Cell type-specific proteome labeling by genetic code expansion

Eunjin Kim

Department of Medical Science

The Graduate School, Yonsei University

Cell type-specific proteome labeling by genetic code expansion

Directed by Professor Hosung Jung

The Master's Thesis
submitted to the Department of Medical Science,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Master of Medical Science

Eunjin Kim

December 2015

This certifies that the Master's Thesis of
Eunjin Kim is approved.

Thesis Supervisor: Hosung Jung

Thesis Committee Member#1: Hoguen Kim

Thesis Committee Member#2: Hyunsoo Lee

The Graduate School
Yonsei University

December 2015

ACKNOWLEDGEMENTS

아직도 서툴고 배울 것이 많지만 2 년 이라는 짧은 시간 동안 조금이나마 성장할 수 있었습니다. 모든 분들께 감사의 말씀 드립니다. 항상 저를 말없이 지켜봐 주시는 부모님과 가족들 덕분에 대학원 석사 과정을 순조롭게 시작하고 끝낼 수 있었습니다. 부족하고 서툴렀던 저를 인내를 갖고 이끌어 주신 정호성 지도교수님, 감사합니다. 철 없던 행동들과 실수들에 죄송하고 부끄러운 마음 밖에 없습니다. 교수님의 제자에 대한 열정과 헌신 덕분에 많은 것들을 배웠습니다. 학위 논문을 검토해주시고 많은 조언 주신 김호근, 이현수 교수님께도 감사의 말씀을 드립니다. 언제나 따스한 미소로 인사에 답해 주시는 해부학 교실의 박경아, 이원택, 김명희, 이종은, 이혜연, 복진웅 교수님께도 감사의 말씀 드립니다. 그리고 가족보다 더 많은 시간들을 함께 지냈던 실험실 박사님들,언니, 오빠, 후배들에게도 감사의 말씀을 드립니다. 곁에서 응원해 주시고 충고와 격려를 해주셨던 모든 분들 덕분에 저는 2 년 전의 제 모습에서 발전 하였고 2 년 동안의 노력들을 한 편의 이야기로 만들 수 있었습니다.

마지막으로 바빠서 잘 놀아주지 못하는 것에 대해 미안함을 가지며, 두 고양이들에게 감사의 말씀을 드립니다.

TABLE OF CONTENTS

List of Figures

ABSTRACT

Cell type-specific proteome labeling by genetic code expansion

Eunjin Kim

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Hosung Jung)

Analyzing gene expression of specific cells in a living organism is crucial in understanding complex biological systems. It has not been possible, however, to isolate the proteome of specific cells in vertebrates. Recently, advance in genetic code expansion has provided a new potential direction in achieving this aim. By using the tool of genetic code expansion, I expanded the genetic code of mammalian cells by introducing the pyrrolysyl-tRNA synthetase (PylRS)/$tRNA^{Pyl}$ system of *Methanosarcina mazei*, a methane-producing genus of archaea. Although the amber (UAG) codon is a stop codon to most of organisms, *Methanosarcina* uses it as a sense codon that encodes pyrrolysine (Pyl), the 22^{nd} amino acid. Unlike eukaryotes where the amber codon is recognized by eukaryotic release factor 1 (eRF1) which helps the termination of mRNA translation, *Methanosarcina* expresses the tRNA ($tRNA^{Pyl}$) whose anticodon recognizes the amber codon. PylRS esterifies $tRNA^{Pyl}$ and Pyl,

generating the Pyl-tRNA^{Pyl} that site-specifically incorporates Pyl into the growing polypeptides during mRNA translation. In this thesis, I expressed PylRS and tRNA^{Pyl} in mammalian cells and showed that the *M. mazei* PylRS/tRNA^{Pyl} system is functional in mammalian cells and orthogonal to endogenous tRNAs and tRNA synthetases. I also showed that the anticodon of tRNA^{Pyl} can be changed to recognize other stop and sense codons without compromising their compatibility with PylRS and Pyl. Furthermore, I showed that Pyl-derivatives containing bio-orthogonal functional group such as alkyne can be site-specifically incorporated into the endogenous proteome only when the genetic code was expanded by the PylRS/tRNA^{Pyl} system. By using azide-alkyne cycloaddition, the *de novo* proteome of genetic code-expanded cells could be specifically visualized. The results of this thesis show a promising direction for a new technique to isolate the proteome of a specific cell type from a living animal.

Key words: *De novo* proteome, Genetic code expansion, Pyrrolysine, Azide-alkyne cycloaddition.

Cell type-specific proteome labeling by genetic code expansion

Eunjin Kim

Department of Medical Science
The Graduate School, Yonsei University

(Directed by Professor Hosung Jung)

Ⅰ. INTRODUCTION

A conventional method to study the proteome is largely based on metabolic labeling, which labels all proteins in all cells in the same sample (e.g. tissue or cell culture) using tagged amino acids (*e.g.* 35S-methionine) or their analogs (*e.g.* L-azidohomoalanine, or AHA) that can be used for protein synthesis. For example, AHAs treated in the sample enter the cells and compete with methionines for the enzyme, methionine aminoacyl-tRNA synthetase (MetRS). MetRS then loads the tRNA that recognizes the AUG codon ($tRNA^{Met}$) to methionine or AHA. AHA-$tRNA^{Met}$ generated in this way is then incorporated into the AUG codon of an mRNA during mRNA translation[1,2]. As such, this method is intrinsically not cell-selective and labels all proteins being synthesized (*i.e. de novo* proteome) in all cells in the sample[3]. Genetic code expansion is a growing technique mainly investigated for site-specific incorporation of unnatural amino acids (UAAs) into a protein. The UAG stop codon (also known as amber) is one of the three universal stop codons together with UGA (also known as opal) and UAA (also known as ochre). In eukaryotes, these stop

codons are not recognized by the tRNAs but by the eukaryotic release factor 1 (eRF1) which terminates mRNA translation (Fig. 1A). Interestingly, a genus of *euryarchaeota archaea*, *Methanosarcina* uses the amber codon as a sense codon to encode pyrrolysine (Pyl), the 22nd amino acid (Fig. 1B).

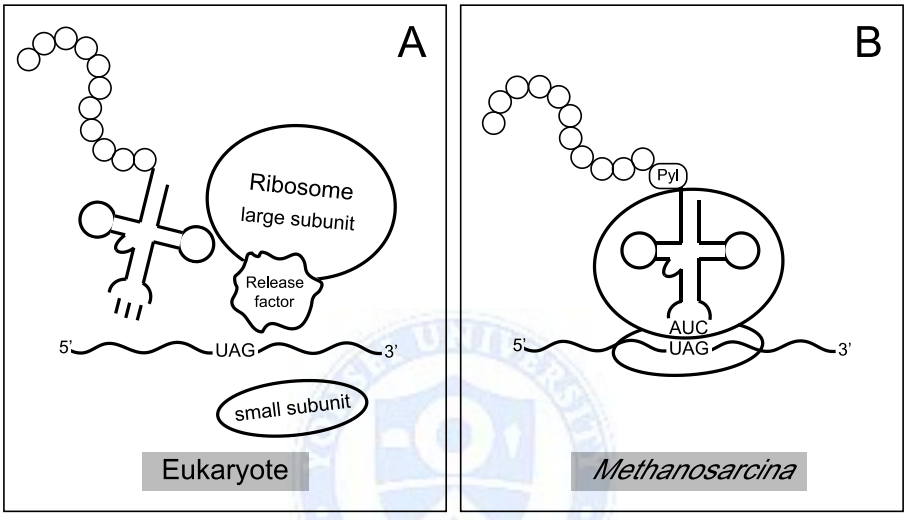
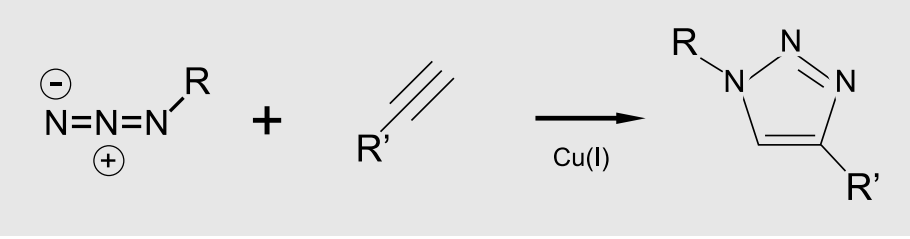


Figure 1. mRNA translation at the amber (UAG) codon in eukaryote and *Methanosarcina*. (A) Most organisms use the UAG codon as a stop codon. Translation is terminated at the amber, opal and ochre (UAG, UGA and UAA) codons. (B) *Methanosarcina* species use the UAG codon as a sense codon. At the UAG codon, the 22nd amino acid, Pyl, is incorporated into the polypeptide and translation goes on.

M. barkeri and *M. mazei* can use the amber codon as a sense codon because they have pyrrolysyl-tRNA synthetase (PylRS) and $tRNA^{Pyl}$ genes. PylRS esterifies $tRNA^{Pyl}$ and Pyl, generating Pyl-$tRNA^{Pyl}$. The anticodon of $tRNA^{Pyl}$ recognizes the amber codon and Pyl is incorporated into the polypeptides during mRNA translation. Importantly, PylRS and $tRNA^{Pyl}$ can function in other hosts when ectopically expressed, and are orthogonal to eukaryotic tRNAs and aminoacyl-tRNA synthetases. In other words, they do not show cross-reactivity with bacterial or eukaryotic tRNAs, aminoacyl-tRNA synthetases, and amino acids. Therefore, if a eukaryotic cell is transfected with the PylRS/$tRNA^{Pyl}$ pair and supplied with Pyl, the Pyl-$tRNA^{Pyl}$ is synthesized and competes with eRF1, resulting in a stochastic incorporation of Pyl at the amber codon. By using Pyl derivatives that are structurally similar enough to Pyl to be recognized by the PylRS/$tRNA^{Pyl}$ system yet at the same time bear a bio-orthogonal functional group, it has now become possible to incorporate such UAAs into the amber codon in *Escherichia coli*, yeast, mammalian cells, *Caenorhabditis elegans* and *Drosophila melanogaster*[4,5,6,7].

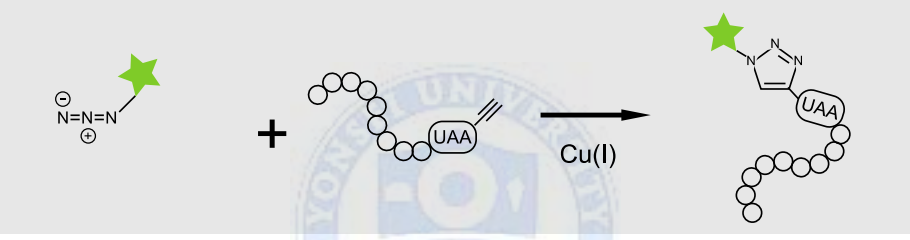
In this thesis, I applied the PylRS/$tRNA^{Pyl}$ system into mammalian cells to expand the genetic code and label *de novo* proteome of specific cell populations (*i.e.* transfected cells). I confirmed that the PylRS/$tRNA^{Pyl}$ pair of *M. mazei* suppresses the amber stop codon in mammalian cells by incorporating Pyl derivatives into this codon. To label *de novo* proteome in genetic code expanded cells, an alkyne-bearing Pyl derivative was added to cells as a Pyl analog. And by using copper-catalyzed azide-alkyne cycloaddition (CuAAC)[8], the Pyl derivative can be connected to fluorescent molecules bearing an azide group (also known as "click chemistry") (Fig. 2). I successfully 'clicked' fluorescence molecules to the *de novo* proteome bearing UAAs, and visualized the labeling of endogenous proteome specifically in genetic code-expanded cells.

A. Principle of Copper-catalyzed azide-alkyne cycloaddition



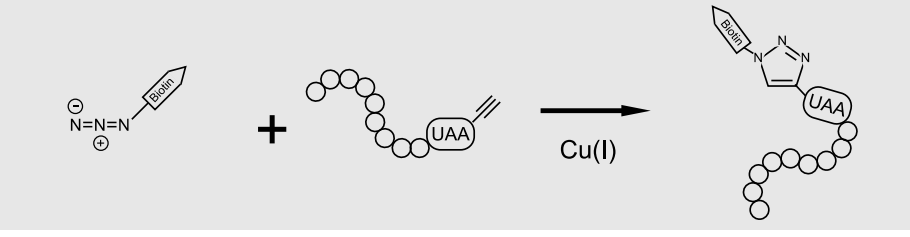
Azide Alkyne Triazole

B. Visualization of labeled protein by CuAAC



Fluorophore-azide UAA incorporated polypeptide-alkyne Fuorophore-polypeptide

C. Purification of labeled protein by CuAAC



Biotin-azide UAA incorporated polypeptide-alkyne Biotin-polypeptide

Figure 2. Visualization and purification of UAA labeled proteins by CuAAC. (A) A chemical reaction generally used in studying organism. An alkyne reacts with an azide and makes a triazole in the catalysis of copper. Cycloaddition of alkyne and azide is one of the reactions commercially used as click chemistry. This reaction is simply termed as CuAAC (Copper-catalyzed azide-alkyne cycloaddition). (B) When alkyne in UAA meets fluorophore tagged azide, they react and create a triazole. Triazole between alkyne and azide is interchangeable. Genetic code expanded polypeptides can be visualized by expressing the fluorophore of azide. (C) If azide is tagged with biotin, alkyne-UAA incorporated polypeptides can be purified by the biotin-streptavidin interaction.

After establishing the method to specific incorporation of the UAA at the amber stop codon by PylRS/$tRNA^{Pyl}$ in mammalian cells, I changed the anticodon of $tRNA^{Pyl}$ to recognize not only amber stop codon but also two other stop codons and four sense codons. Furthermore, I found that expressing multiple $tRNA^{Pyl}$s with differently engineered-anticodons in the same cells allows more efficient proteome labeling. The results of this thesis would provide the basis to develop a new method to study the *de novo* proteome of specific cell population in a living organism at specific time.

Ⅱ. MATERIALS AND METHODS

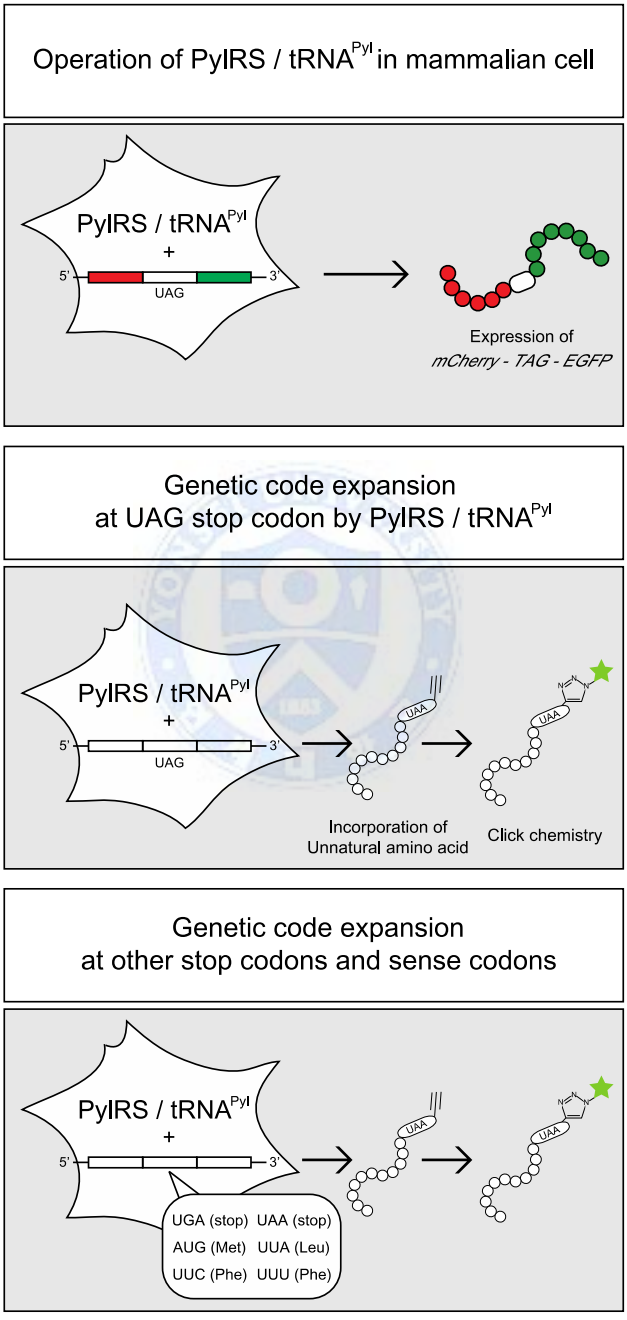
1. Materials

Dulbecco's low glucose modified eagle's medium (DMEM-low glucose medium) and Fetal bovine serum (FBS) were purchased from Hyclone. Media contains 4.0 mM L-Glutamine and 110 mg/sodium pyruvate. Opti-MEM, Antimycotic antibiotic, 10 X hank's balanced salt solution (HBSS), Click-iT cell reaction buffer kit (Catalog #C10269) and Click-iT protein reaction buffer kit (Catalog #C10276) were purchased from Invitrogen. Poly-L-lysine and paraformaldehyde were purchased from Sigma. Mammalian cells were plated in 12 well and 6 well culture plate from SPL. Reagent for transfection was Fugene 6 transfection reagent (Catalog #E2693) from Promega. 1 X phosphate buffered saline (PBS) for washing buffer was from 20 X PBS of Biosesang. As analog of Pyrrolysine (Pyl), unnatural amino acids (UAAs) were Nε-(tert-butyloxycarbonyl)-L-lysine from Bachem and N6-[(2-propynyloxy) carbonyl]-L-lysine from Sichem. Mounting solution was aqua-poly/mount purchased from polyscience.

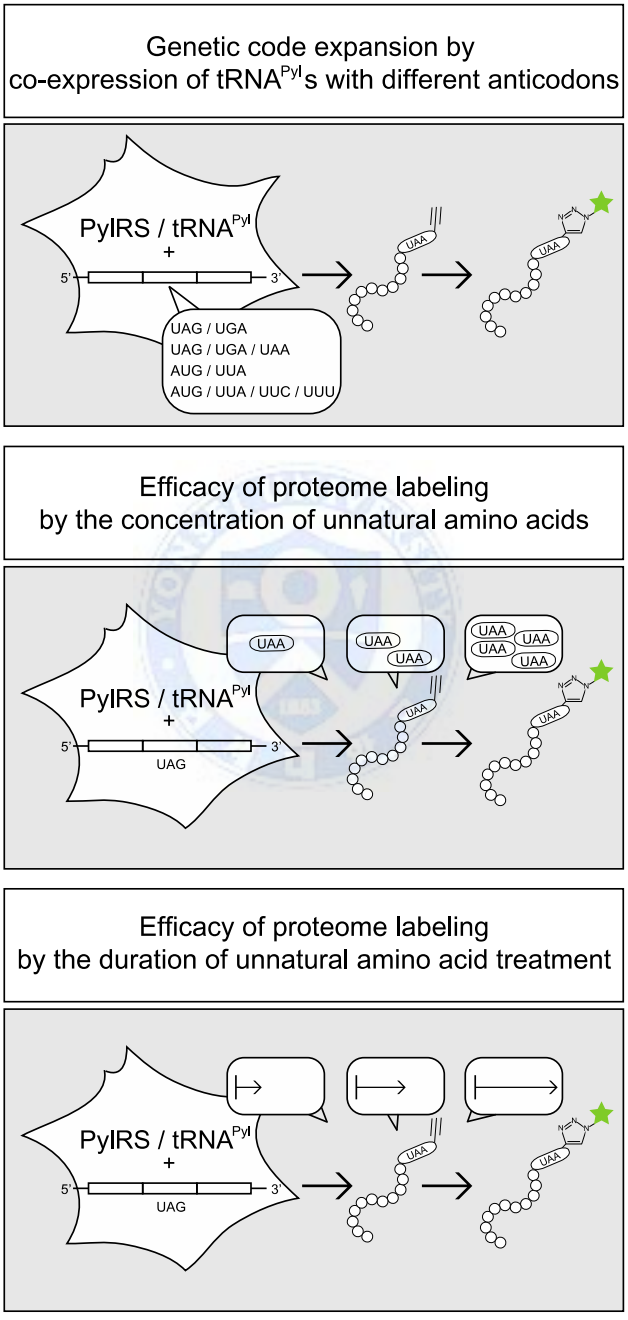
2. Mammalian cell culture

Mammalian cells in this study were human embryonic kidney 293T (HEK293T) cell line. Cells were grown in 1 X antimycotic antibiotic, 10 % FBS in DMEM-low glucose. For fluorescent imaging analysis, cells were plated on coverslip. Glass coverslip was sterilized by 100 % EtOH and placed in 12 well culture plate. They coated with 0.1 % Poly-L-lysine (w/v) in H_2O for 1 hour at room temperature. Confluency of cell was 20 % of 12 well culture plate when it was plated. Incubation was done for overnight in 37 ℃, 5 % CO_2 incubator. An outline of the experimental schemes is illustrated in Figure 3 (Fig. 3).

A


Operation of PylRS / tRNA^{Pyl} in mammalian cell
PylRS / tRNA^{Pyl}
+
5'
3'
UAG
Expression of
mCherry - TAG - EGFP
Genetic code expansion
at UAG stop codon by PylRS / tRNA^{Pyl}
PylRS / tRNA^{Pyl}
+
5'
3'
UAG
UAA
UAA
Incorporation of
Unnatural amino acid
Click chemistry
Genetic code expansion
at other stop codons and sense codons
PylRS / tRNA^{Pyl}
+
5'
3'
UAA
UAA
UGA (stop) UAA (stop)
AUG (Met) UUA (Leu)
UUC (Phe) UUU (Phe)

B


Genetic code expansion by
co-expression of tRNA^{Pyl}s with different anticodons
PylRS / tRNA^{Pyl}
+
5'
3'
UAG / UGA
UAG / UGA / UAA
AUG / UUA
AUG / UUA / UUC / UUU
UAA
Efficacy of proteome labeling
by the concentration of unnatural amino acids
PylRS / tRNA^{Pyl}
+
5'
3'
UAG
UAA
Efficacy of proteome labeling
by the duration of unnatural amino acid treatment
PylRS / tRNA^{Pyl}
+
5'
3'
UAG

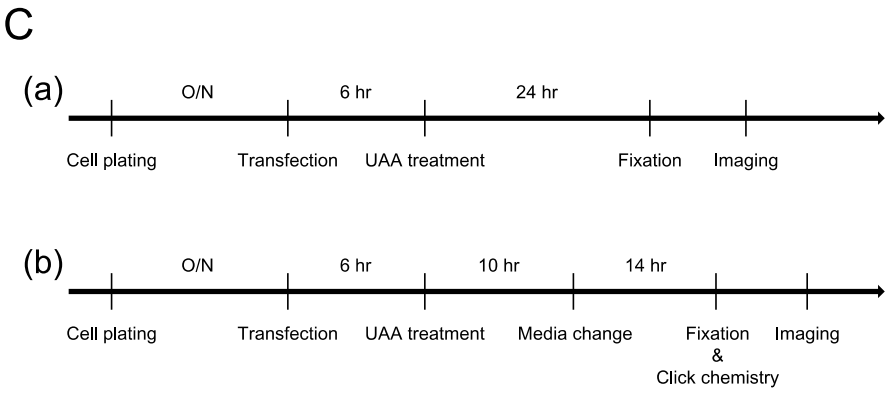


Figure 3. Experimental schemes for studying genetic code expansion in mammalian cells. (A) and (B) Summary about this study. Prior to labeling the proteome of mammalian cells by the PylRS/tRNA^{Pyl}, operation of PylRS/tRNA^{Pyl} in the mammalian cells should be confirmed. If PylRS/tRNA^{Pyl} functions in the mammalian cells, proteome labeling of mammalian cells by the genetic code expansion is possible. Labeled proteins by PylRS/tRNA^{Pyl} can be visualized by click chemistry. To increase the efficiency of proteome labeling, I applied point mutation to anticodon of tRNA^{Pyl} whether it can recognize other stop codons and even the sense codons. Also, I applied multiple tRNA^{Pyl}s with differently changed anticodons into the cells which means multiple genetic code expansion can happen at stop codons and sense codons. Because the concentration and duration of UAA treatment determine the efficiency of proteome labeling, I also measured the optimal condition of UAA treatment. (C) Cells were plated the day before transfection and plated on poly-L-lysine coated coverslip. After checking the confluency and viability of the cell, DNAs that were needed to expand the genetic code were transfected into the cells. (a) To confirm the operation of genetic system of *Methanosarcina* in HEK293T cells, *mCherry –TAG–EGFP* was transfected as a reporter plasmid with *PylRS/tRNA^{Pyl}*. (b) And to label the genetic code expanded proteins in HEK293T cells, *mCherry* was transfected as another reporter plasmid with *PylRS/tRNA^{Pyl}*. Genetic code expanded and UAA incorporated proteins were visualized by the click chemistry.

3. Transient transfection of mammalian cells

Before one-hour transfection, media was replaced with antimycotic antibiotic-free, 10 % FBS in DMEM-low glucose to increase the efficiency of transfection. In transfection complex, the ratio of transfection reagent and DNA was 3:1 ratio. A plasmid for genetic code expansion (*e.g*. *PylRS/tRNA*^{Pyl}) and a reporter plasmid (*e.g*. *mCherry-TAG-EGFP* for suppression assays, or *mCherry* for proteome labeling) were transfected in 4:1 ratio. The complex was incubated for 20 min at room temperature and added to cells in a drop-wise manner. After 6 hr of transfection, media was changed to the media that contained UAAs.

4. Treatment of unnatural amino acids

Nε-(tert-butyloxycarbonyl)-L-lysine (Fig. 4A) was dissolved in 0.1 M NaOH solution. And then to adjust the pH, 5 M HCl was added. N6-[(2-propynyloxy) carbonyl]-L-lysine was dissolved in water but has high pH which cells can't live with in. So 1 M HCl was added to adjust the pH. 5 mM of these UAAs (Fig. 4B) was used to the cell culture media as working solution. Stock solution was 130 mM. Cells were incubated for at least 24 hr within the UAA contained media.

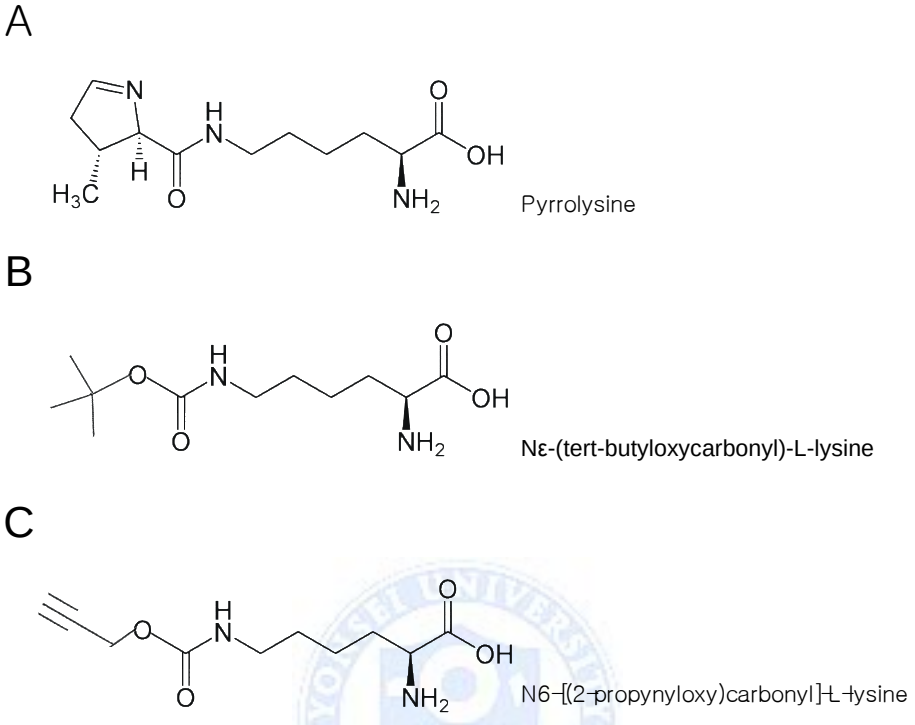


Figure 4. Pyrrolysine (Pyl) and unnatural amino acids (UAAs) used in this study. (A) Pyl is the 22nd amino acid that archaea use in their metabolism. Two UAAs, (B) Nε-(tert-butyloxycarbonyl)-L-lysine and (C) N6-[(2-propynyloxy) carbonyl]-L-lysine, were typically used as Pyl analogs. Especially, N6-[(2-propynyloxy) carbonyl]-L-lysine has an alkyne. In the catalysis of copper, an alkyne and an azide make a triazole (Fig. 2) which is basic reaction for click chemistry.

5. Click chemistry

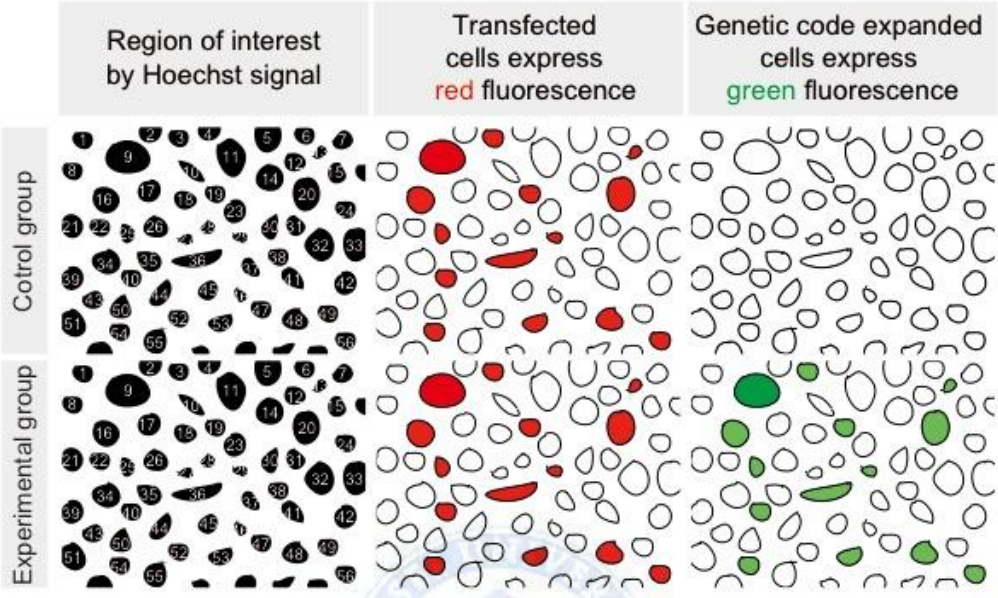
A) Quantitative fluorescence imaging

The alkyne-containing proteome was labeled by BDP-FL-azide by click chemistry using Click-iT cell reaction buffer kit (Molecular probes, USA) with slight modifications. Specifically, Component A was replaced with 0.1 M Tris-HCl (pH 8.5). Component C was replaced with 5 mM ascorbic acid. Instead of Alexa fluor 488, BDP FL azide that is a green fluorophore (maximum excitation :503 nm, maximum emission :509 nm) was used as a counterpart with alkyne which can make copper-catalyzed triazole formation. Cells that were plated on coverslip were washed three times with 1 X HBSS (with calcium) and fixed for 15 min with 4 % paraformaldehyde (PFA) in 1 X PBS. Cells were washed three times with 1 X PBS. While washing the cell, click chemistry reaction cocktail was prepared. The important thing when preparing the cocktail is adding the reagents as the order in the protocol (For example, 0.1 M Tris-HCl, $CuSO_4$, ascorbic acid and BDP FL azide) and the cocktail should be used within 15 min. Samples were incubated in the cocktail for 30 min at room temperature with an avoidance of the light. After that, cells were washed three times with 1 X PBS and mounted with aqua-poly/mount. For quantitative imaging, all images were taken using a laser scanning confocal microscope (Zeiss LSM 700, Carl Zeiss, NY, USA) equipped with a 10 x (N.A. 0.3) and a 20 x (N.A. 0.8) objectives. The samples to be compared were prepared and imaged in identical conditions. Image acquisition settings were non-saturating and identical.

B) Quantitative image analysis

To confirm the proteome labeling by genetic code expansion in mammalian cells, imaging was analyzed by ImageJ. For unbiased analysis, all nuclei in one sample were selected as regions of interest (ROIs) in thresholded Hoechst 33342 stained image. Background-subtracted mean green (labeled proteome) and red

fluorescence intensities per ROI were then calculated. Genetic code-expanded cell nuclei were then selected by the intensity of red fluorescence protein encoded by co-transfected control plasmid (*e.g*. *mCherry*). Then green-to-red intensity ratios were calculated for all transfected cells and plotted. All cells in all samples were numbered based on nucleus staining by Hoechst 33342 (Molecular probes) staining. And then each fluorescence was calculated (Fig. 5). Red fluorescent signal was considered as transfection efficiency. Green fluorescent signal was considered as genetic code expansion efficiency.



$$\frac{G_{(i)}}{R_{(i)}}$$

Green-to-Red intensity ratio

	Intensities of red fluorescence			Intensities of green fluorescence		
	Number	Area	Mean	Number	Area	Mean
Control group	1	0.126	14621	1	0.126	1411
	2	0.322	9957	2	0.322	1957
	3	0.111	65536	3	0.111	5536
	4	0.351	8472	4	0.351	2472
	5	0.549	15632	5	0.549	1632
	6	0.471	1253	6	0.471	1253
	7	0.391	5221	7	0.391	5221
	8	0.482	23873	8	0.482	2873
	9	0.953	5674	9	0.953	1674
	10	0.352	14002	10	0.352	1402
Experimental group	1	0.213	14521	1	0.213	1621
	2	0.732	1957	2	0.732	1572
	3	0.035	6526	3	0.035	55536
	4	0.768	65536	4	0.768	2411
	5	0.233	13231	5	0.233	3532
	6	0.561	253	6	0.561	2531
	7	0.222	7421	7	0.222	2226
	8	0.635	10000	8	0.635	3293
	9	0.221	56274	9	0.221	65536
	10	0.226	11202	10	0.226	1222

Figure 5. Quantification of fluorescence in genetic code expanded cells by ImageJ. Cells were numbered based on Hoechst signal. To count cells automatically, black and white 'binary image' was required. 16 bit of Hoechst images were converted to greyscale images. After threshold the image, pixels under the threshold are changed to black and pixels above the threshold are changed to white. One black dot was considered as a cell. Transfected cells would express red fluorescent protein (RFP)s by the transfection of reporter plasmid. Expression of the green fluorescence means the expansion of the genetic code in mammalian cells. When the translation went on in the stop codon in the reporter (*mCherry-TAG-EGFP*), fluorescent fusion protein would be made. Also, when the stop codon of the cell was expanded by PylRS/$tRNA^{Pyl}$ and UAA was incorporated into there, green fluorescence would be expressed because of the click reaction between alkyne in UAA and green fluorophore tagged azide. Efficiency of genetic code expansion and UAA incorporation was calculated by the intensity ratio of green–to–red fluorescence.

6. Statistical analysis

Experiments were expressed as the mean ± SEM. Statistical analysis was performed by one-way analysis of variance (ANOVA) followed by Dunnett's or Bonferroni's post analysis. Statistical significance was defined as * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.0001$.

Ⅲ. RESULTS

1. Pyrrolysyl-tRNA synthetase and $tRNA^{Pyl}$ can function in mammalian cells

Before labeling the proteome in mammalian cells by using PylRS and $tRNA^{Pyl}$, I confirmed the operation of PylRS/$tRNA^{Pyl}$ system of *Methanosarcina* in mammalian cells. To visualize the suppression at the UAG stop codon by the PylRS/$tRNA^{Pyl}$ pair, DNA plasmid in which UAG stop codon is located between red fluorescent protein (mCherry) and green fluorescent protein (EGFP) (*mCherry-TAG-EGFP*) was transfected into HEK293T cells. Generally, it is called as 'reporter' because it reports the genetic code expansion at the stop codon by expressing fluorescent protein. 'Suppressor' plasmid DNA that encodes PylRS and $tRNA^{Pyl}$ was also transfected with the reporter plasmid (Fig. 6A). After transfection, unnatural amino acids (UAAs) (Fig. 6B) that act as Pyl analog were treated into the cells. Nε-(tert-butyloxycarbonyl)-L-lysine, also called as Lys(Boc) is treated with 0 and 5 mM of concentration into the control and experimental group. By genetic code expansion, this UAA would be recognized at the UAG stop codon and incorporated into the polypeptides. The expected outcome was that only the group which had PylRS and $tRNA^{Pyl}$ and has been treated with UAAs would express the EGFP signals (Fig. 6D). Because of the translation at the UAG stop codon, mCherry-EGFP fluorescent fusion protein would be made. This result identified that translating system of Pyl of *Methanosarcina* worked in mammalian cell system and expanded the genetic code of mammalian cells.

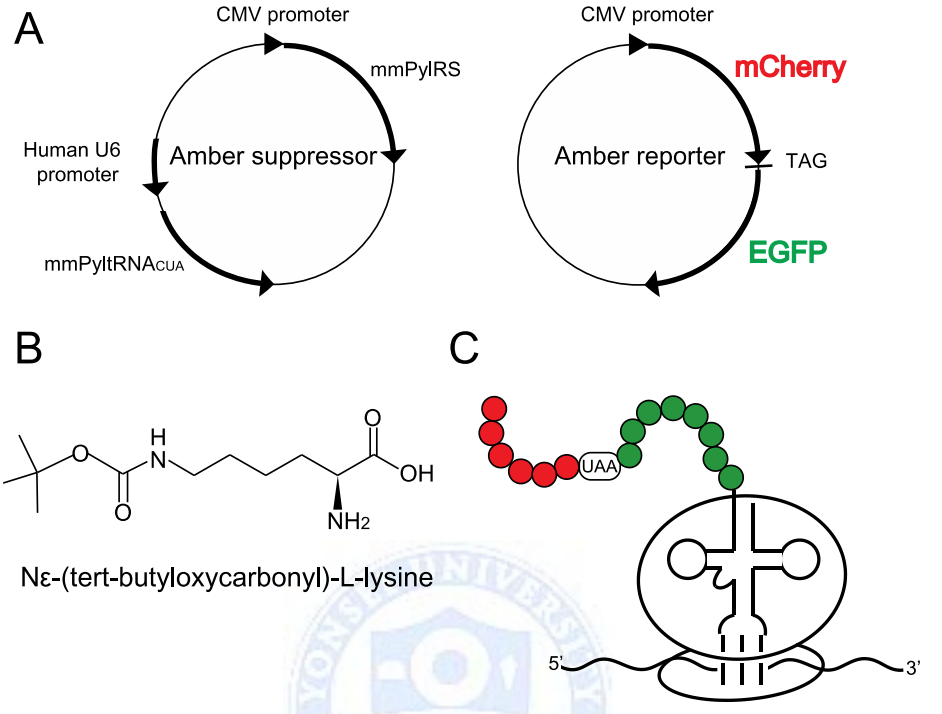


Figure 6. Amber suppression assay to visualize the expansion of the genetic code in HEK293T cells. (A) DNA plasmid constructs used for genetic code expansion in HEK293T cells. Amber suppressor has pyrrolysyl-tRNA (*mmPyltRNA*$_{CUA}$) and tRNA synthetase (*mmPylRS*) of *M. mazei*. Amber reporter has UAG stop codon (amber stop codon) between the genetic code of red fluorescent protein (mCherry) and green fluorescent protein (EGFP) (*mCherry-TAG-EGFP*). (B) Nε-(tert-butyloxycarbonyl)-L-lysine, also called as Lys(Boc), was used instead of Pyl. (C) An expected process in HEK293T cells. If the PylRS and tRNA^{Pyl} work in the protein synthesis and suppress the UAG stop codon in HEK293T cells, EGFP in the amber reporter will be translated followed by mCherry.

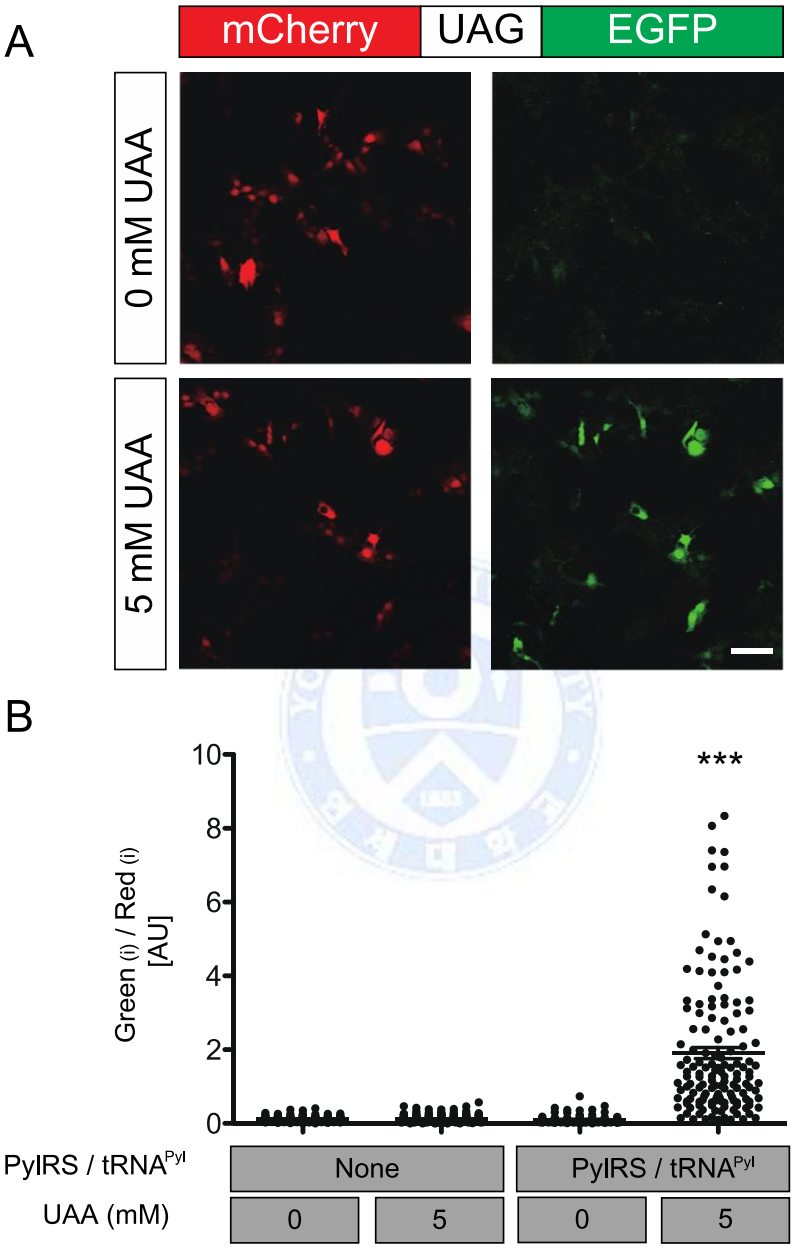

A
mCherry
UAG
EGFP
0 mM UAA
5 mM UAA
B
Green (i) / Red (i) [AU]
10
8
6
4
2
0
***
PylRS / tRNAPyl
None
PylRS / tRNAPyl
UAA (mM)
0
5
0
5

Figure 7. UAA-dependent suppression of the amber stop codon in genetic code-expanded HEK293T cells. (A) Images of genetic code expanded HEK293T cells and control group cells. Two groups were transfected with amber reporter and amber suppressor. After, they were treated with 0 or 5 mM Lys(Boc) which is Pyl analog. EGFP was significantly expressed only at genetic code expanded and UAA treated group. (B) Expression of fluorescent protein was showed as a graph. A single dot means a cell. Cells were numbered by Hoechst signal. Expression of mCherry was considered as efficiency of transfection. Cells that expressed red fluorescence below the standard were excluded in analysis. The expression value of genetic code expansion was measured as the value of the green fluorescence divided by the value of the red fluorescence in all cells in each group. UAA, unnatural amino acid. scale bar, 100 μm. AU, arbitrary unit. Statistical analysis was performed by one-way ANOVA followed by Dunnett's post analysis. ***$p < 0.001$.

2. The anticodon of $tRNA^{Pyl}$ can be engineered to recognize other stop codons

Genetic codes almost frequently used in the organism consist of 64 codons. With some exception, three nucleotides make a codon. 61 codons specify different amino acids. These codons are 'sense codons'. 3 codons are designated as 'nonsense, termination or stop codons' because they take a role in signaling of termination of translation. Sequence of 3 stop codons are UAG, UGA and UAA. For the enhancement of the efficiency of genetic code expansion, I changed the anticodon of $tRNA^{Pyl}$ by the point mutation to expand the genetic code of mammalian cells not only at the UAG stop codon but also at the rest stop codons (Fig. 8). The anticodon of $tRNA^{Pyl}$ which originally recognizes UAG codon was changed to recognize other two stop codons. Accordingly, the stop codons between *mCherry* and *EGFP* in the reporters were also changed. Each converted $tRNA^{Pyl}$s worked by recognizing and expanding their cognate stop codons (Fig. 9). These results showed that mutation of $tRNA^{Pyl}$ didn't influence the operation of PylRS to esterify specific amino acid to the compatible $tRNA^{Pyl}$. And they showed the possibility of genetic code expansion at the 'sense codon'.

Genetic code expansion at all three stop codons

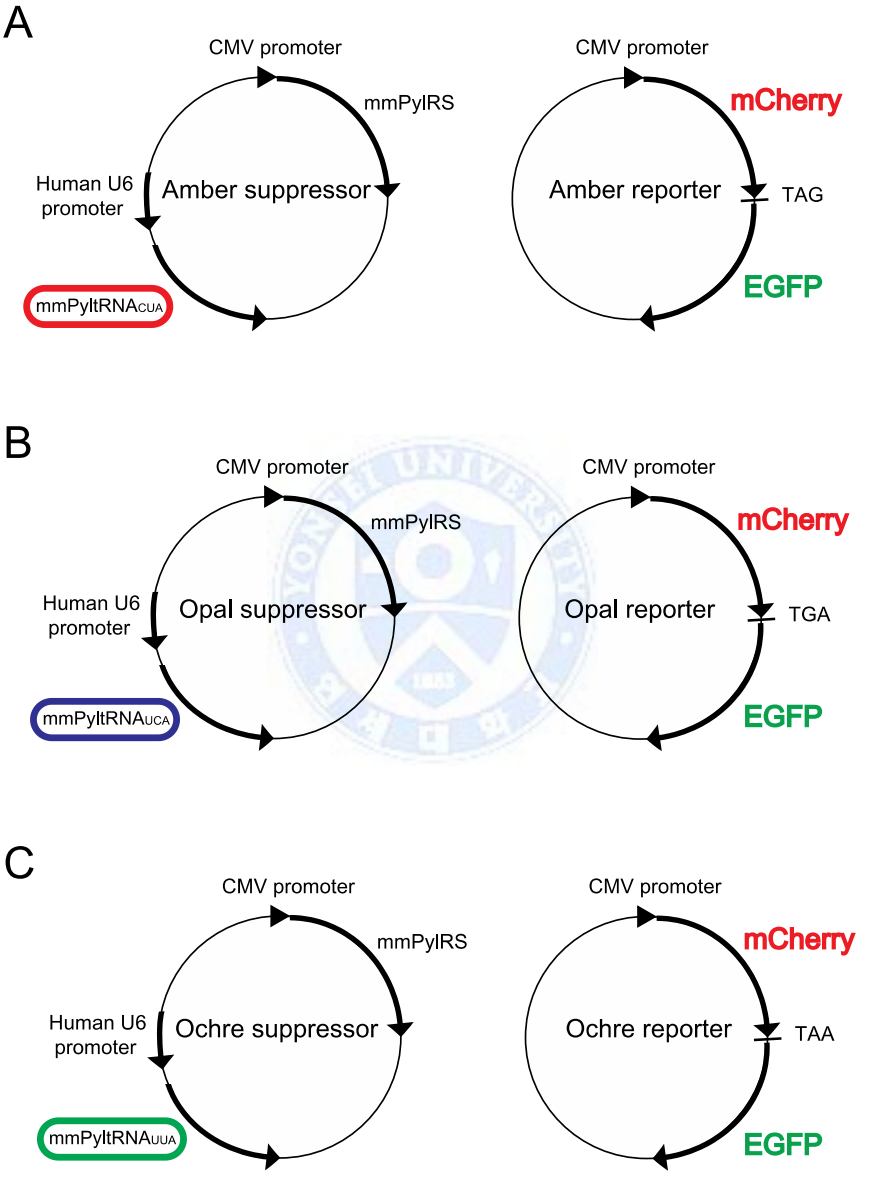
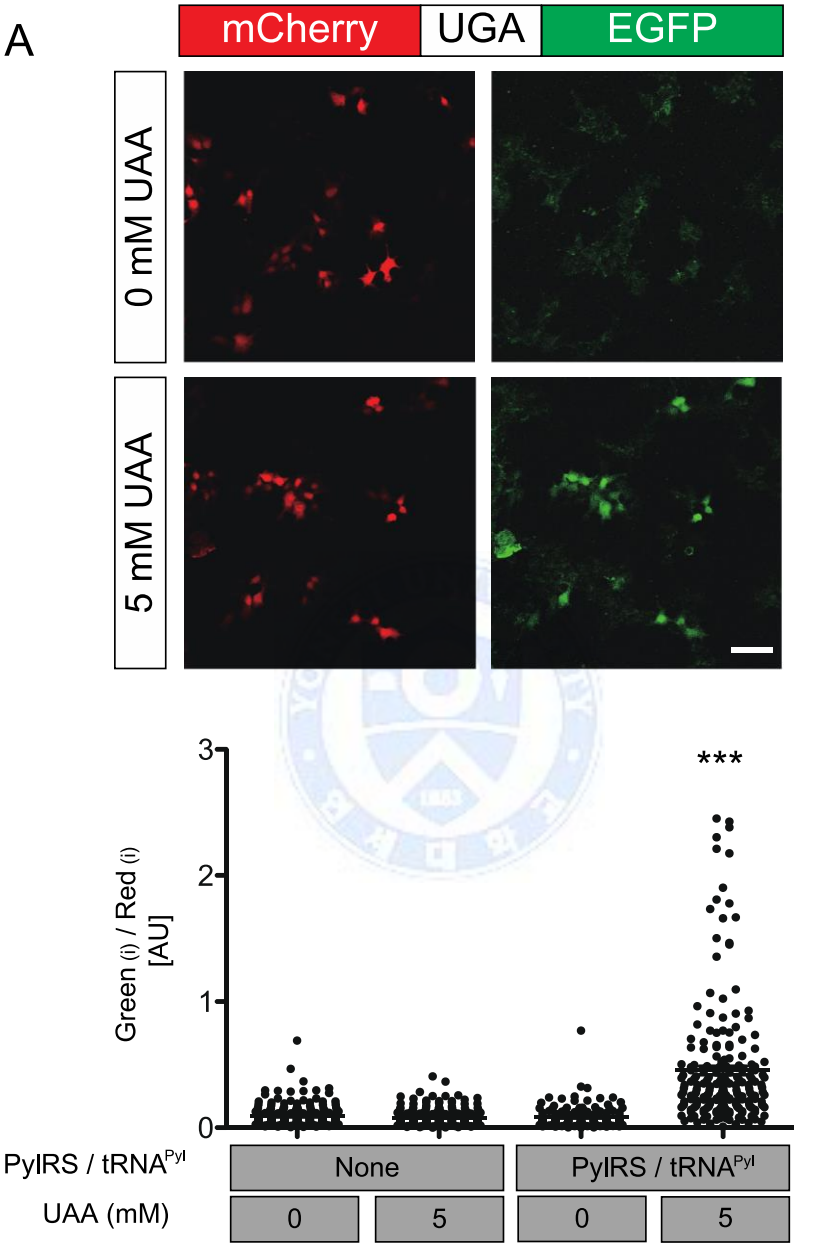


Figure 8. Point mutation of the anticodon in tRNA^{Pyl} to recognize and suppress other stop codons. (A) Anticodons of tRNA^{Pyl}s in opal and ochre suppressor were changed by the point mutation from the original anticodon of tRNA^{Pyl} in amber suppressor which recognizes only UAG stop codon (amber codon). (B) Mutation of tRNA^{Pyl}s in opal and ochre suppressor enabled to recognize UGA stop codon (opal codon) in opal reporter and (C) UAA stop codon (ochre codon) in ochre reporter.


A
mCherry
UGA
EGFP
0 mM UAA
5 mM UAA
Green (i) / Red (i) [AU]
3
2
1
0
***
PylRS / tRNAPyl
None
PylRS / tRNAPyl
UAA (mM)
0
5
0
5

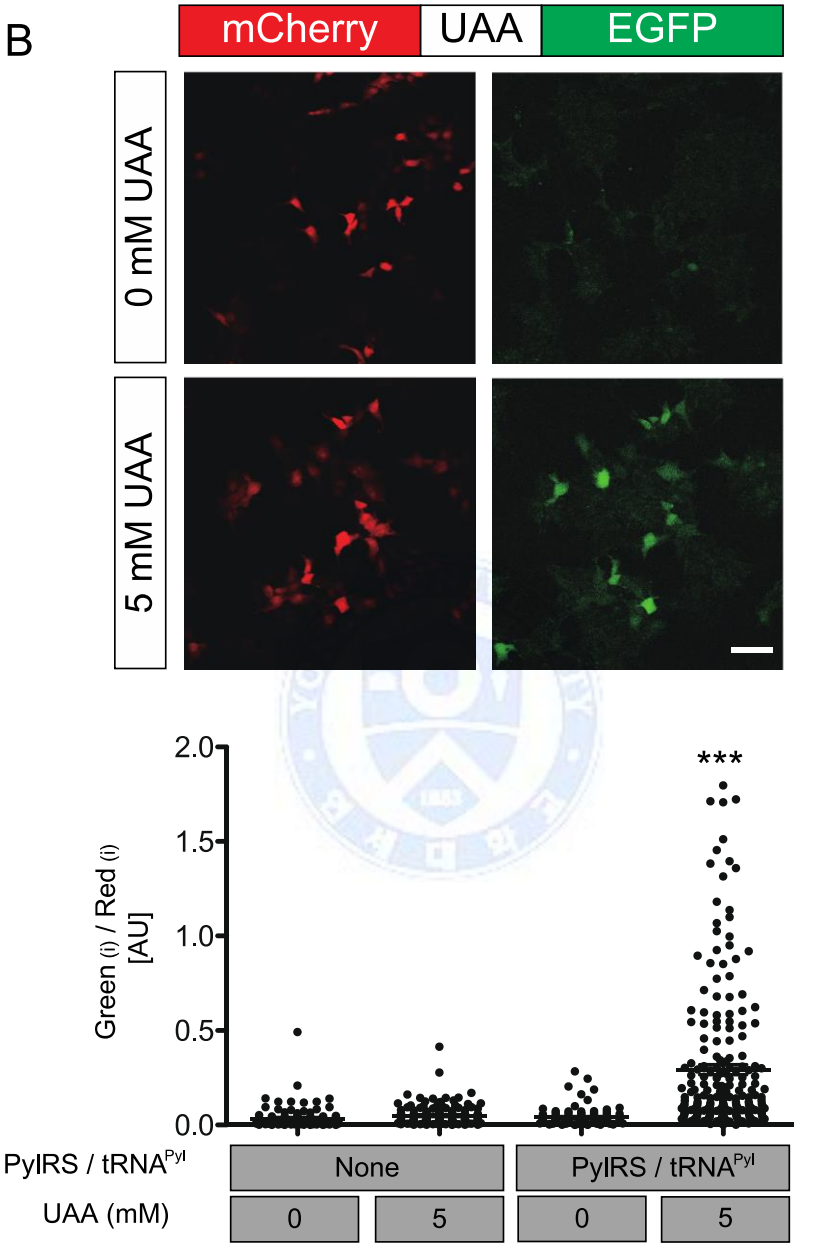
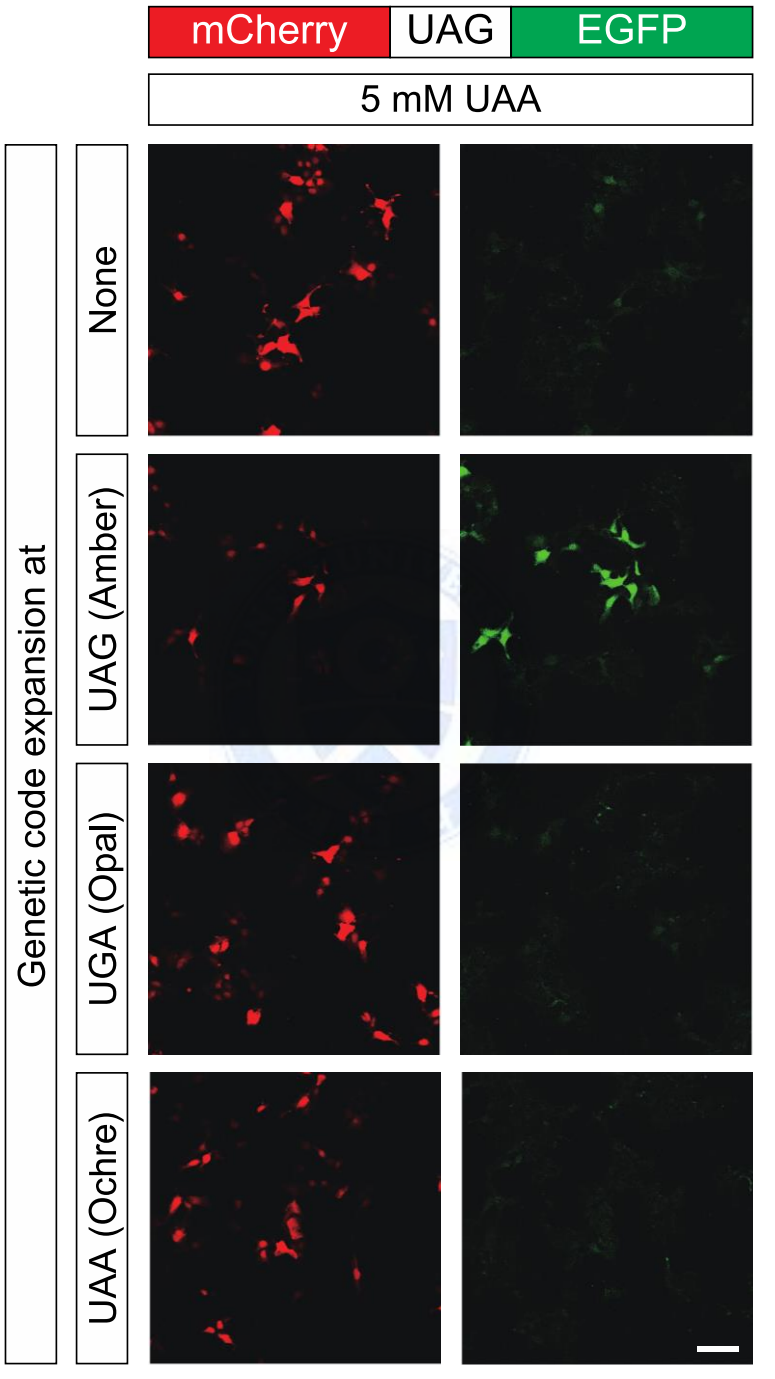

B
mCherry
UAA
EGFP
0 mM UAA
5 mM UAA
Green (i) / Red (i) [AU]
2.0
1.5
1.0
0.5
0.0
***
PylRS / tRNAPyl
None
PylRS / tRNAPyl
UAA (mM)
0
5
0
5

Figure 9. UAA-dependent suppression of the opal and ochre stop codons in genetic code-expanded HEK293T cells. When opal or ochre reporter and suppressor were transfected into the cells and UAAs were treated, genetic code expansion was occurred as in the amber stop codon suppression. (A) EGFP was expressed only in cells that have PylRS/$tRNA^{Pyl}$ and UAAs. UGA stop codon in the opal reporter was expanded to incorporate UAA into the polypeptides instead of completion of protein synthesis. (B) Genetic code expansion in the ochre reporter. UAA stop codon in ochre reporter was suppressed by ochre suppressor which has PylRS and $tRNA^{Pyl}$ to recognize and expand UAA stop codon. Statistical analysis was performed by one-way ANOVA followed by Dunnett's post analysis. UAA, unnatural amino acid. scale bar, 100 μm. AU, arbitrary unit. ***$p < 0.001$.

3. Anticodon-engineered $tRNA^{Pyl}$s specifically recognize their compatible cognate stop codons in mammalian cells

To confirm the specificity of PylRS/$tRNA^{Pyl}$, three stop codons were exposed to three differently anticodon-engineered $tRNA^{Pyl}$s. One of the three reporters where UAG, UGA or UAA stop codon is between the mCherry-EGFP fusion protein was transfected into HEK293T cells with three stop codon suppressors (amber, opal and ochre suppressor). And 5mM of UAA was treated into the cells and incubated for 24 hr. Results of this experiment showed that PylRS and $tRNA^{Pyl}$ were specifically expanded their compatible stop codons and UAA was incorporated into the polypeptides in HEK293T cells. For instance, UAG stop codon in amber reporter was expanded only by the amber suppressor which had $tRNA^{Pyl}$ (of which anticodon was CUA) and PylRS (Fig. 10A and B). HEK293T cells expressed EGFP only in amber reporter-suppressor transfected group. Other two stop codons in opal and ochre reporter were also expanded only by opal and ochre suppressor which implied point mutation of $tRNA^{Pyl}$ didn't change the specificity of it (Fig. 10B).

A


mCherry
UAG
EGFP
5 mM UAA
Genetic code expansion at
None
UAG (Amber)
UGA (Opal)
UAA (Ochre)

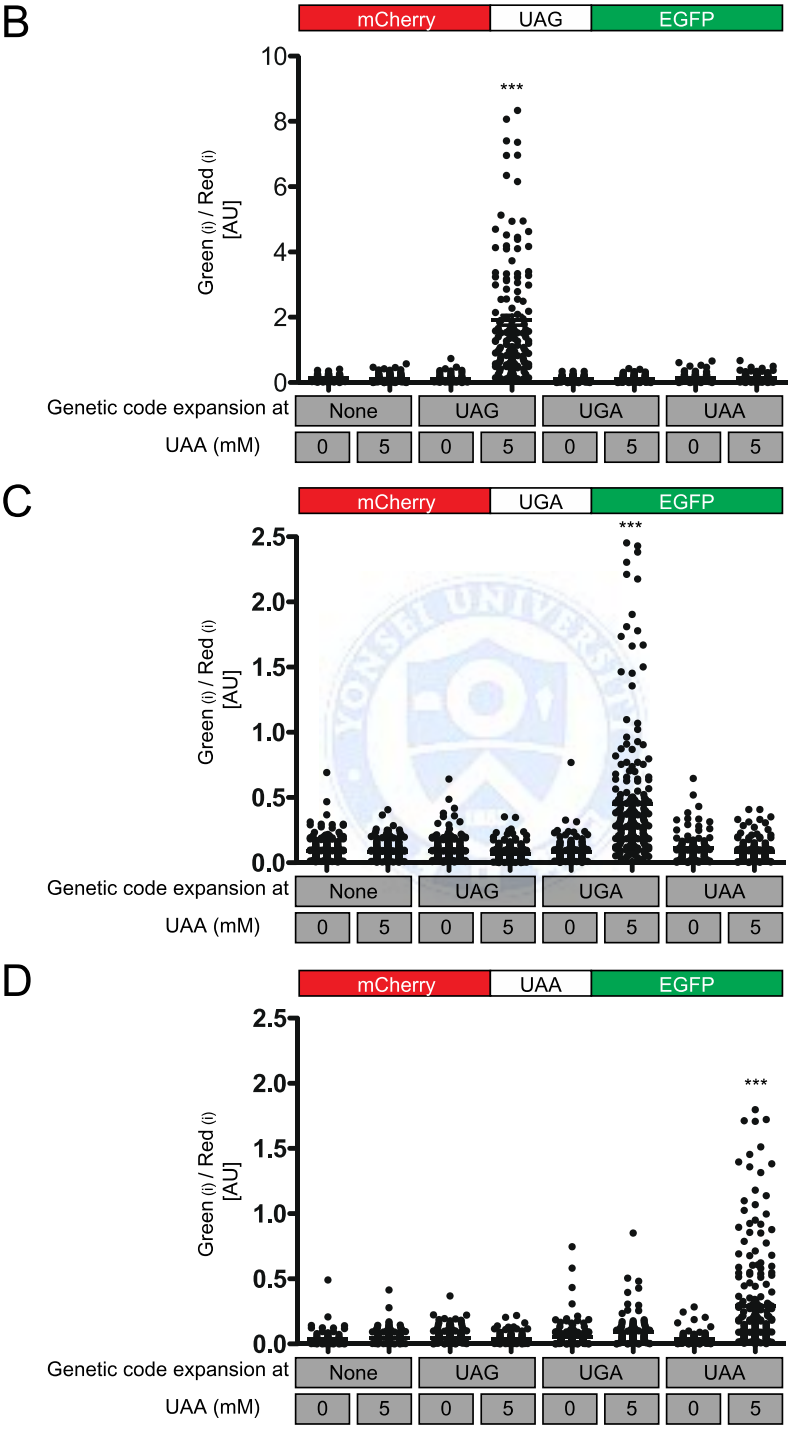

B
mCherry
UAG
EGFP
Green (i) / Red (i) [AU]
10
8
6
4
2
0
***
Genetic code expansion at
None
UAG
UGA
UAA
UAA (mM)
0
5
0
5
0
5
0
5
C
mCherry
UGA
EGFP
Green (i) / Red (i) [AU]
2.5
2.0
1.5
1.0
0.5
0.0
***
Genetic code expansion at
None
UAG
UGA
UAA
UAA (mM)
0
5
0
5
0
5
0
5
D
mCherry
UAA
EGFP
Green (i) / Red (i) [AU]
2.5
2.0
1.5
1.0
0.5
0.0
***
Genetic code expansion at
None
UAG
UGA
UAA
UAA (mM)
0
5
0
5
0
5
0
5

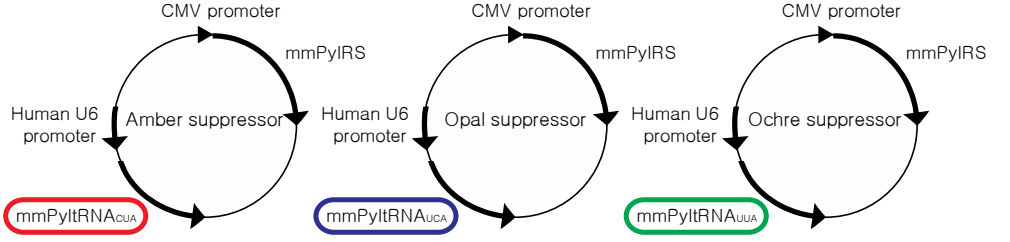
Figure 10. Specificity of differently anticodon-engineered tRNA^{Pyl}s in recognizing their cognate stop codons and site-specific UAA incorporation. (A) and (B) UAG stop codon in amber reporter was expanded only by the amber suppressor that had tRNA^{Pyl} to recognize amber codon (of which anticodon was CUA). mCherry-EGFP fusion protein was expressed only in amber reporter-suppressor group. (C) and (D) Mutation of the anticodon of tRNA^{Pyl} didn't destroy the specificity of tRNA^{Pyl}**.** Opal and Ochre suppressor expanded only their designated stop codons, UGA and UAA codon. UAA, unnatural amino acid. scale bar, 100 μm. AU, arbitrary unit. Statistical analysis was performed by one-way ANOVA followed by Dunnett's post analysis. ***$p < 0.001$.

4. Unnatural amino acids are site-specifically incorporated into all three stop codons by pyrrolysyl-tRNA synthetase and $tRNA^{Pyl}$

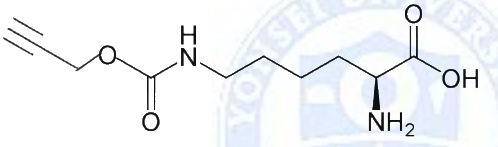
Previous results showed that PylRS and $tRNA^{Pyl}$ of *M. mazei* could function in the mammalian cell system with orthogonality and specificity. It conferred the possibility that genetic code expansion could be a useful tool as labeling proteins in mammalian cells. In this experiment, three different suppressors (a PylRS and three $tRNA^{Pyl}$s) used in previous experiments were transfected into HEK293T cells to expand the stop codons and label the proteins of the mammalian cells. PylRS would esterify the $tRNA^{Pyl}$ and its compatible amino acid. To label the proteins, chemically modified UAA instead of Pyl was treated into the cells (Fig. 11A and B). This UAA, also called as alkyne-lysine briefly, has an alkyne which is bio-orthogonal and can react with an azide and make a ring formation in the catalysis of copper. Chemical reaction between an alkyn and an azide is called as 'Copper-catalyzed azide-alkyne cycloaddition (CuAAC)'. Mechanism of the CuAAC is one of the commercially used reactions in the click chemistry (Fig. 11C). By this theory, I put green fluorescence tagged azide (BDP FL-azide) after treatment of alkyne lysine into the cells. BDP FL-azide reacted with alkyne which was incorporated into the polypeptides and made triazole. Eventually, proteins in genetic code expanded cell expressed green fluorescence. In this experiment, all stop codons were stochastically expanded in HEK293T cells by the PylRS/$tRNA^{Pyl}$ (Fig. 12). By the genetic code expansion in HEK293T cells, two possible theories proceed at stop codon. One is termination of the translation. In eukaryotes, eukaryotic translation termination factor 1 (eRF1) recognizes three stop codons. RF promotes the hydrolysis of the ester bond between the polypeptides and peptidyl site tRNA. The other is elongation of the translation. Because of the existence of PylRS/$tRNA^{Pyl}$, RF competes with aminoacyl-$tRNA^{Pyl}$. When the alkyne lysyl-$tRNA^{Pyl}$ defeats RF, alkyne lysine is incorporated into the nascent polypeptides and translation goes on.

Genetic code expansion at all three stop codons

A

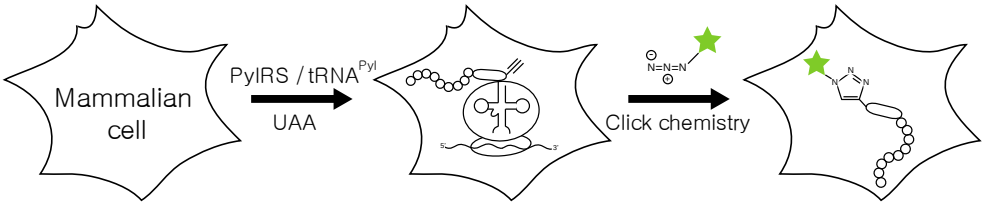


B



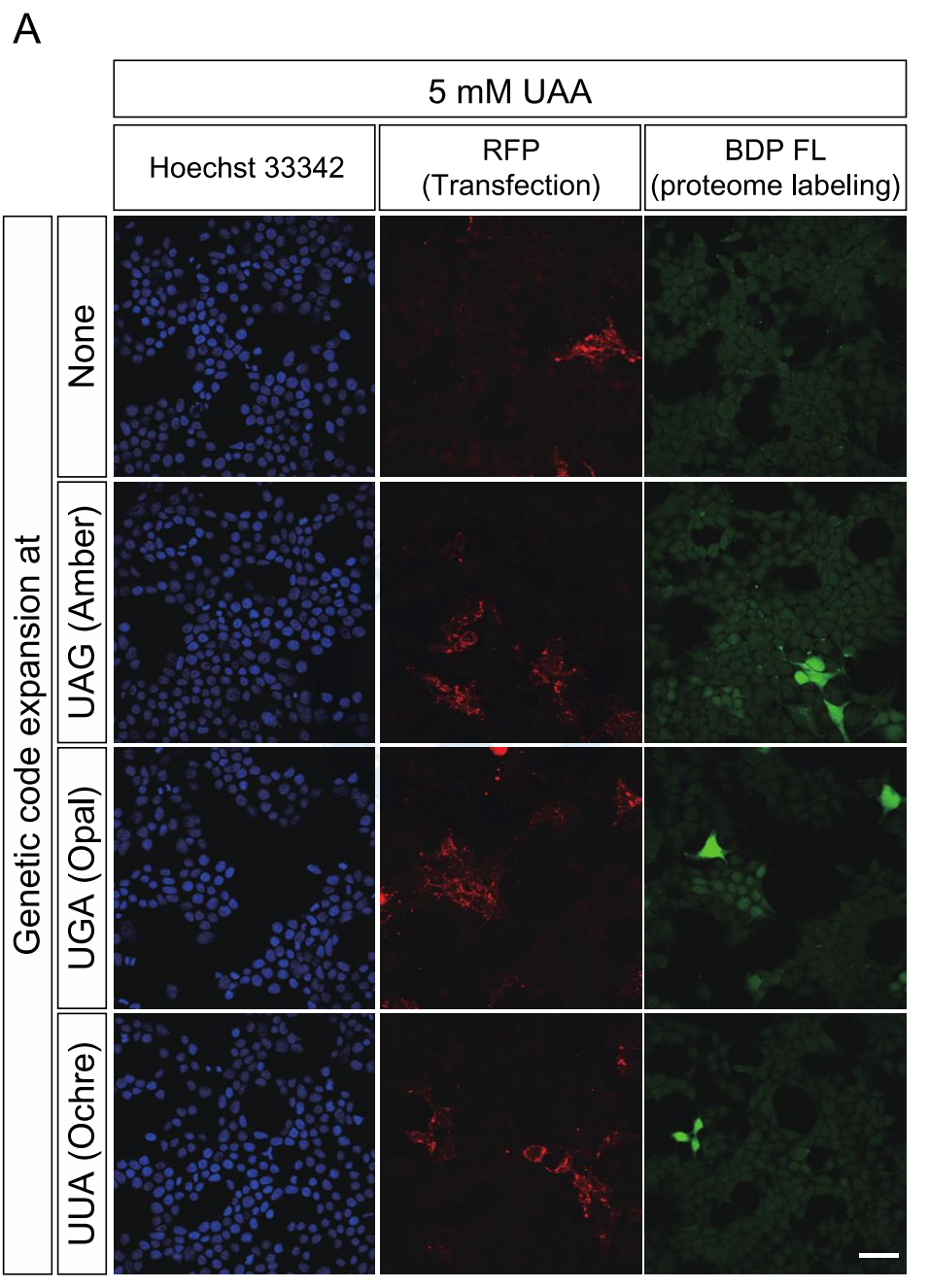
N6-[(2-propynyloxy)carbonyl]-L-lysine

C



Genetic code expansion by PylRS / $tRNA^{Pyl}$	Incorporation of UAA into a polypeptide	Visualization of a protein by click chemistry

Figure 11. Visualization of the genetic code expansion at stop codons and UAA incorporated proteins by click chemistry and fluorescence imaging. (A) Three stop codon suppressors were used to expand the stop codons and label the genetic code expanded proteins in the HEK293T cells. Each suppressor has PylRS and $tRNA^{Pyl}$ which has its designate anticodon about the stop codon. (B) N6-[(2-propynyloxy) carbonyl]-L-lysine, chemically modified UAA that is Pyl analog used to expand the stop codons and label proteins. (C) Click chemistry by CuAAC visualized UAA labeled proteins in genetic code expanded cells.


A
5 mM UAA
Hoechst 33342
RFP
(Transfection)
BDP FL
(proteome labeling)
Genetic code expansion at
None
UAG (Amber)
UGA (Opal)
UUA (Ochre)

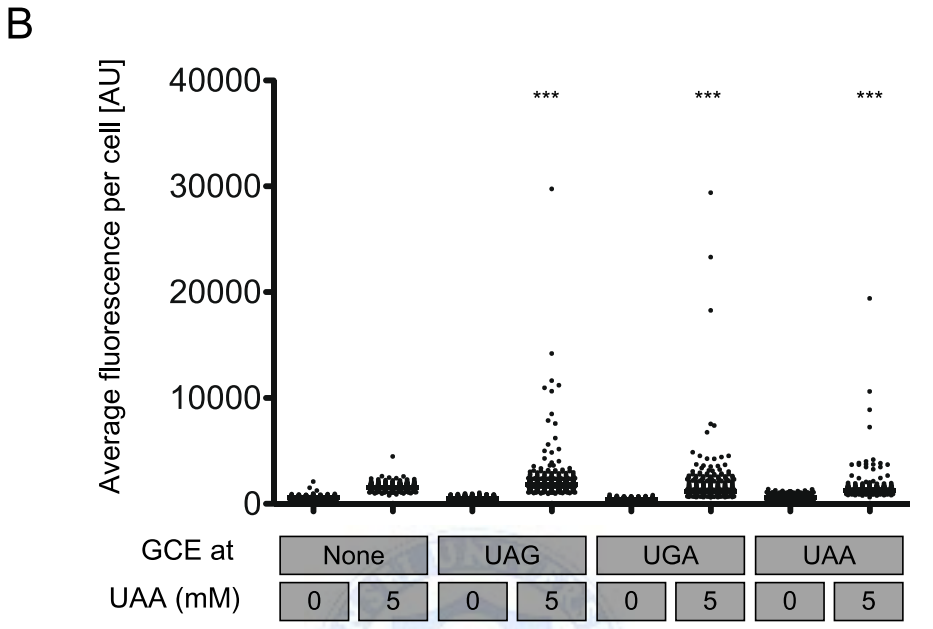


Figure 12. Quantitative analysis of site-specific UAA incorporation into different stop codons in genetic code-expanded HEK293T cells. Suppressor plasmid was transfected into HEK293T cells with *pCS2+GAP DsRed* plasmid which was for identification of transfected cell and efficiency of transfection. After, 5 mM of UAA was treated into the cells. Only the cells that had PylRS and $tRNA^{Pyl}$ could use UAA to label the nascent polypeptides. Genetic code expanded polypeptides in the cells were visualized by click chemistry expressing the green fluorescent signals which were from the BDP FL-azide. scale bar, 50 μm. UAA, unnatural amino acid. GCE, genetic code expansion. AU, arbitrary unit. Statistical analysis was performed by one-way ANOVA followed by Dunnett's post analysis. ***$p < 0.001$.

5. The anticodon of $tRNA^{Pyl}$ can be changed to recognize sense codons

Originally, PylRS/$tRNA^{Pyl}$ pair of *M. mazei* recognizes and expands the UAG stop codon. But the point mutation of anticodon of $tRNA^{Pyl}$ facilitated recognition and expansion of the rest two stop codons. Genetic code expansion at all three stop codons were more efficient than at UAG stop codon only. However, it had some disadvantages. First was high risk of incorrectly formed proteins. Because of the ongoing procedures of translation at the stop codon, eRF1 which recognizes stop codons and orders to releasing of the ribosomes from mRNAs can not act its role. Translation proceeds continuously. Incorrectly synthesized polypeptides could give adverse effects on the cell system. Second was the efficiency of the genetic code expansion at the stop codon. Stop codon exists only one per a polypeptide which means if genetic code expansion occurs it can be once in one polypeptide. Also, UAA labeled polypeptides can be degraded by proteolysis in the cell system because of its abnormal form. In this experiment to settle these problems, the anticodons of $tRNA^{Pyl}$s were point mutated to recognize the sense codons (Figure. 12A). 4 sense codons were selected. AUG, UUA, UUC and UUU which are sense codons for Methionine (Met), Leucine (Leu) and Phenylalanine (Phe). They are well used amino acids in *Xenopus tropicalis* which will be used *in vivo* model in this study. I transfected these suppressors into HEK293T cells. 5 mM of alkyne lysine was treated into the cells and click chemistry was done (Figure. 12B). Among 4 sense codon suppressed groups, Met suppression was the highest efficient (Figure. 13). Phe suppression that genetic code expansion at UUC and UUU showed quite low efficiency of labeling. Although green fluorescent labeling in the Phe (UUU) suppressor group was detected in the imaging analysis compared to the control group, statistical analysis showed non-significant consequence.

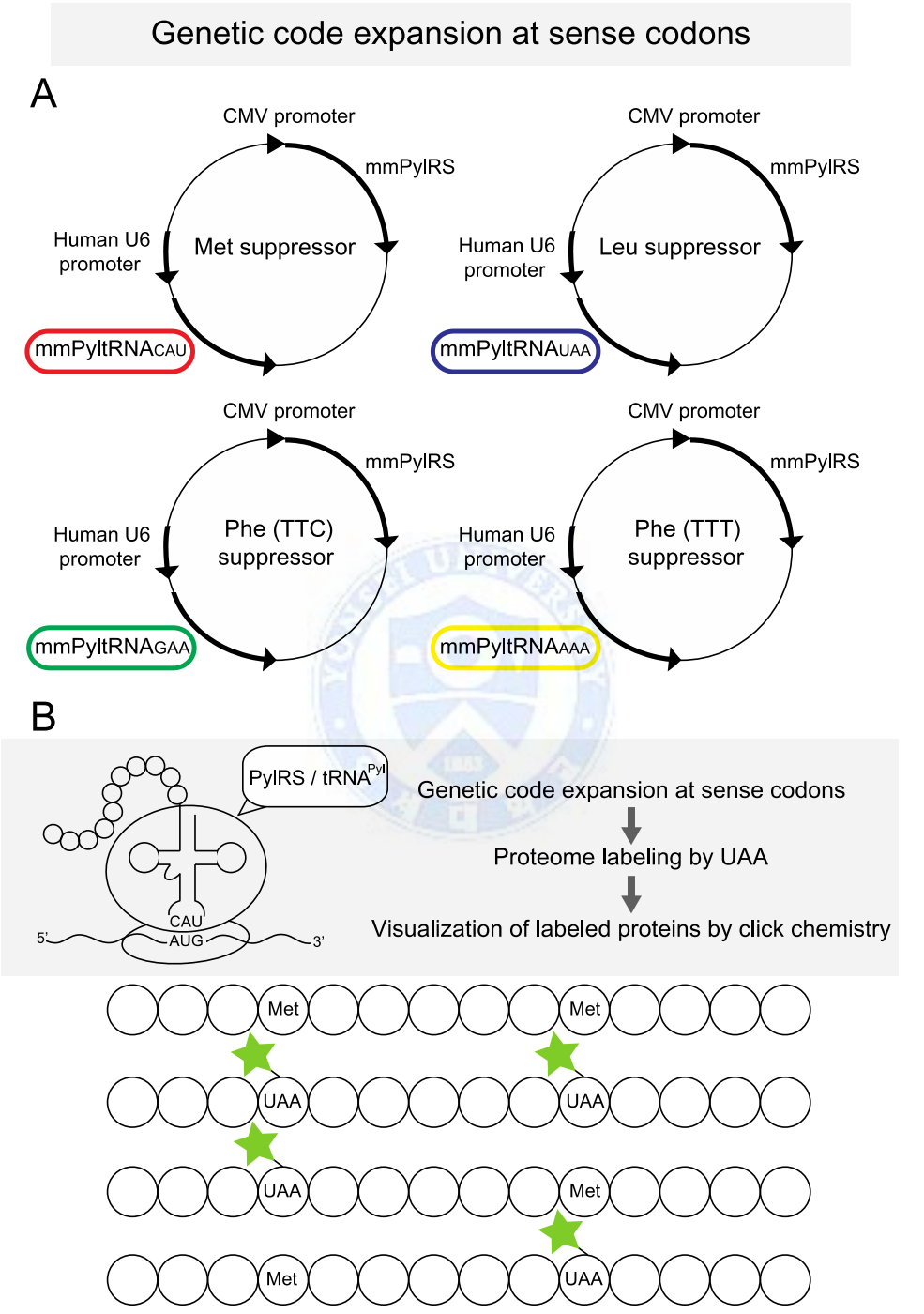
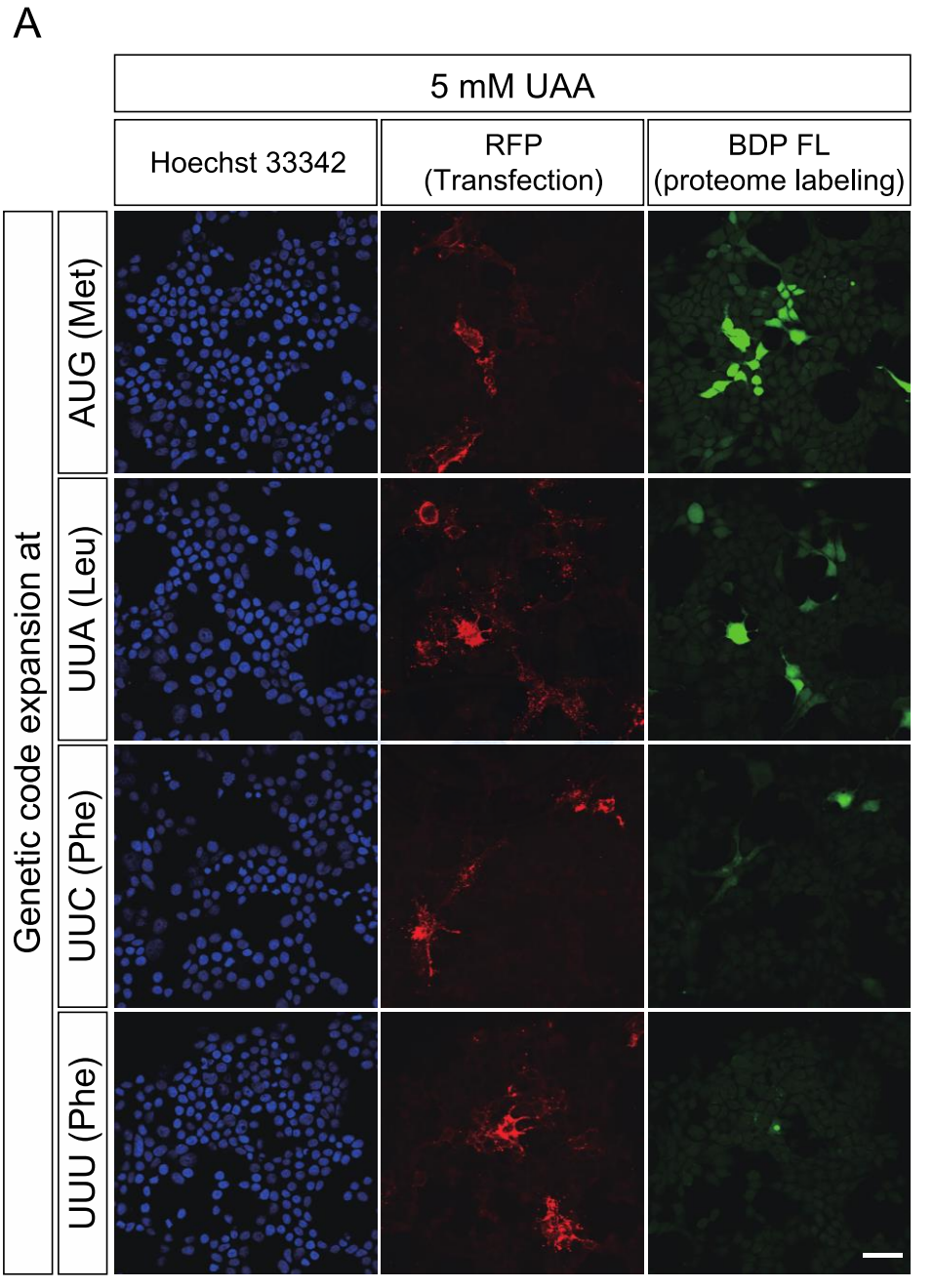

Genetic code expansion at sense codons
A
CMV promoter
mmPylRS
Human U6 promoter
Met suppressor
mmPylRNACAU
CMV promoter
mmPylRS
Human U6 promoter
Leu suppressor
mmPylRNAUAA
CMV promoter
mmPylRS
Human U6 promoter
Phe (TTC) suppressor
mmPylRNAGAA
CMV promoter
mmPylRS
Human U6 promoter
Phe (TTT) suppressor
mmPylRNAAAA
B
PylRS / tRNAPyl
CAU
AUG
5'
3'
Genetic code expansion at sense codons
Proteome labeling by UAA
Visualization of labeled proteins by click chemistry
Met
Met
UAA
UAA
UAA
Met
Met
UAA

Figure 13. Site-specific incorporation of UAA into the AUG sense codon in genetic code-expanded HEK293T cells. (A) Sense codon suppressors have $tRNA^{Pyl}$s which recognize the codon of Met, Leu or Phe. When they were transfected into the cells, PylRS would esterify the Pyl analog to the $tRNA^{Pyl}$ and aminoacyl-$tRNA^{Pyl}$ would recognize the sense codon. (B) when the suppressor was transfected into the mammalian cells, it would expand the sense codon. Some sense codons would incorporate the natural amino acids but others would incorporate the UAAs into the polypeptides. Alkyne in UAA of polypeptides would react with fluorescence tagged azide (BDP FL-azide) and make triazole.


A
5 mM UAA
Hoechst 33342
RFP
(Transfection)
BDP FL
(proteome labeling)
Genetic code expansion at
AUG (Met)
UUA (Leu)
UUC (Phe)
UUU (Phe)

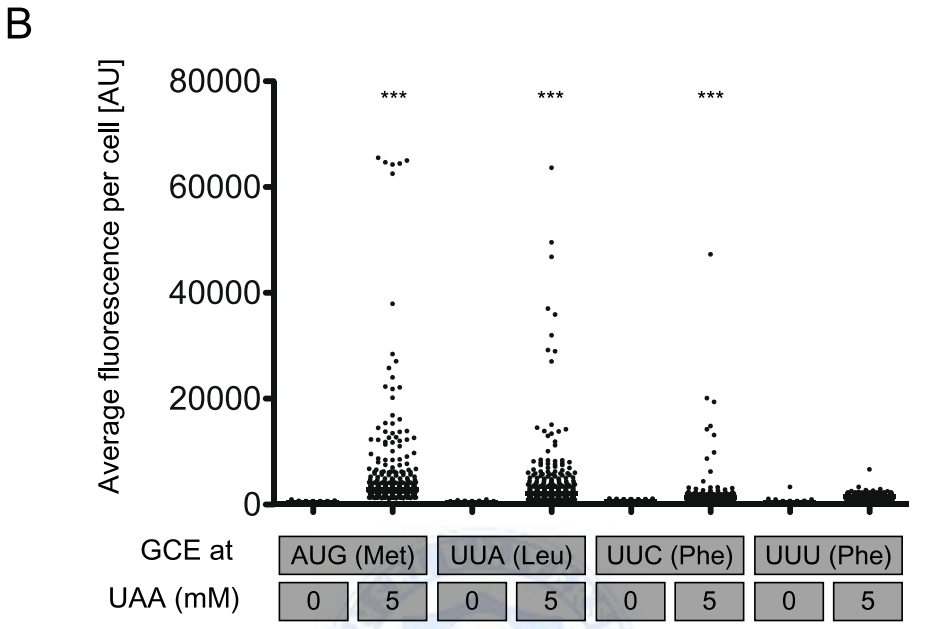


Figure 14. Visualization of the genetic code expansion at sense codons and UAA incorporated proteins by click chemistry and fluorescence imaging. Point mutated anticodon of $tRNA^{Pyl}$ interacted with some sense codons, Met, Leu and Phe in HEK293T cells. Met suppression was the most efficient proteome labeling by genetic code expansion. Phe (UUU) suppression showed poor labeling efficiency. scale bar, 50 μm. UAA, unnatural amino acid. GCE, genetic code expansion. AU, arbitrary unit. Statistical analysis was performed by one-way ANOVA followed by Dunnett's post analysis. ***$p < 0.001$.

6. Co-expression of differently anticodon-engineered $tRNA^{Pyl}$s allows more efficient incorporation of unnatural amino acids and proteome labeling

To elevate the efficiency of genetic code expansion and incorporation of the UAA, the $tRNA^{Pyl}$in the suppressor was inserted two, three or four times more than it originally had (Fig. 15). For example, Amber & Opal suppressor is a DNA plasmid that has encodes for 2 different $tRNA^{Pyl}$s (which recognized UAG or UGA stop codon) and PylRS. Each $tRNA^{Pyl}$ has different anticodons to UAG and UGA stop codon. In this experiment, I inserted multiple $tRNA^{Pyl}$s which had differently engineered-anticodons into a DNA plasmid (Suppressor). I transfected several suppressors that have differently engineered- $tRNA^{Pyl}$s into HEK293T cells and compared the efficiency of proteome labeling by PylRS/$tRNA^{Pyl}$. In the stop codon suppression, 3 different suppressors were transfected into each experimental group (Fig. 16A and C). Suppressors had a single $tRNA^{Pyl}$ (anticodon was CUA), two different $tRNA^{Pyl}$s (anticodons were CUA and UCA) or three different $tRNA^{Pyl}$s (anticodons were CUA, UCA and UUA). Statistical results showed that two or three multiple $tRNA^{Pyl}$s included suppressors more efficiently expressed the green fluorescence than a single $tRNA^{Pyl}$ included suppressor which implied more effective labeling of proteins was done in the multiple $tRNA^{Pyl}$s-suppressors. Sense codon suppression also showed corresponding outcome as the results of stop codon suppression. I transfected sense codon suppressors that had a singe $tRNA^{Pyl}$ (CAU), two different $tRNA^{Pyl}$s (CAU and UAA) and four different $tRNA^{Pyl}$s (CAU, UAA, GAA and AAA) into HEK293T cells (Fig. 15B). Images and graphs represented that multiple $tRNA^{Pyl}$s-suppressors expanded the sense codons more efficiently than a single $tRNA^{Pyl}$-suppressor (Fig. 16B and D). These results indicated that tandem expression of $tRNA^{Pyl}$s of the suppressors which had multiple $tRNA^{Pyl}$s. And co-expression of different $tRNA^{Pyl}$s of the suppressors labeled more proteins than expression of a single $tRNA^{Pyl}$ of the suppressor.

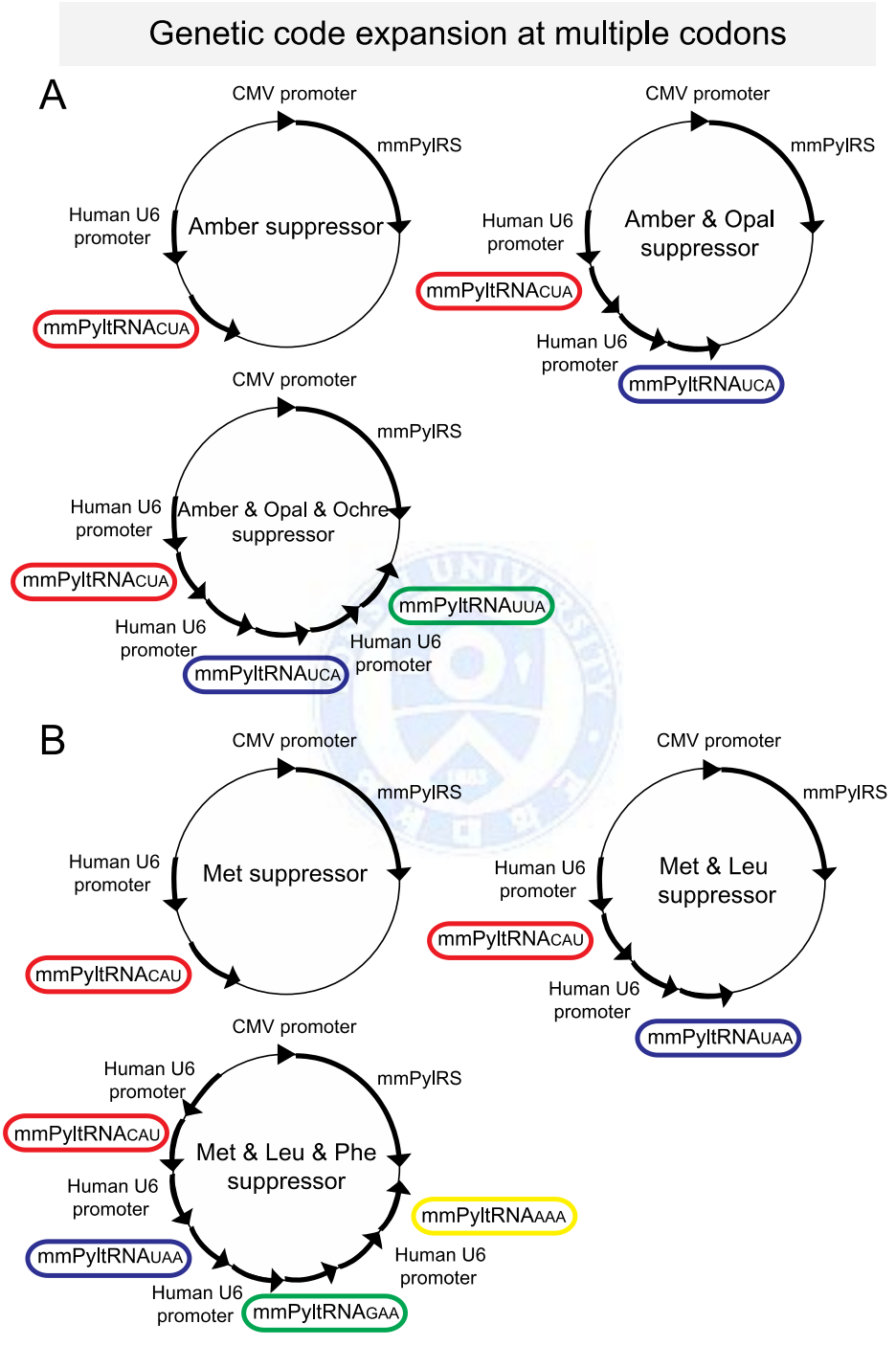
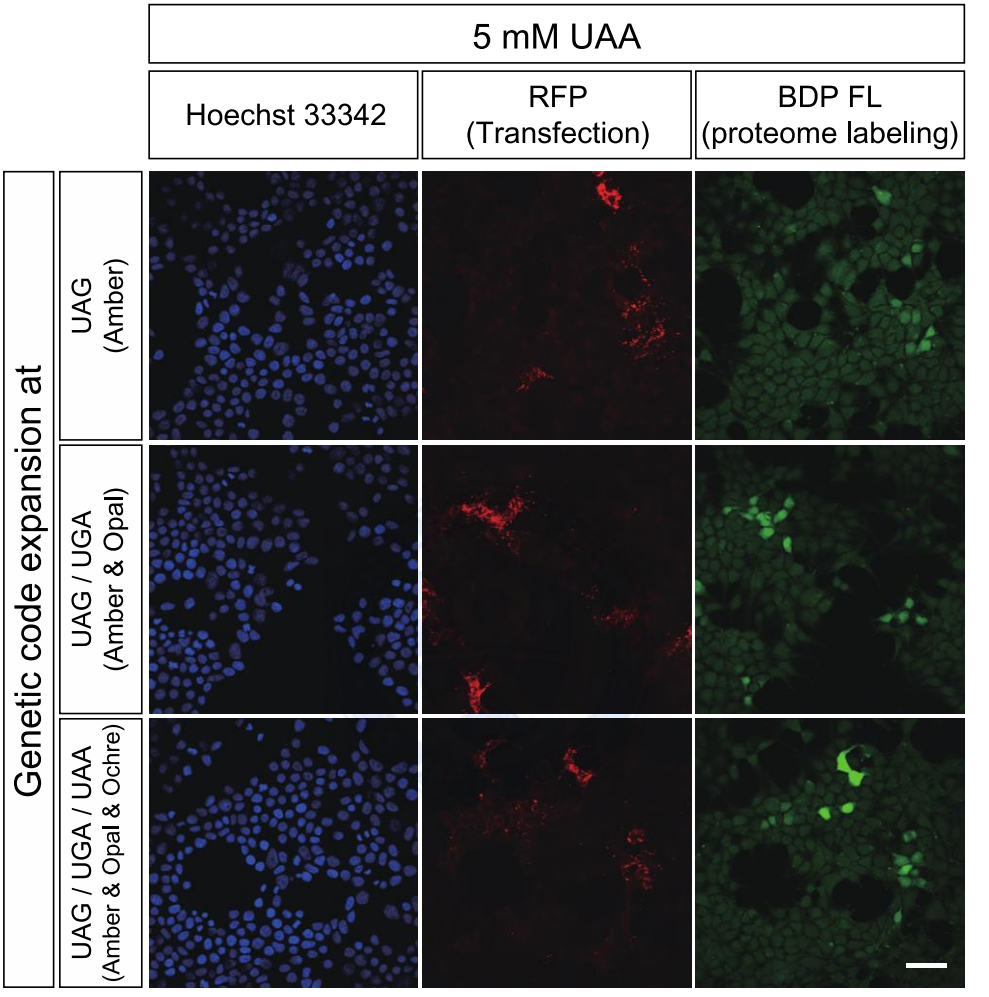
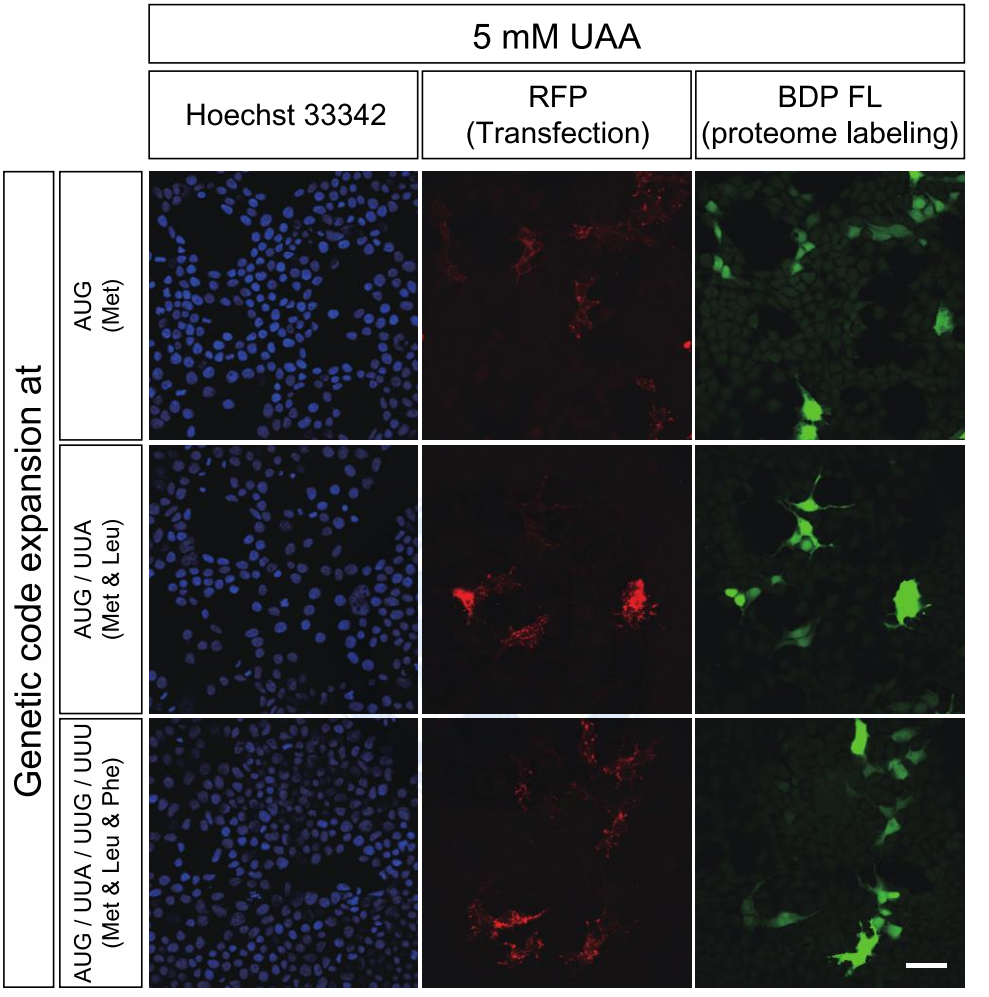

Genetic code expansion at multiple codons
A
CMV promoter
mmPylRS
Human U6 promoter
Amber suppressor
mmPylt RNACUA
CMV promoter
mmPylRS
Human U6 promoter
Amber & Opal suppressor
mmPylt RNACUA
Human U6 promoter
mmPylt RNAUCA
CMV promoter
mmPylRS
Human U6 promoter
Amber & Opal & Ochre suppressor
mmPylt RNACUA
mmPylt RNAUUA
Human U6 promoter
Human U6 promoter
mmPylt RNAUCA
B
CMV promoter
mmPylRS
Human U6 promoter
Met suppressor
mmPylt RNACAU
CMV promoter
mmPylRS
Human U6 promoter
Met & Leu suppressor
mmPylt RNACAU
Human U6 promoter
mmPylt RNAUAA
CMV promoter
Human U6 promoter
mmPylRS
mmPylt RNACAU
Met & Leu & Phe suppressor
Human U6 promoter
mmPylt RNAAAA
mmPylt RNAUAA
Human U6 promoter
Human U6 promoter
mmPylt RNAGAA

Figure 15. Incorporation of UAA into multiple sites using co-expression of different tRNA^{Pyl}s. (A) Amber & Opal suppressor and Amber & Opal & Ochre suppressor have more tRNA^{Pyl}s than Amber suppressor has. Amber & Opal & Ochre suppressor has all tRNA^{Pyl}s to all three stop codons which can hypothetically suppress all stop codons in the cells. (B) Met & Leu suppressor and Met & Leu & Phe (UUC and UUU) suppressor have more tRNA^{Pyl}s than Met suppressor has.

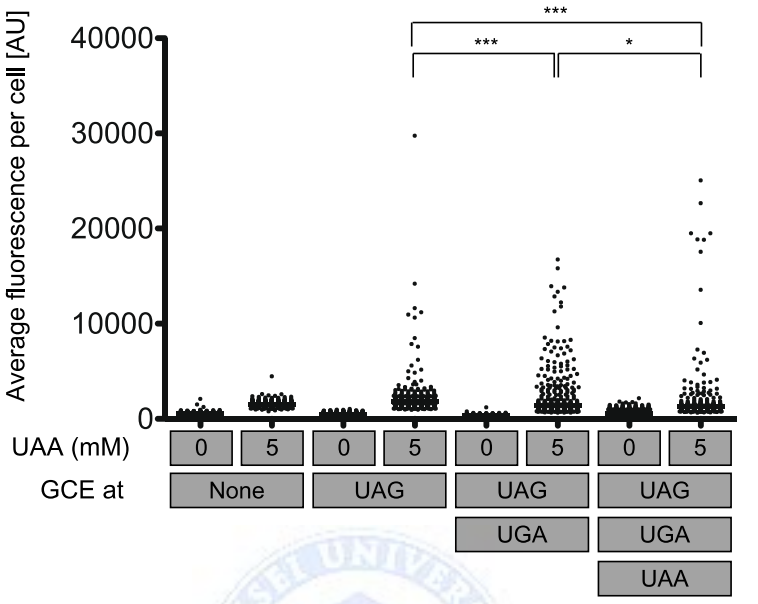
A


5 mM UAA
Hoechst 33342
RFP
(Transfection)
BDP FL
(proteome labeling)
Genetic code expansion at
UAG
(Amber)
UAG / UGA
(Amber & Opal)
UAG / UGA / UAA
(Amber & Opal & Ochre)

B


5 mM UAA
Hoechst 33342
RFP
(Transfection)
BDP FL
(proteome labeling)
Genetic code expansion at
AUG
(Met)
AUG / UUA
(Met & Leu)
AUG / UUA / UUG / UUU
(Met & Leu & Phe)

C


Average fluorescence per cell [AU]
40000
30000
20000
10000
0
***
***
*
UAA (mM)
0
5
0
5
0
5
0
5
GCE at
None
UAG
UAG
UAG
UGA
UGA
UAA

D

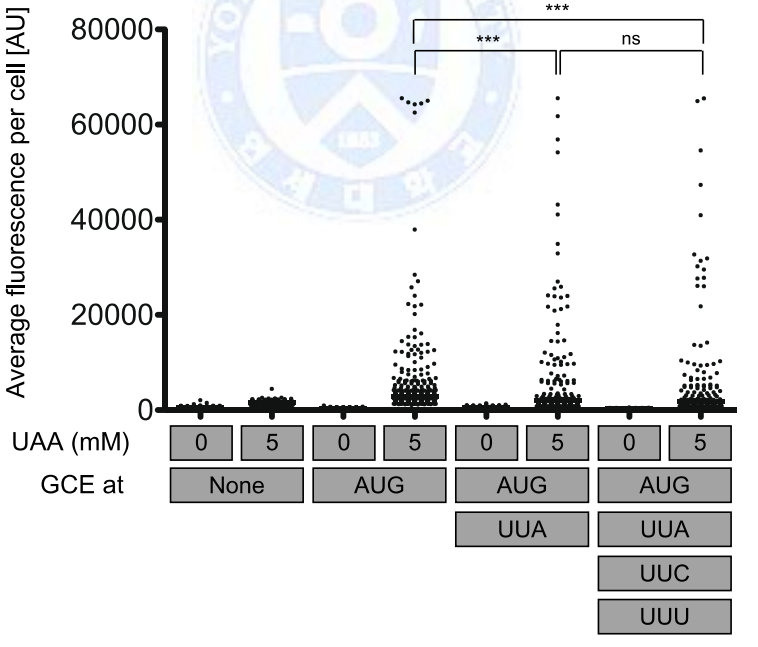

Average fluorescence per cell [AU]
80000
60000
40000
20000
0
***
***
ns
UAA (mM)
0
5
0
5
0
5
0
5
GCE at
None
AUG
AUG
AUG
UUA
UUA
UUC
UUU

Figure 16. Additive effect of parallel expression of different tRNA^{Pyl}s in the proteome labeling in genetic code-expanded HEK293T cells. Multiple tRNA^{Pyl}s in the suppressors more efficiently suppressed than a single tRNA^{Pyl} in the suppressor. Through the statistical analysis, both results in stop and sense codon suppression showed significant increase in multiple suppression than a single suppression. Scale bar, 50 μm. UAA, unnatural amino acid. GCE, genetic code expansion. AU, arbitrary unit. Statistical analysis was performed by one-way ANOVA followed by Bonferroni's post analysis. Statistical significance was defined as ns, not significant, * $p \leq 0.05$, *** $p \leq 0.0001$.

7. Efficacy of proteome labeling by genetic code expansion is dependent on the concentration of unnatural amino acids

When the genetic code is expanded at the stop or sense codon in the mammalian cells by PylRS and $tRNA^{Pyl}$, the 22^{nd} amino acid, Pyl, can be incorporated into the nascent polypeptides. To label and detect the polypeptides, I added to the cells not Pyl but Pyl analog which cell could misunderstand it as Pyl and use it. Pyl analog I used in this study was chemically modified UAA tagged with an alkyne that reacts with an azide in the click chemistry (Fig. 2). To adjust the optimal concentration of UAA to the cells and confirm the lowest concentration of UAA when labeling the proteome of the cells by the genetic code expansion, various concentrations of UAA were tested. HEK293T cells were transfected with Amber & Opal & Ochre suppressor (Fig. 15). After transfection, 12 different concentrations of UAA were treated into each cell group (Fig. 17). UAAs used in this experiment were lysed in water. Concentration of the UAA stock solution was 130 mM. It was diluted by DMEM-low glucose media with 10 % FBS and 1 X antimycotic antibiotic. However, the addition of lysed UAAs into the culture media changed the pH higher than pH 7. To adjust the pH, 5 M HCl was added to the media. Duration of UAA was 10 hr and media was changed to wash the nonspecific UAAs in the cell which didn't participate in proteome labeling. Cells were incubated in UAA-free media for 14 hr. Statistical analysis indicated that treatment of more than 0.75 mM of UAA into the genetic code expanded cells showed significant result compared to negative control (Fig. 18). Treatment of 10 mM of UAA into the cells expressed substantial green fluorescence, meaning high efficiency of proteome labeling. But the signal to ratio also increased with the green fluorescent signals. These results provided that significant concentration of UAA when labeling the proteins is above 0.75 mM of UAA. But the images of confocal microscope detected the signal even at 0.5 mM of UAA treated group. The optimal concentration of UAA was 5 mM in this study. Signal to ratio noise at this point wasn't too high and

cell viability was normal. So I applied to treat 5 mM of UAA in rest experiments in this study.

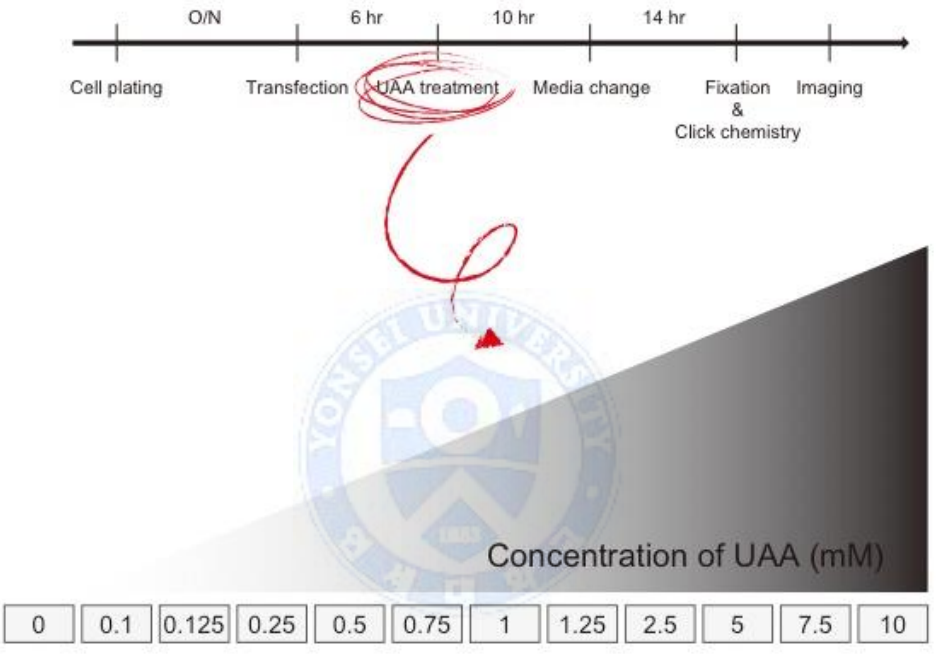
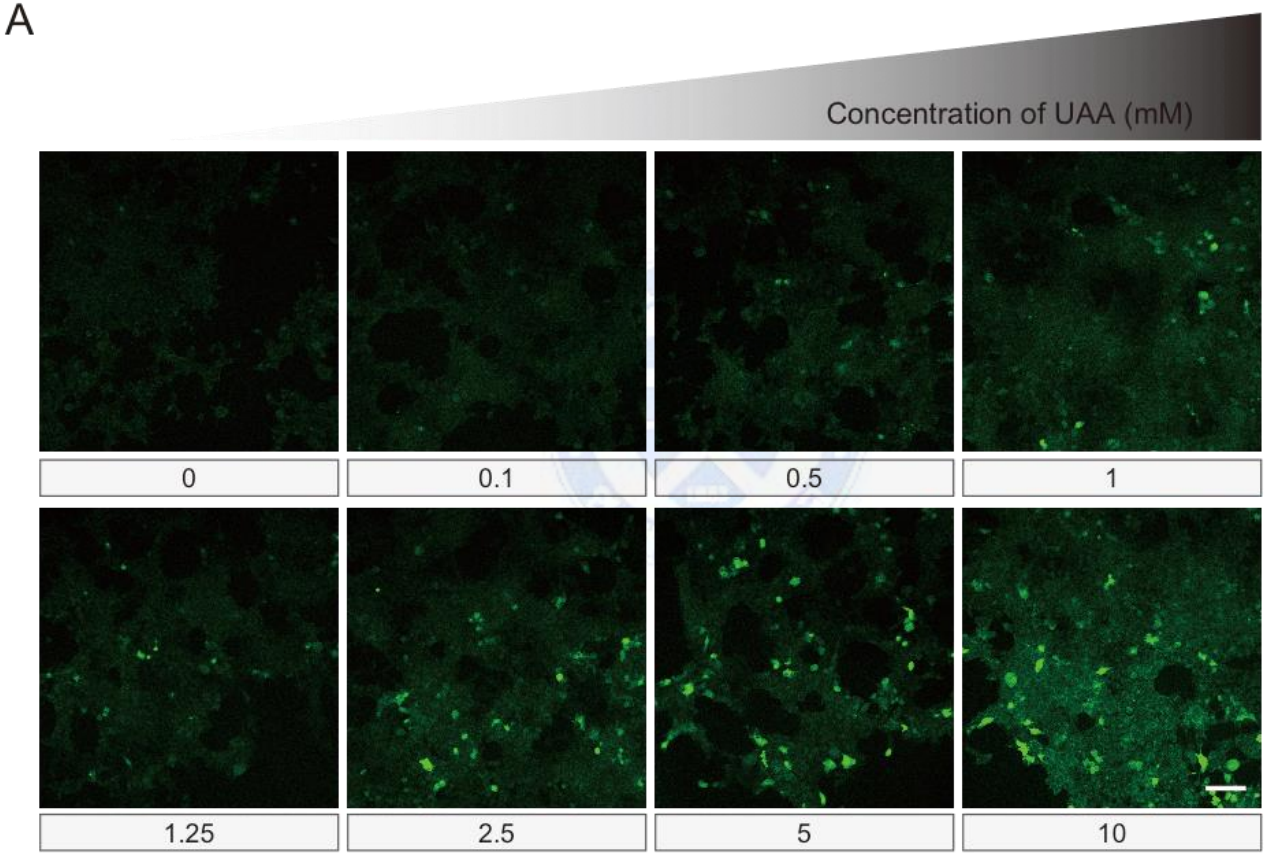


Figure 17. Experimental scheme to determine the concentration of UAA in genetic code expansion-based proteome labeling. After transfection of Amber & Opal & Ochre suppressor which has $tRNA^{Pyl}$s about all three stop codons into the HEK293T cells, 0, 0.1, 0.125, 0.5, 1, 1.25, 2.5, 5, 7.5 and 10 mM of UAA were treated into the cells for 10 hr.


A
Concentration of UAA (mM)
0
0.1
0.5
1
1.25
2.5
5
10

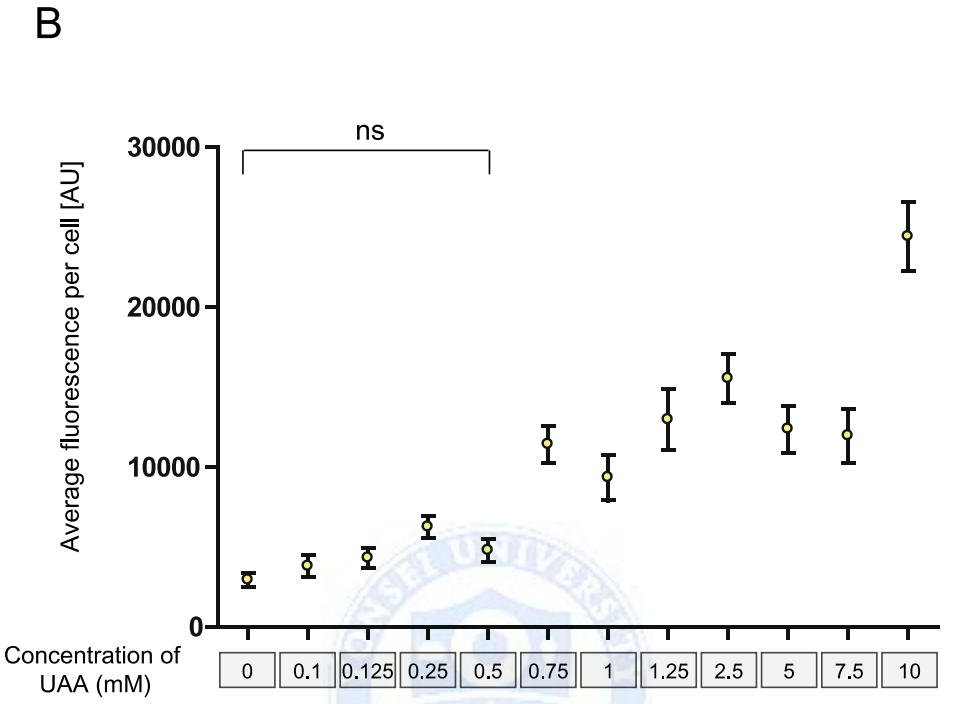


Figure 18. UAA concentration dependency of proteome labeling by genetic code expansion. As the concentration of UAAs increased, green fluorescent intensity also increased. Statistical analysis showed that the signals below 0.5 mM of UAA treated groups were not significant as compared to 0 mM of UAA treated group. Scale bar, 100 μm. UAA, unnatural amino acid. AU, arbitrary unit. Statistical analysis was performed by one-way ANOVA followed by Bonferroni's post analysis. Statistical significance was defined as ns, not significant.

8. Efficacy of proteome labeling by genetic code expansion is dependent on the duration of unnatural amino acid treatment

Not only the concentration of UAA but also the duration of them effects the efficiency of proteome labeling. Amber & Opal & Ochre suppressor was transfected into the cells as the previous experiment (Fig. 17). And 5 mM of UAA was treated into HEK293T cells at different times. The longest time was 32 hr and the shortest was 1 hr. After treatment of UAA, media in all samples were changed to UAA-free media and incubated for 14 hr. Statistically, green fluorescent expressions of 1 and 2 hr duration of UAAs into the cells were not significantly different as the negative control which wasn't treated with UAAs (Fig. 20). These are result from the washing time after the treatment. During this time, there were no UAAs in the media which meant no proteome labeling. So previously UAA labeled proteins had possibility to degraded or dispersed into another region in or outside of the cells. Considering the confluency of the cells as time goes by and the efficiency of proteome labeling, I selected the optimal time of duration of UAAs about 10 or 14 hr in the study.

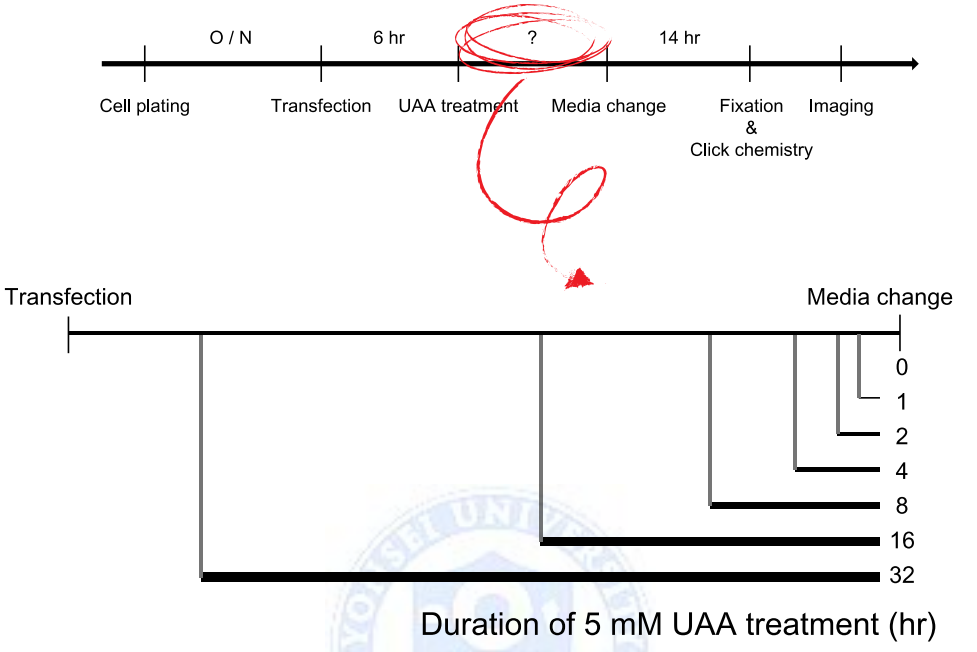
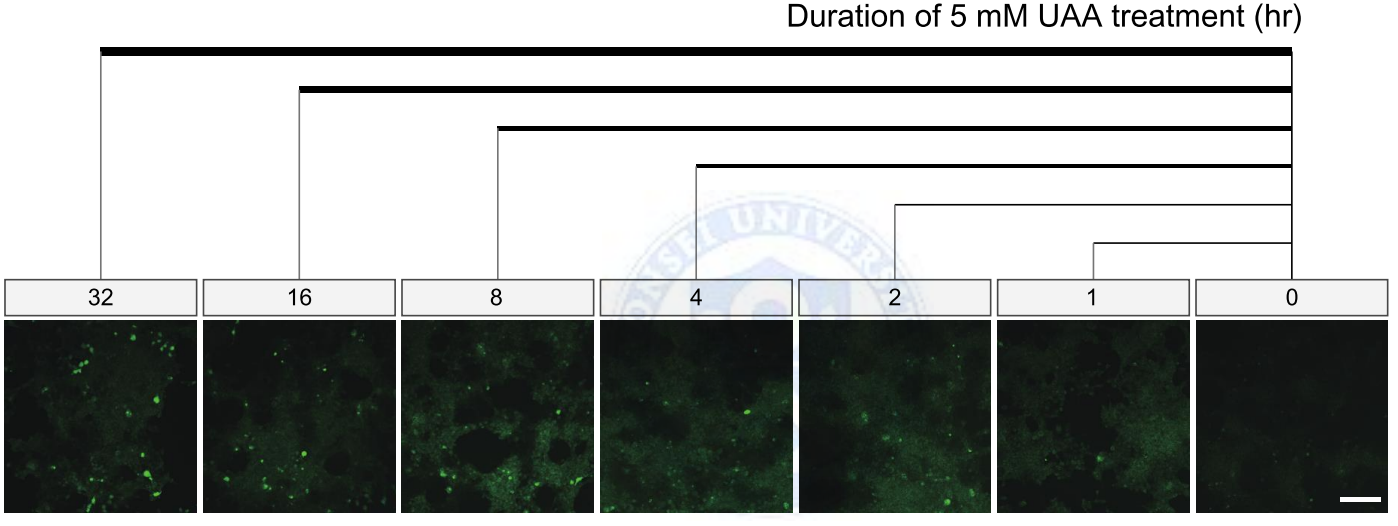


Figure 19. Experimental scheme to determine the time course of UAA in genetic code expansion-based proteome labeling. After transfection of Amber & Opal & Ochre suppressor which has $tRNA^{Pyl}$s about all three stop codons into HEK293T cells, 5 mM of UAA was treated into the cells for 1, 2, 4, 8, 16 and 32 hr.

A



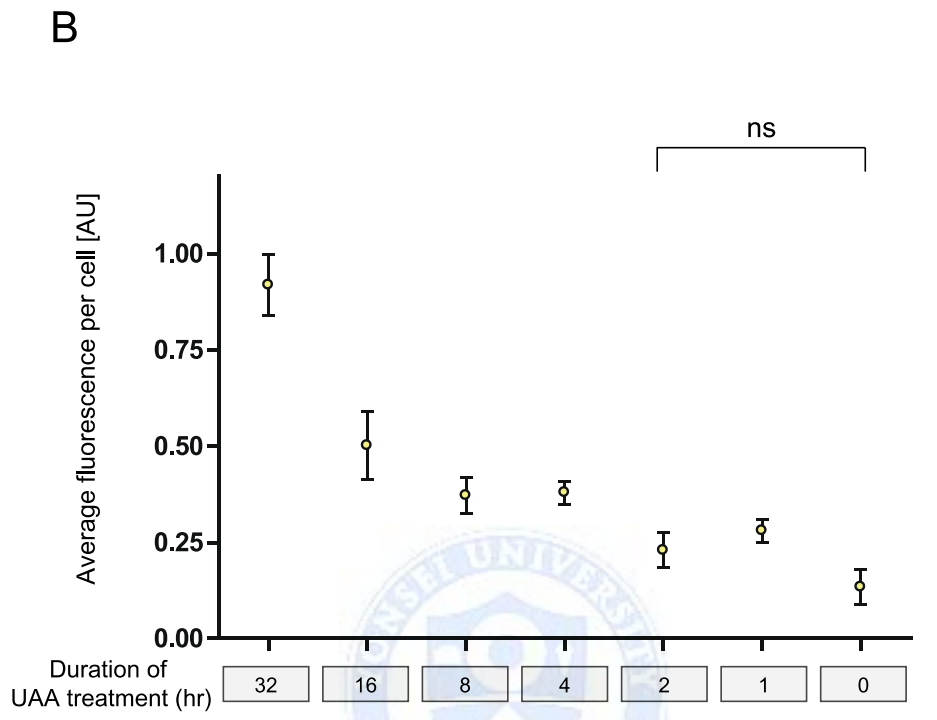


Figure 20. Time dependency of proteome labeling by genetic code expansion. As the duration time of UAA increased, green fluorescent intensity also increased. Statistical analysis showed that the signals below 2 hr of duration of UAA treated groups were not significant as compared to untreated UAA group. Scale bar, 100 μm. UAA, unnatural amino acid. AU, arbitrary unit. Statistical analysis was performed by one-way ANOVA followed by Bonferroni's post analysis. Statistical significance was defined as ns, not significant.

Ⅳ. DISCUSSION

Current proteomic techniques involve labeling and detecting proteins in all cells in the sample, because they use amino acids as probes which can be used by all cells for protein synthesis [9]. One such example is BONCAT (bio-orthogonal noncanonical amino acid tagging), which utilizes methionine analog such as L-azidohomoalanine (AHA). When cells grown in a methionine-free medium are supplied with AHA, it is site-specifically incorporated into the AUG codon during mRNA translation in all cells in the sample (Fig. 21). Another example is SILAC (stable isotope labeling by amino acids in cell culture) [10], which utilizes non-radioactive and stable isotope containing amino acids, such as 13C Arg or 13C Lys. Stable isotope-containing amino acids are then incorporated into the *de novo* proteome during mRNA translation. These two methods efficiently label newly the *de novo* proteome of all cells in the samples but cannot label the proteome of specific cells of interest because the amino acid-based probes can be used by all cells [1,11]. A breakthrough in the effort of cell type-specific proteome labeling has come from an unusual genetic code the methane-producing archaea *Methanosarcina*, which processes the PylRS/$tRNA^{Pyl}$ system that is orthogonal to eukaryotic tRNAs and aminoacyl-tRNA synthetases. After the initial discovery, the main trend in this field of research was to improve this system to incorporate UAAs into a specific site of a protein which is often not possible by chemical synthesis. In this thesis, I applied this system for cell type-specific proteome labeling [1,12]. Up to now, the highest organism used to label proteome by the genetic code expansion is *D. melanogaster* [6]. PylRS and $tRNA^{Pyl}$ suppressed the stop and sense codon in a transgenic *D. melanogaster*. Chemically modifiable UAAs were incorporated into the *D. melanogaster* which was induced to express PylRS only in germ line cells. It resulted in tissue and stage specific proteome labeling in flies.

In this study, I applied genetic code expansion into the mammalian cells to label cell type-specific proteome. The PylRS/tRNA^{Pyl} pair of *Methanosarcina* expanded the genetic code of mammalian cells and enabled site-specific incorporation of UAAs into the *de novo* proteome only in genetic code-expanded cells. It has been previously shown that the PylRS/tRNA^{Pyl} system can be used to incorporate the UAAs into the specific sites of a protein by introducing the UAG stop codon to the specific site of an mRNA. In this thesis, I improved this method in several ways to label the endogenous *de novo* proteome of a specific cell population. First, I engineered the anticodon of tRNA^{Pyl} so that it can simultaneously incorporate UAAs into different codons. I improved this system to incorporate UAAs into all three stop and several sense codons, overcoming the limitation of the current the PylRS/tRNA^{Pyl} system in labeling the proteins encoded by mRNAs without the amber stop codon. The codons that whose site-specific UAA incorporation was made possible in this thesis UAG, UGA, UAA, AUG, UUA, UUC and UUU. Interestingly, quantitative analysis showed that incorporation of UAA into the UUA sense codon (Leu) was relatively inefficient, suggesting that this codon is not widely used in the proteome of HEK293T cells. Studying codon usage of the cells of interest would be the first step in optimizing the proteome labeling strategy. The dose response analysis showed that UAA of 0.75 mM or higher concentration is needed label proteome. Developing efficient methods to deliver high concentrations of UAA into a living animal would be prerequisite to extend this study into an organismal level. The time dependency analysis showed 8 to 16 hr treatment of UAA is required for efficient proteome labeling. It seemed that proteome labeling by genetic code expansion is less efficient than proteome labeling by AHA treatment. AHA labeling in HEK293T cells was known to be efficient with 25-50 μM of AHA for 1-4 hr (Fig. 21). Developing a PylRS that acts faster and more effectively in eukaryotic cells is required to improve the efficacy of proteome labeling using this method.

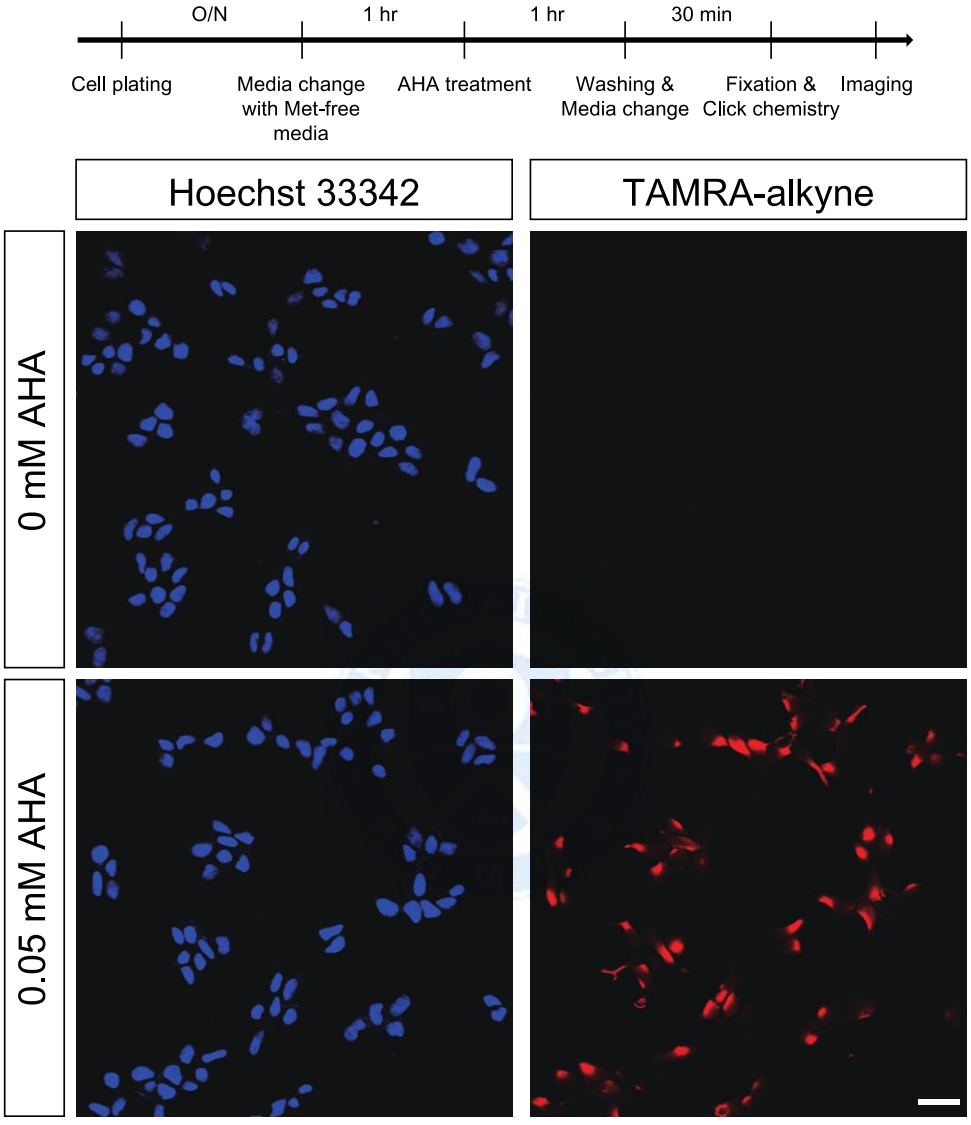
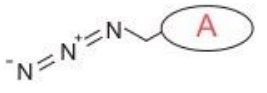


Figure 21. *De novo* proteome labeling by L-azidohomoalanine (AHA) in HEK293T cells. HEK293T cells were plated on poly-L-lysine coated coverslip in 12 well cell culture plate. After overnight incubation, culture media was changed to methionine-free culture media to deplete methionine in the cells. 0.05 mM of AHAs was treated into experimental group for 1 hr. Remained AHAs in the media were washed for 30 min and cells were fixed. Click chemistry reaction was done as in

experiment of proteome labeling by UAA. In click reaction, tetramethylrhodamine-alkyne (TAMRA-alkyne) was counterpart of AHA which results in cycloaddition with Copper catalysis.

UAA used in this study had a terminal alkyne which reacts with an azide and makes a triazole. A form of the copper-catalyzed alkyne-azide cycloaddition (CuAAC) is one of the reactions referred to as click chemistry (Fig. 22) [13,14]. Click chemistry is bio-orthogonal in the sense that it does not react to biological materials. The first bio-orthogonal click reaction was based on the Staudinger ligation. In Staudinger ligation, an azide reacts with a triarylphosphine equipped with an ester. In 2004, Bertozzi group applied this reaction into a living animal [15]. They attached the azide to the cell-surface glycans and monitored the changes in glycosylation by adding phosphine probe comprising flag peptide. However, this reaction showed poor kinetics that high concentration of phosphine was required which leads to high background signal in imaging analysis. With a modification of [3+2] cycloaddition between an acyclic alkyne and an azide developed by Huisgen, Sharpless and co-workers and Meldal and co-workers made CuAAC. Although click chemistry reaction by CuAAC allowed bio-orthogonality and high sensitivity in labeling proteome in living cells, it had toxicity because of Cu catalysis. This drawback restricted analysis of labeled proteins only by lysing or fixing the cell. So the researchers developed new way to settle this challenge. In 2007, Bertozzi and co-workers developed copper-free click chemistry by using an azide and a cyclooctyne [13]. Cyclooctyne is a small cyclo-alkyne which induces cycloaddition with an azide. Cyclooctyne instead of terminal alkyne didn't need Cu catalysis when it reacted with azide. To elevate the rate of cycloaddition, they slightly modified cyclooctyne by adding two fluorine atom to the ring which is called as difluorinated cyclooctyne (DIFO). They added an analog of sugar that contains azide into the cell and fluorescent DIFO probe. Cells used the analog when they made cell surface glycans. Metabolic labeling and detection of glycan by the copper-free click reaction enabled the imaging of glycan internalization and trafficking in live cells. In 2008, Boons and co-workers also developed another copper-free click chemistry [16]. They made a reaction between a dibenzocyclooctyne (DIBO) and an azide. Like Bertozzi group who used azide/DIFO click chemistry, they also applied the

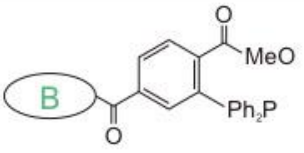
azide/DIBO click reaction into the study of the cell surface glycoconjugates. Both of two copper-free click reactions showed high sensitivity as CuAAC with high biocompatibility in a living organism. Some researchers exploited the reaction between a tetrazine and a trans-cyclooctene (TCO). Click reaction between these two molecules showed faster reactivity than any other copper-free click chemistry [14,17]. Recent study resulted in the site-specific proteome labeling by the tetrazine/TCO click reaction. Fox and chin used PylRS and $tRNA^{Pyl}$ to expand the genetic code and tetrazine/TCO click reaction to detect site-specifically labeled proteins [18]. They demonstrated rapid proteome labeling by incorporating TCO-containing amino acids which acted as Pyl analogs and reacted with tetrazine fluorophores into *E.coli* and live mammalian cells.



Azide-functionalized molecule A

1

3

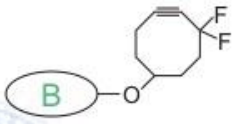


Phoshpine-functionalized molecule B

4

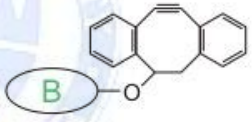
Alkyne-functionalized molecule B

5

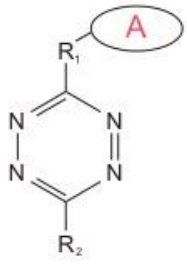


DIFO-functionalized molecule B

6



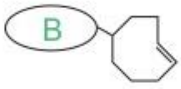
DIBO-functionalized molecule B



Tetrazine-functionalized molecule A

2

7



TCO-functionalized molecule B

Figure 22. Click reactions used in biochemical studies. Various click reactions used in biochemistry. Azide (1) is one of the most popularly used bio-orthogonal chemical group. Azide has small size and shows inert function in a living organism. Staudinger ligation is interaction between an azide (1) and phosphines (3). Phosphine is absent from living system which is suitable material of click reaction. However, this reaction suffers from slow kinetics. It cannot be used in studying fast biological procedure or detecting low abundant molecules. Click reaction between an azide and a terminal alkyne (4) is known as copper-catalyzed azide-alkyne cycloaddition (CuAAC). It is commercially known and sold widely as a kit. Although CuAAC is faster than Staudinger ligation, it needs catalysis of copper which is toxic to the organism. So the CuAAC cannot be applied to study living organism. Difluorinated cyclooctyne (DIFO, 5) and dibenzocyclooctyne (DIBO, 6) are strains of cyclooctyne. Click reaction between an azide and a cyclooctyne doesn't need catalysis. Both of reactions with azide not only have biocompatibility shown in the Staudinger ligation but also have rapid process shown in the CuAAC. Click reaction between a tetrazine (2) and a trans-cyclooctene (TCO, 7) shows faster kinetics than any other click reactions. However, photoisomerization is concerned because of TCO (alkene).

These results in this study showed genetic system of *Methanosarcina* could be used to expand the genetic system of mammalian cells. By inserting PylRS and $tRNA^{Pyl}$ *M. mazei* and incorporating chemically modified UAAs into the HEK293T cells, *de novo* proteome labeling can be done with cell type-specifically. The mutation of anticodon of $tRNA^{Pyl}$s to recognize other stop or even sense codons enhanced the efficiency of incorporation of UAAs. Co-expression of different $tRNA^{Pyl}$s also extended the efficiency of proteome labeling. These results provided a promising direction for a new technique to isolate the proteome of a specific cell type in a living animal (Fig. 23). Also, the application of copper-free click chemistry will broaden the cell type-specific *de novo* proteome labeling by genetic code expansion in real time.

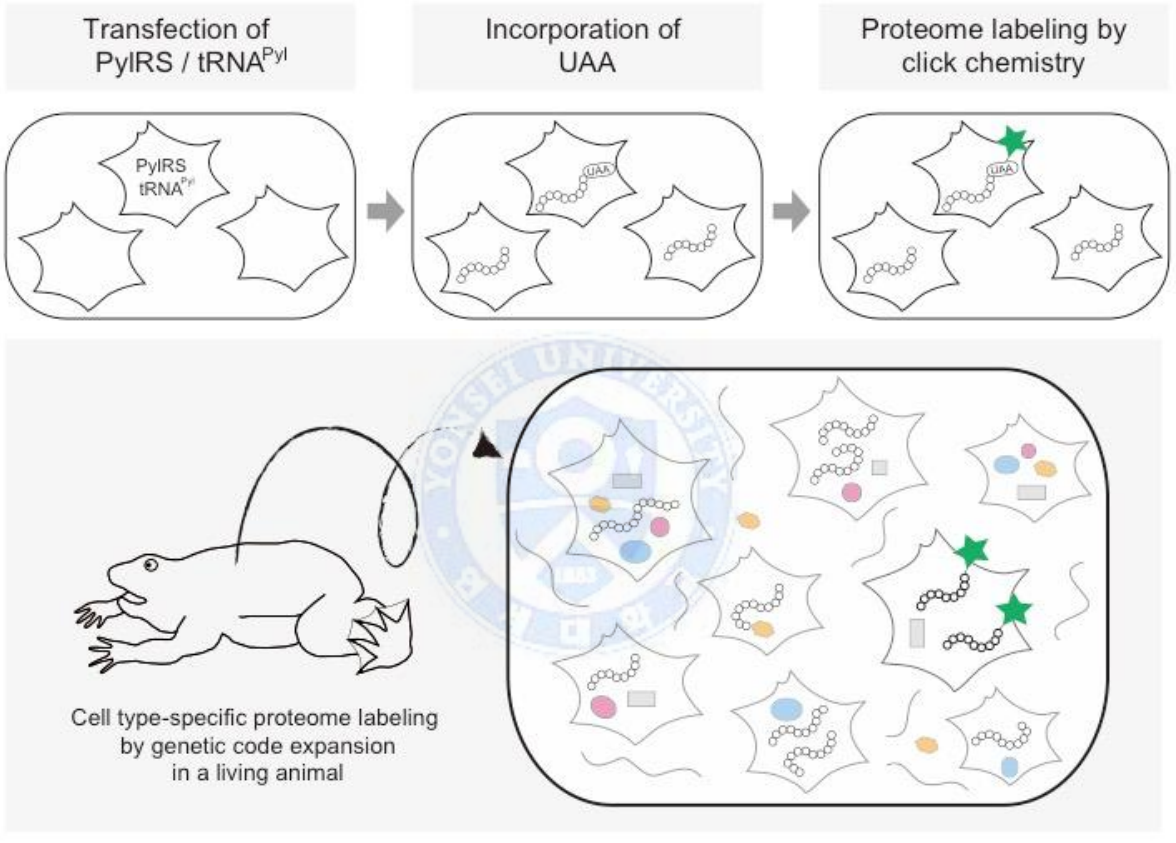

Transfection of
PylRS / tRNA^Pyl
Incorporation of
UAA
Proteome labeling by
click chemistry
PylRS
tRNA^Pyl
UAA
UAA
Cell type-specific proteome labeling
by genetic code expansion
in a living animal

Figure 23. Future direction: cell type-specific *de novo* proteome by genetic code expansion in a living organism. *De novo* proteome labeling by genetic code expansion in the cells proposed the possibility of proteome labeling in a living organism, especially in vertebrates. By inducing PylRS/$tRNA^{Pyl}$ and incorporating UAA into the organism, proteome of a specific cell in specific location can be labeled and identified. Furthermore, click chemistry reactions enabled the visualization and purification of specifically labeled proteome.

Ⅴ. CONCLUSION

This present study shows that genetic code expansion by pyrrolysine tRNA and tRNA synthetase of *Methanosarcina* in mammalian cells. And using genetic code expansion, proteome in the mammalian cells can be labeled by incorporating unnatural amino acids. The major findings of this study are summarized as below:

1. An orthogonal pyrrolysine tRNA and tRNA synthetase pair of *Methanosarcina* expands the genetic code of mammalian cells.
2. Incorporation of unnatural amino acids labels the proteins in genetic code expanded mammalian cells.
3. Genetic code expansion of mammalian cells can happen at all stop and some sense codons.
4. Click chemistry allows specific visualization of the *de novo* proteome of genetic code-expanded mammalian cells.

In conclusion, the *de novo* proteome of a specific cell population in the sample can be specifically labeled by genetic code expansion and incorporation of unnatural amino acids.

REFERENCES

1. Greiss S, Chin JW. Expanding the Genetic Code of an Animal. J Am Chem Soc 2011: 14196-14199.
2. Wang L, Schultz PG. Expanding the genetic code. Angew Chem Int Ed Engl 2004; 44: 34-66.
3. Kim E, Jung H. Local protein synthesis in neuronal axons: why and how we study. BMB Rep 2015; 48 (3): 139-146.
4. Normanly J, Kleina LG, Masson J, Abelson J, Miller JH. Construction of *Escherichia coli* amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency. J Mol Biol 1990; 213(4): 705-717.
5. Liu W, Brock A, Chen S, Chen S, Schultz PG. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat Meth 2007; 4 (3): 239-244.
6. Elliott TS, Townsley FM, Bianco A, Ernst RJ, Sachdeva A, Elsässer SJ, et al. Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. Nat Biotech 2014; 32 (5): 465-472.
7. Ullrich M, Liang V, Chew YL, Banister S, Song X, Zaw T, et al. Bio-orthogonal labeling as a tool to visualize and identify newly synthesized proteins in *Caenorhabditis elegans*. Nat Protoc 2014; 9 (9): 2237-2255.
8. Nikić I, Kang J, Girona GE, Aramburu IV, Lemke EA, et al. Labeling proteins on live mammalian cells using click chemistry. Nat Protoc 2015; 10 (5): 780-791.
9. Borrmann A, Van Hest JMC. Bioorthogonal chemistry in living organisms. Chemical Science 2014; 5 (6): 2123-2134.
10. Ong S, Mann M. A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC). Nat Protoc 2006; 1 (6): 2650-2660.
11. Dieterich DC, Lee JJ, Link AJ, Graumann J, Tirrell DA, Schuman EM. Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. Nat Protoc 2007; 2 (3): 532-540.

12. Wang L. Amersham Prize winner. Expanding the genetic code. Science 2003; 302 (5645): 584-585.
13. Baskin JM, Prescher JA, Laughlin ST, Agard NJ, Chang PV, Miller IA, et al. Copper-free click chemistry for dynamic in vivo imaging. Proc Natl Acad Sci USA 2007; 104 (43): 16793-16797.
14. Jewett JC, Bertozzi CR. Cu-free click cycloaddition reactions in chemical biology. Chem Soc Rev 2010; 39 (4): 1272-1279.
15. Prescher JA, Dube DH, Bertozzi CR. Chemical remodelling of cell surfaces in living animals. Nature 2004; 430 (7002): 873-877.
16. Ning X, Guo J, Wolfert MA, Boons GJ. Visualizing metabolically labeled glycoconjugates of living cells by copper-free and fast huisgen cycloadditions. Angew Chem Int Ed Engl 2008; 47 (12): 2253-2255.
17. Blackman ML, Royzen M, Fox JM. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. J Am Chem Soc 2008; 130 (41): 13518-13519.
18. Lang K, Davis L, Wallace S, Mahesh M, Cox DJ, Blackman ML, et al. Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels–Alder Reactions. J Am Chem Soc 2012; 134 (25): 10317-10320.

ABSTRACT (IN KOREAN)

유전 암호 확장을 이용한 세포 특이적 단백질체 표지법 개발

< 지도교수 정호성 >

연세대학교 대학원 의과학과

김 은 진

단백질의 합성은 매우 정교하게 조절되고 있으며 주위 환경에 미세하게 반응하여 세포의 성장, 분화, 신진대사뿐만 아니라 거의 모든 과정들을 제어한다. 이 때문에 단백질체 연구의 필요성이 커졌고, 여러 조직 내 다양한 세포의 단백질 발현을 연구하는 것이 큰 과제로 떠올랐다. 단백질체를 분석하기 위해서는 조직 혹은 배양된 세포의 단백질 전체를 분리하여 질량분석기 등으로 동정하거나, 대사 표지법을 이용하여 신생 단백질체만을 표지한 후 동정하기도 한다. 그러나 이러한 실험 기법은 시료에 있는 모든 세포의 단백질체를 분석하게 되므로, 배양된 세포와 같이 균일한 세포로 구성되어 있는 시료를 분석하는 데에는 적합하지만 다양한 세포들 간의 상호작용으로 일어나는 생명 현상을 연구하는 데에는 적합하지 않다. 우리 몸은 여러 가지 다른 세포로 구성되어 있고 각 세포의 단백질체는 모두 다르기 때문이다. 따라서 우리가 원하는 세포의 단백질체만을 생물체에서 표지하고 분리하는 방법이 개발되어야 한다.

이에 본 연구는 유전 암호 확장법을 적용하여 세포 특이적 단백질체 표지법을 개발하고자 한다. 보편적으로 생물체가 사용하는 유전 암호는 20개의 아미노산과 3개의 종결 신호를 지정한다. 유전 암호 확장은 생물체가 유전 암호를 이용하여 단백질을 합성할 때, 외부의 유전 암호 체계를 유입함으로써 일어난다. 특정 유전 암호에 대해 다른 아미노산을 지정하고 있는 외부 유전 암호 체계를 이용한다면 생물체는 자신이 암호화하고 있는 아미노산이 아닌 다른 아미노산을 단백질 합성에 참여하도록 할 수 있다. 이처럼 유전 암호 확장법을 이용하면 원하는 세포의 단백질을 표지 하는 것이 가능하다.

구체적으로 본 연구에서는 세포 혹은 생물체에 선별적으로 고세균의 Pyrrolysine 유전 암호 체계를 유입시킨 후 이 유전 암호에 적용 가능하고 표지 가능한 표지용 아미노산을 이용하여 세포 특이적 단백질체 표지법을 개발하고자 한다.

핵심 되는 말 : 유전 암호 확장, 단백질체, 표지용 아미노산