycine Peptide and Peptoid Oligomers. *Drug Dev. Res.* **35**, 20-32, doi:DOI 10.1002/ddr.430350105 (1995).

118 Sheehy, B. A. *et al.* Transport of model peptides across Ascaris suum cuticle. *Mol. Biochem. Parasitol.* **105**, 39-49 (2000).

119 Frankel, A., Millward, S. W. & Roberts, R. W. Encodamers: unnatural peptide oligomers encoded in RNA. *Chem. Biol.* **10**, 1043-1050 (2003).

120 Rezai, T., Yu, B., Millhauser, G. L., Jacobson, M. P. & Lokey, R. S. Testing the conformational hypothesis of passive membrane permeability using synthetic cyclic peptide diastereomers. *J. Am. Chem. Soc.* **128**, 2510-2511, doi:10.1021/ja0563455 (2006).

121 Biron, E. *et al.* Improving oral bioavailability of peptides by multiple N-methylation: somatostatin analogues. *Angew. Chem. Int. Ed. Engl.* **47**, 2595-2599, doi:10.1002/anie.200705797 (2008).

122 Ovadia, O. *et al.* The effect of multiple N-methylation on intestinal permeability of cyclic hexapeptides. *Mol. Pharm.* **8**, 479-487, doi:10.1021/mp1003306 (2011).

123 White, T. R. *et al.* On-resin N-methylation of cyclic peptides for discovery of orally bioavailable scaffolds. *Nat. Chem. Biol.* **7**, 810-817, doi:10.1038/nchembio.664 (2011).

124 Beck, J. G. *et al.* Intestinal permeability of cyclic peptides: common key backbone motifs identified. *J. Am. Chem. Soc.* **134**, 12125-12133, doi:10.1021/ja303200d (2012).

125 Nemoto, N., Miyamoto-Sato, E., Husimi, Y. & Yanagawa, H. In vitro virus: bonding of mRNA bearing puromycin at the 3'-terminal end to the C-terminal end of its encoded protein on the ribosome in vitro. *FEBS Lett.* **414**, 405-408 (1997).

126 Roberts, R. W. & Szostak, J. W. RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12297-12302 (1997).

127 Ohta, A., Murakami, H., Higashimura, E. & Suga, H. Synthesis of polyester by means of genetic code reprogramming. *Chem. Biol.* **14**, 1315-1322, doi:10.1016/j.chembiol.2007.10.015 (2007).

128 Goto, Y. *et al.* Reprogramming the translation initiation for the synthesis of physiologically stable cyclic peptides. *ACS chemical biology* **3**, 120-129, doi:10.1021/cb700233t (2008).

129 Goto, Y., Murakami, H. & Suga, H. Initiating translation with *D*-amino acids. *RNA* **14**, 1390-1398, doi:10.1261/rna.1020708 (2008).

130 Hayashi, Y., Morimoto, J. & Suga, H. *In vitro* selection of anti-Akt2 thioether-macrocyclic peptides leading to isoform-selective inhibitors. *ACS chemical biology* **7**, 607-613, doi:10.1021/cb200388k (2012).

131 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, doi:10.1002/anie.201108118 (2012).

132 Yamagata, K. *et al.* Structural basis for potent inhibition of SIRT2 deacetylase by a macrocyclic peptide inducing dynamic structural change. *Structure* **22**, 345-352, doi:10.1016/j.str.2013.12.001 (2014).

133 Tanaka, Y. *et al.* Structural basis for the drug extrusion mechanism by a MATE multidrug transporter. *Nature* **496**, 247-251, doi:10.1038/nature12014 (2013).

134 Terasaka, N. & Suga, H. Flexizymes-facilitated Genetic Code Reprogramming Leading to the Discovery of Drug-like Peptides. *Chem. Lett.* **43**, 11-19, doi:10.1246/cl.130910 (2014).

135 Hipolito, C. J., Tanaka, Y., Katoh, T., Nureki, O. & Suga, H. A Macrocyclic Peptide that Serves as a Cocrystallization Ligand and Inhibits the Function of a MATE Family Transporter. *Molecules* **18**, 10514-10530, doi:10.3390/molecules180910514 (2013).

136 Ito, K. *et al.* Artificial human Met agonists based on macrocycle scaffolds. *Nat*

Commun **6**, 6373, doi:10.1038/ncomms7373 (2015).

137 Craik, D. J., Fairlie, D. P., Liras, S. & Price, D. The future of peptide-based drugs. *Chem Biol Drug Des* **81**, 136-147, doi:10.1111/cbdd.12055 (2013).

138 Heinis, C. & Winter, G. Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol.* **26**, 89-98, doi:10.1016/j.cbpa.2015.02.008 (2015).

139 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108-160, doi:10.1039/c2np20085f (2013).

140 Horswill, A. R. & Benkovic, S. J. Cyclic peptides, a chemical genetics tool for biologists. *Cell Cycle* **4**, 552-555 (2005).

141 Murphy, F. V. t. & Ramakrishnan, V. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nat. Struct. Mol. Biol.* **11**, 1251-1252, doi:10.1038/nsmb866 (2004).

142 Schlippe, Y. V., Hartman, M. C., Josephson, K. & Szostak, J. W. *In vitro* selection of highly modified cyclic peptides that act as tight binding inhibitors. *J. Am. Chem. Soc.* **134**, 10469-10477, doi:10.1021/ja301017y (2012).

143 Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783-8798 (1987).

144 Pape, T., Wintermeyer, W. & Rodnina, M. Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* **18**, 3800-3807, doi:10.1093/emboj/18.13.3800 (1999).

145 Handschumacher, R. E., Harding, M. W., Rice, J., Drugge, R. J. & Speicher, D. W. Cyclophilin: a specific cytosolic binding protein for cyclosporin A. *Science* **226**, 544-547 (1984).

146 Takahashi, N., Hayano, T. & Suzuki, M. Peptidyl-Prolyl Cis-Trans Isomerase Is the Cyclosporin-a-Binding Protein Cyclophilin. *Nature* **337**, 473-475, doi:DOI 10.1038/337473a0 (1989).

147 Walsh, C. T., Zydowsky, L. D. & McKeon, F. D. Cyclosporin A, the cyclophilin class of peptidylprolyl isomerases, and blockade of T cell signal transduction. *J. Biol. Chem.* **267**, 13115-13118 (1992).

148 Altschuh, D., Vix, O., Rees, B. & Thierry, J. C. A conformation of cyclosporin A in aqueous environment revealed by the X-ray structure of a cyclosporin-Fab complex. *Science* **256**, 92-94 (1992).

149 Pavlov, M. Y. *et al.* Slow peptide bond formation by proline and other

N-alkylamino acids in translation. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 50-54, doi:10.1073/pnas.0809211106 (2009).

150 Wang, J., Kwiatkowski, M., Pavlov, M. Y., Ehrenberg, M. & Forster, A. C. Peptide formation by N-methyl amino acids in translation is hastened by higher pH and tRNA(Pro). *ACS chemical biology* **9**, 1303-1311, doi:10.1021/cb500036a (2014).

151 Ieong, K. W., Pavlov, M. Y., Kwiatkowski, M., Ehrenberg, M. & Forster, A. C. A tRNA body with high affinity for EF-Tu hastens ribosomal incorporation of unnatural amino acids. *RNA* **20**, 632-643, doi:10.1261/rna.042234.113 (2014).

152 Schrader, J. M., Chapman, S. J. & Uhlenbeck, O. C. Tuning the affinity of aminoacyl-tRNA to elongation factor Tu for optimal decoding. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 5215-5220, doi:10.1073/pnas.1102128108 (2011).

153 Iwane, Y. *et al.* Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat Chem* **8**, 317-325, doi:10.1038/nchem.2446 (2016).

154 Pleiss, J. A. & Uhlenbeck, O. C. Identification of thermodynamically relevant interactions between EF-Tu and backbone elements of tRNA. *J. Mol. Biol.* **308**, 895-905, doi:10.1006/jmbi.2001.4612 (2001).

List of achievements

Publication

1. Y. Iwane, A. Hitomi, Y. Goto, T. Katoh, H. Murakami, and H. Suga. Accurate assignment of a nonproteinogenic amino acid and valine to the valine GUN codon box. *Peptide Science*, **2015**, 141–142.
2. N. Terasaka, Y. Iwane, A. S. Geiermann, Y. Goto, and H. Suga. Recent developments of engineered translational machineries for the incorporation of non-canonical amino acids into polypeptides. *Int. J. Mol. Sci.*, **2015**, 16, 6513–6531.
3. Y. Iwane, A. Hitomi, H. Murakami, T. Katoh, Y. Goto, and H. Suga. Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat. Chem.*, **2016**, 8, 317–325.
4. Y. Iwane, T. Katoh, Y. Goto, and H. Suga. Artificial division of codon boxes for expansion of the amino acid repertoire of ribosomal polypeptide synthesis. *submitted.*

Oral presentation

1. Y. Iwane. Tuning of binding affinity between EF-Tu and aminoacyl-tRNA can improve the efficiency and accuracy of *N*-methyl amino acid incorporation into peptides. The 22nd ZESTY Network Seminar, Tokyo, Nov. 2015.

Poster presentation

1. Y. Iwane, A. Hitomi, Y. Goto, T. Katoh, H. Murakami, and H. Suga. Encoding multiple nonproteinogenic amino acids besides proteinogenic ones by artificial codon box division. 第 51 回ペプチド討論会，徳島県，2014 年 10 月.
2. Y. Iwane, A. Hitomi, Y. Goto, T. Katoh, H. Murakami, and H. Suga. Artificial codon box division to encode multiple non-proteinogenic amino acids besides 20 proteinogenic ones. 25th tRNA Conference, Kyllini, Greece, Sep. 2014.

3. 岩根由彦，人見梓，後藤佑樹，加藤敬行，村上裕，菅裕明.コドンボックス人工分割による翻訳基質アミノ酸の拡大. 「細胞を創る」研究会 7.0，東京都，2014 年 11 月
4. Y. Iwane, A. Hitomi, H. Murakami, T. Katoh, Y. Goto, and H. Suga. Artificial division of codon boxes to encode nonproteinogenic amino acids along with 20 proteinogenic ones. Pacifichem 2015, Hawaii, USA, Dec. 2015.
5. Y. Iwane, A. Hitomi, H. Murakami, T. Katoh, Y. Goto, and H. Suga. Artificial division of codon boxes to encode non-proteinogenic amino acids besides 20 proteinogenic ones. 第 3 回バイオ関連化学シンポジウム若手フォーラム，熊本県，2015 年 9 月
6. Y. Iwane, A. Hitomi, H. Murakami, T. Katoh, Y. Goto, and H. Suga. Artificial division of codon boxes to encode multiple nonproteinogenic amino acids along with 20 proteinogenic ones. 第 9 回バイオ関連化学シンポジウム，熊本県，2015 年 9 月
7. Y. Iwane, A. Hitomi, H. Murakami, T. Katoh, Y. Goto, and H. Suga. Encoding 23 amino acids in the genetic code by artificial division of codon boxes. 「細胞を創る」研究会 8.0，大阪府，2015 年 11 月
8. 岩根由彦，人見梓，村上裕，加藤敬行，後藤佑樹，菅裕明. Artificial division of codon boxes to expand amino acid repertoire in ribosomal polypeptide synthesis. 創薬懇話会 2016，長野県，2016 年 6 月.

Awards and fellowships

1. JPS ポスター賞. 第 51 回ペプチド討論会（2014 年 10 月）
2. 優秀ポスター賞. 創薬懇話会 2016 (2016 年 6 月)
3. フジサンケイビジネスアイ主催第 30 回先端技術大賞にて最優秀賞（文部科学大臣賞）受賞. 岩根由彦. コドンボックス人工分割法の開発～リボソーム翻訳における基質アミノ酸の種類拡大～（2016 年 7 月）
4. 日本学術振興会特別研究員 DC1 (2014–2016 年度)

Acknowledgement

This research has been conducted under the supervision of Professor Hiroaki Suga. I appreciate him for his kind guidance, suggestions, education, and encouragement throughout this work. I am deeply grateful to Associate Professor Naokazu Kano, Associate Professor Yuki Goto, Assistant professor Takayuki Katoh, and Project Assistant Professor Toby Passioura and all laboratory members for their kind advices and discussion.

I thank T. Suzuki and K. Miyauchi for a gift of isolated native tRNAs. I also thank M. E. Harris and M. Ohuchi for a gift of plasmids coding E. coli M1 RNA and C5 protein and their products. This research was supported by the Grants-in-Aid for JSPS Fellows (26-9576). This research was supported by the Japan Science and Technology Agency (JST) Core Research for Evolutional Science and Technology (CREST) of Molecular Technologies to Professor Hiroaki Suga, Japan Society for the Promotion of Science (JSPS) Grant-in-Aid for Young Scientists (B) to Associate Professor Yuki Goto, and Grants-in-Aid for JSPS Fellows (26-9576) to Yoshihiko Iwane.