Doctoral Dissertation

博士論文

Profiling of nascent peptide sequences that regulate peptidyl-tRNA drop-off and investigation on expression of proteins lacking the N-terminus due to the drop-off

(ペプチジル-tRNA 脱落を制御するペプチド新生鎖配列のプロファイリング

と脱落に伴う N 末端欠損タンパク質発現の検証)

A Dissertation Submitted for the Degree of Doctor of Philosophy
December 2019

令和元年 12 月博士（理学）申請

Department of Chemistry, Graduate School of Science,
The University of Tokyo

東京大学大学院理学系研究科

化学専攻

Kenya Tajima

田島　研也

Abstract

Introduction

Ribosomal translation elongates polypeptide chains, in which elongation factor G (EF-G) mediates translocation of the A-site peptidyl-tRNA and the P-site deacyl-tRNA into P site and A site, respectively. Although a short peptidyl-tRNA accidentally drops off from the translating ribosome, the actual mechanisms and molecules that induce peptidyl-tRNA drop-off have remained elusive. It is believed that certain mRNA contexts or translation factors, such as EF-G, are involved in the drop-off. Ribosome stalling is one of the plausible causes of drop-off. Certain nascent peptides induce ribosome stalling via formation of inactive state for peptidyl transfer by interaction with the ribosome exit tunnel. A stretch of proline residues in nascent peptides also causes a ribosome stalling by its poor peptidyl donor and acceptor ability. The strength of stalling depends on the number of prolines and neighboring amino acids. In my master course, utilizing ribosomes stalled by four kinds of peptidyl-diprolyl-tRNA, I revealed that peptidyl-tRNA dropped off from the ribosome and C-terminal peptides with N-terminal truncation were generated due to translation restart from the middle of open reading frames. The drop-off was stimulated in dose-dependent manner of EF-G and regulated by nature of the nascent peptide sequences prior to proline stretches.

In my PhD course, I comprehensively profiled 8420 kinds of nascent peptide sequences against regulation of drop-off of peptidyl-polyprolyl-tRNA utilizing mRNA display-based approach. It was revealed that the peptidyl-tRNA drop-off depended on polarity of amino acids preceding the proline stretch, which was suppressed by elongation factor P (EF-P). Finally, I expressed and purified *E. coli* proteins and putative proteins containing consecutive prolines at proximal to their N-termini, which were sequenced by LC-ESI MS/MS after protease digestion. In the case of putative protein YhhM, peptide fragments derived from both of full-length YhhM and YhhM lacking its N-terminal amino acids were identified. Co-expression of EF-P reduced the ratio of fragments derived from YhhM lacking its N-terminal amino acids. My result demonstrated that regulation of peptidyl-tRNA drop-off was attributed to the side chain polarity of its nascent peptide in addition of EF-G concentration and suggested that the drop-off and generation of truncated peptides occurred on the stalled ribosome in *in vivo* protein synthesis.

Results and Discussion

Comprehensive profiling of nascent peptide sequences on peptidyl-tRNA drop-off

In order to comprehensively analyze the effect of nascent peptide sequences prior to the proline stretch on peptidyl-tRNA drop-off, I performed saturation mutagenesis on from one to three amino acid residues in nascent peptides utilizing Random non-standard Peptide Integrated Discovery (RaPID) system. I prepared an mRNA library to express 8420 kinds of peptide sequences with N-terminal biotinyl-phenylalanine, installed by genetic code reprogramming for pull-down, followed by one to three proteinogenic amino acids (X_4, X_3X_4, $X_2X_3X_4$), three consecutive prolines, and a linker peptide. Nascent peptides were expressed with 0.03, 0.26, or 10 μM of EF-G in the absence or presence of EF-P, among which mRNAs translated into full-length peptides were selectively recovered and analyzed by next generation sequencing. Peptidyl-tRNA drop-off would generate C-terminal peptide-mRNA-cDNA conjugates, which were not recovered. Sequencing data were converted to frequencies of a nascent peptide sequence i (N_i) in the initial library ($F_{Ni,\ Initial}$) or in libraries pulled down in the absence or presence of EF-P ($F_{Ni,\ Pull\text{-}down,\ EF\text{-}P\text{-}/+}$), and interpreted as enrichment of the N_i, indicating frequencies of peptidyl-tRNA drop-off ($E_{Ni,\ EF\text{-}P\text{-}/+}$). Suppression strength of drop-off by EF-P ($W_{Ni} = E_{Ni,\ EF\text{-}P+} - E_{Ni,\ EF\text{-}P-}$) was also calculated. $E_{Ni,\ EF\text{-}P\text{-}/+}$ and W_{Ni} were

sorted into four groups in terms of the side chain polarity [Polar (DERKHQN), Small (STGCA), Hydrophobic (VILMFYW), and Proline (P)] at X_2, X_3, and X_4 positions respectively and plotted as histograms and heatmaps.

$E_{Ni,\ EF\text{-}P-}$ and $E_{Ni,\ EF\text{-}P+}$ of N_i having prolines at X_2, X_3, or X_4 showed less than zero, indicating that additional prolines in the random region further induced peptidyl-tRNA drop-off. Focusing on residues except proline, at X_2 and X_3, there was no significant difference of drop-off frequency depending on amino acid residues. However, at X_4, polar residues showed significantly lower distribution in $E_{Ni,\ EF\text{-}P-}$, inducing peptidyl-tRNA drop-off. $E_{Ni,\ EF\text{-}P+}$ at X_4 did not show polarity-dependent difference, resulted in higher distribution of W_{Ni} of polar residues at X_4, meaning drop-off was suppressed by EF-P. At X_4, small residues showed higher incorporation probably due to their flexibility in the exit tunnel. The peak shift of $E_{Ni,\ EF\text{-}P-}$ distribution of polar residue from that of others showed the does-dependent manner of EF-G, demonstrating EF-G-mediated peptidyl-tRNA drop-off for the wide range of sequences.

In order to confirm reliability of the profiling system, the nascent peptides appeared in the nine highest and eight lowest $E_{Ni,\ EFP-}$ were individually expressed *in vitro* with 10 μM EF-G in the absence of EF-P and quantified by LC-ESI MS. Nascent peptides with higher $E_{Ni,\ EF\text{-}P-}$ yielded the full-length peptides with negligible levels of the C-terminal peptide. On the other hand, nascent peptides with lower $E_{Ni,\ EF\text{-}P-}$ did not give the full-length peptides but only the C-terminal peptide. Good correlation between $E_{Ni,\ EF\text{-}P-}$ and quantity by LC-ESI MS demonstrated the reliability of the system. It was concluded that frequency of peptidyl-tRNA drop-off was regulated by the nascent peptide sequences and EF-G concentration.

Truncated polypeptide generation in *in vitro* expression of N-terminal regions of *E. coli*'s proteins and putative proteins and detection of polypeptides lacking their N-terminal amino acids in *in vivo* expression

Finally, *in vitro* and *in vivo*, I demonstrated generation of a C-terminal polypeptides or those lacking their N-terminal amino acids caused by peptidyl-tRNA drop-off. I chose 16 *E. coli* proteins or putative proteins containing a stretch of prolines near their N-termini. In order to screen nascent peptides inducing the drop-off, the N-terminal sequences of the candidate proteins were expressed *in vitro* in the absence or presence of EF-P and PTH, analyzed by tricine-SDS PAGE and LC-ESI MS. Among them, six proteins and putative proteins (YjgZ, PrpR, YhhM, RutD, YdcO, and Flk) exhibited peptidyl-tRNA drop-off. They were cloned into plasmids, expressed in *E. coli* Δ*efp* and *E. coli* Δ*efp* complemented of *efp* on another plasmid, purified, digested by proteases (Glu-C, Lys-N, Lys-C, Asp-N, or Arg-C), and sequenced by LC-ESI MS/MS. As a result, after digestion by Lys-N, polypeptide fragments derived from full-length yhhM and yhhM lacking its N-terminal amino acids were identified. Co-expression of EF-P reduced the ratio of the fragment derived from yhhM lacking its N-terminal amino acids. In conclusion, it was demonstrated that the drop-off and translation restart events were novel pathways of abnormal protein synthesis *in vivo*.

Conclusion

Utilizing RaPID display, 8420 kinds of nascent peptide sequences were comprehensively profiled on frequency of peptidyl-tRNA drop-off. The reliability of developed system was independently demonstrated and side chain polarity just prior to the stretch of prolines was the determinant of drop-off. To the best of my knowledge, this work covered the widest sequences for quantification of peptidyl-tRNA drop-off frequency, which provided insight into regulation of peptidyl-tRNA drop-off depending on the peptide nascent sequences. Demonstration of generation of C-terminal peptides *in vitro* and polypeptides lacking their N-terminal amino acids *in vivo* suggested that the drop-off and translation restart were the novel abnormal translation event in *in vitro* and *in vivo* protein synthesis.

Table of contents

Abbreviation list

aa-tRNA	aminoacyl-tRNA	SSU	small subunit
anti-SD	anti-Shine-Dalgarno	TBS	tris-buffered saline
aaRS	aminoacyl-tRNA synthetase	TFA	trifluoroacetic acid
CME	cyanomethyl ester	Tris	tris(hydroxymethyl)aminomethane
DNase	deoxyribonuclease	tRNA	transfer RNA
DBE	dinitrobenzyl ester	UTR	untranslated region
dFx	dinitro-flexizyme	A	Adenine
dNTP	deoxyribonucleoside triphosphate	C	Cytosine
E. coli	escherichia coli	G	Guanine
EF	elongation factor	T	Thymine
EF-Ts	elongation factor thermo stable	U	Uracil
EF-Tu	elongation factor thermo unstable	mG	2′-O-methylguanine
eFx	enhanced flexizyme	ATP	adenosine triphosphate
FD	formyl donor	GTP	guanosine triphosphate
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	UTP	uridine triphosphate
		CTP	cytosine triphosphate
IF	initiation factor	GMP	guanosine monophosphate
LSU	large subunit	Ala, A	L-alanine
LB	loading buffer	Arg, R	L-arginine
LB	Luria-Bertani	Asn, N	L-asparagine
LC-ESI	liquid chromatography-electron spray ionization	Asp, D	L-aspartic acid
MALDI-TOF	matrix assisted laser desorption/ionization-time of flight	Cys, C	L-cysteine
		Gln, Q	L-glutamine
mRNA	messenger RNA	Glu, E	L-glutamic acid
MS	mass spectrometry	Gly, G	glycine
MTF	methionyl-tRNA formyltransferase	His, H	L-histidine
NTP	ribonucleoside triphosphate	Ile, I	L-isoleucine
ORF	open reading frame	Leu, L	L-leucine
PAGE	poly-acrylamide gel electrophoresis	Lys, K	L-lysine
PCR	polymerase chain reaction	Met, M	L-methionine
PTC	peptidyl transferase center	fMet, fM	formyl-L-methionine
PTH	peptidyl-tRNA hydrolase	Phe, F	L-phenylalanine
RF	release factor	Pro, P	L-proline
RNase	ribonuclease	Ser, S	L-serine
RP	ribosomal protein	Thr, T	L-threonine
RRF	ribosome recycling factor	Trp, W	L-tryptophan
rRNA	ribosomal RNA	Tyr, Y	L-tyrosine
SD	Shine-Dalgarno	Val, V	L-valine
SDS	sodium dodecyl sulfate	X	one of the proteinogenic amino acids

Chapter 1
General Introduction

1.1. Introduction

In this chapter, I would like to introduce my PhD thesis, other recent researches, and my master thesis to help to introduce my PhD thesis. Ribosome translation system synthesizes polypeptides in the manner directed by mRNA sequences, which sometimes faces unsuccessful polypeptide synthesis. Among them, I focused on peptidyl-tRNA drop-off and ribosome stalling, of which mechanism of action remained elusive. Based on previous studies related to the mechanism of prokaryotic translation as well as peptidyl-tRNA drop-off[1] and ribosome stalling[2,3], I hypothesized the following three conditions synergistically induces peptidyl-tRNA drop-off: 1) ribosome stalling due to unsuccessful peptidyl transfer, 2) EF-G-mediated unintentional translocation moving the P-site peptidyl-tRNA and the A-site aminoacyl-tRNA, and 3) a peptidyl-tRNA bearing a short nascent peptide sequence less interactive to the ribosome tunnel. Based on the above hypothesis, I investigated the effects of EF-G, EF-P, mRNA, and nascent peptide sequences on ribosome stalling induced by three consecutive prolines in my master course. As a result, peptidyl-tRNA dropped off in the dose-dependent manner of EF-G and the frequency of the drop-off was regulated by the nature of nascent peptide sequences prior to the prolines. Furthermore, a C-terminal peptide truncated between 2nd and 3rd proline residues was observed. In my PhD study, I further investigated how the nascent peptide sequences regulated the drop-off by saturation mutagenesis and attempted to demonstrate the drop-off event in the polypeptide synthesis in *E. coli*. In order to help to understand my thesis, I would like to introduce summaries of former studies on (1) mechanism of the prokaryotic translation system, (2) peptidyl-tRNA drop-off from ribosome and its causes, and (3) ribosome stalling and its causes. Then, I will introduce my master thesis explaining that peptidyl-diprolyl-tRNA drop-off and generation of a C-terminal peptide induced by EF-G and nascent peptide sequences *in vitro*. Finally, I will introduce my PhD thesis.

1.2. Ribosome translation system and genetic code

In translation system, ribosome and translation factors translate an mRNA sequence to polymerize corresponding amino acids into a polypeptide directed by the genetic code[4]. Translation reaction consists of initiation, elongation, termination, and recycling (**Figure 1**). The standard genetic code is defined by aminoacyl-tRNAs (aa-tRNAs) generated by aa-tRNA synthetases (aaRS) that charge cognate amino acids onto corresponding tRNAs.

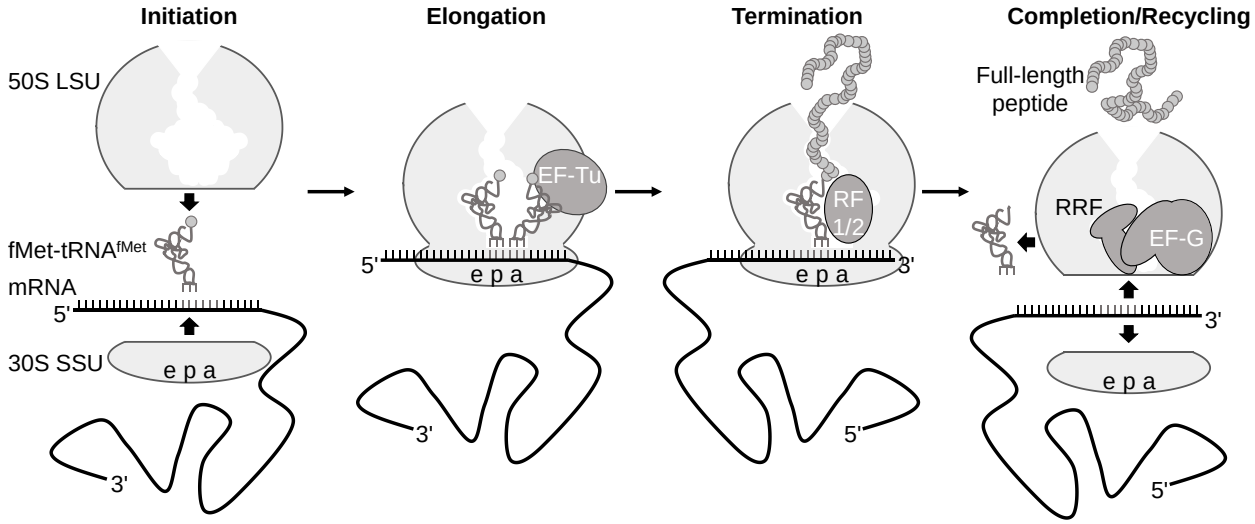


Figure 1 Overview of the bacterial translation.

In translation, initiation, elongation, termination, and recycling take place in this order. Initiation associates the mRNA, fMet-tRNA^{fMet}, 30S small subunit (SSU), and 50S large subunit (LSU), elongation polymerizes amino acids on the nascent peptide into a full-length peptide, termination releases the full-length peptide from the tRNA, and recycling dissociates the 70S ribosome complex into mRNA, tRNA, SSU, and LSU for next round of translation.

1.2.1. Initiation

In prokaryotic initiation, mRNA, fMet-tRNA^{fMet}, 30S ribosomal small subunit (SSU), and 50S ribosomal large subunit (LSU) are assembled into 70S ribosome complex[5]. IF1, IF2, and IF3 position the AUG start codon of mRNA ORF and fMet-tRNA^{fMet} at the P site of 30S SSU (**Figure 2**). In 5' untranslated region (UTR) of AUG start codon in an mRNA, a pyrimidine-rich region called epsilon (ε) element[6], and a purine-rich region called Shine-Dalgarno (SD) sequence[7] interact with the ribosomal protein S1[8] and the anti-Shine-Dalgarno sequence (aSD) of 3' end of 16S rRNA, which promote the correct positioning of the AUG initiation codon at the P site of 30S SSU[9]. After formation of 30S preinitiation complex (PIC) consisting of fMet-tRNA^{fMet}-mRNA-30S SSU-IF1-IF2-IF3, 50S LSU is recruited to the 30S PIC and induces spontaneous GTP hydrolysis by IF2 to dissociate all the IFs, which forms the 70S ribosome-fMet-tRNA^{fMet}-mRNA complex.

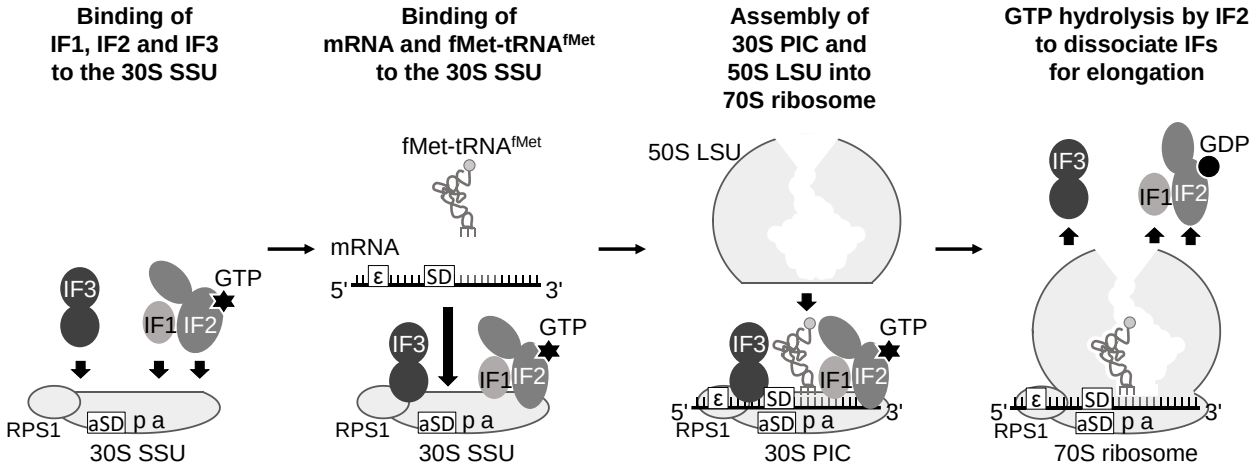


Figure 2 Molecular mechanism of bacterial translation initiation.

First, IF1, IF2·GTP, and IF3 bind to the 30S SSU. IF2 promotes the binding of mRNA and fMet-tRNA^{fMet} to the 30S SSU, forming the 30S preinitiation complex (PIC). In the 5´ UTR of mRNA, the epsilon (ε) element and the SD sequence interact with the ribosomal protein S1 and the anti-SD (aSD) sequence of 3´end of 16S rRNA, which correctly positions the AUG initiation codon at P site of 30S SSU. IF2 recruit the 50S LSU to associate with the 30S PIC, which followed by GTP hydrolysis by IF2 to induce the dissociation of all of IFs and formation of 70S ribosome-fMet-tRNA^{fMet}-mRNA complex for elongation. RPS1: ribosomal protein S1, SD: Shine-Dalgarno.

1.2.2. Elongation

In elongation of nascent polypeptides, accommodation, peptidyl transfer, and translocation are repeated in this order (**Figure 3**). In accommodation, an aa-tRNA is delivered by EF-Tu into decoding center (DC) of the ribosomal A site exposing a codon complementary to the anticodon of the aa-tRNA and enters the A site after release from EF-Tu by hydrolysis of a GTP[10]. Binding affinity between the aa-tRNA and EF-Tu is modulated by the structure of the amino acid, the T stem sequence of the tRNA and binding surface of EF-Tu[11,12]. In peptidyl transfer, PTC in ribosome catalyzes the peptide bond formation reaction between a nascent peptide on the P-site peptidyl-tRNA and an amino acid on the A-site aa-tRNA, resulting in the P-site deacyl-tRNA and the A-site peptidyl-tRNA elongated by one amino acid[13]. In translocation, EF-G mediates the movement of the P-site deacyl-tRNA and the A-site peptidyl-tRNA into E site and P site, respectively, along with the mRNA by hydrolysis of a GTP[14,15]. Then next aa-tRNA is accommodated into A site to repeat the three reactions to synthesize the full-length peptide encoded in the mRNA. Elongating nascent peptides go through the ribosome exit tunnel in the LSU, in which nascent peptides folded into their secondary structure[13]. After emerging from the exit tunnel, the polypeptides are folded into native secondary and tertiary structures spontaneously, which can be mediated by chaperon molecules or even processed by proteases and post-translational modification enzymes.

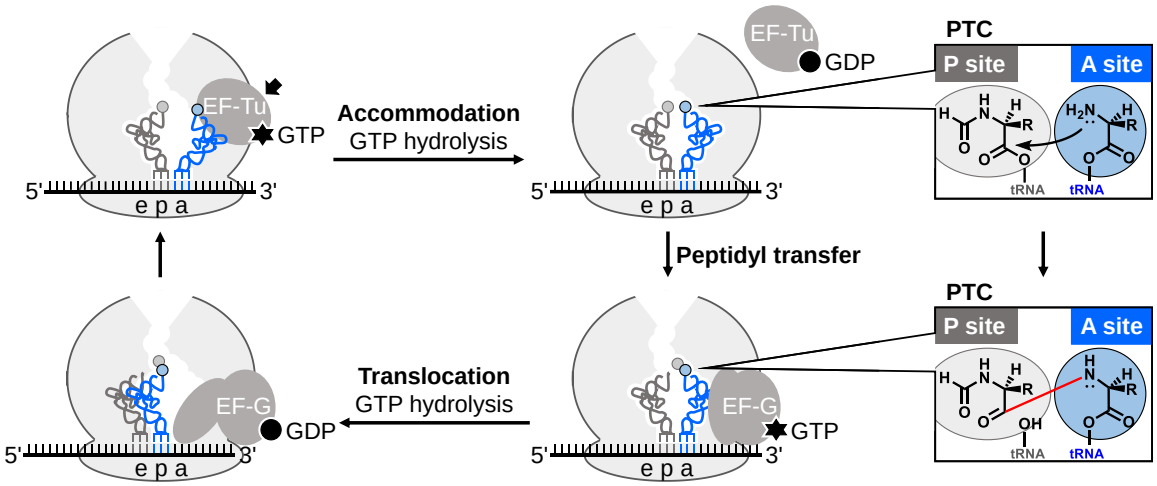


Figure 3 Molecular mechanism of elongation.

In elongation phase of translation, accommodation, peptidyl transfer, and translocation take place in this order. aa-tRNAs are supplied by aaRSs.

1.2.3. Inter subunit rotation of 70S ribosome complex and Translocation

After completion of peptidyl transfer and before translocation, 70S ribosome fluctuate between two conformations in ratchet-like rotation between its 30S SSU and 50S LSU: the classical state and the hybrid state (**Figure 4A, B**)[16]. In the classical state, the anticodon and the CCA end of the peptidyl-tRNA located at A site in the DC of 30S SSU and the PTC of 50S LSU respectively (a/A peptidyl-tRNA, a lower case indicating the anticodon position in the DC and an upper case indicating the CCA end's position in the PTC) and the anticodon and the CCA end of the deacylated-tRNA located at P site in the DC and the PTC respectively (p/P tRNA) (**Figure 4A**)[17]. In the hybrid state, counterclockwise-inter-subunit rotation of 30S SSU move the CCA ends of the peptidyl-tRNA and the tRNA into P site and E site on the 50S LSU, resulted in a/P peptidyl-tRNA and p/E tRNA (**Figure 4B**)[18]. EF-G can bind to both of the classical and hybrid state[19] in its compacted form (**Figure 4C, D**)[20] and partially extends its domain IV to lock the 30S SSU in the hybrid state (**Figure 4E**)[21]. After GTP hydrolysis, EF-G completely extends its domain IV and clockwise rotation of the 30S SSU translocate the peptidyl-tRNA and the tRNA into p/P state and e/E state along with mRNA (**Figure 4F**)[22]. After dissociation of EF-G, the 70S ribosome complex is ready to accommodate next aa-tRNA into vacant A site.

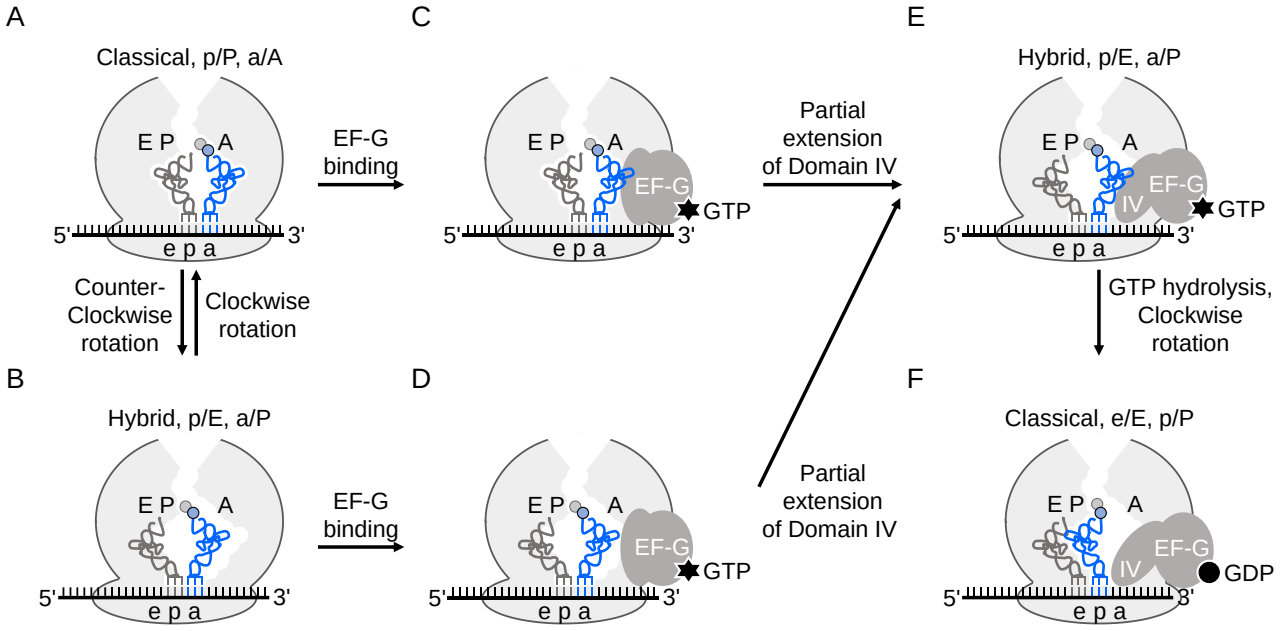


Figure 4 Inter-subunit fluctuation of 70S ribosome between the classical and hybrid state and EF-G-mediated translocation

(A) A 70S ribosome complex in the classical state with an a/A peptidyl-tRNA and a p/P deacylated-tRNA.

(B) A 70S ribosome complex in the hybrid state with an a/P peptidyl-tRNA and a p/E deacylated tRNA. The 70S ribosome complex fluctuates between the classical and hybrid state by inter-subunit rotation.

(C), (D) Binding of EF-G on the classical and hybrid state 70S ribosome complexes.

(E) Partial extension of EF-G domain IV followed by GTP hydrolysis and clockwise rotation of 30S SSU.

(F) A 70S ribosome complex in the classical state with p/P peptidyl-tRNA and e/E deacylated-tRNA.

Before completion of peptidyl transfer, 70S ribosome mainly adopts in the classical state with an a/A aa-tRNA and a p/P peptidyl-tRNA (**Figure 5A**), not the hybrid state with the a/P aa-tRNA and the p/E peptidyl-tRNA (**Figure 5B**)[23]. In the canonical peptide chain elongation, because the rate of peptidyl transfer is so fast that EF-G targets only the 70S ribosome in the classical state with the a/A peptidyl-tRNA and the p/P deacyl-RNA or that in the hybrid state with the a/P peptidyl-tRNA and the p/E deacyl-tRNA (**Figure 4C, D**). However, in the case of the slow peptidyl transfer, such as incorporation of consecutive prolines in the absence of EF-P, EF-G seems to be able to bind the 70S ribosome in the classical state with the a/A aa-tRNA and the p/P peptidyl-tRNA (**Figure 5C**) and may induce inter-subunit rotation and translocation (**Figure 5D, E**), which remained elusive.

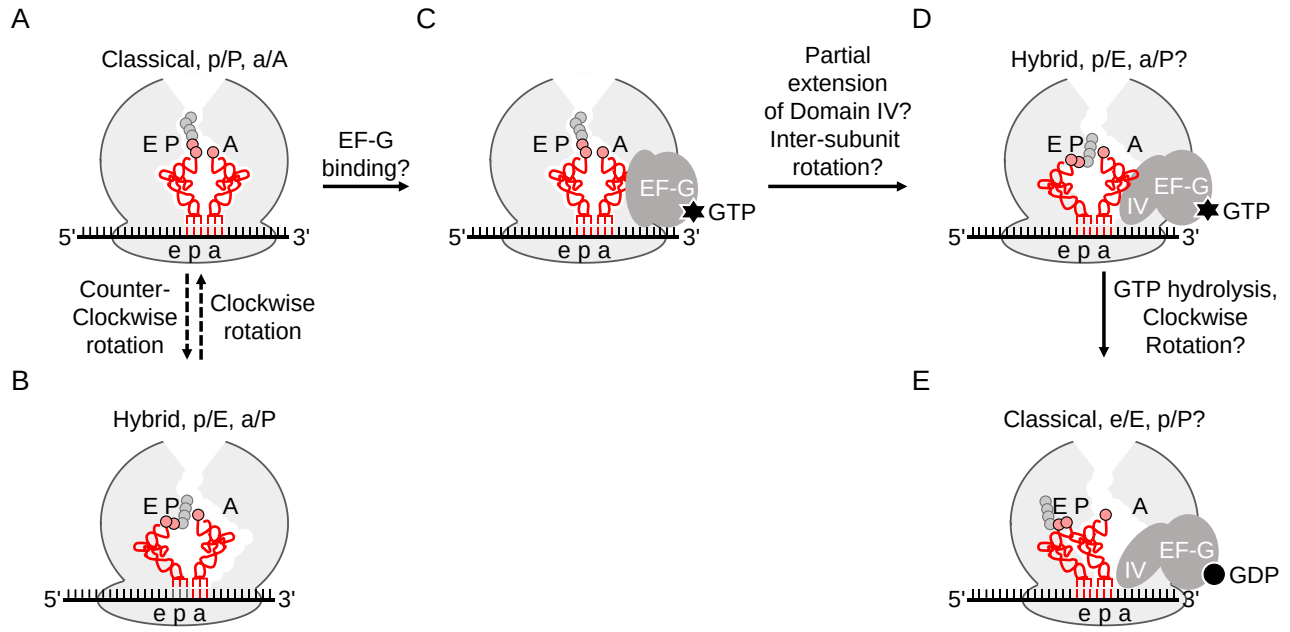


Figure 5 A hypothetical pathway of EF-G binding on 70S ribosome before completion of peptidyl transfer

(A) A 70S ribosome complex in the classical state with an a/A aa-tRNA and a p/P peptidyl-tRNA.

(B) A hypothetical 70S ribosome complex in the hybrid state with an a/P aa-tRNA and a p/E peptidyl-tRNA before completion of peptidyl transfer, which would not be formed.

(C) EF-G binding on the classical 70S ribosome complex with the a/A aa-tRNA and the p/P peptidyl-tRNA.

(D) Partial extension of EF-G domain IV, counter-clockwise rotation of 30S SSU, and hybrid state formation with the a/P aa-tRNA and the p/E peptidyl-tRNA.

(E) Mis-translocation mediated by EF-G forming the 70S ribosome complex in the classical state with the p/P aa-tRNA and the e/E peptidyl-tRNA after GTP hydrolysis and clockwise-inter-subunit rotation.

1.2.4. Termination and recycling

In termination (**Figure 6**), one of three stop codons (UAG, UAA, or UGA) in the A site is recognized by class I release factor RF1 (for UAG or UAA codons) or RF2 (for UAA or UGA codons), which hydrolyze an ester linkage between the full-length peptide and the tRNA[24]. RF1 and RF2 dissociates from the ribosome induced by class II release factor RF3 by hydrolysis of a GTP[25]. After termination, RRF entering the P site and EF-G cooperatively dissociate 70S ribosome complex into 30S SSU, 50S LSU, mRNA, and tRNA for the next round of translation[26].

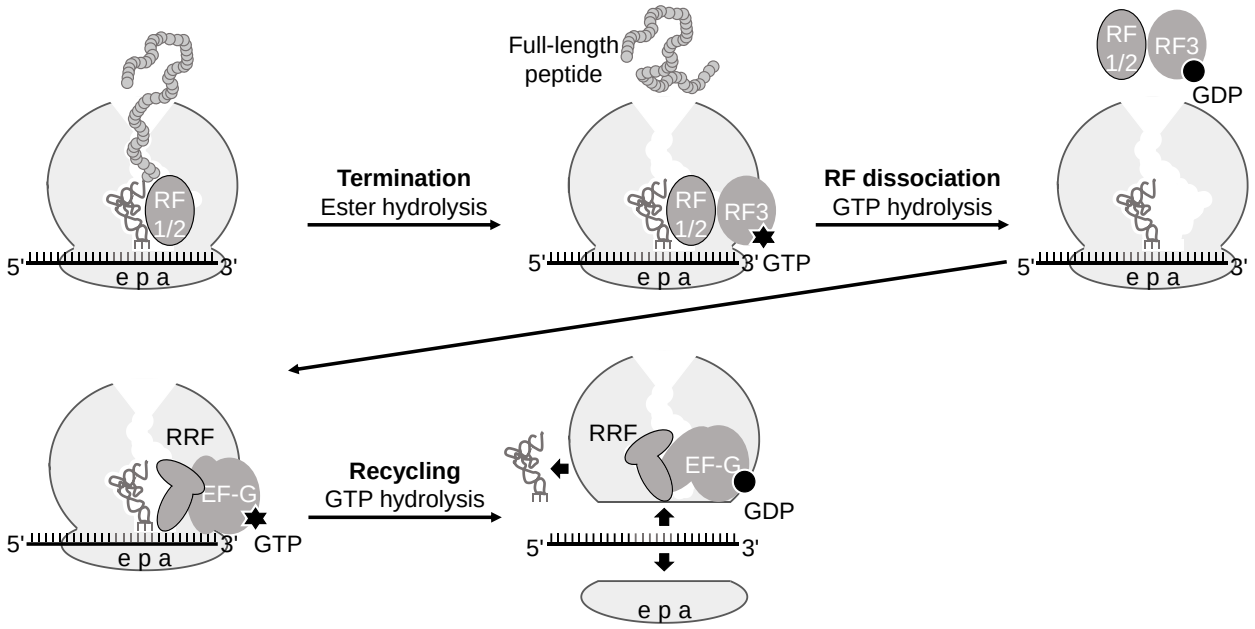


Figure 6 Termination of translation and recycling of the translation apparatus.

When one of the three stop codons appears in the ribosomal A site, corresponding class I release factor RF1 or RF2 release the full-length polypeptide by hydrolyzing the ester linkage between the nascent peptide and the tRNA. RF1/2 dissociates from the ribosome by help of class II release factor RF3 by hydrolysis of a GTP. The 70S ribosome-mRNA-tRNA complex is recycled by RRF and EF-G with a GTP hydrolysis, which dissociates into SSU, LSU, mRNA, and tRNA for the next round of translation.

1.2.5. Genetic code

Genetic codes assign the amino acids and stop signals to corresponding codons mediated by the adopter molecule tRNAs and RF1/RF2, respectively. In nature, an aaRS recognizes and conjugates the cognate amino acid and the cognate tRNA to form an aa-tRNA, which defines the genetic code (**Table 1**). Following this genetic code, polypeptides synthesis utilizes 20 kinds of proteinogenic amino acids including glycine and 19 kinds of L-amino acids and three kinds of termination signals.

Table 1 Standard codon table defined by aa-tRNA.

In the standard genetic code, 20 kinds of proteinogenic amino acids and three termination signals are redundantly assigned to 64 kinds of codon triplets. 20 proteinogenic amino acids are indicated by distinct color based on the nature of the side chains: hydrophobic amino acids with blue, polar and hydrophilic amino acids with orange, and small amino acids with gray.

		2nd base									
		U	aa	C	aa	A	aa	G	aa		
1st base	U	UUU	**Phe**	CUU	**Ser**	UAU	**Tyr**	UGU	**Cys**	U	3rd base
		UUC	**Phe**	CUC	**Ser**	UAC	**Tyr**	UGC	**Cys**	C	
		UUA	**Leu**	CUA	**Ser**	UAA	**Stop**	UGA	**Stop**	A	
		UUG	**Leu**	CUG	**Ser**	UAG	**Stop**	UGG	**Trp**	G	
	C	CUU	**Leu**	CCU	**Pro**	CAU	**His**	CGU	**Arg**	U	
		CUC	**Leu**	CCC	**Pro**	CAC	**His**	CGC	**Arg**	C	
		CUA	**Leu**	CCA	**Pro**	CAA	**Gln**	CGA	**Arg**	A	
		CUG	**Leu**	CCG	**Pro**	CAG	**Gln**	CGG	**Arg**	G	
	A	AUU	**Ile**	ACU	**Thr**	AAU	**Asn**	AGU	**Ser**	U	
		AUC	**Ile**	ACC	**Thr**	AAC	**Asn**	AGC	**Ser**	C	
		AUA	**Ile**	ACA	**Thr**	AAA	**Lys**	AGA	**Arg**	A	
		AUG	**Met**	ACG	**Thr**	AAG	**Lys**	AGG	**Arg**	G	
	G	GUU	**Val**	GCU	**Ala**	GAU	**Asp**	GGU	**Gly**	U	
		GUC	**Val**	GCC	**Ala**	GAC	**Asp**	GGC	**Gly**	C	
		GUA	**Val**	GCA	**Ala**	GAA	**Glu**	GGA	**Gly**	A	
		GUG	**Val**	GCG	**Ala**	GAG	**Glu**	GGG	**Gly**	G	

1.3. Peptidyl-tRNA drop-off

In translation elongation, a peptidyl-tRNA could accidentally drop off from a translating ribosome. In the early stage of the peptide chain elongation, a short peptidyl-tRNA drops off more frequently due to poor interaction of the nascent peptide with the exit tunnel[1,27]. Three kinds of potential causes are reported for peptidyl-tRNA drop-off. (1) Certain mRNA contexts such as cluster of rare codons scarce a rare tRNA isoacceptor, which induce ribosome stalling and drop-off[28,29]. (2) Two groups of translation factors, IF1/IF2 and EF-G/RF3/RRF[30,31], remove the peptidyl-tRNA from 70S ribosome complex and from 70S ribosome complex stalled by a stop codon at A site, respectively. (3) Macrolide antibiotics binding to the exit tunnel[32] clash with the elongating short peptidyl-tRNA. In the former studies, peptidyl-tRNA "drop-off" phenomenon included two distinct phenomena degenerately: peptidyl-tRNA drop-off from the 70S ribosome-mRNA-peptidyl-tRNA complex without dissociation into SSU and LSU, and peptidyl-tRNA release from the 70S ribosome complex along with its dissociation into SSU and LSU. In this section, I distinguish the former "drop-off" as peptidyl-tRNA drop-off and latter "drop-off" as peptidyl-tRNA release and classify former studies on peptidyl-tRNA drop-off. The fate of ribosome after peptidyl-tRNA drops off remained elusive.

1.3.1. Peptidyl-tRNA hydrolase (PTH)

The accumulation of peptidyl-tRNA depletes the tRNA from the translation system, which is lethal for the protein synthesis and cell[33]. Peptidyl-tRNA hydrolase (PTH) breaks an ester linkage between the nascent peptide and the tRNA of a peptidyl-tRNA not binding to a ribosome in order to recycle the tRNA for the next rounds of aminoacylation[34,35]. In the former studies, *E. coli* having a temperature-sensitive-deficient mutation in PTH was widely utilized for the direct detection of a dropped peptidyl-tRNA[36].

1.3.2. Peptidyl-tRNA drop-off induced by mRNA contexts

E. coli mutants defective in PTH activity are unable to maintain the expression of a short ORF *barI* of phage λ minigenes consisting of AUG-AUA-UAA three triplets[28]. The authors proposed that, when ribosome stalled at a sense codon prior to a stop codon, peptidyl-tRNA "dropped off" and inhibit protein synthesis by starvation of the tRNA sequestered in the peptidyl-tRNA (**Figure 7A**). The cell lethality induced by minigenes in PTH-deficient cell were rescued by the over expression of the tRNA sequestered as the peptidyl-tRNA[37]. The pathway of peptidyl-tRNA sequestering can be interpreted as both of drop-off and release of peptidyl-tRNA.

In the case of expression of minigene encoding a fMet-Arg-Arg tripeptide, all the four kinds of ORF were expressed in which Arg was encoded either of rare codons, CGA or AGA, which were decoded by abundant $tRNA^{Arg2}$ and rare $tRNA^{Arg4}$ isoacceptors, respectively (**Figure 7B**)[29]. In PTH deficient *E. coli*, deleterious minigenes to the cell had AGA codon for the 2nd Arg and led to the accumulation of peptidyl-$tRNA^{Arg4}$. Minigene-mediated toxicity and accumulation of peptidyl-tRNA were suppressed by only over expression of $tRNA^{Arg4}$ [29]. Because PTH is only active on the peptidyl-tRNA not bound to the ribosome[34], the authors concluded that cell toxicity mainly derived from the peptidyl-$tRNA^{Arg4}$ accumulated in the stalled ribosome and small fraction of "dropped" peptidyl-$tRNA^{Arg4}$. This "drop-off" can be explained by both of drop-off and release of peptidyl-tRNA.

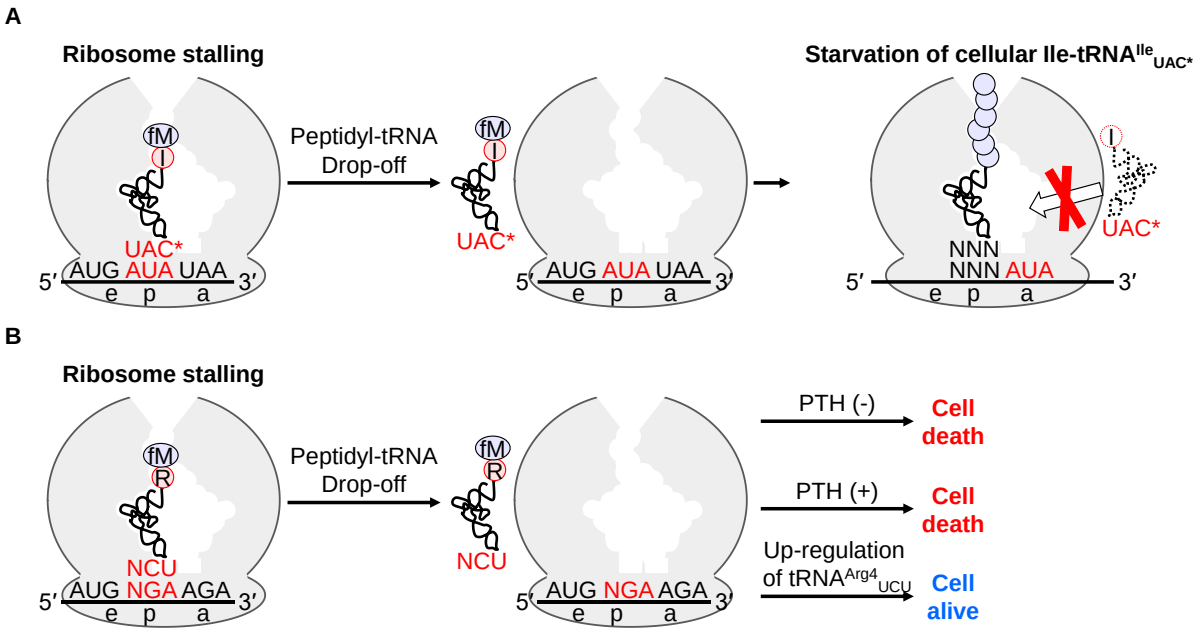


Figure 7 Peptidyl-tRNA drop-off induced by mRNA contexts

(A) Peptidyl-tRNA drop-off induced by AUG-AUA-UAA three triplets in *barl* minigene[28]. Scarce of tRNA^{Ile} in the sequestered peptidyl-tRNA^{Ile} inhibits cellular protein synthesis. C*: lysidine

(B) Peptidyl-tRNA drop-off induced by expression of a fMet-Arg-Arg peptide[29]. In the presence or absence of PTH, expression of the fMet-Arg-Arg peptide was lethal to the cell, which was rescued by only over expression of tRNA$^{Arg4}_{UCU}$.

1.3.3. Peptidyl-tRNA drop-off induced by translation factors

It was reported that peptidyl-tRNA "drop-off" was stimulated by the groups of translation factors, IF1 and IF2[30], and EF-G, RF3, and RRF[31]. Ribosome-mRNA-peptidyl-tRNA complex stalled by UAA stop codons were purified and mixed with IF1 and IF2, which resulted in the accumulation of peptidyl-tRNA in dose-dependent manner of IF1 and IF2. Over expression of both of IF1 and IF2 in the PTH-deficient *E. coli* resulted in impaired growth, which was not observed in *E. coli* with PTH. The authors rationalized that like the 30S PIC of IF1-IF2-SSU-fMet-tRNA^{fMet}, IF1 and IF2 could bind to the A site of 70S ribosome-mRNA-peptidyl-tRNA complex, which induced the dissociation of a peptidyl-tRNA (**Figure 8A**). This phenomenon can be explained by the drop-off of peptidyl-tRNA.

Inactivation of RF3 and suppression mutation of RRF alleviated the temperature-sensitive lethality of PTH deficient *E. coli*[31]. Addition of EF-G, EF3, and RRF stimulated 10-fold higher peptidyl-tRNA drop-off from fMet-Phe-Leu-tRNA-mRNA-ribosome complex stalled by UAA stop codon or AUC Ile codon in the absence of RF1 and RF2 *in vitro*. Because the three factors involve displacement of the P-site deacyl-tRNA and subunit dissociation in ribosome recycling, the authors proposed that recycling-like event released the peptidyl-tRNA on both of elongation ribosome and terminated ribosome (**Figure 8B**).

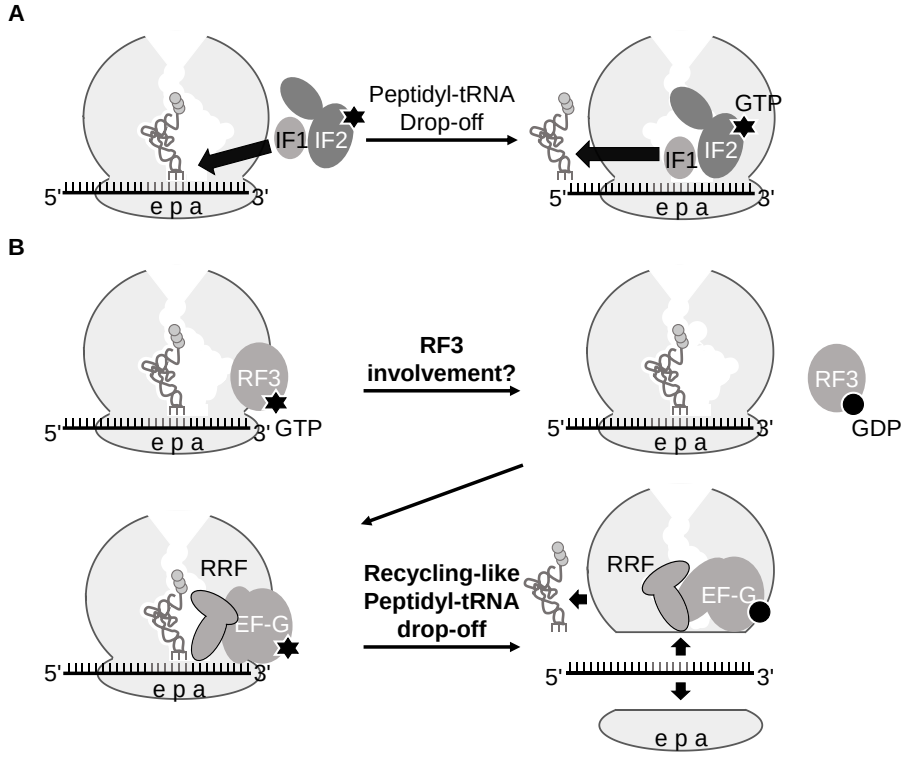


Figure 8 Peptidyl-tRNA drop-off induced by translation factors

(A) A proposed mechanism of peptidyl-tRNA drop-off induced by IF1 and IF2[30]. Like entering into 30S SSU in initiation, IF1 and IF2 would enter into the A site of translating ribosome, which induced the dissociation of a peptidyl-tRNA.

(B) A proposed mechanism of peptidyl-tRNA drop-off induced by EF-G, RF3, and RRF[31]. Like recycling of terminated 70S ribosome, EF-G, RF3, and RRF would remove the peptidyl-tRNA from the ribosome, which could be explained as release of a peptidyl-tRNA from the 70S ribosome complex.

1.3.4. Peptidyl-tRNA drop-off induced by macrolides, lincosamides, and streptogramin B

Macrolide antibiotics are 14-16 membered-ring polyketide compounds directly binding to the exit tunnel of bacterial ribosomes to inhibit the protein synthesis[38–40]. Inhibition works in the manner depending on the sequences of nascent peptides. Due to a little distance between binding sites from PTC, ribosome bound to a macrolide is allowed to synthesize short length of nascent peptides, which drop off from the ribosome when reach to the macrolides[41]. Structure of ribosome-macrolide complexes revealed the drugs bound distinct sites between PTC and entrance of the exit tunnel[40,42]. Consequent biochemical study revealed that three kinds of macrolides (erythromycin, josamycin, and spiramycin), lincosamides, and streptogramin B forced drop-off of peptidyl-tRNA with two to 10 residues of nascent peptides, which corresponded to the length between the binding sties of drugs and PTC[32]. A further study on kinetics revealed that josamycin slowed down peptidyl transfer and induced drop-off of a peptidyl-tRNA from A site after peptidyl transfer or from P site after translocation[43]. This “drop-off” phenomenon could be explained by both of drop-off and release of peptidyl-tRNA.

1.4. Ribosome stalling induced by certain mRNA contexts and nascent peptide sequences

Ribosome stalling on the mRNA is induced by certain mRNA contexts and nascent peptide sequences, from which a peptidyl-tRNA can drop off. Here shows the three kinds of causes of ribosome stalling. 1) Ribosome stalling can be induced by certain mRNA contexts such as the internal SD-like sequence[44], biased codon usage and abundance of aa-tRNA isoacceptors[45], and mRNA secondary structures including stem-loops and pseudoknots[46,47]. 2) A stretch of prolines in the nascent peptides stalls the ribosome by decelerating the peptidyl transfer between prolines via destabilization of the peptidyl-tRNA due to incompatibility of the distinct helix structure of polyproline in the exit tunnel[48]. Translation factor EF-P recognizes and stabilizes the P-site polyprolyl-tRNA^{Pro} [48,49], which alleviate the stalling[50,51]. Incorporation efficiency of a stretch of prolines also depends on its prior and posterior nascent peptide sequences[52,53]. 3) Certain nascent peptide sequences stall the ribosome via interaction with the exit tunnel[54–60], which remodel the PTC into the deficient state for peptidyl transfer[61,62] or peptide chain release[63,64]. Together with small molecules, nascent peptides often auto-regulate expression of their cistronic operon genes which encodes proteins involving biosynthesis, catabolism, or transportation of the small molecules[60,65,66]. In spite of the overlap of the causes, peptidyl-tRNA drop-off and ribosome stalling have been separately described so far. The phenomenon consequent on ribosome stalling remained elusive too, although it was predicted that a short peptidyl-polyprolyl-tRNA drops off from the ribosome in the case of ribosome stalling induced by a stretch of prolines[50].

1.4.1. mRNA elements inducing ribosome stalling

Translation speed is regulated by mRNA elements such as internal SD-like sequences, biased codon usage in ORF, abundance of aa-tRNA decoders, and its secondary structure[67]. Internal SD-like sequences, which promote the translation initiation when it is located at prior to the initiation AUG codon, induces ribosome stalling by base-paring with 3' end of bacterial 16S rRNA (**Figure 9A**)[44]. Except Met and Trp, all the 18 proteinogenic amino acids are redundantly assigned to codons, and usage of codons is biased in the whole ORF according to organisms[68,69]. The codon usage correlates with cellular abundance of corresponding tRNA isoacceptors[45]. Therefore, the synonymous codons are efficiently or inefficiently decoded by the corresponding tRNA isoacceptors, which modulates the elongation speed of nascent peptides to folded into native structure co-translationally[70] and mRNA half-life from RNase degradation[71]. mRNA can form secondary structures by its intra-molecular base-pairing. Although ribosome can unwind the mRNA secondary structure by its helicase activity[72], more complicated secondary structures such as pseudoknots inhibit the progression on mRNA (**Figure 9B**)[46,47]. Translation regulation by mRNA sequences enables the co-translational protein folding correctly or, in the extreme cases, combination of mRNA elements including internal SD-like sequences inducing ribosome stalling, slippery sequences with consecutive homopolymeric bases, and downstream stem-loop structure cooperatively induce the programmed frameshift to produce the full-length proteins (**Figure 9C**)[46,73–78].

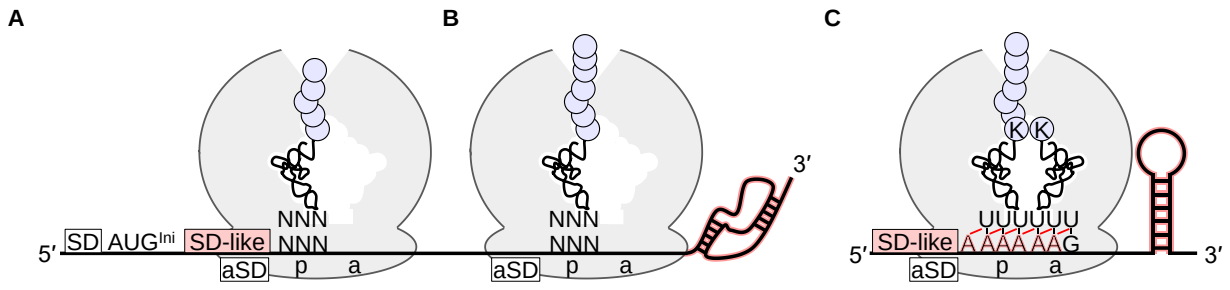


Figure 9 mRNA elements inducing ribosome stalling and their combination utilized for regulation of cellular gene expression

(A) Internal SD-like sequences. Purine-rich SD-like sequence in ORF can stall the translating ribosome via interaction with pyrimidine-rich 3′ end of 16S rRNA[44].

(B) mRNA secondary structures. Even though ribosome can unfold the mRNA secondary structure such as stem-loops, pseudoknots resist the unfolding activity and can stall the ribosome on the mRNA[46,47].

(C) Regulation of cellular gene expression by mRNA elements. Combination of internal SD-like sequence, mRNA secondary structure such as stem-loops, and a slippery sequence with consecutive homopolymeric bases can induce the programmed frameshifting in translation of an ORF[74]. Black: Base-paring between 0 frame codons and anticodons, red: base-paring between -1 codons and anticodons.

1.4.2. A stretch of prolines in polypeptides which induces ribosome stalling

Among the 20 kinds of proteinogenic amino acids utilized for the translation reaction, only proline has the secondary amino group in its five-membered ring (**Figure 10A**). Although the *trans*-peptide bond is far more energetically favorable than the *cis*-peptide bond among primary amino acids, proline can adopt into both *trans*- and *cis*-peptide bond[79] due to lower energy gap derived from the unique five-membered ring and the imide peptide bond[80]. This feature enables the consecutive proline residue to adopt in the distinct polyproline I (PPI) helix and polyproline II (PPII) helix[81]. Although sequence occurrence of consecutive prolines is evolutionally repressed because of their strong ribosome stalling induction, they play important role in the polypeptide sequences[82]. A stretch of proline often appears in the N-terminal 50 residues and the beginning or end of the protein secondary structure. The former feature is thought to prevent jamming of too many ribosomes in an ORF or slow down the polypeptide elongation to facilitate their correct folding into native form, which play the similar role as clusters of rare codons at the beginning of ORF. The latter feature is due to the distinct PPI and PPII helical structures, which break protein secondary structures[83].

In ribosomal incorporation of consecutive proline, PPI or PPII helices in nascent peptides alone are incompatible to the structure of the exit tunnel. The incompatibility can be relieved by (1) destabilizing the CCA end of the nascent peptidyl-tRNA, which induce the ribosome stalling (**Figure 10B**), or (2) stabilizing the CCA end by EF-P (**Figure 10C**)[48]. EF-P protein in bacteria and its homologs in eukaryotes and archaea alleviate the ribosome stalling in incorporation of consecutive prolines into nascent peptides, which achieve the uniform incorporation of proteinogenic amino acids[50,51,84]. EF-P and its homologs are post-translationally modified in a species-specific manner. In *E. coli*, for example, K34 is β-lysyl-hydroxylated by EpmA (previously called as PoxA or YjeA)[85], EpmB

(YjeK)[86], and EpmC (YfcM)[87] enzymes. β-lysylation conducted by EpmA and EpmB enzymes is essential for the EF-P activity and the hydroxylation by EpmC is not essential[48,50,52]. In spite of the uniform incorporation of a stretch of prolines in the presence of EF-P, incorporation efficiency in the absence of EF-P was highly dependent on the amino acid residues neighboring the prolines[52,53]. Although it was predicted that ribosome stalling induced by a stretch of prolines induced the peptidyl-tRNA drop-off from the ribosome[50], the actual phenomenon has not been not confirmed so far.

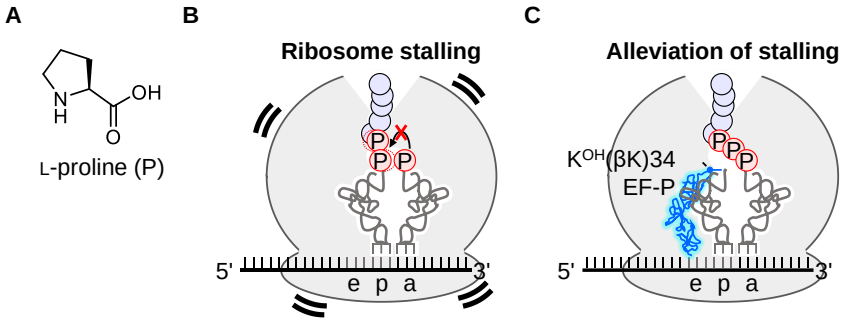


Figure 10 Ribosome stalling induced by a stretch of prolines and its alleviation by EF-P

(A) Chemical structure of L-proline.

(B) Ribosome stalling induced by a stretch of prolines in the nascent peptidyl-tRNA. More than two of consecutive prolines prefer the conformation incompatible to the ribosomal PTC structure, which destabilizes the CCA end of the peptidyl-tRNA and stall the ribosome.

(C) Alleviation of ribosome stalling by EF-P. ε(R)-β-lysyl-hydroxylated K34 ($K^{OH}(\beta K)34$) of EF-P reaches to the CCA end of the P-site peptidyl-tRNA for its stabilization, which promotes the peptidyl transfer and alleviate the ribosome stalling.

1.4.3. Arrest peptides regulating ribosome stalling

In both of bacteria and eukaryote, certain nascent peptides on tRNAs can induce ribosome stalling from the inside via interaction with the exit tunnel[3]. These arrest peptides alone or together with small molecules induce ribosome stalling, which usually regulate expression of ORF of their cistronic operon. Here shows four examples of prokaryotic arrest peptides. 1) In *E. coli*, SecM nascent peptide stalls the ribosome translating its cistronic mRNA, which expose the SD sequence of *secA* gene in its downstream (**Figure 11A**)[88,89]. SecA is the ATPase that drives the protein export across the SecYEG translocon. Expression of SecA drives export of SecM protein via SecYEG translocon, which alleviates the stalling induced by SecM and turn off the expression of *secA* by protecting its SD sequence (**Figure 11B**)[54,88,90,91]. 2) ErmC is one of the methyltransfrerases that dimethylate A2058 of 23S rRNA, which confers the antibiotics resistance against such as erythromycin. Its leader peptide ErmCL induces the expression of *ermC* by conjunction with erythromycin[66]. An inducible concentration of erythromycin arrests translation of *ermCL*, in which stalled ribosome disrupts the mRNA secondary structure to expose the SD of *ermC* (**Figure 11C**). Cryo-EM structure of 50S LSU-ErmCL nascent peptide on the P-site tRNA-erythromycin complex revealed that erythromycin directly bound to both of ErmCL nascent peptide and rRNA bases of the exit tunnel (**Figure 11D**)[92]. 3) Acidic nascent peptides can stall and destabilize the translating ribosome from within (**Figure**

11E). MgtL peptide, a leader peptide of a Mg^{2+} transporter (MgtA) consisting of the acidic residues intermitted by altering prolines, sensitizes a decline in Mg^{2+} concentration, which induced the splitting of 70S ribosome and rearrangement of the cistronic mRNA secondary structure to upregulate expression of MgtA (**Figure 11F**)[60]. 4) The *E. coli tna* operon encodes its leader peptide TnaC, a tryptophanase *tnaA*, and a tryptophan permease *tnaB,* which are inducible by tryptophan. The transcription of *tna* operon is regulated by tryptophan-sensing TnaC peptide interacting with the exit tunnel and tryptophan to inhibit translation termination[65]. With high level of tryptophan, TnaC induces ribosome stalling and inhibition of access of Rho to the transcription termination signal on the mRNA, which synthesizes the whole *tnaAB* transcript. When the tryptophan level becomes low, transcription is terminated prematurely by a Rho-dependent terminator before *tnaA* transcript.

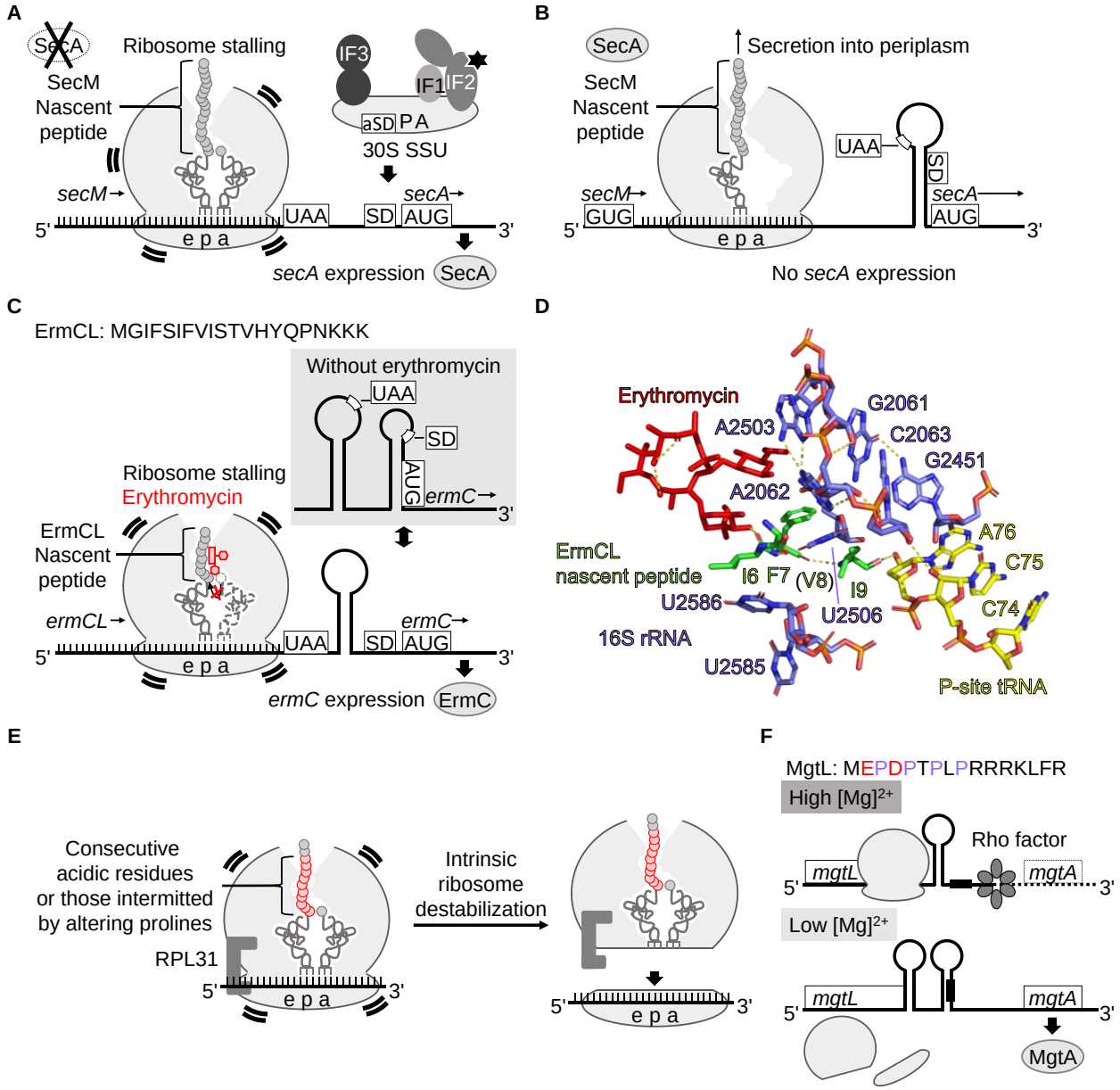


Figure 11 Arrest peptides.

(A) Ribosome stalling in expression of *secM* in the absence of SecA[88]. The stalled ribosome exposes the

SD sequence and the initiation AUG codon of *secA* ORF. Expressed SecA helps SecM to be secreted into the periplasm, which alleviate the ribosome stalling.

(B) Expression of *secM* and its secretion into periplasm in the presence of SecA. Since no ribosome stalling occur, ribosome cannot access to the SD sequence and the initiation AUG codon of *secA* in the mRNA stem-loop.

(C) Expression regulation of *ermC* by ErmCL leader peptide sensitizing erythromycin[56]. In the presence of erythromycin, ErmCL nascent peptide inhibits the peptidyl transfer by allosteric conformation change in the PTC. Ribosome stalling alters the secondary structure of bicistronic *ermCL*/*ermC* mRNA, which exposes the SD sequence and the AUG initiation codon of *ermC*.

(D) Cryo-EM structure of 50S LSU (Purple) complexed with ErmCL nascent peptide (Green) on the P-site tRNA (Yellow) and erythromycin (Red) (PDB ID: 3J7Z)[92]. Erythromycin directly interacts with ErmCL nascent peptide in the exit tunnel to induces allosteric conformational change of proximal rRNA bases, which destabilizes the aa-tRNA binding into A site and inhibits peptidyl transfer.

(E) Intrinsic ribosome destabilization (IRD)[60]. 70S ribosome can be split into 30S SSU and 50S LSU by nascent peptides including consecutive acidic residues or those intermitted by altering prolines. Inter-subunit bridge protein RPL31 harnesses SSU and LSU.

(F) Sensitization of Mg^{2+} concentration by MgtL peptide including the IRD motif. MgtL nascent peptide induces the IRD in the lower Mg^{2+} concentration, which alter the secondary structure of downstream mRNA to elongate the bicistronic mRNA completely and express the Mg^{2+} transporter MgtA.

1.5. Genetic code reprogramming

The standard genetic code can be expanded or reprogrammed by special procedure. Introduction of an engineered tRNA/aaRS pair can expand the genetic code by assigning non-proteinogenic amino acids into codons in *in vitro* or *in vivo*[93]. The genetic code can be reprogrammed via depletion of proteinogenic amino acids/corresponding aaRS pairs and introduction of corresponding non-proteinogenic aa-tRNA pre-charged chemically or enzymatically. The reconstituted cell-free translation system (PURE system)[94] enables the custom-made translation system for any genetic code reprogramming.

1.5.1. The reconstituted cell-free translation system

The reconstituted cell-free translation (PURE) system is mixture of purified components of translation apparatus including ribosome, translation factors, tRNAs, energy regeneration systems, and small molecules[94]. PURE system is a powerful procedure because the any unnecessary components can be depleted from the system and any components required can be added in the system in order to reprogram the genetic code and modify polypeptides enzymatically and chemically.

1.5.2. Genetic code reprogramming for incorporation of non-proteinogenic amino acids

Introduction of an engineered tRNA/aaRS pair can expand the genetic code by assigning non-proteinogenic amino acids into UAG amber codon in *in vitro* or *in vivo*[93]. Genetic code can be reprogrammed via depletion of

proteinogenic amino acids and corresponding aaRSs to make codons vacant and introduction of corresponding tRNA chemically or enzymatically pre-charged by non-proteinogenic amino acids. Artificial aminoacylation ribozyme "Flexizymes" recognize the CCA-3' end of any tRNAs and activated aminoacyl ester moieties to form the aa-tRNAs, which results in aminoacylation versatility to form any kinds of aa-tRNAs theoretically[95]. Recent engineering of the tRNA[96–98] and optimization of factor concentration of the reconstituted cell-free translation system[99] dramatically improved ribosomal incorporation efficiency of non-proteinogenic amino acids and expanded the repertoire of amino acids compatible to the ribosomal translation system. T stem sequences of tRNAs were optimized in order to achieve sufficient binding affinity of non-proteinogenic aa-tRNAs to EF-Tu. D arm shape of tRNA was optimized in order to be recognized by EF-P, which stabilizes the P-site peptidyl-tRNA^{Pro} [48,50,51]. By utilizing Flexible *In vitro* Translation (FIT) system combining genetic code reprogramming enabled by Flexizymes[95], engineered tRNAs, and the optimized reconstituted cell-free translation system, various kinds of nonproteinogenic amino acids such as *N*-substituted α-amino acids[100,101], α-hydroxy acids[102], α-thio acids[103,104], D-α-amino acids[98,99], β-amino acids[105,106], and helical foldamer with backbone aromatic rings[107] can be ribosomally incorporated into polypeptides.

1.6. My master course study

Considering the above conditions of peptidyl-tRNA drop-off, ribosome stalling induced by nascent peptides, and translocation mediated by EF-G, I hypothesized that the following three conditions synergistically induced peptidyl-tRNA drop-off in contrast to the canonical elongation (**Figure 12A, B**): 1) ribosome stalling due to unsuccessful peptidyl transfer, 2) EF-G-mediated unintentional translocation moving the P-site peptidyl-tRNA and the A-site aa-tRNA into E site and P site respectively, and 3) a short peptidyl-tRNA less interactive to the exit tunnel.

At first, I titrated the EF-G concentration against the peptidyl-diprolyl-tRNA-ribosome complex in the reconstituted cell-free translation system without EF-P. Peptidyl-tRNA drop-off was observed in the dose-dependent manner of EF-G, and a C-terminal peptide truncated between 2nd and 3rd prolines was generated (**Figure 12B**). Individual titration of concentration of RF2, RF3, and RRF did not induce generation of the C-terminal peptide. Next, by changing the nascent peptide sequences prior to the proline stretch, it was revealed that the drop-off and generation of C-terminal peptides were suppressed by two kinds of hydrophobic nascent peptides but induced by two kinds of polar/hydrophilic nascent peptides (**Figure 12C**). Finally, in order to distinguish whether the drop-off depended on nascent peptide sequences or mRNA sequences, polar/hydrophilic nascent peptides were expressed from the original purine-rich codons and alternative pyrimidine-rich codons utilizing genetic code reprogramming. As a result, it was revealed that polar/hydrophilic nascent peptide sequences solely induced the drop-off, not the mRNA sequences.

These data suggested following two points. First, EF-G could bind to the 70S ribosome complex which might be in the classical state with an a/A aa-tRNA and a p/P peptidyl-tRNA and induce the translocation of the aa-tRNA and the peptidyl-tRNA into the p/P and e/E state, respectively (**Figure 12B**). After drop-off of the peptidyl-tRNA, accommodation of next aa-tRNA into 70S ribosome with p/P aa-tRNA restarted the translation middle from the ORF, which generated the C-terminal peptide. Second, because nascent peptides can interact with the exit tunnel and polar nascent peptide sequences induced the drop-off and translation restart, interaction or repulsion between polar nascent peptide and the exit tunnel could enable the 70S ribosome to rotate into the hybrid state, which further rotated by EF-G to induce the drop-off of peptidyl-tRNA and translocation of aa-tRNA into p/P state.

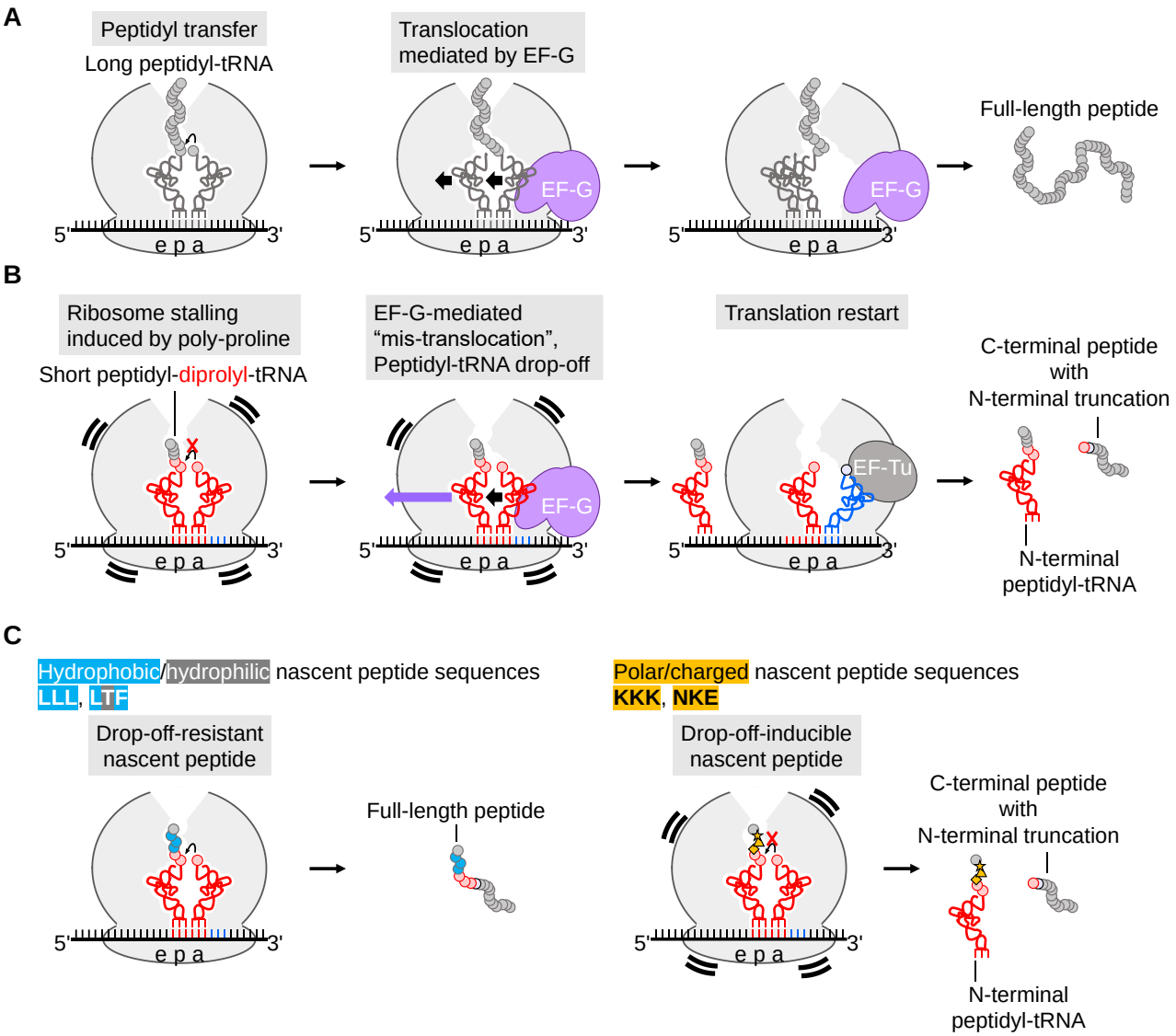


Figure 12 Peptidyl-diprolyl-tRNA drop-off and generation of a C-terminal peptide with N-terminal truncation mediated by EF-G and nascent peptide sequences.

(A) Canonical ribosomal elongation of nascent peptides. After completion of peptidyl transfer, translocation mediated by EF-G moves the P-site deacyl-tRNA and the A-site peptidyl-tRNA into E site and P site, which elongates the nascent peptide into the full-length peptide.

(B) Peptidyl-tRNA drop-off and generation of a C-terminal peptide with N-terminal truncation. On the ribosome stalled by two consecutive prolines in the absence of EF-P, the P-site peptidyl-diprolyl-tRNA dropped off from the ribosome in the dose-dependent manner of EF-G. 70S ribosome complex having the P-site prolyl-tRNA restarted the translation middle of the ORF, generating the C-terminal peptide.

(C) Regulation of peptidyl-tRNA drop-off and generation of C-terminal peptides by nascent peptide sequences. (Left) Three consecutive prolines were well incorporated into hydrophobic nascent peptide、synthesizing only the full-length peptides. (Right) In contrast, they were poorly incorporate into polar/hydrophilic nascent peptide, resulting in the drop-off and generation of a C-terminal peptide.

1.7. My PhD course study

In my PhD course study, I further investigated more nascent peptide sequences in terms of regulation of peptidyl-tRNA drop-off and attempted to demonstrate the generation of C-terminal polypeptide with N-terminal truncation due to the peptidyl-tRNA drop-off in *in vivo* polypeptide synthesis. In order to monitor the peptidyl-tRNA drop-off bearing more sequence variety, I performed the saturated mutagenesis of nascent peptide sequences prior to three consecutive prolines in the presence of 0.03, 0.26, and 10 μM of EF-G and presence and absence of EF-P utilizing mRNA display method combined with genetic code reprogramming[108] and next generation sequencing[109] (**Figure 13A, B, C**). The deep scanning results of 8420 kinds of nascent peptide sequences demonstrated that peptidyl-tRNA drop-off was induced by the amino acids bearing polar side chains just prior to the stretch of prolines in the absence of EF-P. Presence of EF-P suppressed the peptidyl-tRNA drop-off with all the EF-G concentration, which indicated the global suppression of peptidyl-tRNA drop-off. In higher EF-G concentration, the drop-off frequency of polar nascent peptidyl-tRNA was enhanced, which suggested the EF-G-mediated drop-off was a common feature among the polar nascent peptides. In order to demonstrate that the peptidyl-tRNA drop-off in *in vivo* polypeptide synthesis, I expressed *E. coli*'s proteins and putative proteins containing a stretch of prolines proximal to their N-termini. They were expressed in the *E.coli* in which chromosomal *efp* gene was deleted (*E. coli* Δ*efp*) and that was complemented of the *efp* gene by a plasmid (*E. coli* Δ*efp*/+*efp*) (**Figure 13D**). Among six protein constructs expressed, YhhM and RutD were successfully expressed and purified. After protease digestion, peptide fragments derived from YhhM and RutD were sequenced by LC-ESI MS/MS. Fragments derived from full-length YhhM and YhhM lacking its N-terminal amino acids were detected and the ratio of latter was reduced by co-expression of EF-P. This result demonstrated that peptidyl-tRNA drop-off from the ribosome and generation of a C-terminal peptide *in vivo* were the novel abnormal translation events.

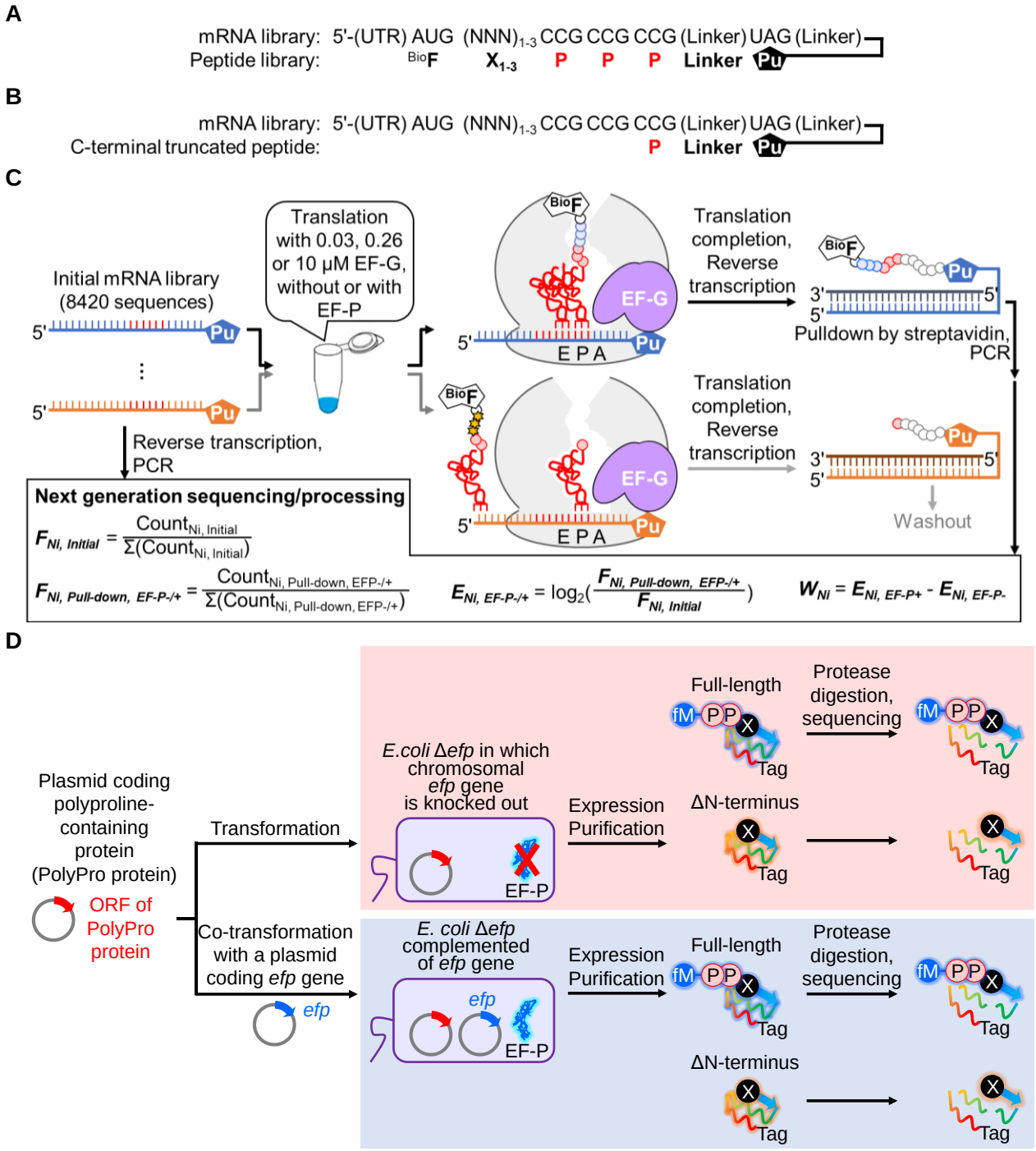


Figure 13 Graphical abstract of Ph.D. thesis

(A) mRNA sequences and peptide sequences used for the saturation mutagenesis of nascent peptides prior to the three consecutive prolines. mRNA sequences consist of the initiation AUG codon, from one to three NNN random codons, and three consecutive proline's CCG codons ligated with a puromycin (Pu) linker, which are translated into N-terminal biotinyl phenylalanine (^{Bio}F) installed for pull-down, from one to three residues of one of 20 proteinogenic amino acids, three consecutive prolines, and a linker peptide.

(B) C-terminal peptide-mRNA conjugates generated by peptidyl-tRNA drop-off.

(C) Graphical scheme of saturation mutagenesis analysis. The mRNA library was translated with 0.03, 0.26, or 10 μM of EF-G in the presence or absence of EF-P in the reconstituted cell-free translation system. Successful incorporation of prolines synthesized the full-length peptide, which was conjugated to the mRNA via puromycin. After reverse transcription, the ^{Bio}F-full-length peptide-mRNA-cDNA conjugates would be selectively recovered by the streptavidin immobilized on magnetic beads. Unsuccessful incorporation due to peptidyl-tRNA drop-off would result in generation of the C-terminal peptide-mRNA-cDNA conjugate without biotin moiety, which would not be pulled down. cDNA library pulled down and initial cDNA library revers-transcribed from the initial mRNA library were amplified by PCR and analyzed by the next generation sequencer. The deep sequencing data were converted to the enrichment of a nascent peptide sequences i (Ni) in the absence or presence of EF-P ($E_{Ni,\ EF\text{-}P\text{-}/+}$) and the suppression strength of drop-off by EF-P (W_{Ni}) calculated by the equations shown.

(D) Demonstration of peptidyl-tRNA drop-off and generation of proteins lacking its N-terminal amino acids in the *in vivo* protein synthesis. Plasmids were designed to code *E. coli*'s proteins and putative proteins bearing a stretch of prolines near the N-termini. Using the plasmids, *E. coli* in which chromosomal *efp* gene was deleted (*E. coli* Δ*efp*) was transformed or co-transformed together with a plasmid coding *efp* gene for its complementation. Polyproline-containing proteins were expressed, affinity-purified, and digested by protease for sequencing of its N-terminal peptide fragments by LC-ESI MS/MS.

Chapter2
Profiling of nascent peptide sequences that regulate the drop-off of peptidyl-tRNA

2.1. IntroductionRibosome stalling induced by a stretch of prolines and its alleviation by EF-P

Among the 20 kinds of amino acids used as building blocks of polypeptides in the ribosomal translation system, proline is the exceptional amino acid due to its secondary amino group in the five-membered ring. This special feature of proline induces the β-turn structure breaking secondary structures in polypeptides, which gives the distinct domains[83]. However, due to the rigid structure, the stretch of prolines in the nascent peptide is incompatible to the exit tunnel, which destabilizes the CCA end of the P-site peptidyl-tRNA to prevent the efficient peptidyl transfer in the absence of EF-P[48,50,51]. EF-P at E site recognizes D-arm of $tRNA^{Pro}$ at P site (**Figure 14A, B**)[49] and stabilizes its CCA end by a post-translationally modified residue of EF-P, which is ε(R)-β-lysyl-hydroxylysine 34 in *E. coli* (**Figure 14C, D**). In the absence of EF-P, the incorporation efficiency of the stretch of prolines depends on the prior and following amino acid residues[2,52,53].

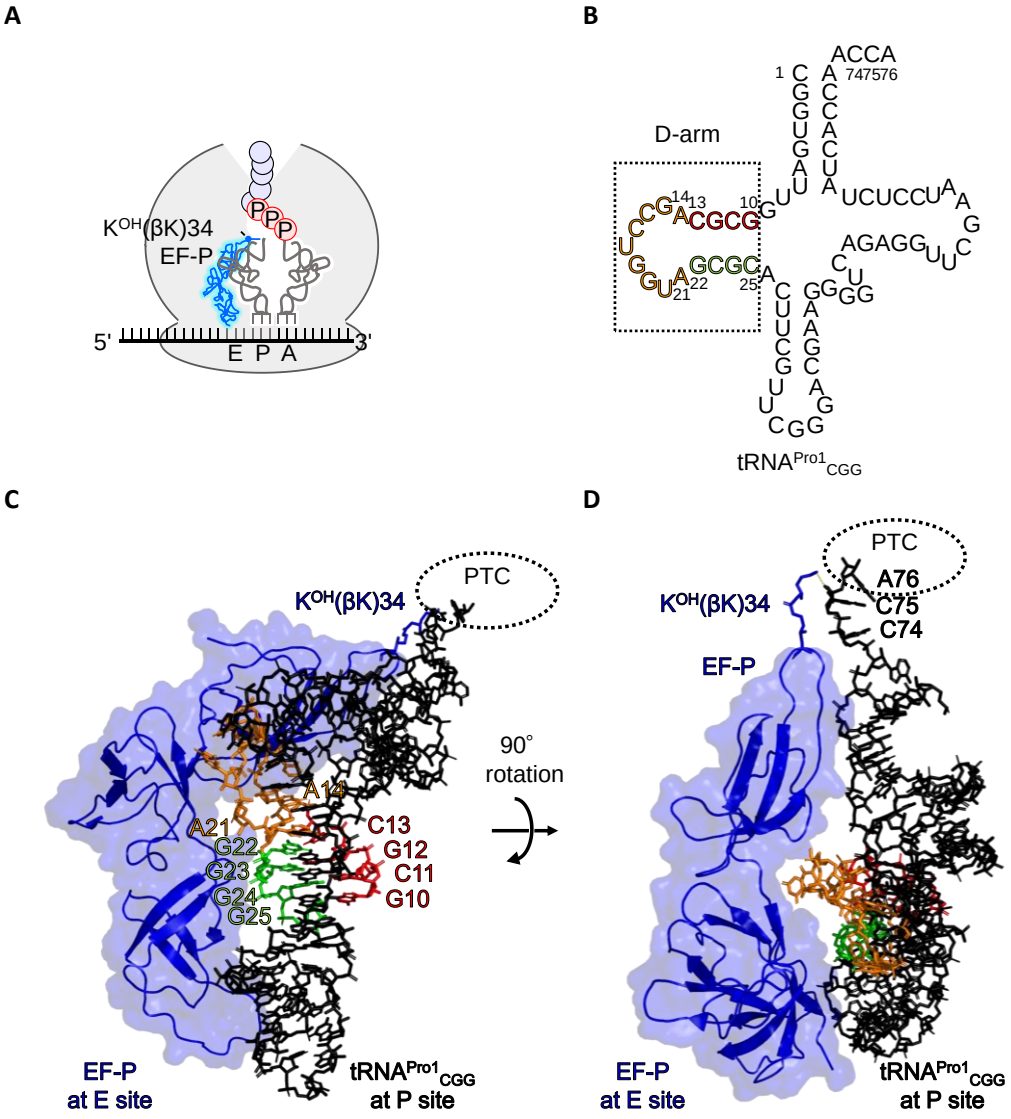


Figure 14 Stabilization of the CCA end of the P-site peptidyl-diprolyl-$tRNA^{Pro}$ by EF-P.

(A) Schematic representation of the complex of 70S ribosome-E-site EF-P-P-site peptidyl-diprolyl-$tRNA^{Pro}$-A-site prolyl-$tRNA^{Pro}$.

(B) Secondary structure of $tRNA^{Pro1}_{CGG}$. The D-arm structure required for EF-P binding is indicated and colored. Base modifications were omitted for clarity.

(C), (D) Cryo-EM structure of EF-P binding to the peptidyl-$tRNA^{Pro1}_{CGG}$ in 70S ribosome (PDB ID: 6ENJ)[48]. Only EF-P and P-site $tRNA^{Pro1}_{CGG}$ are shown for clarity. The P-site $tRNA^{Pro1}_{CGG}$ is stained in the same manner as in B.

2.1.2. Investigation of ribosome stalling strength modulated by nascent peptide sequences

Several literatures have elucidated the relation between amino acids in nascent peptide sequences and ribosome stalling (

Figure 15). Peil et al systematically quantified expression levels of proteins containing PPP, XPP, and PPX motifs by utilizing stable isotope labelling by amino acids in cell culture (SILAC) coupled with high-resolution MS[110] in *E. coli* whose *efp*, *epmA*, *epmB,* or *epmC* gene were knocked out[52]. Because the β-lysylated EF-P is essential for the incorporation of a stretch of prolines, the expression levels of proteins containing PPP motif were significantly down-regulated in the Δ*efp*, Δ*epmA,* and Δ*epmB* strains. Furthermore, expression levels were down-regulated in proteins containing certain XPP and XPP motifs such as PPN, APP, DPP, PPG, and SPP. β-galactosidase activity of *lacZ* transplanted these XPP and PPX motifs showed good correlation with the median of the enrichment quantified by SILAC-MS. In conclusion, by quantifying the expression levels of up to 1300 kinds of proteins, they identified the ribosome stalling motifs in the absence of β-lysylated EF-P as PPX motifs where X = P, W, D, N, and G, and XPP motifs where X= P, D, and A. Woolstenhulme et al systematically quantified the ribosome stalling in *E. coli* lacking *efp* gene by ribosome profiling method[2]. Ribosome profiling quantifies the ribosome density on whole cellular mRNAs to elucidate the translation frequency and translation speed in mRNAs with a single nucleotide resolution utilizing the next generation sequencer[111,112]. High ribosome density on an ORF indicates ribosome stalling. They analyzed the ribosome stalling site in the absence of EF-P, which revealed the strong pausing at PPX motifs where X = W, N, D, and P and XPP motifs where X = P, G, D, A, and S. Chadani et al systematically screened the polypeptide-sequence-dependent ribosome stalling against 1038 kinds of *E. coli* proteins in terms of accumulation of their peptidyl-tRNA[59]. They expressed proteins in PURE system and visualized the puromycin- or RNase A-sensitive bands as peptidyl-tRNA. In the same manner, they visualized RNase A-sensitive proteins from expression in *E. coli*, identified, and categorized the stalling induced by certain nascent peptide sequences. Stalling induced by a stretch of prolines were detected in agreement with the former literatures, yielding the peptidyl-tRNA in the absence of EF-P. They further demonstrated the cluster of acidic residues and those intermitted by altered prolines intrinsically induced the dissociation of 70S ribosome into LSU and SSU (IRD in section 1.4.3)[60]. However, among these literatures, the ribosome stalling has been never connected peptidyl-tRNA drop-off and generation of C-terminal peptides with N-terminal truncation. The sequence coverage was limited only up to the 1600 kinds of natural protein sequences regulating the ribosome stalling and/or peptidyl-tRNA drop-off, which was not comprehensive.

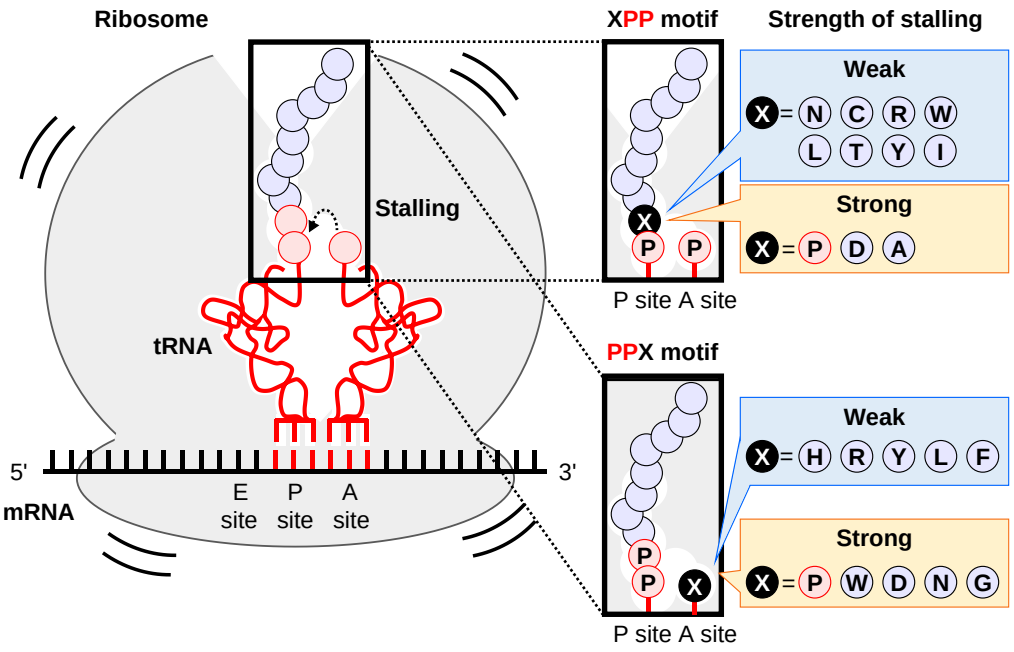


Figure 15 XPP and PPX motifs in nascent peptides inducing ribosome stalling depending on the prior and posterior amino acid residues[2,52].

2.1.3. Saturation mutagenesis to elucidate the polypeptide structure-activity relationship

In order to elucidate the structure-activity relationship of proteins, saturation mutagenesis analysis is widely accepted[109,113,114]. At beginning, saturation mutagenesis required the laborious expression and purification of each mutant for elucidation of its activity. Several procedures resolved the laborious work and enabled the comprehensive elucidation of structure-activity relationship *in vitro* and *in vivo*. Fowler et al performed affinity selection of over 10^5 variants of human Yes-associated protein 65 WW domain (~50 proteinogenic amino acid residues) against its cognate ligand utilizing phage display method[113]. The enrichment of certain protein variants was elucidated by the recovery of the corresponding cDNA quantified by the next generation sequencer. By comprehensively analyzing the sequence enrichment and target affinity, they identified variable residues and conserved residues which tolerate and deny the amino acid mutation against binding, respectively. Their work demonstrated the comprehensive and quantitative elucidation of protein structure-activity relationship. Although polypeptide-display-based method is beneficial to physically conjugate the genotypic cDNA and phenotypic polypeptide for sequencing, the building blocks of polypeptides has been limited to the 20 kinds of proteinogenic amino acids. Rogers et al broke the limitation and elucidated the structure-activity relationship of peptide ligands mutated into proteinogenic and non-proteinogenic amino acids utilizing Random non-standard Peptide Integrated Discovery (RaPID) system combining FIT system and mRNA display method[109]. They revealed the structure-activity relationship of a linear peptide (BH3 domain of PUMA protein binding to folded protein MCL1) and a *de novo* macrocyclic peptide (CP2 peptide inhibiting histone demethylase KDM4A) against their binding targets. 1360 and 468 single mutants of PUMA BH3 domain and CP2 were ribosomally synthesized *in vitro*, in which each amino acid residue was mutated into 20 kinds of proteinogenic and 21 kinds of non-proteinogenic amino acids utilizing genetic code reprogramming. After mRNA conjugation via the puromycin linker and reverse-transcription, peptide-mRNA-cDNA conjugates were challenged to bind to

corresponding targets. Recovered cDNAs of target binders were analyzed by next generation sequencer, interpreted as the enrichment of the sequences through the affinity selection. The depicted structure-activity map revealed that non-proteinogenic amino acids mutations maintained the affinity, which could be even improved. By combination of five mutations improving the affinity, the PUMA protein mutant was improved its affinity (K_D < 0.16 nM from 4 nM). In summary, the deep mutational scanning of peptides enabled the further optimization of the binding affinity of peptides and identification of residues which could be mutated into non-proteinogenic amino acids, which would improve the physical property of peptide such as membrane permeability and protease resistance.

2.1.4. Development of a profiling system of nascent peptide sequences that regulate peptidyl-tRNA drop-off

In order to comprehensively analyze the nature of the nascent peptides that regulates the ribosome stalling and/or peptidyl-tRNA drop-off, I devised a novel nascent peptide profiling system based on RaPID system (**Figure 16**). Saturation mutagenesis analysis was performed against from one to three residues in the nascent peptide prior to three consecutive prolines, up to 8420 kinds of nascent peptide sequences. Sequence-activity relationship was elucidated in terms of incorporation efficiency of consecutive prolines and frequency of peptidyl-tRNA drop-off in various translation conditions, in the absence or presence EF-P, with different concentrations of EF-G, and in incorporation of three consecutive glycines instead of prolines. Results were summarized in heat maps and histograms in terms of polarity of one of three amino acids in the nascent peptides [Polar residues (DERKHQN), Small residues (STGCA), Hydrophobic residues (VILMFYW), and Proline (P)]. In the absence of EF-P, nascent peptides containing additional prolines in the mutated region were less enriched, which corresponded to the former results that more prolines induced the ribosome stalling. The amino acid just prior to the consecutive prolines significantly regulated the frequency of peptidyl-tRNA drop-off, among which polar residues significantly induced the peptidyl-tRNA drop-off but small or hydrophobic residues showed moderate frequency. Higher EF-G concentration induced polar nascent peptidyl-tRNA drop-off more strongly. Addition of EF-P suppressed peptidyl-tRNA drop-off regardless of polarity of nascent peptides, which strongly suppressed the drop-off of peptidyl-tRNA bearing polar amino acids prior to the prolines.

2.2. Results and discussion

2.2.1. Design of the profiling system of nascent peptides that regulate peptidyl-tRNA drop-off

Here shows the experimental design. The peptide sequence was designed as N-terminal biotinylated phenylalanine (^{Bio}F) initiator, installed by the genetic code reprogramming for pull down, followed by from one to three residues of 20 kinds of proteinogenic amino acids, three consecutive prolines, and a linker peptide sequence (**Figure 16A**). Corresponding cDNA and mRNA libraries were designed and prepared by the extension reaction, PCR, and *in vitro* transcription. The initiation and elongation AUG codons were assigned to ^{Bio}F and Met respectively by depletion of 10-HCO-H4folate used for formylation of Met-tRNA$^{fMet}_{CAU}$ to promote initiation (**Table 2, Supplementary Figure 2**), which enabled the saturation mutagenesis analysis utilizing the codon table containing all the 20 kinds of proteinogenic amino acids without any vacant codons. Here shows the experimental scheme of profiling (**Figure 16B**). Prepared mRNA libraries were ligated with a puromycin linker and translated in a reconstituted cell-free translation system in the absence or presence of EF-P with various EF-G concentrations.

Successful incorporation of consecutive prolines would elongate the nascent peptides into full-length peptides, followed by mRNA conjugation via the puromycin and reverse transcription. Resulting N-terminally biotinylated full-length peptide-mRNA-cDNA conjugates were selectively pulled down by streptavidin immobilized on magnetic beads. If the peptidyl-tRNA drops off from the ribosome, a C-terminal peptide-mRNA-cDNA conjugates would be formed, which would not be pulled down by streptavidin. Recovered cDNA and initial cDNA reverse-transcribed from the initial mRNA library were amplified by PCR and analyzed by the next generation sequencer MiSeq (Illumina). The sequencing data were converted to the frequency of a nascent peptide sequence i (Ni) in the initial library ($F_{Ni,\ Initial}$) or those in a library pulled down in the absence or presence of EF-P ($F_{Ni,\ Pull\text{-}down,\ EF\text{-}P\text{-}}$, $F_{Ni,\ Pull\text{-}down,\ EF\text{-}P+}$), and interpreted as enrichment of the nascent peptide sequence i, indicating frequency of peptidyl-tRNA drop-off as $E_{Ni,\ EF\text{-}P\text{-}}$ and $E_{Ni,\ EF\text{-}P+}$. $E_{Ni,\ EF\text{-}P\text{-}/+}$ less than 0 indicates that consecutive prolines are incorporated into Ni less efficiently, which causes more peptidyl-tRNA drop-off and generation of a C-terminal peptide. $E_{Ni,\ EF\text{-}P\text{-}/+}$ more than 0 indicates that consecutive prolines were incorporated into Ni more efficiently, which synthesizes the full-length peptides. Suppression strength of drop-off by EF-P (W_{Ni}) defined as $W_{Ni} = E_{Ni,\ EF\text{-}P+} - E_{Ni,\ EF\text{-}P\text{-}}$ was also calculated. $E_{Ni,\ EF\text{-}P\text{-}/+}$ and W_{Ni} were plotted as heat maps of all the nascent peptide sequences sorted by their side chain polarity at respective position in each EF-G concentration. Then, the $E_{Ni,\ EF\text{-}P\text{-}/+}$ and W_{Ni} were sorted into 4 groups in terms of the side chain polarity [Polar residues (DERKHQN), Small residues (STGCA), Hydrophobic residues (VILMFYW), and Proline (P)] at X_2, X_3, and X_4 positions respectively and plotted as histogram.

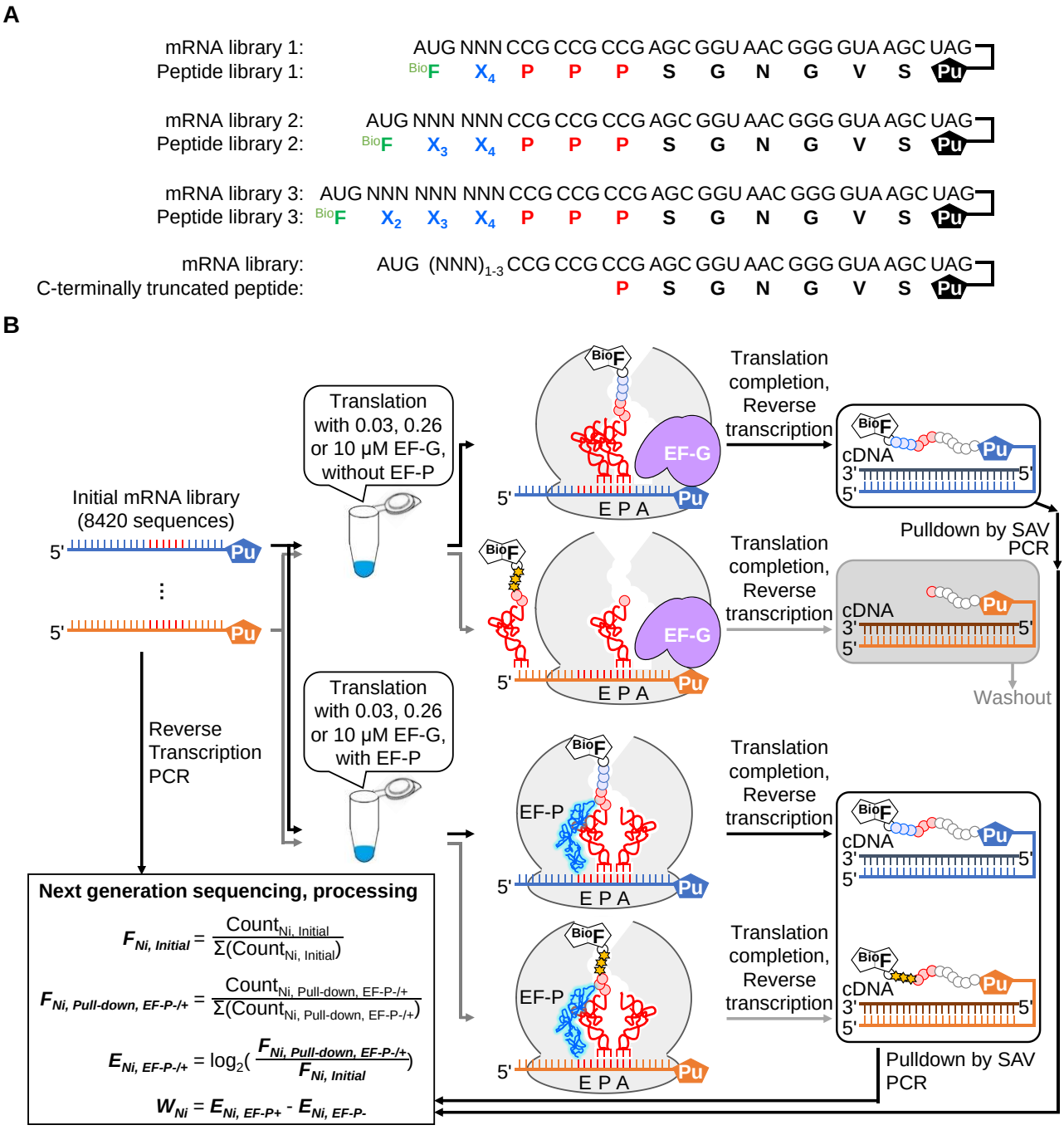


Figure 16 Experimental procedure of nascent peptide profiling system

(A) Designed mRNA and peptide libraries. Each library contains from one to three NNN random codons corresponding to the one of 20 kinds of proteinogenic amino acids, which have 20, 400, and 8000 kinds of nascent peptide sequences theoretically, followed by the three consecutive prolines and a linker peptide. The initiation AUG codon is assigned to ^{Bio}F for pull-down. From all the libraries, peptidyl-tRNA drop-off would generate the C-terminal truncated peptide.

(B) Overview of profiling system. mRNA libraries ligated with a puromycin linker were translated in the reconstituted cell-free translation system with various EF-G concentrations in the absence or presence of EF-P. Successful incorporation of three consecutive prolines elongated the nascent peptides into

full-length. Following mRNA conjugation via a puromycin and reverse-transcription generate the N-terminally biotinylated full-length peptide-mRNA-cDNA conjugates, which were selectively pulled down by streptavidin immobilized on magnetic beads. If the peptidyl-tRNA drops off, C-terminal peptide-mRNA-cDNA conjugates would be generated and washed out in the pull-down step. Recovered cDNA and initial cDNA reverse-transcribed from the initial mRNA were amplified by PCR and analyzed by the next generation sequencer. The read counts of nascent peptide sequence i (Ni) were converted to frequencies in the initial library and library pulled down, $F_{Ni,\ Initial}$, and $F_{Ni,\ Pull\text{-}down,\ EF\text{-}P\text{-}/+,}$ which converted to the enrichment of Ni through the pull-down process, $E_{Ni\ EF\text{-}P\text{-}/+}$. Finally, suppression strength of drop-off by EF-P, W_{Ni}, was calculated.

Table 2 The reprogrammed codon table utilized for the nascent peptide profiling.

By depleting the 10-HCO-H4folate, the initiation AUG codon (AUG^{Ini}) and elongation AUG codon (AUG^{Elo}) were assigned to ^{Bio}F by genetic code reprogramming and Met by cognate MetRS respectively. This genetic code enabled the incorporation of N-terminal ^{Bio}F for pull-down and saturation mutagenesis analysis utilizing 20 kinds of proteinogenic amino acids without any vacant codons.

		2nd base									
		U	AA	C	AA	A	AA	G	AA		
1st base	U	UUU	**Phe**	CUU	**Ser**	UAU	**Tyr**	UGU	**Cys**	U	3rd base
		UUC	**Phe**	CUC	**Ser**	UAC	**Tyr**	UGC	**Cys**	C	
		UUA	**Leu**	CUA	**Ser**	UAA	**Stop**	UGA	**Stop**	A	
		UUG	**Leu**	CUG	**Ser**	UAG	**Stop**	UGG	**Trp**	G	
	C	CUU	**Leu**	CCU	**Pro**	CAU	**His**	CGU	**Arg**	U	
		CUC	**Leu**	CCC	**Pro**	CAC	**His**	CGC	**Arg**	C	
		CUA	**Leu**	CCA	**Pro**	CAA	**Gln**	CGA	**Arg**	A	
		CUG	**Leu**	CCG	**Pro**	CAG	**Gln**	CGG	**Arg**	G	
	A	AUU	**Ile**	ACU	**Thr**	AAU	**Asn**	AGU	**Ser**	U	
		AUC	**Ile**	ACC	**Thr**	AAC	**Asn**	AGC	**Ser**	C	
		AUA	**Ile**	ACA	**Thr**	AAA	**Lys**	AGA	**Arg**	A	
		$AUG^{Ini/Elo}$	**^{Bio}F/Met**	ACG	**Thr**	AAG	**Lys**	AGG	**Arg**	G	
	G	GUU	**Val**	GCU	**Ala**	GAU	**Asp**	GGU	**Gly**	U	
		GUC	**Val**	GCC	**Ala**	GAC	**Asp**	GGC	**Gly**	C	
		GUA	**Val**	GCA	**Ala**	GAA	**Glu**	GGA	**Gly**	A	
		GUG	**Val**	GCG	**Ala**	GAG	**Glu**	GGG	**Gly**	G	

2.2.2. Results of the nascent peptide profiling for incorporation of three consecutive prolines

Nascent peptide sequences containing proline in the random region always showed $E_{Ni,\ EF\text{-}P\text{-}}$ and $E_{Ni,\ EF\text{-}P+}$ values less than 0 in all the EF-G concentrations, indicating that more prolines preceding the fixed three consecutive prolines further induced peptidyl-tRNA drop-off (**Figure 17A, B, C, Supplementary Figure 3, 4, 5, 6, 7, 8,** with 0.03, 0.26, 10 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-},\ X2\ =\ P}$ = -1.1, -0.8, -1.0, $E_{Ni,\ EF\text{-}P+,\ X2\ =\ P}$ = -0.5, -0.5, -0.8, $E_{Ni,\ EF\text{-}P\text{-},\ X3\ =\ P}$ = -1.7, -1.7, -2.1, $E_{Ni,\ EF\text{-}P+,\ X3\ =\ P}$ = -1.6, -1.5, -2.2, $E_{Ni,\ EF\text{-}P\text{-},\ X4\ =\ P}$ = -0.6, -0.9, -2.3, $E_{Ni,\ EF\text{-}P+,\ X4\ =\ P}$ = -0.2, -0.3, -0.9). $E_{Ni,\ EF\text{-}P+}$ showed lower half-value of width than that of $E_{Ni,\ EF\text{-}P\text{-}}$, indicating EF-P suppressed the drop-off and promoted the uniform incorporation of the consecutive prolines into nascent peptide sequences. Focusing on residues except proline, at positions X_2 and X_3, there was no significant difference of drop-off frequency depending on nascent peptide sequences (**Figure 17A**, left and middle column, **Supplementary Figure 3, 4, 5, 6, 7, 8,** peak top of polar, small, hydrophobic, at X_2 with 0.03 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = 0.2, 0.2, 0.2, $E_{Ni,\ EF\text{-}P+}$ = 0, 0.2, 0.5, with 0.26 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = 0.2, -0.1, 0.3, $E_{Ni,\ EF\text{-}P+}$ = 0.1, 0.1, 0.3, with 10 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = 0.3, 0.5, 0.7, $E_{Ni,\ EF\text{-}P+}$ = 0.1, 0, 0.5. At X_3 with 0.03 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = 0.1, -0.2, 0.2, $E_{Ni,\ EF\text{-}P+}$ = 0.1, 0, 0.4, with 0.26 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = 0.2, -0.4, 0.4, $E_{Ni,\ EF\text{-}P+}$ = 0.1, 0, 0.2, with 10 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = 0.5, 0, 0.7, $E_{Ni,\ EF\text{-}P\text{-}}$ = 0.3, -0.1, 0.4). However, at position X_4 just before the proline stretch, polar amino acids showed significantly lower distribution in $E_{Ni,\ EF\text{-}P\text{-}}$ (**Figure 17A,** top of right column**, B,** top**, C,** top, peak top of polar, small, hydrophobic, with 0.03 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = -0.3, 0.5, 0.2, with 0.26 μM EF-G, $E_{Ni,\ EF\text{-}P+}$ = -0.5, 0.5, 0.4, with 10 μM EF-G, $E_{Ni,\ EF\text{-}P\text{-}}$ = -0.8, 0.7, 0.4), which would induce peptidyl-tRNA drop-off. $E_{Ni,\ EF\text{-}P+}$ at position X_4 did not show polarity-dependent difference (**Figure 17A,** middle of right column, **B,** middle, **C,** middle, peak top of polar, small, hydrophobic, with 0.03 μM EF-G, $E_{Ni,\ EF\text{-}P+}$ = 0.2, 0, 0.3, with 0.26 μM EF-G, $E_{Ni,\ EF\text{-}P+}$ = 0.3, 0.1, 0.3, with 10 μM EF-G, $E_{Ni,\ EF\text{-}P+}$ = 0.3, 0, 0.4) and resulted in distribution of W_{Ni} in higher range, meaning the significant suppression of drop-off of peptidyl-tRNA by EF-P. At X_4 position, small residues showed the higher incorporation probably due to their flexibility in the exit tunnel and the PTC (**Figure 17A**, **Supplementary Figure 3, 4, 5, 6, 7, 8,** peak top of polar, small, hydrophobic, with 0.03 μM EF-G, W_{Ni} = 0.3, -0.4, 0.2, with 0.26 μM EF-G, W_{Ni} = 0.6, -0.5, 0, with 10 μM EF-G, W_{Ni} = 0.8, -0.6, 0). Among the hydrophobic residues, aromatic residues showed the relatively higher $E_{Ni,\ EFP\text{-}}$ (**Supplementary Figure 3A, 3D, 3G, 3J, 3M, 3P, 4A, 5A, 6A,** peak tops with 0.03, 0.26, 10 μM EF-G, $E_{Ni,\ EF\text{-}P,\ X2\ =\ F}$ = 0.2, 0.2, 0.2, $E_{Ni,\ EF\text{-}P,\ X3\ =\ F}$ = 0, 0.3, 0, $E_{Ni,\ EF\text{-}P,\ X4\ =\ F}$ = 0.1, 0.5, 0.4, $E_{Ni,\ EF\text{-}P,\ X2\ =\ Y}$ = 0.2, 0.2, 0.2, $E_{Ni,\ EF\text{-}P,\ X3\ =\ Y}$ = 0.1, 0.6, 0.7, $E_{Ni,\ EF\text{-}P,\ X4\ =\ Y}$ = 0, 0.2, 0.4, $E_{Ni,\ EF\text{-}P,\ X2\ =\ W}$ = 0.2, 0.2, 1.1, $E_{Ni,\ EF\text{-}P,\ X3\ =\ W}$ = -0.5, -0.2, 0.1, $E_{Ni,\ EF\text{-}P,\ X4\ =\ W}$ = -0.3, -0.9, 0.5). It was predicted that aromatic residues preceding proline could promote the *cis*-proline conformation by interaction between the π aromatic face and the polarized C-H bonds of prolines[115], which may promote the stabilization of peptidyl-tRNA in the suitable conformation for the next peptidyl transfer. The peak shift of polar residue's $E_{Ni,\ EF\text{-}P\text{-}}$ distribution from that of other residues showed the does-dependent manner of EF-G, suggesting EF-G concentration contributed the drop-off frequency of peptidyl-tRNA (**Figure 17A,** top of right column, **B,** top, **C,** top, $\Delta E_{Ni,\ EF\text{-}P\text{-}}$ from small, hydrophobic, with 0.03 μM EF-G, -0.8, -0.5, with 0.26 μM EF-G, -1.0, -0.9, with 10 μM EF-G, -1.5, -1.2).

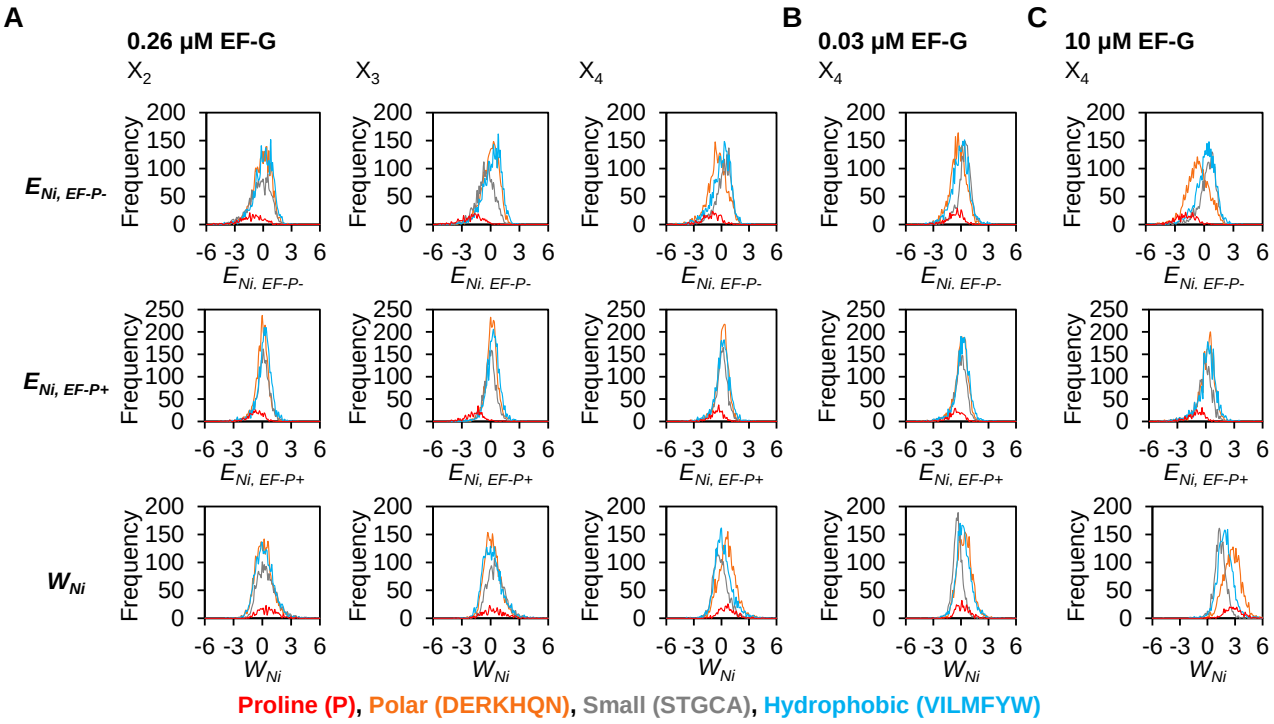


Figure 17 Histograms of the enrichment values through the nascent peptide profiling in terms of side chain polarity of nascent peptide sequences.

(A) Histograms of $E_{Ni,\ EF\text{-}P\text{-}}$, $E_{Ni,\ EF\text{-}P+}$, and W_{Ni} sorted by the side chain polarity of position X_2, X_3, and X_4 with 0.26 μM EF-G.

(B) Histograms of $E_{Ni,\ EF\text{-}P\text{-}}$, $E_{Ni,\ EF\text{-}P+}$, and W_{Ni} sorted by the side chain polarity of position X_4 with 0.03 μM EF-G.

(C) Histograms of $E_{Ni,\ EF\text{-}P\text{-}}$, $E_{Ni,\ EF\text{-}P+}$, and W_{Ni} sorted by the side chain polarity of position X_4 with 10 μM EF-G.

2.2.3. Confirmation of reliability of nascent peptide profiling system

In order to confirm reliability of the nascent peptide profiling system, the nascent peptides appeared in the nine highest and eight lowest $E_{Ni,\ EFP\text{-}}$ (**Figure 18,** indicated by black arrows) were individually expressed *in vitro* with 10 μM EF-G in the absence and presence of EF-P and quantified by LC-ESI MS (**Supplementary Figure 10**). In the absence of EF-P, nascent peptides showed higher $E_{Ni,\ EF\text{-}P\text{-}}$ efficiently synthesized the corresponding full-length peptides with negligible levels of a C-terminal peptide (**Supplementary Figure 10A, 10B**, blue sequences). On the other hand, nascent peptides showed lower $P_{Ni,\ EF\text{-}P\text{-}}$ could not synthesize the corresponding full-length peptides but only the C-terminal peptide (**Supplementary Figure 10A, 10B**, orange sequences) The well correlation between $E_{Ni,\ EF\text{-}P\text{-}}$ and quantification result by LC-ESI-MS confirmed the reliability of the developed profiling system (**Supplementary Figure 10C**). In the presence of EF-P, all the full-length peptides were expressed (**Supplementary Figure 10C, 10D**). As a conclusion, it was revealed that frequency of peptidyl-tRNA drop-off was dynamically affected by the nascent peptide sequences by analyzing up to 8420 sequences with various EF-G concentrations.

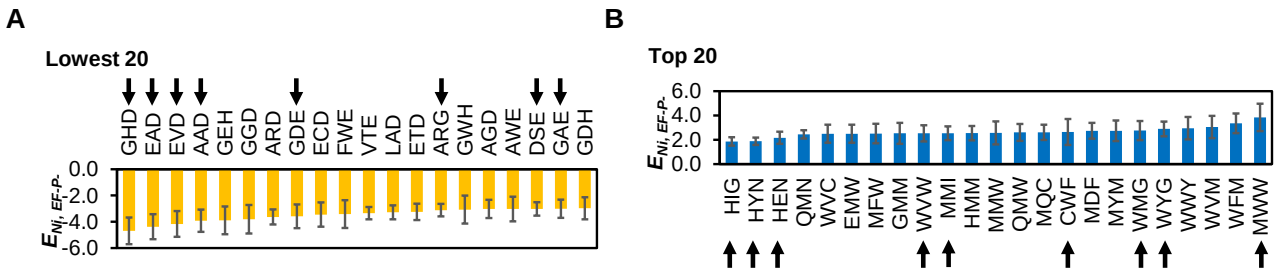


Figure 18 Nascent peptide sequences showing (A) lowest and (B) top 20 $E_{Ni,\ EF\text{-}P\text{-}}$ values.

The sequences indicated by arrows were individually expressed in the reconstituted cell-free translation system for the confirmation of the reliability of the profiling system. Error bars are the Poisson's 95% confidential interval.

2.2.4. Nascent peptide profiling for incorporation of three consecutive glycines

Because proline is the only secondary amino acid among the 20 proteinogenic amino acids, the nascent peptide profiling was applied to the incorporation of three consecutive glycines as a representative of primary amino acids. Judging from the former literature[50,51] and my master course results, incorporation of three consecutive glycines into nascent peptides would not induce ribosome stalling which forced the peptidyl-tRNA drop-off and generation of C-terminal peptides and would not be accelerated by EF-P. Therefore, enrichment of nascent peptide sequences through the profiling system should depend on the abundance of their corresponding aa-tRNA in the system, their rate of accommodation, their rate of peptidyl transfer, and interactions between their side chains and the exit tunnel. Peptide libraries were designed in the same manner as that of prolines, which have N-terminal ^{Bio}F initiator, installed by the genetic code reprogramming for pull down, followed by from one to three residues of 20 kinds of proteinogenic amino acids, three consecutive glycines, and a linker peptide sequence (**Figure 19**). Utilizing the same procedure as incorporation of three consecutive prolines, nascent peptide sequences were profiled against incorporation of three consecutive glycines. mRNA libraries were ligated to the puromycin linker, translated in the reconstituted cell-free translation system with 0.03, 0.26, or 10 μM of EF-G in the absence or presence of EF-P, conjugated with peptides, and reverse-transcribed to generate peptide-mRNA-cDNA conjugates. After selective recovery of N-terminally biotinylated full-length peptide-mRNA-cDNA conjugates by streptavidin immobilized on beads, the resulting cDNA was amplified by PCR and analyzed by next generation sequencer.

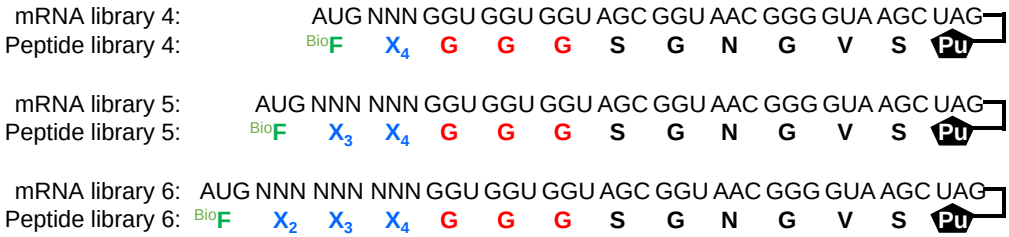


Figure 19 mRNA libraries and peptide libraries for nascent peptide profiling of incorporation of three consecutive glycines.

As a result, the ratio of recovered cDNA divided by the amounts of initial cDNA was higher than in the case of proline, which indicated full-length peptides were dominantly synthesized without peptidyl-tRNA drop-off (**Supplementary table 6**). The $E_{Ni,\ EF\text{-}P\text{-}}$ and $E_{Ni,\ EF\text{-}P+}$ of glycine showed the less biased distribution regardless of the nascent peptide sequences, EF-G concentrations, and the absence or presence of EF-P than those of proline (**Supplementary Figure 11, 12, 13, 14, 15, 16, 17**), which indicated that three consecutive gylcines were incorporated more uniformly than prolines in the above translation conditions. Nascent peptides containing glycines in the randomized regions showed the lower $E_{Ni,\ EF\text{-}P\text{-}}$ and $E_{Ni,\ EF\text{-}P+}$ than those of other nascent peptide sequences, plausibly due to scarcity of Gly-tRNA in the translation system (**Supplementary Figure 11A, 11B, 11D, 11E, 11G, 11H, 11J, 11K, 11M, 11N, 11P, 11Q, 11A, 11B, 12A, 12B**). Nascent peptides containing polar residues showed slightly lower $E_{Ni,\ EF\text{-}P\text{-}}$ and $E_{Ni,\ EF\text{-}P+}$ regardless of EF-G concentration and presence of EF-P, which might indicate that one or several translation reactions shown above were perturbed by the polar amino acids.

2.2.5. Peptidyl-diprolyl-tRNA drop-off controlled by nascent peptides and suppressed by EF-P

To date, several researches have independently investigated the effect of the amino acid sequences around the stretch of prolines on ribosomal stalling in the absence of EF-P[2,52,53]. In these researches, they only tested the peptide sequences derived from the native proteins and investigated only one position preceding or following the prolines. Here, I used 8420 kinds of nascent peptide sequences consisting of from one to three residues of 20 kinds of proteinogenic amino acids, the largest sequence coverage investigated ever. By tracking all the nascent peptides, conclusive result was obtained that drop-off frequency depended on the side chain polarity of nascent peptides.

Also, we revealed that EF-P globally suppressed peptidyl-tRNA drop-off in incorporation of proline stretch into 8420 kinds of nascent peptide sequences (**Figure 17, Supplementary Figure 7**). In the absence of EF-P, drop-off frequency of nascent peptide broadly distributed depending on their sequences, especially on the side chain polarity of the amino acid prior to the proline stretch. In incorporation of a stretch of prolines, EF-P recognizes the peptidyl-tRNA^{Pro} in P site and stabilizes its CCA end for shaping suitable geometry for peptidyl transfer[48]. It is still unknown whether the drop-off event is suppressed by acceleration of peptidyl transfer or direct obstruction of drop-off by EF-P in E site. Here we revealed that EF-P suppresses the drop-off of peptidyl-tRNA and generation of a C-terminal peptide.

2.2.6. Structural insight of interaction of nascent peptides and the exit tunnel

The X_4 position in $^{Bio}FX_2X_3X_4PP$-tRNA, which was profiled as the regulator of peptidyl-diproyl-tRNA drop-off and generation of the C-terminal peptides (**Figure 17A**) seems to be the most proximal to the components of the exit tunnel (**Figure 20**). Considering the Cryo-EM structure of 70S ribosome complex stalled by TnaC leader peptide (**Figure 20A**)[55], Ala22 in the modeled TnaC nascent peptide corresponding to the X_4 position of $^{Bio}FX_2X_3X_4PP$-tRNA is proximal to the A2062 and U2586 forming the exit tunnel (**Figure 20B, C**). Although the actual conformation of $^{Bio}FX_2X_3X_4PP$ nascent peptides are not resolved, polar residues appearing at X_4 could be electrostatically attracted or repulsed by A2062 or U2586 within 5Å distance from the α-carbon atom of Ala22 (**Figure 20B**). The re-orientation of nascent peptides might allow the inter-subunit rotation of the pre-peptidyl-transfer 70S ribosome, in which EF-G would induce the drop-off of the peptidyl-tRNA and translocation of aa-tRNA

into P site. In the case of nascent peptides resisting the drop-off, they might not interact with rRNA bases electrostatically and lock the 70S ribosome in the hybrid state, which could not allow the formation of hybrid state by inter-subunit rotation. Therefore, even in the presence of high concentration of EF-G, only the full-length peptides were synthesized form them without peptidyl-tRNA drop-off and translation restart.

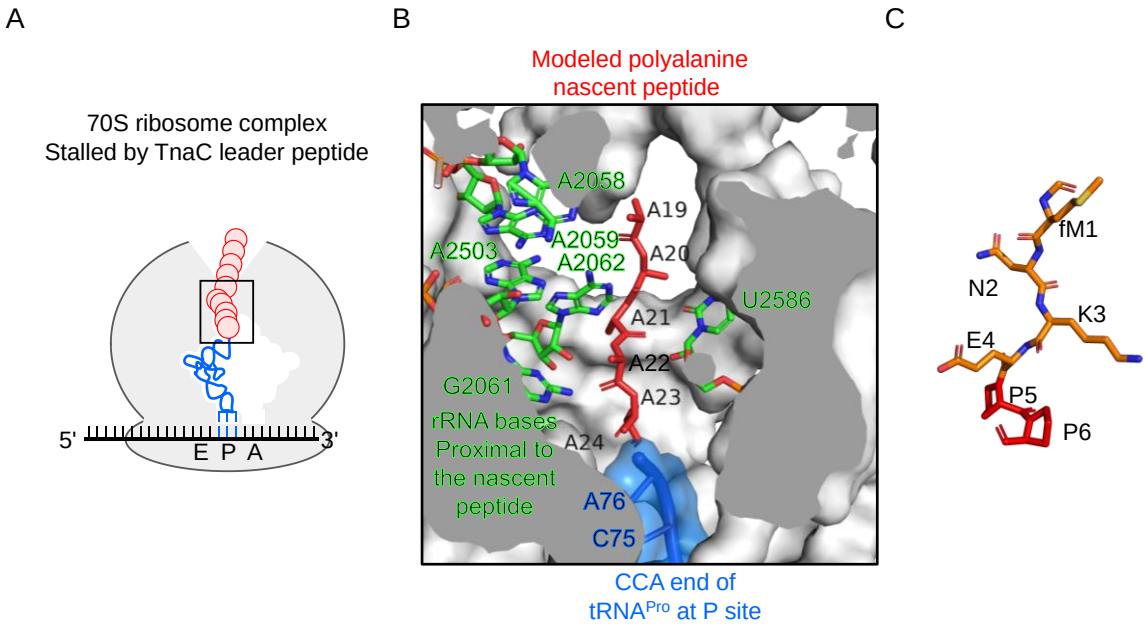


Figure 20 Cryo-EM structure of PTC and the exit tunnel with modeled nascent peptide chains.

(A) The 70S ribosome complex stalled by TnaC leader peptide resolved in the Cryo-EM image. Boxed region was magnified in (B).

(B) Cryo-EM structure of PTC and entrance of the exit tunnel including CCA end of P-site tRNA (Blue) and a modeled polyalanine nascent peptide (Red) (PDB ID: 2WWQ)[55]. rRNAs proximal to the nascent peptide are shown as green stick and others are shown as surface representation. From the original nascent peptide containing 24 alanines, only the six alanines connected to the P-site tRNA are shown.

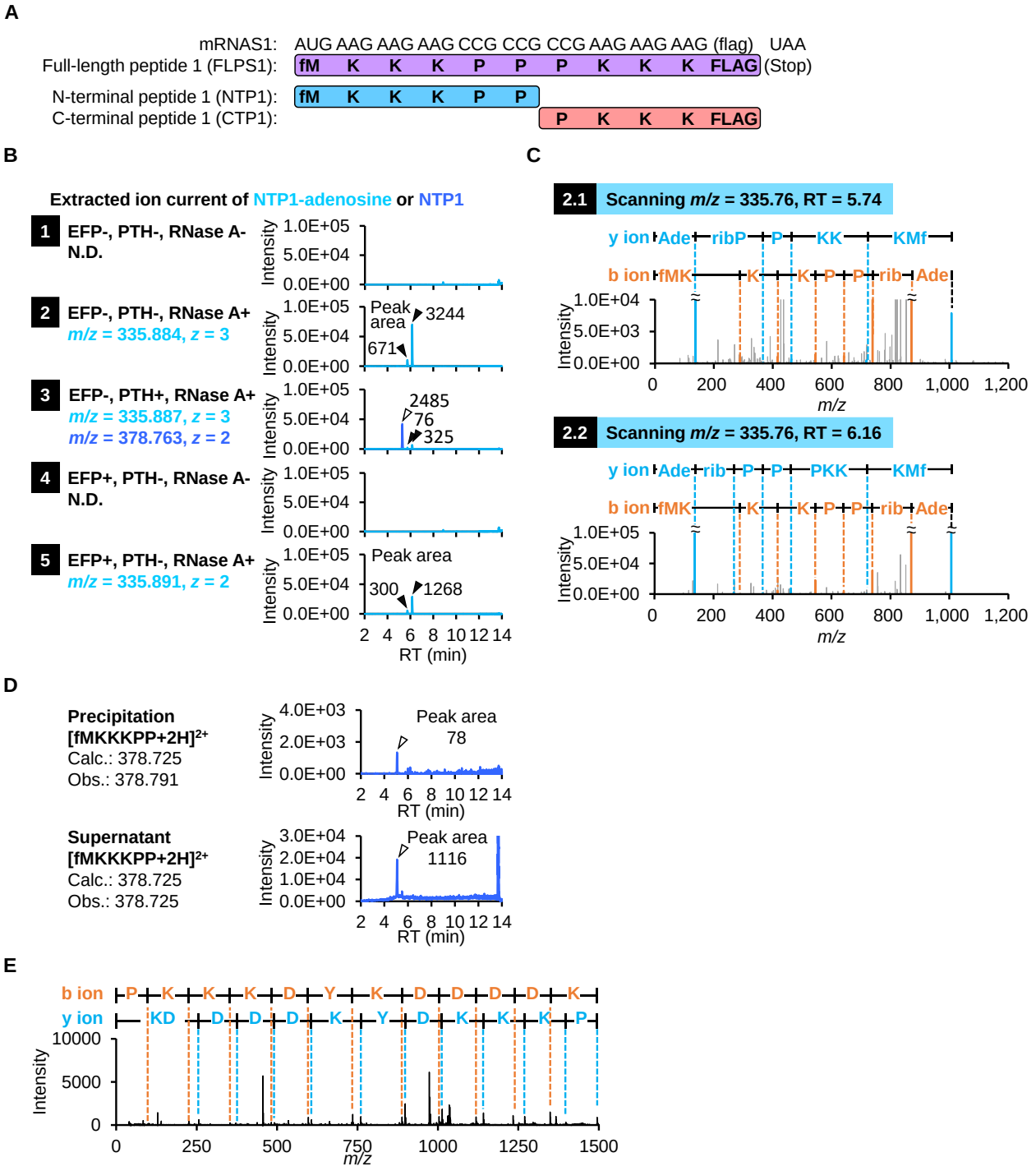
(C) A model structure of nascent peptide fMNKEPP inducing the peptidyl-tRNA drop-off. The model was generated by PyMOL and aligned with the polyalanine nascent peptide in (B).

2.3. Conclusion

The novel system was developed to profile the nascent peptide sequences against peptidyl-tRNA drop-off and generation of C-terminal peptide on the basis of the combination of the mRNA display, genetic code reprogramming, and the next generation sequencer. Taking advantage of this system, regulation of peptidyl-tRNA drop-off was monitored upon incorporation of three consecutive prolines or three consecutive glycines into 8420 kinds of nascent peptide sequences with various EF-G concentrations in the presence or absence of EF-P. In the case of proline incorporation, peptidyl-tRNA drop-off was induced by the nature of the side chain of nascent peptides, which revealed that polar residues just prior to the consecutive proline enhanced the peptidyl-tRNA drop-off. In the absence of EF-P, drop-off of polar nascent peptidyl-tRNA was induced in the dose-dependent manner of EF-G. Presence of EF-P

suppressed the peptidyl-tRNA drop-off especially for the nascent peptide containing additional prolines or polar residues just prior to the consecutive prolines, which achieved the uniform incorporation of three consecutive prolines into nascent peptides. Three consecutive glycines were incorporated into nascent peptides more efficiently and less biased than prolines, as expected. This study revealed that peptidyl-tRNA drop-off occurred depending on the nascent peptide sequences by comprehensive analysis of the broadest coverage up to 8420 kinds of sequence, to the best of my knowledge. It is worth investigating the structural determinant of peptidyl-tRNA drop-off depending on nascent peptide sequences, for example, which induce inappropriate conformation of PTC to repulse the peptidyl-tRNA or which hold the peptidyl-tRNA based on the strength of interaction between nascent peptides and the exit tunnel. The obtained results afford the novel insight of translation regulation mechanism that short peptidyl-diprolyl-tRNA dropped off from the ribosome.

2.4. Supplementary results



Supplementary Figure 1 Identification of a dropped peptidyl-tRNA and a C-terminal peptide with N-terminal truncation generated in the incorporation of three consecutive prolines.

(A) Translated mRNA sequences and expressed peptide sequences. The nascent peptide of a dropped peptidyl-tRNA and the C-terminal peptide observed were shown. FLAG is an octa peptide, Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys.

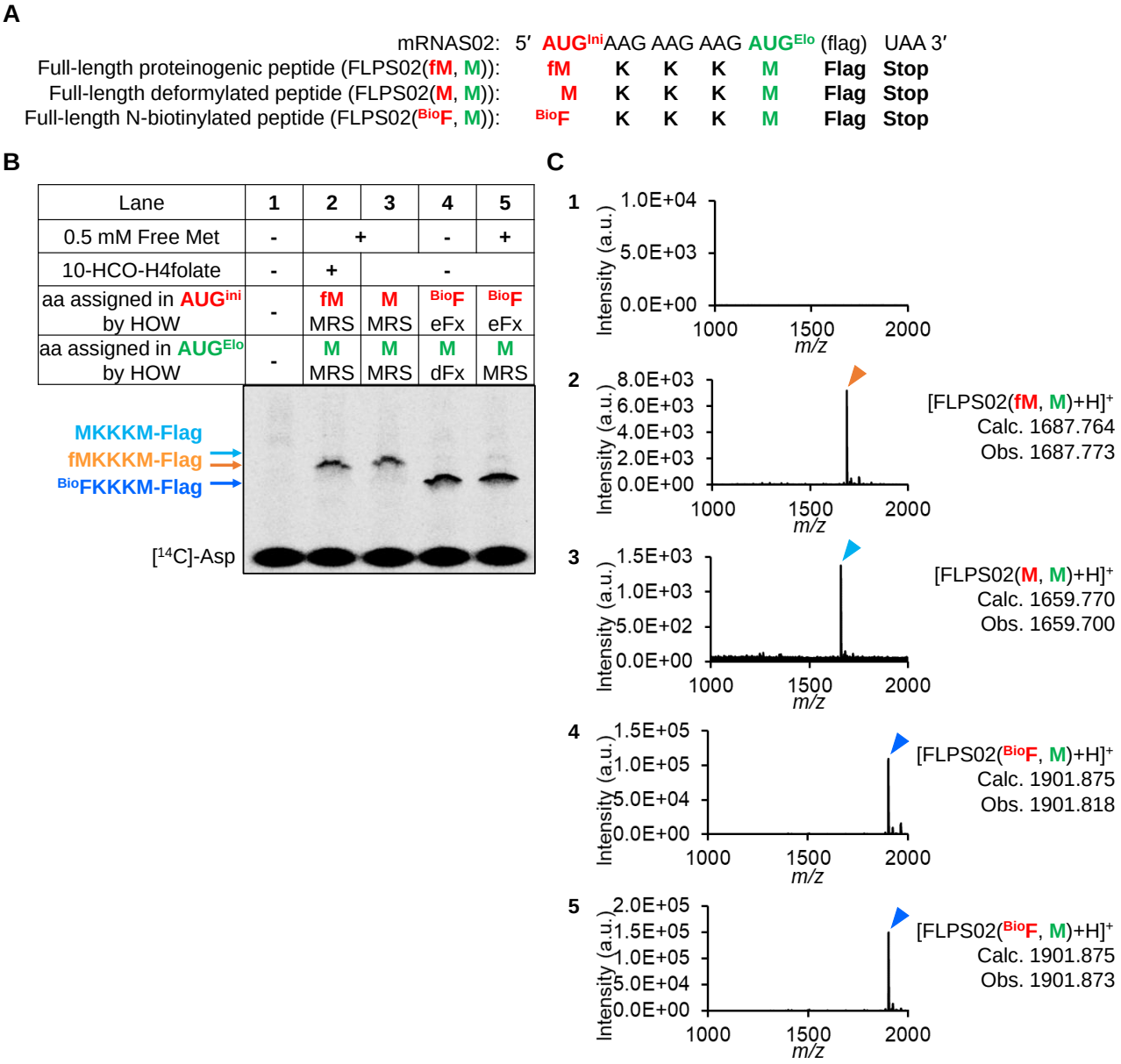
(B) Extracted ion current chromatograms (XICs) of NTP2-tRNA^{Pro} derivatives degraded by PTH or RNase A. Translation products were treated with the reagents shown and analyzed by LC-ESI MS. Black

arrow-heads are corresponding to the peaks of NTP2-Ade and white arrow-head is corresponding to NTP2.

(C) Deconvoluted LC-ESI MS/MS spectra of NTP2-adenosine observed at RT = 5.74 (2.1) and 6.16 (2.2) in the XIC 2 in Fig 1D.

(D) PTH and RNase A addition after ribosomal subunit dissociation to confirm that peptidyl-tRNA remains in stalled ribosomes. After MeOH centrifugation, supernatant was treated with RNase A and precipitation was treated with H_2O for resolve, EDTA, kanamycin, $Mg(OAc)_2$, PTH, and RNase A.

(E) MALDI-TOF MS/MS identification of a C-terminal peptide generated in the translation of mRNA2 with 0.26 μM EF-G.

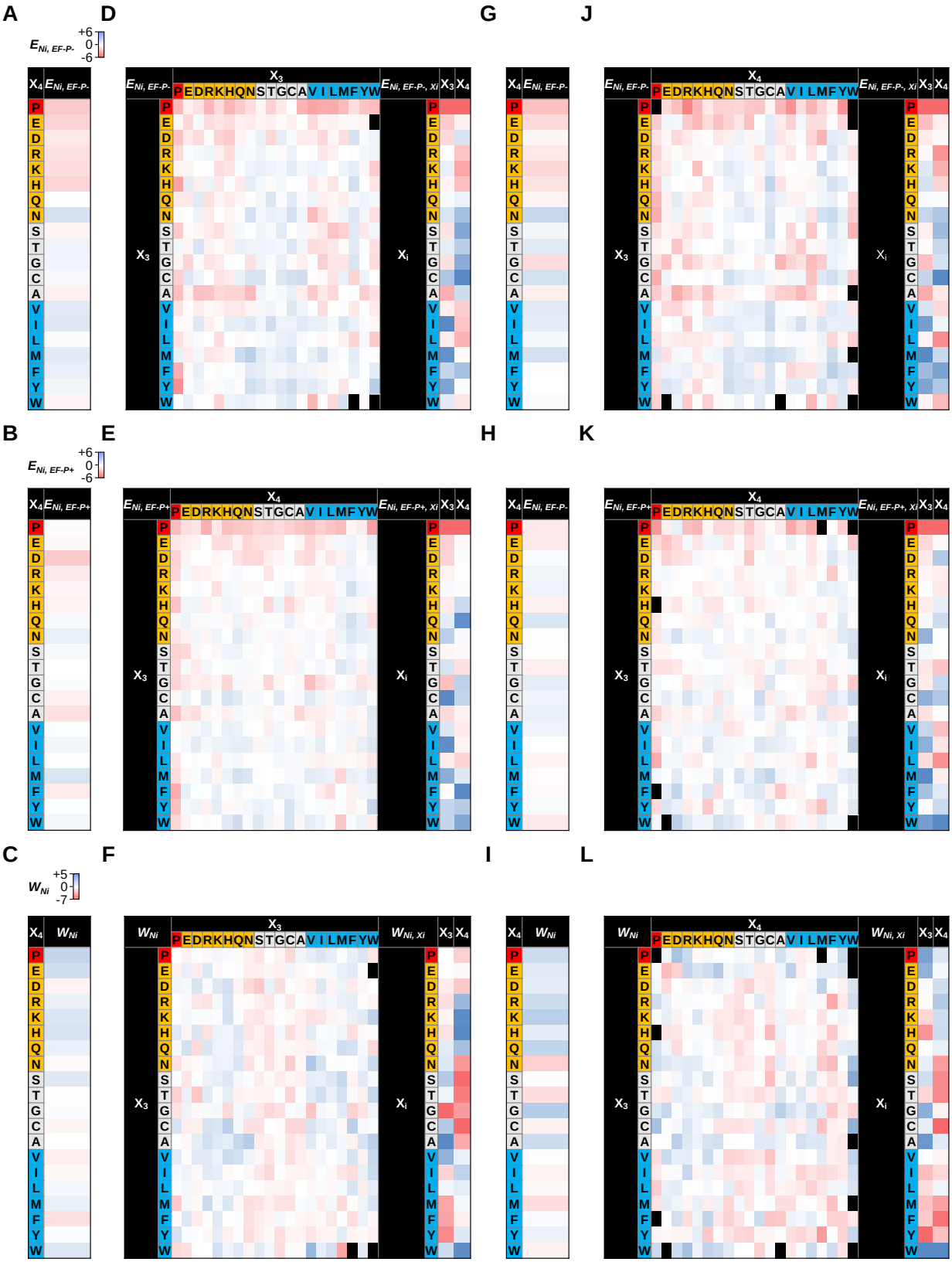


Supplementary Figure 2 Coding different amino acids in the initiation and elongation AUG codons.

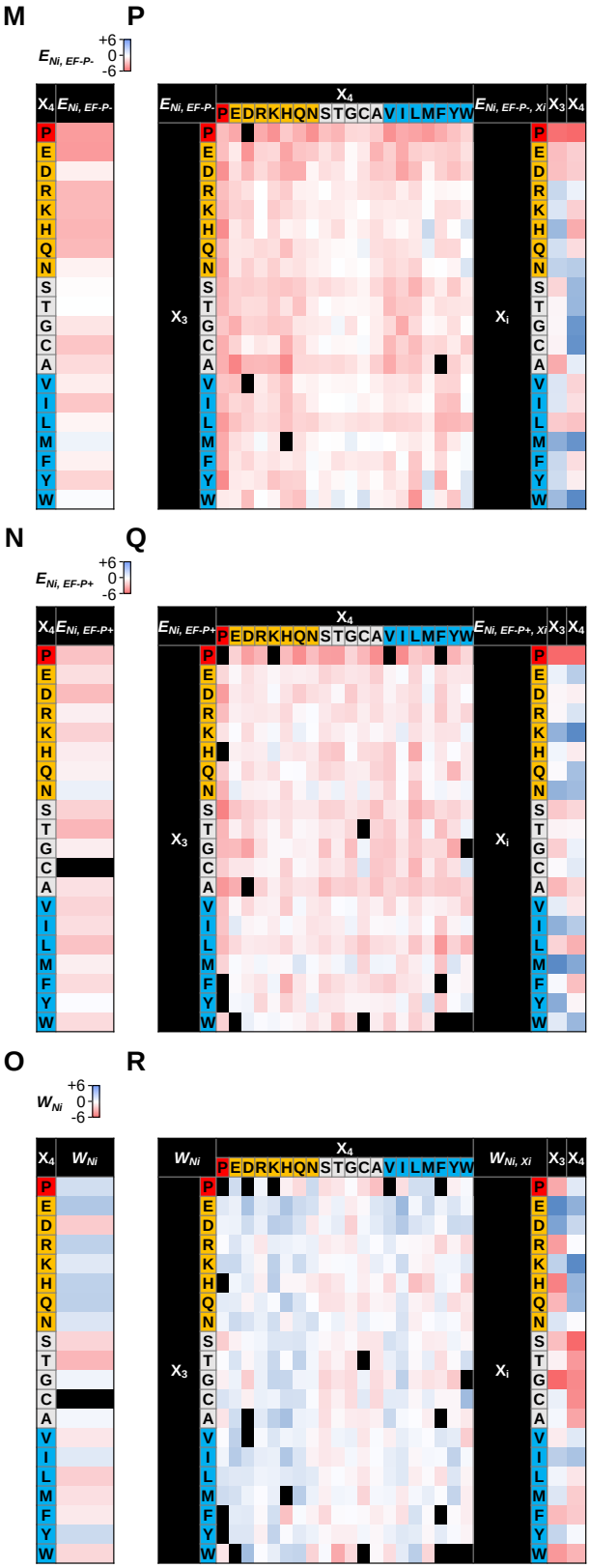
(A) mRNA sequence and peptide sequences used.

(B) Expression of peptides by assigning different amino acids in the initiation AUG codon (AUG^{Ini}) and the elongation AUG codon (AUG^{Elo}). mRNAS02 having both AUG^{Ini} and AUG^{Elo} was expressed without Met and 10-HCO-H4folate for formylation of Met-$tRNA^{fMet}$ (Control for no expression, Lane1), with both of Met and 10-HCO-H4folate in the presence of MetRS (Control for coding fMet and Met into AUG^{Ini} and AUG^{Elo}, Lane 2), with Met but without 10-HCO-H4folate (Control for coding Met into both of AUG^{Ini} and AUG^{Elo}, Lane 3), with pre-charged ^{Bio}F-$tRNA^{fMet}$ and Met-$tRNA^{AsnE2}$ (Coding ^{Bio}F and Met into AUG^{Ini} and AUG^{Elo}, Lane 4), and pre-charged ^{Bio}F-$tRNA^{fMet}$ and Met charged onto $tRNA^{Met}$ *in situ* (Coding ^{Bio}F and Met into AUG^{Ini} and AUG^{Elo}, Lane 5).

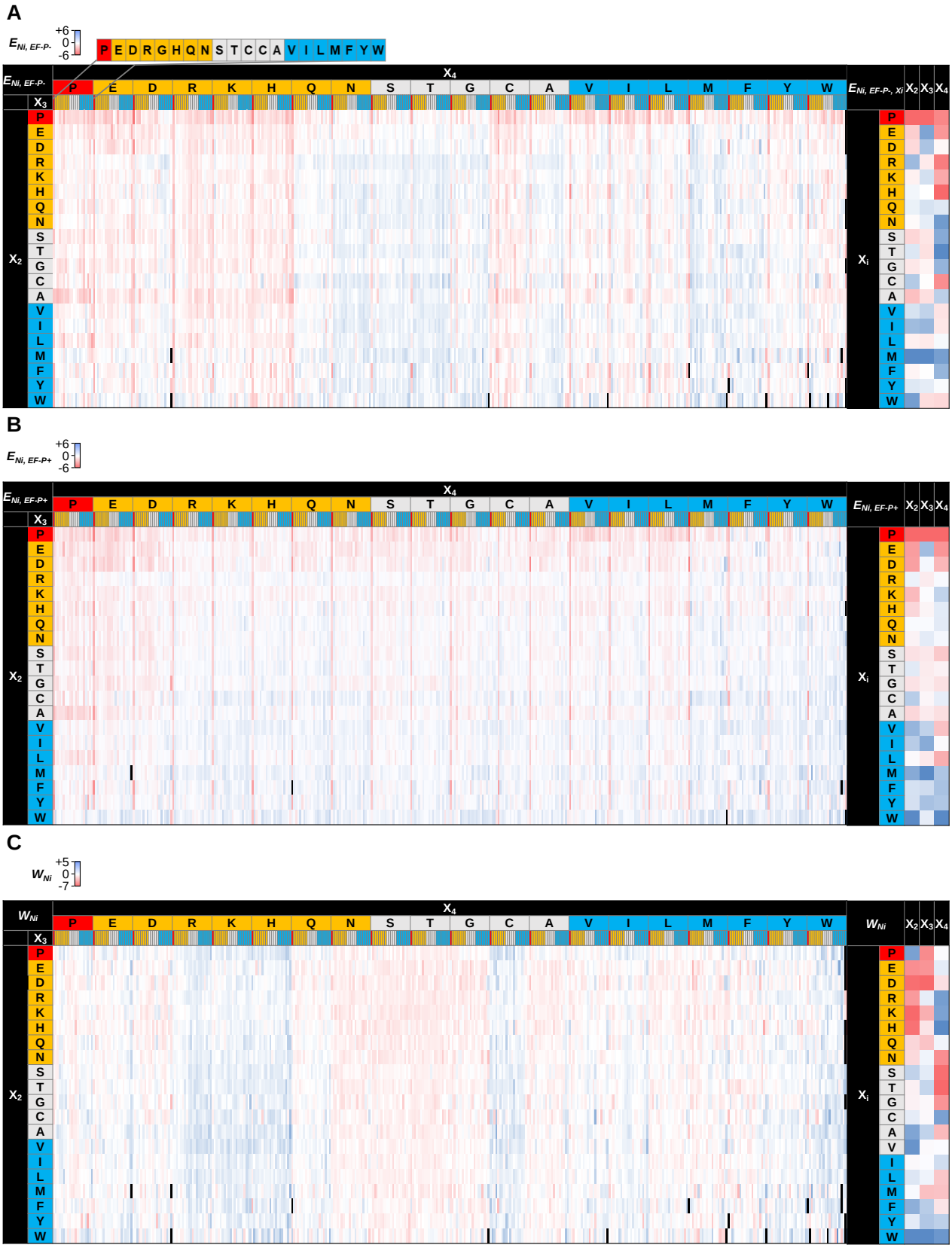
(C) Identification of expressed peptides by MALDI-TOF MS. Same nomenclature as in (B).



Supplementary Figure 3 Profiling of nascent peptide $^{Bio}FX_4$ and $^{Bio}FX_3X_4$ for tri-proline incorporation.
(A)-(C) Heat maps of (A) $E_{Ni,\ EF\text{-}P\text{-}}$, (B) $E_{Ni,\ EF\text{-}P+}$, and (C) W_{Ni}. against $^{Bio}FX_4$ with 0.03 μM EF-G.
(D)-(f) Heat maps of (D) $E_{Ni,\ EF\text{-}P\text{-}}$, (E) $E_{Ni,\ EF\text{-}P+}$, and (F) W_{Ni}. against $^{Bio}FX_3X_4$ with 0.03 μM EF-G.
(G)-(I) Heat maps of (G) $E_{Ni,\ EF\text{-}P\text{-}}$, (H) $E_{Ni,\ EF\text{-}P+}$, and (I) W_{Ni}. against $^{Bio}FX_4$ with 0.26 μM EF-G.
(J)-(L) Heat maps of (J) $E_{Ni,\ EF\text{-}P\text{-}}$, (K) $E_{Ni,\ EF\text{-}P+}$, and (L) W_{Ni}. against $^{Bio}FX_3X_4$ with 0.26 μM EF-G.
Peptide sequences which were not detected by the next generation sequencer are filled with black.



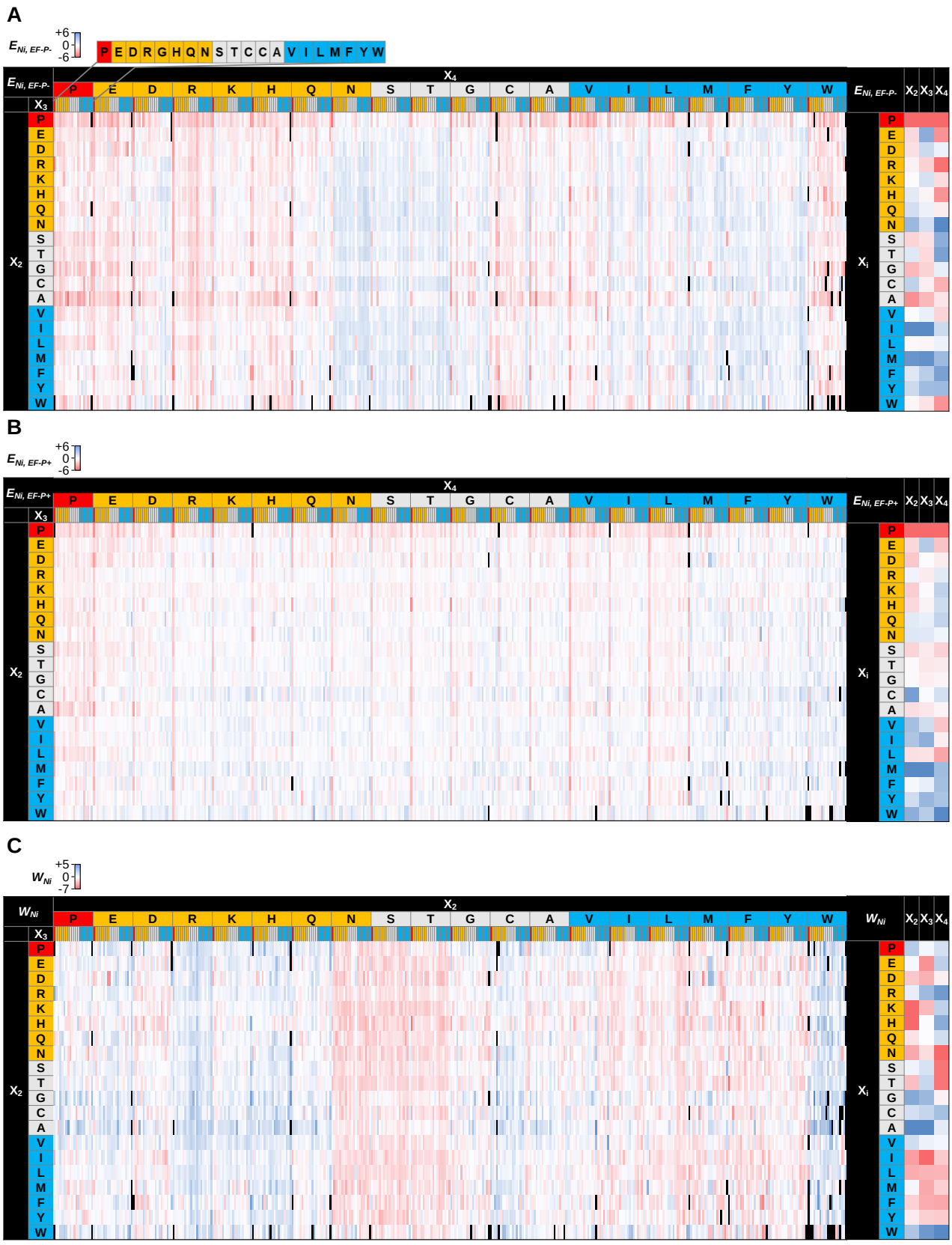
Supplementary figure 3 (continued)
(M)-(O) Heat maps of (M) $E_{Ni,\ EF\text{-}P\text{-}}$, (N) $E_{Ni,\ EF\text{-}P+}$, and (O) W_{Ni}. against $^{Bio}FX_4$ with 10 μM EF-G.
(P)-(R) Heat maps of (P) $E_{Ni,\ EF\text{-}P\text{-}}$, (Q) $E_{Ni,\ EF\text{-}P+}$, and (R) W_{Ni}. against $^{Bio}FX_3X_4$ with 10 μM EF-G.
Peptide sequences which were not detected by the next generation sequencer are filled with black.



Supplementary Figure 4 Profiling of the nascent peptide $^{Bio}FX_2X_3X_4$ for tri-proline incorporation with 0.03 μM EF-G.

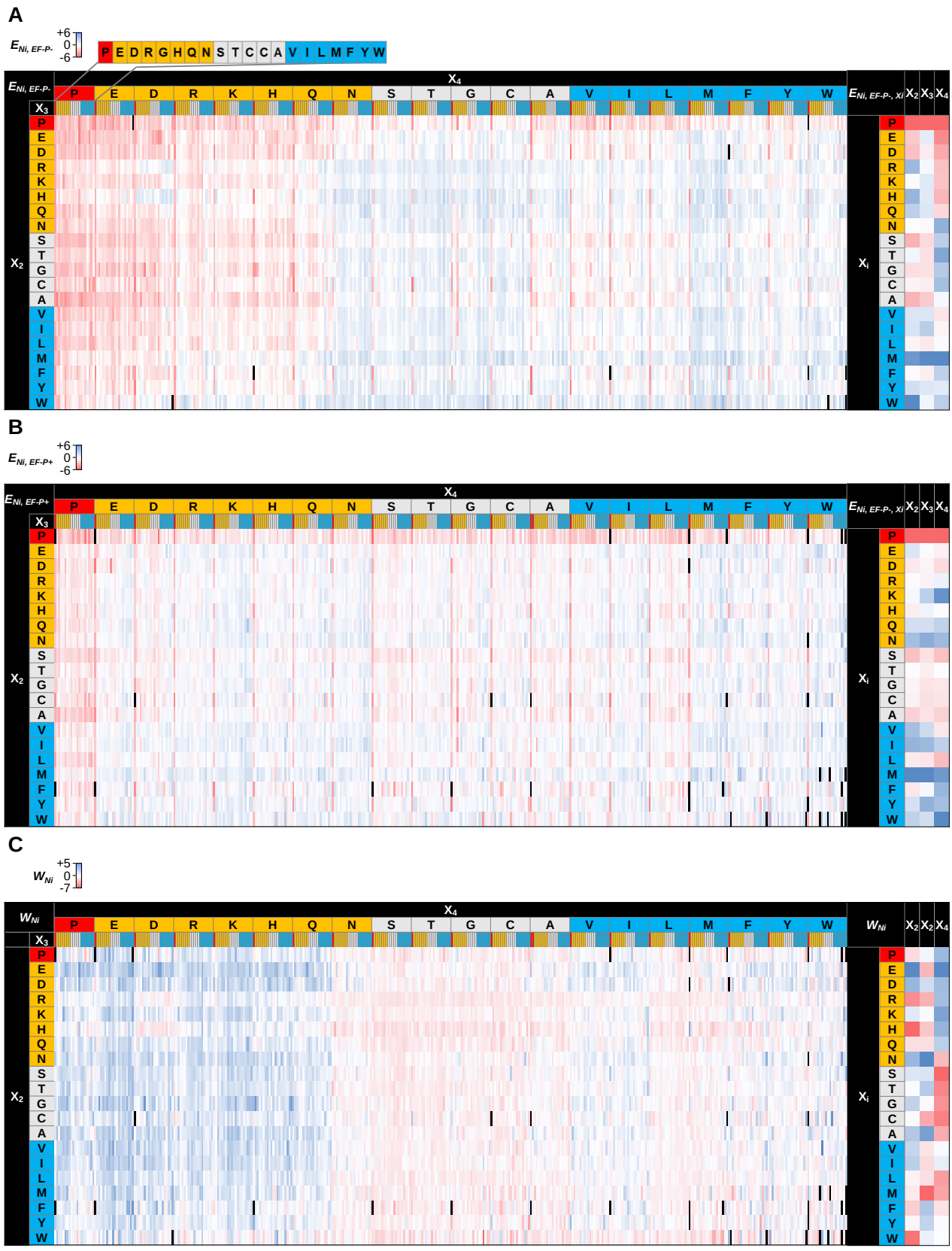
(A) Heat map of $E_{Ni,\ EF\text{-}P\text{-}}$, (B) Heat map of $E_{Ni,\ EF\text{-}P+}$, (C) Heat map of W_{Ni}.

Peptide sequences which were not detected by the next generation sequencer are filled with black.



Supplementary Figure 5 Profiling of the nascent peptide $^{Bio}FX_2X_3X_4$ for tri-proline incorporation with 0.26 μM EF-G.

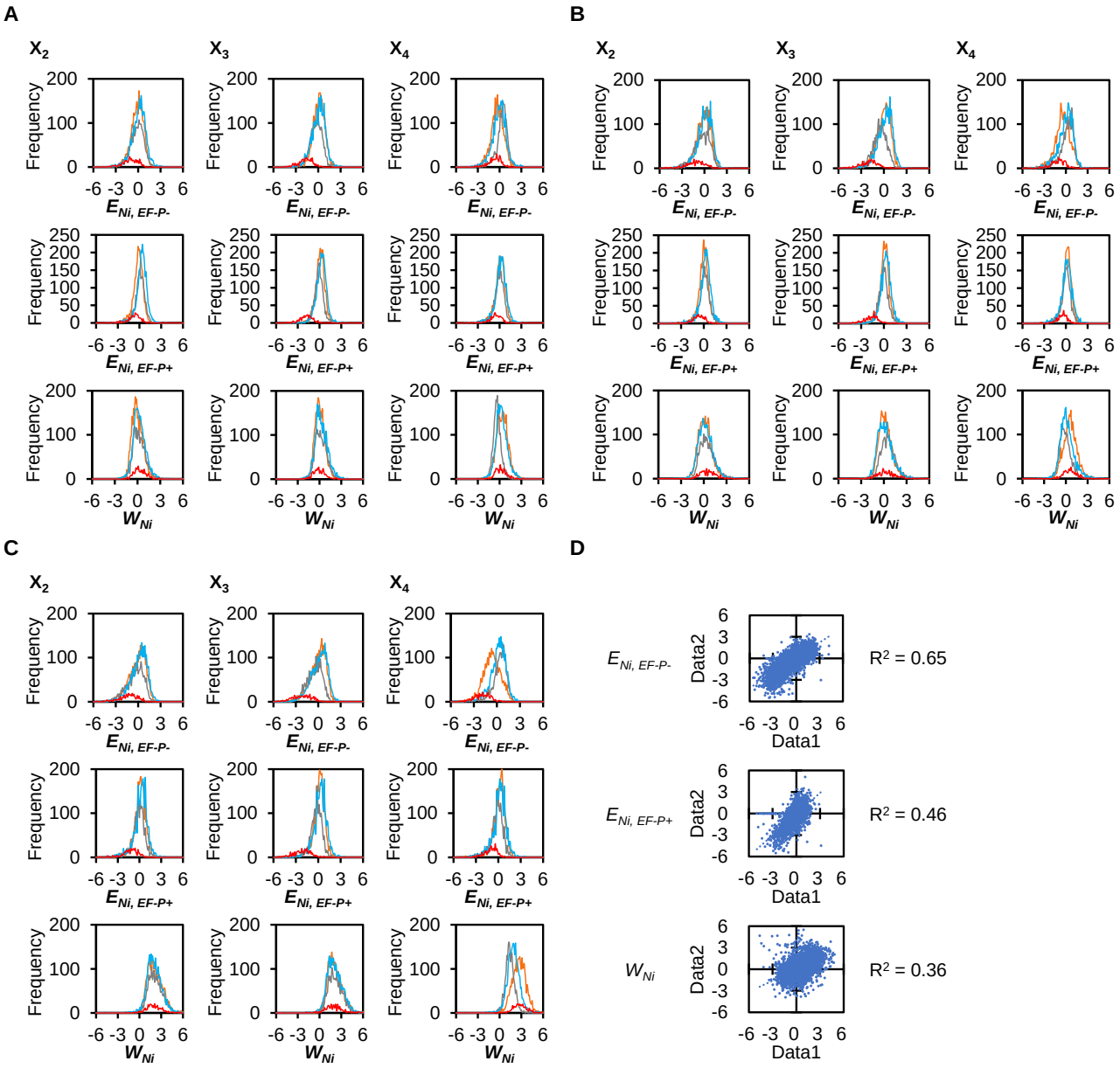
(A) Heat map of $E_{Ni,\ EF\text{-}P\text{-}}$, (B) Heat map of $E_{Ni,\ EF\text{-}P+}$, (C) Heat map of W_{Ni}.
Peptide sequences which were not detected by the next generation sequencer are filled with black.



Supplementary Figure 6 Profiling of the nascent peptide $^{Bio}FX_2X_3X_4$ for tri-proline incorporation with 10 μM EF-G.

(A) Heat map of $E_{Ni,\ EF\text{-}P\text{-}}$, (B) Heat map of $E_{Ni,\ EF\text{-}P+}$ (C) Heat map of W_{Ni}

Peptide sequences which were not detected by the next generation sequencer are filled with black.



Supplementary Figure 7 Histograms of enrichment values of profiling of nascent peptides for the three consecutive proline incorporation.

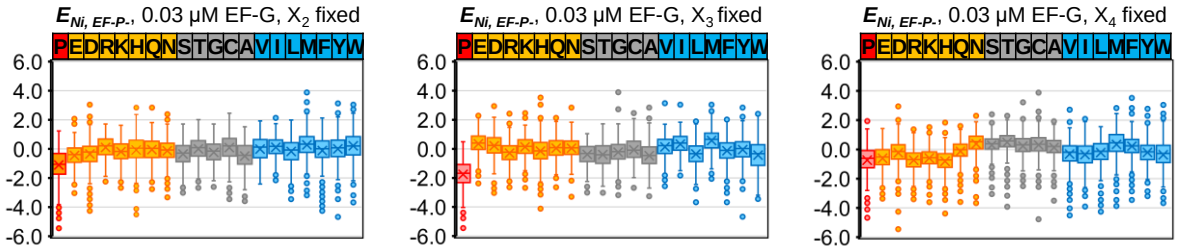
(A) Histograms of enrichment values in the presence of 0.03 μM EF-G. The histograms were sorted in terms the polarity of one of three amino acids of the nascent peptides, orange: Polar residues (DERKHQN), grey: small residues (STGCA), sky-blue: hydrophobic residues (VILMFYW), and red: proline (P)] at X_2, X_3, and X_4 positions.

(B) Histograms of enrichment values in the presence of 0.26 μM EF-G.

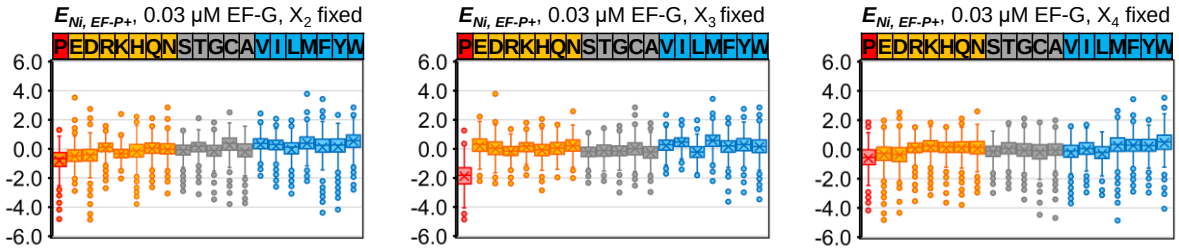
(C) Histograms of enrichment values in the presence of 10 μM EF-G.

(D) Reproducibility of nascent peptide profiling system in the presence of 10 μM EF-G. $E_{Ni,\ EF\text{-}P\text{-}}$, $E_{Ni,\ EF\text{-}P\text{+}}$, and W_{Ni} from two independent experiments were plotted and their correlation were calculated.

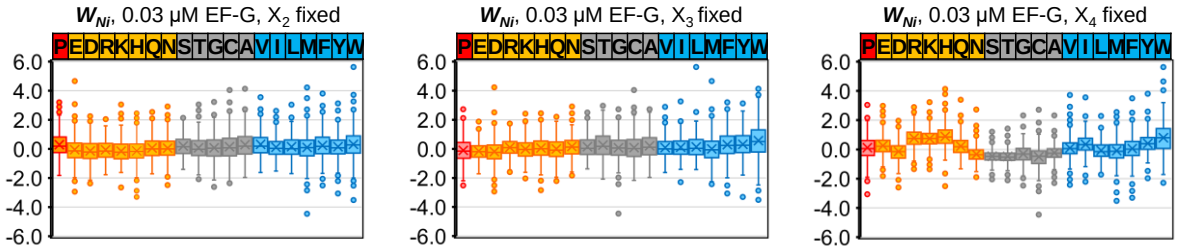
A


$E_{Ni,\ EF\text{-}P\text{-}}$, 0.03 μM EF-G, X_2 fixed
$E_{Ni,\ EF\text{-}P\text{-}}$, 0.03 μM EF-G, X_3 fixed
$E_{Ni,\ EF\text{-}P\text{-}}$, 0.03 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

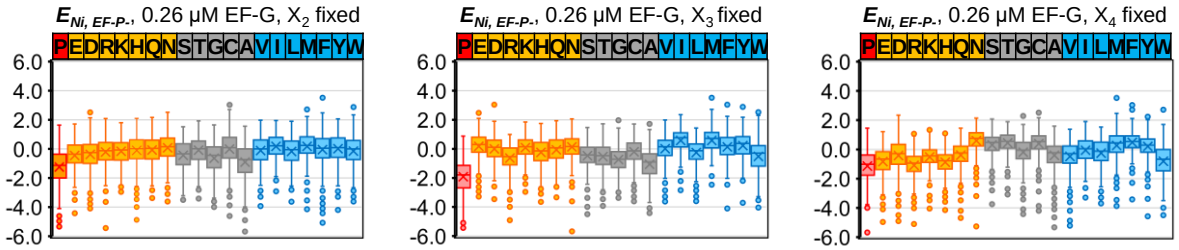
B


$E_{Ni,\ EF\text{-}P+}$, 0.03 μM EF-G, X_2 fixed
$E_{Ni,\ EF\text{-}P+}$, 0.03 μM EF-G, X_3 fixed
$E_{Ni,\ EF\text{-}P+}$, 0.03 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

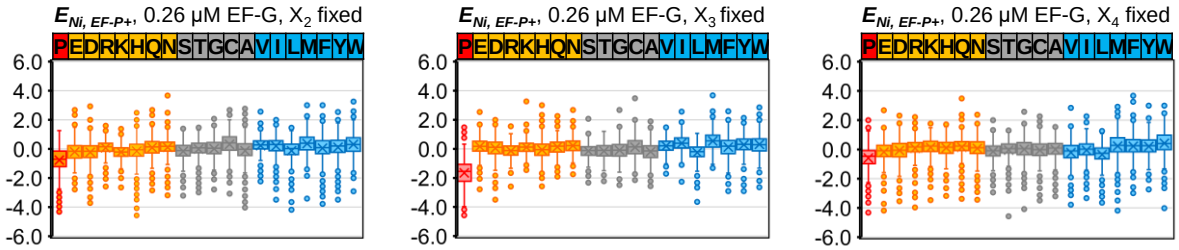
C


W_{Ni}, 0.03 μM EF-G, X_2 fixed
W_{Ni}, 0.03 μM EF-G, X_3 fixed
W_{Ni}, 0.03 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

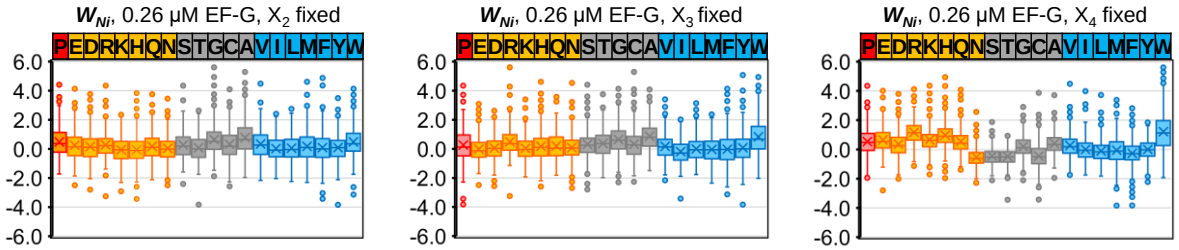
D

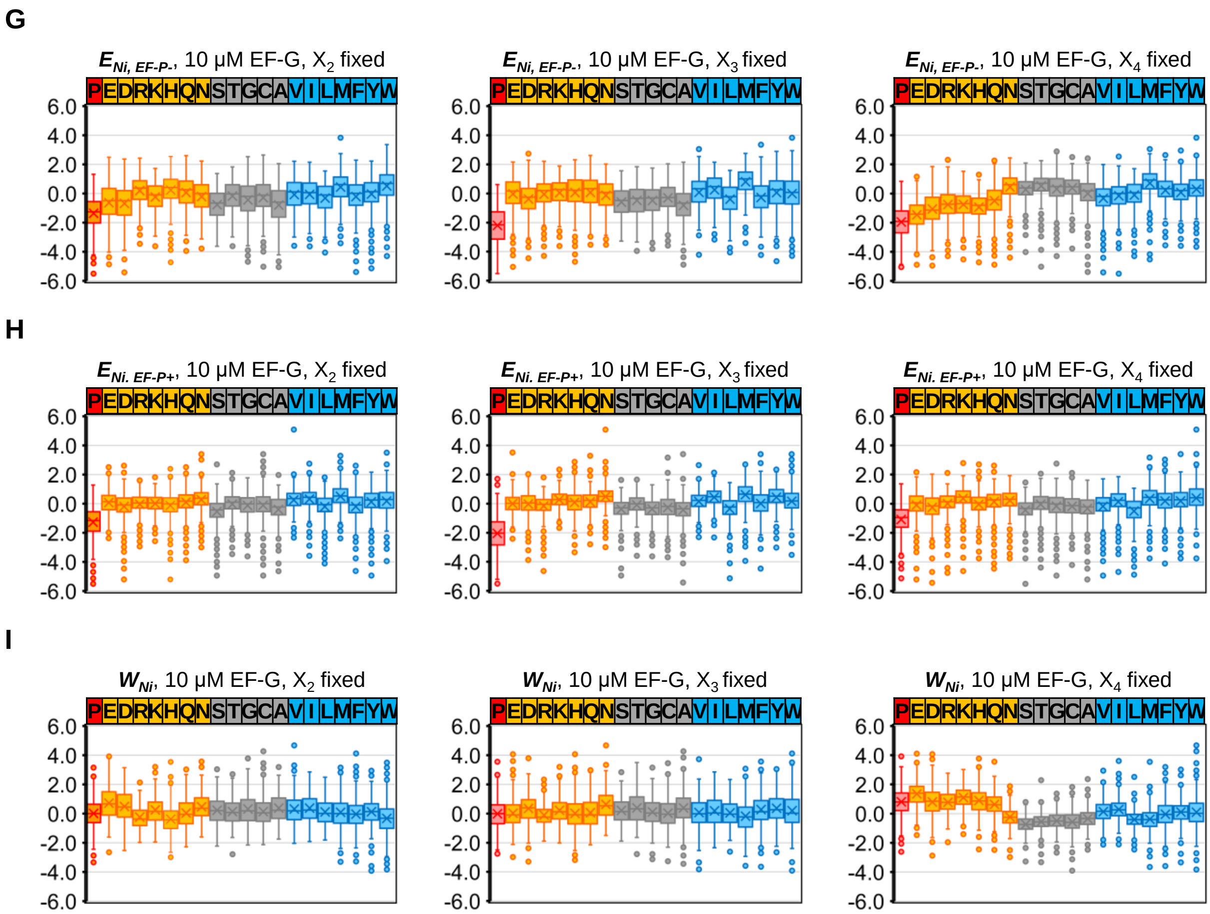

$E_{Ni,\ EF\text{-}P\text{-}}$, 0.26 μM EF-G, X_2 fixed
$E_{Ni,\ EF\text{-}P\text{-}}$, 0.26 μM EF-G, X_3 fixed
$E_{Ni,\ EF\text{-}P\text{-}}$, 0.26 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

E


$E_{Ni,\ EF\text{-}P+}$, 0.26 μM EF-G, X_2 fixed
$E_{Ni,\ EF\text{-}P+}$, 0.26 μM EF-G, X_3 fixed
$E_{Ni,\ EF\text{-}P+}$, 0.26 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

F


W_{Ni}, 0.26 μM EF-G, X_2 fixed
W_{Ni}, 0.26 μM EF-G, X_3 fixed
W_{Ni}, 0.26 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0



Supplementary Figure 8 Box plots of enrichment values of nascent peptide profiling for the three consecutive proline incorporation.

(E) Box plots of $E_{Ni,\ EF\text{-}P-}$ with 0.03 μM EF-G for 8420 kinds of nascent peptide sequences sorted by 20 kinds of proteinogenic amino acids at X_2 (Left), X_3 (Middle), and X_4 (Right). The upper and lower lines of boxes indicate the upper and lower quartiles respectively and upper and lower whiskers indicate the maximum and minimum values respectively. Any points outside those lines or whiskers are considered as outliers. The horizontal lien and cross in a box indicate the average and median, respectively.

(F) Box plots of $E_{Ni,\ EF\text{-}P+}$ with 0.03 μM EF-G.

(G) Box plots of W_{Ni} with 0.03 μM EF-G.

(H) Box plots of $E_{Ni,\ EF\text{-}P-}$ with 0.26 μM EF-G.

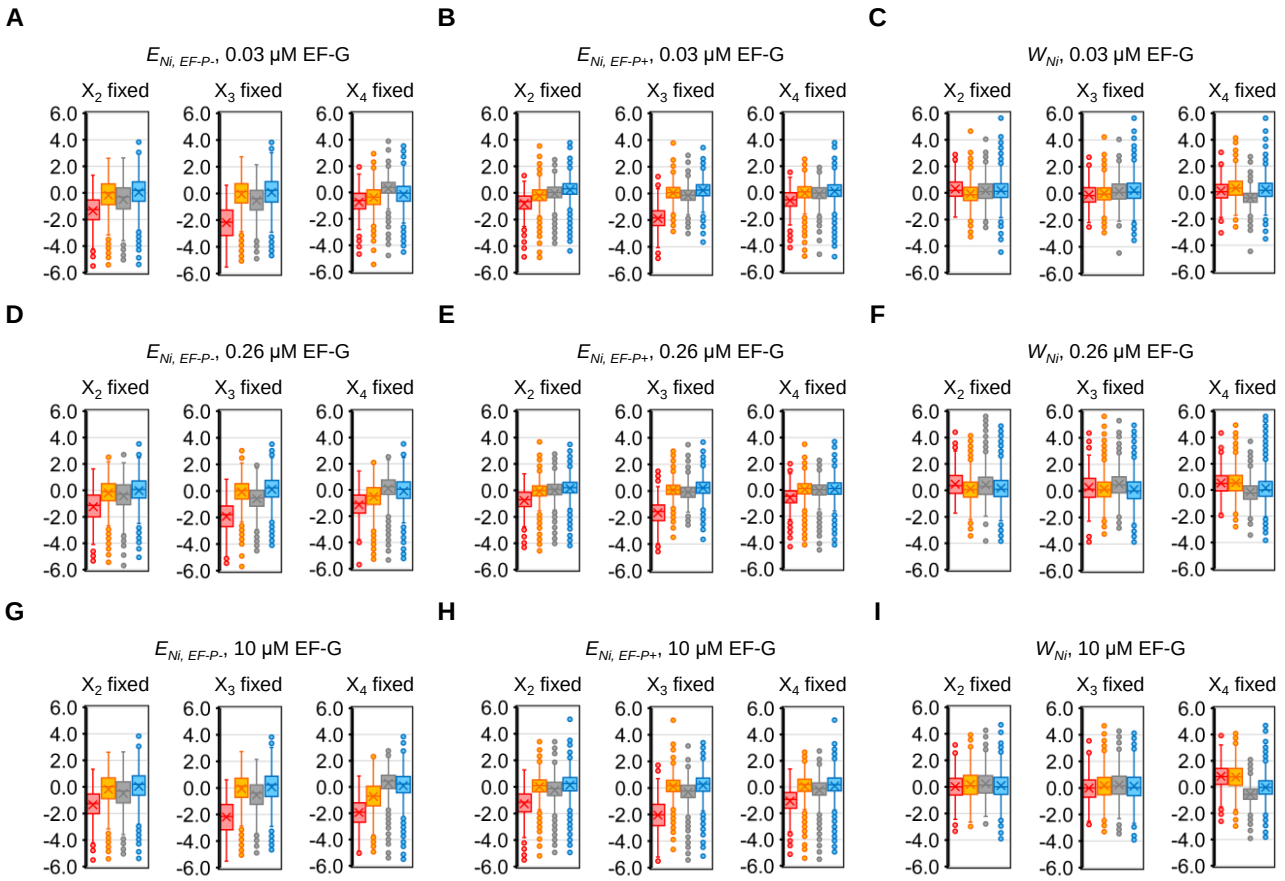
(I) Box plots of $E_{Ni,\ EF\text{-}P+}$ with 0.26 μM EF-G.

(J) Box plots of W_{Ni} with 0.03 μM EF-G.

(K) Box plots of $E_{Ni,\ EF\text{-}P-}$ with 10 μM EF-G.

(L) Box plots of $E_{Ni,\ EF\text{-}P+}$ with 10 μM EF-G.

(M) Box plots of W_{Ni} with 10 μM EF-G.



Supplementary Figure 9 Box plots of enrichment values of nascent peptide profiling for the three consecutive proline incorporation sorted by the polarity of the amino acids.

(A) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 0.03 μM EF-G for 8420 kinds of nascent peptide sequences sorted by 4 groups [Proline (Red), Polar residues (DERKHQN, Orange), Small residues (STGCA, Grey), Hydrophobic residues (VILMFYW, Sky-blue)] in terms of polarity of amino acids at X_2 (Left), X_3 (Middle), and X_4 (Right). The upper and lower lines of boxes indicate the upper and lower quartiles respectively and upper and lower whiskers indicate the maximum and minimum values respectively. Any points outside those lines or whiskers are considered as outliers. The horizontal lien and cross in a box indicate the average and median, respectively.

(B) Box plots of $E_{Ni,\ EF\text{-}P\text{+}}$ with 0.03 μM EF-G.

(C) Box plots of W_{Ni} with 0.03 μM EF-G.

(D) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 0.26 μM EF-G.

(E) Box plots of $E_{Ni,\ EF\text{-}P\text{+}}$ with 0.26 μM EF-G.

(F) Box plots of W_{Ni} with 0.26 μM EF-G.

(G) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 10 μM EF-G.

(H) Box plots of $E_{Ni,\ EF\text{-}P\text{+}}$ with 10 μM EF-G.

(I) Box plots of W_{Ni} with 10 μM EF-G.

A

mRNAS##:	AUG	NNN	NNN	NNN	CCG	CCG	CCG	AGC	GGU	AAC	(flag)	UAA
Full-length peptide##:	^{Bio}F	X_2	X_3	X_4	P	P	P	S	G	N	FLAG	(Stop)
C-terminal peptide 2 (CTP2):							P	S	G	N	FLAG	

B

mRNAS03 MWW
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.857
[FLPS03+3H]$^{3+}$ 813.673 814.026

mRNAS09 HEN
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.857
[FLPS09+3H]$^{3+}$ 772.655 772.675

mRNAS15 KKK
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.752
[FLPS15+3H]$^{3+}$ 774.035 N.D.

mRNAS21 AAD
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.761
[FLPS21+3H]$^{3+}$ 731.641 N.D.

mRNAS04 WYG
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.866
[FLPS04+3H]$^{3+}$ 781.328 781.346

mRNAS10 HYN
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.857
[FLPS10+3H]$^{3+}$ 783.995 783.990

mRNAS16 NKE
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.761
[FLPS16+3H]$^{3+}$ 769.667 N.D.

mRNAS22 EVD
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.752
[FLPS22+3H]$^{3+}$ 760.319 N.D.

mRNAS05 WMG
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.870
[FLPS05+3H]$^{3+}$ 770.654 770.674

mRNAS11 HIG
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.866
[FLPS11+3H]$^{3+}$ 748.328 748.329

mRNAS17 GAE
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.853
[FLPS17+3H]$^{3+}$ 731.641 N.D.

mRNAS23 EAD
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.757
[FLPS23+3H]$^{3+}$ 750.976 N.D.

mRNAS06 CWF
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.875
[FLPS06+3H]$^{3+}$ 791.326 791.349

mRNAS12 LIF
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.761
[FLPS12+3H]$^{3+}$ 770.352 770.186

mRNAS18 DSE
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.840
[FLPS18+3H]$^{3+}$ 756.307 N.D.

mRNAS24 GHD
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: N.D.
[FLPS24+4H]$^{4+}$ 562.236 561.65

mRNAS07 MMI
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.870
[FLPS07+3H]$^{3+}$ 770.995 771.015

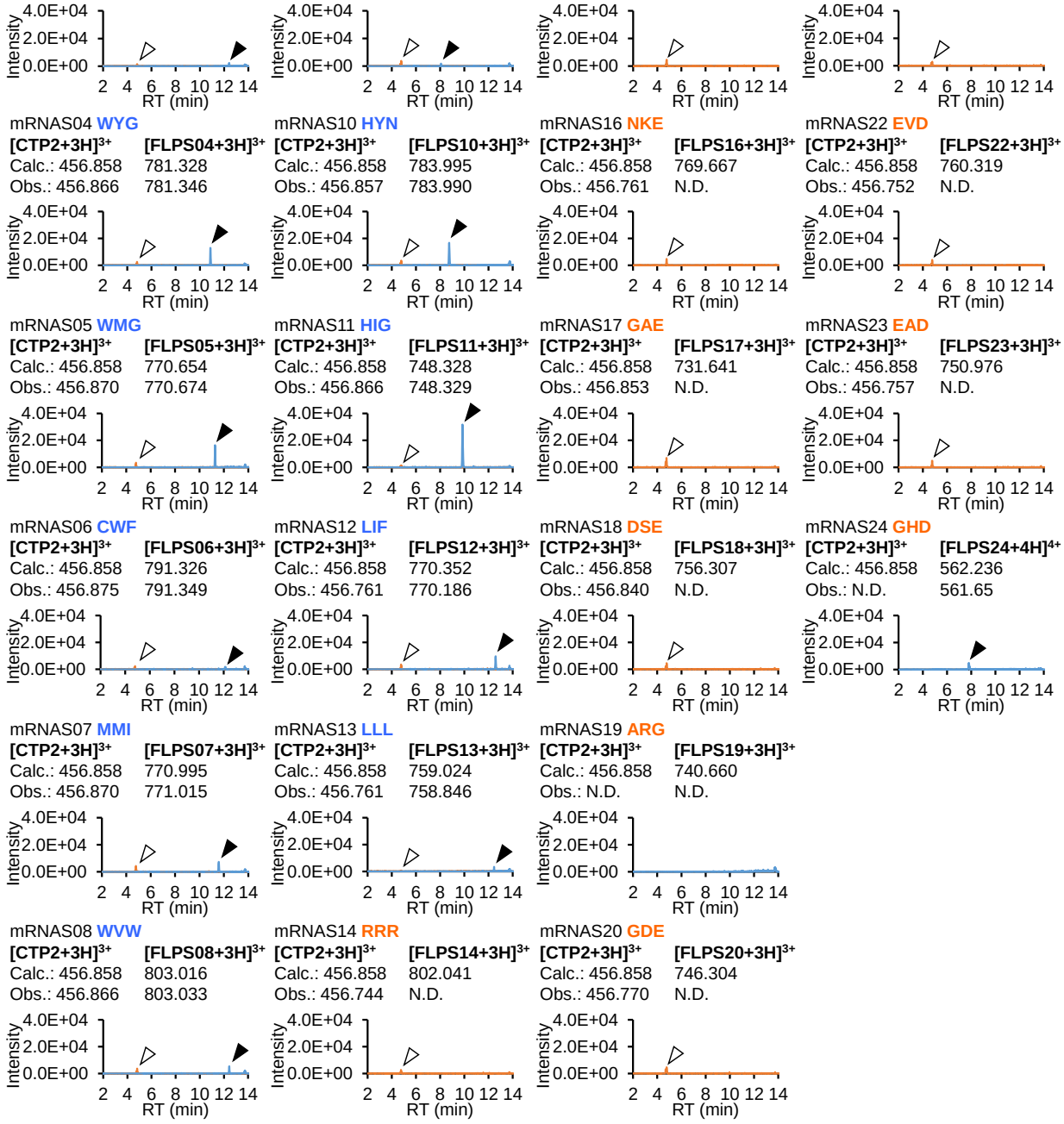
mRNAS13 LLL
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.761
[FLPS13+3H]$^{3+}$ 759.024 758.846

mRNAS19 ARG
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: N.D.
[FLPS19+3H]$^{3+}$ 740.660 N.D.

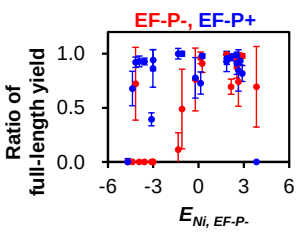
mRNAS08 WVW
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.866
[FLPS08+3H]$^{3+}$ 803.016 803.033

mRNAS14 RRR
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.744
[FLPS14+3H]$^{3+}$ 802.041 N.D.

mRNAS20 GDE
[CTP2+3H]$^{3+}$ Calc.: 456.858 Obs.: 456.770
[FLPS20+3H]$^{3+}$ 746.304 N.D.



C



D

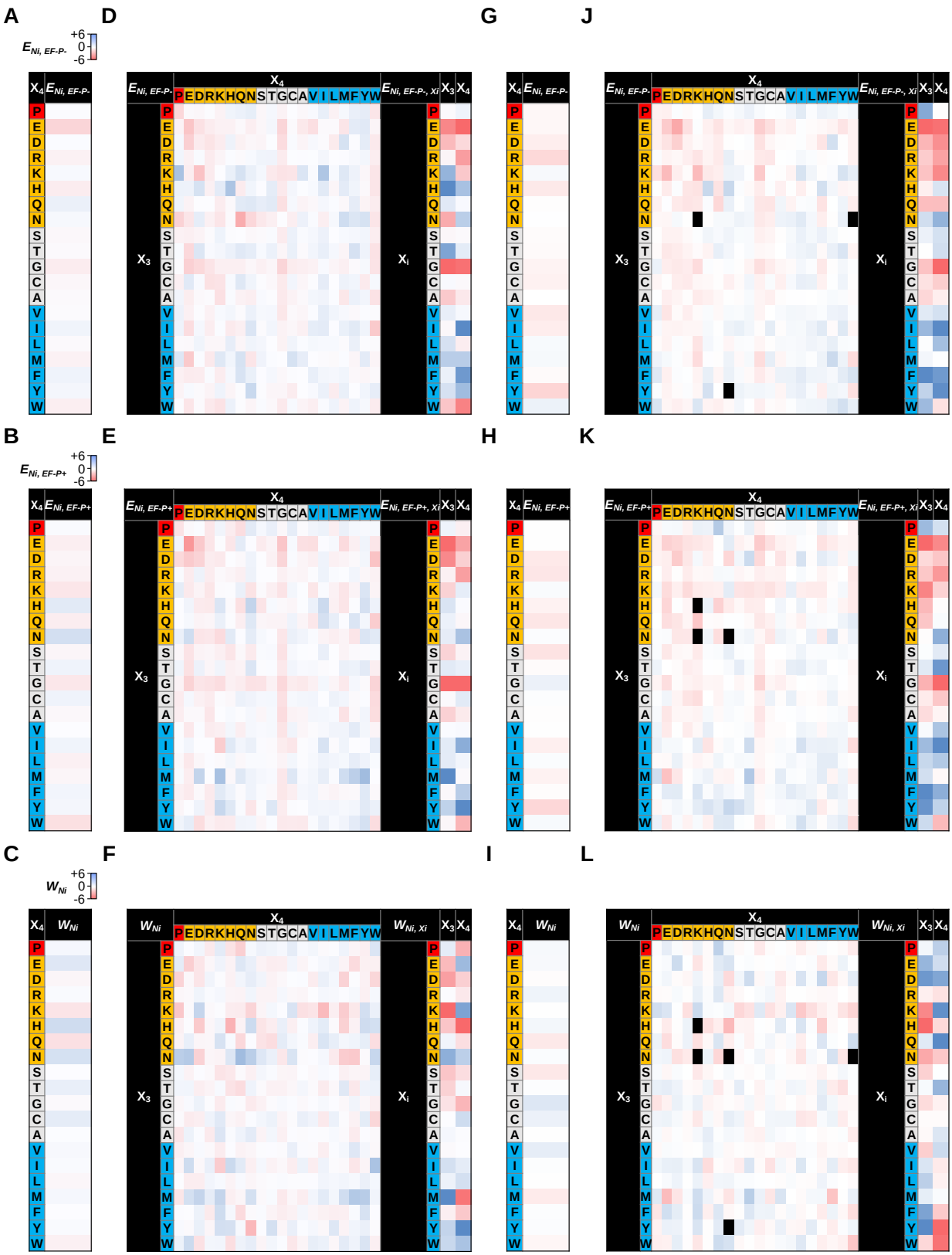
Supplementary Figure 10 XICs of full-length peptides and a C-terminal peptide expressed by incorporation of three consecutive prolines into representative nascent peptide.

(A) mRNA and peptide sequences used for the individual expression.

(B) XICs of peptides expressed with 10 μM EF-G and 0 μM EF-P. Sky-blue and orange XICs were extracted by using *m/z* values of full-length peptides and C-terminal peptide 2 (CTP2). Closed triangles and open triangles indicate peaks corresponding to the full-length peptides CTP2. N.D.: Not detected.

(C) Correlation between deep sequencing results of profiling system and quantification of LC-ESI MS

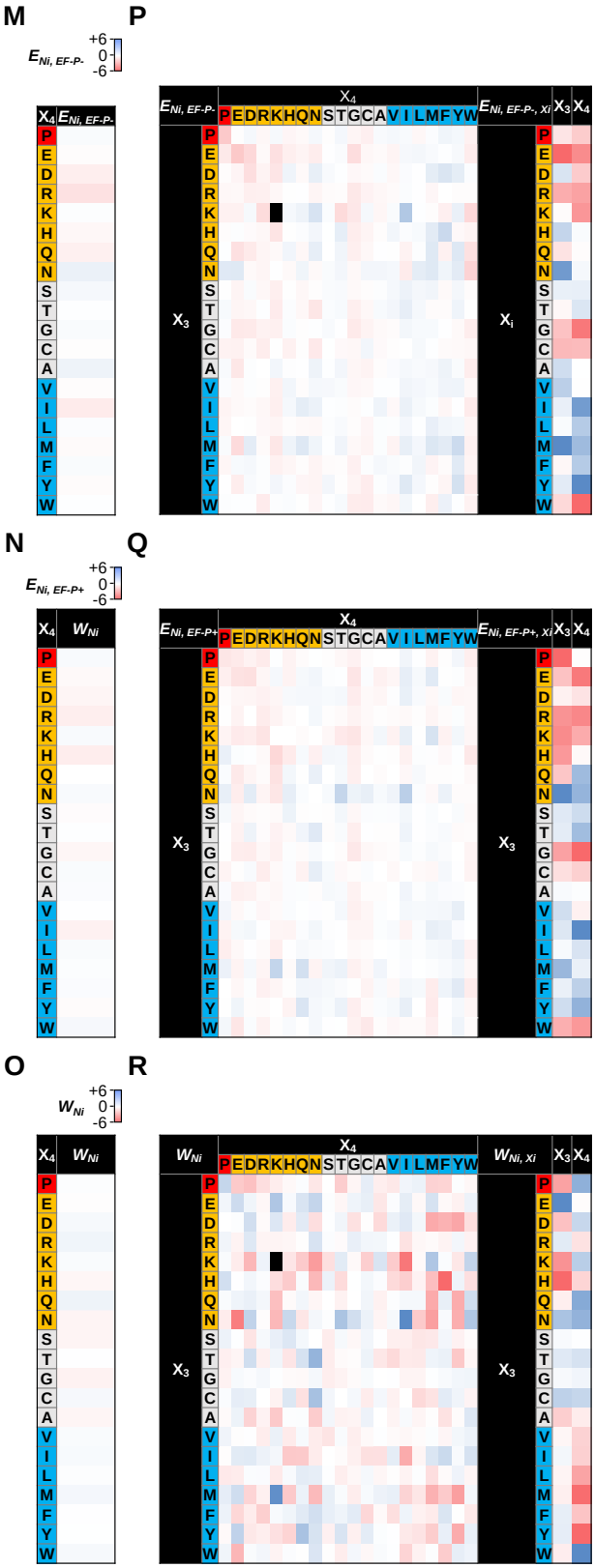
(D) XICs of peptides expressed in the presence of EF-P and 10 μM EF-G.



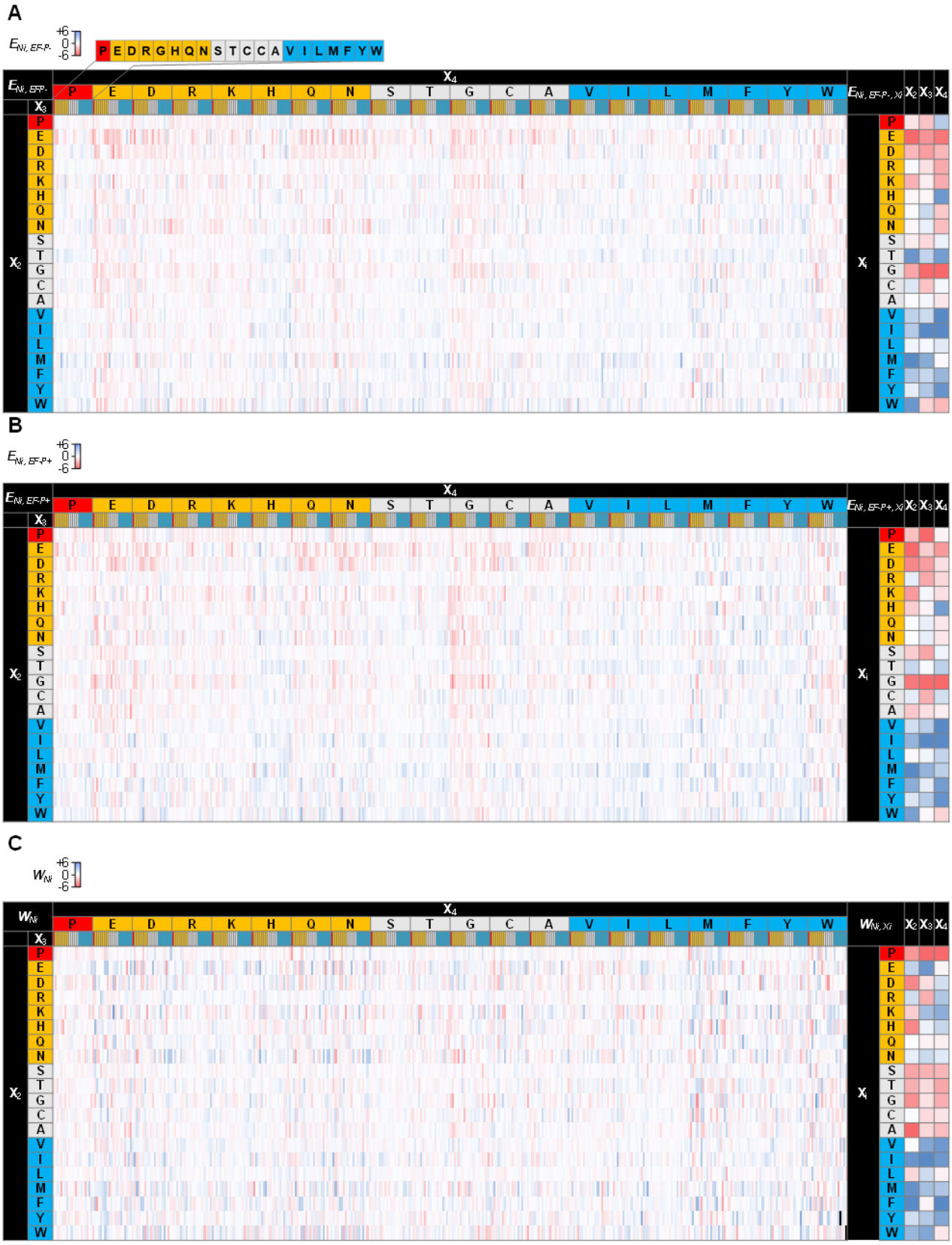
Supplementary Figure 11 Profiling of nascent peptide $^{Bio}FX_4$ and $^{Bio}FX_3X_4$ for tri-glycine incorporation.

(A)-(C) Heat maps of (A) $E_{Ni,\ EF\text{-}P\text{-}}$, (B) $E_{Ni,\ EF\text{-}P+}$, and (C) W_{Ni}. against $^{Bio}FX_4$ with 0.03 μM EF-G.
(D)-(f) Heat maps of (D) $E_{Ni,\ EF\text{-}P\text{-}}$, (E) $E_{Ni,\ EF\text{-}P+}$, and (F) W_{Ni}. against $^{Bio}FX_3X_4$ with 0.03 μM EF-G.
(G)-(I) Heat maps of (G) $E_{Ni,\ EF\text{-}P\text{-}}$, (H) $E_{Ni,\ EF\text{-}P+}$, and (I) W_{Ni}. against $^{Bio}FX_4$ with 0.26 μM EF-G.
(J)-(L) Heat maps of (J) $E_{Ni,\ EF\text{-}P\text{-}}$, (K) $E_{Ni,\ EF\text{-}P+}$, and (L) W_{Ni}. against $^{Bio}FX_3X_4$ with 0.26 μM EF-G.

Peptide sequences which were not detected by the next generation sequencer are filled with black.



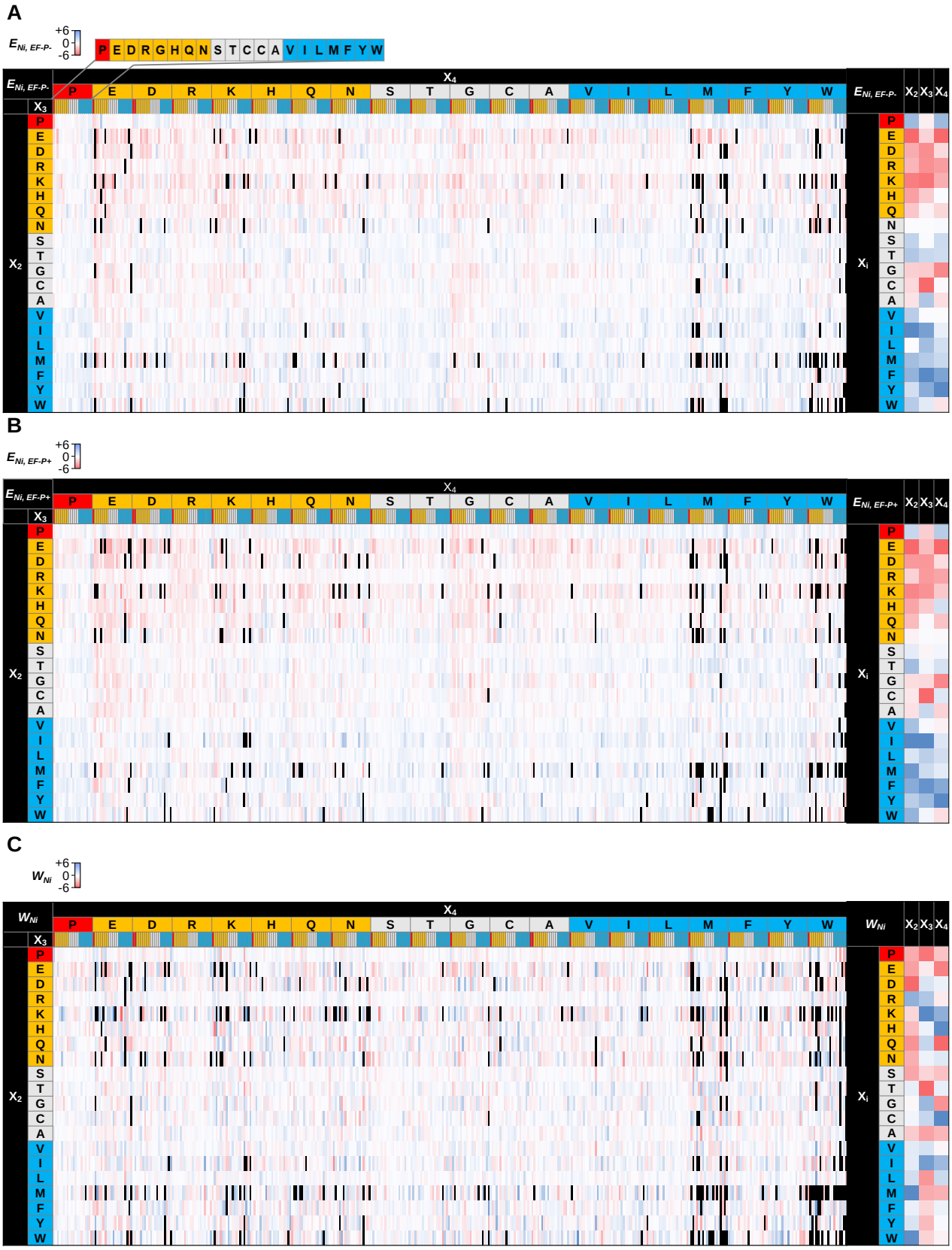
Supplementary figure 11 (continued)
(M)-(O) Heat maps of (M) $E_{Ni,\ EF\text{-}P\text{-}}$, (N) $E_{Ni,\ EF\text{-}P+}$, and (O) W_{Ni}. against $^{Bio}FX_4$ with 10 μM EF-G.
(P)-(R) Heat maps of (P) $E_{Ni,\ EF\text{-}P\text{-}}$, (Q) $E_{Ni,\ EF\text{-}P+}$, and (R) W_{Ni}. against $^{Bio}FX_3X_4$ with 10 μM EF-G.
Peptide sequences which were not detected by the next generation sequencer are filled with black.



Supplementary Figure 12 Profiling of the nascent peptide $^{Bio}FX_2X_3X_4$ for tri-glycine incorporation with 0.03 μM EF-G.

(A) Heat map of $E_{Ni,\ EF\text{-}P\text{-}}$, (B) Heat map of $E_{Ni,\ EF\text{-}P+}$ (C) Heat map of W_{Ni}

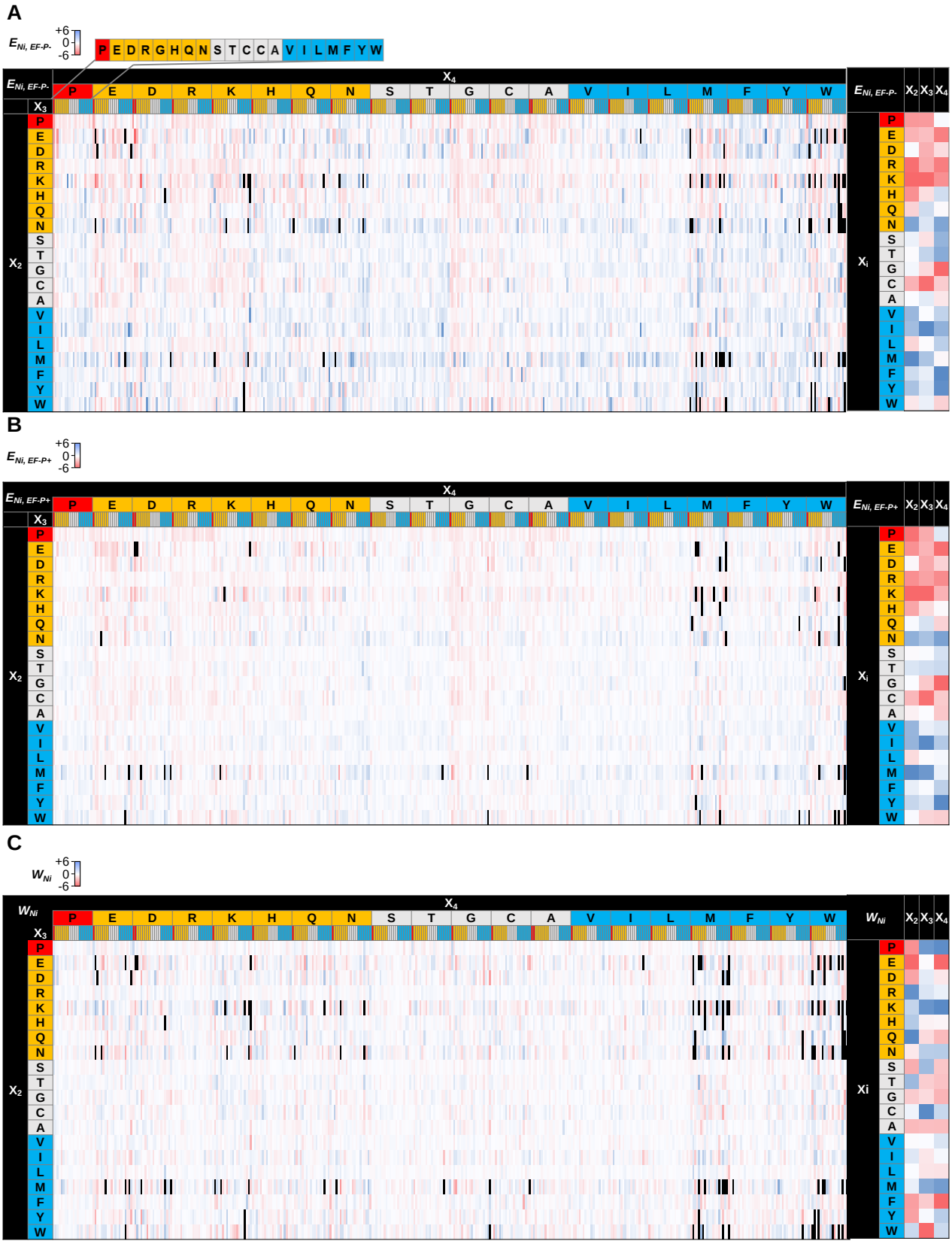
Peptide sequences which were not detected by the next generation sequencer are filled with black.



Supplementary Figure 13 Profiling of nascent peptide $^{Bio}FX_2X_3X_4$ for tri-glycine incorporation with 0.26 μM EF-G.

(B) Heat map of $E_{Ni,\ EF\text{-}P\text{-}}$, (B) Heat map of $E_{Ni,\ EF\text{-}P+}$ (C) Heat map of W_{Ni}

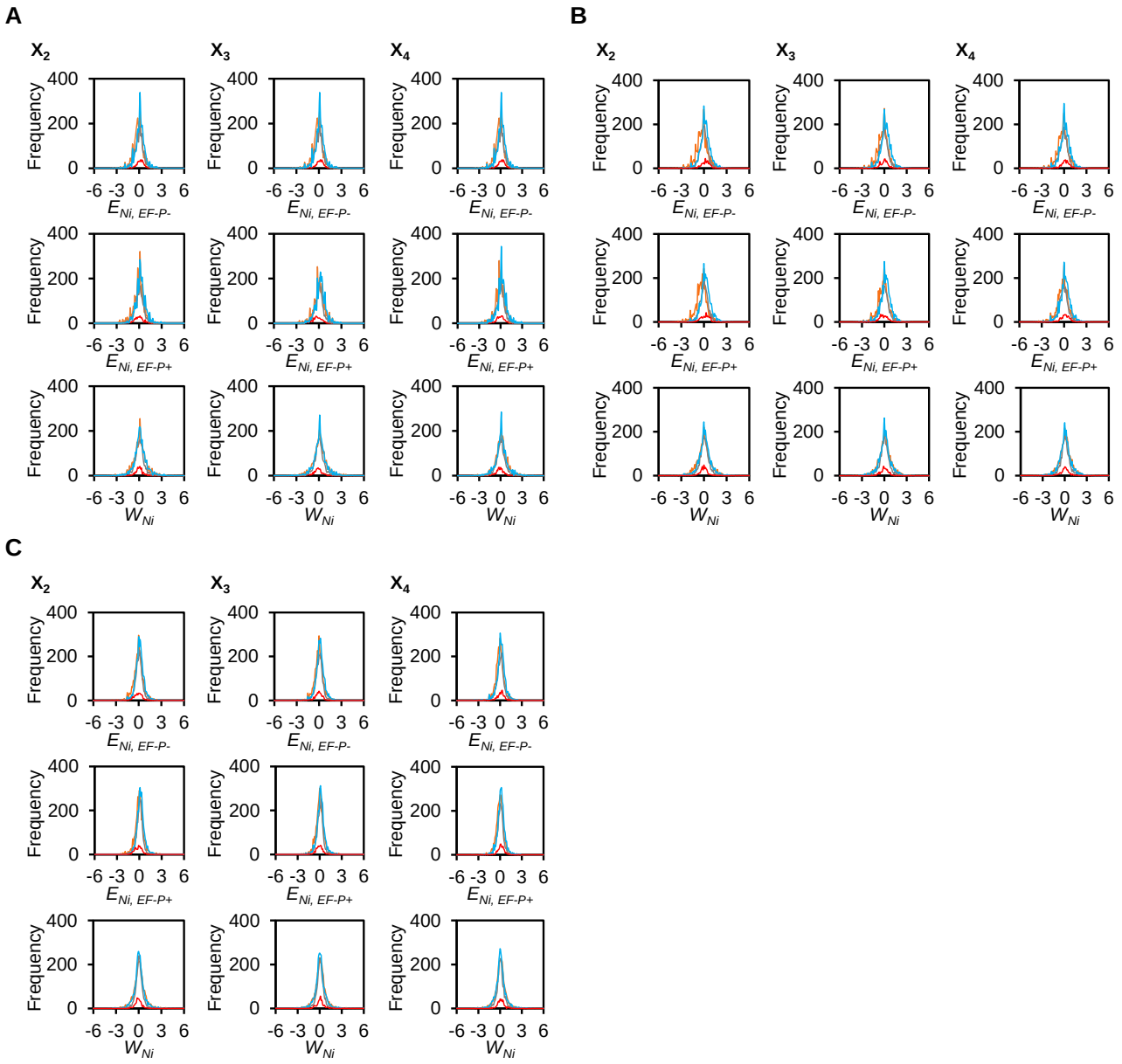
Peptide sequences which were not detected by the next generation sequencer are filled with black.



Supplementary Figure 14 Profiling of nascent peptide $^{Bio}FX_2X_3X_4$ for tri-glycine incorporation with 10 μM EF-G.

(C) Heat map of $E_{Ni,\ EF\text{-}P\text{-}}$, (B) Heat map of $E_{Ni,\ EF\text{-}P+}$ (C) Heat map of W_{Ni}

Peptide sequences which were not detected by the next generation sequencer are filled with black.



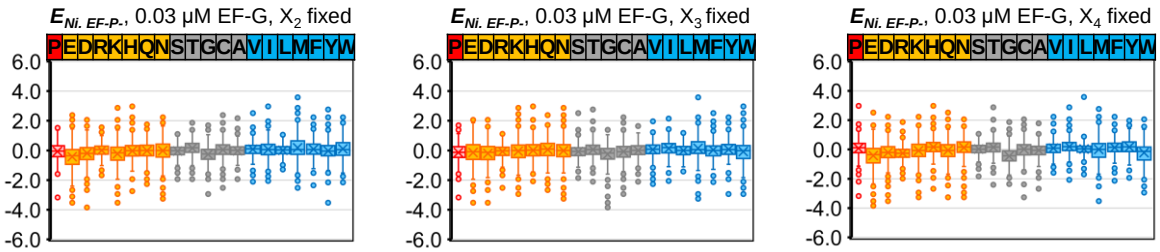
Supplementary Figure 15 Histograms of enrichment values of profiling of nascent peptide sequences for incorporation of three consecutive glycines.

(A) Histograms of enrichment values in the presence of 0.03 μM EF-G. The histograms were sorted in terms the polarity of one of three amino acids of the nascent peptides, orange: Polar residues (DERKHQN), grey: small residues (STGCA), sky-blue: hydrophobic residues (VILMFYW), and red: proline (P)] at X_2, X_3, and X_4 positions.

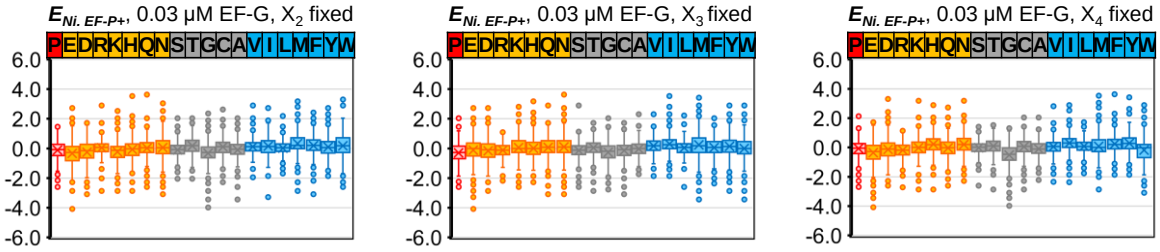
(B) Histograms of enrichment values in the presence of 0.26 μM EF-G.

(C) Histograms of enrichment values in the presence of 10 μM EF-G.

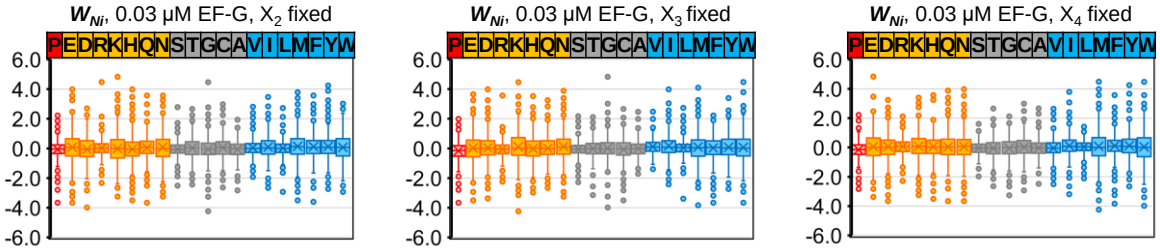
A


$E_{Ni.\ EF\text{-}P\text{-}}$, 0.03 μM EF-G, X_2 fixed
$E_{Ni.\ EF\text{-}P\text{-}}$, 0.03 μM EF-G, X_3 fixed
$E_{Ni.\ EF\text{-}P\text{-}}$, 0.03 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

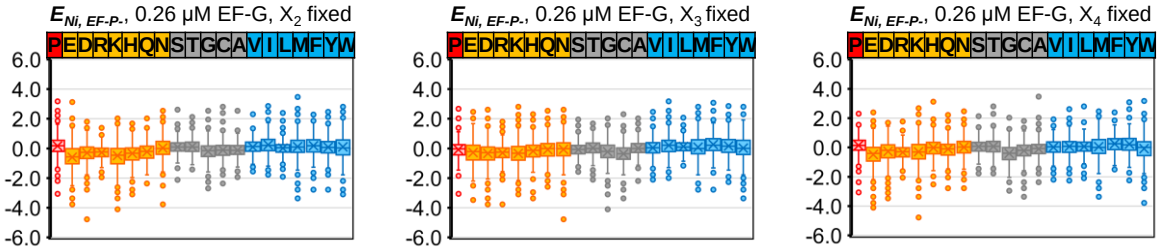
B


$E_{Ni.\ EF\text{-}P+}$, 0.03 μM EF-G, X_2 fixed
$E_{Ni.\ EF\text{-}P+}$, 0.03 μM EF-G, X_3 fixed
$E_{Ni.\ EF\text{-}P+}$, 0.03 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

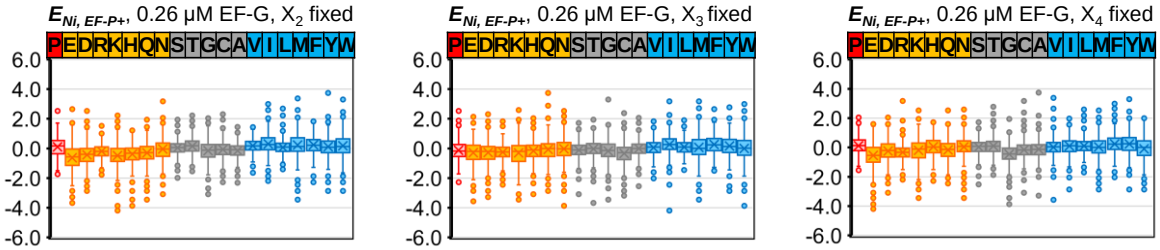
C


W_{Ni}, 0.03 μM EF-G, X_2 fixed
W_{Ni}, 0.03 μM EF-G, X_3 fixed
W_{Ni}, 0.03 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

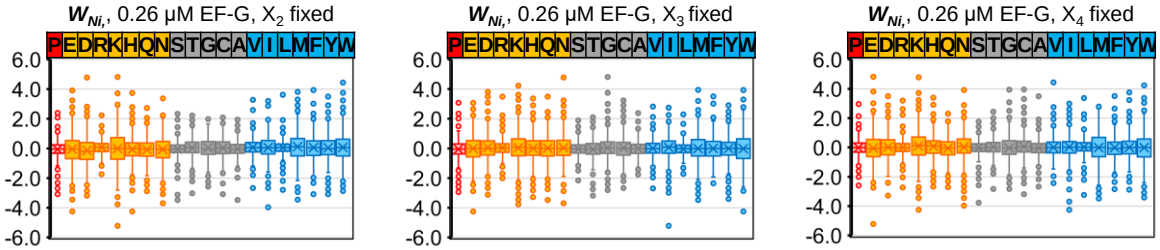
D

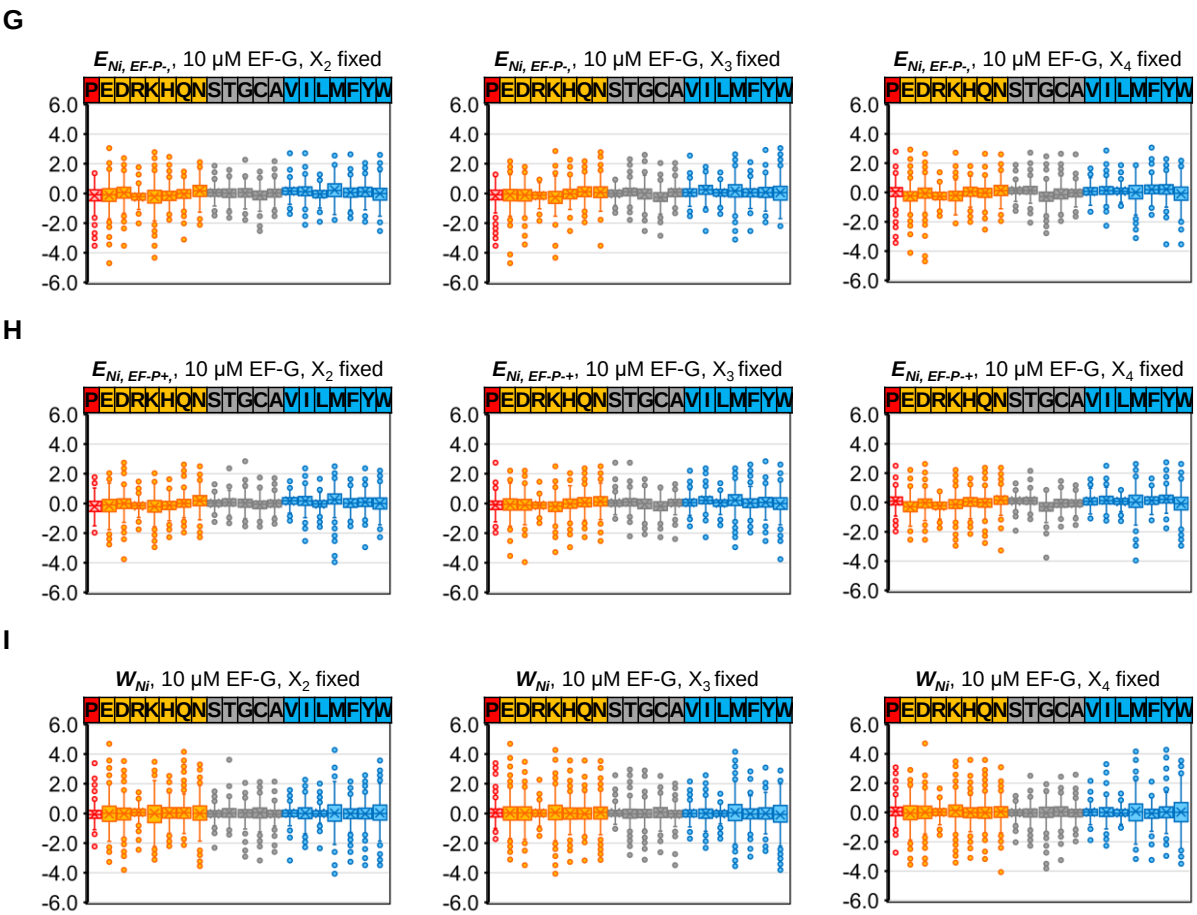

$E_{Ni,\ EF\text{-}P\text{-}}$, 0.26 μM EF-G, X_2 fixed
$E_{Ni,\ EF\text{-}P\text{-}}$, 0.26 μM EF-G, X_3 fixed
$E_{Ni,\ EF\text{-}P\text{-}}$, 0.26 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

E


$E_{Ni,\ EF\text{-}P+}$, 0.26 μM EF-G, X_2 fixed
$E_{Ni,\ EF\text{-}P+}$, 0.26 μM EF-G, X_3 fixed
$E_{Ni,\ EF\text{-}P+}$, 0.26 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0

F


$W_{Ni,}$, 0.26 μM EF-G, X_2 fixed
$W_{Ni,}$, 0.26 μM EF-G, X_3 fixed
$W_{Ni,}$, 0.26 μM EF-G, X_4 fixed
PEDRKHQNSTGCAVILMFYW
6.0
4.0
2.0
0.0
-2.0
-4.0
-6.0



Supplementary Figure 16 Box plots of enrichment values of nascent peptide profiling for the three consecutive glycine incorporation.

(A) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 0.03 μM EF-G for 8420 kinds of nascent peptide sequences sorted by 20 kinds of proteinogenic amino acids at X_2 (Left), X_3 (Middle), and X_4 (Right). The upper and lower lines of boxes indicate the upper and lower quartiles respectively and upper and lower whiskers indicate the maximum and minimum values respectively. Any points outside those lines or whiskers are considered as outliers. The horizontal lien and cross in a box indicate the average and median, respectively.

(B) Box plots of $E_{Ni,\ EF\text{-}P+}$ with 0.03 μM EF-G.

(C) Box plots of W_{Ni} with 0.03 μM EF-G.

(D) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 0.26 μM EF-G.

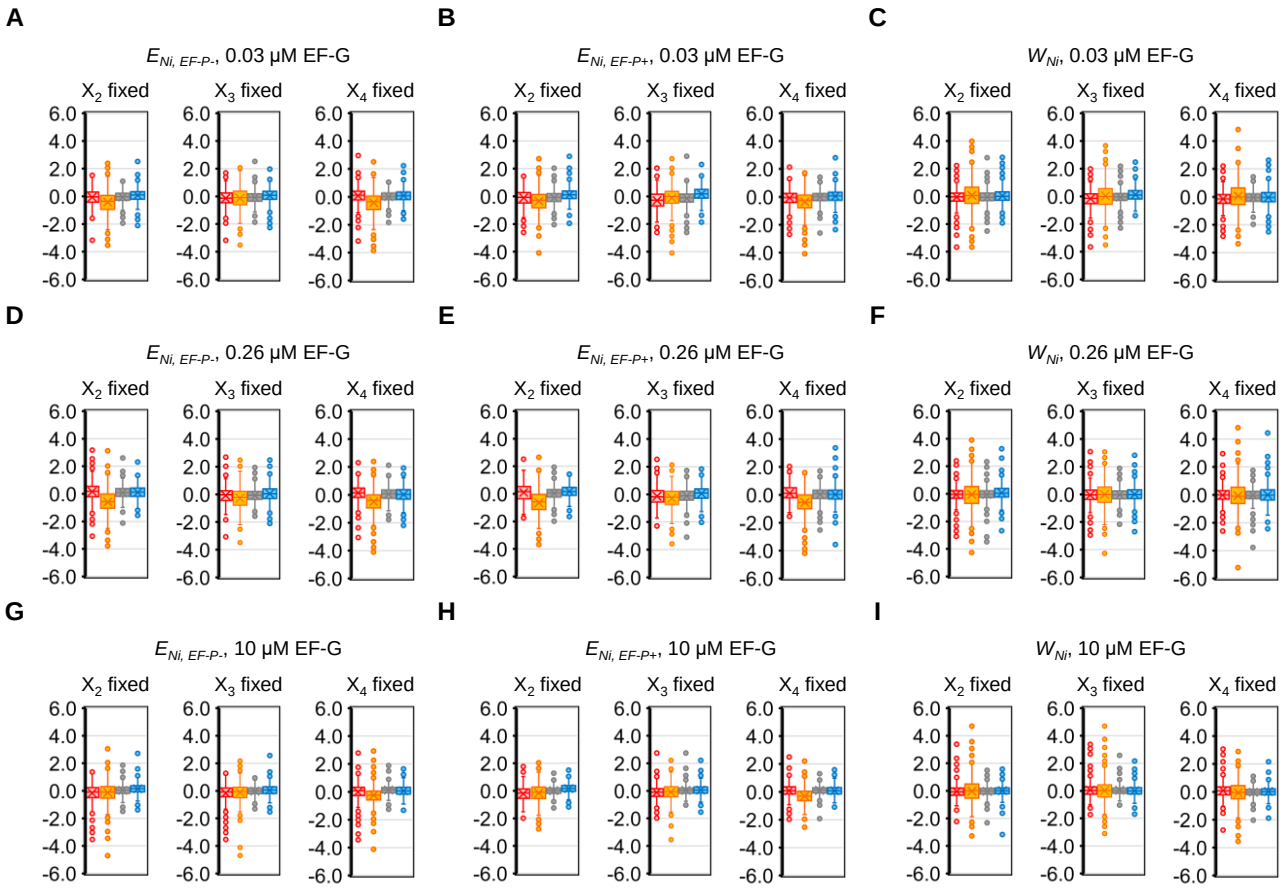
(E) Box plots of $E_{Ni,\ EF\text{-}P+}$ with 0.26 μM EF-G.

(F) Box plots of W_{Ni} with 0.03 μM EF-G.

(G) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 10 μM EF-G.

(H) Box plots of $E_{Ni,\ EF\text{-}P+}$ with 10 μM EF-G.

(I) Box plots of W_{Ni} with 10 μM EF-G.



Supplementary Figure 17 Box plots of enrichment values of nascent peptide profiling for the three consecutive glycine incorporation sorted by the polarity of the amino acids.

(A) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 0.03 μM EF-G for 8420 kinds of nascent peptide sequences sorted by 4 groups [Proline (Red), Polar residues (DERKHQN, Orange), Small residues (STGCA, Grey), Hydrophobic residues (VILMFYW, Sky-blue)] in terms of polarity of amino acids at X_2 (Left), X_3 (Middle), and X_4 (Right). The upper and lower lines of boxes indicate the upper and lower quartiles respectively and upper and lower whiskers indicate the maximum and minimum values respectively. Any points outside those lines or whiskers are considered as outliers. The horizontal lien and cross in a box indicate the average and median, respectively.

(B) Box plots of $E_{Ni,\ EF\text{-}P\text{+}}$ with 0.03 μM EF-G.

(C) Box plots of W_{Ni} with 0.03 μM EF-G.

(D) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 0.26 μM EF-G.

(E) Box plots of $E_{Ni,\ EF\text{-}P\text{+}}$ with 0.26 μM EF-G.

(F) Box plots of W_{Ni} with 0.26 μM EF-G.

(G) Box plots of $E_{Ni,\ EF\text{-}P\text{-}}$ with 10 μM EF-G.

(H) Box plots of $E_{Ni,\ EF\text{-}P\text{+}}$ with 10 μM EF-G.

(I) Box plots of W_{Ni} with 10 μM EF-G.

2.5. Methods

Preparation of cDNA templates, mRNA templates, Flexizymes, and tRNAs

cDNA templates, mRNA templates, Flexizymes (dFx and eFx), and tRNAs used for translation and pre-charging of activated amino acids were prepared by PCR and following *in vitro* transcription by T7 RNA polymerase. Template DNAs that have a T7 promoter prior to transcripts were prepared by extension of forward and reverse primer pairs, followed by PCR using forward and reverse PCR primers. DNA primer sequences and combination of primers used in extension and PCR reactions are summarized in the supplementary table. PCR was conducted in 1x PCR reaction mixture containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1v/v% Triton X-100, 2.5 mM MgCl2, 250 μM each dNTPs, 0.5 μM forward and reverse primers, and 1.5v/v% 1x Taq DNA polymerase with pH 9.0. The PCR products were extracted by phenol/chloroform, precipitated by ethanol, and used for *in vitro* transcription at 37◦C for at least 3 hours in a 100-2000 uL reaction mixture. For mRNA and Flexizymes, *in vitro* transcription reaction mixture contains 40 mM Tris-HCl, 1 mM spermidine, 0.01v/v% Triton X-100, 10 mM DTT, 30 mM $MgCl_2$, 5 mM NTPs, 30 mM KOH, 10v/v% template DNA solution, 0.12 μM T7 RNA polymerase with pH 8.0. For tRNAs, reaction mixture contains 40 mM Tris-HCl, 1 mM spermidine, 0.01v/v% Triton X-100, 10 mM DTT, 22.5 mM $MgCl_2$, 3.75 mM NTPs, 5 mM GMP 22.5 mM KOH, 10v/v% template DNA solution, 0.12 μM T7 RNA polymerase with pH 8.0. The resulting RNA transcripts were treated with RQ1 DNase (Promega) for 1 hour at 37◦C. The RNA transcripts were precipitated by adding 10%v/v 3M NaCl and twice volume of isopropanol followed by centrifuge. After wash with 70% EtOH, the RNA pellet was dissolved into H_2O with 10%v/v volume of *in vitro* transcription reaction and equivalent volume of 2x RNA loading buffer (8M urea, 2 mM $Na_2EDTA \cdot 2H_2O$, 2 mM Tris-HCl, pH 7.5) was added. After heating at 95◦C for 2 min, RNA was purified by 8% polyacrylamide gel containing 6 M urea. The RNA band was visualized by UV light, cut out, and extracted in the 0.3 M NaCl solution for at least 5 hours. After removal of gel, twice volume of EtOH was added and RNA was precipitated by centrifuge at 13000 rpm for 15 min. After wash with 70% EtOH and drying, 20 μM RNA solution was prepared by dissolving into H_2O.

Aminoacylation of tRNAs

Activated amino acids (L-proline-3,5-dinitrobenzyl ester (Pro-DBE), L-Lysine-3,5-dinitrobenzyl ester (Lys-DBE), L-asparagine-3,5-dinitrobenzyl ester (Asn-DBE), L-glutamate-3,5-dinitrobenzyl ester (Glu-DBE), L-leucine-3,5-dinitrobenzyl ester (Leu-DBE), L-isoleucine-3,5-dinitrobenzyl ester (Ile-DBE), L-phenylalanine-cyanomethyl ester (Phe-CME), and N-biotinylated-L-phenylalanine cyanomethyl ester (^{Bio}F-CME)) were synthesized by previously reported methods. Aminoacylation was varied out at 0◦C in reaction mixture containing 50 mM HEPES-KOH (pH 7.5), 600 mM $MgCl_2$, 20% DMSO, 25 μM dFx or eFx, 25 μM tRNA, and 5 mM activated amino acids. eFx was used for Phe-CME and ^{Bio}F-CME and dFx for the other amino acids. Reaction time was 2 hours for all the activated amino acids. The aa-tRNAs were recovered by ethanol precipitation, and then pellet was washed twice with 70% ethanol containing 0.1 M sodium acetate (pH 5.2), once with 70% ethanol and dissolved in 1 mM sodium acetate (pH 5.2).

Preparation of EF-P for the reconstituted cell-free translation system

E. coli efp gene was cloned into a modified pET28a(+) vector that has PreScission protease recognition site

instead of thrombin site. *E. coli epmA* and *epmB* genes were cloned into pETDuet-1 vector. These vectors were co-introduced into *E. coli* Rosetta2 (DE3) pLysS. The cells were cultured in LB broth containing 100 μg/mL of ampicillin, 50 μg/mL of kanamycin, and 30 μg/mL of chloramphenicol. Protein expression was induced by 0.5 mM IPTG for 2 h at 37◦C and lysed by sonication. After clarification by centrifuge at 15000 g for 15 min, the cell lysate was applied to a Co^{2+} column (HiTrap TALON crude, GE health care) to purify the histidine-tagged EF-P. The column was washed with buffer A (20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 2 mM imidazole, and 1 mM 2-mercaptoethanol), and then the histidine-tagged EF-P was eluted by buffer A containing 500 mM imidazole. Turbo3C protease was added to the eluent for cleaving the histidine-tag and the mixture was dialyzed against the buffer A at 4◦C overnight. The sample was applied on the Co^{2+} column, and the flow-through and wash fractions were recovered as EF-P without histidine-tag. Then, the protein was concentrated by Amicon ultra 10k centrifugal filter (Merck Millipore). Modification of EF-P was confirmed by MALDI-TOF MS analysis after digestion by Glu-C protease.

Expression of peptides in reconstituted cell-free translation system

Expression of peptides were carried out utilizing modified FIT (Flexible *in vitro* translation system) including only the necessary components and non-necessary amino acids, aaRSs, and EF-P were withdrawn from the following recipe. The complete FIT system contains 2 μM template mRNA, 0.5 mM each 20 proteinogenic L-amino acids (Ala, Cys, Asp, Glu, Phe, Gly, His, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Ser, Thr, Val, Trp, Tyr), 2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM Creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM Potassium acetate, 12 mM Magnesium acetate, 2 mM Spermidine, 1 mM DTT, 0.1 mM 10-HCO-H4folate, 1.5 mg/mL *E. coli* tRNAs, 0.3 mM Magnesium acetate, 1.2 μM *E. coli* ribosome, 0.6 μM MTF, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 0.26 μM EF-G, 10 μM EF-Tu, 0.66 μM EF-Ts, 5 μM EF-P0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 4 μg/mL Creatine kinase, 3 μg/mL Myokinase, 0.1 μM Inorganic pyrophosphatase, 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 μM IleRS, 0.04 μM LeuRS, 0.11 μM LysRS, 0.03 μM MetRS, 0.68 μM PheRS, 0.16 μM ProRS, 0.04 μM SerRS, 0.09 μM ThrRS, 0.03 μM TrpRS, 0.02 μM TryRS, and 0.02 μM ValRS. The concentration of EF-G was changed if necessary. In distinguish whether drop-off is caused by mRNA or nascent peptides, peptides were expressed from corresponding mRNA utilizing corresponding pre-charged aa-tRNAs as follows instead of free amino acids and cognate aaRSs: FLPS02(^{Bio}F, M) was expressed from mRNAS02 with 100 μM ^{Bio}F-tRNA$^{fMet}_{CAU}$ and 50 μM Met-tRNA$^{AsnE2}_{CAU}$ without free Met, MetRS, and 10-HCO-H4folate or with 100 μM ^{Bio}F-tRNA$^{fMet}_{CAU}$, 0.5 mM free Met, and MetRS without 10-HCO-H4folate. In expression of representative peptides with various kinds of nascent peptide sequences prior to the three consecutive prolines, peptides were expressed with 12.5 μM ^{Bio}F-tRNA$^{fMet}_{CAU}$ without 10-HCO-H4folate. The reaction mixture was incubated at 37◦C for 20, 30, or 80 min. In the identification of dropped peptidyl-tRNA or its nascent chain, reaction mixture was incubated with f.c. 10 μM of kanamycin (Nacalai tesque) for 5 min at 37◦C, followed by incubation with f.c. 1 μM of PTH for 10 min at 37◦C, if necessary. For sample preparation for LC-ESI MS, the reaction mixture was mixed with equivalent volume of MeOH, centrifuged for 3 min at 13000 rpm, and the supernatant was separated from the precipitation. The precipitation was re-resolved into 1x FIT buffer (20 mM Tris-HCl (pH 7.6), 50 mM KCl, and 1 mM DTT), mixed with f.c. 12.8 mM EDTA and f.c. 10 μM kanamycin, mixed with f.c. 15.8 mM $Mg(Oac)_2$ and f.c. 1 μM PTH, and incubated at 37◦C for

10, 30, or 60 min. Finally, the reaction mixture was incubated with f.c. 50 μg/mL of RNase A (Funakoshi) for 30 min at 37◦C if necessary.

In the identification by MALDI-TOF MS, the translation products were purified by anti FLAG antibody if necessary. After addition of equal amount of 2xTBS buffer (100 mM Tris-HCl, 300 mM NaCl, pH 7.6), the reaction mixture was incubated with anti FLAG antibody agarose gel (Sigma) for 1 hour at room temperature, followed by wash with 1x TBS buffer (50 mM Tris-HCl and 150 mM NaCl, pH 7.6) twice and eluted with 0.1v/v% TFA solution. Reaction mixture and elution were desalted with SPE C-tip (Nikkyo Technos) and eluted with 1.2 μL of 80v/v% acetonitrile, 0.5v/v% acetic acid solution containing 50w/v% saturated I-cyano-4-hydroxycinnamic acid. MALDI-TOF MS analysis was performed using ultrafleXtreme (Bruker Daltonics) in reflector positive mode. Peptide calibration standard II (Bruker Daltonics) was used for the external mass calibration.

In the quantification of peptide yields, 0.5 mM [^{12}C]-Asp, [^{12}C]-Lys, or [^{32}S]-Met in the FIT system were replaced with radio-labelled 0.05 mM [^{14}C]-Asp, 0.05 mM [^{14}C]-Lys, or 0.8683 μM [^{35}S]-Met. After incubation, reaction mixture was mixed with equal amount of 2x tricine-SDS-PAGE loading solution (0.9 M Tris-HCl, 8w/v% SDS, 30v/v% glycerol, and 0.001w/v% xylene cyanol, pH 8.45) and incubated at 95◦C for 5 min. The solution was applied for the tricine-SDS gel composed of 4% stacking gel and 15% separation gel and run for 40 min at 150 V with cathode buffer (0.1 mM Tris-HCl , 0.1 mM tricine, and 0.1w/v% SDS) and anode buffer (0.2 mM Tris-HCl). After drying, the gel was analyzed by Typhoon FLAG 7000 (GE Health Care).

In the identification and quantification by LC-ESI MS, proteins were precipitated by incubation of the reaction solution with equal volume of methanol on ice for 5 min, centrifuged at 13000 rpm for 3 min, and centrifuged with four equivalent volume of 1v/v% trifluoroacetic acid at 13000 rpm for 3 min. Resulting solution was injected to the LC-ESI MS (Waters XevoTM G2-XS) assembled with Aquity UPLC BEH C18 column (1.7 μm, 300Å, 2.1x150 mm waters) and eluted by gradient of H_2O/MeCN containing 0.1v/v% formic acid.

Comprehensive elucidation of effect of nascent chain on incorporation of amino acids

Nascent chain-dependent incorporation efficiency of prolines and glycines was elucidated utilizing RaPID system combining FIT system and mRNA display, followed by deep-sequencing. 1.2 μM mRNA conjugated with a puromycin linker was translated in 2.5 μL of FIT system in the presence of 12.5 μM of L-^{Bio}F-tRNA and in the absence of 10-HCO-H4folate, with 0.03, 0.26, or 10 μM EF-G and in the presence or absence of EF-P. The reaction mixture was incubated at 37◦C for 10 min, followed by incubation with f.c. 16.7 mM EDTA at 37◦C for 5 min and reverse transcription by adding 1.72 μL of RT solution (0.73 mM dNTPs, 73.4 mM Tris-HCl pH 8.3, 4.9 μM reverse primer, 29 mM KOH, 43.7 mM $Mg(OAc)_2$, and 14.6 U/μl RTase (Promega)) at 42◦C for 1 hour. The reaction was quenched by addition of 0.5 μL of 100 mM EDTA and 0.55 μL of 0.2 M HCl. 0.5 μL of reaction mixture was aliquoted as "initial" fraction, diluted into 1x PCR mixture (50 mM KCl 10 mM Tris-HCl (pH 9.0) 0.1v/v% Triton X-100, 2.5 mM $MgCl_2$, 0.25 mM dNTPs, and 0.25 μM of forward and reverse primer) by 100-fold. Remaining solution was incubated at room temperature for 10 min with 2.5 μL slurry of Dynabeads M-280 streptavidin (Veritas) washed by 1x TBST buffer (50 mM Tris-HCl, 150 mM NaCl, and 0.05v/v% Tween-20, pH 7.6) three times. After removal of supernatant, the Dynabeads was washed with 100 μL of 1x TBST and transferred to the new tube 10 times. The cDNA of L-^{Bio}F-peptide-mRNA-cDNA conjugates were eluted by 1x PCR mix by incubation at 95◦C for 5 min as "Pull-

down” fraction. After addition of Taq DNA polymerase into 1x PCR mixture, the amount of cDNA in the “initial” fraction and “Pull-down” fraction were quantified by qPCR. qPCR was conducted in the qPCR reaction mix consisting of PCR reaction mix containing 0.25 μM forward and reverse primers and 1/200000 SYBR® Green I Nucleic Acid Gel Stain (Takara). 1 μL of samples was added to 19 μL of qPCR reaction mix. Real time PCR using LightCycler® Nano (Roche) was conducted under following conditions: 95 ºC 30 sec, (95 ºC 15 sec → 61 ºC 15 sec → 72 ºC 30 sec (0.5ºC/sec)) x35 cycles. The cDNA was amplified by PCR in optimum thermal cycles determined by qPCR. Amplified cDNAs were purified by phenol/chloroform extraction and ethanol precipitation. Purified cDNAs were amplified by two tailed PCR to add an adopter sequence for deep sequencing and purified by NucleoSpin Gel and PCR Clean-up (Macherey-Nagel). The purified cDNAs were deep sequenced by MiSeq and MiSeq reagent kit v3 (150 cycles) (Illumina).

Data processing

The raw data were processed by python program written by Dr. Naoya Ozawa and further processed in order to remove the sequences mutated in the ORF, such as base insertion or deletion, and count the read number of each sequence.

Python script:

```
#!/usr/bin/env python3

# Copyright © 2016-2017 Naoya Ozawa

import json
import os
import platform
import re
import subprocess
import sys

print()

script_dir = os.path.dirname(os.path.realpath(__file__))
on_windows = (platform.system() ‘= 'Wind’ws')

config_path = sys.argv[1]
input_paths = sys.argv[2:-1]
output_dir = sys.argv[-1]
for path in [config_path] + input_paths:
    if not os.path.isfile(path):
        pri‘t('Error: File {} does not exi’t.'.format(path))
```

```
            sys.exit(1)
if not os.path.isdir(output_dir):
    pri‘t('Error: Folder {} does not exi’t.'.format(output_dir))
    sys.exit(1)
if not config_path.endswi‘h('.j’on'):
    pri‘t('Error: File {} must be a .json fi’e.'.format(config_path))
    sys.exit(1)
for path in input_paths:
    if not (path.endswi‘h('.fastq’gz') or path.endswi‘h('.fa’tq')):
        pri‘t('Error: File {} must be a .fastq(.gz) fi’e.'.format(path))
        sys.exit(1)


def create_filename(dir_path, name):
    root, ext = os.path.splitext(name)
    filename = os.path.join(dir_path, name)
    i = 1
    while os.path.exists(filename):
        filename = os.path.join(dir_pat‘, '{}-{’{}'.format(roo©, i, ext))
        i += 1
    return filename


output_total_path = create_filename(output_d‘‘, 'total_reads’’sv')
output_q_path = create_filename(output_d‘‘, 'counts_by_Q’’sv')
output_dna_path = create_filename(output_d‘‘, 'counts_by_dna’’sv')

pr‘‘t('Please check the following paramet’’s.')
pr‘‘t('    Config file:’’{}'.format(config_path))
pr‘‘t('    Input files:’’{}'.format(input_paths[0]))
for path in input_paths[1:]:
    pr‘‘t('                        ’’{}'.format(path))
pr‘‘t('    Output files’’{}'.format(output_total_path))
pr‘‘t('                         ’’{}'.format(output_q_path))
pr‘‘t('                         ’’{}'.format(output_dna_path))
print()
ans = in‘‘t('Continue? Please t‘‘pe ””es"‘‘o”””””‘‘: ').lower()
print()
```

```
if ans not ‘n‘’[’‘', ’’es']:
    pr‘‘t('Cance’’d.')
    sys.exit(0)

# Check config file

with open(config_pa‘h’’’rt') as f:
    try:
        config = json.load(f)
    except json.JSONDecodeError as err:
        pr‘‘t('Error detected in the config f’’e.')
        pr‘‘t('    line {} colum’’{}'.format(err.lineno, err.colno))
        pr‘‘t('   ’’{}'.format(err.msg))
        pr‘‘t('Please fix the error and re’’y.')
        sys.exit(1)

try:
    pattern = re.compile(con‘‘g['pat’’rn'])
except KeyError:
    pr‘‘t('Error detected in the config f’’e.')
    pr‘‘t('    ““y "pat””rn" is requi’’d.')
    pr‘‘t('Please fix the error and re’’y.')
    sys.exit(1)
except:
    pr‘‘t('Error detected in the config f’’e.')
    pr‘‘t('    The value““f "pat””rn" must be a valid regular express’’n.')
    pr‘‘t('Please fix the error and re’’y.')
    sys.exit(1)

try:
    min_q = con‘‘g['m’’_q']
except KeyError:
    pr‘‘t('Error detected in the config f’’e.')
    pr‘‘t('    ““y "m””_q" is requi’’d.')
    pr‘‘t('Please fix the error and re’’y.')
    sys.exit(1)
if not isinstance(min_q, int):
    pr‘‘t('Error detected in the config f’’e.')
```

```
        pr‘‘t('       The value‘‘‘f "m’’’_q" must be an inte’’r.')
        pr‘‘t('Please fix the error and re’’y.')
        sys.exit(1)

try:
        initiator_aa = con‘‘g['initiato’’aa']
except KeyError:
        pr‘‘t('Error detected in the config f’’e.')
        pr‘‘t('       ‘‘‘y "initiato’’’aa" is requi’’d.')
        pr‘‘t('Please fix the error and re’’y.')
        sys.exit(1)
if not isinstance(initiator_aa, str):
        pr‘‘t('Error detected in the config f’’e.')
        pr‘‘t('       The value‘‘‘f "initiato’’’aa" must be a str’’g.')
        pr‘‘t('Please fix the error and re’’y.')
        sys.exit(1)

try:
        codon_table = con‘‘g['codon_t’’le']
except KeyError:
        pr‘‘t('Error detected in the config f’’e.')
        pr‘‘t('       “« y "codon_t” »le" is requi’’d.')
        pr‘‘t('Please fix the error and re’’y.')
        sys.exit(1)
if not isinstance(codon_table, dict):
        pr‘‘t('Error detected in the config f’’e.')
        pr‘‘t('       The value‘‘‘f "codon_t’’’le" must be a diction’’y.')
        pr‘‘t('Please fix the error and re’’y.')
        sys.exit(1)
for codon in codon_table.keys():
        if not re.ma‘‘h('^[ATGC]’’}$', codon):
                pr‘‘t('Error detected in the config f’’e.')
                pr‘‘t('       Invalid c odon’’{}'.format(codon))
                pr‘‘t('       A codon consists of three letters from A’’C.')
                pr‘‘t('Please fix the error and re’’y.')
                sys.exit(1)
for aa in codon_table.values():
        if not isinstance(aa, str):
```

```
            pr‘‘t('Error detected in the config f’’e.')
            pr‘‘t('       Invalid amino acid’’{}'.format(aa))
            pr‘‘t('       Amino acids must be a str’’g.')
            pr‘‘t('Please fix the error and re’’y.')
            sys.exit(1)

pr‘‘t('Processing the data. This may take several minu’’s.')
print()

# counttotal.py

pr‘‘t('---- Total re‘‘s '.ljust(‘0‘‘‘'-'))
with open(output_total_pa‘h’’’wb') as f:
    subprocess.run(
        ‘ ’’’py'] if on_windows else [])
            + [os.path.join(script_d‘‘, 'counttota’’py')]
            + input_paths,
            stdout=f, stderr=sys.stderr)
pr‘‘t('Output written to ’’ .'.format(output_total_path))
pr‘n‘‘‘'-'*80)
print()

# countq.py

pr‘‘t('---- Q-filter and cou‘‘s '.ljust(‘0‘‘‘'-'))
with open(output_q_pa‘h’’’wb') as f:
    subprocess.run(
        ‘ ’’’py'] if on_windows else [])
            + [os.path.join(script_d‘‘, 'count’’py'), config_path]
            + input_paths,
            stdout=f, stderr=sys.stderr)
pr‘‘t('Output written to ’’ .'.format(output_total_path))
pr‘n‘‘‘'-'*80)
print()

# countdna.py

pr‘‘t('---- Filtering, counting and translat‘‘g '.ljust(‘0‘‘‘'-'))
```

```
with open(output_dna_pa‘h’’wb') as f:
    subprocess.run(
      ‘ ’’py'] if on_windows else [])
        + [os.path.join(script_d‘‘, 'countdn’’py'), config_path]
        + input_paths,
        stdout=f, stderr=sys.stderr)
pr‘‘t('Output written to ’’ .'.format(output_total_path))
pr‘n‘‘‘-'*80)

print()
pr‘‘t('==== WARN‘‘G '.ljust(‘0‘’’='))
pr‘‘t('THE AUTHOR DOES NOT GUARANTEE THE CORRECTNESS OF THE OUT’’T.')
pr‘n‘’’='*80)

print()
pr‘‘t('All finis’’d!')
pr‘‘t('You may close this win’’w.')
```

The ratio of a nascent peptide ch©in i (Ni) in the “Initial” or “Pull-down” fraction, $F_{Ni,\ Initial}$ and $F_{Ni\ Pull\text{-}down,\ EF\text{-}P\text{-}/+}$, enrichment of n efficiency of Ni in the presence or absence of EF-P through the selection process, $E_{Ni,EF\text{-}P\text{-}/+}$, and incorporation enhancement by EF-P, $W_{Ni,}$ were calculated as following equations;

$$F_{Ni,Initial} = \frac{Count_{Ni,Initial}}{\sum_i^{8420} Count_{Ni,Initial}}$$

$$F_{Ni,Pull-down,EF-P-/+} = \frac{Count_{Ni,Pull-down,EF-P-/+}}{\sum_i^{8420} Count_{Ni,Pull-down,EF-P-/+}}$$

$$F_{Ni,EF-P-/+} = \log_2\left(\frac{F_{Ni,Pull-down,EF-P-/+}}{F_{Ni,Initial}}\right)$$

$$W_{Ni} = F_{Ni,EF-P+} - F_{Ni,EF-P-}$$

2.6. Supplementary tables

Supplementary table 1 DNA primer sequences used.

Primer Name	DNA sequence
Primer01.F46	TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATAAATA
Primer02.R58	CCGTCCTAGCTTACCCCGTTACCGCTCGGCGGCGGNNNCATATTTATATCTCCTTCTT
Primer03.R36	TTTCCGCCCCCCGTCCTAGCTTACCCCGTTACCGCT
Primer04.R61	CCGTCCTAGCTTACCCCGTTACCGCTCGGCGGCGGNNNNNNCATATTTATATCTCCTTCTT
Primer05.R64	CCGTCCTAGCTTACCCCGTTACCGCTCGGCGGCGGNNNNNNNNNCATATTTATATCTCCTTCTT
Primer06.F40	AGAAGGAGATATAAATATGNNNGGTGGTGGTAGCGGTAAC
Primer07.R45	TTTCCGCCCCCCGTCCTAGCTTACCCCGTTACCGCTACCACCACC
Primer08.F43	AGAAGGAGATATAAATATGNNNNNNGGTGGTGGTAGCGGTAAC
Primer09.F46	AGAAGGAGATATAAATATGNNNNNNNNNGGTGGTGGTAGCGGTAAC
Primer10.F52	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGAAGAAG
Primer11.R58	5'GTCGTCGTCCTTGTAGTCCTTCTTCTTACCACCACCCTTCTTCTTCATGTTTTTCTCC3'
Primer12.F22	GGCGTAATACGACTCACTATAG
Primer13.R33	CGAAGCTTACTTGTCGTCGTCGTCCTTGTAGTC
Primer14.F46	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATG
Primer15.R42	GTCGTCGTCCTTGTAGTCCATCTTCTTCTTCATGTTTTTCTC
Primer16.R42	CTTGTCGTCGTCGTCCTTGTAGTCGTTACCGCTCGGCGGCGG
Primer17.F46	AGAAGGAGATATAAATATGATGTGGTGGCCGCCGCCGAGCGGTAAC
Primer18.F46	AGAAGGAGATATAAATATGTGGTATGGACCGCCGCCGAGCGGTAAC
Primer19.F46	AGAAGGAGATATAAATATGTGGATGGGACCGCCGCCGAGCGGTAAC
Primer20.F46	AGAAGGAGATATAAATATGTGCTGGTTCCCGCCGCCGAGCGGTAAC
Primer21.F46	AGAAGGAGATATAAATATGATGATGATACCGCCGCCGAGCGGTAAC
Primer22.F46	AGAAGGAGATATAAATATGTGGGTATGGCCGCCGCCGAGCGGTAAC
Primer23.F46	AGAAGGAGATATAAATATGCATGAAAACCCGCCGCCGAGCGGTAAC
Primer24.F46	AGAAGGAGATATAAATATGCATTACAATCCGCCGCCGAGCGGTAAC
Primer25.F46	AGAAGGAGATATAAATATGCACATAGGACCGCCGCCGAGCGGTAAC
Primer26.F46	AGAAGGAGATATAAATATGCTCATCTTCCCGCCGCCGAGCGGTAAC
Primer27.F46	AGAAGGAGATATAAATATGCTCCTCCTCCCGCCGCCGAGCGGTAAC
Primer28.F46	AGAAGGAGATATAAATATGCGCCGCCGCCCGCCGCCGAGCGGTAAC
Primer29.F46	AGAAGGAGATATAAATATGAAGAAGAAGCCGCCGCCGAGCGGTAAC
Primer30.F46	AGAAGGAGATATAAATATGAACAAGGAGCCGCCGCCGAGCGGTAAC
Primer31.F46	AGAAGGAGATATAAATATGGGTGCCGAACCGCCGCCGAGCGGTAAC
Primer32.F46	AGAAGGAGATATAAATATGGACAGTGAACCGCCGCCGAGCGGTAAC
Primer33.F46	AGAAGGAGATATAAATATGGCACGAGGACCGCCGCCGAGCGGTAAC
Primer34.F46	AGAAGGAGATATAAATATGGGAGATGAACCGCCGCCGAGCGGTAAC
Primer35.F46	AGAAGGAGATATAAATATGGCAGCAGACCCGCCGCCGAGCGGTAAC
Primer36.F46	AGAAGGAGATATAAATATGGAAGTAGATCCGCCGCCGAGCGGTAAC
Primer37.F46	AGAAGGAGATATAAATATGGAAGCAGACCCGCCGCCGAGCGGTAAC
Primer38.F46	AGAAGGAGATATAAATATGGGGCATGACCCGCCGCCGAGCGGTAAC
Primer39.F36	GTAATACGACTCACTATAGGATCGAAAGATTTCCGC
Primer40.R46	ACCTAACGCCATGTACCCTTTCGGGGATGCGGAAATCTTTCGATCC
Primer41.F22	GGCGTAATACGACTCACTATAG
Primer42.R19	ACCTAACGCCATGTACCCT
Primer43.R45	ACCTAACGCTAATCCCCTTTCGGGGCCGCGGAAATCTTTCGATCC
Primer44.R18	ACCTAACGCTAATCCCCT
Primer45.F49	GTAATACGACTCACTATAGGCGGGGTGGAGCAGCCTGGTAGCTCGTCGG
Primer46.R44	GAACCGACGATCTTCGGGTTATGAGCCCGACGAGCTACCAGGCT
Primer47.R38	TGGTTGCGGGGGCCGGATTTGAACCGACGATCTTCGGG
Primer48.R20	TmGGTTGCGGGGGCCGGATTT
Primer49.F49	GTAATACGACTCACTATAGGCTCTGTAGTTCAGTCGGTAGAACGGCGGA
Primer50.R43	GAACCAGTGACATACGGATTATGAGTCCGCCGTTCTACCGACT
Primer51.R38	TGGCGGCTCTGACTGGACTCGAACCAGTGACATACGGA
Primer52.R20	TmGGCGGCTCTGACTGGACTC

Supplementary table 2 Combinations of DNA primers for the extension and PCR reaction for preparation of mDNAs, flexizymes, and tRNAs.

	Extension		PCR	
Name	**Forward primer**	**Reverse primer**	**Forward primer**	**Reverse primer**
cDNA library 1	Prim01.F46	Prim02.R58	Prim01.F46	Prime03.R36
cDNA library 2	Prim01.F46	Prime04.R61	Prim01.F46	Prime03.R36
cDNA library 3	Prim01.F46	Prime05.R64	Prim01.F46	Prime03.R36
cDNA library 4	Prime06.F40	Prime07.R45	Prim01.F46	Prime07.R45
cDNA library 5	Prime08.F43	Prime07.R45	Prim01.F46	Prime07.R45
cDNA library 6	Prime09.F46	Prime07.R45	Prim01.F46	Prime07.R45
cDNAS01	Prime10.F52	Prime11.R58	Prime12.F22	Prime13.R33
cDNAS02	Prime14.F46	Prime15.R42	Prime12.F22	Prime13.R33
cDNAS03	Prime17.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS04	Prime18.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS05	Prime19.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS06	Prime20.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS07	Prime21.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS08	Prime22.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS09	Prime23.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS10	Prime24.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS11	Prime25.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS12	Prime26.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS13	Prime27.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS14	Prime28.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS15	Prime29.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS16	Prime30.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS17	Prime31.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS18	Prime32.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS19	Prime33.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS20	Prime34.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS21	Prime35.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS22	Prime36.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS23	Prime37.F46	Prime16.R42	Prime12.F22	Prime13.R33
cDNAS24	Prime38.F46	Prime16.R42	Prime12.F22	Prime13.R33

	Extension		1st PCR		2nd PCR	
RNA Name	**Forward primer**	**Reverse primer**	**Forward primer**	**Reverse primer**	**Forward primer**	**Reverse primer**
eFx	Prime39.F36	Prime40.R46	Prime41.F22	Prime42.R19		
dFx	Prime39.F36	Prime43.R45	Prime41.F22	Prime44.R18		
tRNAini	Prime45.F49	Prime46.R44	Prime41.F22	Prime47.R38	Prime41.F22	Prime48.R20
tRNAAsnE2cau	Prime49.F49	Prime50.R43	Prime41.F22	Prime51.R38	Prime41.F22	Prime52.R20

Supplementary table 3 Sequences of Flexizymes and tRNAs.

RNA Name	**RNA**
eFx	GGATCGAAAGATTTCCGCGGCCCCGAAAGGGGATTAGCGTTAGGT
dFx	GGATCGAAAGATTTCCGCATCCCCGAAAGGGTACATGGCGTTAGGT
tRNAini	GGCGGGGTGGAGCAGCCTGGTAGCTCGTCGGGCTCATAACCCGAAGATCGTCGGTTCAAATCCGGCCCCCGCAACCA
tRNAAsnE2cau	GGCUCUGUAGUUCAGUCGGUAGAACGGCGGACUCAUAAUCCGUAUGUCACUGGUUCGAGUCCAGUCAGAGCCGCCA

Supplementary table 4 Sequences of cDNA, mRNA, Flexizymes, tRNAs, and peptides.

	DNA sequence upstream of initiation ATG codon
	TAATACGACTCACTATAGGGTTAACTTTAAGAAGGAGATATAAAT
Name	**DNA sequence downstream from initiation ATG codon**
cDNA library 1	ATGNNNCCGCCGCCGAGCGGTAACGGGGTAAGCTAGGACGGGGGGCGGAAA
cDNA library 2	ATGNNNNNNCCGCCGCCGAGCGGTAACGGGGTAAGCTAGGACGGGGGGCGGAAA
cDNA library 3	ATGNNNNNNNNNCCGCCGCCGAGCGGTAACGGGGTAAGCTAGGACGGGGGGCGGAAA
cDNA library 4	ATGNNNGGTGGTGGTAGCGGTAACGGGGTAAGCTAGGACGGGGGGCGGAAA
cDNA library 5	ATGNNNNNNGGTGGTGGTAGCGGTAACGGGGTAAGCTAGGACGGGGGGCGGAAA
cDNA library 6	ATGNNNNNNNNNGGTGGTGGTAGCGGTAACGGGGTAAGCTAGGACGGGGGGCGGAAA
	DNA sequence upstream of initiation ATG codon
	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAAC
cDNAS01	ATGAAGAAGAAGCCGCCGCCGAAGAAGAAGGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS02	ATGAAGAAGAAGCCGCCGCCGAAGAAGAAGGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS03	ATGATGTGGTGGCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS04	ATGTGGTATGGACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS05	ATGTGGATGGGACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS06	ATGTGCTGGTTCCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS07	ATGATGATGATACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS08	ATGTGGGTATGGCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS09	ATGCATGAAAACCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS10	ATGCATTACAATCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS11	ATGCACATAGGACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS12	ATGCTCATCTTCCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS13	ATGCTCCTCCTCCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS14	ATGCGCCGCCGCCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS15	ATGAAGAAGAAGCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS16	ATGAACAAGGAGCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS17	ATGGGTGCCGAACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS18	ATGGACAGTGAACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS19	ATGGCACGAGGACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS20	ATGGGAGATGAACCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS21	ATGGCAGCAGACCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS22	ATGGAAGTAGATCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS23	ATGGAAGCAGACCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG
cDNAS24	ATGGGGCATGACCCGCCGCCGAGCGGTAACGACTACAAGGACGACGACGACAAGTAAGCTTCG

	mRNA sequence upstream of initiation AUG codon	
	GUUAACUUUAAGAAGGAGAUAUAAAU	
Name	**mRNA sequence downstream from initiation AUG codon**	**Peptide sequence**
mRNA library 1	AUGNNNCCGCCGCCGAGCGGUAACGGGGUAAGCUAGGACGGGGGGCGGAAA	MXPPPSGNGVS
mRNA library 2	AUGNNNNNNCCGCCGCCGAGCGGUAACGGGGUAAGCUAGGACGGGGGGCGGAAA	MXXPPPSGNGVS
mRNA library 3	AUGNNNNNNNNNCCGCCGCCGAGCGGUAACGGGGUAAGCUAGGACGGGGGGCGGAAA	MXXXPPPSGNGVS
mRNA library 4	AUGNNNGGUGGUGGUAGCGGUAACGGGGUAAGCUAGGACGGGGGGCGGAAA	fMXGGGSGNGVS
mRNA library 5	AUGNNNNNNGGUGGUGGUAGCGGUAACGGGGUAAGCUAGGACGGGGGGCGGAAA	fMXXGGGSGNGVS
mRNA library 6	AUGNNNNNNNNNGGUGGUGGUAGCGGUAACGGGGUAAGCUAGGACGGGGGGCGGAAA	fMXXXGGGSGNGVS
	mRNA sequence upstream of initiation AUG codon	
	GUUAACUUUAACAAGGAGAAAAAC	
mRNAS01	AUGAAGAAGAAGCCGCCGCCGAAGAAGAAGGACUACAAGGACGACGACGACAAGUAAGCUUCG	MKKKPPPKKKDYKDDDDK
mRNAS02	AUGAAGAAGAAGCCGCCGCCGAAGAAGAAGGACUACAAGGACGACGACGACAAGUAAGCUUCG	MKKKMDYKDDDDK
mRNAS03	AUGAUGUGGUGGCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MMWWPPPSGNDYKDDDDK
mRNAS04	AUGUGGUAUGGACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MWYGPPPSGNDYKDDDDK
mRNAS05	AUGUGGAUGGGACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MWMGPPPSGNDYKDDDDK
mRNAS06	AUGUGCUGGUUCCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MCWFPPPSGNDYKDDDDK
mRNAS07	AUGAUGAUGAUACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MMMIPPPSGNDYKDDDDK
mRNAS08	AUGUGGGUAUGGCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MWVWPPPSGNDYKDDDDK
mRNAS09	AUGCAUGAAAACCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MHENPPPSGNDYKDDDDK
mRNAS10	AUGCAUUACAAUCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MHYNPPPSGNDYKDDDDK
mRNAS11	AUGCACAUAGGACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MHIGPPPSGNDYKDDDDK
mRNAS12	AUGCUCAUCUUCCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MLIFPPPSGNDYKDDDDK
mRNAS13	AUGCUCCUCCUCCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MLLLPPPSGNDYKDDDDK
mRNAS14	AUGCGCCGCCGCCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MRRRPPPSGNDYKDDDDK
mRNAS15	AUGAAGAAGAAGCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MKKKPPPSGNDYKDDDDK
mRNAS16	AUGAACAAGGAGCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MNKEPPPSGNDYKDDDDK
mRNAS17	AUGGGUGCCGAACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MGAEPPPSGNDYKDDDDK
mRNAS18	AUGGACAGUGAACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MDSEPPPSGNDYKDDDDK
mRNAS19	AUGGCACGAGGACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MARGPPPSGNDYKDDDDK
mRNAS20	AUGGGAGAUGAACCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MGDEPPPSGNDYKDDDDK
mRNAS21	AUGGCAGCAGACCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MAADPPPSGNDYKDDDDK
mRNAS22	AUGGAAGUAGAUCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MEVDPPPSGNDYKDDDDK
mRNAS23	AUGGAAGCAGACCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MEADPPPSGNDYKDDDDK
mRNAS24	AUGGGGCAUGACCCGCCGCCGAGCGGUAACGACUACAAGGACGACGACGACAAGUAAGCUUCG	MGHDPPPSGNDYKDDDDK

Supplementary table 5 Numerical values of nascent peptide profiling for three proline incorporation.

(A) Statistical values of raw and filtered read counts of nascent peptides. SAV: Streptavidin.

(B) Statistical values of calculated $E_{Ni,\ EF\text{-}P\text{-}}$, $E_{Ni,\ EF\text{-}P+}$, and W_{Ni}.

A

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Date	180307	180307	180317	180317	180307	180307	180307	180307	171205	171205	180131	180131	191217	191217	191217	191217
EF-G (μM)	0.03	0.03	0.03	0.03	0.26	0.26	0.26	0.26	10	10	10	10	10	10	10	10
EF-P (μM)	0	0	5	5	0	0	5	5	0	0	5	5	0	0	5	5
Sample type	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV
Amount of cDNA (molecule)	2.2E+12	7.8E+10	2.1E+12	3.8E+10	1.6E+12	3.4E+10	2.5E+12	5.3E+10	3.2E+12	1.6E+10	3.3E+12	1.7E+11	6.8E+11	8.0E+9	8.9E+11	4.5E+10
Total reads	2680447	1249045	2168423	1280101	1545161	869421	2254602	849908	1857321	938587	717251	686498	1493664	1141195	1755412	2538417
Reads Q>30	377347	623767	645461	790265	728330	409487	393507	429683	1177802	600014	395211	406657	810685	591137	940630	1401804
Sequences as designed without stop codon	302117	550761	516842	716536	586013	358182	316730	396836	938744	526663	318770	391503	648465	535987	759409	1343863
Average reads	35.9	65.4	61.4	85.1	69.6	42.5	37.6	47.1	111.5	62.5	37.9	46.5	77.0	63.6	90.2	159.6
Variance	953	4475	2759	5830	3553	2099	1052	1651	9261	4793	1173	1772	4370.9	8028.0	5895.3	17729.9
Median	27	45	46	61	53	28	28	35	83	39	28	34	58	32	68	122
Maximum read	239	575	396	568	458	434	247	336	768	649	861	327	566	1007	647	1836
Minimum read	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
Detected nascent peptide sequences	8406	8409	8417	8419	8418	8349	8405	8402	8419	8408	8405	8384	8421	8421	8421	8421
Sequence coverage (%)	99.83	99.87	99.96	99.99	99.98	99.16	99.82	99.79	99.99	99.86	99.82	99.57	1.000	1.000	1.000	1.000

B

No.	17	18	19	20	21	22	23	24	25	26	27	28
EFG (μM)	0.03	0.03	0.26	0.26	10	10	10	10	0.03	0.26	10	10
EFP (μM)	0	5	0	5	0	5	0	5	N/A	N/A	N/A	N/A
X of $\log_2(X)$	$E_{Ni,\ EF\text{-}P\text{-}}$	$E_{Ni,\ EF\text{-}P+}$	$E_{Ni,\ EF\text{-}P\text{-}}$	$E_{Ni,\ EF\text{-}P+}$	$E_{Ni,\ EF\text{-}P\text{-}}$	$E_{Ni,\ EF\text{-}P+}$	$E_{Ni,\ EF\text{-}P\text{-}}$	$E_{Ni,\ EF\text{-}P+}$	W_{Ni}	W_{Ni}	W_{Ni}	W_{Ni}
Average	-0.15	-0.04	-0.24	0.00	-0.24	-0.07	-0.47	-0.01	0.1	0.2	0.2	0.46
Variance	0.88	0.64	1.15	0.62	1.45	0.96	1.88	0.45	0.66	0.95	1.00	1.46
Median	-0.05	0.03	-0.12	0.06	-0.05	0.06	-0.43	0.09	0.03	0.15	0.07	0.47
Maximum	3.89	3.78	3.52	3.67	3.83	5.10	4.23	3.14	5.64	5.61	4.68	4.83
Minimum	-5.45	-4.86	-5.68	-4.57	-5.51	-5.51	-5.53	-4.99	-4.46	-3.84	-6.70	-4.63

Supplementary table 6 Numerical values of nascent peptide profiling for three glycine incorporation.

(A) Statistical values of raw and filtered read counts of nascent peptides. SAV: Streptavidin.

(B) Statistical values of calculated $E_{Ni,\ EF\text{-}P\text{-}}$, $E_{Ni,\ EF\text{-}P+}$, and W_{Ni}.

A

No.	1	2	3	4	5	6	7	8	9	10	11	12
Date	191217	191217	191217	191217	180810	180810	180810	180810	180810	180810	180810	180810
EF-G (μM)	0.03	0.03	0.03	0.03	0.26	0.26	0.26	0.26	10	10	10	10
EF-P (μM)	0	0	5	5	0	0	5	5	0	0	5	5
Sample type	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV	Input	Pull-down by SAV
Amount of cDNA (molecule)	3.7E+11	1.8E+10	3.6E+11	1.1E+10	1.1E+10	5.6E+9	2.2E+11	5.0E+9	1.5E+11	2.3E+10	1.8E+11	2.4E+10
Total reads	1291411	753100	1669616	504254	653214	619421	1242984	1376234	1003566	1292726	1788690	1670874
Reads Q > 30	744457	480285	939799	313594	341310	307948	708379	741372	500248	682057	940595	929163
Sequences as designed without stop codon	236930	454695	241621	293465	196365	192824	674563	704710	313491	460126	906494	893791
Average reads	153.4	89.5	198.3	59.9	23.3	22.9	80.1	83.7	37.2	54.6	107.6	106.1
Variance	1132.5	4232.6	1198.9	1725.7	830	823	10872	11904	2119	4513	18170	17099
Median	17	32	17	21	14	13	44	46	21	32	62	62
Maximum read	345	629	329	407	276	272	1138	1080	458	674	1340	1308
Minimum read	0	0	0	0	0	0	0	0	0	0	0	0
Detected nascent peptide sequences	8419	8419	8419	8419	8251	8237	8397	8406	8339	8375	8415	8413
Sequence coverage (%)	99.98	99.98	99.98	99.98	97.98	97.81	99.71	99.82	99.03	99.45	99.93	99.90

B

No.	13	14	15	16	17	18	19	20	21
EFG (μM)	0.03	0.03	0.26	0.26	10	10	0.03	0.26	10
EFP (μM)	0	5	0	5	0	5	N/A	N/A	N/A
X of $\log_2(X)$	$E_{Ni,\ EF\text{-}P\text{-}}$	$E_{Ni,\ EF\text{-}P+}$	$E_{Ni,\ EF\text{-}P\text{-}}$	$E_{Ni,\ EF\text{-}P+}$	$E_{Ni,\ EF\text{-}P\text{-}}$	$E_{Ni,\ EF\text{-}P+}$	W_{Ni}	W_{Ni}	W_{Ni}
Average	-0.030	-0.002	-0.069	-0.062	-0.006	-0.001	0.028	0.007	0.005
Variance	0.44	0.51	0.51	0.54	0.36	0.30	0.71	0.70	0.49
Median	-0.002	0.041	-0.037	-0.022	0.001	0.009	0.005	0.000	0.001
Maximum	3.58	3.63	3.47	3.74	3.05	2.85	4.83	4.82	4.70
Minimum	-3.85	-4.09	-4.78	-4.19	-4.70	-3.96	-4.25	-5.22	-4.07

Chapter3
Demonstration of expression of proteins lacking N-terminus due to the drop-off

3.1. Introduction

Obtaining *in vitro* results described in the Chapter 2, I further elucidated the C-terminal peptide synthesis induced by peptidyl-tRNA drop-off in *in vivo* polypeptide synthesis. Peptidyl-tRNA drop-off induced by various factors has been detected *in vivo* by utilizing *E. coli* having temperature-sensitive (ts) mutation in PTH. However, generation of C-terminal peptides has never been investigated. By expressing *E. coli*'s proteins and putative proteins containing a stretch of prolines proximal to their N- termini in *E. coli* Δ*efp* and *E. coli* Δ*efp* complemented of *efp* on a plasmid, I investigated generation of C-terminal peptides or proteins lacking their N-terminal amino acids in the *E. coli*'s endogenous protein synthesis system. In the case of putative protein YhhM, N-terminal peptide fragments derived from full-length YhhM and YhhM lacking amino acids from N-terminal Met to 2nd proline were detected, and the relative quantity of latter was suppressed by co-expression of EF-P.

3.1.1. Direct and indirect detection of peptidyl-tRNA drop-off in the PTH-deficient *E. coli* mutant

Peptidyl-tRNA drop-off induced by various factors has been directly or indirectly detected *in vivo*. Menninger directly showed the accumulation of dropped peptidyl-tRNA in *E. coli* having ts mutant PTH[27]. He extracted the mixture of naked, aminoacyl-, and peptidyl-tRNAs from the *E. coli* mutant and evaluated the amino acid accepting ability of the tRNA mixture after PTH treatment. Addition of PTH enhanced the amino acid accepting ability, indicating that PTH unmasked the peptidyl-tRNA dropped off from ribosome for the next round of aminoacylation. Accumulation of peptidyl-tRNA has been indirectly detected in the above *pth* ts *E. coli* by such as ochre suppressor mutation of $tRNA^{Lys}$ isoacceptor[116], expression of minigenes coding rare codons in their short OFRs[28,29,37,117–122], gene knockout of the RF3 and RRF[31], and overexpression of IF1 and IF2[30]. However, accumulation of peptidyl-tRNA by knocking out *pth* gene could induce the several artifacts, such as ribosome stalling induced by starvation of tRNA trapped in peptidyl-tRNA and spontaneous release of peptidyl-tRNA from the dissociated 70S ribosomes.

3.1.2. Investigation of ribosome stalling induced by a stretch of prolines *in vivo*

Ribosome stalling induced by a stretch of prolines have been well characterized *in vivo* by measurement of change in cell proliferation time, SILAC[52], and ribosome profiling[2]. However, it has never been monitored whether a stretch of prolines generates a C-terminal peptide due to peptidyl-tRNA drop-off. More than thousands of *E. coli* proteins have consecutive proline motifs and some of them are located within 50 amino acid residues from the N-terminal Met[82]. According to the former literature, because shorter nascent peptidyl-tRNAs dropped off more frequently, I assumed that the short peptidyl-polyprolyl-tRNA could drop off from the ribosome *in vivo*.

3.1.3. Investigation of C-terminal peptide generation by expressing polyproline-containing proteins in *E. coli* Δ*efp* and *E. coli* Δ*efp* completed of *efp* on a plasmid

In order to detect generation of proteins lacking their N-terminal amino acid residues due to the peptidyl-tRNA drop-off and avoid artifacts due to *pth* gene mutation, I expressed proteins containing polyproline motif within 10 amino acid residues from the N-terminal Met in the presence of PTH and in the absence or presence of EF-P. At first, ~20 amino acid residues from the N-terminal Met of candidate proteins containing a stretch of prolines were expressed in the reconstituted cell-free translation system and the peptidyl-tRNA drop-off was monitored by LC-ESI

MS and autoradiography after tricine-SDS PAGE. The proteins showing peptidyl-tRNA drop-off were cloned into plasmid vectors and expressed in *E. coli* Δ*efp*[123] and in that complemented of *efp* on another plasmid. After affinity purification, candidate proteins were digested by proteases and sequenced by LC-ESI MS/MS. As a result, from YhhM-FLAG-His_6 digested by Lys-N protease, N-terminal fragment derived from both of full-length of YhhM-FLAG-His_6 and YhhM-FLAG-His_6 lacking three, four, or five amino acids from the N-terminal Met were identified. The quantity of YhhM-FLAG-His_6 lacking its N-terminal amino acids were suppressed in the presence of EF-P.

3.2. Results and discussion

3.2.1. *In vitro* screening of *E. coli*'s polyproline-containing proteins inducing peptidyl-tRNA drop-off

In order to perform a preliminary screening of peptidyl-tRNA drop-off, I chose 16 kinds of *E. coli* proteins and putative proteins having two or three consecutive prolines within 20 residues from its N-terminal Met for *in vitro* expression. Among them, 13 sequences (YhfK, HofN, YidD, YjgZ, PrpR, YnjA, YhhM, SfsA, RecT, MdtB, YggW, AldB, and YaiA) have two consecutive prolines and three sequences (RutD, YdcO, and Flk) have less than four consecutive prolines. Protein derivatives consist of native sequences from N-terminal Met to 10 residues downstream from the consecutive prolines and C-terminal FLAG tag (**Figure 21A**). Protein derivatives were expressed in the reconstituted cell-free translation system with or without EF-P and their N-terminal fMet were labelled with [^{35}S]-Met for detection of full-length peptides, peptidyl-tRNA, and N-terminal peptides hydrolyzed by PTH from dropped peptidyl-tRNA. After expression, reaction mixture was treated with PTH or not treated. Then, peptides and peptidyl-tRNA were subjected to tricine-SDS PAGE and visualized by autoradiography (**Figure 21B**). Among the 16 kinds of protein derivatives, six (FLP04, FLP05, FLP07, FLP14, FLP15, and FLP16 derived from YjgZ, PrpR, YhhM, RutD, YdcO, and Flk, respectively) showed the PTH-sensitive bands in the absence of EF-P, which should be dropped peptidyl-tRNAs (**Figure 21B**). In order to identify nascent peptide sequences on dropped peptidyl-tRNA and C-terminal peptides, protein derivatives were expressed in the presence of cold [^{32}S]-Met instead of [^{35}S]-Met and analyzed by LC-ESI MS after treatment of PTH (**Supplementary Figure 19-21**).

A

Peptide name	Original protein	Peptide construct
FLP01	YhfK	fMF**PP**MWRRLIYHPD-FLAG
FLP02	HofN	fMN**PP**INFLPWRQQR-FLAG
FLP03	YidD	fMA**PP**LSPGSRVLIA-FLAG
FLP04	YjgZ	fML**PP**GPLLVLPVGA-FLAG
FLP05	PrpR	fMAH**PP**RLNDDKPVIW-FLAG
FLP06	YnjA	fMGL**PP**LSKIPLILRP-FLAG
FLP07	YhhM	fMSK**PP**LFFIVIIGLI-FLAG
FLP08	SfsA	fMEFS**PP**LQRATLIQRY-FLAG
FLP09	RecT	fMTKQ**PP**IAKADLQKTQ-FLAG
FLP10	MdtB	fMQVL**PP**SSTGGPSRLF-FLAG
FLP11	YggW	fMVKL**PP**LSLYIHIPWC-FLAG
FLP12	AldB	fMTNN**PP**SAQIKPGEYG-FLAG
FLP13	YaiA	fMPTK**PP**YPREAYIVTI-FLAG
FLP14	RutD	fMKLSLS**PPP**YADAPVVVLI-FLAG
FLP15	YdcO	fMRLFSI**PPP**TLLAGFLAVL-FLAG
FLP16	Flk	fMIQPISG**PPP**GQ**PP**GQGDNL-FLAG

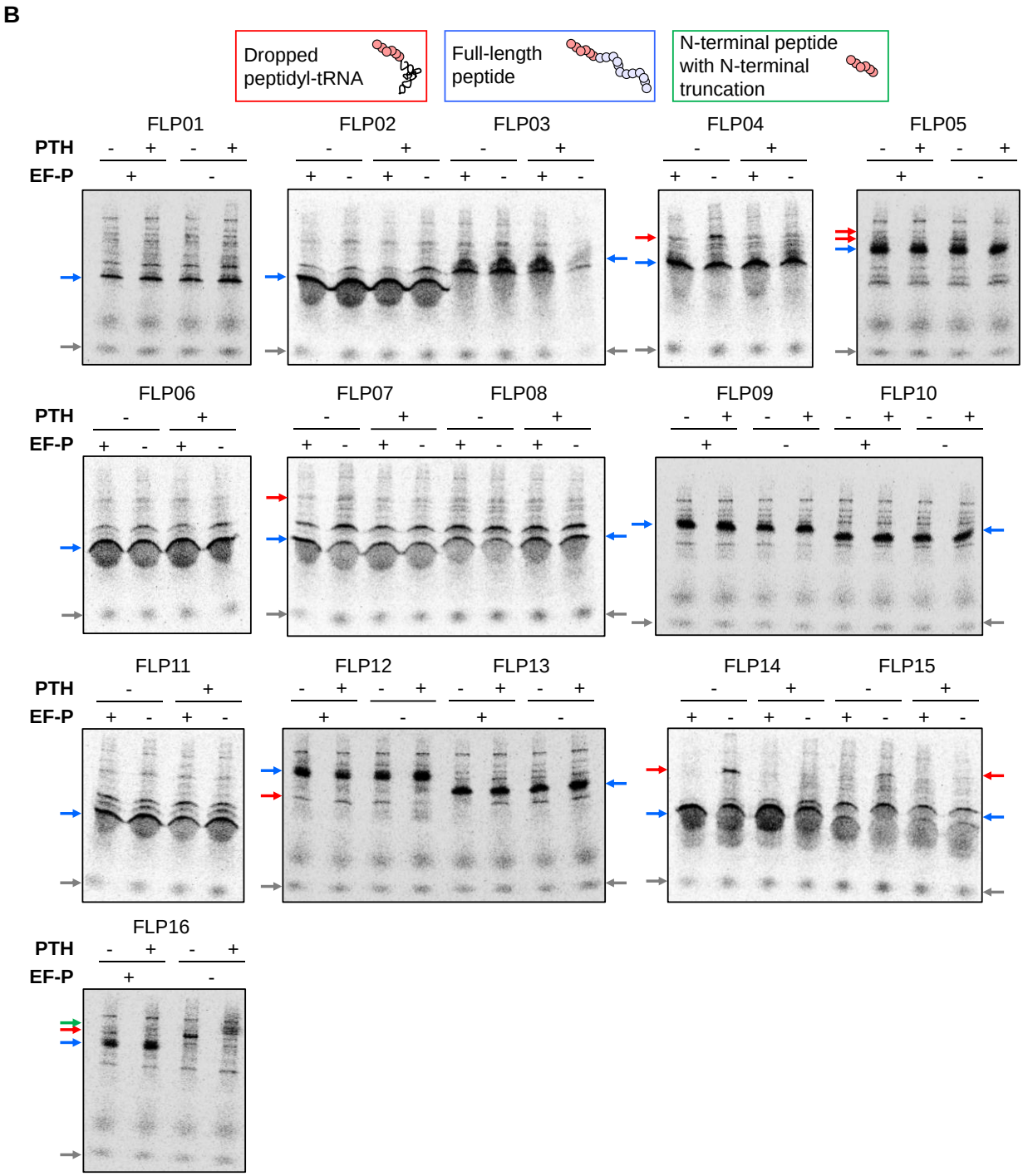


Figure 21 *In vitro* expression of N-terminal sequences of *E. coli*'s polyproline-containing proteins.

(A) 16 kinds of polyproline-containing peptides derived from N-terminal sequences of *E. coli* proteins/putative proteins were used. Their N-terminal sequences were C-terminally FLAG-tagged.

(B) Detection of peptidyl-tRNA drop-off in *in vitro* expression of polyproline-containing peptides. Peptides were N-terminally radioisotope-labelled by [^{35}S]-Met with or without EF-P and treated with or without PTH. Bands corresponding to dropped peptidyl-tRNA were identified by comparing PTH(-)/EF-P(-) lanes and PTH(+)/EF-P(-) lanes. Bands corresponding to peptidyl-tRNA in the stalled ribosome were identified by comparing PTH(+)/EF-P(+) lanes and PTH(+)/EF-P(-) lanes.

3.2.2. Preparation of plasmids coding polyproline-containing proteins and coding EF-P

As the preliminary data indicated that peptidyl-tRNA dropped off in synthesis of protein derivative *in vitro*, I further investigated the drop-off event and generation of C-terminal peptides in the *E. coli*'s endogenous protein synthesis system. Five proteins and putative proteins (YjgZ, PrpR, YhhM, RutD, and YdcO) showing drop-off were cloned into pBAD-HisA (P_{BAD} promotor, ampicillin resistance, colE1 ori), pUC18 (P_{lac} promotor), and pSF-TAC (P_{tac} promotor) plasmids and expressed in the *E. coli* Δ*efp* (JW4107 strain) in which chromosomal *efp* gene was knocked out[123]. Since *E. coli* Δ*efp* does not express T7 RNA polymerase, I chose above plasmids having non-T7 promoters transcribed by *E. coli*'s endogenous RNA polymerases. DNA coding ORF of above proteins and plasmid bodies were amplified by adopter PCR and inversion PCR respectively and assembled by homologous recombination (**Figure 22**). Among the three plasmids, proteins were successfully cloned into pBAD-HisA. In order to complement of *efp* gene in *E. coli* Δ*efp* , *efp* gene was cloned into modified pBAD-*efp*-CmR-p15A plasmid (P_{BAD} promotor, chloramphenicol resistance (Cm^R), p15A ori), in which antibiotics resistance and replication origin were derived from pLysS plasmid to co-express the polyproline-containing proteins and EF-P by utilizing P_{BAD} promotors (**Figure 22**).

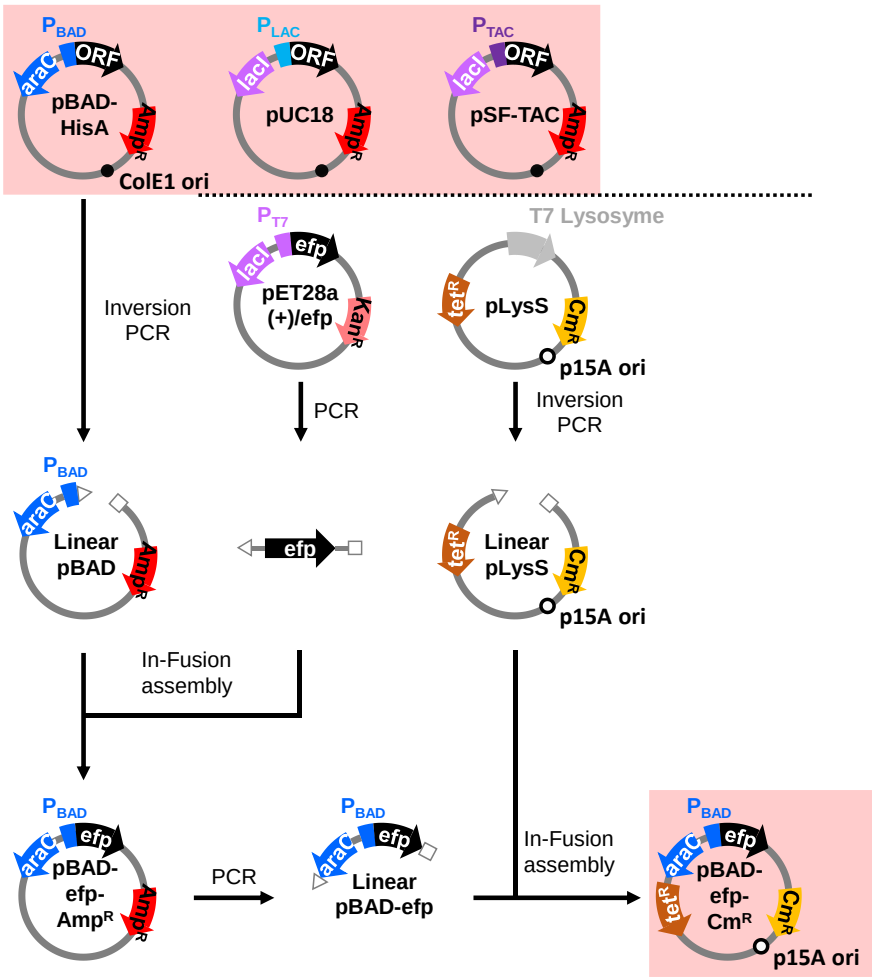


Figure 22 Preparation of plasmids coding polyproline-containing proteins and EF-P.
pBAD-HisA, pUC18, and pSF-TAC plasmids coding polyproline-containing proteins and pBAD-*efp*-Cm^R-p15A Ori were prepared by homologous recombination (In-Fusion) of ORF DNAs and plasmid bodies.

3.2.3. Expression of polyproline-containing proteins without/with EF-P

E. coli Δ*efp* was transformed by pBAD-*efp*-CmR-p15A ori for complementation of EF-P. In resulting *E. coli* Δ*efp*/+*efp*, expression and post-translational modification state of complemented EF-P was confirmed by digestion by Glu-C followed by identification by LC-ESI MS (**Figure 23**). As a result, expression of EF-P was induced by L-arabinose inducer without leaky expression (**Figure 23A**). LC-ESI MS analysis revealed that K34 of EF-P was ε(R)-β-lysyl-hydroxylated, which is native and active form (**Figure 23B**).

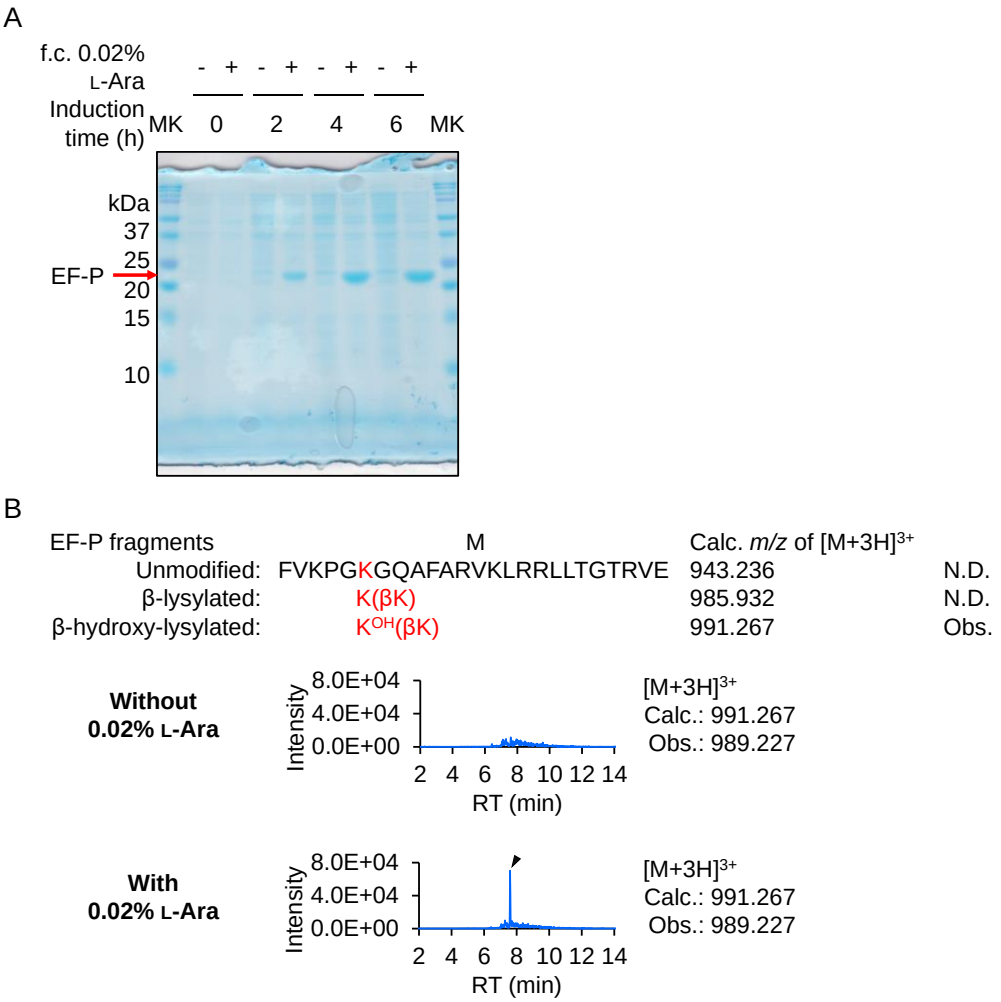


Figure 23 Expression of EF-P from a complemental plasmid and confirmation of its post-translational modification state by Glu-C digestion and LC-ESI MS.

(A) EF-P expression induced by L-arabinose. There was no leaky expression of EF-P regulated by P_{BAD} promotor. L-Ara: L-arabinose, MK: protein molecular weight marker.

(B) Post-translational modification state of EF-P was confirmed by LC-ESI MS after digestion by Glu-C. Without induction by L-arabinose, no peak was detected. However, with induction by L-arabinose, only the ε(R)-β-lysyl-hydroxylated EF-P fragment was detected.

Then, polyproline-containing proteins encoded in pBAD-HisA were expressed in *E. coli* Δ*efp* and *E. coli*/+*efp*. Initially, proteins were expressed with C-terminal His_6 tag (**Figure 24A**) and purified by a Ni NTA-immobilized resin but resulted in quite low yield and low purity due to quite low expression level of proteins plausibly caused by the ribosome stalling and peptidyl-tRNA drop-off induced by consecutive prolines and lack of EF-P. In order to improve their purity, the proteins were dual-tagged by C-terminal PreScission tag followed by His_6 tag or C-terminal FLAG tag followed by His_6 tag (**Figure 24A**). Proteins tagged with PreScission-His_6 were purified by Ni NTA-immobilized resin, digested by Turbo3C protease to remove His_6 tag and recovered from flow-through fractions of 2^{nd} Ni NTA-immobilized resin (**Figure 24B**). Proteins tagged with FLAG-His_6 were purified by Ni NTA-immobilized resin followed by purification by Anti-FLAG M2 antibody agarose (**Figure 24B**). As a result, two putative proteins (YhhM and RutD) were expressed and purified sufficiently as the FLAG-His_6 constructs (**Figure 25A** Lane10: YhhM-FLAG-His6 expressed without EF-P, Lane12: YhhM-FLAG-His6 co-expressed with EF-P, **Figure 25C** Lane23: RutD-FLAG-His_6 expressed without EF-P, Lane24: RutD-FLAG-His_6 co-expressed with EF-P). Dual affinity purification greatly improved the purity of resulting proteins. RutD-FLAG-His_6 was poorly expressed without EF-P but its expression level was recovered by co-expression of EF-P (**Figure 25C**, Lane23 and Lane24). Because of small difference in molecular weight between full-length proteins and proteins lacking N-terminal amino acids, they could not be separated in glycine-SDS PAGE which further investigated by LC-ESI MS/MS after protease digestion.

A

YhhM-His$_6$	(14.4 kDa)	: YhhM-HHHHHH
YhhM-PreScission-His$_6$	(15.5 kDa)	: YhhM-LEVLFQGP-SS-HHHHHH
YhhM-FLAG-His$_6$	(15.6 kDa)	: YhhM-DYKDDDDK-SS-HHHHHH
RutD-His$_6$ (29 kDa)	(29.7 kDa)	: RutD-HHHHHH
RutD-PreScission-His$_6$	(30.8 kDa)	: RutD-LEVLFQGP-SS-HHHHHH
RutD-FLAG-His$_6$	(31.0 kDa)	: RutD-DYKDDDDK-SS-HHHHHH

B

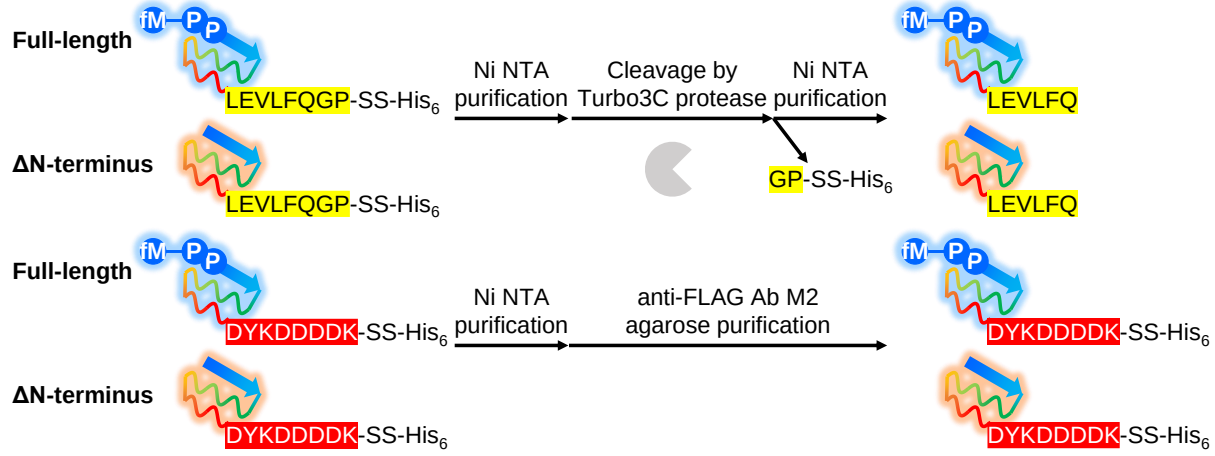


Figure 24 Constructs of YhhM and RutD proteins.

(A) Three kinds of YhhM and RutD protein constructs were prepared. YhhM and RutD were C-terminally tagged with His$_6$, PreScission-His$_6$, and FLAG-His$_6$ peptides. PreScission tag is cleaved by Turbo3C protease at the C-terminus of glutamine (Q). Molecular weights of full-length constructs are shown.

(B) Purification procedures after expression of polyproline-containing proteins without or with EF-P. His$_6$ tagged proteins were purified by only Ni NTA-immobilized resin. For PreScission-His$_6$ tagged constructs, after 1st Ni NTA purification, His$_6$ tag was removed by Turbo3C protease and flow-through fraction of 2nd Ni NTA purification was recovered. For FLAG-His$_6$ tagged constructs, proteins were purified in two step, Ni NTA purification followed by anti-FLAG M2 agarose gel purification.

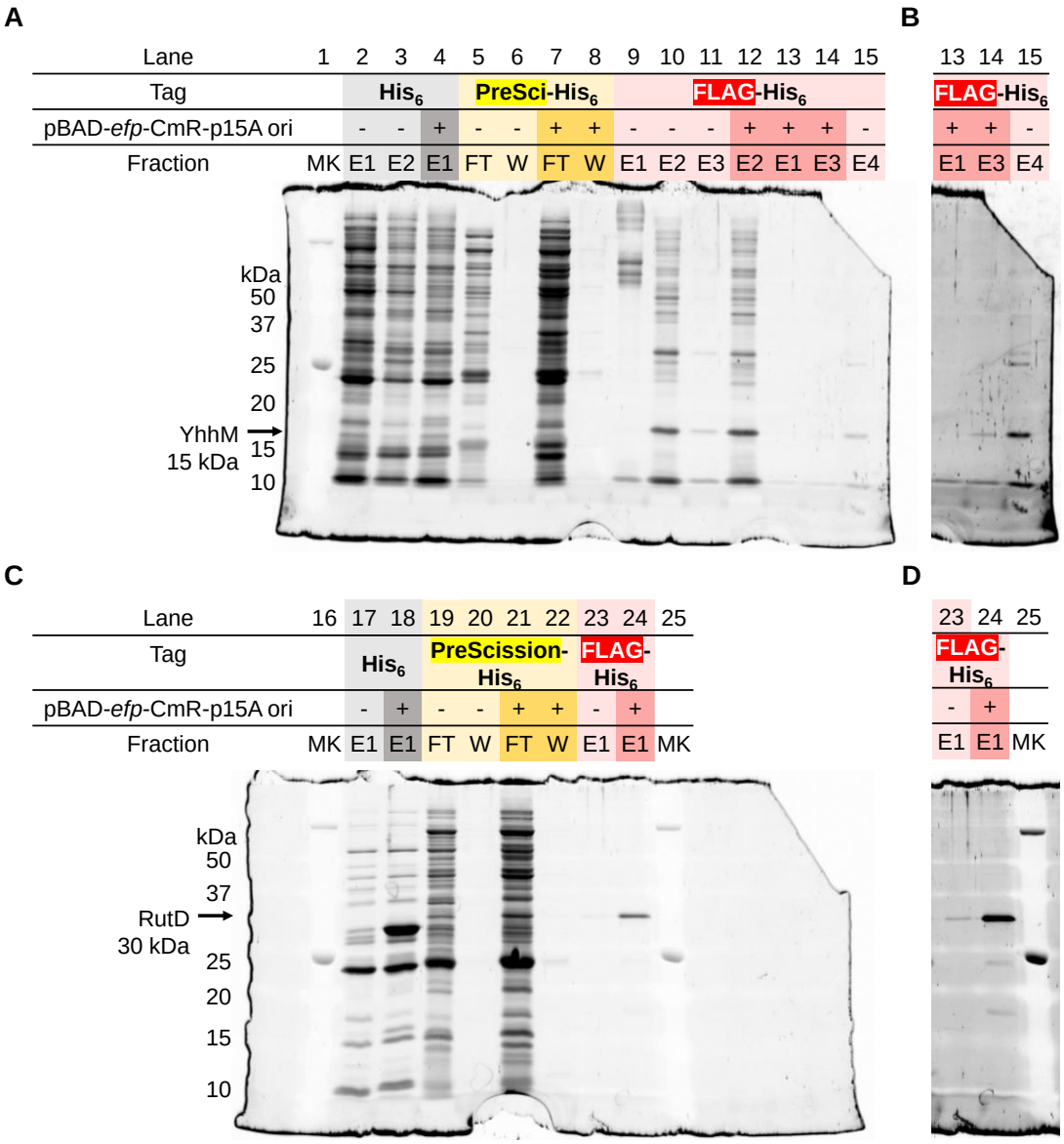


Figure 25 Affinity purification results of YhhM and RutD constructs.

(A) Purified fractions of YhhM constructs visualized in Gly-SDS-PAGE stained by SYPRO Ruby (Thermo Fisher Invitrogen). MK; Precision Plus Protein Dual Color marker (BioRad), E: elution, FT: flow-through, W: wash.

(B) High-contrasted gel image of YhhM-FLAG-His_6 constructs.

(C) Purified fractions of RutD constructs visualized in Gly-SDS-PAGE stained by SYPRO Ruby.

(D) High-contrasted gel image of RutD-FLAG-His_6 constructs.

3.2.4. LC-ESI MS/MS sequencing of N-terminus of polyproline-containing proteins

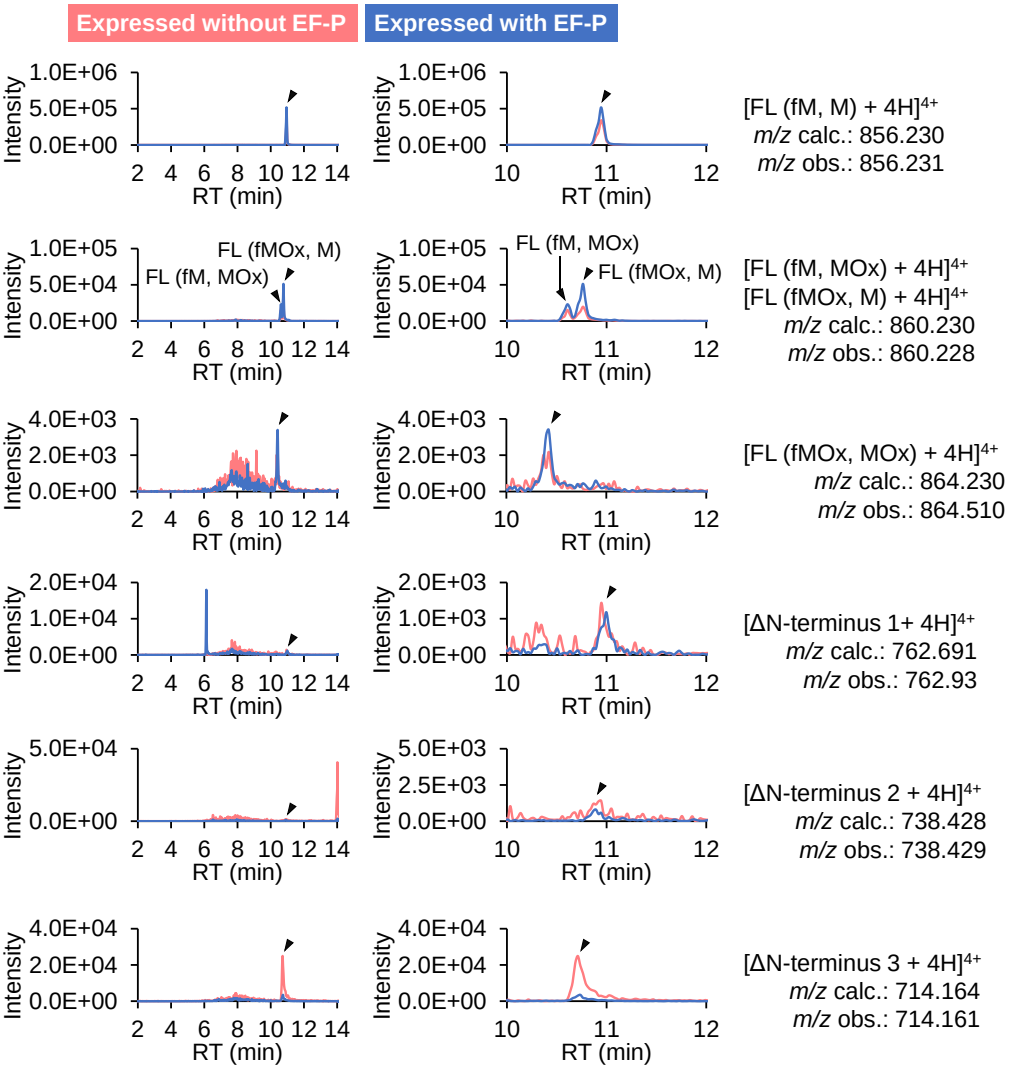
In order to demonstrate the peptidyl-tRNA drop-off and generation of a polypeptides lacking its N-terminal amino acids, the purified YhhM-FLAG-His_6 (**Figure 25A** Lane10: YhhM-FLAG-His6 expressed without EF-P, Lane12: YhhM-FLAG-His6 co-expressed with EF-P) and RutD-FLAG-His_6 (**Figure 25C** Lane23: RutD-FLAG-His_6 expressed without EF-P, Lane24: RutD-FLAG-His_6 co-expressed with EF-P) were digested by proteases and

sequenced by LC-ESI MS/MS. In initial screening by utilizing five proteases (Glu-C, Asp-N, Lys-N, Arg-C, and Lys-C), peptide fragments of YhhM-FLAG-His_6 digested by Lys-N (**Figure 26A, B, Figure 29**) and peptide fragments of RutD-FLAG-His_6 digested by Glu-C (**Figure 30A, B, Figure 33**) were successfully detected by LC-ESI MS. Further LC-ESI MS/MS sequencing identified N-terminal peptide fragments derived from the full-length YhhM-FLAG-His_6 and YhhM-FLAG-His_6 lacking three, four, or five amino acid residues from the N-terminal Met (**Figure 27, Figure 28**), along with almost all of peptide fragments derived from Lys-N-digested YhhM-FLAG-His_6 (**Figure 29**). In the case of RutD-FLAG-His_6, peptide fragments derived from only the full-length polypeptide was identified (**Figure 30, Figure 31, Figure 32**). Actually, in the full-length YhhM-FLAG-His_6, the most N-terminal lysine residues was not cleaved by Lys-N (**Figure 27A-D, Figure 28A-D**). Plausibly, the N-terminal sequence prior to the lysine was too short and flexible to be cleaved by Lys-N. Although quantity of the fragment derived from full-length YhhM-FLAG-His_6 was comparable between expression in the absence and presence of EF-P (**Figure 26C,** Ion response (a.u.): $4.9 \pm 1.1\times10^4$ vs $11 \pm 6\times10^4$, **Figure 26D**, Fraction: $93 \pm 3\%$ vs $99 \pm 4\%$), the quantity of the fragment derived from the YhhM-FLAG-His_6 lacking amino acids from the N-terminal Met to the 2nd proline was significantly high in the case of YhhM-FLAG-His_6 expressed in the absence of EF-P (**ΔN-terminus 3, Figure 26B, 26C,** Ion response (a.u.): $3.1 \pm 0.9\times10^3$ vs $6 \pm 4\times10^2$, $p < 0.05$ in Welch's t test, **26D,** Fraction: $6 \pm 3\%$ vs $0.5 \pm 0.3\%$, $p < 0.1$ in Welch's t test). This suggested that whereas the full-length YhhM-FLAG-His_6 successfully synthesized in the absence or presence of EF-P, peptidyl-tRNA dropped off from the ribosome more frequently in the absence of EF-P than in the presence, which resulted in the restart of translation from the middle of the mRNA. Among YhhM-FLAG-His_6 lacking its N-terminal amino acids, the dominant species were truncated fragment between the 2nd proline and posterior leucine (**Figure 26C,** Ion response of ΔN-terminus 1/2/3 expressed in the absence of EF-P: $1.6 \pm 1.2\times10^2/0.8 \pm 0.3\times10^2/3.1 \pm 0.9\times10^3$, those in the presence of EF-P: $1.5 \pm 1.5\times10^2/1.4 \pm 0.9\times10^2/6 \pm 4\times10^2$, **26D,** Fraction of ΔN-terminus 1/2/3 expressed in the absence of EF-P: $0.3 \pm 0.2\%/0.15 \pm 0.05\%/6 \pm 3\%$, those in the presence of EF-P: $0.2 \pm 0.1\%/0.12 \pm 0.07\%/0.5 \pm 0.3\%$). This would be because peptidyl transfer between peptidyl-diprolyl-tRNA and leucyl-tRNA should be the slowest reaction in ribosomal synthesis of YhhM-FLAG-His_6 in the absence of EF-P, which induced the ribosome stalling and drop-off of peptidyl-diprolyl-tRNA. In conclusion, in protein synthesis in *E. coli*, ribosome restarted the translation from the middle of the mRNA after peptidyl-tRNA drop-off. EF-P facilitated the synthesis of full-length YhhM-FLAG-His_6 and suppressed generation of YhhM-FLAG-His_6 lacking N-terminal amino acids, which was consistent with the *in vitro* result. Even in the presence of EF-P, YhhM-FLAG-His_6 lacking its N-terminal amino acids was detected, indicating that a short peptidyl-tRNA is prone to drop-off from the ribosome in the canonical translation.

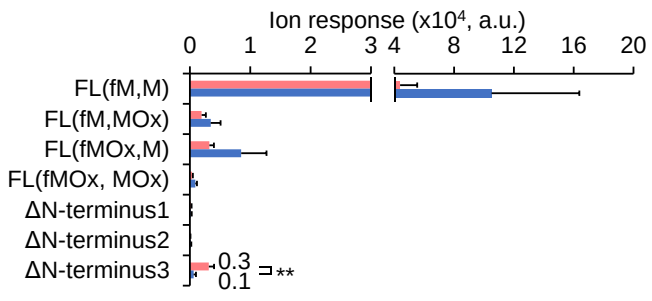
A

Name	M	*z*	*m/z* of $[M + 4H]^{4+}$
FL (fM, M)	fMSKPPLFFIVIIGLIVVAASFRFMQQRRE	4	856.231
FL (fM, MOx)	fMSKPPLFFIVIIGLIVVAASFRFM(Ox)QQRRE	4	860.228
FL (fMOx, M)	fM(Ox)SKPPLFFIVIIGLIVVAASFRFMQQRRE	4	860.228
FL (fMOx, MOx)	fM(Ox)SKPPLFFIVIIGLIVVAASFRFM(Ox)QQRRE	4	864.510
ΔN-terminus 1	PPLFFIVIIGLIVVAASFRFMQQRRE	4	762.930
ΔN-terminus 2	PLFFIVIIGLIVVAASFRFMQQRRE	4	738.429
ΔN-terminus 3	LFFIVIIGLIVVAASFRFMQQRRE	4	714.161

B



C



D

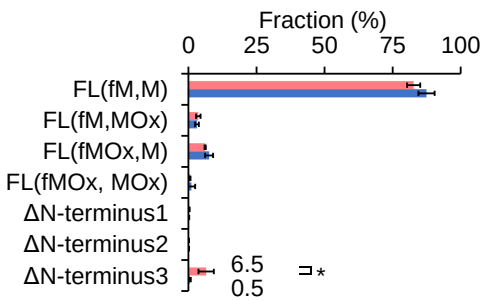


Figure 26 LC-ESI MS screening of N-terminal fragments of YhhM-FLAG-His$_6$ expressed in the

absence or presence of EF-P.

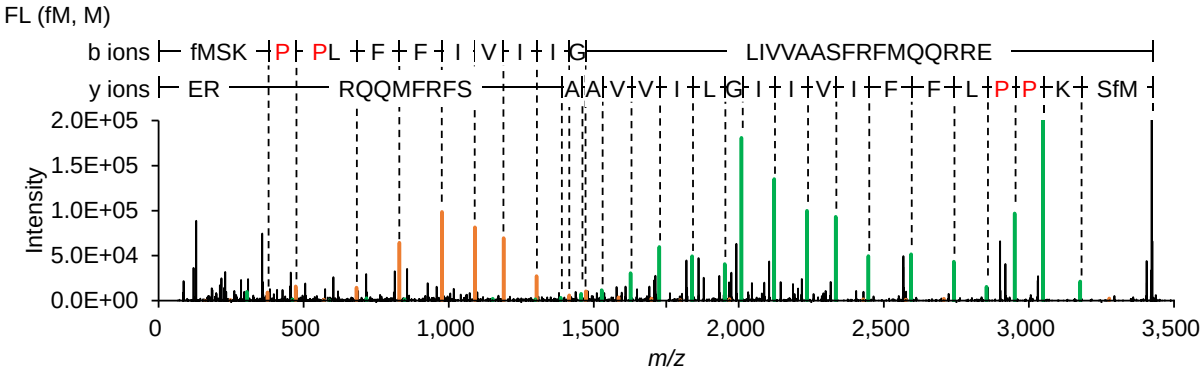
(A) Identified peptide sequences derived from Lys-N digestion of full-length (FL) YhhM-FLAG-His_6 and YhhM-FLAG-His_6 lacking its N-terminal amino acids. MOx: Methionine oxidized on its sidechain.

(B) Full-range (Left) and magnified (middle) XICs of N-terminal peptide fragments derived from Lys-N digestion of YhhM-FLAG-His_6 expressed alone (Red) and co-expressed with EF-P (Blue).

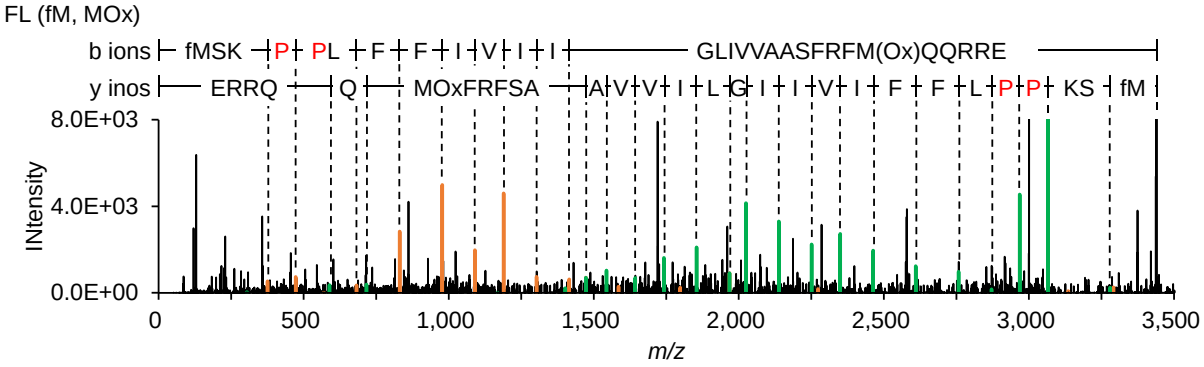
(C) LC-ESI MS quantification of ion response of N-terminal fragments generated by Lys-N digestion of YhhM-FLAG-His_6 expressed in the absence or presence of EF-P. SD of three independent experiment of digestion of YhhM-FLAG-His_6 expressed without EF-P and SD of four independent experiment of digestion of YhhM-FLAG-His_6 expressed with EF-P are shown. **: $p < 0.05$ in Welch's t test.

(D) LC-ESI MS quantification of ratios of N-terminal fragments generated by Lys-N digestion of YhhM-FLAG-His_6 expressed in the absence or presence of EF-P. Ratios were calculated as (an ion response of a N-terminal fragments)/(sum of ion responses of all the N-terminal fragments). SD of three independent experiments of digestion of YhhM-FLAG-His_6 expressed without EF-P and SD of four independent experiments of digestion of YhhM-FLAG-His_6 expressed with EF-P are shown. *: $p < 0.1$ in Welch's t test.

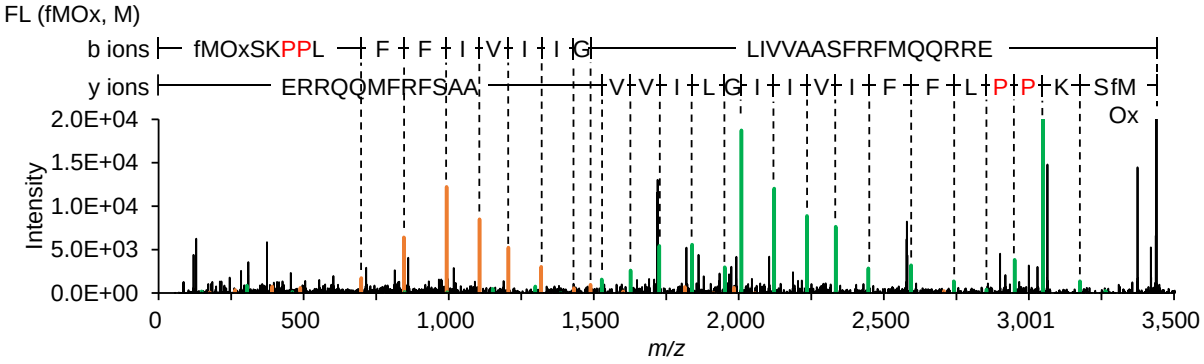
A


FL (fM, M)
b ions fMSK P PL F F I V I I G LIVVAASFRFMQQRRE
y ions ER RQQMFRFS AAVVVIILGIIVIFFLPPKSfM
Intensity
2.0E+05
1.5E+05
1.0E+05
5.0E+04
0.0E+00
0
500
1,000
1,500
2,000
2,500
3,000
3,500
m/z

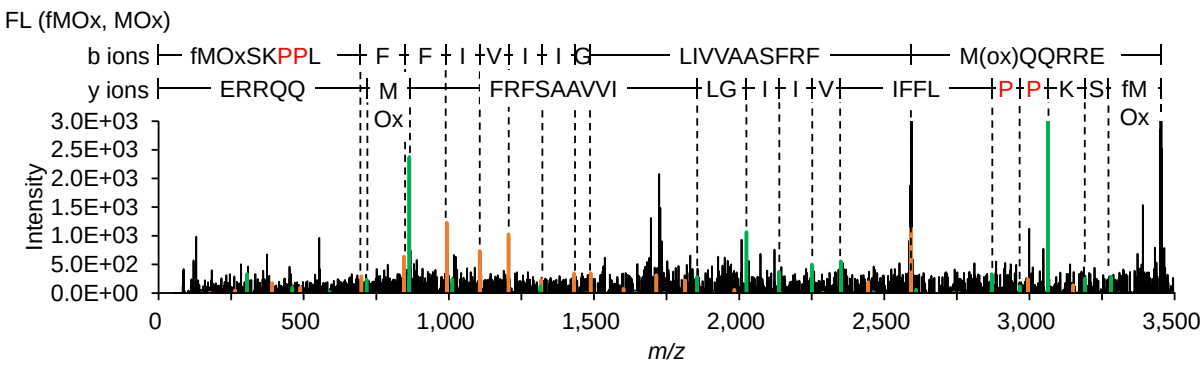
B


FL (fM, MOx)
b ions fMSK P PL F F I V I I GLIVVAASFRFM(Ox)QQRRE
y inos ERRQ Q MOxFRFSA AVVVIILGIIVIFFLPPKS fM
INtensity
8.0E+03
4.0E+03
0.0E+00
0
500
1,000
1,500
2,000
2,500
3,000
3,500
m/z

C


FL (fMOx, M)
b ions fMOxSKPPL F F I V I I G LIVVAASFRFMQQRRE
y ions ERRQQMFRFSAA VVIILGIIVIFFLPPKSfM
Ox
Intensity
2.0E+04
1.5E+04
1.0E+04
5.0E+03
0.0E+00
0
500
1,000
1,500
2,000
2,500
3,001
3,500
m/z

D


FL (fMOx, MOx)
b ions fMOxSKPPL F F I V I I G LIVVAASFRF M(ox)QQRRE
y ions ERRQQ M FRFSAAVVI LG I I V IFFL PPKS fM
Ox
Ox
Intensity
3.0E+03
2.5E+03
2.0E+03
1.5E+03
1.0E+03
5.0E+02
0.0E+00
0
500
1,000
1,500
2,000
2,500
3,000
3,500
m/z

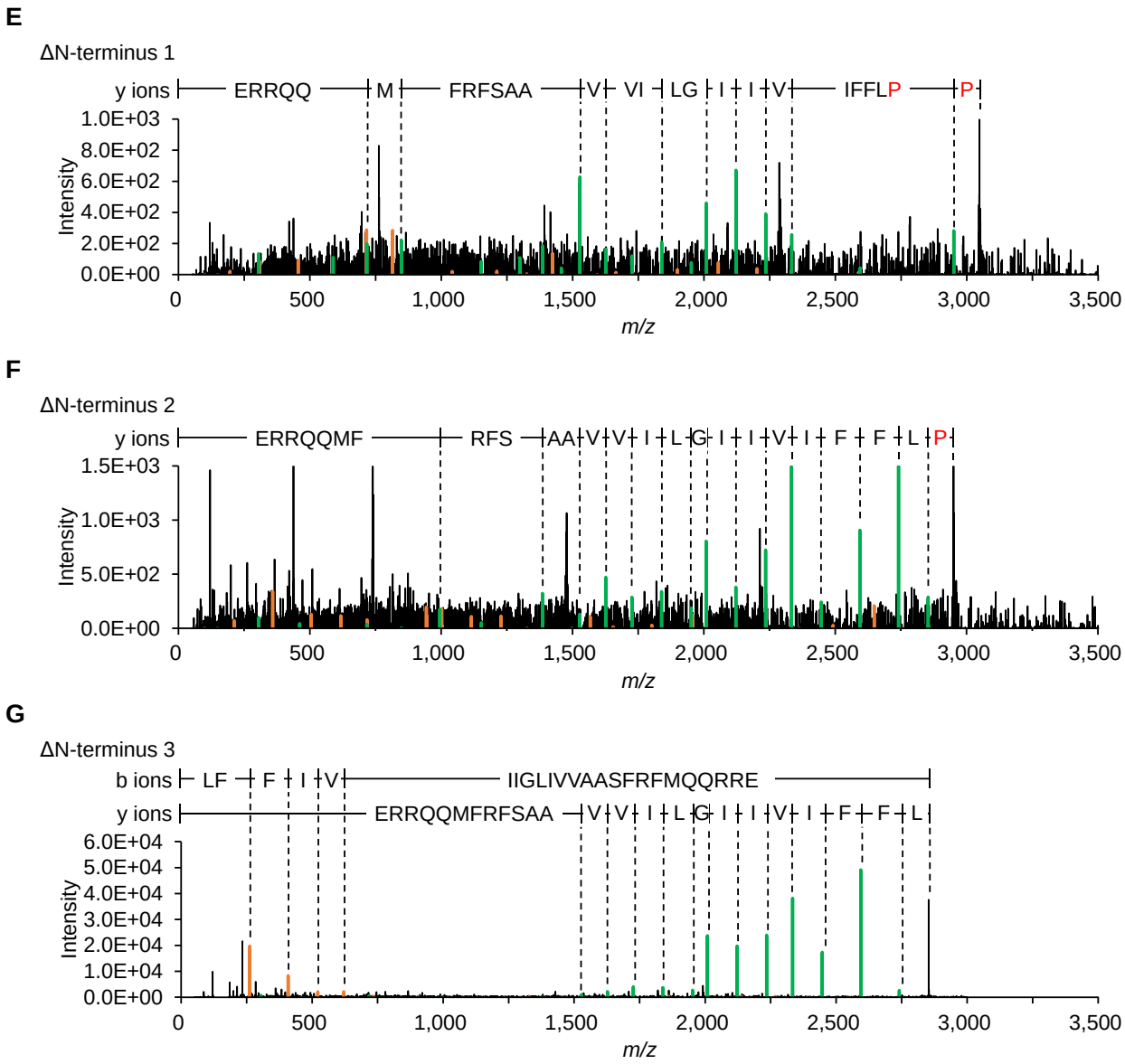


Figure 27 Lys-N digestion and LC-ESI MS/MS sequencing of N-terminal fragments of YhhM-FLAG-His_6 expressed without EF-P.

(A) Deconvoluted LC-ESI MS/MS spectrum of FL (fM, M) at RT = 10.95 min. Green peaks are corresponding to daughter y ions and orange peaks are corresponding to daughter b ions.

(B) Deconvoluted LC-ESI MS/MS spectrum of FL (fM, MOx) at RT = 10.62 min.

(C) Deconvoluted LC-ESI MS/MS spectrum of FL (fMOx, M) at RT = 10.77 min.

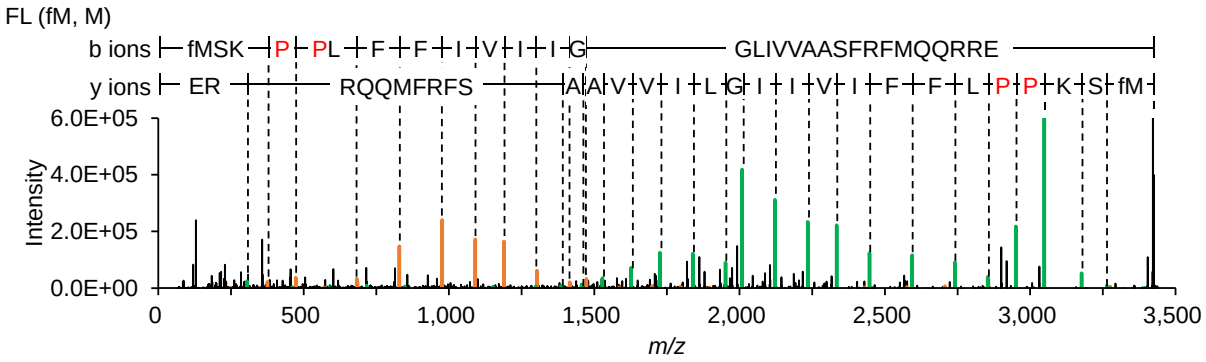
(D) Deconvoluted LC-ESI MS/MS spectrum of FL (fMOx, MOx) at RT = 10.41 min.

(E) Deconvoluted LC-ESI MS/MS spectrum of ΔN-terminus 1 at RT = 10.96 min.

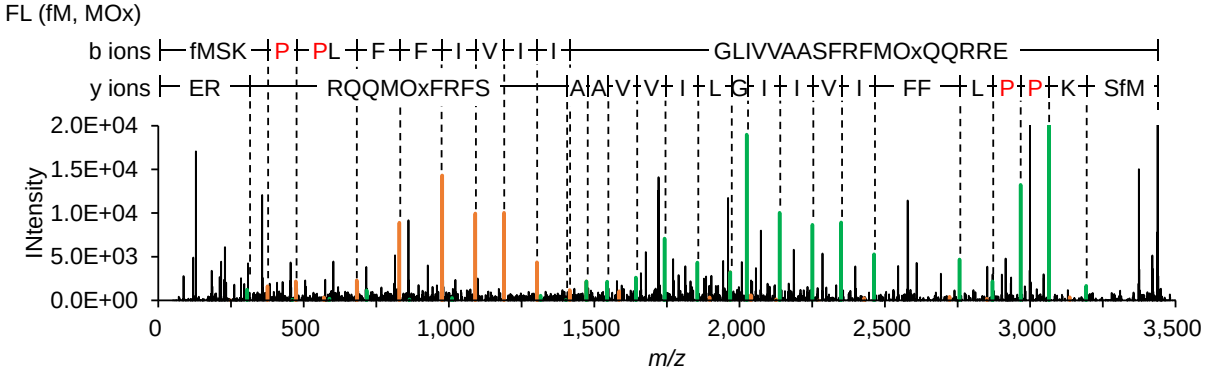
(F) Deconvoluted LC-ESI MS/MS spectrum of ΔN-terminus 2 at RT = 10.86 min.

(G) Deconvoluted LC-ESI MS/MS spectrum of ΔN-terminus 3 at RT = 10.72 min.

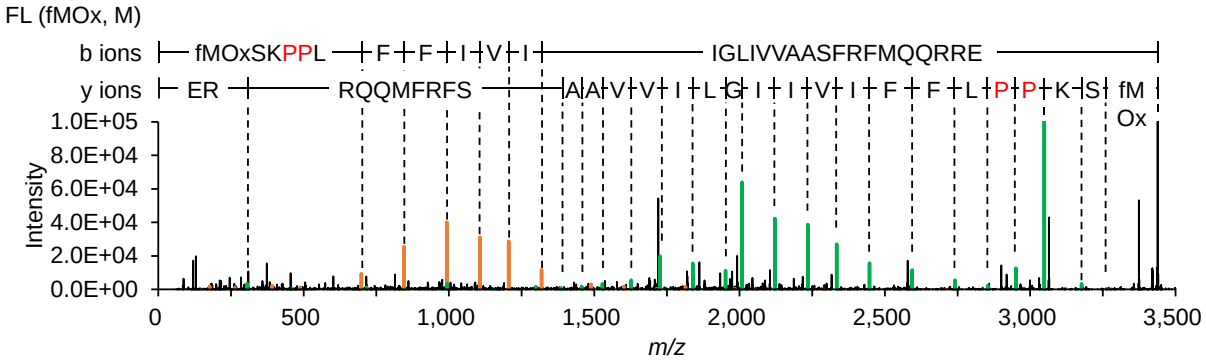
A


FL (fM, M)
b ions
y ions
Intensity
m/z

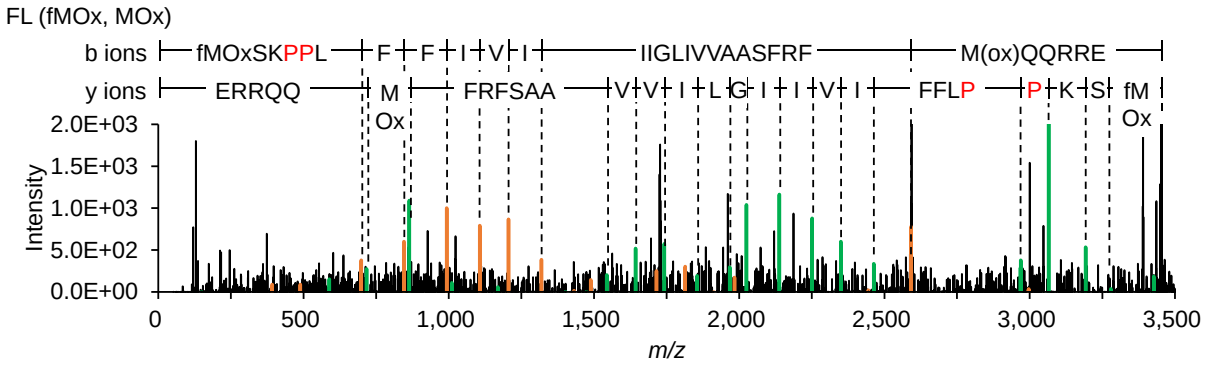
B


FL (fM, MOx)
b ions
y ions
INtensity
m/z

C


FL (fMOx, M)
b ions
y ions
Intensity
m/z

D


FL (fMOx, MOx)
b ions
y ions
Intensity
m/z

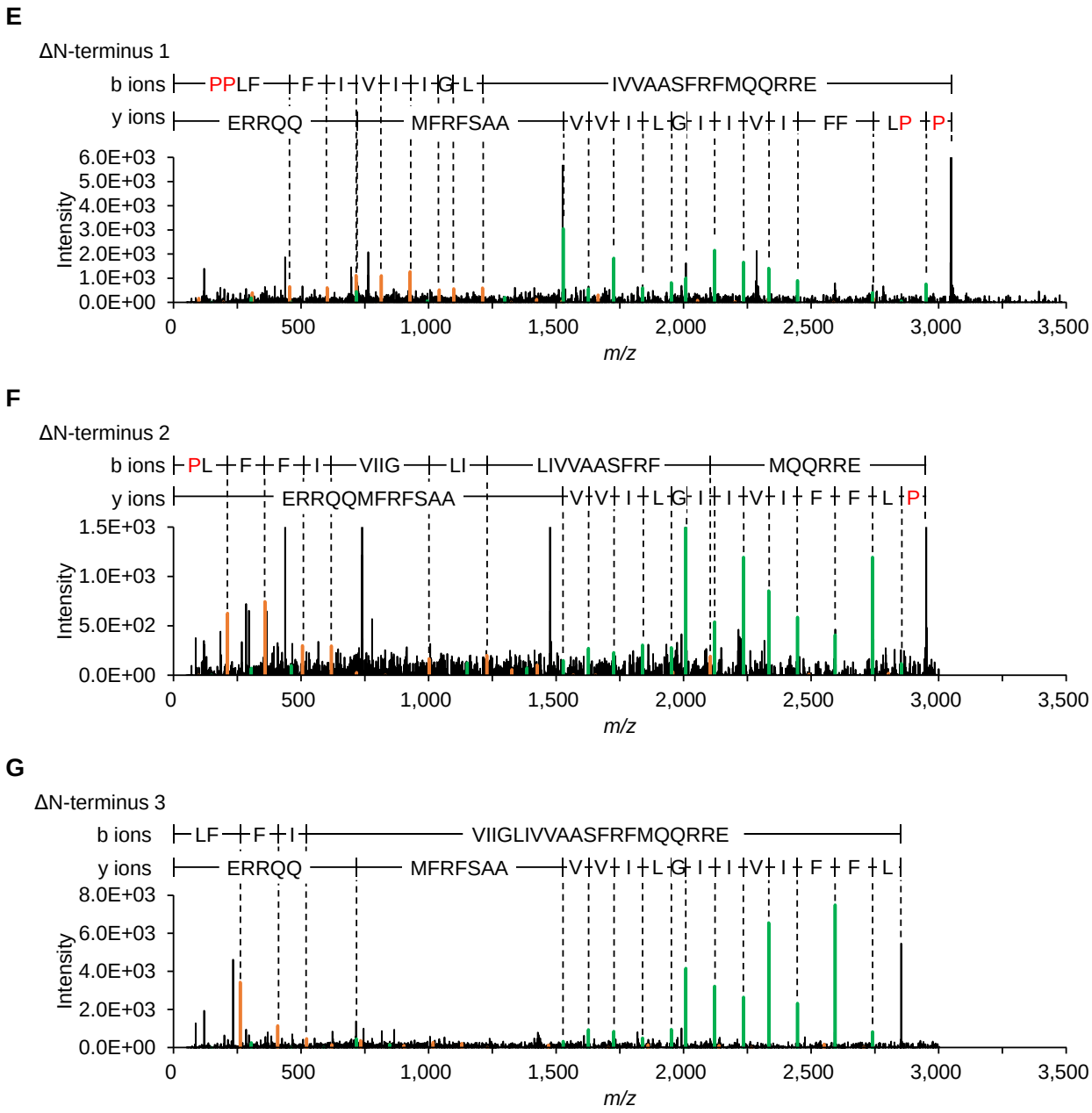


Figure 28 Lys-N digestion and LC-ESI MS/MS sequencing of N-terminal fragments of YhhM-FLAG-His_6 co-expressed with EF-P.

(A) Deconvoluted LC-ESI MS/MS spectrum of FL (fM, M) at RT = 10.95 min. Green peaks are corresponding to daughter y ions and orange peaks are corresponding to daughter b ions.

(B) Deconvoluted LC-ESI MS/MS spectrum of FL (fM, MOx) at RT = 10.62 min.

(C) Deconvoluted LC-ESI MS/MS spectrum of FL (fMOx, M) at RT = 10.77 min.

(D) Deconvoluted LC-ESI MS/MS spectrum of FL (fMOx, MOx) at RT = 10.41 min.

(E) Deconvoluted LC-ESI MS/MS spectrum of ΔN-terminus 1 at RT = 10.96 min.

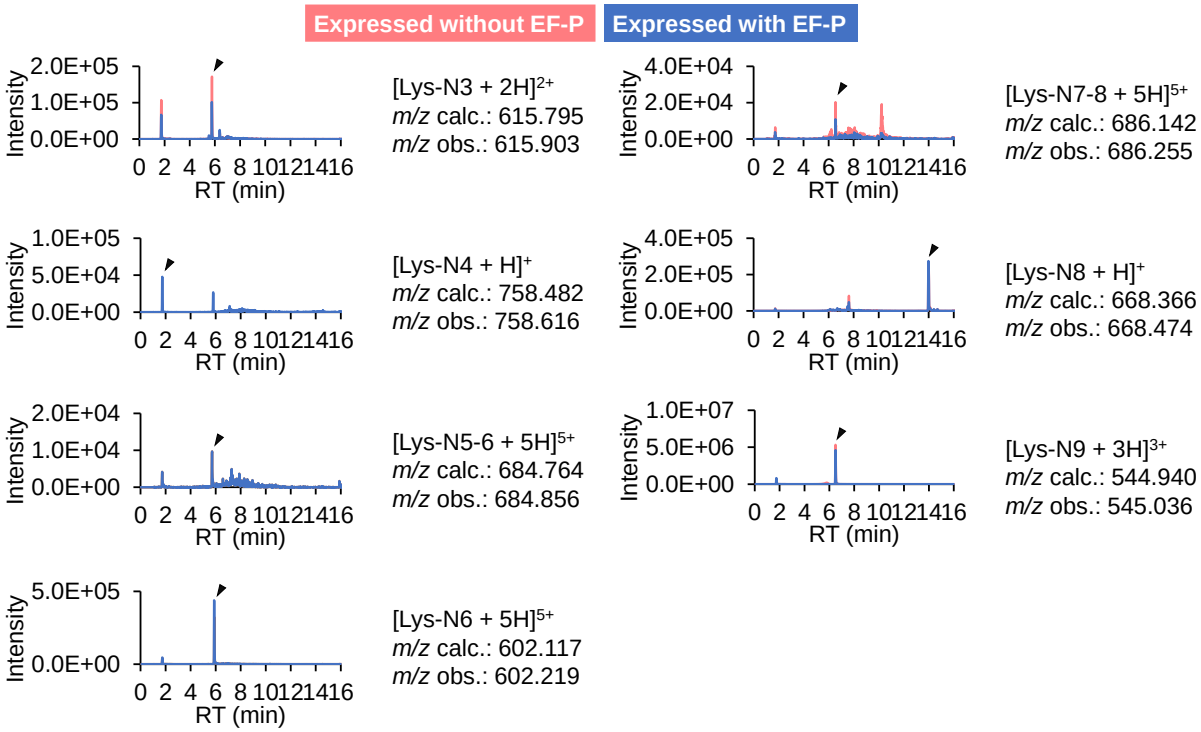
(F) Deconvoluted LC-ESI MS/MS spectrum of ΔN-terminus 2 at RT = 10.89 min.

(G) Deconvoluted LC-ESI MS/MS spectrum of ΔN-terminus 3 at RT = 10.74 min.

A

Fragment	Peptide	*z*	*m/z* of $[M + zH]^{z+}$	RT (min)
Lys-N3	KADNDMAPLQQ	2	615.903	5.801
Lys-N4	KLVVVSN	1	758.616	1.731
Lys-N5-6	KREKPINDRRSRQQEVTPAGTSIRYEASF	5	684.856	5.759
Lys-N6	KPINDRRSRQQEVTPAGTSIRYEASF	5	602.219	5.896
Lys-N7-8	KPQSGGMEQTFRLDAQQYHALTVGDKGTLSY	5	686.255	6.566
Lys-N8	KGTLSY	1	668.474	13.989
Lys-N9	KGTRFVSFVGEQDY	3	545.036	6.494

B



C

MS KPPLFFIVIIGLIVVAASFRFMQQRRE KADNDMAPLQQ KLVVVSN KRE KPINDRRSRQQEVTPAGTSIRYEASF KPQSGGMEQTFRLDAQQYHALTVGD KGTLSY KGTRFVSFVGEQDY KDDDD KSSHHHHHH*

D

MS KPPLFFIVIIGLIVVAASFRFMQQRRE KADNDMAPLQQ KLVVVSN KRE KPINDRRSRQQEVTPAGTSIRYEASF KPQSGGMEQTFRLDAQQYHALTVGD KGTLSY KGTRFVSFVGEQDY KDDDD KSSHHHHHH*

Figure 29 Sequence coverage of YhhM-FLAG-His$_6$ expressed in the absence and presence of EF-P detected by LC-ESI MS.

(A) Identified peptide fragments derived from Lys-N digestion of YhhM-FLAG-His$_6$.

(B) XICs of peptide fragments derived from Lys-N digestion of YhhM-FLAG-His$_6$ expressed in the absence of EF-P (Red) and presence of EF-P (Blue).

(C) Sequence coverage of identified peptide fragments derived from Lys-N digestion of YhhM-FLAG-His$_6$ expressed in the absence of EF-P. *: Stop codon.

(D) Sequence coverage of identified peptide fragments derived from Lys-N digestion of YhhM-FLAG-His$_6$ expressed in the presence of EF-P. *: Stop codon.

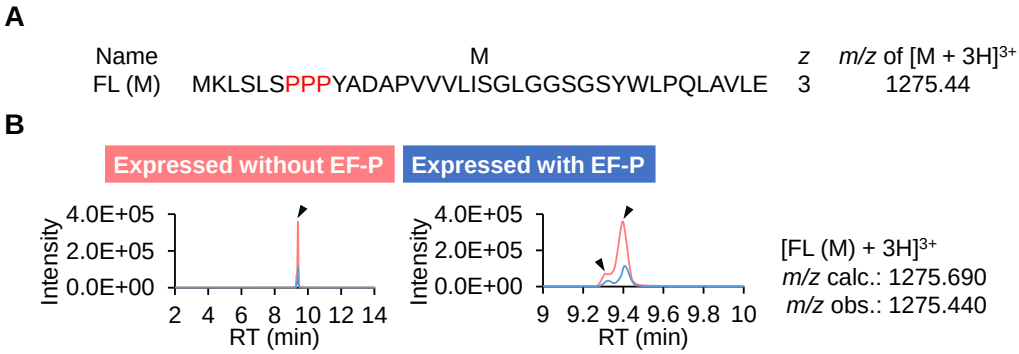


Figure 30 LC-ESI MS screening of N-terminal fragments of RutD-FLAG-His$_6$ expressed in the absence or presence of EF-P.

(A) Identified peptide sequences derived from Glu-C digestion of full-length (FL) RutD-FLAG-His$_6$.

(B) Full-range (Left) and magnified (middle) XICs of N-terminal peptide fragments derived from Glu-C digestion of RutD-FLAG-His$_6$ expressed alone (Red) and co-expressed with EF-P (Blue).

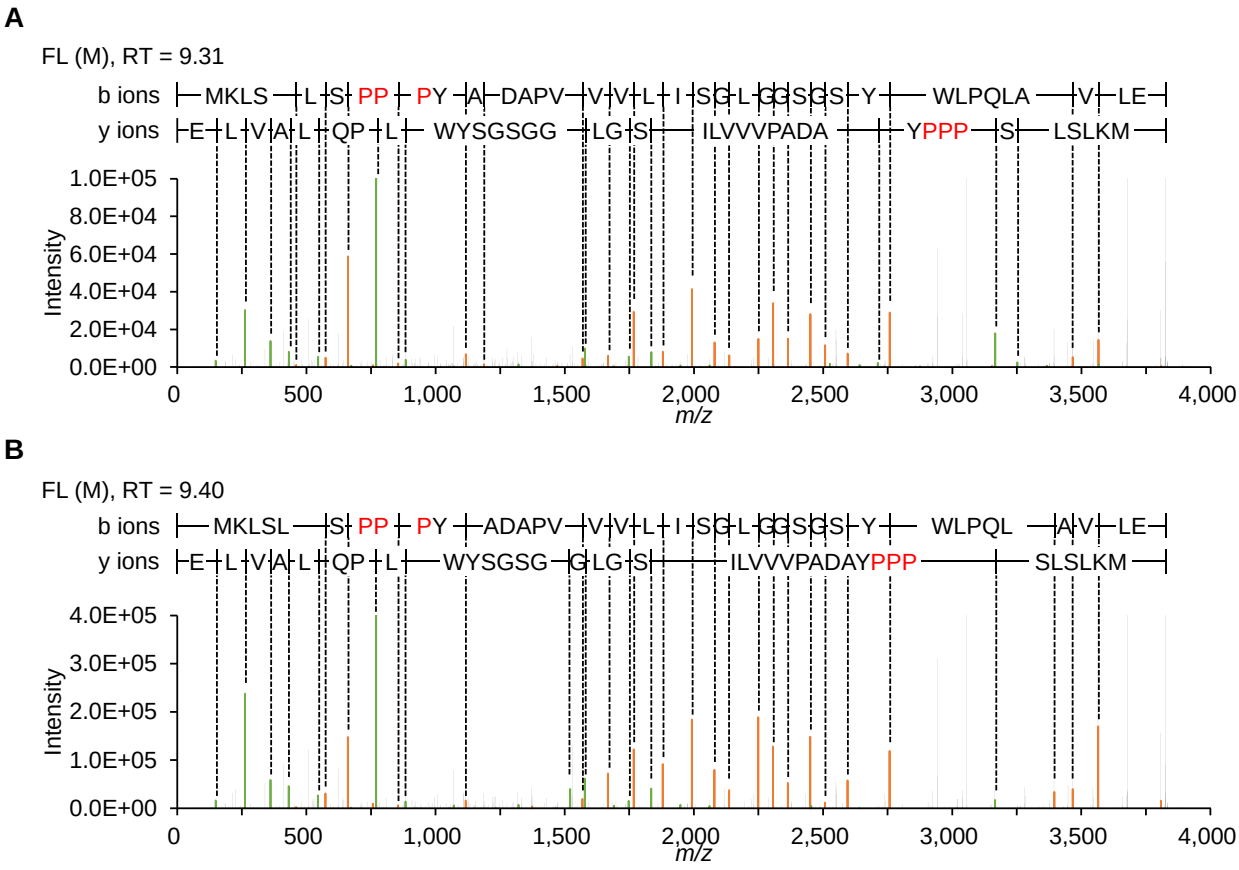


Figure 31 Glu-C digestion and LC-ESI MS/MS sequencing of N-terminal fragments of RutD-FLAG-His$_6$ expressed in the absence of EF-P.

(A) Deconvoluted LC-ESI MS/MS spectrum of FL (M) at RT = 9.31 min. Green peaks are corresponding to daughter y ions and orange peaks are corresponding to daughter b ions.

(B) Deconvoluted LC-ESI MS/MS spectrum of FL (M) at RT = 9.40 min.

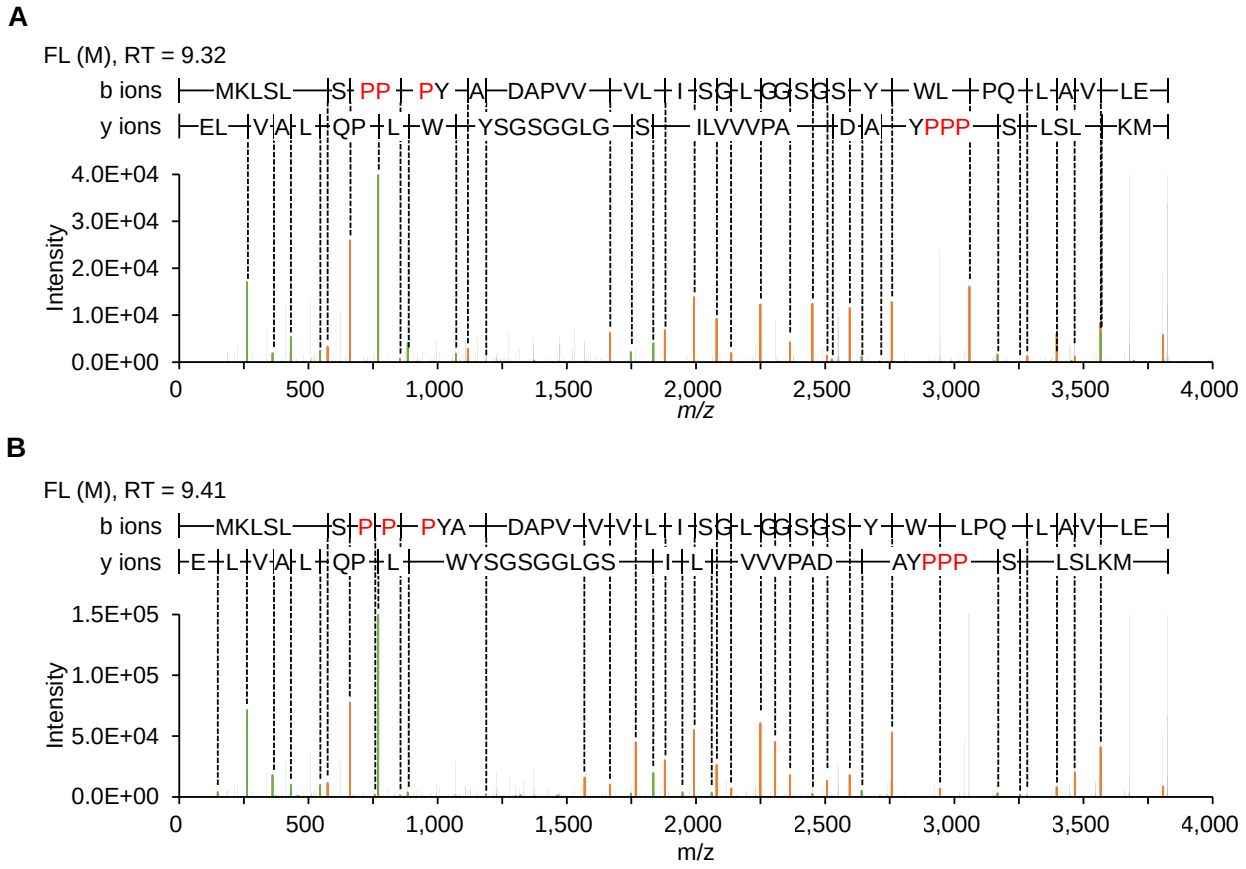


Figure 32 Glu-C digestion and LC-ESI MS/MS sequencing of N-terminal fragments of RutD-FLAG-His_6 co-expressed with EF-P.

(A) Deconvoluted LC-ESI MS/MS spectrum of FL (M) at RT = 9.32 min. Green peaks are corresponding to daughter y ions and orange peaks are corresponding to daughter b ions.

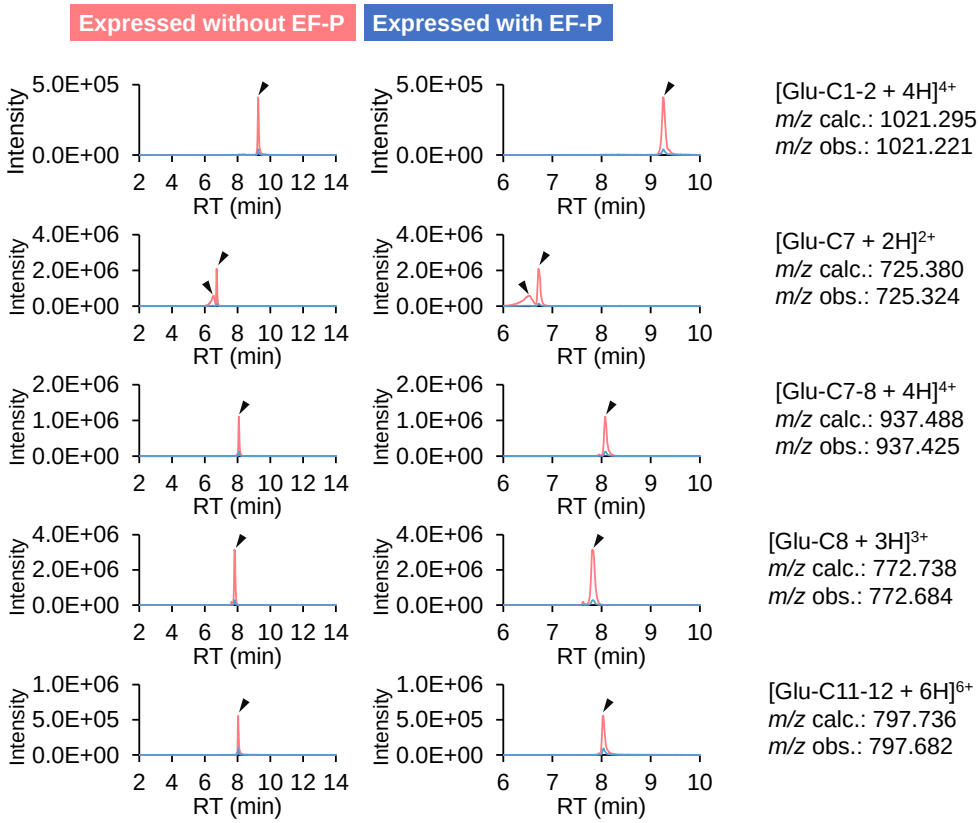
(B) Deconvoluted LC-ESI MS/MS spectrum of FL (M) at RT = 9.41 min.

A

Fragment	Peptide	*z*
Glu-C1-2	MKLSLSPPPYADAPVVVLISGLGGSGSYWLPQLAVLEQE	4
Glu-C7	RLLYSGGAQAWVE	2
Glu-C7-8	RLLYSGGAQAWVE AQPLFLYPADWMAARAPRLE	4
Glu-C8	AQPLFLYPADWMAARAPRLE	3
Glu-C11-12	LHAALPDSQKMVMPYGGHACNVTDPE TFNALLLNGLASLLHHRE	6

Fragment	*m/z* of $[M + zH]^{z+}$	RT (min)
Glu-C1-2	1021.295	9.26
Glu-C7	725.380	6.52, 6.71
Glu-C7-8	937.488	8.07
Glu-C8	772.738	7.81
Glu-C11-12	797.736	8.03

B



C

MKLSLSPPPYADAPVVVLISGLGGSGSYWLPQLAVLE QE YQVVCYDQRGTGNNPDTLAE DYSIAQMAAE
LHQALVAAGIE HYAVVGHALGALVGMQLALDYPASVTVLISVNGWLRINAHTRRCFQVRE RLLYSGGAQAWVE
AQPLFLYPADWMAARAPRLE AE DALALAHFQGKNNLLRRLNALKRADFSHHADRIRCPVQIICASDDLL
VPTACSSE LHAALPDSQKMVMPYGGHACNVTDPE TFNALLLNGLASLLHHRE AALDYKDDDDKSSHHHHHH

D

MKLSLSPPPYADAPVVVLISGLGGSGSYWLPQLAVLE QE YQVVCYDQRGTGNNPDTLAE DYSIAQMAAE
LHQALVAAGIE HYAVVGHALGALVGMQLALDYPASVTVLISVNGWLRINAHTRRCFQVRE RLLYSGGAQAWVE
AQPLFLYPADWMAARAPRLE AE DALALAHFQGKNNLLRRLNALKRADFSHHADRIRCPVQIICASDDLL
VPTACSSE LHAALPDSQKMVMPYGGHACNVTDPE TFNALLLNGLASLLHHRE AALDYKDDDDKSSHHHHHH

Figure 33 Sequence coverage of YhhM-FLAG-His$_6$ expressed in the absence and presence of EF-P detected by LC-ESI MS.

(A) Identified peptide fragments derived from Lys-N digestion of YhhM-FLAG-His$_6$.

(B) XICs of peptide fragments derived from Lys-N digestion of YhhM-FLAG-His$_6$ expressed in the absence

of EF-P (Red) and presence of EF-P (Blue).

(C) Sequence coverage of identified peptide fragments derived from Lys-N digestion of YhhM-FLAG-His_6 expressed in the absence of EF-P. *: Stop codon.

(D) Sequence coverage of identified peptide fragments derived from Lys-N digestion of YhhM-FLAG-His_6 expressed in the presence of EF-P. *: Stop codon.

3.2.5. Physiological indication of peptidyl-tRNA drop-off and translation restart compared to known protein quality control and degradation

Translation restart from the middle of the mRNA is a unique phenomenon compared to other known ribosome stalling rescued by quality control molecules. Besides the dropped peptidyl-tRNA which can be recycled by PTH into immature short peptides and tRNA[35], known ribosome rescue factors, tmRNA/SmpB[124,125], ArfA[126], and ArfB[127] seem not to be able to rescue the stalled ribosome to be generating C-terminal peptides. This is because the both of ribosomal A site and P site where the rescue factors should enter are occupied by peptidyl-diprolyl-tRNA and aminoacyl-tRNA, respectively. RRF is known to induce peptidyl-tRNA drop-off *in vivo* in cooperation of with EF-G and RF3[31]. However, RRF also require vacant A site for entering the 70S ribosome complex, which seems not to involve in the above peptidyl-tRNA drop-off pathway. If stalled 70S ribosomes can be dissociated into two subunits, generation of C-terminal peptides can be terminated and ribosome quality control pathway may degrade the peptidyl-tRNA[128]. However, judging from the *in vitro* and *in vivo* results, the stalled ribosome is still stable to restart translation from the middle of mRNA. Then, the ribosome continues the translation until it reaches to a stop codon, and releases a protein lacking N-terminal amino acids. The released protein would be degraded by the cellular proteases according to the N-end rule pathway[129], in which the recognition pattern would be changed between full-length proteins bearing the N-terminal fM and C-terminal proteins bearing a N-terminal non-fM amino acid. In *E. coli*, there are as many as 2115 proteins and putative proteins containing more than two consecutive proline residues. Among them, 94% of the proline stretch locates far more than 24 amino acids residues from the initiation methionine[82], which can traverse whole the exit tunnel to stabilize the peptidyl-polyprolyl-tRNA. This may reflect the high drop-off frequency of peptidyl-tRNA bearing a short nascent peptide even in the presence of EF-P and they were evolutionally removed due to the instability of expression levels.

3.3. Conclusion

Based on the preliminary result of the peptidyl-tRNA drop-off and generation of a C-terminal peptide *in vitro*, I demonstrated that a putative protein lacking its N-terminal amino acids was expressed plausibly as a consequence of peptidyl-tRNA drop-off induced by ribosome stalling due to consecutive prolines in the nascent peptide (**Figure 34**). LC-ESI MS quantification revealed that full-length YhhM was dominantly synthesized up to 90% ratio among all the YhhM species even in the absence of EF-P, plausibly because ribosome stalling was induced weakly by proline-proline-leucine residues of YhhM in nascent peptides, which previously categorized[2,52]. However, the "weak" ribosome stalling was strong enough to force the peptidyl-tRNA drop-off from the ribosome. Among the YhhM species lacking their N-terminal amino acids, YhhM truncated between 2nd proline and posterior leucine was dominantly observed in the absence of EF-P, which was consistent with ribosome stalling induced by slow peptidyl transfer between P-site nascent peptide-diprolyl-tRNA and A-site aa-tRNA.

Co-expression of EF-P improved the yield of full-length YhhM and reduced the ratio of YhhM lacking its N-terminal amino acids plausibly by alleviation of ribosome stalling. However, even in the presence of EF-P, YhhM lacking its N-terminal amino acids was detected. This observation suggested that in the very early stage of peptide chain elongation, peptidyl-diprolyl-tRNAs dropped off from ribosome more frequently because nascent peptides was too short to interact with the ribosome tunnel to stabilize themselves. It is known that a short peptidyl-tRNA drops off more frequently than the longer ones, which might induce the translation restart from the mRNA and generation of proteins lacking its N- terminal amino acids. This might be an overlooked abnormal translation event to be further investigated. It remained elusive that how lack of N-terminal amino acids affects the protein structure and its function, and any quality control mechanism monitor the synthesized proteins.

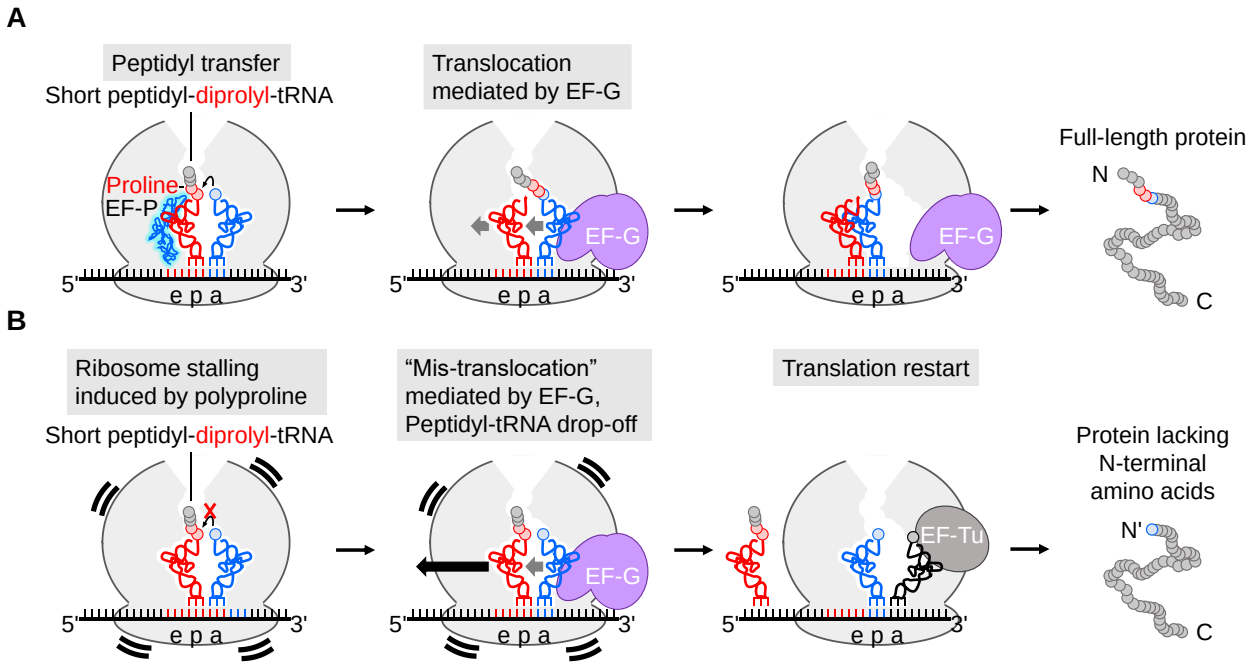


Figure 34 Plausible pathways of (A) synthesis of full-length protein containing polyproline stretch and (B) generation of the protein lacking N-terminal amino acids.

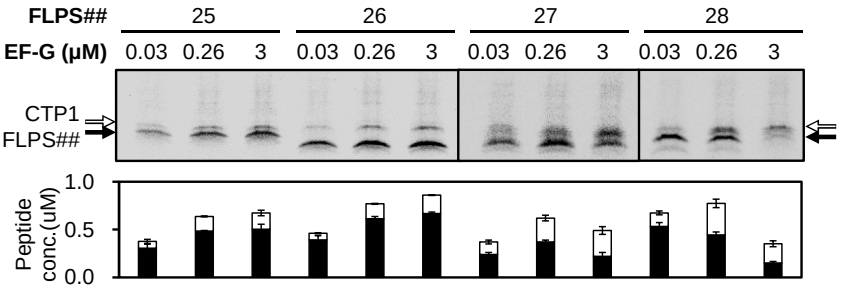
In the expression of YhhM in the presence of EF-P, pathway A was promoted.

3.4. Supplementary results

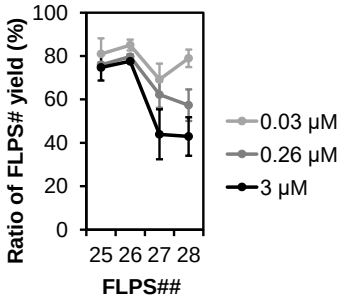
A

Full-length peptide (FLPS25) : fMRFCLILITALLLAGCSHHKAPPPKKK-FLAG
Full-length peptide (FLPS26) : fMITALLLAGCSHHKAPPPKKK-FLAG
Full-length peptide (FLPS27) : fMGCSHHKAPPPKKK-FLAG
Full-length peptide (FLPS28) : fMHKAPPPKKK-FLAG
C-Truncated peptide (CTP1) : PKKK-FLAG

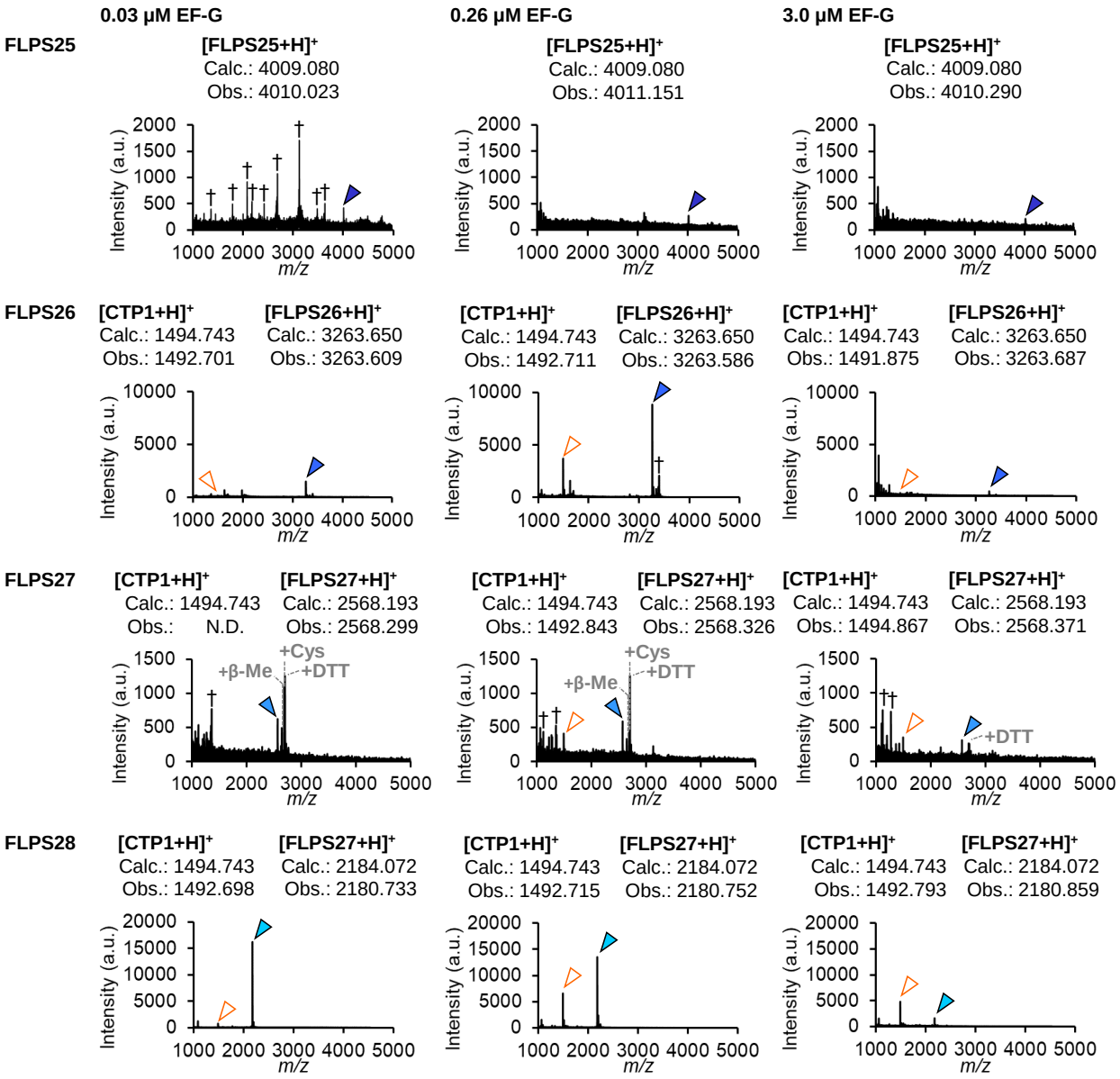
B



C



D



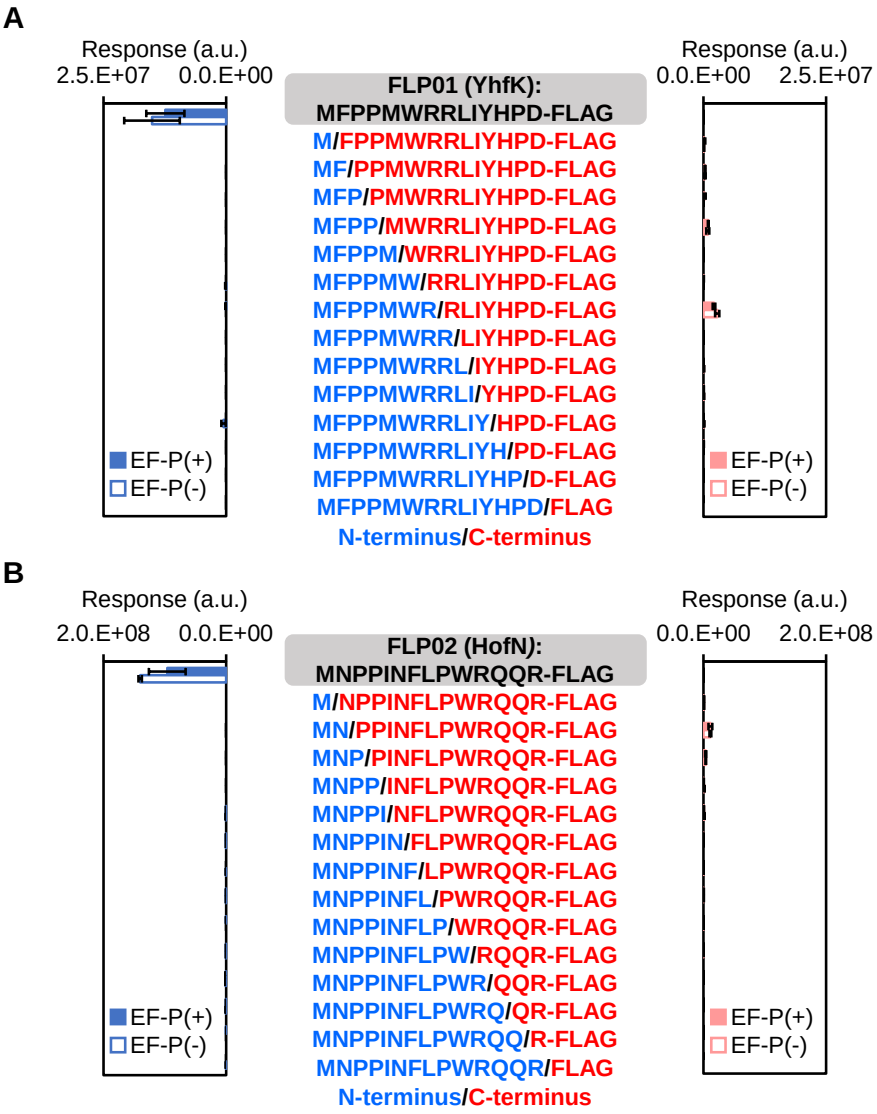
Supplementary Figure 18 Peptidyl-tRNA destabilization, induction of peptidyl-tRNA drop-off, and generation of a C-terminal peptide by shortening nascent peptide chains.

(A) Peptide sequences used. A series of truncated peptides derived from the N-terminal region of NlpC protein including PPP motif were fused with three Lys and FLAG peptide. FLPS3: Full-length peptide derived from NlpC. FLPS4-6: truncated FLPS3. FLAG is an octa peptide, Asp-Tyr-Lys-Asp-Asp-Asp-Asp-Lys.

(B) Expression and EF-G titration against a series of N-terminally truncated N-termini of NlpC in the absence of EF-P. FLPS3-6 peptides were radioisotope-labelled by the [^{14}C]-Asp in the FLAG sequences and separated by the tricine-SDS PAGE. FLPS peptides (Blue arrow) encoded in the mRNA and C-terminal peptide CTP2 (red arrow) were quantified as shown right. Tris(2-carboxyethyl)phosphine) (TCEP) were added into translation reaction mix to reduce the sulfhydlyl group of Cys. Standard deviations of three independent experiments are shown.

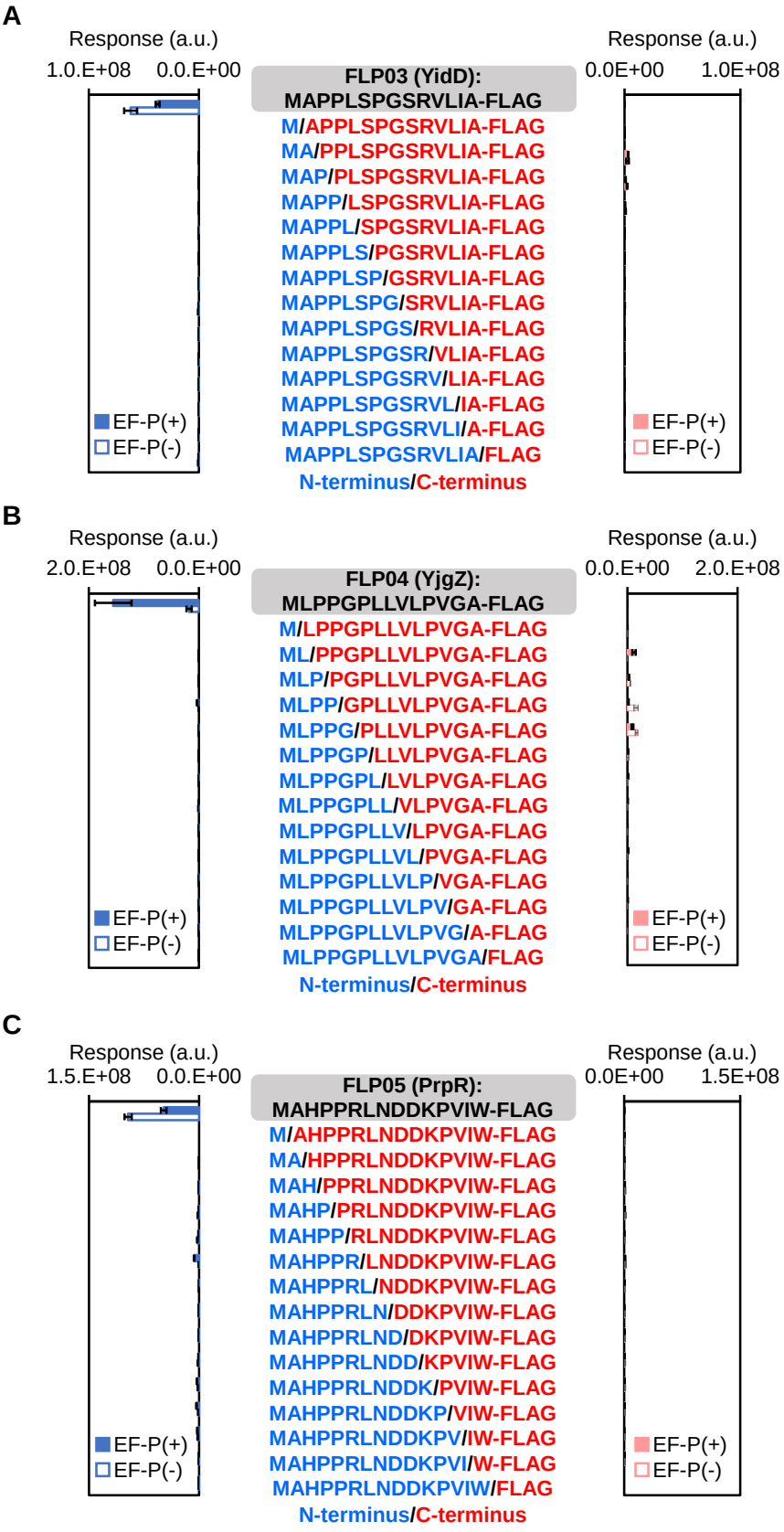
(C) Peptidyl-tRNA stabilization depending on nascent peptide chain length. Ratio of FLPS# yield was calculated as (yield of FLPS#)/{(yield of FLPS#) + (yields of CTP2)} from (B) and mapped in terms of length of nascent peptide chains and EF-G concentration. Standard deviations of three independent experiments are shown.

(D) Identification of translation products by MALDI-TOF MS. Peptides were expressed with 0.03, 0.26, or 3.0 μM EF-G in the absence of EF-P for 20 min, purified by anti-FLAG M2 agarose gel, desalted by the C18 column, and applied for the MALDI-TOF MS. Blue closed triangles: FLPS3-6, orange open triangle: CTP2, *: Na^+ or K^+ adduct peaks, †: Unidentified peaks, +β-Me: β-Mercaptoethanol adduct on a cysteine of peptides, +Cys: Cysteine adduct on a cysteine of peptides, +DTT: Dithiothreitol adducts on a cysteine of peptides.

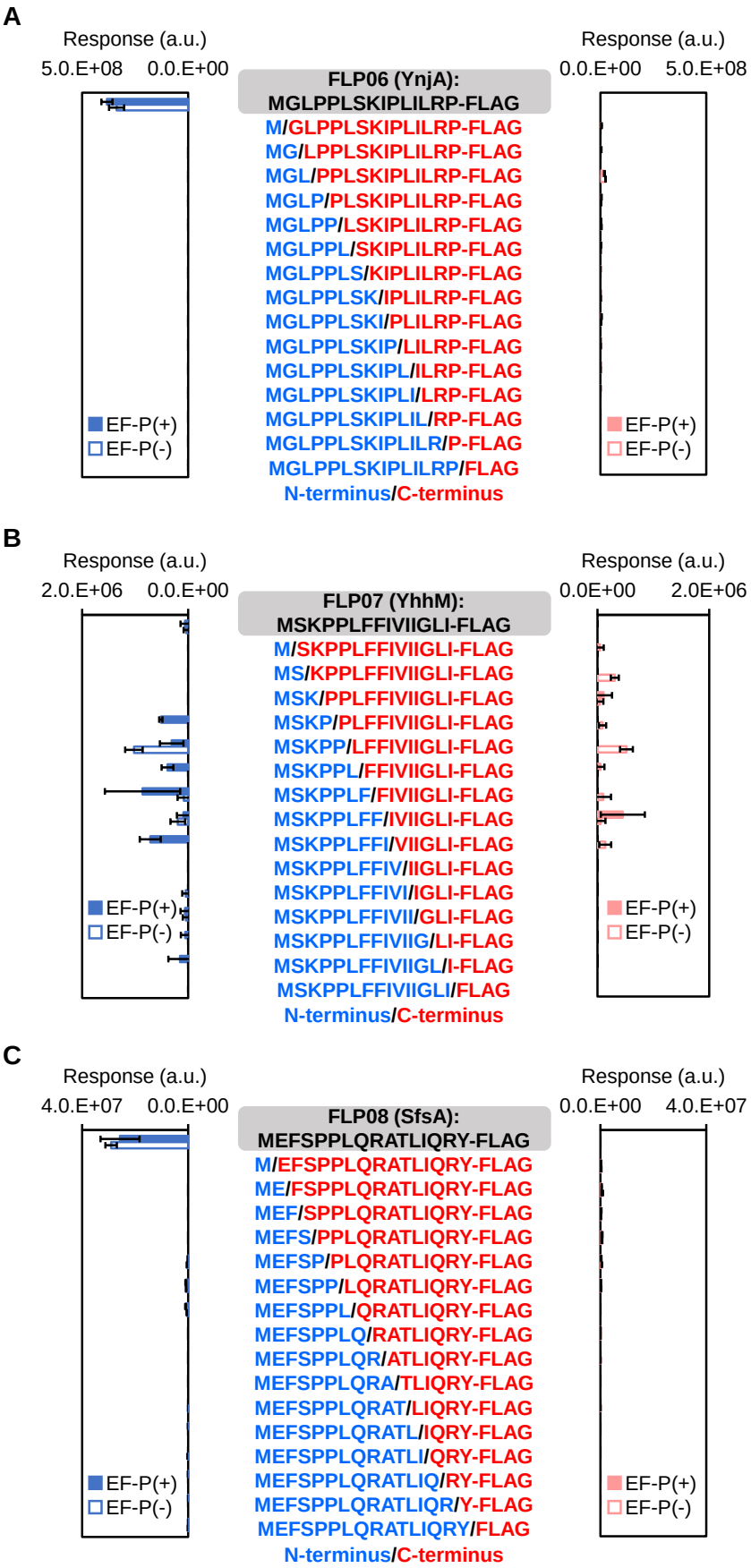


Supplementary Figure 19 LC-ESI MS quantification of (A) YhfK and (B) HofN N-terminal peptides expressed *in vitro*.

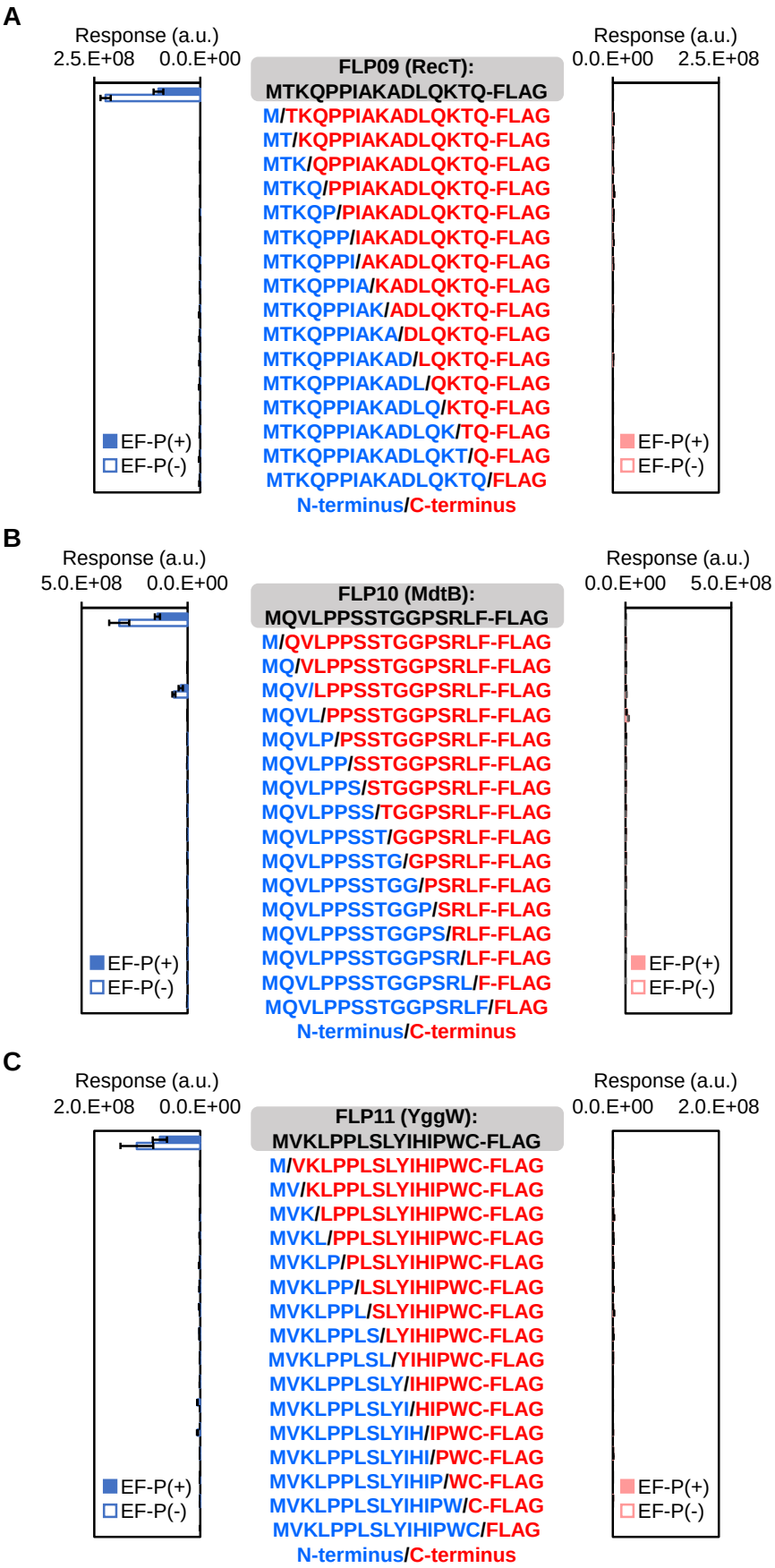
N-terminal sequences of (A) YhfK and (B) HofN were expressed *in vitro* with or without EF-P, treated with PTH and all of their possible N-terminal and C-terminal peptides were quantified. Black: Full-length peptides, Blue: N-terminal peptides, Red: C-terminal peptides. Close bars: peptide quantity with EF-P, Open bars: peptides quantity without EF-P. Standard deviation from 3 independent experiments are shown. Following supplementary figures showed quantification results analyzed in the same manner.



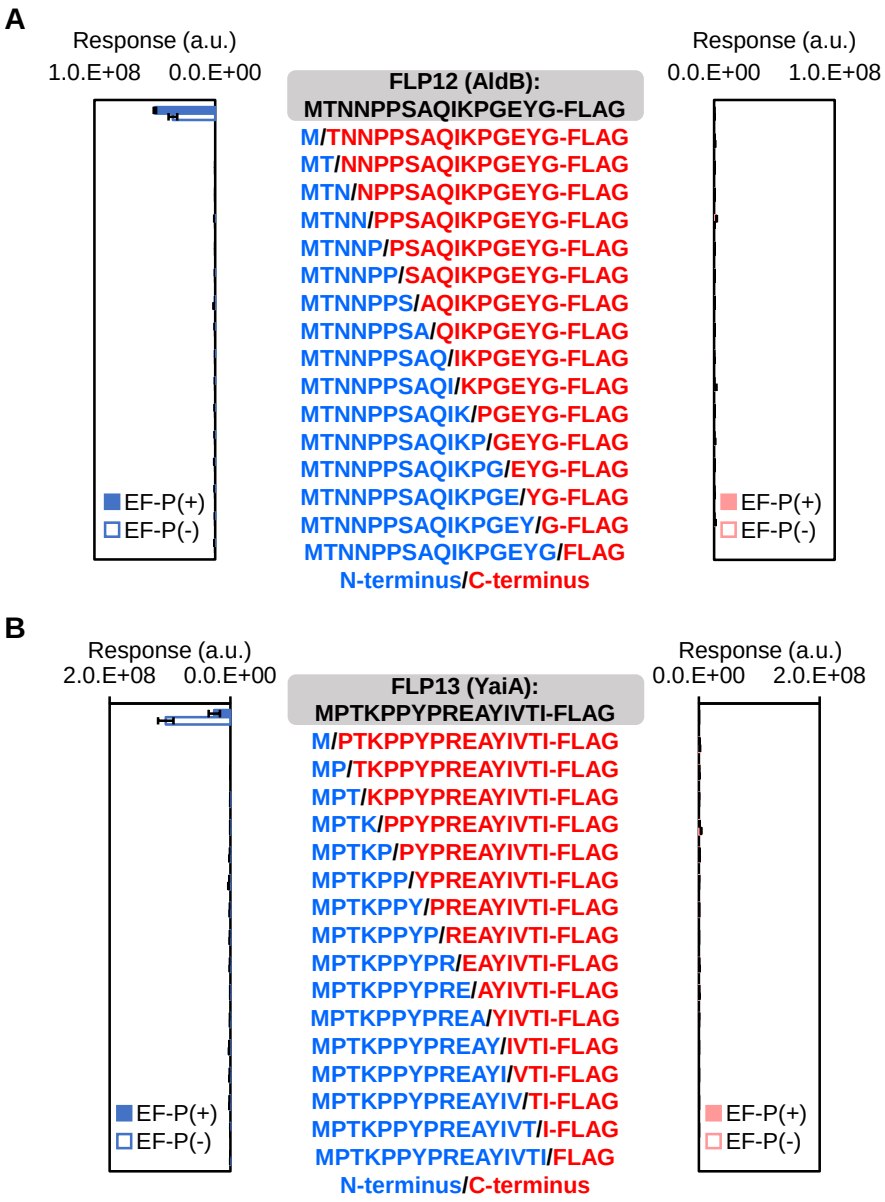
Supplementary Figure 20 LC-ESI MS quantification of (A) YidD, (B) YjgZ, and (C) PrpR N-terminal peptides expressed *in vitro*.



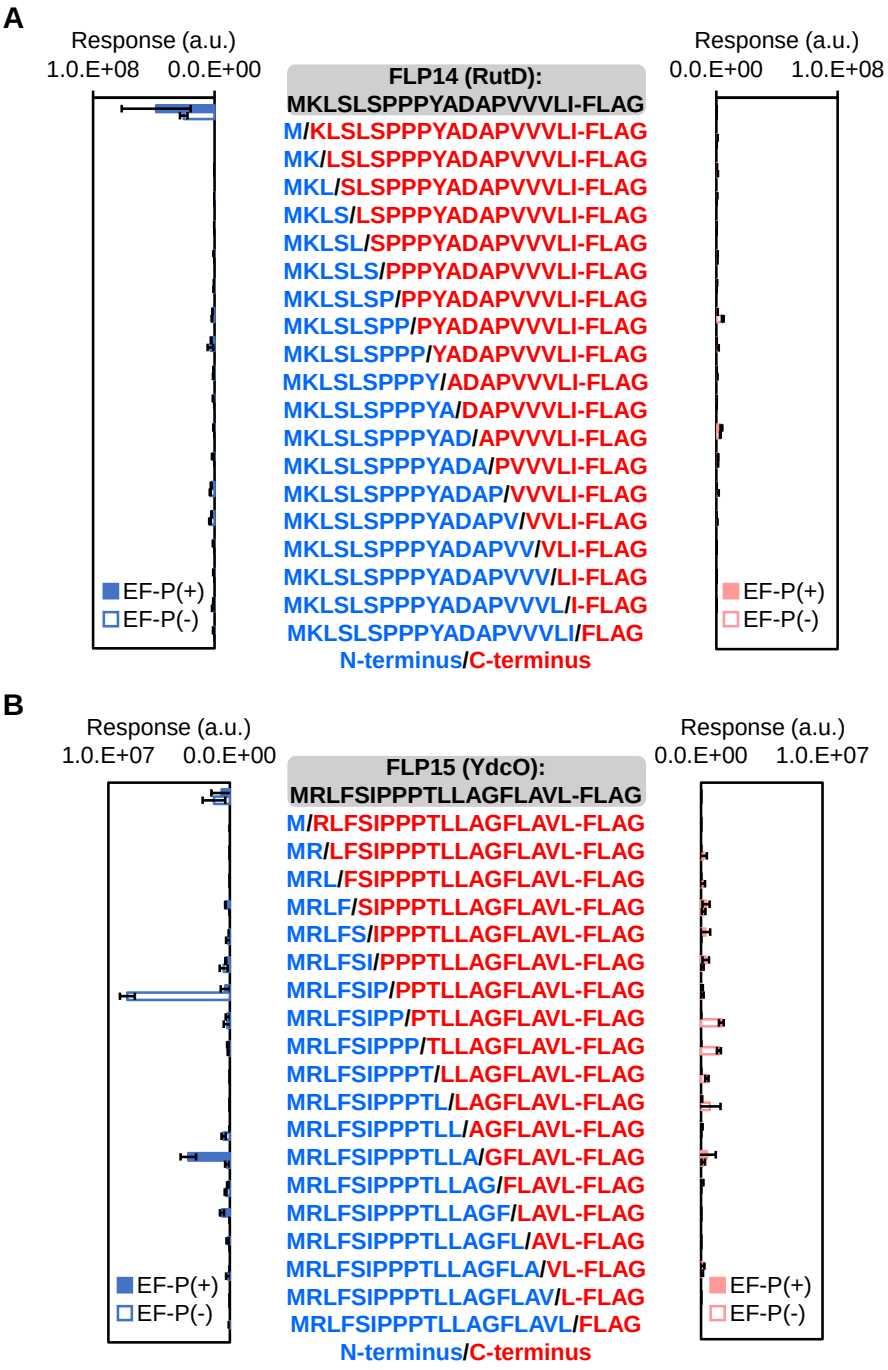
Supplementary Figure 21 LC-ESI MS quantification of (A) YnjA, (B) YhhM, and (C) SfsA N-terminal peptides expressed *in vitro*.



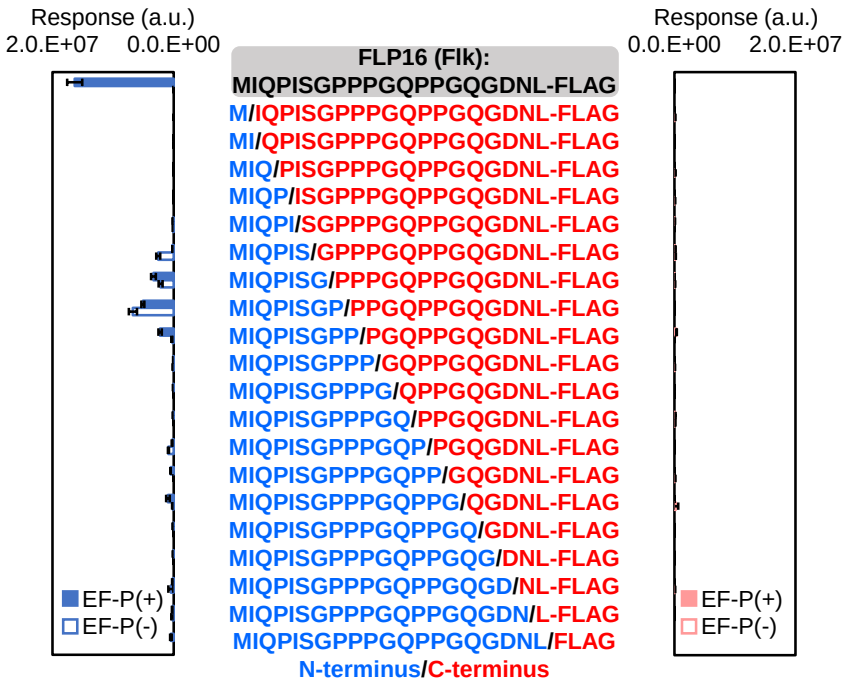
Supplementary Figure 22 LC-ESI MS quantification of (A) RecT, (B) MdtB, and (C) YggW N-terminal peptides expressed *in vitro*.



Supplementary Figure 23 LS-ESI MS quantification of (A) AldB and (B) YaiA N-terminal peptides expressed *in vitro*.



Supplementary Figure 24 LC-ESI MS quantification of (A) RutD and (B) YdcO N-terminal peptides expressed *in vitro*.



Supplementary Figure 25 LC-ESI MS quantification of Flk N-terminal peptide expressed *in vitro*.

3.5. Method

Preparation of cDNA templates and mRNA templates

cDNA templates and mRNA templates were prepared by PCR and following *in vitro* transcription by T7 RNA polymerase. Template DNAs that have a T7 promoter prior to transcripts were prepared by extension of forward and reverse primer pairs, followed by PCR using forward and reverse PCR primers. DNA primer sequences and combination of primers used in extension and PCR reactions are summarized in the supplementary table. PCR was conducted in 1x PCR reaction mixture containing 50 mM KCl, 10 mM Tris-HCl (pH 9.0), 0.1v/v% Triton X-100, 2.5 mM $MgCl_2$, 250 μM each dNTPs, 0.5 μM forward and reverse primers, and 1.5v/v% 1x Taq DNA polymerase with pH 9.0. The PCR products were extracted by phenol/chloroform, precipitated by ethanol, and used for *in vitro* transcription at 37◦C for at least 3 hours in a 100 uL reaction mixture. Reaction mixture contains 40 mM Tris-HCl, 1 mM spermidine, 0.01v/v% Triton X-100, 10 mM DTT, 30 mM $MgCl_2$, 5 mM NTPs, 30 mM KOH, 10v/v% template DNA solution, and 0.12 μM T7 RNA polymerase with pH 8.0. The RNA transcripts were precipitated by adding 10v/v% 3 M NaCl and twice volume of isopropanol followed by centrifuge. After wash with 70% EtOH, the RNA pellet was dissolved into H_2O with 10v/v% of *in vitro* transcription reaction and equivalent volume of 2x RNA loading buffer (8M urea, 2 mM $Na_2EDTA \cdot 2H_2O$, and 2 mM Tris-HCl, pH 7.5) was added. After heating at 95◦C for 2 min, RNA was purified by 8% polyacrylamide gel containing 6 M urea. The RNA band was visualized by UV light, cut out, and extracted in the 0.3 M NaCl solution for at least 5 hours. After removal of gel, twice volume of EtOH was added and RNA was precipitated by centrifuge at 13000 rpm for 15 min. After wash with 70% EtOH and drying, 20 μM RNA solution was prepared by dissolving into H_2O.

Expression of peptides in reconstituted cell-free translation system

Expression of peptides were carried out utilizing modified FIT (Flexible *in vitro* translation) system including only the necessary components and non-necessary amino acids, aaRSs, and EF-P were withdrawn from the following recipe. The complete FIT system contains 2 μM template mRNA, 0.5 mM each 20 proteinogenic L-amino acids (Ala, Cys, Asp, Glu, Phe, Gly, His, Ile, Lys, Leu, Met, Asn, Pro, Gln, Arg, Ser, Thr, Val, Trp, and Tyr), 2 mM ATP, 2 mM GTP, 1 mM CTP, 1 mM UTP, 20 mM Creatine phosphate, 50 mM HEPES-KOH pH 7.6, 100 mM Potassium acetate, 12 mM Magnesium acetate, 2 mM Spermidine, 1 mM DTT, 0.1 mM 10-HCO-H4folate, 1.5 mg/mL *E. coli* tRNAs, 0.3 mM Magnesium acetate, 1.2 μM *E. coli* ribosome, 0.6 μM MTF, 2.7 μM IF1, 0.4 μM IF2, 1.5 μM IF3, 0.26 μM EF-G, 10 μM EF-Tu, 0.66 μM EF-Ts, 5 μM EF-P0.25 μM RF2, 0.17 μM RF3, 0.5 μM RRF, 4 μg/mL Creatine kinase, 3 μg/mL Myokinase, 0.1 μM Inorganic pyrophosphatase, 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 μM IleRS, 0.04 μM LeuRS, 0.11 μM LysRS, 0.03 μM MetRS, 0.68 μM PheRS, 0.16 μM ProRS, 0.04 μM SerRS, 0.09 μM ThrRS, 0.03 μM TrpRS, 0.02 μM TryRS, and 0.02 μM ValRS. The concentration of EF-G was changed if necessary. In distinguish whether drop-off is caused by mRNA or nascent peptides, peptides were expressed from corresponding mRNA utilizing corresponding pre-charged aa-tRNAs as follows instead of free amino acids and the cognate aaRSs.

In the quantification of peptide yields, 0.5 mM [^{12}C]-Asp, [^{12}C]-Lys, or [^{32}S]-Met in the FIT system were replaced with radio-labelled 0.05 mM [^{14}C]-Asp, 0.05 mM [^{14}C]-Lys, or 0.8683 μM [^{35}S]-Met. After incubation, reaction mixture was mixed with equal amount of 2x tricine-SDS-PAGE loading solution (0.9 M Tris-HCl, 8w/v% SDS, 30v/v% glycerol, and 0.001w/v% xylene cyanol, pH 8.45) and incubated at 95◦C for 5 min. The solution was applied for the tricine-SDS gel composed of 4% stacking gel and 15% separation gel and run for 40 min at 150 V with cathode buffer (0.1 mM Tris-HCl , 0.1 mM tricine, and 0.1w/v% SDS, pH 8.45) and anode buffer (0.2 mM Tris-HCl, pH 8.9). After drying, the gel was analyzed by Typhoon FLA 7000 (GE Health Care).

In the identification and quantification by LC-ESI MS, proteins were precipitated by incubation of the reaction solution with equal volume of methanol on ice for 5 min, centrifuged at 13000 rpm for 3 min, and centrifuged with 4 equivalent volume of 1% trifluoroacetic acid at 13000 rom for 3 min. Resulting solution was injected to the LC-ESI MS (Waters XevoTM G2-XS) assembled with Aquity UPLC BEH C18 column (1.7 μm, 300Å, 2.1x150 mm waters) and eluted by gradient of H_2O/MeCN containing 0.1% formic acid.

Preparation of plasmid coding polyproline containing proteins or *efp* under a non-T7 promotors

Plasmids were designed to code proteins containing polyprolines and EF-P under regulation of non-T7 promoters. 5 proteins containing a stretch of prolines proximal to the N-terminus (YjgZ, PrpR, YhhM, RutD, and YdcO) were cloned into pBAD-HisA (Thermo Fisher Invitrogen), pUC18 (Takara Bio), and pSF-TAC (Sigma-Aldrich). Plasmids were prepared as following procedure. cDNA corresponding to the ORF of proteins (Purchased from Integrated DNA Technologies) were amplified by adopter PCR and above plasmids were amplified by inversion PCR. Both of PCR products were designed to have 15 bp of homologous region. DNA primer sequences and plasmid DNA sequences are summarized in supplementary table. PCR reaction contained 50v/v% KODone premix (TOYOBO), 0.3 μM of forward and reverse primers, and 1.00 ng/μL of temple linear or circular DNA. PCR reaction

was performed as 94◦C 120 sec, repeat of 94◦C 10 sec, 56◦C 30 sec, and 68◦C for 60 sec/1 kb. PCR products were separated by 1w/v% agarose gel electrophoresis, extracted, and purified by NucleoSpin Gel and PCR Clean-up (Macherey-Nagel). Resulting cDNA and linear plasmids were assembled by In-Fusion HD cloning kit (Takara bio). Transformation was carried on 50 μL of *E. coli* DH5α competent cells with 2.5 μL of In-Fusion reaction mixture, kept on ice for 10 min, heat-shocked for 45 sec at 42◦C, kept on ice for 2 min, and spread on LB plates with 100 μg/mL ampicillin for pBAD-HisA derivatives and pUC18 derivatives or 50 μg/mL kanamycin for pSF-TAC derivatives at 37◦C for overnight. Single colony were cultivated in 4 mL LB broth with 100 μg/mL ampicillin for pBAD-HisA derivatives and pUC18 derivatives or with 50 μg/mL kanamycin for pSF-TAC derivatives at 37◦C for overnight. From the resulting DH5α, plasmid was extracted by fast gene plasmid mini kit (Nippon Gene) and sequenced (Fasmaq).

For the complementation of EF-P, pBAD-*efp*-CmR-p15A ori plasmid was prepared as follows. First, pBAD-*efp*-ColE1 ori plasmid was prepared by In-Fusion assembly (Takara bio) of an inversion PCR product of pBAD-HisA (Thermo Fisher Invitrogen) and PCR product of *efp* gene from pET28a(+)-*efp* plasmid. Transformation was carried on 50 μL of *E. coli* DH5α competent cells with 2.5 μL of In-Fusion reaction mixture, following the same procedure as preparation of plasmid coding polyproline containing proteins using ampicillin. Second, pBAD-*efp*-CmR-p15A ori plasmid was prepared by In-Fusion reaction of a DNA fragment containing *araC*-P_{BAD} promotor-*efp* amplified from the pBAD-*efp*-ColE1 ori plasmid and a DNA fragment containing p15A ori-*cat* amplified from pLysS plasmid. Transformation was carried on 50 μL of *E. coli* DH5α competent cells with 2.5 μL of In-Fusion reaction mixture, following the same procedure as preparation of plasmids coding polyproline-containing proteins using 30 μg/mL chloramphenicol in LB. The recombinant plasmid was prepared and sequenced following the procedure shown above.

Preparation of competent *E. coli* Δ*efp* cells and its complementation by pBAD-*efp*-p15A ori plasmid

E. coli JW4107 strain lacking its chromosomal *efp* gene (*E. coli* Δ*efp*, BW25113 but *efp*::Km, National BioResource Project, *E. coli* and *B. subtilis*, National Institute of Genetics) was made chemical competent. *E. coli* Δ*efp* was cultivated in 8 mL of LB with 50 μg/mL kanamycin at 37◦C for overnight, mixed with equal amount of 50v/v% glycerol and stored at -80◦C. Chemical competent *E. coli* Δ*efp* was prepared as follows. Glycerol-stocked *E. coli* Δ*efp* was pre-cultivated in 2 mL LB with 50 μg/mL kanamycin at 37◦C for overnight with 250 rpm. 1 mL of pre-culture was transferred into 100 mL of SOB broth (2w/v% Bacto Trypton, 0.5w/v% Yeast Extract, 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, and 10 mM MgSO4 (anhydrous))) with 50 μg/mL kanamycin and cultivated at 18◦C with 150 rpm until the OD_{580} reach 0.5. The broth was cooled on ice for 15 min and rotated at 4◦C for 15 min at 3000 rpm. After removal of supernatant, *E. coli* pellet was suspended with 33.5 mL of TB buffer (10 mM PIPES, 15 mM $CaCl_2$, 150 mM KCl, and 55 mM $MnCl_2$, pH 6.7). The *E. coli* suspension was cooled on ice for 10 min and rotated at 4◦C for 15 min at 3000 rpm. After removal of supernatant, the *E. coli* pellet was suspended in 8 mL of TB buffer, mixed with 0.6 mL DMSO, cooled on ice for 15 min, and 250 μL aliquots were kept at -80◦C.

In *E. coli* Δ*efp* transformed by pBAD-*efp*-CmR-p15A ori, EF-P was expressed, and its post-translational modification state was confirmed by LC-ESI MS. *E. coli* Δ*efp* competent cell was transformed with pBAD-*efp*-CmR-p15A ori plasmid as same as above. Complemented EF-P was expressed as follow. *E. coli* Δ*efp*/pBAD-*efp*-CmR-p15A ori was cultivated at 37◦C in 3 mL LB with 50 μg/mL kanamycin and 30 μg/mL chloramphenicol until OD_{580}

reaches 0.8. EF-P was induced addition of f.c. 0.02w/v% L-arabinose for 6 hours at 37◦C. The *E. coli* pellet was collected by centrifuge at 37◦C for 2 min with 13000 rpm, lysed into 50 μL of 1x TBST (50 mM Tris-HCl, 150 mM NaCl, and 0.05v/v% Tween20, pH 7.6), digested by 0.1 μg/μL Glu-C protease (Sigma-Aldrich) in 50 mM Tris-HCl pH 8 for 2 hours at 37◦C. After MeOH precipitation and 1v/v% TFA precipitation, digestion solution was applied to the LC-ESI MS (Waters Xevo^{TM} G2-XS) assembled with Aquity UPLC BEH C18 column (1.7 μm, 300Å, 2.1x150 mm waters)) and eluted by gradient of H_2O/MeCN containing 0.1v/v% formic acid.

Resulting *E. coli* Δ*efp* /pBAD-*efp*-p15A ori was made into chemical competent cell following above procedure using 50 μg/mL kanamycin and 30 μg/mL chloramphenicol antibiotics in a LB plate and SOB broth.

Expression and purification of polyproline-containing proteins

E. coli Δ*efp* competent cell and *E. coli* Δ*efp*/pBAD-*efp*-p15A ori competent cell were transformed by pBAD-HisA and pUC18 plasmid coding YjgZ, PrpR, YhhM, RutD, or YdcO proteins, and the proteins were expressed and purified. After transformation of *E. coli* Δ*efp* and *E. coli* Δ*efp*/BAD-*efp*-p15A ori stain, obtained single colony was pre-cultivated in 20 mL of LB with 100 μg/mL ampicillin and 50 μg/mL of kanamycin for the *E. coli* Δ*efp* and with additional 30 μg/mL chloramphenicol for *E. coli* Δ*efp*/pBAD-*efp*-p15A ori at 37◦C for overnight. After proliferation, preculture broth was added into 2 L of LB broth with the same antibiotics as the preculture and cultivated at 37◦C until OD_{580} reaches 0.8 and induced by f.c. 0.02v/v% L-arabinose for pBAD-HisA or 0.5 mM IPTG for pUC18 at 37◦C for 4 hours. *E. coli* pellet was precipitated by centrifuge at 8000 rpm for 5 min, flash-frozen, and kept at -80◦C. After suspending with lysis buffer (20 mM Tis-HCl, 200 mM NaCl, 1 mM DTT, 0.1 mM PMSF), *E. coli* was sonicated, centrifuged at 15000g for 10 min, and supernatant was recovered. Supernatant were applied into pre-equilibrated 1 mL slurry of His Super Flow (GE health care) and eluted by imidazole gradient. The fractions containing target proteins were visualized by Glycine-SDS PAGE, using Precision Plus Protein Dual Color marker (BioRad). For proteins containing PreScission protease tag, 0.1v/v% of Turbo3C protease (2 U/μL, Acceleragen) were added to the eluent, digested, and dialyzed in 1 L of Equilibration buffer (20 mM Tis-HCl, 200 mM NaCl, and 1 mM DTT, pH 7.4) at 4˚C for overnight. Dialyzed solution applied to 1 mL slurry of His Super Flow (GE health care) and flow-through fractions were recovered. For proteins containing FLAG tag, the proteins were purified by affinity against anti-FLAG M2 agarose gel. The eluent from His Super Flow was concentrated less than 2.5 mL by amiconultra 10k MW cut-off (Merck Millipore) and exchanged the buffer by PD10 (GE health care) in order to remove DTT for the FLAG purification. 500 μL slurry of anti-FLAG M2 affinity agarose gel was drained in a column, washed with 2.5 mL of 1x TBS twice, 0.5 mL 0.1 M Gly-HCl pH 3.5 three times, and 2.5 mL 1x TBS. Proteins purified by His Super Flow was applied in the column, and flow-through was re-applied column for 5 times. The gel was washed with 10 mL of 1x TBS pH 7.6 and proteins were eluted with 0.5 mL of 0.1 M Gly-HCl pH3.5 nine times. The eluents were immediately neutralized with 10 μL of 1M Tris-HCl pH 8.0. After identification of fractions containing target protein by Glycine-SDS PAGE, the fractions were concentrated by amiconultra 10k MW cut-off (Merck Millipore).

Sequence identification by LC-ESI MS/MS of polyproline-containing proteins digested by proteases

Purified proteins were subjected to the LC-ESI MS or LC-ESI MS/MS after protease digestion. At first,

optimum proteases were screened to detect protein fragment by LC-ESI MS. YhhM-FLAG-His_6 expressed in the absence or presence of EF-P were digested by Glu-C (Sigma-Aldrich), Lys-N (Thermo Fisher Invitrogen), Lys-C (Sigma-Aldrich), Asp-N (Sigma-Aldrich)), and Arg-C (Promega). ~0.5 μg/μL of purified proteins were digested in the following 4 μL of reaction mixtures at 37 ◦C for 4 hours: 0.027 μg/μL Glu-C with 0.05 M NH_4HCO_3 pH 8.2, 0.027 μg/μL Lys-N with 0.05 M NH_4HCO_3 pH 8.2, 0.003 μg/μL Lys-C, with 0.025 M NH_4HCO_3 pH 8.2 and 0.001 M EDTA pH 8.0, 0.01 μg/μL Asp-N with 0.05 M NH_4HCO_3 pH 8.2, and 0.01 μg/μL Arg-C with 0.05 M Tris HCl pH 7.6 and 0.005 M DTT. RutD-FLAG-His_6 expressed in the absence or presence of EF-P were digested by Glu-C (Sigma-Aldrich) and Arg-C (Promega). ~0.5 μg/μL of purified proteins were digested in the following 4 μL of reaction mixtures at 37 ◦C for 4 hours: 0.027 μg/μL GluC with 0.05 M NH_4HCO_3 pH 8.2 and 0.01 μg/μL Arg-C with 0.05 M Tris HCl pH 7.6 and 0.005 M DTT. Reaction mixture were desalted by SPE C-tip (Nikkyo Technos) and eluted with 30 μL of 80v/v% acetonitrile with 0.5v/v% acetic acid solution. The eluents were analyzed by LC-ESI MS (Waters Xevo^{TM} G2-XS) assembled with Aquity UPLC BEH C18 column (1.7 μm, 300Å, 2.1x150 mm, waters) and eluted by gradient of H_2O/MeCN containing 0.1v/v% formic acid.

Second, using optimum proteases, the protein fragments were sequenced by LC-ESI MS/MS and LC-ESI MS^E. 0.44 μg/μL of YhhM-FLAG-His_6 expressed in the absence of EF-P and 0.85 μg/μL of YhhM-FLAG-His_6 expressed in the presence of EF-P was digested by 0.027 μg/μL Lys-N with 0.05 M NH_4HCO_3 pH 8.2 in 30 μL reaction mixture at 37 ◦C for 16 hours. ~0.067 μg/μL of RutD-FLAG-His_6 expressed in the absence or presence of EF-P were digested by 0.01 μg/μL Arg-C with 0.05 M Tris HCl pH 7.6 and 0.005 M DTT. Reaction mixture were desalted by SPE C-tip (Nikkyo Technos) and eluted with 30 μL of 80v/v% acetonitrile with 0.5v/v% acetic acid solution. The eluents were analyzed by LC-ESI MS/MS (Waters Xevo^{TM} G2-XS) assembled with Aquity UPLC BEH C18 column (1.7 μm, 300Å, 2.1x150 mm, waters) and eluted by gradient of H_2O/MeCN containing 0.1v/v% formic acid. First, the parent ions were screened by MS^E mode. *m/z* values corresponding to parent peptide fragments were screened by UniFi software (waters). In order to better fragmentation and fragment identification, parent ions were fragmented by MS/MS mode. In MS/MS analysis, parent ions having a specific *m/z* value was selected by quadrupole and fragmented by ramping of collision energy from 10 V to 40 V. The daughter ions were identified by Masslynx software (waters) after deconvolution of raw MS/MS spectra.

3.6. Supplementary tables

Supplementary table 7 DNA primer sequences used.

Primer Name	DNA sequence
Primer53.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACTTGTTTCCCCCGATGTGGC
Primer54.R66	CTTGTCGTCGTCGTCCTTGTAGTCATCGGGGTGATAAATCAGTCTGCGCCACATCGGGGGAAACAA
Primer55.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGAACCCGCCAATTAATT
Primer56.R66	CTTGTCGTCGTCGTCCTTGTAGTCGCGTTGCTGTCGCCAGGGCAAAAAATTAATTGGCGGGTTCAT
Primer57.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGGCGCCGCCACTGTCGC
Primer58.R66	CTTGTCGTCGTCGTCCTTGTAGTCGGCTATCAGGACCCGCGAGCCAGGCGACAGTGGCGGCGCCAT
Primer59.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACTTGTTACCTCCTGGACCGC
Primer60.R66	CTTGTCGTCGTCGTCCTTGTAGTCCGCTCCGACTGGAAGGACAAGAAGCGGTCCAGGAGGTAACAA
Primer61.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGGCACATCCACCACGGC
Primer62.R69	CTTGTCGTCGTCGTCCTTGTAGTCCCAGATAACCGGTTTGTCGTCATTAAGCCGTGGTGGATGTGCCAT
Primer63.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGGGTTTACCGCCGCTTA
Primer64.R69	CTTGTCGTCGTCGTCCTTGTAGTCTGGACGTAAAATTAAAGGAATTTTGCTAAGCGGCGGTAAACCCAT
Primer65.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGAGCAAACCACCTCTTT
Primer66.R69	CTTGTCGTCGTCGTCCTTGTAGTCAATTAAGCCAATGATAACAATAAAGAAAAGAGGTGGTTTGCTCAT
Primer67.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGGAATTTTCTCCCCCTC
Primer68.R70	TGTCGTCGTCGTCCTTGTAGTCGTAACGCTGAATTAGCGTCGCGCGCTGTAGAGGGGGAGAAAATTCCAT
Primer69.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGACTAAGCAACCACCAA
Primer70.R70	TGTCGTCGTCGTCCTTGTAGTCCTGAGTTTTTTGCAGATCGGCTTTTGCGATTGGTGGTTGCTTAGTCAT
Primer71.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGCAGGTGTTACCCCCGA
Primer72.R70	TGTCGTCGTCGTCCTTGTAGTCAAACAGGCGCGACGGGCCGCCTGTGCTGCTCGGGGGTAACACCTGCAT
Primer73.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGGTTAAATTACCTCCGC
Primer74.R70	TGTCGTCGTCGTCCTTGTAGTCGCACCACGGGATGTGAATGTAGAGACTCAGCGGAGGTAATTTAACCAT
Primer75.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGACCAATAATCCCCCTT
Primer76.R70	TGTCGTCGTCGTCCTTGTAGTCACCATACTCGCCGGGCTTAATCTGTGCTGAAGGGGGATTATTGGTCAT
Primer77.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGCCAACGAAACCGCCTT
Primer78.R70	TGTCGTCGTCGTCCTTGTAGTCAATCGTCACTATATATGCTTCACGAGGATAAGGCGGTTTCGTTGGCAT
Primer79.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGAAACTTTCACTCTCAC
Primer80.F70	CGTCCTTGTAGTCAATCAACACCACTACGGGCGCATCAGCATAAGGGGGAGGTGAGAGTGAAAGTTTCAT
Primer81.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGCGTCTGTTTTCTATTC
Primer82.F70	CGTCCTTGTAGTCTAATACCGCCAGAAACCCCGCCAGTAGCGTGGGTGGAGGAATAGAAAACAGACGCAT
Primer83.F62	TAATACGACTCACTATAGGGTTAACTTTAACAAGGAGAAAAACATGATACAACCTATTTCCG
Primer84.F70	CCTTGTAGTCCAGATTATCTCCCTGACCTGGTGGTTGCCCAGGAGGAGGGCCGGAAATAGGTTGTATCAT
Primer85.F70	CTTTAACAAGGAGAAAAACATGCGTTTCTGCCTTATTTTGATCACAGCACTGCTTCTGGCCGGGTGTAGC
Primer86.F52	CTTTAACAAGGAGAAAAACATGATCACAGCACTGCTTCTGGCCGGGTGTAGC
Primer87.F50	CTTTAACAAGGAGAAAAACATGGGGTGTAGCCACCATAAAGCACCGCCGC
Primer88.F38	CTTTAACAAGGAGAAAAACATGCATAAAGCACCGCCGC
Primer89.R70	GTCGTCGTCGTCCTTGTAGTCCTTCTTCTTTGGCGGCGGTGCTTTATGGTGGCTACACCCGGCCAGAAGC
Primer90.R48	GTCGTCGTCGTCCTTGTAGTCCTTCTTCTTTGGCGGCGGTGCTTTATG
Primer91.R42	CGAAGCTTACTTGTCGTCGTCGTCCTTGTAGTCCTTCTTCTT
Primer92.F40	CAGGAGGAATTAACCTTGTTACCTCCTGGACCGCTTCTTG
Primer93.R61	CTGATTTAATCTGTATCAGTGGTGGTGGTGGTGGTGCCCGAAGAAGGTAAAGCCCTCAAGA
Primer94.R44	GGTTAATTCCTCCTGTTAGCCCAAAAAACGGGTATGGAGAAACA
Primer95.F44	TACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAACAGAAT
Primer96.F40	CAGGAGGAATTAACCATGTTACCTCCTGGACCGCTTCTTG
Primer97.F40	CAGGAGGAATTAACCATGGCACATCCACCACGGCTTAATG
Primer98.R58	CTGATTTAATCTGTATCAGTGGTGGTGGTGGTGGTGGCTTTTCAGCCGCCGCCAGAAC
Primer99.F40	CAGGAGGAATTAACCATGAGCAAACCACCTCTTTTCTTTA
Primer100.R61	CTGATTTAATCTGTATCAGTGGTGGTGGTGGTGGTGTTGTTCGCCTACAAAGCTGACAAAG
Primer101.F40	CAGGAGGAATTAACCATGAAACTTTCACTCTCACCTCCCC
Primer102.R57	CTGATTTAATCTGTATCAGTGGTGGTGGTGGTGGTGCAGGGCGGCTTCACGGTGATG
Primer103.F40	CAGGAGGAATTAACCATGCGTCTGTTTTCTATTCCTCCAC
Primer104.R61	CTGATTTAATCTGTATCAGTGGTGGTGGTGGTGGTGATATCGGTTTCTGTCAGCGATTAAA
Primer105.F35	TGATACAGATTAAATCAGAACGCAGAAGCGGTCTG
Primer106.R34	CATGGTTAATTCCTCCTGTTAGCCCAAAAAACGG
Primer107.F44	GAGGAATTAACCATGGCAACGTACTATAGCAACGATTTTCGTGC
Primer108.R43	ATTTAATCTGTATCACTTCACGCGAGAGACGTATTCACCAGAG
Primer109.F42	CGGGAGCTGCATGTGTTTTTACGGTTCCTGGCCTTTTGCTGG
Primer110.R39	TAAGAATTAATTCATCATTGGAAAACGTTCTTCGGGGCG
Primer111.F31	ATGAATTAATTCTTACGCCCCGCCCTGCCAC
Primer112.R48	CACATGCAGCTCCCGACAACTTATATCGTATGGGGCTGACTTCAGGTG
Primer113.F40	AGAAGGAGATATACCTTGTTACCTCCTGGACCGCTTCTTG
Primer114.R58	AGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCCCGAAGAAGGTAAAGCCCTCAAGA
Primer115.F30	TGAGATCCGGCTGCTAACAAAGCCCGAAAG
Primer116.R40	GGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAG
Primer117.F40	AGAAGGAGATATACCATGTTACCTCCTGGACCGCTTCTTG
Primer118.F40	AGAAGGAGATATACCATGGCACATCCACCACGGCTTAATG
Primer119.R55	AGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGGCTTTTCAGCCGCCGCCAGAAC
Primer120.F40	AGAAGGAGATATACCATGAGCAAACCACCTCTTTTCTTTA
Primer121.R58	AGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGTTGTTCGCCTACAAAGCTGACAAAG
Primer122.F40	AGAAGGAGATATACCATGAAACTTTCACTCTCACCTCCCC
Primer123.R54	AGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGCAGGGCGGCTTCACGGTGATG
Primer124.F40	AGAAGGAGATATACCATGCGTCTGTTTTCTATTCCTCCAC
Primer125.R58	AGCAGCCGGATCTCAGTGGTGGTGGTGGTGGTGATATCGGTTTCTGTCAGCGATTAAA

Primer Name	DNA sequence
Primer126.R48	CCCCTGGAACAGAACTTCCAGCCCGAAGAAGGTAAAGCCCTCAAGAGG
Primer127.F67	GTTCTGTTCCAGGGGCCCAGCAGCCATCATCATCATCATCATTGATACAGATTAAATCAGAACGCAG
Primer128.R43	CCCCTGGAACAGAACTTCCAGGCTTTTCAGCCGCCGCCAGAAC
Primer129.R48	CCCCTGGAACAGAACTTCCAGTTGTTCGCCTACAAAGCTGACAAAGCG
Primer130.R42	CCCCTGGAACAGAACTTCCAGCAGGGCGGCTTCACGGTGATG
Primer131.R57	CCCCTGGAACAGAACTTCCAGATATCGGTTTCTGTCAGCGATTAAATTCAACACCAC
Primer132.R48	GTCGTCGTCGTCCTTGTAGTCCCCGAAGAAGGTAAAGCCCTCAAGAGG
Primer133.F63	AAGGACGACGACGACAAGAGCAGCCATCATCATCATCATCATTGATACAGATTAAATCAGAAC
Primer134.R43	GTCGTCGTCGTCCTTGTAGTCGCTTTTCAGCCGCCGCCAGAAC
Primer135.R48	GTCGTCGTCGTCCTTGTAGTCTTGTTCGCCTACAAAGCTGACAAAGCG
Primer136.R48	GTCGTCGTCGTCCTTGTAGTCCAGGGCGGCTTCACGGTGATGTAACAG
Primer137.R57	GTCGTCGTCGTCCTTGTAGTCATATCGGTTTCTGTCAGCGATTAAATTCAACACCAC
Primer138.F40	ACACAGGAAACAGCTTTGTTACCTCCTGGACCGCTTCTTG
Primer139.R61	TCGGGGCTGGCTTAATCAGTGGTGGTGGTGGTGGTGCCCGAAGAAGGTAAAGCCCTCAAGA
Primer140.R25	TTAAGCCAGCCCCGACACCCGCCAA
Primer141.F44	AGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACAC
Primer142.F51	TCACACAGATCTTTGAGGAAACAGCTTTGTTACCTCCTGGACCGCTTCTTG
Primer143.R61	ACAACTTCTAGAAACCTAGTGGTGGTGGTGGTGGTGCCCGAAGAAGGTAAAGCCCTCAAGA
Primer144.R44	CAAAGATCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAT
Primer145.F44	GTTTCTAGAAGTTGTCTCCTCCTGCACTGACTGACTGATACAAT
Primer146.F40	ACACAGGAAACAGCTATGTTACCTCCTGGACCGCTTCTTG
Primer147.F51	TCACACAGATCTTTGAGGAAACAGCTATGTTACCTCCTGGACCGCTTCTTG
Primer148.F40	ACACAGGAAACAGCTATGGCACATCCACCACGGCTTAATG
Primer149.R58	TCGGGGCTGGCTTAATCAGTGGTGGTGGTGGTGGTGGCTTTTCAGCCGCCGCCAGAAC
Primer150.F51	TCACACAGATCTTTGAGGAAACAGCTATGGCACATCCACCACGGCTTAATG
Primer151.R58	ACAACTTCTAGAAACCTAGTGGTGGTGGTGGTGGTGGCTTTTCAGCCGCCGCCAGAAC
Primer152.F40	ACACAGGAAACAGCTATGAGCAAACCACCTCTTTTCTTTA
Primer153.R61	TCGGGGCTGGCTTAATCAGTGGTGGTGGTGGTGGTGTTGTTCGCCTACAAAGCTGACAAAG
Primer154.F51	TCACACAGATCTTTGAGGAAACAGCTATGAGCAAACCACCTCTTTTCTTTA
Primer155.R61	ACAACTTCTAGAAACCTAGTGGTGGTGGTGGTGGTGTTGTTCGCCTACAAAGCTGACAAAG
Primer156.F40	ACACAGGAAACAGCTATGAAACTTTCACTCTCACCTCCCC
Primer157.R57	TCGGGGCTGGCTTAATCAGTGGTGGTGGTGGTGGTGCAGGGCGGCTTCACGGTGATG
Primer158.F51	TCACACAGATCTTTGAGGAAACAGCTATGAAACTTTCACTCTCACCTCCCC
Primer159.R57	ACAACTTCTAGAAACCTAGTGGTGGTGGTGGTGGTGCAGGGCGGCTTCACGGTGATG
Primer160.F40	ACACAGGAAACAGCTATGCGTCTGTTTTCTATTCCTCCAC
Primer161.R61	TCGGGGCTGGCTTAATCAGTGGTGGTGGTGGTGGTGATATCGGTTTCTGTCAGCGATTAAA
Primer162.F51	TCACACAGATCTTTGAGGAAACAGCTATGCGTCTGTTTTCTATTCCTCCAC
Primer163.R61	ACAACTTCTAGAAACCTAGTGGTGGTGGTGGTGGTGATATCGGTTTCTGTCAGCGATTAAA

Supplementary table 8 Combinations of DNA primers for the extension and PCR reaction for preparation of cDNAs and mRNAs.

	Extension		1st PCR		2nd PCR	
Name	**Forward primer**	**Reverse primer**	**Forward primer**	**Reverse primer**	**Forward primer**	**Reverse primer**
cDNA01	Primer53.F62	Primer54.R66	Primer41.F22	Primer13.R33		
cDNA02	Primer55.F62	Primer56.R66	Primer41.F22	Primer13.R33		
cDNA03	Primer57.F62	Primer58.R66	Primer41.F22	Primer13.R33		
cDNA04	Primer59.F62	Primer60.R66	Primer41.F22	Primer13.R33		
cDNA05	Primer61.F62	Primer62.R69	Primer41.F22	Primer13.R33		
cDNA06	Primer63.F62	Primer64.R69	Primer41.F22	Primer13.R33		
cDNA07	Primer65.F62	Primer66.R69	Primer41.F22	Primer13.R33		
cDNA08	Primer67.F62	Primer68.R70	Primer41.F22	Primer13.R33		
cDNA09	Primer69.F62	Primer70.R70	Primer41.F22	Primer13.R33		
cDNA10	Primer71.F62	Primer72.R70	Primer41.F22	Primer13.R33		
cDNA11	Primer73.F62	Primer74.R70	Primer41.F22	Primer13.R33		
cDNA12	Primer75.F62	Primer76.R70	Primer41.F22	Primer13.R33		
cDNA13	Primer77.F62	Primer78.R70	Primer41.F22	Primer13.R33		
cDNA14	Primer79.F62	Primer80.F70	Primer41.F22	Primer13.R33		
cDNA15	Primer81.F62	Primer82.F70	Primer41.F22	Primer13.R33		
cDNA16	Primer83.F62	Primer84.F70	Primer41.F22	Primer13.R33		
cDNAS25	Primer85.F70	Primer89.R70	Primer14.F46	Primer91.R42	Primer41.F22	Primer91.R42
cDNAS26	Primer86.F52	Primer89.R70	Primer14.F46	Primer91.R42	Primer41.F22	Primer91.R42
cDNAS27	Primer87.F50	Primer90.R48	Primer14.F46	Primer91.R42	Primer41.F22	Primer91.R42
cDNAS28	Primer88.F38	Primer90.R48	Primer14.F46	Primer91.R42	Primer41.F22	Primer91.R42

Supplementary table 9 Sequences of cDNA, mRNA, and peptides.

	mRNA sequence upstream of initiation AUG codon
	GUUAACUUUAAGAAGGAGAUAUAAAU
Name	**mRNA sequence downstream from initiation AUG codon**
mRNA01	UUGUUUCCCCCGAUGUGGCGCAGACUGAUUUAUCACCCCGAUGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA02	AUGAACCCGCCAAUUAAUUUUUUGCCCUGGCGACAGCAACGCGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA03	AUGGCGCCGCCACUGUCGCCUGGCUCGCGGGUCCUGAUAGCCGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA04	UUGUUACCUCCUGGACCGCUUCUUGUCCUUCCAGUCGGAGCGGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA05	AUGGCACAUCCACCACGGCUUAAUGACGACAAACCGGUUAUCUGGGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA06	AUGGGUUUACCGCCGCUUAGCAAAAUUCCUUUAAUUUUACGUCCAGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA07	AUGAGCAAACCACCUCUUUUCUUUAUUGUUAUCAUUGGCUUAAUUGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA08	AUGGAAUUUUCUCCCCCUCUACAGCGCGCGACGCUAAUUCAGCGUUACGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA09	AUGACUAAGCAACCACCAAUCGCAAAAGCCGAUCUGCAAAAAACUCAGGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA10	AUGCAGGUGUUACCCCCGAGCAGCACAGGCGGCCCGUCGCGCCUGUUUGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA11	AUGGUUAAAUUACCUCCGCUGAGUCUCUACAUUCACAUCCCGUGGUGCGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA12	AUGACCAAUAAUCCCCCUUCAGCACAGAUUAAGCCCGGCGAGUAUGGUGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA13	AUGCCAACGAAACCGCCUUAUCCUCGUGAAGCAUAUAUAGUGACGAUUGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA14	AUGAAACUUUCACUCUCACCUCCCCCUUAUGCUGAUGCGCCCGUAGUGGUGUUGAUUGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA15	AUGCGUCUGUUUUCUAUUCCUCCACCCACGCUACUGGCGGGGUUUCUGGCGGUAUUAGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNA16	AUGAUACAACCUAUUUCCGGCCCUCCUCCUGGGCAACCACCAGGUCAGGGAGAUAAUCUGGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNAS25	AUGCGUUUCUGCCUUAUUUUGAUCACAGCACUGCUUCUGGCCGGGUGUAGCCACCAUAAAGCACCGCCGCCAAAGAAGAAGGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNAS26	AUGAUCACAGCACUGCUUCUGGCCGGGUGUAGCCACCAUAAAGCACCGCCGCCAAAGAAGAAGGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNAS27	AUGGGGUGUAGCCACCAUAAAGCACCGCCGCCAAAGAAGAAGGACUACAAGGACGACGACGACAAGUAAGCUUCG
mRNAS28	AUGCAUAAAGCACCGCCGCCAAAGAAGAAGGACUACAAGGACGACGACGACAAGUAAGCUUCG

Name	**Peptide sequence**
mRNA01	MFPPMWRRLIYHPDDYKDDDDK
mRNA02	MNPPINFLPWRQQRDYKDDDDK
mRNA03	MAPPLSPGSRVLIADYKDDDDK
mRNA04	MLPPGPLLVLPVGADYKDDDDK
mRNA05	MAHPPRLNDDKPVIWDYKDDDDK
mRNA06	MGLPPLSKIPLILRPDYKDDDDK
mRNA07	MSKPPLFFIVIIGLIDYKDDDDK
mRNA08	MEFSPPLQRATLIQRYDYKDDDDK
mRNA09	MTKQPPIAKADLQKTQDYKDDDDK
mRNA10	MQVLPPSSTGGPSRLFDYKDDDDK
mRNA11	MVKLPPLSLYIHIPWCDYKDDDDK
mRNA12	MTNNPPSAQIKPGEYGDYKDDDDK
mRNA13	MPTKPPYPREAYIVTIDYKDDDDK
mRNA14	MKLSLSPPPYADAPVVVLIDYKDDDDK
mRNA15	MRLFSIPPPTLLAGFLAVLDYKDDDDK
mRNA16	MIQPISGPPPGQPPGQGDNLDYKDDDDK
mRNAS25	MRFCLILITALLLAGCSHHKAPPPKKKDYKDDDDK
mRNAS26	MITALLLAGCSHHKAPPPKKKDYKDDDDK
mRNAS27	MGCSHHKAPPPKKKDYKDDDDK
mRNAS28	MHKAPPPKKKDYKDDDDK

Supplementary table 10 DNA sequences of plasmids and protein sequences encoded in them

pBAD-HisA (Underlined: ORF)

AAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATT
AAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGA
TTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTAC
TGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTAACCATGGGGGGTTCTCATCATCATCATCATCATGGTATGGCTAGCATGACT
GGTGGACAGCAAATGGGTCGGGATCTGTACGACGATGACGATAAGGATCGATGGGGATCCGAGCTCGAGATCTGCAGCTGGTACCATATG
GGAATTCGAAGCTTGGCTGTTTTGGCGGATGAGAGAAGATTTTCAGCCTGATACAGATTAAATCAGAACGCAGAAGCGGTCTGATAAAAC
AGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTAGCGCCGATGGTAGTGTG
GGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGCCTTTCGTTTTATCTGTTG
TTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCCCGGAGGGTGGCGGGCA
GGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTCTACAAACTCTTTTGTTTA
TTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATT
CAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGC
TGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTT
TCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACAC
TATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCC
ATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGG
GATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATG
GCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTT
GCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTG
CAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGA
TCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTT
AATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCC
CGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGT
GGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTG
TAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTG
GCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACA
CAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAA
AGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAG
TCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
GGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGC
CTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGG
TATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTAT
ACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCC
CGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAG
CAGATCAATTCGCGCGCGAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGGACGAAGCAGGGATTCTGCAAACCCTATGCTACTCC
GTCAAGCCGTCAATTGTCTGATTCGTTACCAATTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGC
TCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCA
TCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAA
GATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACT
GACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATT
TATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCA
TCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTG
GTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAA
TTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAA
TCGAGATAACCGTTGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCT
TCAGCCATACTTTTCATACTCCCGCCATTCAGAG

pUC18 (Underlined: ORF)

TCGCGCGTTTCGGTGATGACGGTGAAAACCTCTGACACATGCAGCTCCCGGAGACGGTCACAGCTTGTCTGTAAGCGGATGCCGGGAGC
AGACAAGCCCGTCAGGGCGCGTCAGCGGGTGTTGGCGGGTGTCGGGGCTGGCTTAACTATGCGGCATCAGAGCAGATTGTACTGAGAGT
GCACCATATGCGGTGTGAAATACCGCACAGATGCGTAAGGAGAAAATACCGCATCAGGCGCCATTCGCCATTCAGGCTGCGCAACTGTTG
GGAAGGGCGATCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAG
GGTTTTCCCAGTCACGACGTTGTAAAACGACGGCCAGTGCCAAGCTTGCATGCCTGCAGGTCGACTCTAGAGGATCCCCGGGTACCGAG
CTCGAATTCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGT

GTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCT
CGGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAAGGCGGTAATACGGTTATCCACAGAATCAGGGGATAACGCAGGAAAGAAC
ATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTTTCCATAGGCTCCGCCCCCCTGACGAGCA
TCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCCTGGAAGCTCCCTCGTGC
GCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCGTGGCGCTTTCTCATAGCTCACGCTGTAG
GTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCAGCCCGACCGCTGCGCCTTATCCGGTAA
CTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAACAGGATTAGCAGAGCGAGGTATGTAGG
CGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGGTATCTGCGCTCTGCTGAAGCCAGTTAC
CTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGTTTGCAAGCAGCAGATTACGCG
CAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAACGAAAACTCACGTTAAGGGATTTTGGT
CATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTC
TGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCCGTCGTGTAGA
TAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAAT
AAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCCAGTCTATTAATTGTTGCCGGGAAGCTAG
AGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGGTATGGCTTCAT
TCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGT
CAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTG
TGACTGGTGAGTACTCAACCAAGTCATTCTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGATAATACCGC
GCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAG
TTCGATGTAACCCACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAA
TGCCGCAAAAAAGGGAATAAGGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATT
GTCTCATGAGCGGATACATATTTGAATGTATTTAGAAAAATAAACAAATAGGGGTTCCGCGCACATTTCCCCGAAAAGTGCCACCTGACGT
CTAAGAAACCATTATTATCATGACATTAACCTATAAAAATAGGCGTATCACGAGGCCCTTTCGTCGACGAAAGGGCCTCGTGATACGCCTAT
TTTTATAGGTTAATGTCATGATAATAATGGTTTCTTAGACGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTT
TCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTATGAGTATTCAA
CATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCTGGTGAAAGTAAAAGATGCTGA
AGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTTTTCGCCCCGAAGAACGTTTTCC
AATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTATTGACGCCGGGCAAGAGCAACTCGGTCGCCGCATACACTAT
TCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAGAGAATTATGCAGTGCTGCCATA
ACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGCTTTTTTGCACAACATGGGGGAT
CATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACACCACGATGCCTGTAGCAATGGC
AACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGGATGGAGGCGGATAAAGTTGC
AGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGTGGGTCTCGCGGTATCATTGCA
GCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATGGATGAACGAAATAGACAGATC
GCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGATTGATTTAAAACTTCATTTTTAA
TTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCG
TAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGG
TTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTA
GCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGGC
GATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACA
GCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAG
GCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCC
TGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGG
CCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCT
TTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAA
ACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAA
TTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAA
TTTCACACAGGAAACAGCTATGACCATGATTACGAATTCGAGCTCGGTACCCGGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTT
GGCACTGGCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAG
CTGGCGTAATAGCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCCTGATGCGGTATTTTCT
CCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGCCCCGACACC
CGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGT
GTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGA

pSF-TAC (Underlined: ORF)

GCGATCGCGGCTCCCGACATCTTGGACCATTAGCTCCACAGGTATCTTCTTCCCTCTAGTGGTCATAACAGCAGCTTCAGCTACCTCTCAAT

TCAAAAAACCCCTCAAGACCCGTTTAGAGGCCCCAAGGGGTTATGCTATCAATCGTTGCGTTACACACACAAAAAACCAACACACATCCA
TCTTCGATGGATAGCGATTTTATTATCTAACTGCTGATCGAGTGTAGCCAGATCTAAGCTGTGGTATGGCTGTCCAGGTCGTAAATCACTGC
ATAATTCGTGTCGCTCAAGGCGCACTCCCGTTCTGGATAATGTTTTTTGCGCCGACATCATAACGGTTCTGGCAAATATTCTGAAATGAGCT
GGAGCTGTTGACAATTAATCATCGGCTCGTATAATGTGTGGAATTGTGAGCGGATAACAATTTCACACAGATCTTTGTCGATCCTACCATCC
ACTCGACACACCCGCCAGCGGCCGCTGCCAAGCTTCCGAGCTCTCGAATTCAAAGGAGGTACCCACC<u>ATGGGGTACCGCGATATCTACCT</u>
<u>CGAGGTTTCTAGAAGTTGTCTCCTCCTGCACTGA</u>CTGACTGATACAATCGATTTCTGGATCCGCAGGCCTCTGCTAGCTTGACTGACTGAG
ATACAGCGTACCTTCAGCTCACAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTAT
TTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCA
GGTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTGGTATTGGCCCATCTCTATCGGTATCGTAGCATAAC
CCCTTGGGGCCTCTAAACGGGTCTTGAGGGGTTTTTTGTGCCCCTCGGGCCGGATTGCTATCTACCGGCATTGGCGCAGAAAAAAATGCC
TGATGCGACGCTGCGCGTCTTATACTCCCACATATGCCAGATTCAGCAACGGATACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCCTC
CTTACCAGAAATTTATCCTTAAGGTCGTCAGCTATCCTGCAGGCGATCTCTCGATTTCGATCAAGACATTCCTTTAATGGTCTTTTCTGGAC
ACCACTAGGGGTCAGAAGTAGTTCATCAAACTTTCTTCCCTCCCTAATCTCATTGGTTACCTTGGGCTATCGAAACTTAATTAACCAGTCAA
GTCAGCTACTTGGCGAGATCGACTTGTCTGGGTTTCGACTACGCTCAGAATTGCGTCAGTCAAGTTCGATCTGGTCCTTGCTATTGCACCC
GTTCTCCGATTACGAGTTTCATTTAAATCATGTGAGCAAAAGGCCAGCAAAAGGCCAGGAACCGTAAAAAGGCCGCGTTGCTGGCGTTTT
TCCATAGGCTCCGCCCCCCTGACGAGCATCACAAAAATCGACGCTCAAGTCAGAGGTGGCGAAACCCGACAGGACTATAAAGATACCAG
GCGTTTCCCCCTGGAAGCTCCCTCGTGCGCTCTCCTGTTCCGACCCTGCCGCTTACCGGATACCTGTCCGCCTTTCTCCCTTCGGGAAGCG
TGGCGCTTTCTCATAGCTCACGCTGTAGGTATCTCAGTTCGGTGTAGGTCGTTCGCTCCAAGCTGGGCTGTGTGCACGAACCCCCCGTTCA
GCCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGTAAGACACGACTTATCGCCACTGGCAGCAGCCACTGGTAA
CAGGATTAGCAGAGCGAGGTATGTAGGCGGTGCTACAGAGTTCTTGAAGTGGTGGCCTAACTACGGCTACACTAGAAGAACAGTATTTGG
TATCTGCGCTCTGCTGAAGCCAGTTACCTTCGGAAAAAGAGTTGGTAGCTCTTGATCCGGCAAACAAACCACCGCTGGTAGCGGTGGTTT
TTTTGTTTGCAAGCAGCAGATTACGCGCAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTACGGGGTCTGACGCTCAGTGGAA
CGAAAACTCACGTTAAGGGATTTTGGTCATGAGATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCA
ATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCC
ATAGTTGCATTTAAATTTCCGAACTCTCCAAGGCCCTCGTCGGAAAATCTTCAAACCTTTCGTCCGATCCATCTTGCAGGCTACCTCTCGA
ACGAACTATCGCAAGTCTCTTGGCCGGCCTTGCGCCTTGGCTATTGCTTGGCAGCGCCTATCGCCAGGTATTACTCCAATCCCGAATATCCG
AGATCGGGATCACCCGAGAGAAGTTCAACCTACATCCTCAATCCCGATCTATCCGAGATCCGAGGAATATCGAAATCGGGGCGCGCCTGG
TGTACCGAGAACGATCCTCTCAGTGCGAGTCTCGACGATCCATATCGTTGCTTGGCAGTCAGCCAGTCGGAATCCAGCTTGGGACCCAGG
AAGTCCAATCGTCAGATATTGTACTCAAGCCTGGTCACGGCAGCGTACCGATCTGTTTAAACCTAGATATTGATAGTCTGATCGGTCAACGT
ATAATCGAGTCCTAGCTTTTGCAAACATCTATCAAGAGACAGGATCAGCAGGAGGCTTTCGCATGATTGAACAAGATGGATTGCACGCAG
GTTCTCCGGCGGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGT
CAGCGCAGGGGCGTCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGG
CTGGCGACGACGGGCGTTCCTTGCGCGGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGG
GCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCT
ACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGAC
GAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGTCTATGCCCGACGGCGAGGATCTCGTCGTGACCCACGG
CGATGCCTGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGTCTGGGTGTGGCGGACCGCTATCAG
GACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTTGTGCTTTACGGTATCGCCGCGCCCG
ATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGACCGATTCTAGGTGCATTGGCGCAGAAAAAAATGCCTGATGCGAC
GCTGCGCGTCTTATACTCCCACATATGCCAGATTCAGCAACGGATACGGCTTCCCCAACTTGCCCACTTCCATACGTGTCCTCCTTACCAGA
AATTTATCCTTAAGGTCGTTTAAACTCGACTCTGGCTCTATCGAATCTCCGTCGTTTCGAGCTTACGCGAACAGCCGTGGCGCTCATTTGCT
CGTCGGGCATCGAATCTCGTCAGCTATCGTCAGCTTACCTTTTTGGCA

pET28a(+)-*efp* (Underlined: ORF)

TGGCGAATGGGACGCGCCCTGTAGCGGCGCATTAAGCGCGGCGGGTGTGGTGGTTACGCGCAGCGTGACCGCTACACTTGCCAGCGCCC
TAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTCGCCACGTTCGCCGGCTTTCCCCGTCAAGCTCTAAATCGGGGGCTCCCTTTAGGG
TTCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAAACTTGATTAGGGTGATGGTTCACGTAGTGGGCCATCGCCCTGATAGACGGTTT
TTCGCCCTTTGACGTTGGAGTCCACGTTCTTTAATAGTGGACTCTTGTTCCAAACTGGAACAACACTCAACCCTATCTCGGTCTATTCTTTT
GATTTATAAGGGATTTTGCCGATTTCGGCCTATTGGTTAAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAAC
GTTTACAATTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCAT
GAATTAATTCTTAGAAAAACTCATCGAGCATCAAATGAAACTGCAATTTATTCATATCAGGATTATCAATACCATATTTTTGAAAAAGCCGTT
TCTGTAATGAAGGAGAAAACTCACCGAGGCAGTTCCATAGGATGGCAAGATCCTGGTATCGGTCTGCGATTCCGACTCGTCCAACATCAA
TACAACCTATTAATTTCCCCTCGTCAAAAATAAGGTTATCAAGTGAGAAATCACCATGAGTGACGACTGAATCCGGTGAGAATGGCAAAA
GTTTATGCATTTCTTTCCAGACTTGTTCAACAGGCCAGCCATTACGCTCGTCATCAAAATCACTCGCATCAACCAAACCGTTATTCATTCGT
GATTGCGCCTGAGCGAGACGAAATACGCGATCGCTGTTAAAAGGACAATTACAAACAGGAATCGAATGCAACCGGCGCAGGAACACTGC
CAGCGCATCAACAATATTTTCACCTGAATCAGGATATTCTTCTAATACCTGGAATGCTGTTTTCCCGGGGATCGCAGTGGTGAGTAACCATG
CATCATCAGGAGTACGGATAAAATGCTTGATGGTCGGAAGAGGCATAAATTCCGTCAGCCAGTTTAGTCTGACCATCTCATCTGTAACATC

ATTGGCAACGCTACCTTTGCCATGTTTCAGAAACAACTCTGGCGCATCGGGCTTCCCATACAATCGATAGATTGTCGCACCTGATTGCCCG
ACATTATCGCGAGCCCATTTATACCCATATAAATCAGCATCCATGTTGGAATTTAATCGCGGCCTAGAGCAAGACGTTTCCCGTTGAATATG
GCTCATAACACCCCTTGTATTACTGTTTATGTAAGCAGACAGTTTTATTGTTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGA
GCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCG
CTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACT
GTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGG
CTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGG
GGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCC
CGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTG
GTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAAC
GCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAA
CCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAG
CGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCAT
AGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACG
GGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAA
ACGCGCGAGGCAGCTGCGGTAAAGCTCATCAGCGTGGTCGTGAAGCGATTCACAGATGTCTGCCTGTTCATCCGCGTCCAGCTCGTTGAG
TTTCTCCAGAAGCGTTAATGTCTGGCTTCTGATAAAGCGGGCCATGTTAAGGGCGGTTTTTTCCTGTTTGGTCACTGATGCCTCCGTGTAA
GGGGGATTTCTGTTCATGGGGGTAATGATACCGATGAAACGAGAGAGGATGCTCACGATACGGGTTACTGATGATGAACATGCCCGGTTAC
TGGAACGTTGTGAGGGTAAACAACTGGCGGTATGGATGCGGCGGGACCAGAGAAAAATCACTCAGGGTCAATGCCAGCGCTTCGTTAAT
ACAGATGTAGGTGTTCCACAGGGTAGCCAGCAGCATCCTGCGATGCAGATCCGGAACATAATGGTGCAGGGCGCTGACTTCCGCGTTTCC
AGACTTTACGAAACACGGAAACCGAAGACCATTCATGTTGTTGCTCAGGTCGCAGACGTTTTGCAGCAGCAGTCGCTTCACGTTCGCTCG
CGTATCGGTGATTCATTCTGCTAACCAGTAAGGCAACCCCGCCAGCCTAGCCGGGTCCTCAACGACAGGAGCACGATCATGCGCACCCGT
GGGGCCGCCATGCCGGCGATAATGGCCTGCTTCTCGCCGAAACGTTTGGTGGCGGGACCAGTGACGAAGGCTTGAGCGAGGGCGTGCAA
GATTCCGAATACCGCAAGCGACAGGCCGATCATCGTCGCGCTCCAGCGAAAGCGGTCCTCGCCGAAAATGACCCAGAGCGCTGCCGGCA
CCTGTCCTACGAGTTGCATGATAAAGAAGACAGTCATAAGTGCGGCGACGATAGTCATGCCCCGCGCCCACCGGAAGGAGCTGACTGGG
TTGAAGGCTCTCAAGGGCATCGGTCGAGATCCCGGTGCCTAATGAGTGAGCTAACTTACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC
AGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCCAGGGTGGTTTT
TCTTTTCACCAGTGAGACGGGCAACAGCTGATTGCCCTTCACCGCCTGGCCCTGAGAGAGTTGCAGCAAGCGGTCCACGCTGGTTTGCC
CCAGCAGGCGAAAATCCTGTTTGATGGTGGTTAACGGCGGGATATAACATGAGCTGTCTTCGGTATCGTCGTATCCCACTACCGAGATATC
CGCACCAACGCGCAGCCCGGACTCGGTAATGGCGCGCATTGCGCCCAGCGCCATCTGATCGTTGGCAACCAGCATCGCAGTGGGAACGA
TGCCCTCATTCAGCATTTGCATGGTTTGTTGAAAACCGGACATGGCACTCCAGTCGCCTTCCCGTTCCGCTATCGGCTGAATTTGATTGCG
AGTGAGATATTTATGCCAGCCAGCCAGACGCAGACGCGCCGAGACAGAACTTAATGGGCCCGCTAACAGCGCGATTTGCTGGTGACCCA
ATGCGACCAGATGCTCCACGCCCAGTCGCGTACCGTCTTCATGGGAGAAAATAATACTGTTGATGGGTGTCTGGTCAGAGACATCAAGAA
ATAACGCCGGAACATTAGTGCAGGCAGCTTCCACAGCAATGGCATCCTGGTCATCCAGCGGATAGTTAATGATCAGCCCACTGACGCGTT
GCGCGAGAAGATTGTGCACCGCCGCTTTACAGGCTTCGACGCCGCTTCGTTCTACCATCGACACCACCACGCTGGCACCCAGTTGATCGG
CGCGAGATTTAATCGCCGCGACAATTTGCGACGGCGCGTGCAGGGCCAGACTGGAGGTGGCAACGCCAATCAGCAACGACTGTTTGCCC
GCCAGTTGTTGTGCCACGCGGTTGGGAATGTAATTCAGCTCCGCCATCGCCGCTTCCACTTTTTCCCGCGTTTTCGCAGAAACGTGGCTGG
CCTGGTTCACCACGCGGGAAACGGTCTGATAAGAGACACCGGCATACTCTGCGACATCGTATAACGTTACTGGTTTCACATTCACCACCCT
GAATTGACTCTCTTCCGGGCGCTATCATGCCATACCGCGAAAGGTTTTGCGCCATTCGATGGTGTCCGGGATCTCGACGCTCTCCCTTATGC
GACTCCTGCATTAGGAAGCAGCCCAGTAGTAGGTTGAGGCCGTTGAGCACCGCCGCCGCAAGGAATGGTGCATGCAAGGAGATGGCGCC
CAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCTCATGAGCCCGAAGTGGCGAGCCCGATCTTCCCCAT
CGGTGATGTCGGCGATATAGGCGCCAGCAACCGCACCTGTGGCGCCGGTGATGCCGGCCACGATGCGTCCGGCGTAGAGGATCGAGATCT
CGATCCCGCGAAATTAATACGACTCACTATAGGGGAATTGTGAGCGGATAACAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGG
AGATATACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGAAGTTCTGTTCCAGGGGCCCCTGCATATGGCAACGTACT
ATAGCAACGATTTTCGTGCTGGTCTTAAAATCATGTTAGACGGCGAACCTTACGCGGTTGAAGCGAGTGAATTCGTAAAACCGGGTAAAG
GCCAGGCATTTGCTCGCGTTAAACTGCGTCGTCTGCTGACCGGTACTCGCGTAGAAAAAACCTTCAAATCTACTGATTCCGCTGAAGGCG
CTGATGTTGTCGATATGAACCTGACTTACCTGTACAACGACGGTGAGTTCTGGCACTTCATGAACAACGAAACTTTCGAGCAGCTGTCTG
CTGACGCAAAAGCAATTGGCGACAACGCTAAATGGTTGCTGGATCAGGCAGAGTGTATCGTAACTCTGTGGAATGGTCAGCCGATCTCCG
TTACTCCGCCGAACTTCGTTGAACTGGAAATCGTTGATACCGATCCGGGCCTGAAAGGTGATACCGCAGGTACTGGCGGCAAACCGGCTA
CCCTGTCTACTGGCGCTGTGGTTAAAGTTCCGCTGTTTGTACAAATCGGCGAAGTCATCAAAGTGGATACCCGCTCTGGTGAATACGTCTC
TCGCGTGAAGTGACAAAGCCCGAAAGGAAGCTGAGTTGGCTGCTGCCACCGCTGAGCAATAACTAGCATAACCCCTTGGGGCCTCTAAA
CGGGTCTTGAGGGGTTTTTTGCTGAAAGGAGGAACTATATCCGGAT

His_6-PreScission-EF-P

MGSSHHHHHHSSGLEVLFQGPLHMATYYSNDFRAGLKIMLDGEPYAVEASEFVKPGKGQAFARVKLRRLLTGTRVEKTFKSTDSAEGADVVD
MNLTYLYNDGEFWHFMNNETFEQLSADAKAIGDNAKWLLDQAECIVTLWNGQPISVTPPNFVELEIVDTDPGLKGDTAGTGGKPATLSTGAV
VKVPLFVQIGEVIKVDTRSGEYVSRVK*

pLysS

GAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCTTATTTTTCTTTACGGTCTTTAAAAAG
GCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAAATGTTCTTTACGATGCCATTGGGATAT
ATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGATAACTCAAAAAATACGCCCGGTAGTG
ATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAAAGTTGGCCCAGGGCTTCCCGGTATCA
ACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCAAAGTGCGTCGGGTGATGCTGCCAACT
TACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCCTCCTGTTCAGCTACTGACGGGGTGGT
GCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTGGCACTGATGAGGGTGTCAGTGAAGTG
CTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATATATTCCGCTTCCTCGCTCACTGACTCGCT
ACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGGAAGATGCCAGGAAGATACTTAACAGGG
AAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACGAAATCTGACGCTCAAATCAGTGGTGGC
GAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGCGGCTCCCTCGTGCGCTCTCCTGTTCCTGCCTTTCGGTTTACCGGTGTC
ATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTTCGCTCCAAGCTGGACTGTATGCACGAA
CCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAAAGACATGCAAAAGCACCACTGGCAGCA
GCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGAAAGGACAAGTTTTGGTGACTGCGCTCC
TCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCTGCAAGGCGGTTTTTTCGTTTTCAGAGC
AAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTTCTAGATTTCAGTGCAATTTATCTCTTCA
AATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTAATTCTCATGTTTGACAGCTTATCATCGATAAGCTTTAATGCGGTAGTTTATC
ACAGTTAAATTGCTAACGCAGTCAGGCACCGTGTATGAAATCTAACAATGCGCTCATCGTCATCCTCGGCACCGTCACCCTGGATGCTGTA
GGCATAGGCTTGGTTATGCCGGTACTGCCGGGCCTCTTGCGGGATATCGTCCATTCCGACAGCATCGCCAGTCACTATGGCGTGCTGCTAG
CGCTATATGCGTTGATGCAATTTCTATGCGCACCCGTTCTCGGAGCACTGTCCGACCGCTTTGGCCGCCGCCCAGTCCTGCTCGCTTCGCT
ACTTGGAGCCACTATCGACTACGCGATCATGGCGACCACACCCGTCCTGTGGATCCGGCCCATTGGCTGCCTCCCACACTTGGATATGCCT
CCTCGGAGCCTTATAGAATTGTTTATAAGACTTGCGCATTATTTGACCTCCAATGCGAACAAAGGGAAACCGCTGTGGTCTCCCTTTAGTG
AGTTCAATTAATTATCCACGGTCAGAAGTGACCAGTTCGTTCTTCTCCCACCAACGCTTAAGGTCGAACGAAGGGCAAGCCTTCGGCGCC
ACCTCATGATGGGCGCGAAGACCAGCGCCTTCGTACTTAGCCAGCAGTGTGACAAGCAGTGAGCGAAGGGATTGCATTTGGGCTGGCGT
AAAGTTAGCGTCGAACTTACCTTTATCGTCGATACCACCAACAAGGCAGACGCCGATAGAGTTGTGGTTGTAACCCTTAGCGTGAGAGCC
TACAGCCATCTCATCTCGTCCTGCCTCCACAGTACCGTCTCGCTTGATGATAAAGTGGTATCCCACATCGAGCCAACCCTGCTCTTTGTGCC
ACTGGCGAATCTCACGGACACCAACATTCTGACTTGGCTTGGTAGCCGAGCAGTGAACAAAGATTGCGTCAGTAGATTCACGTTGTTTAA
ACTGTACACGAGCCATTATTTCTTTCCTCCTTTCCTTTTTAATCTATCAAAGGGGACCCGGATCCTCTACGCCGGACGCATCGTGGCCGGCA
TCACCGGCGCCACAGGTGCGGTTGCTGGCGCCTATATCGCCGACATCACCGATGGGGAAGATCGGGCTCGCCACTTCGGGCTCATGAGCG
CTTGTTTCGGCGTGGGTATGGTGGCAGGCCCCGTGGCCGGGGGACTGTTGGGCGCCATCTCCTTGCATGCACCATTCCTTGCGGCGGCGG
TGCTCAACGGCCTCAACCTACTACTGGGCTGCTTCCTAATGCAGGAGTCGCATAAGGGAGAGCGTCGACCGATGCCCTTGAGAGCCTTCA
ACCCAGTCAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCAACTCGTAGGACAGGT
GCCGGCAGCGCTCTGGGTCATTTTCGGCGAGGACCGCTTTCGCTGGAGCGCGACGATGATCGGCCTGTCGCTTGCGGTATTCGGAATCTT
GCACGCCCTCGCTCAAGCCTTCGTCACTGGTCCCGCCACCAAACGTTTCGGCGAGAAGCAGGCCATTATCGCCGGCATGGCGGCCGACG
CGCTGGGCTACGTCTTGCTGGCGTTCGCGACGCGAGGCTGGATGGCCTTCCCCATTATGATTCTTCTCGCTTCCGGCGGCATCGGGATGCC
CGCGTTGCAGGCCATGCTGTCCAGGCAGGTAGATGACGACCATCAGGGACAGCTTCAAGGATCGCTCGCGGCTCTTACCAGCCTAACTTC
GATCACTGGACCGCTGATCGTCACGGCGATTTATGCCGCCTCGGCGAGCACATGGAACGGGTTGGCATGGATTGTAGGCGCCGCCCTATA
CCTTGTCTGCCTCCCCGCGTTGCGTCGCGGTGCATGGAGCCGGGCCACCTCGACCTGAATGGAAGCCGGCGGCACCTCGCTAACGGATTC
ACCACTCCAAGAATTGGAGCCAATCAATTCTTGCGGAGAACTGTGAATGCGCAAACCAACCCTTGGCAGAACATATCCATCGCGTCCGCC
ATCTCCAGCAGCCGCACGCGGCGCATCTCGGGCAGCGTTGGGTCCTGGCCACGGGTGCGCATGATCGTGCTCCTGTCGTTGAGGACCCG
GCTAGGCTGGCGGGGTTGCCTTACTGGTTAGCAGAATGAATCACCGATACGCGAGCGAACGTGAAGCGACTGCTGCTGCAAAACGTCTG
CGACCTGAGCAACAACATGAATGGTCTTCGGTTTCCGTGTTTCGTAAAGTCTGGAAACGCGGAAGTCCCCTACGTGCTGCTGAAGTTGCC
CGCAACAGAGAGTGGAACCAACCGGTGATACCACGATACTATGACTGAGAGTCAACGCCATGAGCGGCCTCATTTCTTATTCTGAGTTAC
AACAGTCCGCACCGCTGTCCGGTAGCTCCTTCCGGTGGGCGCGGGGCATGACTATCGTCGCCGCACTTATGACTGTCTTCTTTATCATGCA
ACTCGTAGGACAGGTGCCGGCAGCGCCCAACAGTCCCCCGGCCACGGGGCCTGCCACCATACCCACGCCGAAACAAGCGCCCTGCACC
ATTATGTTCCGGATCTGCATCGCAGGATGCTGCTGGCTACCCTGTGGAACACCTACATCTGTATTAACGAAGCGCTAACCGTTTTTATCAGG
CTCTGGGAGGCAGAATAAATGATCATATCGTCAATTATTACCTCCACGGGGAGAGCCTGAGCAAACTGGCCTCAGGCATTTGAGAAGCAC
ACGGTCACACTGCTTCCGGTAGTCAATAAACCGGTAAACCAGCAATAGACATAAGCGGCTATTTAACGACCCTGCCCTGAACCGACGACC
GGGTCGAATTTGCTTTCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCACCAGGCGTTTAAGGGCACCAATAACTGC
CTTAAAAAAATTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCTGCCGACATGGAAGCCATCACAGACGGC
ATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATGGTGAAAACGGGGGCGAAGAAGTTGTC
CATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAACATATTCTCAATAAACCCTTTAGGGAA
ATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGAT
GAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACCAGCTCACCGTCTTTCATTGCCATACG

ORF of pBAD-P_{BAD}-*yjgZ*

The ORF of pBAD-HisA was replaced by Following DNA sequence.

TTGTTACCTCCTGGACCGCTTCTTGTCCTTCCAGTCGGAGCGAGAGTTAGCCGGGGCGGACTTTCACCGACGGAAAGATCGTGCATTTCA
AGGCACACACAACAATATCTGCGAGCGTGCTATCCGTCCGGTCGTTATGGGACGAAAGGCCTGGTTGTTCGCTGGTTCACTGGTGGCCGG
GAACCGCGCGGCACAGATAATGAGCCTTCTGGGAACCGCAGTCTGGAGCCACATGCTTGGCTGACGGACGTCCTGACGCGTCTGCCGGA
GTGGCCGGAGGAGAGGTTGGCTGAGTTACTGCCTCTTGAGGGCTTTACCTTCTTCGGGCATCATCATCATCATCATTGA

YjgZ-His$_6$

LLPPGPLLVLPVGARVSRGGLSPTERSCISRHTQQYLRACYPSGRYGTKGLVVRWFTGGREPRGTDNEPSGNRSLEPHAWLTDVLTRLPEWPEE
RLAELLPLEGFTFFGHHHHHH*

pBAD-P$_{BAD}$-*yjgZ'*

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGTTACCTCCTGGACCGCTTCTTGTCCTTCCAGTCGGAGCGAGAGTTAGCCGGGGCGGACTTTCACCGACGGAAAGATCGTGCATTTCA
AGGCACACACAACAATATCTGCGAGCGTGCTATCCGTCCGGTCGTTATGGGACGAAAGGCCTGGTTGTTCGCTGGTTCACTGGTGGCCGG
GAACCGCGCGGCACAGATAATGAGCCTTCTGGGAACCGCAGTCTGGAGCCACATGCTTGGCTGACGGACGTCCTGACGCGTCTGCCGGA
GTGGCCGGAGGAGAGGTTGGCTGAGTTACTGCCTCTTGAGGGCTTTACCTTCTTCGGGCATCATCATCATCATCATTGA

YjgZ'-His$_6$

MLPPGPLLVLPVGARVSRGGLSPTERSCISRHTQQYLRACYPSGRYGTKGLVVRWFTGGREPRGTDNEPSGNRSLEPHAWLTDVLTRLPEWPE
ERLAELLPLEGFTFFGHHHHHH*

pBAD-P$_{BAD}$-*prpR*

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGGCACATCCACCACGGCTTAATGACGACAAACCGGTTATCTGGACGGTATCTGTAACGCGCCTGTTCGAGCTGTTTCGCGATATCAGCC
TCGAGTTTGATCACCTGGCGAACATTACCCCTATTCAGCTTGGCTTTGAAAAAGCAGTGACCTACATCCGCAAGAAACTGGCAAACGAAC
GCTGTGACGCCATCATCGCGGCTGGCTCTAACGGCGCGTACCTGAAAAGCCGCCTGTCAGTGCCAGTTATTTTGATTAAACCGAGCGGCT
ACGATGTGTTACAGGCACTGGCAAAAGCCGGAAAACTCACCTCTTCTATCGGCGTTGTCACCTATCAGGAAACCATTCCAGCGCTGGTGG
CGTTTCAAAAAACCTTTAATTTGCGCCTCGACCAACGTAGCTACATTACCGAGGAAGACGCACGCGGGCAGATTAACGAGCTAAAAGCTA
ACGGCACCGAAGCGGTGGTCGGCGCAGGGCTGATTACCGATTTGGCAGAAGAAGCCGGAATGACCGGAATTTTTATCTATTCCGCCGCCA
CCGTGCGCCAGGCGTTCAGCGATGCGCTGGATATGACGCGCATGTCGTTACGCCATAACACTCACGATGCCACCCGCAACGCCCTGCGTA
CTCGTTACGTGCTGGGCGATATGCTCGGTCAATCACCACAGATGGAACAAGTACGGCAGACTATTTTGCTGTATGCCCGCTCCAGTGCGGC
GGTGTTGATTGAGGGGGAAACGGGGACGGGCAAAGAGCTGGCGGCCCAGGCGATTCATCGGGAATATTTTGCCCGCCACGATGCGCGAC
AGGGCAAAAAGTCGCATCCGTTTGTTGCCGTCAACTGCGGGGCGATTGCCGAATCGCTGCTGGAAGCAGAACTGTTTGGCTATGAGGAA
GGGGCGTTTACCGGCTCGCGACGCGGAGGTCGCGCCGGGCTGTTCGAAATTGCCCACGGCGGTACGCTGTTTCTGGATGAGATTGGCGA
AATGCCGCTACCTTTGCAGACCCGGCTGCTGCGGGTGCTGGAAGAAAAAGAGGTCACCCGCGTCGGCGGGCATCAGCCTGTTCCGGTAG
ATGTACGGGTCATTAGCGCCACTCACTGCAATCTGGAAGAAGATATGCAGCAAGGACGTTTTCGCCGTGATCTGTTTTATCGGCTGAGTAT
TTTGCGTCTGCAATTGCCACCACTGCGCGAGCGGGTGGCGGATATTCTTCCGCTGGCGGAAAGCTTTTTGAAAGTGTCTCTGGCGGCGCT
CTCCGCCCCATTTTCTGCTGCATTACGCCAGGGGTTACAGGCAAGTGAAACTGTGCTGCTGCACTACGACTGGCCAGGCAATATTCGTGA
ACTGCGCAATATGATGGAACGACTGGCGCTGTTTTTAAGTGTGGAACCGACGCCGGATTTAACGCCGCAGTTTATGCAACTGCTACTGCC
GGAACTGGCGCGCGAGTCGGCGAAAACTCCCGCTCCACGCTTACTGACACCACAACAGGCACTGGAGAAATTTAATGGCGATAAAACAG
CAGCGGCGAATTATTTAGGCATCAGCCGGACGACGTTCTGGCGGCGGCTGAAAAGCCATCATCATCATCATCATTGA

PrpR-His$_6$

MAHPPRLNDDKPVIWTVSVTRLFELFRDISLEFDHLANITPIQLGFEKAVTYIRKKLANERCDAIIAAGSNGAYLKSRLSVPVILIKPSGYDVLQA
LAKAGKLTSSIGVVTYQETIPALVAFQKTFNLRLDQRSYITEEDARGQINELKANGTEAVVGAGLITDLAEEAGMTGIFIYSAATVRQAFSDALD
MTRMSLRHNTHDATRNALRTRYVLGDMLGQSPQMEQVRQTILLYARSSAAVLIEGETGTGKELAAQAIHREYFARHDARQGKKSHPFVAVNC
GAIAESLLEAELFGYEEGAFTGSRRGGRAGLFEIAHGGTLFLDEIGEMPLPLQTRLLRVLEEKEVTRVGGHQPVPVDVRVISATHCNLEEDMQQ
GRFRRDLFYRLSILRLQLPPLRERVADILPLAESFLKVSLAALSAPFSAALRQGLQASETVLLHYDWPGNIRELRNMMERLALFLSVEPTPDLTP
QFMQLLLPELARESAKTPAPRLLTPQQALEKFNGDKTAAANYLGISRTTFWRRLKSHHHHHH*

pBAD-P$_{BAD}$-*yhhM*

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGAGCAAACCACCTCTTTTCTTTATTGTTATCATTGGCTTAATTGTCGTCGCCGCATCGTTTCGTTTTATGCAACAGCGACGGGAAAAAGC
TGATAATGATATGGCTCCGCTCCAGCAAAAGCTGGTGGTGGTGAGCAACAAGCGGGAAAAACCGATTAACGATCGCCGTTCGCGCCAGC
AGGAAGTGACTCCGGCAGGCACCAGTATACGCTATGAGGCAAGCTTCAAACCGCAAAGCGGAGGAATGGAGCAGACGTTTCGCCTCGA
CGCCCAGCAGTACCACGCCCTGACAGTGGGCGATAAAGGTACGCTGAGCTATAAAGGAACGCGCTTTGTCAGCTTTGTAGGCGAACAAC
ATCATCATCATCATCATTGA

YhhM-His$_6$

MSKPPLFFIVIIGLIVVAASFRFMQQRREKADNDMAPLQQKLVVVSNKREKPINDRRSRQQEVTPAGTSIRYEASFKPQSGGMEQTFRLDAQQY
HALTVGDKGTLSYKGTRFVSFVGEQHHHHHH*

pBAD-P$_{BAD}$-*rutD*

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGAAACTTTCACTCTCACCTCCCCCTTATGCTGATGCGCCCGTAGTGGTGTTGATTTCGGGTCTTGGGGGTAGCGGCAGTTACTGGTTAC
CGCAACTGGCGGTGCTGGAGCAGGAGTATCAGGTAGTCTGTTACGACCAGCGCGGCACCGGCAATAATCCCGACACGCTGGCAGAAGAT
TACAGTATCGCCCAGATGGCAGCGGAACTGCATCAGGCGCTGGTAGCCGCAGGGATTGAGCATTACGCAGTGGTCGGCCATGCGCTCGGT
GCGCTGGTGGGAATGCAGCTGGCGCTGGATTATCCCGCGTCGGTAACTGTGCTGATCAGCGTTAACGGCTGGCTACGAATAAACGCCCAT
ACGCGCCGCTGTTTTCAGGTTCGCGAACGATTACTGTATAGCGGCGGCGCAGGCATGGGTGGAAGCGCAGCCGTTGTTCCTCTATCCC
GCCGACTGGATGGCGGCCCGCGCACCTCGCCTGGAGGCAGAAGACGCGCTGGCACTGGCGCATTTTCAGGGCAAAAATAATTTACTGCG
TCGACTTAACGCCCTCAAACGCGCTGACTTTAGTCACCATGCGGATCGCATCCGCTGCCCGGTGCAAATCATCTGCGCCAGTGATGATCTG
CTGGTGCCAACAGCATGTTCCAGTGAACTTCATGCCGCCCTGCCCGATAGCCAGAAAATGGTGATGCCCTATGGCGGACACGCCTGCAAC
GTGACCGATCCCGAAACGTTTAATGCTCTGTTACTCAACGGGCTTGCCAGCCTGTTACATCACCGTGAAGCCGCCCTGCATCATCATCATC
ATCATTGA

RutD-His$_6$

MKLSLSPPPYADAPVVVLISGLGGSGSYWLPQLAVLEQEYQVVCYDQRGTGNNPDTLAEDYSIAQMAAELHQALVAAGIEHYAVVGHALGAL
VGMQLALDYPASVTVLISVNGWLRINAHTRRCFQVRERLLYSGGAQAWVEAQPLFLYPADWMAARAPRLEAEDALALAHFQGKNNLLRRL
NALKRADFSHHADRIRCPVQIICASDDLLVPTACSSELHAALPDSQKMVMPYGGHACNVTDPETFNALLLNGLASLLHHREAALHHHHHH*

pBAD-P$_{BAD}$-*ydcO*

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGCGTCTGTTTTCTATTCCTCCACCCACGCTACTGGCGGGGTTTCTGGCGGTATTAATTGGCTACGCCAGTTCAGCGGCAATAATCTGGCA
AGCAGCGATTGTCGCCGGAGCCACCACTGCACAAATCTCTGGCTGGATGACGGCGCTGGGGCTGGCAATGGGCGTCAGTACGCTGACTC
TGACATTATGGTATCGCGTACCTGTTCTCACCGCATGGTCAACGCCTGGCGCGGCTTTGTTGGTCACCGGATTGCAGGGACTAACACTTAA
CGAAGCCATCGGCGTTTTTATTGTCACCAACGCGCTAATAGTCCTCTGCGGCATAACGGGACTCTTTGCTCGTCTGATGCGCATTATTCCGC
ACTCGCTTGCGGCGGCAATGCTTGCCGGGATTTTATTACGCTTTGGTTTACAGGCGTTTGCCAGTCTGGACGGTCAATTTACGTTGTGTGG
AAGTATGTTGCTGGTATGGCTGGCAACCAAGGCCGTTGCGCCGCGCTATGCGGTAATTGCCGCGATGATTATTGGGATCGTGATCGTCATC
GCGCAAGGTGACGTTGTCACAACTGATGTTGTCTTTAAACCCGTTCTCCCCACTTATATTACCCCTGATTTTTCGTTTGCTCACAGCCTGAG
CGTTGCACTCCCCCTTTTTCTGGTGACGATGGCATCGCAAAACGCACCGGGTATCGCAGCAATGAAAGCAGCTGGATATTCGGCTCCTGTT
TCGCCATTAATTGTATTTACTGGATTGCTGGCACTGGTTTTTTCCCCTTTCGGCGTTTATTCCGTCGGTATTGCGGCAATCACCGCGGCTATT
TGCCAAAGCCCGGAAGCGCATCCGGATAAAGATCAACGTTGGCTGGCCGCTGCCGTTGCAGGCATTTTCTATTTGCTCGCAGGTCTGTTT
GGTAGTGCCATTACCGGGATGATGGCTGCCCTGCCCGTAAGTTGGATCCAGATGCTGGCAGGTCTGGCGCTGTTAAGTACCATCGGCGGC
AGTTTGTATCAGGCGCTGCATAATGAGCGTGAGCGAGACGCGGCGGTGGTGGCATTTCTGGTAACGGCAAGTGGATTGACGCTGGTCGG
GATTGGTTCTGCGTTTTGGGGATTAATTGCCGGAGGCGTTTGTTACGTGGTGTTGAATTTAATCGCTGACAGAAACCGATATCATCATCATC
ATCATCATTGA

YdcO-His$_6$

MRLFSIPPPTLLAGFLAVLIGYASSAAIIWQAAIVAGATTAQISGWMTALGLAMGVSTLTLTLWYRVPVLTAWSTPGAALLVTGLQGLTLNEAIG
VFIVTNALIVLCGITGLFARLMRIIPHSLAAAMLAGILLRFGLQAFASLDGQFTLCGSMLLVWLATKAVAPRYAVIAAMIIGIVIVIAQGDVVTTD
VVFKPVLPTYITPDFSFAHSLSVALPLFLVTMASQNAPGIAAMKAAGYSAPVSPLIVFTGLLALVFSPFGVYSVGIAAITAAICQSPEAHPDKDQR
WLAAAVAGIFYLLAGLFGSAITGMMAALPVSWIQMLAGLALLSTIGGSLYQALHNERERDAAVVAFLVTASGLTLVGIGSAFWGLIAGGVCYV
VLNLIADRNRYHHHHHH*

pBAD-P$_{BAD}$-*efp*-AmpR (Underlined: ORF)

AAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATT
AAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGA
TTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTAC
TGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTAACCATGGCAACGTACTATAGCAACGATTTTCGTGCTGGTCTTAAAATCATG
TTAGACGGCGAACCTTACGCGGTTGAAGCGAGTGAATTCGTAAAACCGGGTAAAGGCCAGGCATTTGCTCGCGTTAAACTGCGTCGTCTG
CTGACCGGTACTCGCGTAGAAAAAACCTTCAAATCTACTGATTCCGCTGAAGGCGCTGATGTTGTCGATATGAACCTGACTTACCTGTACA
ACGACGGTGAGTTCTGGCACTTCATGAACAACGAAACTTTCGAGCAGCTGTCTGCTGACGCAAAAGCAATTGGCGACAACGCTAAATGG
TTGCTGGATCAGGCAGAGTGTATCGTAACTCTGTGGAATGGTCAGCCGATCTCCGTTACTCCGCCGAACTTCGTTGAACTGGAAATCGTTG
ATACCGATCCGGGCCTGAAAGGTGATACCGCAGGTACTGGCGGCAAACCGGCTACCCTGTCTACTGGCGCTGTGGTTAAAGTTCCGCTGT

TTGTACAAATCGGCGAAGTCATCAAAGTGGATACCCGCTCTGGTGAATACGTCTCTCGCGTGAAGTGATACAGATTAAATCAGAACGCAG
AAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTA
GCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGC
CTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCC
CGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTC
TACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAA
AGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCT
GGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAGCACTTTTAAAGTTCTGCTATGTGGCGCGGTATTATCCCGTGTTGACGCCGGGCAAGAGCA
ACTCGGTCGCCGCATACACTATTCTCAGAATGACTTGGTTGAGTACTCACCAGTCACAGAAAAGCATCTTACGGATGGCATGACAGTAAG
AGAATTATGCAGTGCTGCCATAACCATGAGTGATAACACTGCGGCCAACTTACTTCTGACAACGATCGGAGGACCGAAGGAGCTAACCGC
TTTTTTGCACAACATGGGGGATCATGTAACTCGCCTTGATCGTTGGGAACCGGAGCTGAATGAAGCCATACCAAACGACGAGCGTGACAC
CACGATGCCTGTAGCAATGGCAACAACGTTGCGCAAACTATTAACTGGCGAACTACTTACTCTAGCTTCCCGGCAACAATTAATAGACTGG
ATGGAGGCGGATAAAGTTGCAGGACCACTTCTGCGCTCGGCCCTTCCGGCTGGCTGGTTTATTGCTGATAAATCTGGAGCCGGTGAGCGT
GGGTCTCGCGGTATCATTGCAGCACTGGGGCCAGATGGTAAGCCCTCCCGTATCGTAGTTATCTACACGACGGGGAGTCAGGCAACTATG
GATGAACGAAATAGACAGATCGCTGAGATAGGTGCCTCACTGATTAAGCATTGGTAACTGTCAGACCAAGTTTACTCATATATACTTTAGAT
TGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTT
CCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAA
ACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACC
AAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTAC
CAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGA
ACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCA
CGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAA
ACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATG
GAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTG
TGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCG
GAAGAGCGCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGC
CGCATAGTTAAGCCAGTATACACTCCGCTATCGCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCC
TGACGGGCTTGTCTGCTCCCGGCATCCGCTTACAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCA
CCGAAACGCGCGAGGCAGCAGATCAATTCGCGCGCGAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGGACGAAGCAGGGATTCT
GCAAACCCTATGCTACTCCGTCAAGCCGTCAATTGTCTGATTCGTTACCAATTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCAC
AACCGGCACGGAACTCGCTCGGGCTGGCCCCGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAACATTGCGAC
CGACGGTGGCGATAGGCATCCGGGTGGTGCTCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCC
CTAACTGCTGGCGGAAAAGATGTGACAGACGCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCC
AGGTGATCGCTGATGTACTGACAAGCCTCGCGTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTA
ACAATTGCTCAAGCAGATTTATCGCCAGCAGCTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAA
ATGCGGCTGGTGCGCTTCATCCGGGCGAAAGAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGG
ACGAAAGTAAACCCACTGGTGATACCATTCGCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATC
ACCCGGTCGGCAAACAAATTCTCGTCCCTGATTTTTCACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGC
GGTCGGTCGATAAAAAAATCGAGATAACCGTTGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGC
AGGGGATCATTTTGCGCTTCAGCCATACTTTTCATACTCCCGCCATTCAGAG

EF-P

MATYYSNDFRAGLKIMLDGEPYAVEASEFVKPGKGQAFARVKLRRLLTGTRVEKTFKSTDSAEGADVVDMNLTYLYNDGEFWHFMNNETFE
QLSADAKAIGDNAKWLLDQAECIVTLWNGQPISVTPPNFVELEIVDTDPGLKGDTAGTGGKPATLSTGAVVKVPLFVQIGEVIKVDTRSGEYVS
RVK*

pBAD-P$_{BAD}$-*efp*-CmR-p15A (Underlined: ORF)

AAGAAACCAATTGTCCATATTGCATCAGACATTGCCGTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCAAACCGGTAACCCCGCTTATT
AAAAGCATTCTGTAACAAAGCGGGACCAAAGCCATGACAAAAACGCGTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGA
TTATTTGCACGGCGTCACACTTTGCTATGCCATAGCATTTTTATCCATAAGATTAGCGGATCCTACCTGACGCTTTTTATCGCAACTCTCTAC
TGTTTCTCCATACCCGTTTTTTGGGCTAACAGGAGGAATTAACCATGGCAACGTACTATAGCAACGATTTTCGTGCTGGTCTTAAAATCATG
TTAGACGGCGAACCTTACGCGGTTGAAGCGAGTGAATTCGTAAAACCGGGTAAAGGCCAGGCATTTGCTCGCGTTAAACTGCGTCGTCTG
CTGACCGGTACTCGCGTAGAAAAAACCTTCAAATCTACTGATTCCGCTGAAGGCGCTGATGTTGTCGATATGAACCTGACTTACCTGTACA
ACGACGGTGAGTTCTGGCACTTCATGAACAACGAAACTTTCGAGCAGCTGTCTGCTGACGCAAAAGCAATTGGCGACAACGCTAAATGG
TTGCTGGATCAGGCAGAGTGTATCGTAACTCTGTGGAATGGTCAGCCGATCTCCGTTACTCCGCCGAACTTCGTTGAACTGGAAATCGTTG
ATACCGATCCGGGCCTGAAAGGTGATACCGCAGGTACTGGCGGCAAACCGGCTACCCTGTCTACTGGCGCTGTGGTTAAAGTTCCGCTGT
TTGTACAAATCGGCGAAGTCATCAAAGTGGATACCCGCTCTGGTGAATACGTCTCTCGCGTGAAGTGATACAGATTAAATCAGAACGCAG

AAGCGGTCTGATAAAACAGAATTTGCCTGGCGGCAGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAACTCAGAAGTGAAACGCCGTA
GCGCCGATGGTAGTGTGGGGTCTCCCCATGCGAGAGTAGGGAACTGCCAGGCATCAAATAAAACGAAAGGCTCAGTCGAAAGACTGGGC
CTTTCGTTTTATCTGTTGTTTGTCGGTGAACGCTCTCCTGAGTAGGACAAATCCGCCGGGAGCGGATTTGAACGTTGCGAAGCAACGGCC
CGGAGGGTGGCGGGCAGGACGCCCGCCATAAACTGCCAGGCATCAAATTAAGCAGAAGGCCATCCTGACGGATGGCCTTTTTGCGTTTC
TACAAACTCTTTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAA
AGGAAGAGTATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTTTTTGCTCACCCAGAAACGCT
GGTGAAAGTAAAAGATGCTGAAGATCAGTTGGGTGCACGAGTGGGTTACATCGAACTGGATCTCAACAGCGGTAAGATCCTTGAGAGTT
TTCGCCCCGAAGAACGTTTTCCAATGATGAATTAATTCTTACGCCCCGCCCTGCCACTCATCGCAGTACTGTTGTAATTCATTAAGCATTCT
GCCGACATGGAAGCCATCACAGACGGCATGATGAACCTGAATCGCCAGCGGCATCAGCACCTTGTCGCCTTGCGTATAATATTTGCCCATG
GTGAAAACGGGGGCGAAGAAGTTGTCCATATTGGCCACGTTTAAATCAAAACTGGTGAAACTCACCCAGGGATTGGCTGAGACGAAAAA
CATATTCTCAATAAACCCTTTAGGGAAATAGGCCAGGTTTTCACCGTAACACGCCACATCTTGCGAATATATGTGTAGAAACTGCCGGAAA
TCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTTTGCTCATGGAAAACGGTGTAACAAGGGTGAACACTATCCCATATCACC
AGCTCACCGTCTTTCATTGCCATACGGAATTCCGGATGAGCATTCATCAGGCGGGCAAGAATGTGAATAAAGGCCGGATAAAACTTGTGCT
TATTTTTCTTTACGGTCTTTAAAAAGGCCGTAATATCCAGCTGAACGGTCTGGTTATAGGTACATTGAGCAACTGACTGAAATGCCTCAAA
ATGTTCTTTACGATGCCATTGGGATATATCAACGGTGGTATATCCAGTGATTTTTTTCTCCATTTTAGCTTCCTTAGCTCCTGAAAATCTCGAT
AACTCAAAAAATACGCCCGGTAGTGATCTTATTTCATTATGGTGAAAGTTGGAACCTCTTACGTGCCGATCAACGTCTCATTTTCGCCAAA
AGTTGGCCCAGGGCTTCCCGGTATCAACAGGGACACCAGGATTTATTTATTCTGCGAAGTGATCTTCCGTCACAGGTATTTATTCGGCGCA
AAGTGCGTCGGGTGATGCTGCCAACTTACTGATTTAGTGTATGATGGTGTTTTTGAGGTGCTCCAGTGGCTTCTGTTTCTATCAGCTGTCCC
TCCTGTTCAGCTACTGACGGGGTGGTGCGTAACGGCAAAAGCACCGCCGGACATCAGCGCTAGCGGAGTGTATACTGGCTTACTATGTTG
GCACTGATGAGGGTGTCAGTGAAGTGCTTCATGTGGCAGGAGAAAAAAGGCTGCACCGGTGCGTCAGCAGAATATGTGATACAGGATAT
ATTCCGCTTCCTCGCTCACTGACTCGCTACGCTCGGTCGTTCGACTGCGGCGAGCGGAAATGGCTTACGAACGGGGCGGAGATTTCCTGG
AAGATGCCAGGAAGATACTTAACAGGGAAGTGAGAGGGCCGCGGCAAAGCCGTTTTTCCATAGGCTCCGCCCCCCTGACAAGCATCACG
AAATCTGACGCTCAAATCAGTGGTGGCGAAACCCGACAGGACTATAAAGATACCAGGCGTTTCCCCTGGCGGCTCCCTCGTGCGCTCTCC
TGTTCCTGCCTTTCGGTTTACCGGTGTCATTCCGCTGTTATGGCCGCGTTTGTCTCATTCCACGCCTGACACTCAGTTCCGGGTAGGCAGTT
CGCTCCAAGCTGGACTGTATGCACGAACCCCCCGTTCAGTCCGACCGCTGCGCCTTATCCGGTAACTATCGTCTTGAGTCCAACCCGGAA
AGACATGCAAAAGCACCACTGGCAGCAGCCACTGGTAATTGATTTAGAGGAGTTAGTCTTGAAGTCATGCGCCGGTTAAGGCTAAACTGA
AAGGACAAGTTTTGGTGACTGCGCTCCTCCAAGCCAGTTACCTCGGTTCAAAGAGTTGGTAGCTCAGAGAACCTTCGAAAAACCGCCCT
GCAAGGCGGTTTTTTCGTTTTCAGAGCAAGAGATTACGCGCAGACCAAAACGATCTCAAGAAGATCATCTTATTAATCAGATAAAATATTT
CTAGATTTCAGTGCAATTTATCTCTTCAAATGTAGCACCTGAAGTCAGCCCCATACGATATAAGTTGTCGGGAGCTGCATGTGTTTTTACGG
TTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCT
GATACCGCTCGCCGCAGCCGAACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCTGATGCGGTATTTTCTCCTTAC
GCATCTGTGCGGTATTTCACACCGCATATGGTGCACTCTCAGTACAATCTGCTCTGATGCCGCATAGTTAAGCCAGTATACACTCCGCTATC
GCTACGTGACTGGGTCATGGCTGCGCCCCGACACCCGCCAACACCCGCTGACGCGCCCTGACGGGCTTGTCTGCTCCCGGCATCCGCTTA
CAGACAAGCTGTGACCGTCTCCGGGAGCTGCATGTGTCAGAGGTTTTCACCGTCATCACCGAAACGCGCGAGGCAGCAGATCAATTCGC
GCGCGAAGGCGAAGCGGCATGCATAATGTGCCTGTCAAATGGACGAAGCAGGGATTCTGCAAACCCTATGCTACTCCGTCAAGCCGTCA
ATTGTCTGATTCGTTACCAATTATGACAACTTGACGGCTACATCATTCACTTTTTCTTCACAACCGGCACGGAACTCGCTCGGGCTGGCCC
CGGTGCATTTTTTAAATACCCGCGAGAAATAGAGTTGATCGTCAAAACCAACATTGCGACCGACGGTGGCGATAGGCATCCGGGTGGTGC
TCAAAAGCAGCTTCGCCTGGCTGATACGTTGGTCCTCGCGCCAGCTTAAGACGCTAATCCCTAACTGCTGGCGGAAAAGATGTGACAGAC
GCGACGGCGACAAGCAAACATGCTGTGCGACGCTGGCGATATCAAAATTGCTGTCTGCCAGGTGATCGCTGATGTACTGACAAGCCTCGC
GTACCCGATTATCCATCGGTGGATGGAGCGACTCGTTAATCGCTTCCATGCGCCGCAGTAACAATTGCTCAAGCAGATTTATCGCCAGCAG
CTCCGAATAGCGCCCTTCCCCTTGCCCGGCGTTAATGATTTGCCCAAACAGGTCGCTGAAATGCGGCTGGTGCGCTTCATCCGGGCGAAA
GAACCCCGTATTGGCAAATATTGACGGCCAGTTAAGCCATTCATGCCAGTAGGCGCGCGGACGAAAGTAAACCCACTGGTGATACCATTC
GCGAGCCTCCGGATGACGACCGTAGTGATGAATCTCTCCTGGCGGGAACAGCAAAATATCACCCGGTCGGCAAACAAATTCTCGTCCCTG
ATTTTTCACCACCCCCTGACCGCGAATGGTGAGATTGAGAATATAACCTTTCATTCCCAGCGGTCGGTCGATAAAAAAATCGAGATAACCG
TTGGCCTCAATCGGCGTTAAACCCGCCACCAGATGGGCATTAAACGAGTATCCCGGCAGCAGGGGATCATTTTGCGCTTCAGCCATACTTT
TCATACTCCCGCCATTCAGAG

EF-P

MATYYSNDFRAGLKIMLDGEPYAVEASEFVKPGKGQAFARVKLRRLLTGTRVEKTFKSTDSAEGADVVDMNLTYLYNDGEFWHFMNNETFE
QLSADAKAIGDNAKWLLDQAECIVTLWNGQPISVTPPNFVELEIVDTDPGLKGDTAGTGGKPATLSTGAVVKVPLFVQIGEVIKVDTRSGEYVS
RVK*

pBAD-P_{BAD}-*yjgZ'*-PreScission-His_6

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGTTACCTCCTGGACCGCTTCTTGTCCTTCCAGTCGGAGCGAGAGTTAGCCGGGGCGGACTTTCACCGACGGAAAGATCGTGCATTTCA
AGGCACACACAACAATATCTGCGAGCGTGCTATCCGTCCGGTCGTTATGGGACGAAAGGCCTGGTTGTTCGCTGGTTCACTGGTGGCCGG
GAACCGCGCGGCACAGATAATGAGCCTTCTGGGAACCGCAGTCTGGAGCCACATGCTTGGCTGACGGACGTCCTGACGCGTCTGCCGGA
GTGGCCGGAGGAGAGGTTGGCTGAGTTACTGCCTCTTGAGGGCTTTACCTTCTTCGGGCTGGAAGTTCTGTTCCAGGGGCCCAGCAGCC

ATCATCATCATCATCATTGA

YjgZ'-PreScission-His$_6$

MLPPGPLLVLPVGARVSRGGLSPTERSCISRHTQQYLRACYPSGRYGTKGLVVRWFTGGREPRGTDNEPSGNRSLEPHAWLTDVLTRLPEWPE
ERLAELLPLEGFTFFGLEVLFQGPSSHHHHHH*

pBAD-P$_{BAD}$-*prpR*-PreScission-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGGCACATCCACCACGGCTTAATGACGACAAACCGGTTATCTGGACGGTATCTGTAACGCGCCTGTTCGAGCTGTTTCGCGATATCAGCC
TCGAGTTTGATCACCTGGCGAACATTACCCCTATTCAGCTTGGCTTTGAAAAAGCAGTGACCTACATCCGCAAGAAACTGGCAAACGAAC
GCTGTGACGCCATCATCGCGGCTGGCTCTAACGGCGCGTACCTGAAAAGCCGCCTGTCAGTGCCAGTTATTTTGATTAAACCGAGCGGCT
ACGATGTGTTACAGGCACTGGCAAAAGCCGGAAAACTCACCTCTTCTATCGGCGTTGTCACCTATCAGGAAACCATTCCAGCGCTGGTGG
CGTTTCAAAAAACCTTTAATTTGCGCCTCGACCAACGTAGCTACATTACCGAGGAAGACGCACGCGGGCAGATTAACGAGCTAAAAGCTA
ACGGCACCGAAGCGGTGGTCGGCGCAGGGCTGATTACCGATTTGGCAGAAGAAGCCGGAATGACCGGAATTTTTATCTATTCCGCCGCCA
CCGTGCGCCAGGCGTTCAGCGATGCGCTGGATATGACGCGCATGTCGTTACGCCATAACACTCACGATGCCACCCGCAACGCCCTGCGTA
CTCGTTACGTGCTGGGCGATATGCTCGGTCAATCACCACAGATGGAACAAGTACGGCAGACTATTTTGCTGTATGCCCGCTCCAGTGCGGC
GGTGTTGATTGAGGGGGAAACGGGGACGGGCAAAGAGCTGGCGGCCCAGGCGATTCATCGGGAATATTTTGCCCGCCACGATGCGCGAC
AGGGCAAAAAGTCGCATCCGTTTGTTGCCGTCAACTGCGGGGCGATTGCCGAATCGCTGCTGGAAGCAGAACTGTTTGGCTATGAGGAA
GGGGCGTTTACCGGCTCGCGACGCGGAGGTCGCGCCGGGCTGTTCGAAATTGCCCACGGCGGTACGCTGTTTCTGGATGAGATTGGCGA
AATGCCGCTACCTTTGCAGACCCGGCTGCTGCGGGTGCTGGAAGAAAAAGAGGTCACCCGCGTCGGCGGGCATCAGCCTGTTCCGGTAG
ATGTACGGGTCATTAGCGCCACTCACTGCAATCTGGAAGAAGATATGCAGCAAGGACGTTTTCGCCGTGATCTGTTTTATCGGCTGAGTAT
TTTGCGTCTGCAATTGCCACCACTGCGCGAGCGGGTGGCGGATATTCTTCCGCTGGCGGAAAGCTTTTTGAAAGTGTCTCTGGCGGCGCT
CTCCGCCCCATTTTCTGCTGCATTACGCCAGGGGTTACAGGCAAGTGAAACTGTGCTGCTGCACTACGACTGGCCAGGCAATATTCGTGA
ACTGCGCAATATGATGGAACGACTGGCGCTGTTTTTAAGTGTGGAACCGACGCCGGATTTAACGCCGCAGTTTATGCAACTGCTACTGCC
GGAACTGGCGCGCGAGTCGGCGAAAACTCCCGCTCCACGCTTACTGACACCACAACAGGCACTGGAGAAATTTAATGGCGATAAAACAG
CAGCGGCGAATTATTTAGGCATCAGCCGGACGACGTTCTGGCGGCGGCTGAAAAGCCTGGAAGTTCTGTTCCAGGGGCCCAGCAGCCAT
CATCATCATCATCATTGA

PrpR-PreScission-His$_6$

MAHPPRLNDDKPVIWTVSVTRLFELFRDISLEFDHLANITPIQLGFEKAVTYIRKKLANERCDAIIAAGSNGAYLKSRLSVPVILIKPSGYDVLQA
LAKAGKLTSSIGVVTYQETIPALVAFQKTFNLRLDQRSYITEEDARGQINELKANGTEAVVGAGLITDLAEEAGMTGIFIYSAATVRQAFSDALD
MTRMSLRHNTHDATRNALRTRYVLGDMLGQSPQMEQVRQTILLYARSSAAVLIEGETGTGKELAAQAIHREYFARHDARQGKKSHPFVAVNC
GAIAESLLEAELFGYEEGAFTGSRRGGRAGLFEIAHGGTLFLDEIGEMPLPLQTRLLRVLEEKEVTRVGGHQPVPVDVRVISATHCNLEEDMQQ
GRFRRDLFYRLSILRLQLPPLRERVADILPLAESFLKVSLAALSAPFSAALRQGLQASETVLLHYDWPGNIRELRNMMERLALFLSVEPTPDLTP
QFMQLLLPELARESAKTPAPRLLTPQQALEKFNGDKTAAANYLGISRTTFWRRLKSLEVLFQGPSSHHHHHH*

pBAD-P$_{BAD}$-*yhhM*-PreScission-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGAGCAAACCACCTCTTTTCTTTATTGTTATCATTGGCTTAATTGTCGTCGCCGCATCGTTTCGTTTTATGCAACAGCGACGGGAAAAAGC
TGATAATGATATGGCTCCGCTCCAGCAAAAGCTGGTGGTGGTGAGCAACAAGCGGGAAAAACCGATTAACGATCGCCGTTCGCGCCAGC
AGGAAGTGACTCCGGCAGGCACCAGTATACGCTATGAGGCAAGCTTCAAACCGCAAAGCGGAGGAATGGAGCAGACGTTTCGCCTCGA
CGCCCAGCAGTACCACGCCCTGACAGTGGGCGATAAAGGTACGCTGAGCTATAAAGGAACGCGCTTTGTCAGCTTTGTAGGCGAACAAC
TGGAAGTTCTGTTCCAGGGGCCCAGCAGCCATCATCATCATCATCATTGA

YhhM-PreScission-His$_6$

MSKPPLFFIVIIGLIVVAASFRFMQQRREKADNDMAPLQQKLVVVSNKREKPINDRRSRQQEVTPAGTSIRYEASFKPQSGGMEQTFRLDAQQY
HALTVGDKGTLSYKGTRFVSFVGEQLEVLFQGPSSHHHHHH*

pBAD-P$_{BAD}$-*rutD*-PreScission-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGAAACTTTCACTCTCACCTCCCCCTTATGCTGATGCGCCCGTAGTGGTGTTGATTTCGGGTCTTGGGGGTAGCGGCAGTTACTGGTTAC
CGCAACTGGCGGTGCTGGAGCAGGAGTATCAGGTAGTCTGTTACGACCAGCGCGGCACCGGCAATAATCCCGACACGCTGGCAGAAGAT
TACAGTATCGCCCAGATGGCAGCGGAACTGCATCAGGCGCTGGTAGCCGCAGGGATTGAGCATTACGCAGTGGTCGGCCATGCGCTCGGT
GCGCTGGTGGGAATGCAGCTGGCGCTGGATTATCCCGCGTCGGTAACTGTGCTGATCAGCGTTAACGGCTGGCTACGAATAAACGCCCAT
ACGCGCCGCTGTTTTCAGGTTCGCGAACGATTACTGTATAGCGGCGGCGCGCAGGCATGGGTGGAAGCGCAGCCGTTGTTCCTCTATCCC
GCCGACTGGATGGCGGCCCGCGCACCTCGCCTGGAGGCAGAAGACGCGCTGGCACTGGCGCATTTTCAGGGCAAAAATAATTTACTGCG
TCGACTTAACGCCCTCAAACGCGCTGACTTTAGTCACCATGCGGATCGCATCCGCTGCCCGGTGCAAATCATCTGCGCCAGTGATGATCTG
CTGGTGCCAACAGCATGTTCCAGTGAACTTCATGCCGCCCTGCCCGATAGCCAGAAAATGGTGATGCCCTATGGCGGACACGCCTGCAAC

GTGACCGATCCCGAAACGTTTAATGCTCTGTTACTCAACGGGCTTGCCAGCCTGTTACATCACCGTGAAGCCGCCCTGCATCATCATCATC
ATCATTGA

RutD-PreScission-His$_6$

MKLSLSPPPYADAPVVVLISGLGGSGSYWLPQLAVLEQEYQVVCYDQRGTGNNPDTLAEDYSIAQMAAELHQALVAAGIEHYAVVGHALGAL
VGMQLALDYPASVTVLISVNGWLRINAHTRRCFQVRERLLYSGGAQAWVEAQPLFLYPADWMAARAPRLEAEDALALAHFQGKNNLLRRL
NALKRADFSHHADRIRCPVQIICASDDLLVPTACSSELHAALPDSQKMVMPYGGHACNVTDPETFNALLLNGLASLLHHREAALLEVLFQGPS
SHHHHHH*

pBAD-P$_{BAD}$-*ydcO*-PreScission-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGCGTCTGTTTTCTATTCCTCCACCCACGCTACTGGCGGGGTTTCTGGCGGTATTAATTGGCTACGCCAGTTCAGCGGCAATAATCTGGCA
AGCAGCGATTGTCGCCGGAGCCACCACTGCACAAATCTCTGGCTGGATGACGGCGCTGGGGCTGGCAATGGGCGTCAGTACGCTGACTC
TGACATTATGGTATCGCGTACCTGTTCTCACCGCATGGTCAACGCCTGGCGCGGCTTTGTTGGTCACCGGATTGCAGGGACTAACACTTAA
CGAAGCCATCGGCGTTTTTATTGTCACCAACGCGCTAATAGTCCTCTGCGGCATAACGGGACTCTTTGCTCGTCTGATGCGCATTATTCCGC
ACTCGCTTGCGGCGGCAATGCTTGCCGGGATTTTATTACGCTTTGGTTTACAGGCGTTTGCCAGTCTGGACGGTCAATTTACGTTGTGTGG
AAGTATGTTGCTGGTATGGCTGGCAACCAAGGCCGTTGCGCCGCGCTATGCGGTAATTGCCGCGATGATTATTGGGATCGTGATCGTCATC
GCGCAAGGTGACGTTGTCACAACTGATGTTGTCTTTAAACCCGTTCTCCCCACTTATATTACCCCTGATTTTTCGTTTGCTCACAGCCTGAG
CGTTGCACTCCCCCTTTTTCTGGTGACGATGGCATCGCAAAACGCACCGGGTATCGCAGCAATGAAAGCAGCTGGATATTCGGCTCCTGTT
TCGCCATTAATTGTATTTACTGGATTGCTGGCACTGGTTTTTTCCCCTTTCGGCGTTTATTCCGTCGGTATTGCGGCAATCACCGCGGCTATT
TGCCAAAGCCCGGAAGCGCATCCGGATAAAGATCAACGTTGGCTGGCCGCTGCCGTTGCAGGCATTTTCTATTTGCTCGCAGGTCTGTTT
GGTAGTGCCATTACCGGGATGATGGCTGCCCTGCCCGTAAGTTGGATCCAGATGCTGGCAGGTCTGGCGCTGTTAAGTACCATCGGCGGC
AGTTTGTATCAGGCGCTGCATAATGAGCGTGAGCGAGACGCGGCGGTGGTGGCATTTCTGGTAACGGCAAGTGGATTGACGCTGGTCGG
GATTGGTTCTGCGTTTTGGGGATTAATTGCCGGAGGCGTTTGTTACGTGGTGTTGAATTTAATCGCTGACAGAAACCGATATCTGGAAGTT
CTGTTCCAGGGGCCCAGCAGCCATCATCATCATCATCATTGA

YdcO-PreScission-His$_6$

MRLFSIPPPTLLAGFLAVLIGYASSAAIIWQAAIVAGATTAQISGWMTALGLAMGVSTLTLTLWYRVPVLTAWSTPGAALLVTGLQGLTLNEAIG
VFIVTNALIVLCGITGLFARLMRIIPHSLAAAMLAGILLRFGLQAFASLDGQFTLCGSMLLVWLATKAVAPRYAVIAAMIIGIVIVIAQGDVVTTD
VVFKPVLPTYITPDFSFAHSLSVALPLFLVTMASQNAPGIAAMKAAGYSAPVSPLIVFTGLLALVFSPFGVYSVGIAAITAAICQSPEAHPDKDQR
WLAAAVAGIFYLLAGLFGSAITGMMAALPVSWIQMLAGLALLSTIGGSLYQALHNERERDAAVVAFLVTASGLTLVGIGSAFWGLIAGGVCYV
VLNLIADRNRYLEVLFQGPSSHHHHHH*

pBAD-P$_{BAD}$-*yjgZ'*-FLAG-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGTTACCTCCTGGACCGCTTCTTGTCCTTCCAGTCGGAGCGAGAGTTAGCCGGGGCGGACTTTCACCGACGGAAAGATCGTGCATTTCA
AGGCACACACAACAATATCTGCGAGCGTGCTATCCGTCCGGTCGTTATGGGACGAAAGGCCTGGTTGTTCGCTGGTTCACTGGTGGCCGG
GAACCGCGCGGCACAGATAATGAGCCTTCTGGGAACCGCAGTCTGGAGCCACATGCTTGGCTGACGGACGTCCTGACGCGTCTGCCGGA
GTGGCCGGAGGAGAGGTTGGCTGAGTTACTGCCTCTTGAGGGCTTTACCTTCTTCGGGGACTACAAGGACGACGACGACAAGAGCAGCC
ATCATCATCATCATCATTGA

YjgZ'-FLAG-His$_6$

MLPPGPLLVLPVGARVSRGGLSPTERSCISRHTQQYLRACYPSGRYGTKGLVVRWFTGGREPRGTDNEPSGNRSLEPHAWLTDVLTRLPEWPE
ERLAELLPLEGFTFFGDYKDDDDKSSHHHHHH*

pBAD-P$_{BAD}$-*prpR*-FLAG-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGGCACATCCACCACGGCTTAATGACGACAAACCGGTTATCTGGACGGTATCTGTAACGCGCCTGTTCGAGCTGTTTCGCGATATCAGCC
TCGAGTTTGATCACCTGGCGAACATTACCCCTATTCAGCTTGGCTTTGAAAAAGCAGTGACCTACATCCGCAAGAAACTGGCAAACGAAC
GCTGTGACGCCATCATCGCGGCTGGCTCTAACGGCGCGTACCTGAAAAGCCGCCTGTCAGTGCCAGTTATTTTGATTAAACCGAGCGGCT
ACGATGTGTTACAGGCACTGGCAAAAGCCGGAAAACTCACCTCTTCTATCGGCGTTGTCACCTATCAGGAAACCATTCCAGCGCTGGTGG
CGTTTCAAAAAACCTTTAATTTGCGCCTCGACCAACGTAGCTACATTACCGAGGAAGACGCACGCGGGCAGATTAACGAGCTAAAAGCTA
ACGGCACCGAAGCGGTGGTCGGCGCAGGGCTGATTACCGATTTGGCAGAAGAAGCCGGAATGACCGGAATTTTTATCTATTCCGCCGCCA
CCGTGCGCCAGGCGTTCAGCGATGCGCTGGATATGACGCGCATGTCGTTACGCCATAACACTCACGATGCCACCCGCAACGCCCTGCGTA
CTCGTTACGTGCTGGGCGATATGCTCGGTCAATCACCACAGATGGAACAAGTACGGCAGACTATTTTGCTGTATGCCCGCTCCAGTGCGGC
GGTGTTGATTGAGGGGGAAACGGGGACGGGCAAAGAGCTGGCGGCCCAGGCGATTCATCGGGAATATTTTGCCCGCCACGATGCGCGAC
AGGGCAAAAAGTCGCATCCGTTTGTTGCCGTCAACTGCGGGGCGATTGCCGAATCGCTGCTGGAAGCAGAACTGTTTGGCTATGAGGAA
GGGGCGTTTACCGGCTCGCGACGCGGAGGTCGCGCCGGGCTGTTCGAAATTGCCCACGGCGGTACGCTGTTTCTGGATGAGATTGGCGA
AATGCCGCTACCTTTGCAGACCCGGCTGCTGCGGGTGCTGGAAGAAAAAGAGGTCACCCGCGTCGGCGGGCATCAGCCTGTTCCGGTAG

ATGTACGGGTCATTAGCGCCACTCACTGCAATCTGGAAGAAGATATGCAGCAAGGACGTTTTCGCCGTGATCTGTTTTATCGGCTGAGTAT
TTTGCGTCTGCAATTGCCACCACTGCGCGAGCGGGTGGCGGATATTCTTCCGCTGGCGGAAAGCTTTTTGAAAGTGTCTCTGGCGGCGCT
CTCCGCCCCATTTTCTGCTGCATTACGCCAGGGGTTACAGGCAAGTGAAACTGTGCTGCTGCACTACGACTGGCCAGGCAATATTCGTGA
ACTGCGCAATATGATGGAACGACTGGCGCTGTTTTTAAGTGTGGAACCGACGCCGGATTTAACGCCGCAGTTTATGCAACTGCTACTGCC
GGAACTGGCGCGCGAGTCGGCGAAAACTCCCGCTCCACGCTTACTGACACCACAACAGGCACTGGAGAAATTTAATGGCGATAAAACAG
CAGCGGCGAATTATTTAGGCATCAGCCGGACGACGTTCTGGCGGCGGCTGAAAAGCGACTACAAGGACGACGACGACAAGAGCAGCCA
TCATCATCATCATCATTGA

PrpR-FLAG-His$_6$

MAHPPRLNDDKPVIWTVSVTRLFELFRDISLEFDHLANITPIQLGFEKAVTYIRKKLANERCDAIIAAGSNGAYLKSRLSVPVILIKPSGYDVLQA
LAKAGKLTSSIGVVTYQETIPALVAFQKTFNLRLDQRSYITEEDARGQINELKANGTEAVVGAGLITDLAEEAGMTGIFIYSAATVRQAFSDALD
MTRMSLRHNTHDATRNALRTRYVLGDMLGQSPQMEQVRQTILLYARSSAAVLIEGETGTGKELAAQAIHREYFARHDARQGKKSHPFVAVNC
GAIAESLLEAELFGYEEGAFTGSRRGGRAGLFEIAHGGTLFLDEIGEMPLPLQTRLLRVLEEKEVTRVGGHQPVPVDVRVISATHCNLEEDMQQ
GRFRRDLFYRLSILRLQLPPLRERVADILPLAESFLKVSLAALSAPFSAALRQGLQASETVLLHYDWPGNIRELRNMMERLALFLSVEPTPDLTP
QFMQLLLPELARESAKTPAPRLLTPQQALEKFNGDKTAAANYLGISRTTFWRRLKSDYKDDDDKSSHHHHHH*

pBAD-P$_{BAD}$-*yhhM*-FLAG-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGAGCAAACCACCTCTTTTCTTTATTGTTATCATTGGCTTAATTGTCGTCGCCGCATCGTTTCGTTTTATGCAACAGCGACGGGAAAAAGC
TGATAATGATATGGCTCCGCTCCAGCAAAAGCTGGTGGTGGTGAGCAACAAGCGGGAAAAACCGATTAACGATCGCCGTTCGCGCCAGC
AGGAAGTGACTCCGGCAGGCACCAGTATACGCTATGAGGCAAGCTTCAAACCGCAAAGCGGAGGAATGGAGCAGACGTTTCGCCTCGA
CGCCCAGCAGTACCACGCCCTGACAGTGGGCGATAAAGGTACGCTGAGCTATAAAGGAACGCGCTTTGTCAGCTTTGTAGGCGAACAAG
ACTACAAGGACGACGACGACAAGAGCAGCCATCATCATCATCATTGA

YhhM-FLAG-His$_6$

MSKPPLFFIVIIGLIVVAASFRFMQQRREKADNDMAPLQQKLVVVSNKREKPINDRRSRQQEVTPAGTSIRYEASFKPQSGGMEQTFRLDAQQY
HALTVGDKGTLSYKGTRFVSFVGEQDYKDDDDKSSHHHHHH*

pBAD-P$_{BAD}$-*rutD*-FLAG-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGAAACTTTCACTCTCACCTCCCCCTTATGCTGATGCGCCCGTAGTGGTGTTGATTTCGGGTCTTGGGGGTAGCGGCAGTTACTGGTTAC
CGCAACTGGCGGTGCTGGAGCAGGAGTATCAGGTAGTCTGTTACGACCAGCGCGGCACCGGCAATAATCCCGACACGCTGGCAGAAGAT
TACAGTATCGCCCAGATGGCAGCGGAACTGCATCAGGCGCTGGTAGCCGCAGGGATTGAGCATTACGCAGTGGTCGGCCATGCGCTCGGT
GCGCTGGTGGGAATGCAGCTGGCGCTGGATTATCCCGCGTCGGTAACTGTGCTGATCAGCGTTAACGGCTGGCTACGAATAAACGCCCAT
ACGCGCCGCTGTTTTCAGGTTCGCGAACGATTACTGTATAGCGGCGGCGCGCAGGCATGGGTGGAAGCGCAGCCGTTGTTCCTCTATCCC
GCCGACTGGATGGCGGCCCGCGCACCTCGCCTGGAGGCAGAAGACGCGCTGGCACTGGCGCATTTTCAGGGCAAAAATAATTTACTGCG
TCGACTTAACGCCCTCAAACGCGCTGACTTTAGTCACCATGCGGATCGCATCCGCTGCCCGGTGCAAATCATCTGCGCCAGTGATGATCTG
CTGGTGCCAACAGCATGTTCCAGTGAACTTCATGCCGCCCTGCCCGATAGCCAGAAAATGGTGATGCCCTATGGCGGACACGCCTGCAAC
GTGACCGATCCCGAAACGTTTAATGCTCTGTTACTCAACGGGCTTGCCAGCCTGTTACATCACCGTGAAGCCGCCCTGGACTACAAGGAC
GACGACGACAAGAGCAGCCATCATCATCATCATTGA

RutD-FLAG-His$_6$

MKLSLSPPPYADAPVVVLISGLGGSGSYWLPQLAVLEQEYQVVCYDQRGTGNNPDTLAEDYSIAQMAAELHQALVAAGIEHYAVVGHALGAL
VGMQLALDYPASVTVLISVNGWLRINAHTRRCFQVRERLLYSGGAQAWVEAQPLFLYPADWMAARAPRLEAEDALALAHFQGKNNLLRRL
NALKRADFSHHADRIRCPVQIICASDDLLVPTACSSELHAALPDSQKMVMPYGGHACNVTDPETFNALLLNGLASLLHHREAALDYKDDDDK
SSHHHHHH*

pBAD-P$_{BAD}$-*ydcO*-FLAG-His$_6$

The ORF of pBAD-HisA was replaced by Following DNA sequence.

ATGCGTCTGTTTTCTATTCCTCCACCCACGCTACTGGCGGGGTTTCTGGCGGTATTAATTGGCTACGCCAGTTCAGCGGCAATAATCTGGCA
AGCAGCGATTGTCGCCGGAGCCACCACTGCACAAATCTCTGGCTGGATGACGGCGCTGGGGCTGGCAATGGGCGTCAGTACGCTGACTC
TGACATTATGGTATCGCGTACCTGTTCTCACCGCATGGTCAACGCCTGGCGCGGCTTTGTTGGTCACCGGATTGCAGGGACTAACACTTAA
CGAAGCCATCGGCGTTTTTATTGTCACCAACGCGCTAATAGTCCTCTGCGGCATAACGGGACTCTTTGCTCGTCTGATGCGCATTATTCCGC
ACTCGCTTGCGGCGGCAATGCTTGCCGGGATTTTATTACGCTTTGGTTTACAGGCGTTTGCCAGTCTGGACGGTCAATTTACGTTGTGTGG
AAGTATGTTGCTGGTATGGCTGGCAACCAAGGCCGTTGCGCCGCGCTATGCGGTAATTGCCGCGATGATTATTGGGATCGTGATCGTCATC
GCGCAAGGTGACGTTGTCACAACTGATGTTGTCTTTAAACCCGTTCTCCCCACTTATATTACCCCTGATTTTTCGTTTGCTCACAGCCTGAG
CGTTGCACTCCCCCTTTTTCTGGTGACGATGGCATCGCAAAACGCACCGGGTATCGCAGCAATGAAAGCAGCTGGATATTCGGCTCCTGTT
TCGCCATTAATTGTATTTACTGGATTGCTGGCACTGGTTTTTTCCCCTTTCGGCGTTTATTCCGTCGGTATTGCGGCAATCACCGCGGCTATT
TGCCAAAGCCCGGAAGCGCATCCGGATAAAGATCAACGTTGGCTGGCCGCTGCCGTTGCAGGCATTTTCTATTTGCTCGCAGGTCTGTTT

GGTAGTGCCATTACCGGGATGATGGCTGCCCTGCCCGTAAGTTGGATCCAGATGCTGGCAGGTCTGGCGCTGTTAAGTACCATCGGCGGC
AGTTTGTATCAGGCGCTGCATAATGAGCGTGAGCGAGACGCGGCGGTGGTGGCATTTCTGGTAACGGCAAGTGGATTGACGCTGGTCGG
GATTGGTTCTGCGTTTTGGGGATTAATTGCCGGAGGCGTTTGTTACGTGGTGTTGAATTTAATCGCTGACAGAAACCGATATGACTACAAG
GACGACGACGACAAGAGCAGCCATCATCATCATCATTGA

YdcO-FLAG-His$_6$

MRLFSIPPPTLLAGFLAVLIGYASSAAIIWQAAIVAGATTAQISGWMTALGLAMGVSTLTLTLWYRVPVLTAWSTPGAALLVTGLQGLTLNEAIG
VFIVTNALIVLCGITGLFARLMRIIPHSLAAAMLAGILLRFGLQAFASLDGQFTLCGSMLLVWLATKAVAPRYAVIAAMIIGIVIVIAQGDVVTTD
VVFKPVLPTYITPDFSFAHSLSVALPLFLVTMASQNAPGIAAMKAAGYSAPVSPLIVFTGLLALVFSPFGVYSVGIAAITAAICQSPEAHPDKDQR
WLAAAVAGIFYLLAGLFGSAITGMMAALPVSWIQMLAGLALLSTIGGSLYQALHNERERDAAVVAFLVTASGLTLVGIGSAFWGLIAGGVCYV
VLNLIADRNRYDYKDDDDKSSHHHHHH*

Supplementary table 11 Combinations of DNA templates and DNA primers for plasmid preparation.

Plasmid name	DNA template for adopter PCR	Forward primer	Reverse primer	DNA template for inversion PCR	Forward primer	Reverse primer
pBAD-PBAD-yjgZ	DNA of yjgZ ORF from IDT	Primer92.F40	Primer93.R61	pBAD-HisA	Primer94.R44	Primer95.F44
pBAD-PBAD-yjgZ'	DNA of yjgZ ORF from IDT	Primer96.F40	Primer93.R61	pBAD-HisA	Primer94.R44	Primer95.F44
pBAD-PBAD-prpR	DNA of prpR ORF from IDT	Primer97.F40	Primer98.R58	pBAD-HisA	Primer94.R44	Primer95.F44
pBAD-PBAD-yhhM	DNA of yhhM ORF from IDT	Primer99.F40	Primer100.R61	pBAD-HisA	Primer94.R44	Primer95.F44
pBAD-PBAD-rutD	DNA of rutD ORF from IDT	Primer101.F40	Primer102.R57	pBAD-HisA	Primer94.R44	Primer95.F44
pBAD-PBAD-ydcO	DNA of ydcO ORF from IDT	Primer103.F40	Primer104.R61	pBAD-HisA	Primer94.R44	Primer95.F44
pBAD-PBAD-efp-AmpR	pBAD-HisA	Primer105.F35	Primer106.R34	pET28a(+)-efp	Primer107.F44	Primer108.R43
pBAD-PBAD-efp-CmR-p15A	pBAD-PBAD-efp-AmpR	Primer109.F42	Primer110.R39	pLysS	Primer111.F31	Primer112.R48
pET28a(+)-yjgZ	DNA of yjgZ ORF from IDT	Primer113.F40	Primer114.R58	pET28a(+)-efp	Primer115.F30	Primer116.R40
pET28a(+)-yjgZ'	DNA of yjgZ ORF from IDT	Primer117.F40	Primer114.R58	pET28a(+)-efp	Primer115.F30	Primer116.R40
pET28a(+)-prpR	DNA of prpR ORF from IDT	Primer118.F40	Primer119.R55	pET28a(+)-efp	Primer115.F30	Primer116.R40
pET28a(+)-yhhM	DNA of yhhM ORF from IDT	Primer120.F40	Primer121.R58	pET28a(+)-efp	Primer115.F30	Primer116.R40
pET28a(+)-rutD	DNA of rutD ORF from IDT	Primer122.F40	Primer123.R54	pET28a(+)-efp	Primer115.F30	Primer116.R40
pET28a(+)-ydcO	DNA of ydcO ORF from IDT	Primer124.F40	Primer125.R58	pET28a(+)-efp	Primer115.F30	Primer116.R40
pBAD-PBAD-yjgZ'-PreScission-His6	pBAD-PBAD-yjgZ'	Primer96.F40	Primer126.R48	pBAD-PBAD-yjgZ'	Primer127.F67	Primer106.R34
pBAD-PBAD-prpR-PreScission-His6	pBAD-PBAD-prpR	Primer97.F40	Primer128.R43	pBAD-PBAD-prpR	Primer127.F67	Primer106.R34
pBAD-PBAD-yhhM-PreScission-His6	pBAD-PBAD-yhhM	Primer99.F40	Primer129.R48	pBAD-PBAD-yhhM	Primer127.F67	Primer106.R34
pBAD-PBAD-rutD-PreScission-His6	pBAD-PBAD-rutD	Primer101.F40	Primer130.R42	pBAD-PBAD-rutD	Primer127.F67	Primer106.R34
pBAD-PBAD-ydcO-PreScission-His6	pBAD-PBAD-ydcO	Primer103.F40	Primer131.R57	pBAD-PBAD-ydcO	Primer127.F67	Primer106.R34
pBAD-PBAD-yjgZ-FLAG-His6	pBAD-PBAD-yjgZ'	Primer96.F40	Primer132.R48	pBAD-PBAD-yjgZ'	Primer133.F63	Primer106.R34
pBAD-PBAD-yjgZ'-FLAG-His6	pBAD-PBAD-prpR	Primer97.F40	Primer134.R43	pBAD-PBAD-prpR	Primer133.F63	Primer106.R34
pBAD-PBAD-prpR-FLAG-His6	pBAD-PBAD-yhhM	Primer99.F40	Primer135.R48	pBAD-PBAD-yhhM	Primer133.F63	Primer106.R34
pBAD-PBAD-yhhM-FLAG-His6	pBAD-PBAD-rutD	Primer101.F40	Primer136.R48	pBAD-PBAD-rutD	Primer133.F63	Primer106.R34
pBAD-PBAD-rutD-FLAG-His6	pBAD-PBAD-ydcO	Primer103.F40	Primer137.R57	pBAD-PBAD-ydcO	Primer133.F63	Primer106.R34
pUC18-yjgZ	DNA of yjgZ ORF from IDT	Primer138.F40	Primer139.R61	pUC18	Primer140.R25	Primer141.F44
pUC18-yjgZ'	DNA of yjgZ ORF from IDT	Primer142.F51	Primer143.R61	pSF-TAC	Primer144.R44	Primer145.F44
pUC18-prpR	DNA of yjgZ' ORF from IDT	Primer146.F40	Primer139.R61	pUC18	Primer140.R25	Primer141.F44
pUC18-yhhM	DNA of yjgZ ORF from IDT	Primer147.F51	Primer143.R61	pSF-TAC	Primer144.R44	Primer145.F44
pUC18-rutD	DNA of prpR ORF from IDT	Primer148.F40	Primer149.R58	pUC18	Primer140.R25	Primer141.F44
pUC18-ydcO	DNA of prpR ORF from IDT	Primer150.F51	Primer151.R58	pSF-TAC	Primer144.R44	Primer145.F44
pSF-TAC-yjgZ	DNA of yhhM ORF from IDT	Primer152.F40	Primer153.R61	pUC18	Primer140.R25	Primer141.F44
pSF-TAC-yjgZ'	DNA of yhhM ORF from IDT	Primer154.F51	Primer155.R61	pSF-TAC	Primer144.R44	Primer145.F44
pSF-TAC-prpR	DNA of rutD ORF from IDT	Primer156.F40	Primer157.R57	pUC18	Primer140.R25	Primer141.F44
pSF-TAC-yhhM	DNA of rutD ORF from IDT	Primer158.F51	Primer159.R57	pSF-TAC	Primer144.R44	Primer145.F44
pSF-TAC-rutD	DNA of ydcO ORF from IDT	Primer160.F40	Primer161.R61	pUC18	Primer140.R25	Primer141.F44
pSF-TAC-ydcO	DNA of ydcO ORF from IDT	Primer162.F51	Primer163.R61	pSF-TAC	Primer144.R44	Primer145.F44

Supplementary table 12 Primary sequence of proteases

Glu-C

VILPNNDRHQITDTTNGHYAPVTYIQVEAPTGTFIASGVVVGKDTLLTNKHVVDATHGDPHALKAFPSAINQDNYPNGGFTAEQITKYSGEGD
LAIVKFSPNEQNKHIGEVVKPATMSNNAETQVNQNITVTGYPGDKPVATMWESKGKITYLKGEAMQYDLSTTGGNSGSPVFNEKNEVIGIHW
GGVPNEFNGAVFINENVRNFLKQNIEDIHFANDDQPNNPDNPDNPNNPDNPNNPDEPNNPDNPNNPDNPDNGDNNNSDNPDAA

Lys-N

TYNGCSSSEQSALAAAASAAQSYVAESLSYLQTHTAATPRYTTWFGSYISSRHSTVLQHYTDMNSNDFSSYSFDCTCTAAGTFAYVYPNRFGT
VYLCGAFWKAPTTGTDSQAGTLVHESSHFTRNGGTKDYAYGQAAAKSLATMDPDKAVMNADNHEYFSENNPAQS

Lys-C

GVSGSCNIDVVCPEGNGHRDVIRSVAAYSKQGTMWCTGSLVNNSANDKKMYFLTANHCGMTTAAIASSMVVYWNYQNSTCRAPGSSSSGA
NGDGSLAQSQTGAVVRATNAASDFTLLELNTAANPAYNLFWAGWDRRDQNFAGATAIHHPNVAEKRISHSTVATEISGYNGATGTSHLHVFWQ
ASGGVTEPGSSGSPIYSPEKRVLGQLHGGPSSCSATGADRSDYYGRVFTSWTGGGTSATRLSDWLDAAGTGAQFIDGLDSTGTPPV

Asp-N

ESNQGYVNSNVGIELARYETTNYTESGSFDTDLARFRGTSDSIHTSRNTYTAADCATGYYSFAHEIGHLQSARDIATDSSTSPYAYGHGYRYEPA
TGWRTIMAYNCTRSCPRLNYWSNPNISYDIGPDNQRVLVNTKATIAAFR

Arg-C

AICWDDSNLDKNGEADCLYMGEISDHLTEKQSVDLLAFDACLMGTAEVAYQYRPGNGGFSADTLVASSPVVWGPGFKYDKIFDRIKAGGGTN
NEDDLTLGGKEQNFDPATITNEQLGALFVEEQRDSTHANGRYDQHLSFYDLKKAESVKRAIDNLAVNLSNENKKSEIEKLRGSGIHTDLMHYF
DEYSEGEWVEYPYFDVYDLCEKINKSENFSSKTKDLASNAMNKLNEMIVYSFGDPSNNFKEGKNGLSIFLPNGDKKYSTYYTSTKIPHWTMQ
SWYNSIDTVKYGLNPYGKLSWCKDGQDPEINKVGNWFELLDSWFDKTNDVTGGVNHYQW

Turbo3C

AFRPCNVNTKIGNAKCCPFVCGKAVTFKDRSTCSTYNLSSSLHHILEEDKRRRQVVDVMSAIFQGPISLDAPPPPAIADLLQSVRTPRVIKYCQII
MGHPAECQVERDLNIANSIIAIIANIISIAGIIFVIYKLFCSLQGPYSGEPKPKTKVPERRVVAQGPEEEFGRSILKNNTCVITTGNGKFTGLGIHDRI
LIIPTHADPGREVQVNGVHTKVLDSYDLYNRDGVKLEITVIQLDRNEKFRDIRKYIPETEDDYPECNLALSANQDEPTIIKVGDVVSYGNILLSG
NQTARMLKYNYPTKSGYCGGVLYKIGQILGIHVGGNGRDGFSAMLLRSYFTGQIKVNKHATECGLPDIQTIHTPSKTKLQPSVFYDVFPGSKEP
AVLTDNDPRLEVNFKEA

Chapter4
General Conclusion

Peptidyl-tRNA drop-off is one of the abnormal events in ribosomal protein synthesis. Causes of peptidyl-tRNA drop-off have been attributed to mRNA contexts[28,29], translation factors[30,31], and macrolide antibiotics[32] but the actual mechanism has remained elusive. Ribosome stalling is plausible cause of peptidyl-tRNA drop-off and caused by mRNA contexts[44], certain nascent peptide sequences[54–60], and consecutive prolines in nascent peptides[48,50,51]. However, peptidyl-tRNA drop-off and ribosome stalling have been independently described. In my master course study, I revealed that peptidyl-tRNA dropped off from the ribosome stalled by consecutive prolines in the nascent peptide by EF-G-mediated translocation, which was regulated by the nature of the nascent peptide sequences. Furthermore, following translation restart from the middle of the mRNA generated C-terminal peptides. In my PhD study, I further investigated the peptidyl-tRNA drop-of induced by nascent peptide sequences by (1) developing of profiling system of nascent peptide sequences against peptidyl-tRNA drop-off based on mRNA display methodology and (2) investigation of C-terminal peptide generation due to peptidyl-tRNA drop-off induced by ribosome stalling in the *E .coli* endogenous protein synthesis system.

In the study described in Chapter 2, the novel methodology was developed to comprehensively profile the nascent peptide sequences that regulate the peptidyl-tRNA drop-off. Combination of the mRNA display method and genetic code reprogramming technique enabled the selective recovery of nascent peptide sequences that allow incorporation of consecutive prolines without peptidyl-tRNA drop-off. The profiling system analyzed 8420 kinds of nascent peptide sequences with various EF-G concentration and in the absence and presence of EF-P, in parallel, which achieved the broadest sequence coverage in my best knowledge. Sequencing results of recovered cDNA revealed the nature of amino acid side chain just prior to the three consecutive proline regulate the frequency of peptidyl-tRNA drop-off, in which polar amino acids induced peptidyl-tRNA drop-off. The frequency of peptidyl-tRNA drop-off was more enhanced in the nascent peptides having polar residues just prior to the consecutive prolines than that having other residues in dose-dependent manner of EF-G. Because the exit tunnel consists of hydrophilic rRNA and ribosomal proteins, electrostatic attraction or repulsion between the exit tunnel and polar nascent peptides would modulate the conformation of the PTC allosterically by propagating through the nascent peptides. In the case of the short peptidyl-tRNA, the allosteric modulation and EF-G might allow the inter-subunit rotation to form hybrid state with p/E peptidyl-tRNA and a/P aa-tRNA, from which the peptidyl-tRNA dropped off.

In the study described in Chapter 3, C-terminal peptide generation after peptidyl-tRNA drop-off was investigated in the *E. coli*'s endogenous protein expression system As a result of expression of *E. coli* proteins and putative proteins containing two or three consecutive prolines in *E. coli* Δ*efp* or *E. coli* Δ*efp*/+ *efp* complemented by a plasmid, the full-length YhhM and YhhM lacking amino acids from N-terminal Met to two consecutive prolines were identified. Although full-length YhhM was dominantly expressed over YhhM lacking its N-terminal amino acids, the expression level of the YhhM lacking its N-terminal amino acids was enhanced in expression in *E. coli* Δ*efp* compared to the that of *E. coli* Δ*efp*/+*efp*. These results demonstrated that the generation of C-terminal peptides after peptidyl-tRNA drop-off was the abnormal protein synthesis in *E. coli* for the first time and suggested that EF-P could possess a function to suppress the drop-off of peptidyl-prolyl-tRNA in addition to the promotion of the consecutive incorporation of prolines. Detection of YhhM lacking its N-terminal amino acids even in the presence of EF-P suggested that protein synthesis system might face to drop-off of short nascent peptidyl-tRNAs from the ribosome, which could continue to generate the remaining C-terminal proteins. Because known bacterial quality

control systems cannot rescue the stalled 70S ribosome complex with A-site aa-tRNA, the unknown quality control system is worth to be investigated if it exists.

In my PhD thesis, I revealed the determinant of the peptidyl-tRNA drop-off by developing the profiling system of nascent peptide sequences on regulation of peptidyl-tRNA drop-off and demonstrated ribosomal generation of a protein lacking its N-terminal amino acids due to the drop-off in *in vivo* protein synthesis system. From the viewpoint of basic research, it is proposed that a protein not corresponding to that of encoded in ORF could be generated due to translation restart after the drop-off and possesses the different function and property from its full-length protein. If the translation machinery fails to synthesize the full-length proteins, the protein quality would be controlled by N-end rule which selectively degrades C-terminal proteins which may have N-terminal amino acids except fMet. From the viewpoint of application, optimization of nascent peptide sequences can suppress the drop-off, which would improve the yield of the full-length proteins *in vivo* and incorporation efficiency of consecutive prolines and other non-proteinogenic amino acids *in vitro*. The regulation of the drop-off by polarity of nascent peptide sequences and *in vivo* generation of the C-terminal protein with N-terminal truncation investigated in my thesis provides the insight into abnormal protein synthesis mechanisms system and is expected to contribute to understand the protein synthesis system and protein quality control.

References

1. Menninger, J. R. The accumulation as peptidyl-transfer RNA of isoaccepting transfer RNA families in Escherichia coli with Temparature-sensitive Peptidyl-transfer RNA Hydrplase. *Jbc* **253**, (1978).
2. Woolstenhulme, C. J., Guydosh, N. R., Green, R. & Buskirk, A. R. High-Precision analysis of translational pausing by ribosome profiling in bacteria lacking EFP. *Cell Rep.* **11**, 13–21 (2015).
3. Ito, K. & Chiba, S. Arrest Peptides: Cis -Acting Modulators of Translation. *Annu. Rev. Biochem.* **82**, 171–202 (2013).
4. Schmeing, T. M. & Ramakrishnan, V. What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242 (2009).
5. Gualerzi, C. O. & Pon, C. L. Initiation of mRNA translation in bacteria: Structural and dynamic aspects. *Cellular and Molecular Life Sciences* **72**, 4341–4367 (2015).
6. Olins, P. O. & Rangwala, S. H. A novel sequence element derived from bacteriophage T7 mRNA acts as an enhancer of translation of the lacZ gene in Escherichia coli. *J. Biol. Chem.* **264**, 16973–16976 (1989).
7. Shine, J. & Dalgarno, L. The 3'-Terminal Sequence of Escherichia coli 16S Ribosomal RNA: Complementarity to Nonsense Triplets and Ribosome Binding Sites. *Proc. Natl. Acad. Sci.* **71**, 1342–1346 (1974).
8. Sengupta, J., Agrawal, R. K. & Frank, J. Visualization of protein S1 within the 30S ribosomal subunit and its interaction with messenger RNA. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 11991–11996 (2001).
9. Kaminishi, T. *et al.* A Snapshot of the 30S Ribosomal Subunit Capturing mRNA via the Shine-Dalgarno Interaction. *Structure* **15**, 289–297 (2007).
10. Rodnina, M. V & Wintermeyer, W. GTP consumption of elongation factor Tu during translation of heteropolymeric mRNAs. *Proc. Natl. Acad. Sci.* **92**, 1945–1949 (1995).
11. Howe, D. K. *et al.* Uniform Binding of Aminoacyl-tRNAs to Elongation Factor Tu by Thermodynamic Compensation. *Science* **294**, 165–169 (2001).
12. Sanderson, L. E. & Uhlenbeck, O. C. Exploring the specificity of bacterial elongation factor Tu for different tRNAs. *Biochemistry* **46**, 6194–6200 (2007).
13. Nissen, P. *et al.* The Structural Basis of Ribosome Activity in Peptide Bond Synthesis. *Science* **289**, 920–931 (2000).
14. Rodnina, M. V., Savelsbergh, A., Katunin, V. I. & Wintermeyer, W. Hydrolysis of GTP by elongation factor G drives tRNA movement on the ribosome. *Nature* (1997).
15. Tourigny, D. S., Fernández, I. S., Kelley, A. C. & Ramakrishnan, V. Elongation factor G bound to the ribosome in an intermediate state of translocation. *Science* **340**, (2013).
16. Moazed, D. & Noller, H. F. Intermediate states in the movement of transfer RNA in the ribosome. *Nature* **342**, 142–148 (1989).
17. Selmer, M. *et al.* Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
18. Dunkle, J. A. *et al.* Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981–984 (2011).
19. Chen, C. *et al.* Single-Molecule Fluorescence Measurements of Ribosomal Translocation Dynamics. *Mol. Cell*

42, 367–377 (2011).

20. Lin, J., Gagnon, M. G., Bulkley, D. & Steitz, T. A. Conformational Changes of Elongation Factor G on the Ribosome during tRNA Translocation. *Cell* **160**, 219–227 (2015).
21. Zhou, J., Lancaster, L., Donohue, J. P. & Noller, H. F. Crystal structures of EF-G - Ribosome complexes trapped in intermediate states of translocation. *Science* **340**, (2013).
22. Gao, Y. G. *et al.* The structure of the ribosome with elongation factor G trapped in the posttranslocational state. *Science* **326**, 694–699 (2009).
23. Cornish, P. V, Ermolenko, D. N., Noller, H. F. & Ha, T. Spontaneous Intersubunit Rotation in Single Ribosomes. *Mol. Cell* **30**, 578–588 (2008).
24. Nakamura, Y., Ito, K. & Uno, M. A tripeptide 'anticodon' deciphers stop codons in messenger RNA. *Nature* **403**, 680–684 (2000).
25. Freistroffer, D. V., Pavlov, M. Y., MacDougall, J., Buckingham, R. H. & Ehrenberg, M. Release factor RF3 in E. coli accelerates the dissociation of release factors RF1 and RF2 from the ribosome in a GTP-dependent manner. *EMBO J.* **16**, 4126–4133 (1997).
26. Pavlov, M. Y., Freistroffer, D. V., MacDougall, J., Buckingham, R. H. & Ehrenberg, M. Fast recycling of Escherichia coli ribosomes requires both ribosome recycling factor (RRF) and release factor RF3. *EMBO J.* **16**, 4134–4141 (1997).
27. Menninger, J. R. Peptidyl Transfer RNA Dissociates during Protein from Ribosomes of Escherichia coli. *J. Biol. Chem.* **251**, 3392–3398 (1976).
28. Ontiveros, C., Gerardo Valadez, J., Hernández, J. & Guarneros, G. Inhibition of Escherichia coli protein synthesis by abortive translation of phage λ minigenes. *J. Mol. Biol.* **269**, 167–175 (1997).
29. Cruz-Vera, L. R., Magos-Castro, M. A., Zamora-Romo, E. & Guarneros, G. Ribosome stalling and peptidyl-tRNA drop-off during translational delay at AGA codons. *Nucleic Acids Res.* **32**, 4462–4468 (2004).
30. Karimi, R., Pavlov, M. Y., Heurgué-Hamard, V., Buckingham, R. H. & Ehrenberg, M. Initiation factors IF1 and IF2 synergistically remove peptidyl-tRNAs with short polypeptides from the P-site of translating Escherichia coli ribosomes. *J. Mol. Biol.* **281**, 241–252 (1998).
31. Heurgué-Hamard, V. *et al.* Ribosome release factor RF4 and termination factor RF3 are involved in dissociation of peptidyl-tRNA from the ribosome. *EMBO J.* **17**, 808–816 (1998).
32. Tenson, T., Lovmar, M. & Ehrenberg, M. The mechanism of action of macrolides, lincosamides and streptogramin B reveals the nascent peptide exit path in the ribosome. *J. Mol. Biol.* **330**, 1005–14 (2003).
33. Menninger, J. R. Accumulation of peptidyl tRNA is lethal to Escherichia coli. *J. Bacteriol.* **137**, 694–696 (1979).
34. Vogel, Z., Vogel, T., Zamir, A. & Elson, D. The Protection by 70S Ribosomes of N-Acyl-aminoacyl-tRNA against Cleavage by Peptidyl-tRNA Hydrolase and its Use to Assay Ribosomal Association. *Eur. J. Biochem.* **21**, 582–592 (1971).
35. Das, G. & Varshney, U. Peptidyl-tRNA hydrolase and its critical role in protein biosynthesis. *Microbiology* **152**, 2191–2195 (2006).
36. Atherly, A. G. & Menninger, J. R. Mutant E. coli strain with temperature sensitive peptidyltransfer RNA hydrolase. *Nat. New Biol.* **240**, 245–246 (1972).

37. Tenson, T., Vega Herrera, J., Kloss, P., Guarneros, G. & Mankin, A. S. Inhibition of translation and cell growth by minigene expression. *J. Bacteriol.* **181**, 1617–1622 (1999).
38. Schlünzen, F. *et al.* Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814–821 (2001).
39. Hansen, J. L. *et al.* The structures of four macrolide antibiotics bound to the large ribosomal subunit. *Mol. Cell* **10**, 117–128 (2002).
40. Kannan, K. *et al.* The general mode of translation inhibition by macrolide antibiotics. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15958–15963 (2014).
41. Otaka, T. & Kaji, A. Release of (oligo)peptidyl tRNA from ribosomes by erythromycin A. *Proc. Natl. Acad. Sci. U. S. A.* **72**, 2649–2652 (1975).
42. Dunkle, J. A., Xiong, L., Mankin, A. S. & Cate, J. H. D. Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17152–17157 (2010).
43. Lovmar, M., Tenson, T. & Ehrenberg, M. Kinetics of Macrolide Action. *J Biol Chem* **279**, 53506–53515 (2004).
44. Li, G. W., Oh, E. & Weissman, J. S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538–541 (2012).
45. Ikemura, T. Correlation between the abundance of Escherichia coli transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the E. coli translational system. *J. Mol. Biol.* **151**, 389–409 (1981).
46. Brierley, I., Digard, P. & Inglis, S. C. Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57**, 537–547 (1989).
47. Namy, O., Moran, S. J., Stuart, D. I., Gilbert, R. J. C. & Brierley, I. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244–247 (2006).
48. Huter, P. *et al.* Structural Basis for Polyproline-Mediated Ribosome Stalling and Rescue by the Translation Elongation Factor EF-P. *Mol. Cell* **68**, 515-527.e6 (2017).
49. Katoh, T., Wohlgemuth, I., Nagano, M., Rodnina, M. V & Suga, H. Essential structural elements in tRNAPro for EF-P-mediated alleviation of translation stalling. *Nat. Commun.* **7**, 1–12 (2016).
50. Doerfel, L. K. *et al.* EF-P is essential for rapid synthesis of proteins containing consecutive proline residues. *Science* **339**, 85–88 (2013).
51. Ude, S. *et al.* Translation elongation factor EF-P alleviates ribosome stalling at polyproline stretches. *Science* **339**, 82–85 (2013).
52. Peil, L. *et al.* Distinct XPPX sequence motifs induce ribosome stalling, which is rescued by the translation elongation factor EF-P. *Proc. Natl. Acad. Sci.* **110**, 15265–15270 (2013).
53. Starosta, A. L. *et al.* Translational stalling at polyproline stretches is modulated by the sequence context upstream of the stall site. *Nucleic Acids Res.* **42**, 10711–10719 (2014).
54. Nakatogawa, H. & Ito, K. Secretion monitor, secM, undergoes self-translation arrest in the cytosol. *Mol. Cell* **7**, 185–192 (2001).
55. Seidelt, B. *et al.* Structural Insight into Nascent Polypeptide Chain–Mediated Translational Stalling. *Science*

1621, 1412–1415 (2009).

56. Vazquez-Laslop, N., Thum, C. & Mankin, A. S. Molecular Mechanism of Drug-Dependent Ribosome Stalling. *Mol. Cell* **30**, 190–202 (2008).
57. Chiba, S. & Ito, K. Multisite Ribosomal Stalling: A Unique Mode of Regulatory Nascent Chain Action Revealed for MifM. *Mol. Cell* **47**, 863–872 (2012).
58. Woolstenhulme, C. J. *et al.* Nascent peptides that block protein synthesis in bacteria. *Proc. Natl. Acad. Sci.* **110**, E878–E887 (2013).
59. Chadani, Y., Niwa, T., Chiba, S., Taguchi, H. & Ito, K. Integrated in vivo and in vitro nascent chain profiling reveals widespread translational pausing. *Proc. Natl. Acad. Sci.* **113**, E829–E838 (2016).
60. Chadani, Y. *et al.* Intrinsic Ribosome Destabilization Underlies Translation and Provides an Organism with a Strategy of Environmental Sensing. *Mol. Cell* **68**, 528-539.e5 (2017).
61. Ramu, H. *et al.* Nascent Peptide in the Ribosome Exit Tunnel Affects Functional Properties of the A-Site of the Peptidyl Transferase Center. *Mol. Cell* **41**, 321–330 (2011).
62. Muto, H. & Ito, K. Peptidyl-prolyl-tRNA at the ribosomal P-site reacts poorly with puromycin. *Biochem. Biophys. Res. Commun.* **366**, 1043–1047 (2008).
63. Luo, Z. & Sachs, M. S. Role of an upstream open reading frame in mediating arginine-specific translational control in Neurospora crassa. *J. Bacteriol.* **178**, 2172–2177 (1996).
64. Bhushan, S. *et al.* Structural basis for translational stalling by human cytomegalovirus and fungal arginine attenuator peptide. *Mol. Cell* **40**, 138–146 (2010).
65. Gong, F., Ito, K., Nakamura, Y. & Yanofsky, C. The mechanism of tryptophan induction of tryptophanase operon expression: Tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNAPro. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 8997–9001 (2001).
66. Ramu, H., Mankin, A. & Vazquez-Laslop, N. Programmed drug-dependent ribosome stalling. *Mol. Microbiol.* **71**, 811–824 (2009).
67. Hanson, G. & Coller, J. Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews Molecular Cell Biology* **19**, 20–30 (2018).
68. Grantham, R., Gautier, C., Gouy, M., Mercier, R. & Pavé, A. Codon catalog usage and the genome hypothesis. *Nucleic Acids Res.* **8**, 197 (1980).
69. Grantham, R., Gautier, C. & Gouy, M. Codon frequencies in 119 individual genes confirm corsistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* **8**, 1893–1912 (1980).
70. Buhr, F. *et al.* Synonymous Codons Direct Cotranslational Folding toward Different Protein Conformations. *Mol. Cell* **61**, 341–351 (2016).
71. Presnyak, V. *et al.* Codon optimality is a major determinant of mRNA stability. *Cell* **160**, 1111–1124 (2015).
72. Takyar, S., Hickerson, R. P. & Noller, H. F. mRNA helicase activity of the ribosome. *Cell* **120**, 49–58 (2005).
73. Dayhuff, T. J., Atkins, J. F. & Gesteland, R. F. Characterization of ribosomal frameshift events by protein sequence analysis. *J. Biol. Chem.* **261**, 7491–7500 (1986).
74. Blinkowa, A. L. & Walker, J. R. Programmed ribosomal frameshifting generates the Escherichia coli DNA polymerase III γ subunit from within the τ subunit reading frame. *Nucleic Acids Res.* **18**, 1725–1729 (1990).

75. Mejlhede, N., Atkins, J. F. & Neuhard, J. Ribosomal-1 frameshifting during decoding of Bacillus subtilis cdd occurs at the sequence CGA AAG. *J. Bacteriol.* **181**, 2930–2937 (1999).
76. Kurian, L., Palanimurugan, R., Gödderz, D. & Dohmen, R. J. Polyamine sensing by nascent ornithine decarboxylase antizyme stimulates decoding of its mRNA. *Nature* **477**, 490–494 (2011).
77. Caliskan, N., Katunin, V. I., Belardinelli, R., Peske, F. & Rodnina, M. V. Programmed -1 frameshifting by kinetic partitioning during impeded translocation. *Cell* **157**, 1619–1631 (2014).
78. Meydan, S. *et al.* Programmed Ribosomal Frameshifting Generates a Copper Transporter and a Copper Chaperone from the Same Gene. *Mol. Cell* **65**, 207–219 (2017).
79. Stewart, D. E., Sarkar, A. & Wampler, J. E. Occurrence and role ofcis peptide bonds in protein structures. *J. Mol. Biol.* **214**, 253–260 (1990).
80. Zimmerman, S. S. & Scheraga, H. A. Stability of Cis, Trans, and Nonplanar Peptide Groups. *Macromolecules* **9**, 408–416 (1976).
81. Adzhubei, A. A., Sternberg, M. J. E. & Makarov, A. A. Polyproline-II helix in proteins: Structure and function. *J. Mol. Biol.* **425**, 2100–2132 (2013).
82. Qi, F., Motz, M., Jung, K., Lassak, J. & Frishman, D. Evolutionary analysis of polyproline motifs in Escherichia coli reveals their regulatory role in translation. *PLOS Comput. Biol.* 1–19 (2018). doi:10.1371/journal.pcbi.1005987
83. Wang, S., Fleming, R. T., Westbrook, E. M., Matsumura, P. & McKay, D. B. Structure of the Escherichia coli FlhDC complex, a prokaryotic heteromeric regulator of transcription. *J. Mol. Biol.* **355**, 798–808 (2006).
84. Saini, P., Eyler, D. E., Green, R. & Dever, T. E. Hypusine-containing protein eIF5A promotes translation elongation. *Nature* **459**, 118–121 (2009).
85. Roy, H. *et al.* The tRNA synthetase paralog PoxA modifies elongation factor-P with (R)-β-lysine. *Nat. Chem. Biol.* **7**, 667–669 (2011).
86. Yanagisawa, T., Sumida, T., Ishii, R., Takemoto, C. & Yokoyama, S. A paralog of lysyl-tRNA synthetase aminoacylates a conserved lysine residue in translation elongation factor P. *Nat. Struct. Mol. Biol.* **17**, 1136–1143 (2010).
87. Peil, L. *et al.* Lys34 of translation elongation factor EF-P is hydroxylated by YfcM. *Nat. Chem. Biol.* **8**, 695–697 (2012).
88. Nakatogawa, H. & Ito, K. The ribosomal exit tunnel functions as a discriminating gate. *Cell* **108**, 629–636 (2002).
89. Rajapandi, T., Dolan, K. M. & Oliver, D. B. The first gene in the Escherichia coli secA operon, gene X, encodes a nonessential secretory protein. *J. Bacteriol.* **173**, 7092–7097 (1991).
90. Muto, H., Nakatogawa, H. & Ito, K. Genetically Encoded but Nonpolypeptide Prolyl-tRNA Functions in the A Site for SecM-Mediated Ribosomal Stall. *Mol. Cell* **22**, 545–552 (2006).
91. Mitra, K. *et al.* Elongation Arrest by SecM via a Cascade of Ribosomal RNA Rearrangements. *Mol. Cell* **22**, 533–543 (2006).
92. Arenz, S. *et al.* Drug sensing by the ribosome induces translational arrest via active site perturbation. *Mol. Cell* **56**, 446–452 (2014).

93. Davis, L. & Chin, J. W. Designer proteins: Applications of genetic code expansion in cell biology. *Nature Reviews Molecular Cell Biology* **13**, 168–182 (2012).
94. Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001).
95. Goto, Y., Katoh, T. & Suga, H. Flexizymes for genetic code reprogramming. *Nat. Protoc.* **6**, 779–790 (2011).
96. 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 (2014).
97. Achenbach, J. *et al.* Outwitting EF-Tu and the ribosome: Translation with D-amino acids. *Nucleic Acids Res.* **43**, 5687–5698 (2015).
98. Katoh, T., Iwane, Y. & Suga, H. Logical engineering of D-arm and T-stem of tRNA that enhances D-amino acid incorporation. *Nucleic Acids Res.* **45**, 12601–12610 (2017).
99. Katoh, T. *et al.* Consecutive Elongation of D-Amino Acids in Translation. *Cell Chem. Biol.* **24**, 46–54 (2017).
100. Kawakami, T., Murakami, H. & Suga, H. Messenger RNA-Programmed Incorporation of Multiple N-Methyl-Amino Acids into Linear and Cyclic Peptides. *Chem. Biol.* **15**, 32–42 (2008).
101. Kawakami, T., Murakami, H. & Suga, H. Ribosomal Synthesis of Polypeptoids and Peptoid - Peptide Hybrids. *J. Am. Chem. Soc.* 16861–16863 (2008).
102. Ohta, A., Murakami, H., Higashimura, E. & Suga, H. Synthesis of Polyester by Means of Genetic Code Reprogramming. *Chem. Biol.* **14**, 1315–1322 (2007).
103. Takatsuji, R. *et al.* Ribosomal Synthesis of Backbone-Cyclic Peptides Compatible with in Vitro Display. *J. Am. Chem. Soc.* (2019). doi:10.1021/jacs.8b05327
104. Ellman, J. A., Ellman, J. A., Mendel, D. & Shiultz, P. G. Site-Specific Incorporation of Novel Backbone Structures into Proteins. *Science* **197**, (1992).
105. Fujino, T., Goto, Y., Suga, H. & Murakami, H. Ribosomal Synthesis of Peptides with Multiple β-Amino Acids. *J. Am. Chem. Soc.* **138**, 1962–1969 (2016).
106. Katoh, T. & Suga, H. Ribosomal Incorporation of Consecutive β-Amino Acids. *J. Am. Chem. Soc.* **140**, 12159–12167 (2018).
107. Rogers, J. M. *et al.* Ribosomal synthesis and folding of peptide-helical aromatic foldamer hybrids. *Nat. Chem.* **10**, 405–412 (2018).
108. Yamagishi, Y. *et al.* Natural product-like macrocyclic N-methyl-peptide inhibitors against a ubiquitin ligase uncovered from a ribosome-expressed de novo library. *Chem. Biol.* **18**, 1562–1570 (2011).
109. Rogers, J. M., Passioura, T. & Suga, H. Nonproteinogenic deep mutational scanning of linear and cyclic peptides. *Proc. Natl. Acad. Sci.* **115**, 10959–10964 (2018).
110. Mann, M. Functional and quantitative proteomics using SILAC. *Nature Reviews Molecular Cell Biology* **7**, 952–958 (2006).
111. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-Wide Analysis in Vivo of Resolution Using Ribosome Profiling. *Science* **324**, 218–23 (2009).
112. Jackson, R. & Standart, N. The awesome power of ribosome profiling. *RNA* **21**, 652–654 (2015).
113. Fowler, D. M. *et al.* High-resolution mapping of protein sequence-function relationships. *Nat. Methods* **7**, 741–

746 (2010).

114. Olson, C. A., Wu, N. C. & Sun, R. A comprehensive biophysical description of pairwise epistasis throughout an entire protein domain. *Curr. Biol.* **24**, 2643–2651 (2014).

115. Zondlo, N. J. Aromatic-proline interactions: Electronically tunable CH/π interactions. *Acc. Chem. Res.* **46**, 1039–1049 (2013).

116. Heurgué-Hamard, V., Mora, L., Guarneros, G. & Buckingham, R. H. The growth defect in Escherichia coli deficient in peptidyl-tRNA hydrolase is due to starvation for Lys-tRNA(Lys). *EMBO J.* **15**, 2826–2833 (1996).

117. Hernández, J., Ontiveros, C., Valadez, J. G., Buckingham, R. H. & Guarneros, G. Regulation of protein synthesis by minigene expression. *Biochimie* **79**, 527–531 (1997).

118. Dinçbas, V., Heurgué-Hamard, V., Buckingham, R. H., Karimi, R. & Ehrenberg, M. Shutdown in protein synthesis due to the expression of mini-genes in bacteria. *J. Mol. Biol.* **291**, 745–759 (1999).

119. Cruz-vera, L. R. *et al.* The rate of peptidyl-tRNA dissociation from the ribosome during minigene expression depends on the nature of the last decoding interaction. *J. Biol. Chem.* **278**, 26065–26070 (2003).

120. Gonzalez De Valdivia, E. I. & Isaksson, L. A. Abortive translation caused by peptidyl-tRNA drop-off at NGG codons in the early coding region of mRNA. *FEBS J.* **272**, 5306–5316 (2005).

121. Delgado-Olivares, L., Zamora-Romo, E., Guarneros, G. & Hernandez-Sanchez, J. Codon-specific and general inhibition of protein synthesis by the tRNA-sequestering minigenes. *Biochimie* **88**, 793–800 (2006).

122. Jacinto-loeza, E., Vivanco-domínguez, S., Guarneros, G. & Hernández-sánchez, J. Minigene-like inhibition of protein synthesis mediated by hungry codons near the start codon. *Nucleic Acids Res.* **36**, 4233–4241 (2008).

123. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame , single-gene knockout mutants : the Keio collection. *Mol. Syst. Biol.* **4474**, (2006).

124. Keiler, K. C., Waller, P. R. H. & Sauer, R. T. Role of a Peptide Tagging System in Degradation of Proteins Synthesized from Damaged Messenger RNA. *Science* **271**, 990–993 (1996).

125. Wali Karzai, A., Susskind, M. M. & Sauer, R. T. SmpB, a unique RNA-binding protein essential for the peptide-tagging activity of SsrA (tmRNA). *EMBO J.* **18**, 3793–3799 (1999).

126. Chadani, Y. *et al.* Ribosome rescue by Escherichia coli ArfA (YhdL) in the absence of trans-translation system. *Mol. Microbiol.* **78**, 796–808 (2010).

127. Chadani, Y., Ono, K., Kutsukake, K. & Abo, T. Escherichia coli YaeJ protein mediates a novel ribosome-rescue pathway distinct from SsrA- and ArfA-mediated pathways. *Mol. Microbiol.* **80**, 772–785 (2011).

128. Lytvynenko, I. *et al.* Alanine Tails Signal Proteolysis in Bacterial Ribosome-Associated Quality Control. *Cell* 1–15 (2019). doi:10.1016/j.cell.2019.05.002

129. Dougan, D. A., Truscott, K. N. & Zeth, K. The bacterial N-end rule pathway: Expect the unexpected. *Mol. Microbiol.* **76**, 545–558 (2010).

Achievement

[Publication related to the thesis]

"Comprehensive elucidation of effect of nascent peptide sequences and EF-G concentration on incorporation of prolines in translation"

Kenya Tajima, Takayuki Katoh, Hiroaki Suga, *Peptide Science 2018*, 2019, 103

[Publication not related to the thesis]

"Genetic code expansion via integration of redundant amino acid assignment by finely tuning tRNA pools"

Kenya Tajima, Takayuki Katoh, Hiroaki Suga, *Current opinion in chemical biology*, 2018, 46: 212-218.

[Poster presentations]

「EF-G とペプチド新生鎖がペプチジル-tRNA 脱落に及ぼす影響の解明」

○田島研也、加藤敬行、菅裕明、バイオ関連化学若手フォーラム、東京、2017 年 9 月

「プロリンの翻訳導入に対するペプチド新生鎖配列と EF-G 濃度の影響の網羅的評価」

○田島研也、加藤敬行、菅裕明、日本ケミカルバイオロジー学会　第 13 回年会、東京、2018 年 6 月

"Comprehensive elucidation of effect of nascent peptide sequences and EF-G concentration on incorporation of prolines in translation"

○Kenya Tajima, Takayuki Katoh, Hiroaki Suga, 4th International Conference on Circular Proteins and Peptides, Kawasaki, November 2018

「プロリンの翻訳導入に対する新生ペプチド配列と EF-G 濃度の影響の網羅的評価」

○田島研也、加藤敬行、菅裕明、第 55 回ペプチド討論会、京都、2018 年 12 月

Acknowledgement

This research has been conducted under the supervision by professor Hiroaki Suga. I appreciate him for his kind education, discussion, guidance, and encourage throughout this study. I am deeply grateful to professor Naokazu Kano, associate professor Takayuki Katoh, associate professor Yuki Goto, Dr Toby Passioura, project assistant professor Naohiro Terasaka, and project assistant professor Alex Vinogradov for helpful advices and discussion for this study. I thank Dr. Naoya Ozawa for preparation of the python manuscript.

I thank National BioResource Project (NIG, Japan): E. coli for furnishing Keio collection JW4107 strain.

I thank Human Frontier Science Program for financial support [RGP0015/2017].

I thank Grants-in-Aid for JSPS Fellows for financial support [18J12342].