Title
Recoding Rare Codons to Unnatural Amino Acids in Bacteria

Author
Wang, Yiyan

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, MERCED

Recoding Rare Codons to Unnatural Amino Acids in Bacteria

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Quantitative and Systems Biology

by

Yiyan Wang

Committee in charge:

Professor Andy LiWang, Chair
Professor Jinah Choi
Professor Henry J. Forman
Professor Meng-Lin Tsao, Supervisor

2013

The dissertation of Yiyan Wang, titled Recoding Rare Codons to Unnatural Amino Acids in Bacteria, is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____ Date _____

Jinah Choi

_____ Date _____

Henry J. Forman

Supervisor _____ Date _____

Meng-Lin Tsao

Chair _____ Date _____

Andy LiWang

University of California, Merced

2013

To My Family

For the mind set on the spirit is life and peace.

Contents

Abstract **vii**

List of Figures **ix**

List of Tables **xiii**

List of Abbreviations **xiv**

Vita **xvii**

1 Current Opinions on Genetically Incorporating Unnatural Amino Acids into Proteins in *E. coli* **1**

1.1 Advantages to genetically incorporate unnatural amino acids into proteins . . . 1
1.2 Strategies to encode unnatural amino acids in *Escherichia coli* 3
1.3 Two commonly used orthogonal aminoacyl-tRNA synthetase /tRNA pairs . . . 5
1.4 Unique 'blank' codons to be recoded as ***UAA***s in *E. coli* 9
1.5 Recent progresses on reassigning multiple codons in *E coli.* 9
1.6 Is it possible to expand the genetic code further? 10

2 Recoding AGG codon to Unnatural Amino Acids **12**

2.1 The two-step selection strategy of ***UAA***-$\mathrm{RS_{AGG}}$/$\mathrm{tRNA_{CCU}}$ 12
2.2 Materials 13
2.3 Methods 14
2.4 Results 29
2.5 Discussion 52
2.6 Conclusion 56

3 Recoding AGA codon to Unnatural Amino Acids **57**

3.1 The two-step selection strategy of ***UAA***-$\mathrm{RS_{AGA}}$/$\mathrm{tRNA_{UCU}}$ 57
3.2 Materials 58

3.3 Methods 59
3.4 Results 67
3.5 Discussion 80
3.6 Conclusion 82

4 Recoding of two codons at once: AGG and UAG **83**
4.1 Two pairs of orthogonal ***UAA***-RS/tRNA pairs were used to reassign AGG and UAG simultaneously 83
4.2 Materials 85
4.3 Methods 85
4.4 Results 91
4.5 Discussion 101
4.6 Conclusion 104

5 Other Endeavors to Expand the Genetic Code **105**
5.1 Recoding AGA and UAG at the same time 105
5.2 Using modified *Mm*PylRS$_{\mathrm{UAG}}$/tRNA to reassign sense codons 116
5.3 Recoding multiple AGA codons in one protein 119

6 Conclusion and Future Directions **122**
6.1 The expansion of the genetic code: sense codons reassigned 122
6.2 Future directions 123

Bibliography **129**

Abstract

Recoding Rare Codons to Unnatural Amino Acids in Bacteria

by

Yiyan Wang

Doctor of Philosophy in Quantitative and Systems Biology

University of California, Merced

Professor Meng-Lin Tsao, Supervisor
Professor Andy LiWang, Chair

The ability to introduce unnatural amino acids (***UAA***s) with novel chemical, physical and biological properties into proteins adds a new dimension to study the structure and function of proteins. In order to site-specifically incorporate ***UAA***s into proteins *in vivo*, one unique 'blank' codon designated for ***UAA***s and an orthogonal aminoacyl-tRNA synthetase (aaRS) /tRNA pair are required. However there are very few 'blank' codons, which dramatically limits the efficiency and diversity of ***UAA*** incorporation. Till now, only three stop codons and few frame-shift quadruplets, such as AGGA, were reported to be used for ***UAA***s. In order to provide more blank codons, we reassigned rare sense codons to unnatural amino acids. We created new ***UAA***-RS/tRNA pairs from *Mj* TyrRS/tRNA to recode two sense codons (AGG or AGG) from arginine to ***UAA***s. Also, multiple ***UAA***s were incorporated into one protein simultaneously in response to a sense codon and the amber codon UAG by using the new ***UAA***-RS/tRNA pair and *Mm* PylRS/$tRNA_{CUA}$. To overcome the difficulties during sense codon reassignment, we adopted the following strategies. First, a sensitive probe was created for the detection of a single amino acid mutation due to the success of codon reassignment. Second, new ***UAA***-RS/tRNA pairs were evolved to effectively and orthogonally recode the sense codon. Third, the ***UAA*** incorporation conditions were optimized to maximize the ***UAA*** incorporation efficiency and to minimize the toxicity caused by global suppressing of the sense codon. Mass spectrometry results showed several ***UAA***s were introduced into proteins in response to AGG or AGA at a single position by using the newly evolved ***UAA***-RS/tRNA pair, and the arginine suppression rates were over 90%. Further, multiple ***UAA***s were simultaneously incorporated into proteins in responding to AGG and UAG, respectively. Double ***UAA*** incorporation rates were about 80% or above. In addition, fluorescent dye tagging, and PEGylation were used to confirm the successful incorporation of target ***UAA***s. In this work, for the first time the resource of 'blank' codons

was successfully expanded to sense codons. The same strategies can be applied to reassign many other sense codons which contain 61 potential candidate codons for 20 canonical amino acids.

List of Figures

1.1 Protein conjugates synthesized with unnatural amino acids and examples of their applications. . . . 2
1.2 The incorporation of unnatural amino acids into proteins. . . . 4
1.3 The comparison of *E. coli* $tRNA^{Tyr}$ and an orthogonal $tRNA^{Tyr}_{CUA}$ in E. coli which was derived from *Mj* $tRNA^{Tyr}$. . . . 6
1.4 The cartoon structure of *Mj* TyrRS/$tRNA^{Tyr}$ dimer. Colored by subunits. PDB 1J1U. (http://www.rcsb.org/) . . . 7
1.5 Left: *Mm* pylRS monomer, pyrrolysine+AMP. Right: Mm $tRNA^{Pyl}_{CUA}$ 8
1.6 The structure of pyrrolysine. . . . 8

2.1 The recoding of AGG codon . . . 12
2.2 The two-step selection strategy of *UAA*-RS_{AGG}/$tRNA_{CCU}$. . . . 13
2.3 Structure of *Mj*$tRNA_{CCU}$ . . . 14
2.4 Mutation sites on *Mj*TyrRS. . . . 15
2.5 pAGG-GFP_{UV} vector. . . . 16
2.6 GFP protein and its flurophore. . . . 17
2.7 pBK2-JYRS library. . . . 18
2.8 pAGG-Zagg3 vector. . . . 22
2.9 The amino acid binding pocket of pCNF-RS. . . . 23
2.10 The structure of four ***UAA***s: AcF, AzF, IF, OMeY. . . . 23
2.11 pAK5 vector. . . . 24
2.12 The mutation of argW gene . . . 25
2.13 DNSH reacts with AcF. . . . 28
2.14 The spectra of dansyl. . . . 29
2.15 Bacterial fluorescence measurements (Top10F'). . . . 31
2.16 Protein green fluorescence measurements without concentration adjustment. . . . 32
2.17 ESI-MS results of Z-7AcF from DH10BW(YW78) and BL21(DE3)(YW95). . . . 34
2.18 The averages of IF incorporation rates vary among different *E. coli* strains in AGG suppression . . . 35

2.19 IF incorporation rates increased with the increasing of IF concentration in AGG suppression. 36
2.20 Temperature effects on ***UAA*** incorporation:37°C is better than 42°C. . . 37
2.21 Different concentrations of arabinose impact the IF incorporation rates. . 38
2.22 ESI-MS results of Z-7IF from three *E. coli* strains. 39
2.23 DH10BW has higher ***UAA*** incorporation rates than DH10B. 41
2.24 pAK3 vector works better than pAK5 vector to incorporate IF into Z domain proteins in response to AGG. 42
2.25 Dansyl hydrazine (DNSH) conjugation of AcF (AGG) 43
2.26 The toxicity caused by AGG global suppression during pAK5+pAGG3-Zagg/GM10 cell growth. 44
2.27 The toxicity caused by AGG global suppression during pAK3+pAGG3-Zagg/DH10BW cell growth 45
2.28 SDS-PAGE gel of GFP 46
2.29 SDS-PAGE electrophoresis results of GFP proteins purified from different strains after all protein concentrations were adjusted to OD_{280}=0.1. . . 46
2.30 Protein green fluorescence measurements after all protein concentrations were adjusted to OD_{280}=0.1. 47
2.31 ESI-MS results of Z-7AcF from DH10B(YW79). 3mM AcF was added during protein induction. 47
2.32 ESI-MS results of Z-7AzF from DH10B(YW73), DH10BW(YW72) and BL21(DE3)(YW96). 3mM AzF was added during protein induction. . . 48
2.33 ESI-MS results of Z-7OmeY from DH10BW (YW76).3mM OmeY was added during protein induction. 48
2.34 ESI-MS results of Z-NpOH from GM10 (YW41). Protein expression condition: pAK7+pAGG3-Zagg/GM10, 1.5mM NpOH 49
2.35 Control: the cell growth curve of pAK5+pAGG3-Zagg/GM10 cell. Only varied concentrations of arabinose were added at the beginning of cell growth. 49

3.1 The recoding of AGA codon. 57
3.2 The orthogonal $tRNA_{UCU}$ from the archaebacteria *M. jannaschii*. 60
3.3 Secondary structure of argU tRNA in DH10B 61
3.4 pAGA3-Zaga vector carries Z-66AGA gene and the optimized $tRNA_{UCU}{}^{opt}$. 62
3.5 pAK3 vector. 64
3.6 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid (NpOH). 64
3.7 ESI results: four ***UAA***s (AzF-YW4, AcF-YW28, OMeY-YW32, IF-YW44) were incorporated into Z domain proteins at position 7. 69
3.8 LC-MS/MS result indicates that IF was incorporated into Z domain protein at the seventh position. 69

3.9 Various ***UAA***s were incorporated into Z domain proteins in respond to AGA. . . . 70
3.10 Optimized $tRNA^{opt}_{AGA}$ improved IF incorporation rates. . . . 71
3.11 Various ***UAA***s were incorporated into Z domain proteins in respond to AGA. . . . 72
3.12 IF were incorporated into Z domain proteins in respond to AGA in different ***E. coli*** strains. . . . 73
3.13 ESI result: NpOH was incorporated into Z domain proteins in respond to AGA (YW46). . . . 74
3.14 Dansyl hydrazine conjugation of AcF (AGA) . . . 76
3.15 The toxicity caused by AGA global suppression . . . 77
3.16 Control: the cell growth curve of pAK3+pAGA3-Zaga/GM10 cell. Only varied concentrations of arabinose were added at the beginning of cell growth, but no ***UAA*** was added. . . . 78

4.1 The recoding of AGG and UAG at once. . . . 83
4.2 The recoding of AGG and UAG simultaneously depends on two sets of orthogonal aaRS/tRNA pairs. . . . 84
4.3 The structure of two ***UAA***s: BocK(left) and AlK(right). . . . 87
4.4 pAK8 vector . . . 88
4.5 pt2Z238GX vector . . . 88
4.6 AlK reacts with methoxypolyethylene glycol azide through opper(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC). . . . 90
4.7 SDS-PAGE gel: the expression of Z-7AGG-38UAG proteins with the presence of different ***UAA***s. . . . 92
4.8 ESI results of Z-7AzF-38AlK (YW92) (DH10BW). . . . 93
4.9 ESI results of Z-7AcF-38AlK from DH10BW (YW93). . . . 94
4.10 ESI results of Z-7AcF-38AlK from BL21(DE3) (YW99). . . . 95
4.11 MALDI-TOF results of Z-7IF-9BocK (YW3M). . . . 96
4.12 MALDI-TOF results of Z-7IF-38BocK (YW4M). . . . 97
4.13 ESI results of Z-7IF-38BocK (YW91). . . . 98
4.14 SDS-PAGE gel: DNSH labeling and PEGylation on Z-7AcF-38AlK proteins. . . . 99
4.15 SDS-PAGE gel: CuAAC reaction of Z-7AzF-38AlK proteins. . . . 100
4.16 The toxicity caused by global AGG+UAG suppression during DBL21(DE3) growth . . . 102

5.1 The recoding of AGA and UAG. ***NAA***, noncanonical amino acids=unnatural amino acids. . . . 105
5.2 pt2Z238 vector . . . 107

5.3 SDS-PAGE gel: Proteins purified from pAK8+pt2Z2/GM10 after protein expressed with IF, or BocK, or both, and different amount of arabinose. 108
5.4 SDS-PAGE results of mutant Z domain proteins with both AGA and UAG recoding. 110
5.5 ESI-MS result of YW90: pAK8+pt2Z238/GM10 with IF+BocK. 110
5.6 ESI-MS result of YW88: pAK8+pt2Z238/GM10 with IF+AlK. 111
5.7 OMeY and BocK were incorporated into Z domain proteins by pAK10 with pt2Z238 in GM10. (YW82, ESI-MS). 112
5.8 Four controls:AlK-only, BocK-only, AzF-only, and AcF-only. For each control, only one ***UAA*** was added into pAK8+pt2Z2/GM10 for protein expressed. SDS-PAGE gel. 113
5.9 The MALDI-TOF result of AcF-only sample (YW2M). 114
5.10 pAGA9-Zaga vector. 117
5.11 ESI-MS result of Z-9BocK from pAK9+pAGA9-Zaga/GM10 (YW86). . 118
5.12 SDS-PAGE gel:trying to recode continuous multiple AGA codons in Z domain protein. 120

List of Tables

2.1 The AGG ambiguity among 7 screen-out clones. 19
2.2 Clones built to clarify AGG ambiguity. 20
2.3 The amino acid sequences of pBK2-10R,pBK2-12R and pBK2-15RR. . . 30
2.4 Molecular weights of different Z domain proteins. 33
2.5 Incorporation of ***UAA***s by pAK3+pAGG3-Zagg in DH10B, DH10BW and BL21(DE3). 40
2.6 Varied ***UAA***s were incorporated into Z domain proteins in response to AGG codon by using pAK5 vector. 50
2.7 Varied ***UAA***s were incorporated into Z domain proteins in response to AGG codon by using pAK3 vector. 51

3.1 Sequencing results of mutation sites of 7 positive strains from pAGA-GFP and pBK4-JYRS library screening. 68
3.2 short 70
3.3 Varied ***UAA***s were incorporated into Z domain proteins in response to AGA codon. 79

4.1 The molecular weights of Z domain proteins when its two amino acids were replaced by ***UAA***s. 92

5.1 Mass Spectrometry of Z domain proteins purified from pAK8+ pt2Z238/GM10. 115

6.1 Rare codon frequencies in *Escherichia coli*. 126
6.2 Rare codon frequencies in *Homo sapiens*. 127

List of Abbreviations

$3NO_2Y$ 3-nitrotyrosine

NAA noncanonical amino acid

UAA unnatural amino acid

Mb *Methanosarcina barkeri*

Mj *Methanococcus Jannaschii*

Mm *Methanosarcina mazei*

aaRS aminoacyl-tRNA synthetase

AcF *p*-acetylphenylalanine

AcK N^{ϵ}-acetyl-L-lysine

ADCs antibodyrug conjugates

Af-tRNA$^{\mathrm{Pro}}$ *Archaeoglobus fulgidus* prolyl-tRNAs

AlK N^{ϵ}-progargyloxycarbonyl-L-lysine

ArgRS arginyl-tRNA synthetase

AzF *p*-azidophenylalanine

BocK N^{ϵ}-Boc-L-lysine

CmR chloramphenicol acetyl transferase

CNF *p*-cyanophenylalnine

CuAAC Copper(I)-catalyzed Azide-Alkyne Cycloaddition

cyc N^{ϵ}-cyclopentyloxycarbonyl-L-lysine

DNSH 5- dimethylaminonaphthalene -1-sulfonyl hydrazine (Dansyl hydrazine)

ESI-MS Electrospray ionization mass spectrometry

IF *p*-iodophenylalanine

IPTG isopropyl-β-D-1-thiogalactopyranoside

LC-MS/MS Liquid chromatography-tandem mass spectrometry

MALDI-TOF MS Matrix-assisted laser desorption-ionization time-of-flight mass spectrometry

NO_2F *p*-nitrophenylalanine

NpOH 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid

OMeY *O*-Methyl-L-tyrosine

PEG polyethylene glycol

PhProRS *Pyrococcus horikoshii* ProRS

PylRS Pyrrolysyl-tRNA synthetase

SO_3Y sulfotyrosine

TyrRS tyrosyl-tRNA synthetase

Acknowledgments

I would like to thank my PhD supervisor and members of my dissertation committee for supervising my researches, for supporting me, for helpful advises and suggestions, especially for their extreme patience in the face of numerous obstacles. I would like to thank my fellow doctoral students and labmates for their support, feedback, and friendship. I would like to thank financial support and scholarships from University of California, Merced. This work was sponsored by UC Merced Start-up fund. My thank also goes to Dr Zhongjie Jiang for assisting fluorescence measurements, Shuo Chen for synthesizing chemicals, and Dr Jian Cai for facilitating data analysis. I am also grateful for materials from Professor Peter Schultz's research group, for DNA sequencing service and mass spectrometry service from UC Berkeley, for LC-MS/MS service from Stanford University, for bacteria strains from The Coli Genetic Stock Center at Yale, and for pSIM5 plasmid from Dr Don L. Court's lab.

Yiyan Wang

ywang34@ucmerced.edu Tel: 209-386-3062

Education

Ph.D. Expected (12/2013) 09/2008 –12/2013
Quantitative & Systems Biology, University of California, Merced, USA
2013 Spring QSB Dissertation Fellowship; QSB summer fellowships

Master of Science, Biology 09/2006- 08/2008
Texas Tech University, USA
The AT&T Chancellor's Fellowship

Postgraduate Certificate 08/2005- 12/2005
Teaching Science & Engineering in English, The Univ. of Manchester, UK;
Sponsored by China Scholarship Council

Master of Science, Biology of Pathogen 09/2001- 07/2004
Huazhong **U**niv of **S**ci & **T**ech (HUST), China
College Fellowship for M.S.

Bachelor of Medicine (M.D. equivalent) 09/1995- 07/2000
Tongji Medical College, HUST, China;
Distinguished Student Scholarship; Excellent Student Scholarships

Research Experiences

University of California, Merced, USA 09/2008 – 12/2013
Ph.D. Research, School of Natural Sciences, Quantitative and Systems Biology
Projects: Recoding Rare Codons to Unnatural Amino Acids in Bacteria

Texas Tech University, USA 08/2006- 07/2008
Master Research, Department of Biological Sciences
Projects: Ectopic expression of PP2A in *Arabidopsis thaliana*

Tongji Medical College, HUST, China 07/2000-08/2006
Master Research, Department of Biology of Pathogens
Research Assistant, Department of Biology of Pathogens, Division of Microbiology
Projects: Morphological study of *P. boydii* & establishment of its systemic infection by intraperitoneal inoculation in mice.

Teaching Experience

University of California, Merced, USA 01/2009-12/20/2013
Teaching Assistant, School of Natural Sciences

Texas Tech University, USA 08/2006-08/2008
Teaching Assistant, Department of Biological Sciences

Tongji Medical College, HUST, China 07/2000-07/2006
Lecturer & Course Coordinator, Department of Biology of Pathogens

Clinical Experiences

Wuhan Union Hospital, Tongji Medical College, HUST, China

Trainee, Department of Dermatology & Sexually Transmitted Diseases	03/2001-04/2001
Internship, Dept of Clinic Laboratories	09/2000-11/2000
Internship as a Medical Doctor	01/1999-01/2000

Publications

- **Yiyan Wang**, Jingui Zhu, Aaron Ortiz, Meng-Lin Tsao, The Recoding of a rare AGG codon in *E. coli*. 2013 (to be submitted)
- **Yiyan Wang**, Meng-Lin Tsao, Expending Genetic Code: Reassign a rare AGA codon to code for unnatural amino acids, 2013 (to be submitted)
- **Yiyan Wang**, Meng-Lin Tsao, Incorporating multiple unnatural amino acids by recoding AGG and UAG codons in *E. coli.* 2013 (to be submitted)
- **Yiyan Wang**, Xukai Zhang, Hanju Huang, Study development of Hantaana Virus Nucleic Acid Vaccine. Herald of Medicine, 2006, 25(7):601 (in Chinese)
- Xukai Zhang, Dongbin Qian, **Yiyan Wang**, Hanju Huang, Prevent and Treat Avian Influenza. Herald of Medicine, 2006,25(7): 607 (in Chinese)
- **Yiyan Wang**, Hanju Huang, Yanqing Wu *et al*, Establishment of *Pseudollescheria boydii* or *Scedosporium apiospermum* Systemic Infection by Intraperitoneal Inoculation in Experimental Murine. *J. Huazhong Univ Sci Tech <Health Sci>*, 2004,33 (2):255 (in Chinese)
- Yuecheng Zheng, Yongbin Zheng, Jiawen Li, Aipin Feng, **Yiyan Wang**, Case report: A child with congenital ichthyosiform erythoderma. Chinese Journal of Dermatology, 2004,37(3):181 (in Chinese)
- **Book chapter,** Medical Microbiology, Jilin Science and Technology Publishing House, China, 2005 February
- Proof reading, Medical Microbiology, Higher Education Press, China, 2004 July (in Chinese)
- **Editorial staff**, A Textbook of Experimental Microbiology, Huazhong University of Science and Technology, 2003 (in Chinese)

Chapter 1

Current Opinions on Genetically Incorporating Unnatural Amino Acids into Proteins in *E. coli*

1.1 Advantages to genetically incorporate unnatural amino acids into proteins

The ability to genetically encode unnatural amino acids(***UAA***s) directly in living cells allows one to manipulate the physiochemical, biological, and pharmacological properties of proteins with exquisite control over structure[1, 2]. To date, over 70 unnatural amino acids[3] have been co-translationally incorporated into proteins with high fidelity and efficiency by new components of the protein biosynthetic machinery[4]. These include metal ion binding, fluorescent, photo cross-linking, and photocaged amino acids, amino acids containing NMR, infrared (IR), and crystallographic probes, and amino acids with orthogonal chemical reactivity[3].

Decorating proteins with some special ***UAA***s may introduce conjugating-potential groups into proteins and then precisely control conjugation sites and stoichiometry[5]. For instance, ketone-containing ***UAA***s and alkoxy-amine derivatives can be linked by forming a stable oxime bond which is stable under physiological conditions[6]. Also, certain relatively mild conjugation condition,such as Copper(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC)[7, 8, 9, 10], can mediate the conjugation of two moieties through the formation of a 1,2,3-triazole linkage between azide and alkyne . Nowadays unnatural amino acids carrying keto group or azide group can be efficiently introduced into proteins[11, 12], and alkoxy-amine or alkyne derivatives are relatively easy to synthesize[10, 5].

Different from traditional chemical modification, site-specific incorporation of chemically orthogonal groups provides a very precise control over the functional and

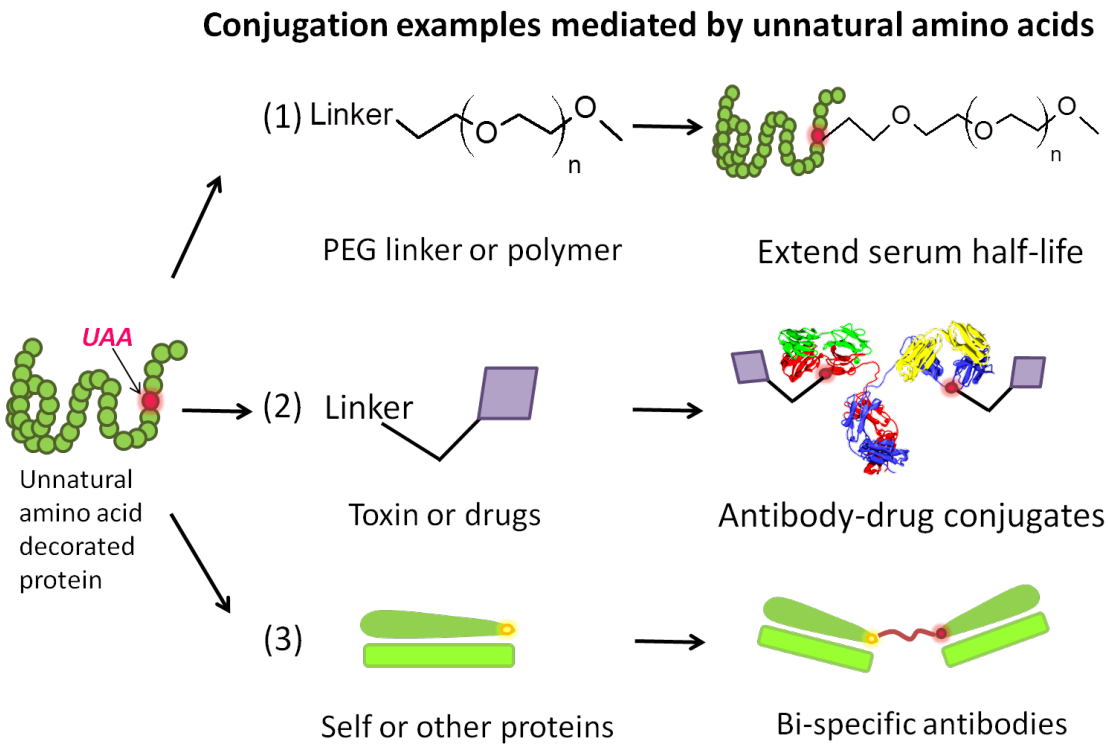


Figure 1.1: Protein conjugates synthesized with unnatural amino acids and examples of their applications.

physical properties of protein therapeutics. Protein conjugates synthesized with unnatural amino acids can be linked with polyethylene glycol (PEG) moiety, or other linkers(e.g. fluorescent tags), or other internal groups within the same protein, or even oligonucleotides[13, 5] (Fig. 1.1). For example, site-specific PEG conjugation is used to extend the serum half-life of proteins, such as human growth hormone[14] and fibroblast growth factor 21[15]. Anther good example of utilizing incorporated ***UAA***s is to synthesize antibody-drug conjugates (ADCs), which preferentially deliver cytotoxic drugs to tumor-associated antigen presented cells to improve drug efficacy and avoid toxicity to healthy tissues[5]. In addition, bi-specific antibodies that can bind two different antigens simultaneously also emerges as a good candidate for next-generation immunotherapeutics for cancer[5]. The bi-specific antibodies can be generated by using ***UAA***s (*e.g.*, AcF) decorated antibodies and flexible bi-functional linkers (*e.g.*, short PEG linkers)[16] .

1.2 Strategies to encode unnatural amino acids in *Escherichia coli*

Requirements to genetically incorporate *UAA*s into proteins

In order to co-translationally incorporate ***UAA***s into proteins at specified sites, there are three necessary components: an unnatural amino acid, a codon that uniquely designates the unnatural amino acid, and an orthogonal aminoacyl-tRNA synthetase (aaRS)/tRNA pair that specifically incorporate the unnatural amino acid into proteins in response to the cognate codon[1, 3] (Fig. 1.2). These components must satisfy a number of criteria as follows[4, 3]. First, there are some requirements for the desired ***UAA***: can be up-taken by cells, and stable and functional inside cells; can be tolerated by EF-Tu and the ribosomes; can not be recognized by any endogenous aaRSs. Second, a unique codon should be available to be recognized by the new tRNA but not by any endogenous tRNAs. Third, the aaRS/tRNA pair must specifically recognize and incorporate the ***UAA***, into proteins in response to the unique codon in the host organism, and must be orthogonal in the context of all endogenous aaRS/tRNA pairs in the organism[17].

Usually the ***UAA***s can either be synthesized in biochemistry labs or commercially available, and are taken up by host cells after added into culture media. Also, natural aaRSs are highly specific for their cognate amino acids after the long time evolution. And the aminoacyl-binding site of elongation factor Tu (EF-Tu) and the ribosomes are reported to be highly compatible to a range of acceptable substrates, including many noncanonical amino acids as well as D-amino acids. So the first criterion can be fulfilled by taken all stated above together[18].

To meet the second criterion, one may choose a 'blank' codon from stop codons, frame-shift codons, or any unused codons, as the unique codon. For example, the amber stop codon UAG[1, 19, 12] was most often used to encode unnatural amino acids without perturbing the growth of a host because UAG is the least used(0.3 per thousand) among the three stop codons in *E. coli*, rarely terminates essential genes, and is efficiently translated by amber suppressor tRNAs *in vivo* and *in vitro*[18]. Nowadays, many efforts were put to expand the selection of the unique 'blank' codons.

It's the third but not the least criterion to find out aaRS/tRNAs with good orthogonality and aminoacylation . Actually real challenges lie on the searching of the suitable aaRS/tRNA pairs and the subsequent modifications on them. The most straightforward way involves importing a heterologous aaRS/tRNA pair from a different domain of life because tRNA recognition by aaRSs is domain or species specific. Further, structure-based mutagenesis and a two-step selection, which will be stated below, can be applied to improve the orthogonality and the specificity[1, 3].

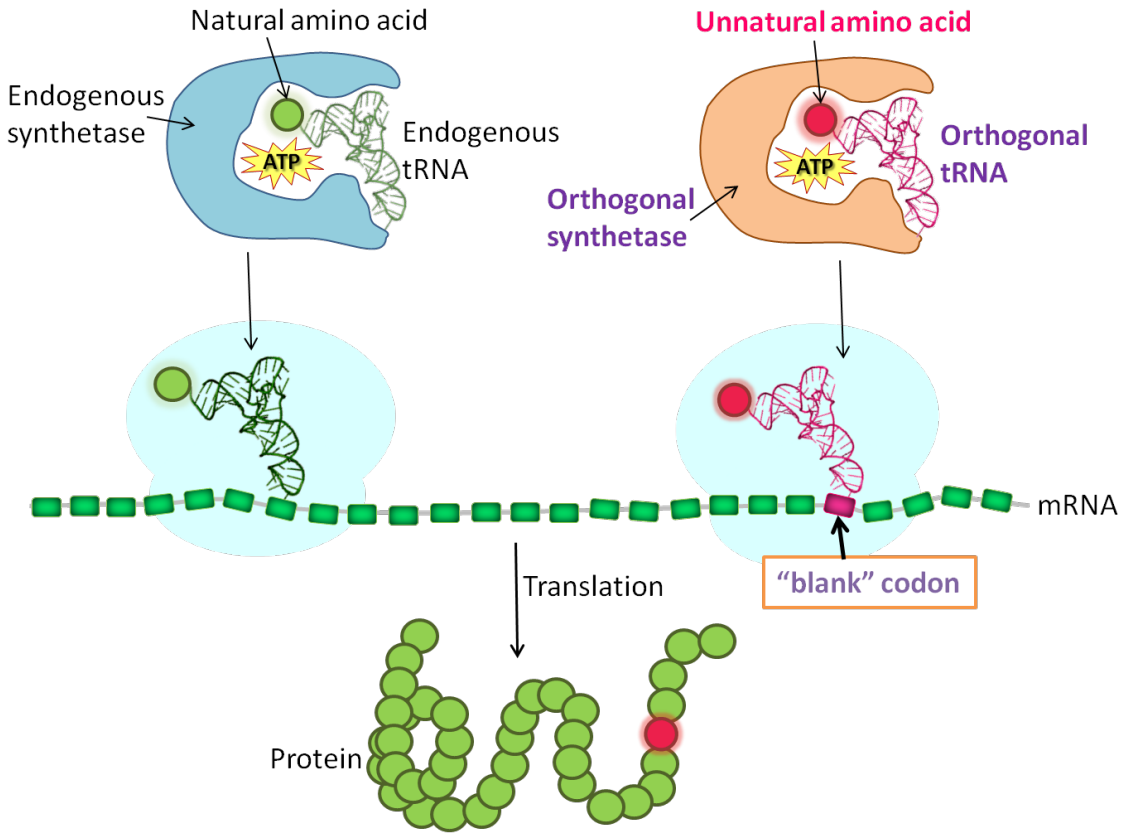


Figure 1.2: The incorporation of unnatural amino acids into proteins.

Screening scheme

One can apply a two-step selection scheme [1, 3] to assist the directed-evolution of the heterologous aaRS which only uniquely recognizes the ***UAA*** of interest [3, 20, 21]. For the tyrosyl system in *E. coli*, the first round is called the positive selection. A library of aaRS mutants with randomized residues in the amino acid-binding pocket is constructed. Then this library is transferred into cells containing chloramphenicol acetyl transferase (CmR) with an amber TAG mutation at a permissive site, and the transformed cells are cultured in media containing chloramphenicol and the unnatural amino acid. The survivals can only be the colonies which contain synthetase variants that incorporate either the unnatural or any of the endogenous amino acids in response to the amber codon TAG. Selected synthetase clones are then transferred into cells containing a toxic barnase gene with amber mutations at permissive sites, and grown in the absence of unnatural amino acid. In this negative selection step, those clones incorporate endogenous amino acids produced full-length barnase protein and died. Repeated rounds of positive and negative selections may result in the

isolation of mutant synthetases that can specifically incorporate the unnatural amino acid in response to the amber codon. In a simplified genetic selection system, an amber-T7/GFP_{UV} was used as an additional reporter in the positive selection[21]. Suppression of amber codons introduced at permissive sites in the T7 RNA polymerase gene produces full-length T7 RNAP, which drives the expression of GFP_{UV}.

1.3 Two commonly used orthogonal aminoacyl-tRNA synthetase /tRNA pairs

Methanococcus Jannaschii(Mj) tyrosyl-tRNA synthetase and $tRNA^{Tyr}$

One chosen orthogonal tRNA-synthetase pair in *E. coli* was derived from *Methanococcus Jannaschii(Mj)* tyrosyl-tRNA synthetase (TyrRS)/$tRNA^{Tyr}$ [4, 22, 21], which does not cross-react with the endogenous synthetases of the *E. coli* [23, 1]. There are following reasons to use this pair.

One reason comes from tRNA itself. The orthogonal $tRNA^{Tyr}_{CUA}$ derived from *Mj* $tRNA^{Tyr}$ is mainly recognized by the synthetase through its discriminator base A73 and the base-pair, C1-G72, in the acceptor stem; while the anticodon triplet is only weakly involved in identity determination, which allows the orthogonal *Mj* TyrRS still to recognize orthogonal *Mj* $tRNA^{Tyr}$ even if the anticodon 5'-GUA-3' is altered to 5'-CUA-3'[1]. On the contrary, in *E. coli* $tRNA^{Tyr}$, the recognition elements are composed by A73, G1-C72, a long variable arm, and the anticodon triplet [1]. Thus *Mj* $tRNA^{Tyr}$ is quite different from endogenous *E. coli* $tRNA^{Tyr}$ and can functional orthogonally in *E. coli*(Fig. 1.3).

Meanwhile, some important characteristics of *Mj* TyrRS make it a very good candidate for unnatural amino acids incorporation. As a class I synthetase, *Mj* TyrRS possesses *an amino acid-binding pocket, which is highly tolerant for substitutions.* As we stated above, *Mj* TyrRS contains a minimalist anticodon loop binding domain, so it still maintain the ability to recognize its *Mj* $tRNA^{Tyr}_{CUA}$ with an alternation of anticodon from 5'-GUA-3' to 5'-CUA-3'[24, 4]. Normally after the two-step screening scheme, the amino acid binding pocket is altered with certain mutations to accommodate the designated ***UAA*** while its ability to charge endogenous amino acid, *i.e.*, tyrosine, is decreased dramatically. Further, *Mj*TyrRS does not possess an editing mechanism that could deacylate and unnatural amino acid [23] (Fig. 1.4). Another reason of *Mj*TyrRS/ $tRNA^{Tyr}_{CUA}$ been widely used is this pair can be expressed at high levels in its functional form within *E. coli*. Schultz *et al.* reported a system with both the encoding *Mj*$tRNA^{Tyr}_{CUA}$ and the gene encoding the suppressor *Mj*TyrRS in a single plasmid compatible with most *E.coli* expression vectors and strains[12]. In

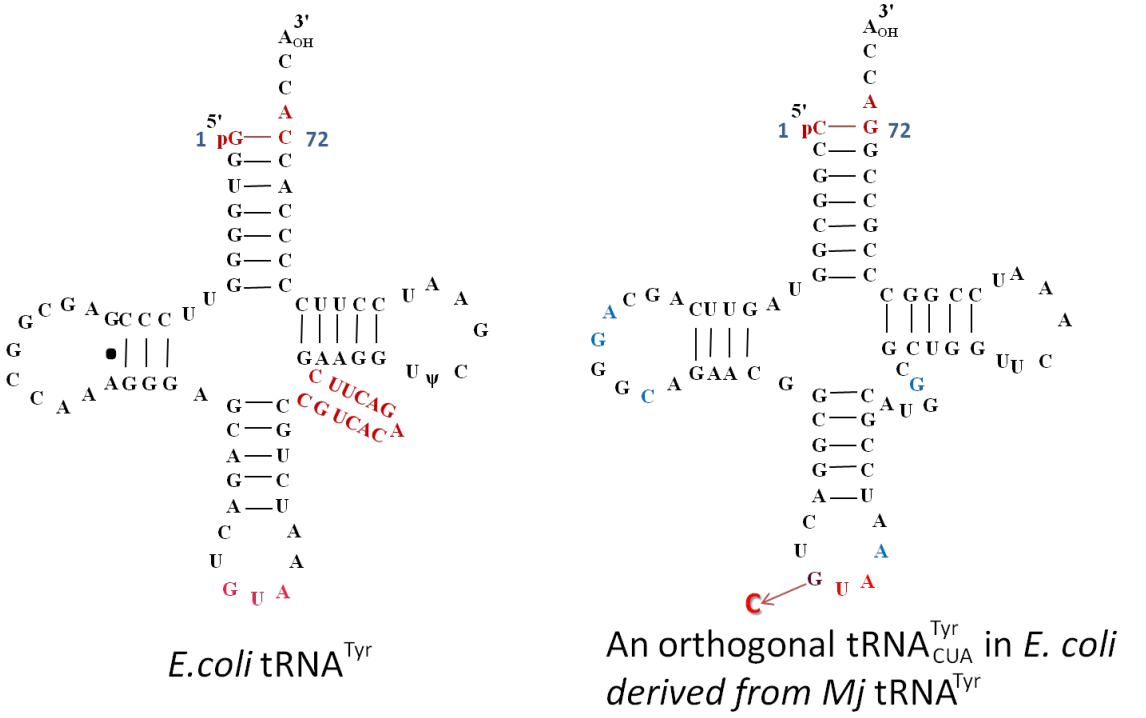


Figure 1.3: The comparison of *E. coli* tRNA^{Tyr} and an orthogonal $\text{tRNA}^{\text{Tyr}}_{\text{CUA}}$ in E. coli which was derived from *Mj* tRNA^{Tyr}.

their system, the suppression efficiencies range from 25% to 50% of wild type protein and translational fidelity ≥99% [3].

Due to the high tolerance to mutations in its amino acid-binding pockets, a large amount of ***UAA***-RSs has been derived from the thermophilic *Mj*TyrRS. For example, there are specific ***UAA***-RSs derived from *Mj*TyrRSfor *p*-benzoylphenylalanine[25] , *p*-bromophenylalanine[26], *p*-acetylphenylalanine(AcF)[27], *p*-azidophenylalanine(AzF) [12, 28], *O*-Methyl-L-tyrosine(OMeY)[29], and so on. Notably, Young *et al* reported an evolved *p*-cyanophenylalnine(CNF) specific aminoacyl-tRNA synthetase (*p*CNF-RS) with atypical polysubstrate specificity[30]. While still being orthogonal to the 20 canonical amino acids, this *Mj*TyrRS originated *p*CNF-RS has high substrate permissivity for 21 types of unnatural amino acids, including *p*CNF, AcF, AzF, OMeY, *p*-iodophenylalanine (IF), *p*-nitrophenylalanine (NO_2F), and etc. The amino acid binding pocket of the pCNF-RS have six mutations (Y32L, L65V, F108W, N109M, D158G, and I159A) compared to the WT *Mj*TyrRS to build a larger cavity for multiple unnatural amino acids[30].

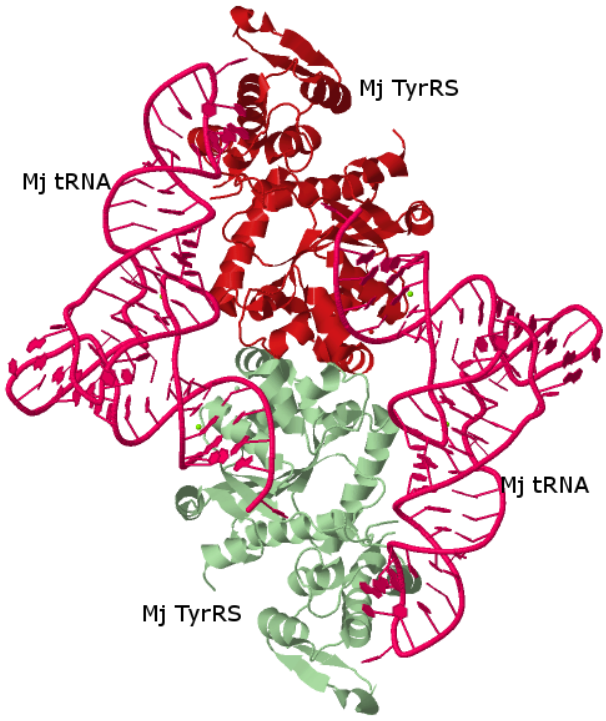


Figure 1.4: The cartoon structure of *Mj* TyrRS/tRNA^{Tyr} dimer. Colored by subunits. PDB 1J1U. (http://www.rcsb.org/)

Methanosarcinaceae pyrrolysyl-tRNA synthetase and tRNA$^{Pyl}_{CUA}$

Pyrrolysyl-tRNA synthetase(PylRS) and tRNA$^{Pyl}_{CUA}$ pair from members of *Methanosarcinaceae* was first found to decode UAG amber stop codon to pyrrolysine, the 22nd genetically encoded amino acid [31, 32, 33]. PylRS, an archaeal class II aminoacyl-tRNA synthetase, directly aminoacylates tRNA$^{Pyl}_{CUA}$ with pyrrolysine(Fig. 1.5), which functions as a free molecule [32, 34]. These PylRS/tRNA$^{Pyl}_{CUA}$ pairs, such as the ones from *Methanosarcina barkeri (Mb)* or *Methanosarcina mazei (Mm)*, can function well within either *E. coli* or mammalian cells[35], and remain their orthogonality without any cross-reactivity with the 20 canonical aaRS/isoacceptor tRNA set pairs[36, 37, 35]. One interesting characteristic of tRNA$^{Pyl}_{CUA}$ is that it has a smaller than normal D-loop, and only one base, instead of the normal two bases, is found between the acceptor and D-stems. Also, tRNA$^{Pyl}_{CUA}$ lacks the highly conserved G-purine sequence in the D loop and TφC sequence in the T loop[33]. Further, the substrate specificity of PylRS is determined by a small number of residues that form the substrate-chain-binding pocket[38].

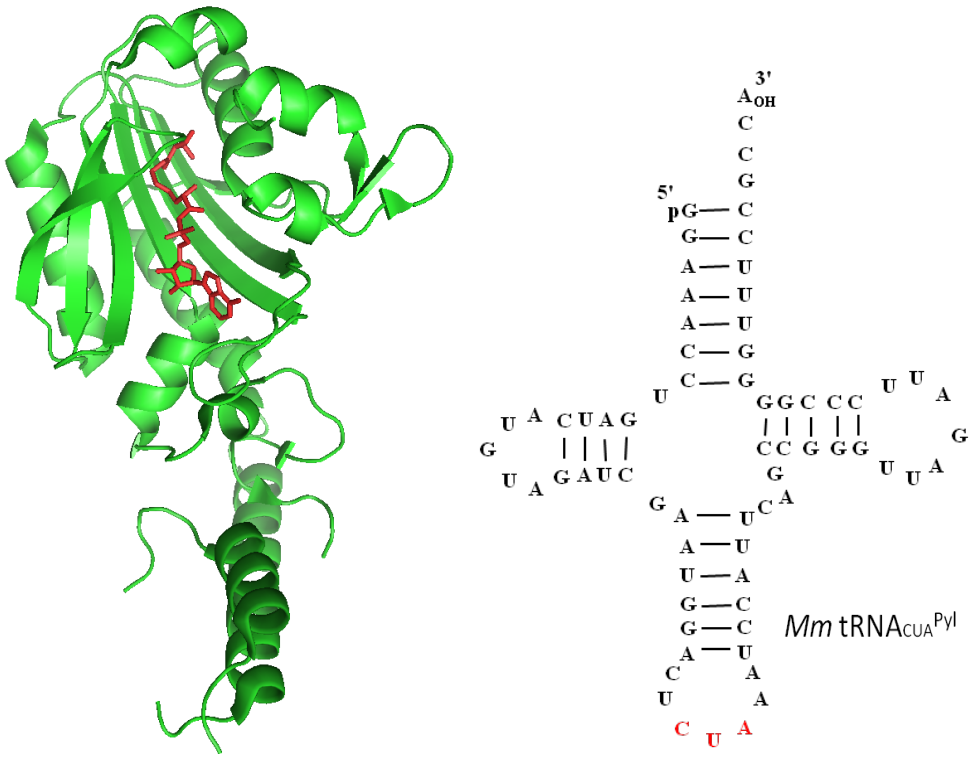


Figure 1.5: Left: *Mm* pylRS monomer, pyrrolysine+AMP. Right: Mm $tRNA^{Pyl}_{CUA}$

In its native contexts, PylRS/tRNA pair is able to incorporate not only pyrrolysine (Fig. 1.6), but also many of its analogs, such as N^{ϵ}-acetyl-L-lysine (AcK), N^{ϵ}-Boc-L-lysine (BocK), N^{ϵ}-progargyloxycarbonyl-L-lysine (AlK), N^{ϵ}-cyclopentyloxycarbonyl-L-lysine(cyc), and etc[39, 40, 10, 35, 41, 42]. Furthermore, with proper modifications and rational screenings, the side chain binding sites of $PylRS_{CUA}$ can be altered to host many other unnatural amino acids, such as OMeY and so on[43].

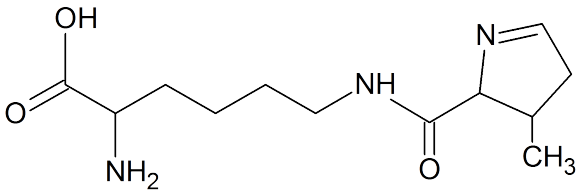


Figure 1.6: The structure of pyrrolysine.

1.4 Unique 'blank' codons to be recoded as *UAA*s in *E. coli*

To encode distinct unnatural amino acids into proteins, the blank codons and mutually orthogonal aminoacyl-tRNA synthetase-tRNA pairs that recognize ***UAA***s and decode the new codons are equally essential. While in nature each of 64 triplet codons is assigned to encode 20 canonical amino acids or protein synthesis termination, there are mainly three types of blank codons which can be borrowed for the incorporation of ***UAA***s genetically. The first group comes from the least used nonsense stop codons, such as amber stop codon TAG, or opal stop codon TGA. As stated above, the amber codon has been proved to specify ***UAA***s very efficiently[3]. Secondly, the frameshift quadruplet codons are another choice for blank codons[3]. For instance, the AGGA quadruplet codons and cognate suppressor tRNAs with expanded anticodon loops were reported to encode additional unnatural amino acids in single polypeptides[44, 45] . In this case, the suppression is in competition with the recognition of frame-shift codons by tRNAs with three-base anticodons, which decreases the co-translational incorporation efficiency. Indeed, the inefficiency with which natural ribosomes decode quadruplet codons is one of the key problems in this approach. To conquer this problem, the evolution of a quadruplet-decoding ribosomes in AGGA frameshift expression system was tested. But it still needs further explore to generalize, simplify and optimize the AGGA frame-shift codon expression system. At last, in theory it is feasible to reassign certain rare codons to ***UAA***s with a decreased degeneracy of the code. However it is not as easy as expected to compete with the natural endogenous system. For example, Liu *et al* pointed out PylRS/$\text{tRNA}^{\text{Pyl}}_{\text{CCU}}$ failed to suppress one rare codon AGG to unnatural amino acids in *E. coli* [46].

1.5 Recent progresses on reassigning multiple codons in *E coli.*

Since unnatural amino acids can bring in different properties to proteins, many efforts nowadays are making to incorporating multiple amino acids into one protein. But some problems are still under investigation.

One major limitation comes from the lack of the resource of 'blank' codons, which is now expanded from UAG amber codon to all three stop codons and few frame-shift quadruplets. In an in *vivo* study, Schultz and co-workers developed an approach to recode a quadruplet codon AGGA to homoglutamine by using a variant *Pyrococcus horikoshii* lysyl-tRNA synthetase/tRNA pair in *E. coli*, and used this variant in combination with a mutant *Mj*TyrRS/ tRNA_{CUA} pair that directs the incorporation of OMeY in response to the amber codon[47, 45]. Later Chatterjee and Schultz *et al* dis-

played that several *Pyrococcus horikoshii* ProRS (PhProRS) variants with mutations within anticodon-binding pocket were able to recognize engineered *Archaeoglobus fulgidus* prolyl-tRNAs (Af-tRNA^{Pro}) with three different anticodons: CUA, AGGG, and CUAG[48]. Also, as mentioned above, Chin and colleagues recoded a frame-shift codon AGGA to unnatural amino acids via evolution of a quadruplet-decoding ribosome, and further encoded two types of ***UAA***s by using both ***UAA***-RS*/tRNA$_{UCCU}$ (a derivative from *Mj*TyrRS/tRNA$_{CUA}$) for AGGA codon and *Mb*PylRS/tRNA$_{CUA}$ for UAG within the special ribosomes modified(ribo-X) *E. coli*[44, 49]. It is a very elegant system to combine a stop codon and a frame-shift quadruplet. Furthermore, a few researches demonstrated that UAG amber codon can be accompanied with other stop codons and serve as 'blank' codons for multiple ***UAA***s incorporation. A mutant *Mm*PylRS/ pylT$_{UUA}$ and *Mj*TyrRS/tRNA$_{CUA}$ were paired to incorporate two different ***UAA***s in one protein by both ochre UAA and amber UAG suppression in *E. coli*[41, 50]. Similarly, the opal stop codon UGA can also be suppressed by the *Mm*mutated PylRS-tRNA$^{Pyl}_{UCA}$ and further serves as another candidate to pair with *Mj*TyrRS/tRNA$_{CUA}$[46].

Also, it used to be difficult to simultaneously reassign multiple UAG codons in one protein because the unnatural amino acids incorporation rates dropped dramatically if more than one UAG codons were introduced. Later several studies demonstrated that it's feasible, but need to be under certain special conditions. Huang *et al* found that if the C-terminal domain of the bacterial large ribosomal subunit protein L11 was over-expressed, then three N-ϵ-acetyl-L-lysine (AcK) would be incorporated into GFP$_{UV}$ by suppressing discontinuous UAG codons[42]. Also it has been shown that a re-engineered EF-Tu, which relaxes its quality-control function and permit RS/tRNA binding, allows the incorporation of two phosphoserine into one protein at two different sites in response to UAGs[51]. Moreover, multiple UAG codons can be suppressed within a special JX33 *E.coli* strain, whose RF1 was deleted of compensating mutations in RF2[52]. Even lack of the ***UAA*** incorporation specificity among these multiple UAG codons, this found is still very exciting because the same mechanism can be applied to reassign other codons, or combinations of different codons.

1.6 Is it possible to expand the genetic code further?

One big limitation of multiple ***UAA***s biological incorporation still comes from the shortage of unique 'blank' codons to specify ***UAA***s. As stated above, so far only three stop codons (majorly UAG amber codon) and few frame-shift quadruplets can be recognized by certain engineered aaRS/tRNA pairs. Meanwhile, the site-specificity of incorporated multiple ***UAA***s is also constrained by the poor orthognalities between

aaRS/tRNA pairs and the designated 'blank' codons. Besides the UAG recoding aaRS/tRNA pairs , there are several aaRS/tRNA pairs to reassign ochre codon UAA or opal codon UGA. But they were usually derived from the pairs used to recode UAG and still possess that ability. These problems dramatically limit the efficiency and diversity of ***UAA***s incorporation.

On the other hand, 61 sense codons have never been successfully reassigned to ***UAA***s till now. This is because certain strict criteria make it even more difficult to expand a sense codon to ***UAA***s than the recoding of stop codons. First, a sensitive probe is needed to detect a single amino acid mutation between the codon's designated amino acid and the target one. Second, an ***UAA***-RS/tRNA pair is required to effectively and orthogonally recode the sense codon. Third, the ***UAA*** incorporation efficiency is expected to be high while the toxicity caused by global suppressing of the sense codon should be reduced to avoid devastating bacterial growth.

Nevertheless, any approach to recode sense codons will add enormous diversities to introduce orthogonally chemical groups to proteins. If succeed, the resource of 'blank' codons would be extended to sense codons, which contain 61 potential candidate codons for 20 canonical amino acids in nature. Further, it would allow the incorporation of multiple ***UAA*** combinations site specifically, such as by suppressing a sense codon and a stop codon together. Moreover, this process of evolving a sense codon from its original meaning to another could provide insights for the evolution process of the universal genetic code.

Among rare sense codons, the AGG codon and AGA codon in *E. coli* are the rarest used which accounts for about 0.21% and 0.24% of the entire bacterium genome and about 4% and 5.1% among all codons for arginine [45, 53]. Due to their low occurrence rates, the consequent probability of toxicity is expected to be low if either AGG or AGA is recoded.

Therefore to provide more codons for ***UAA***s and further expand genetic codons, we **hypothesized** that some sense codons (such as AGG and AGA) could be recoded by engineered *Mj*TyrRS/tRNA pair. Furthermore, by combining one sense codon with UAG amber codon, multiple unnatural amino acids would be incorporated into one protein simultaneously via the newly evolved ***UAA***-RS/tRNA pair and its orthogonal counterpart *Mm*PylRS/tRNA_{CUA}.

Chapter 2

Recoding AGG codon to Unnatural Amino Acids

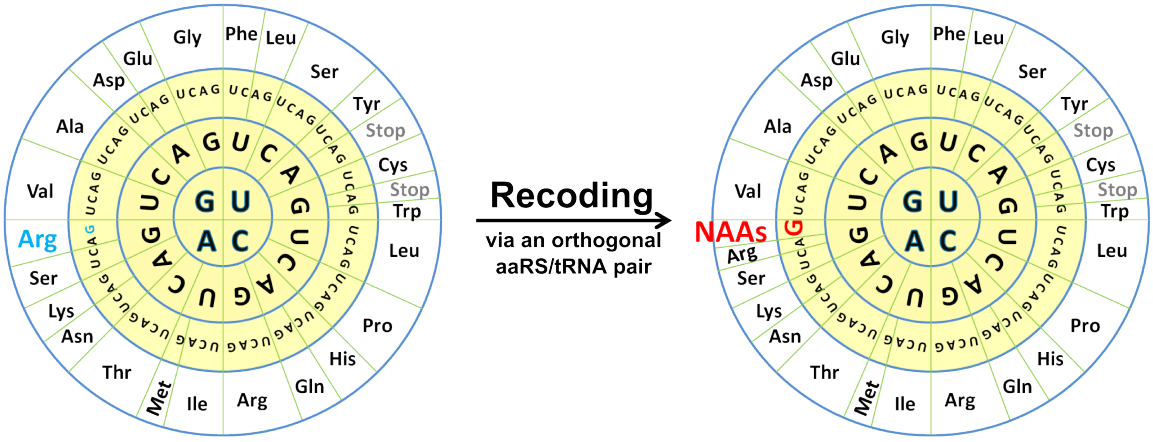


Figure 2.1: The recoding of AGG codon. ***NAA***, the abbreviation of non-canonical (unnatural) amino acid.

2.1 The two-step selection strategy of *UAA*-RS$_{\mathrm{AGG}}$/tRNA$_{\mathrm{CCU}}$

AGG, a rare codon in *E. coli*, can serve as one good candidate blank codon for unnatural amino acids (*UAA*s) because of its low count and low probability of toxicity if recoded. Among 10 rare codons, the AGG codon in *E. coli*(Table 6.1) is the rarest used one which accounts for about 2.1 per thousand of the entire bacterium genome and about 4% among all six codons for arginine[53].

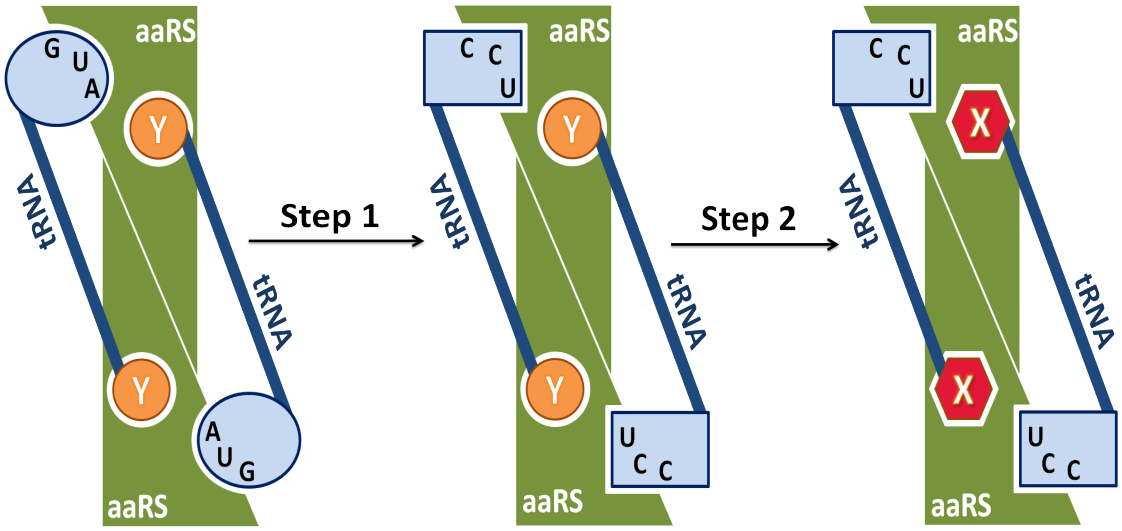


Figure 2.2: The two-step selection strategy of *UAA*-RS$_{AGG}$/tRNA$_{CCU}$.

Here we report the recoding of AGG codon to ***UAA***s in *E. coli* by introducing into bacterial cells a set of evolved orthogonal ***UAA***-RS/tRNA$_{CCU}^{UAA}$ pair that can selectively interact with only the target AGG codon and target amino acid but none of the endogenous components. The discovery of this orthogonal ***UAA***-RS/tRNA pair was achieved by a two-step selection scheme(Fig. 2.2).

The *Methanococcus Jannaschii* tyrosyl-tRNA synthetase (*Mj*TyrRS) / tRNA pair was used as the starting template. We firstly recoded AGG from arginine to tyrosine by evolving the anticodon recognition loop of *Mj*TyrRS to recognize CCU anticodon only. Secondly we reassigned AGG to unnatural amino acids by editing the amino acids binding pocket of the aaRS. To improve ***UAA*** incorporation rates in response to AGG,dual pAK-***UAA***-RS vectors were applied to increase cytoplasmic ***UAA***-RS amounts to obtain more orthogonal ***UAA***-charged tRNA$_{CCU}$s. For the same purpose, optimized *Mj* ***UAA***-tRNA$_{CCU}$s [54] were introduced to bind elongation factor better and the endogenous Arg-tRNA$_{CCU}$s were mutated in certain strains. Also, 3mM ***UAA*** working concentration was proved to be useful to achieve over 90% incorporation rates for certain ***UAA***s in some commonly used E. coli strains, such as BL21(DE3).

2.2 Materials

Top10 *E. coli* (Invitrogen) was used for cloning and amplifying plasmids. DNA was isolated from *E. coli.* using the QIAGEN Plasmid Mini Kit (Qiagen). Expand high fidelity PCR system (Roche) was used for PCR. Restriction enzymes were obtained

from New England Biolabs. BL21(DE3) and DH10B *E. coli* were purchased from Invitrogen. The working antibiotic concentrations were as follows: choramphenicol 50 μg/ml; kanamycin 50 μg/ml and they were purchased from Sigma-Aldrich or Fisher. All cultures were grown in 2YT (Difco) media. All constructed plasmids were confirmed by DNA sequencing at UC Berkeley DNA Sequencing Facility. Primers were synthesized at Integrated DNA Technology (San Diego, CA). Noncannonical amino acids were purchased from Bachem or from Dr Peter G. Schultzs lab. Proteins with his-tag were purified under the native condition with Ni-NTA agarose (Qiagen). Proteins were sent to University of California Berkeley, QB3/College of Chemistry Mass Spectrometry Facility for Electrospray ionization mass spectrometry (ESI-MS).

2.3 Methods

Recoding AGG from arginine to tyrosine

General approach

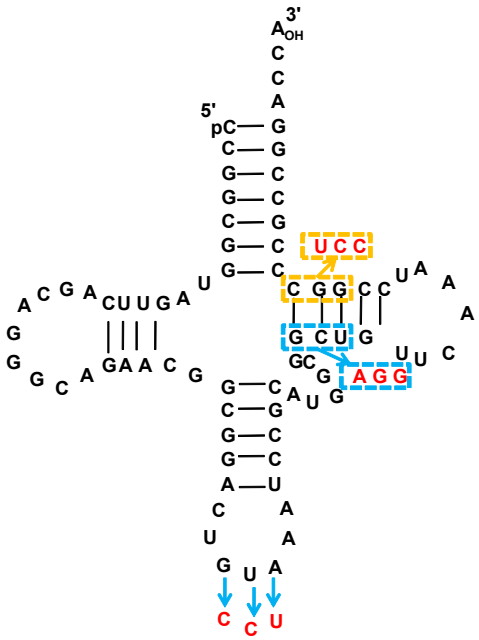


Figure 2.3: Structure of *Mj*tRNA$_{\mathrm{CCU}}$. The anticodon was changed into 5'CCU3' and the optimized mutation sites were shown in rectangular box.

To reassign AGG from arginine to tyrosine, the anti-codon GUA of *Mj*tRNA$^{\mathrm{Tyr}}_{\mathrm{GUA}}$ was altered to CCU (Fig. 2.3). Meanwhile, a *Mj*TyrRS library was established to screen out synthetases to recognize CCU and charge tyrosine into the mutant

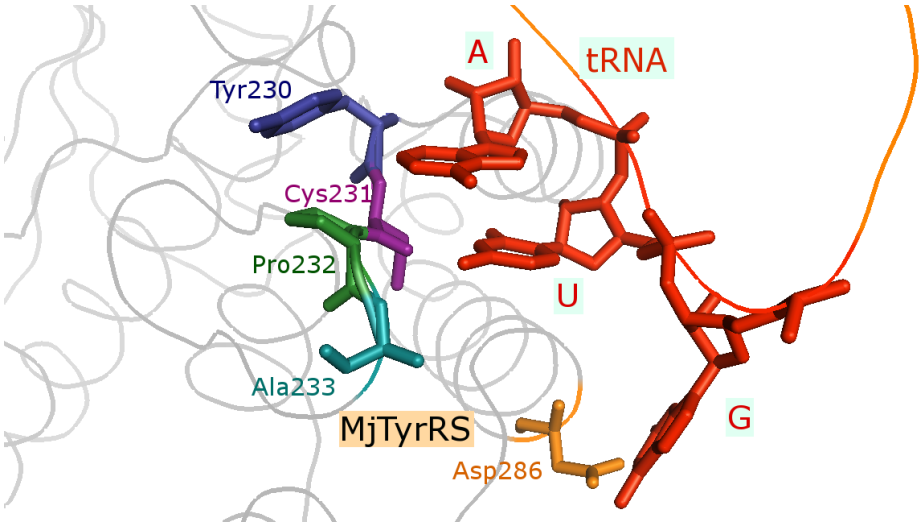


Figure 2.4: Mutation sites on *Mj*TyrRS.

*Mj*tRNA$^{Tyr}_{CCU}$. The library contained the gene of *Mj*TyrRS with four randomly mutations (Tyr230, Cys231, Pro232, and Ala233) within the tRNA anticodon recognizing region(Fig. 2.4). Within the aaRS gene all AGG codons were changed to CGC mutations to avoid any translational ambiguities in response to AGG. Further, a GFP$_{UV}$-Tyr66AGG mutant was developed as a biological reporter to indicate a successful recoding of AGG from arginine to tyrosine. When *Mj*tRNA$^{Tyr}_{CCU}$, *Mj*TyrRS library, and GFP$_{UV}$-Tyr66AGG probes were all expressed in *E. coli* cells, only colonies with strong green fluorescence were selected because they contained desired synthetases which are able to efficiently recode AGG as tyrosine, but without obvious toxicities to bacterial growth.

Construction of pAGG-GFP$_{UV}$ probe vectors

The GFP$_{UV}$-Tyr66AGG reporter was designed to distinguish one single mutation from arginine to tyrosine at position 66. In wild type GFP$_{UV}$ protein Tyr66 is one of the three critical fluorophore amino acid residues, while in the reporter gene GFP$_{UV}$-Tyr66AGG the original codon for Tyr66 was replaced by AGG. Functional GFP$_{UV}$ proteins would only be produced when an evolved aaRS successfully recodes AGG from arginine to tyrosine.

The mutant pAGG-GFP$_{UV}$ probe vectors was derived from pLEIG and pLEIZ (Fig. 2.5). GFP$_{UV}$ -Tyr66AGG(Fig. 2.6) gene was created by PCR with the primers GFP66AGG-f 5- CTT TCT CTA GGG GTG TTC AAT GCT TTT CCC GTT ATC-3 and GFP66AGG-R 5-TGA ACA CCC CTA GAG AAA GTA GTG ACA AGT GTT G-3 and template pLEIG. The GUA anti-codon of *Mj*tRNA, which is coded by the

tRNA gene on the same vector, was altered to CCU (Fig. 2.3) by overlapping PCR with primer 5- GGA CTC CTA ATC CGC ATG GCG CTG GTT C-3 and primer 5- ATG CGG ATT AGG AGT CCG CCG TTC TGC-3. The mutant GFP and wild type GFP were amplified by PCR and cloned into pLEIZ by cutting off the majority part of Z domain gene. The probe vector carries the chloramphenicol acetyltransferase marker.

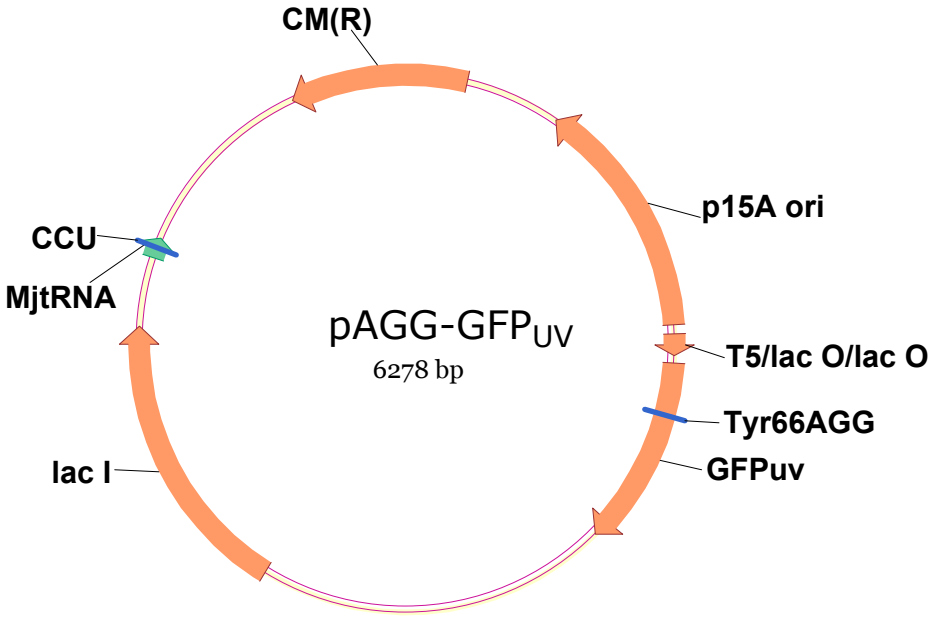


Figure 2.5: pAGG-GFP$_{UV}$ vector.

In pAGG-GFP-his vector, a C-terminal 6His tag was inserted to GFP$_{UV}$ mutant. The 6His tag was carried by mixing primer GFP-7f and GFP-7r. Then the mixture of the primers was then phosphorylated by T4PNK. The pAGG-GFP vectors were cut by BglII and SacI and then ligated with the phorphorylated GFP-7f and GFP-7r mixture.

The DNA sequence of GFP-Tyr66AGG-his is: 5'-atgactagtaaaggagaagaacttttcac tggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtcagtggagagggtgaaggtg atgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggccaacacttgtcacta ctttctctaggggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccg aaggttatgtacaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaagg tgatacccttgttaatcgtatcgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaat acaactataactcacacaatgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgcca caacattgaagatggatccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttac cagacaaccattacctgtcgacacaatctgcccttttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgag tttgtaactgctgctgggattacacatggcatggatgagctctacaaactcgagcaccaccaccaccaccactga-3'

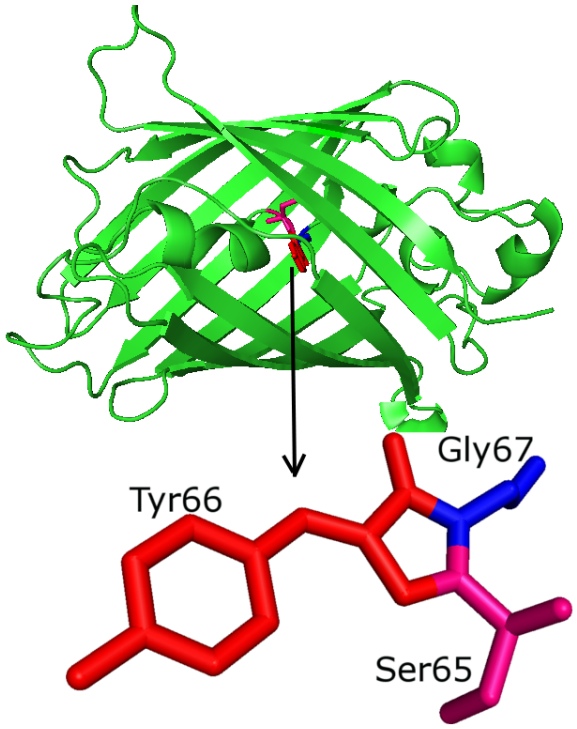


Figure 2.6: GFP protein and its flurophore.

Construction of pBK2-JYRS (pBK-AGG-lib) library

pBK2-JYRS library was derived from pBK-JYRS plasmid (Fig. 2.7). Four randomly mutated sites (Tyr230X, Cys231X, Pro232X, and Ala233X) at the anticodon recognition region of MjTyrRS (Fig. 2.4) were introduced by overlapping PCR with primer pBK2-4f-L-CGC GCT AAG ATA AAG AAA GCA NNK NNK NNK NNK GGA GTT GTT GAA GGA AAT CCA ATA ATG-3 and pBK2-4r 5- CAA ATT TTT CTG GGC GTT TTA TGG TTA AAG GAT ATT CAA GGA AG-3. Five AGG codons on the same vector were altered to CGC by overlapping PCR with primers pBK2-1f/pBK2-1r, pBK2-2f/pBK2-2r, pBK2-3f/pBK2-3r, pBK2-4f/pBK2-4r, pBK2-5f/pBK2-5r. Five AGG codons on the same vector were altered by overlapping PCR to avoid translational ambiguities. The library vector also contains a kanamycin resistant gene.

Screening of pBK2-JYRS library by using pAGG-GFP$_{UV}$ probe vector

The *Mj*TyrRS library and the pAGG-GFP$_{UV}$ vector (including the orthogonal *Mj* tRNA$_{CCU}^{Tyr}$) were co-transformed into *E. coli* to allow the rapid screening for substrate permissivity (either arginine or tyrosine). Transformed cells were inoculated

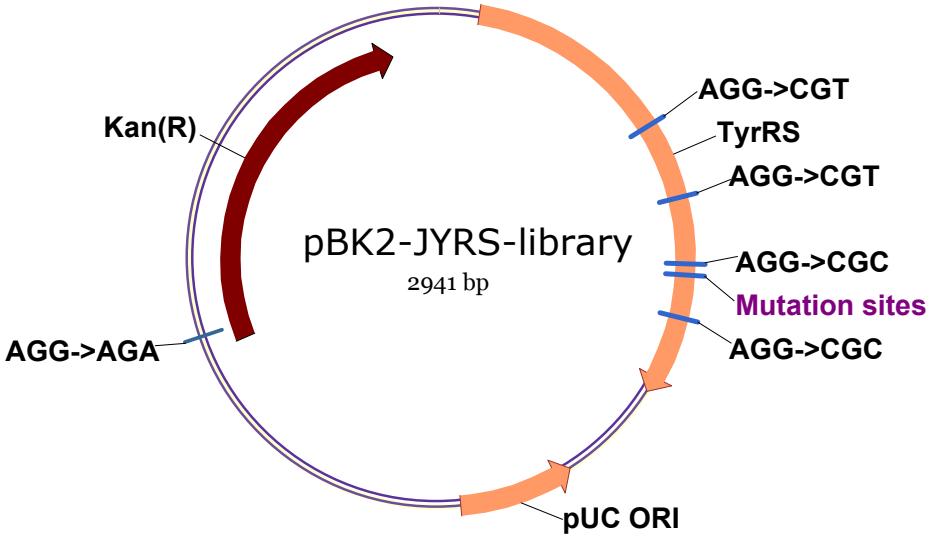


Figure 2.7: pBK2-JYRS library.

on LB agar plates containing kanamycin, chloramphenicol and IPTG (1mM) to culture 24-48 hours. Only the mutated aaRS clones with bright green fluorescence were selected, which indicated the suppression of endogenous arginyl-tRNA RS(ArgRS)/ $tRNA_{CCU}$ pair in response to AGG codon, or mischarging tRNA, mutation to mRNA or DNA.

Clarifying AGG ambiguity of the screening results

We randomly selected 7 (pBK2-9,10,12,11,14,15,16) individual clones which showed strong green fluorescence. The plasmid DNA from these 7 clones was isolated and sequenced. Sequence results indicated that some AGG codons existed among the designated mutation sites within these 7 clones (Table 2.1), and whether these AGG codons were translated into arginine or tyrosine remained unclear. Logically it should be arginine initially, but whether their reassignment to tyrosine would yield a better synthetase needs to be examined. Therefore, in order to eliminate the ambiguity between arginine and tyrosine (R/Y) in response to AGG codon, a serial vectors were created (Table 2.2). Plasmids extracted from the selected clones were used as templates, and the inserted fragments were amplified by overlapping PCR with the following primers: pBK2-9YY, pBK2-9RR, pBK2-9RY, pBK2-9YR, pBK2-10R, pBK2-10Y, pBK2-12R, pBK2-12Y, pBK2-14R, pBK2-14Y, pBK2-15RR, pBK2-15YY, pBK2-15RY, pBK2-15YR. The primers mentioned above coded for either arginine(AGA) or tyrosine(TAC) at the ambiguous mutation sites, separately, and the table "'Clarifying AGG ambiguity" displayed these mutation sites. All PCR prod-

Table 2.1: The AGG ambiguity among 7 screen-out clones.

WT Codon Amino Acid	763TAC 230Tyr	766 TGC 231 Cys	769 CCA 232 Pro	772 GCT 233 Ala
Strain 9	AGG R/Y	ATG Met	AGG R/Y	TGG Trp
Strain 10	AAG Lys	AGG R/Y	ACT Thr	TGG Trp
Strain 11	AGG R/Y	TGG Trp	CAG Gln	AGG R/Y
Strain 12	AGG R/Y	TGG Trp	AAG Lys	CGG Arg
Strain 14	AGG R/Y	GTT Val	GAG Glu	AAG Lys
Strain 15	AGG R/Y	TGG Trp	ACT Thr	AGG R/Y
Strain 16	TGG Trp	TGT Cys	GTT Val	AAG Lys

ucts and their corresponding vectors were double digested by PstI and NdeI and then ligated.

Again, these vectors and the pAGG-GFP probe vectors were co-transformed into Top10 *E. coli* cells to clarify AGG ambiguity for further GFP expression and measurement.Transformed cells were inoculated on LB agar plates containing kanamycin, chloramphenicol and IPTG (1mM) to culture 24-48 hours. If the tRNA/synthetase pairs, then tyrosine will be charged to the tRNA which consequently results in the translation of AGG as tyrosine. Thus the bacterial clone will show fluorescence. Further, the interaction between $tRNA_{CCU}$/synthetases decides the recoding ability of the mutant synthetase . The stronger fluorescence the colony shows, the better-fit for $tRNA_{CCU}$ the synthetase is.

GFP expression and measurement

GFP_{UV} proteins were purified from green fluorescent stains and their fluorescence intensities were measured as well. A single colony from each co-transformation was grown overnight at 37 °C in 2 ml of 2YT media supplemented with kanamycin and chloramphenicol. Overnight cultures were diluted twenty times and grown to OD600=0.6-0.8 before induction. Cultures were induced by 1 mM IPTG and the growth was continued at 37 °C for 10 hours before the cultures were cooled down to room temperature for 30 minutes. Both bacterial green fluorescence intensity and purified GFP protein fluorescence intensity were measured using a Jobin-Yvon

Table 2.2: Clones built to clarify AGG ambiguity.

Strain	Mutation Start position #1	Mutation #1	Mutation Start position #2	Mutation #2
9RY	763	AGG to AGA	769	AGG to TAC
9RR	763	AGG to AGA	769	AGG to AGA
9YR	763	AGG to TAC	769	AGG to AGA
9YY	763	AGG to TAC	769	AGG to TAC
10 R	766	AGG to AGA	n/a	n/a
10Y	766	AGG to TAC	n/a	n/a
12R/14R	763	AGG to AGA	n/a	n/a
12Y/14Y	763	AGG to TAC	n/a	n/a
15RY	763	AGG to AGA	772	AGG to TAC
15YR	763	AGG to TAC	772	AGG to AGA
15YY	763	AGG to TAC	772	AGG to TAC
15RR	763	AGG to AGA	772	AGG to AGA

Fluorolog-3 spectrometer with excitation wavelength 395nm, emission wavelength 509nm, and exposure time 0.5 second. This instrument can perform static photoluminescence measurement and consists of a xenon lamp/double monochromator excitation source and a CCD detector. For bacterial green fluorescence measurement, 1 ml culture was taken and the reading was normalized with its final OD_{600} value.GFP proteins with 6Xhis-tag from different bacterial strains were expressed and extracted under the same condition (the native condition) with Ni-NTA agarose (Qiagen). After Ni-NTA purification, every type of protein was dissolved in the same amount of buffer(50mM Na2HPO4 and 300 mM NaCl, pH=8.0), and the protein concentration was determined by Nano-drop at OD280. Then the green fluorescence intensity was measured under the same condition as bacterial fluorescence measurement with its original concentration and the adjusted concentration OD_{280} value equaled to 0.1.

Recoding AGG to *UAA*s

Approaches to optimize unnatural amino acids incorporation rates

To outcompete with the endogenous ArgRS/$tRNA_{CCU}$ pair and maximize the ***UAA***s incorporation rate, several strategies were adopted. First, the working concentration of ***UAA***s was increased to 3mM, instead of the common 1mM, for the sake of competing with the canonical amino acid arginine. Second, we employed a dual expression vector of aminoacyl-tRNA synthetase (pAK-pCNF-RS_{AGG}) to express *UAA* decorated proteins [55]. The pAK-pCNF-RS_{AGG} vector contains two copies of pCNF-RS_{AGG} genes with two promoters: a modified gln S (constitutive) promoter and an araBAD (arabinose-inducible) promoter. Also, to reduce the competing capability of endogenous ArgRS/$tRNA_{CCU}$ pair, one special *E. coli* strain DH10BW was created, originated from DH10B. We introduced a G1 to A1 mutation in argW gene of DH10BW chromosome, which codes for endogenous $tRNA^{Arg}_{CCU}$, through homologous recombination with the help of pSIM5 plasmid. Last but not the least, the ***UAA*** incorporation rates and protein yields were compared among different *E. coli* strains to find out the best one.

Construction of pAGG3-Zagg vector

pAGG3-Zagg vector is the pAGG-Z-opt vector which carries Z-R7X (X indicates unnatural amino acids, which is coded by AGG codon) gene and the gene for optimized $tRNA^{opt}_{CCU}$ (Fig. 2.8). A mutated Z domain protein (Z-R7X) was used as a model protein to test the ***UAA***s incorporation rate of the evolved synthetase in response to AGG codon. Z domain protein, an engineered analogue of the IgG-binding domain B of Staphylococcal Protein A, has no disulfide bonds and is 69 residues long. To allow a rapid assessment of protein yields, an AGG rare codon was substituted at a permissive site (Lys7) in the gene for the mutant Z domain protein (Z-R7X) with a C-terminal hexameric His tag. In pAGG3-Zagg(the optimized vector) vector, the gene of Z-R7X was under the control of an isopropyl-β-D-1-thiogalactopyranoside (IPTG)-inducible T5 promoter. An optimized $tRNA^{opt}_{CCU}$ was also carried on the same vector with six mutations (GCU→AGG and GGC→CCU) on the T-stem of $tRNA_{CCU}$ (Fig. 2.3)[54, 56]. This vector carries the chloramphenicol acetyltransferase gene as the selection marker.

The sequence of optimized $tRNA^{opt}_{UCC}$ is: 5'-tggtccggcggaggggatttgaacccctgccat gcggattaggagtccgccgttctgccctgctgaactaccgccgg-3'

The sequence of Z-7agg is 5'-atgactagttgactagtgtagacaacaggatcaacaaagaacaacaaa acgccttctatgagatcttacatttacctaacctgaatgaggagcagcgtgatgccttcatccaaagtttaaaagatgac ccaagccaaagcgctaacctttagcagaagctaaaaagctaaatgatgctcaggcgcctaagggatctcatcaccatc accatcactaa-3'.

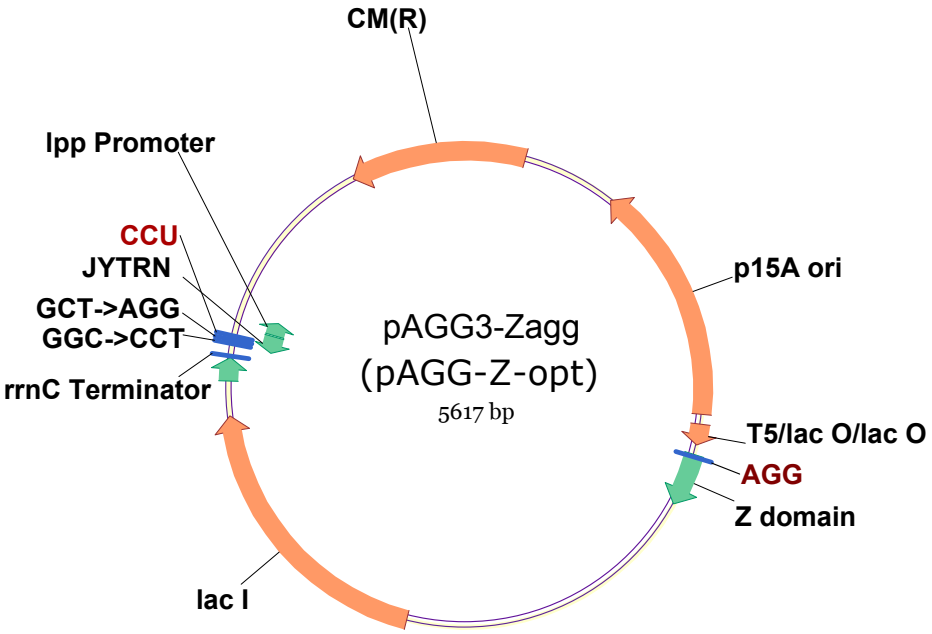


Figure 2.8: pAGG-Zagg3 vector.

The amino acid sequence of Z-7agg is mtsvdnrinkeqqnafyeilhlpnlneeqrdafiqslkdd psqsanllaeakklndaqapkgshhhhhh

To reassign AGG codon to ***UAA***s, we created pCNF-RS$_{AGG}$, of which the amino acid binding pocket was then modified as the following mutations: Tyr32Leu, Leu65Val, Phe108Trp, Asn109Met, Asp158Gly, and Ile159Ala (Fig. 2.9). pCNF-RS$_{AGG}$ is poly-substrate specific against an array of ***UAA***s, including four ***UAA***s which were tested in this study, *p*-acetylphenylalanine (AcF), *p*-azidophenylalanine (AzF), *O*-methyl-tyrosine (OMeY), and p-iodophenylalanine (IF) (Fig. 2.10). Each pAK-pCNF-RS vector contains two copies of pCNF-RS$_{AGG}$ genes with two promoters (a modified gln S (constitutive) promoter and an araBAD (arabinose-inducible) promoter), the araC repressor gene a p15A origin of replication, and the kanamycin resistant gene marker.

The pAK3 vector was obtained from Dr Tsao. Besides the five mutation sites at the amino acid binding pockets which were mentioned in the paragraph above, each pCNF-RS in pAK3 vector contains five mutations, Tyr230Arg, Cys231Trp, Pro232Thr, Ala233Arg, and Asp286Gln[23] (Fig. 2.4)[30] . The first four mutations of amino acids are the same as synthetase isolated from pBK2-15RR.

The pAK5 vector (Fig. 2.11) was derived from pAK3 vector, which was developed during the process of recoding AGA (described in Chapter 3). Every synthetase in pAK5 includes two sets of mutations: one set is consist of five mutations inside the amino acid binding pocket of pCNF-RS, while the other set has four mutations as described in pBK2-12R (Tyr230Arg, Cys231Trp, Pro232Lys, and Ala233Arg)[30].

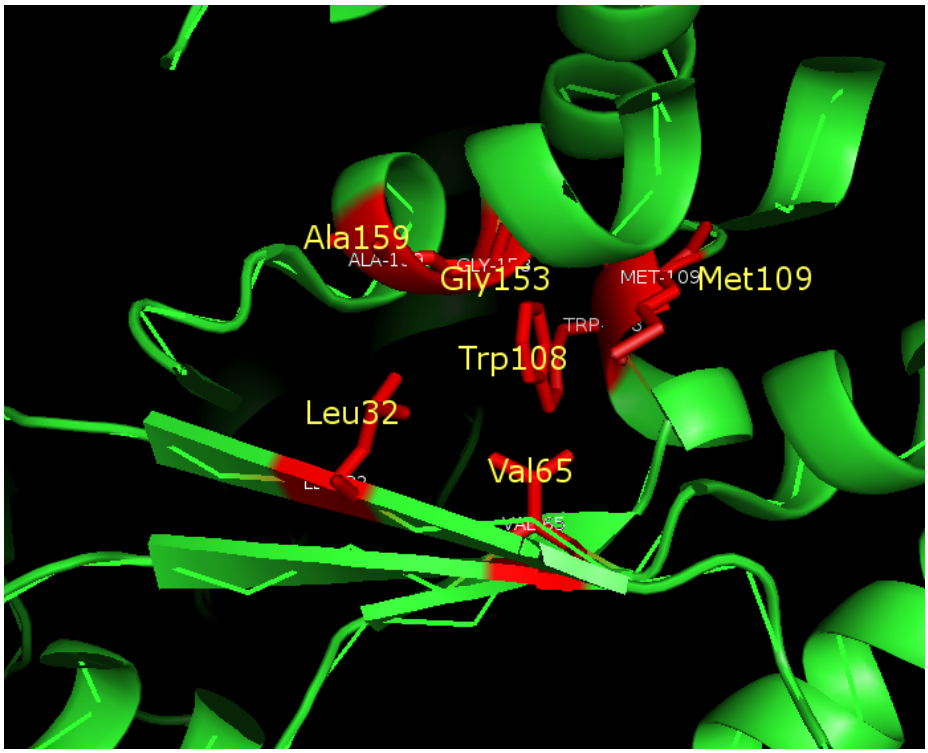


Figure 2.9: The amino acid binding pocket of pCNF-RS.

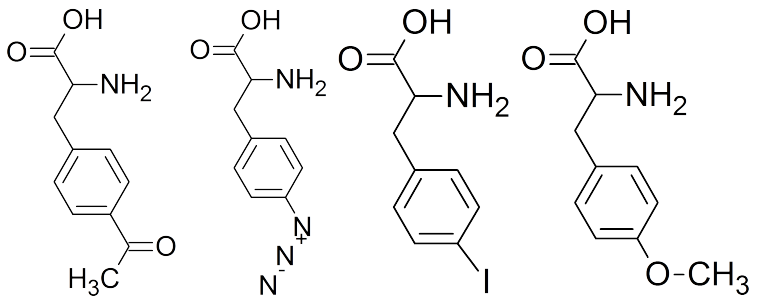


Figure 2.10: The structure of four ***UAA***s: AcF, AzF, IF, OMeY.

Overlapping PCR was used to introduce corresponding mutations for the synthetase gene. In both pAK3 and pAK5 vectors, the synthetase gene after pBAD promoter carries a restriction site (BglII or SacI) at each end, whereas one restriction enzyme (PstI or NedI) cutting site was positioned at each end of the synthetase gene after glnS' promoter. These restriction enzyme cutting sites were utilized for the construction of pAK5 vector from pAK3.

The sequence of pCNF-RS$_{AGA}$ gene in pAK5 is 5'-atggacgaatttgaaatgataaagcgta

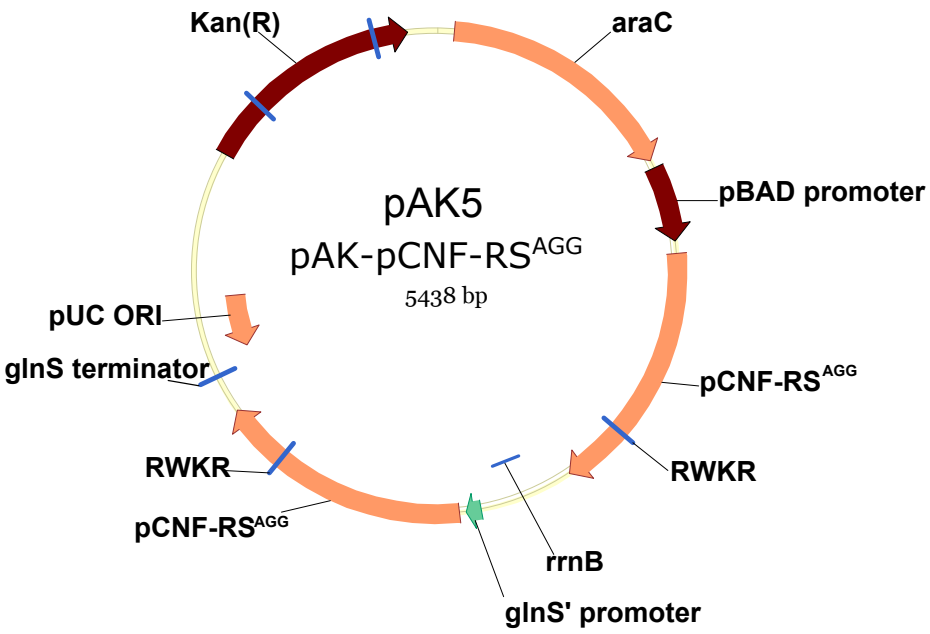


Figure 2.11: pAK5 vector.

acacatctgaaattatcagcgaggaagagttacgtgaggttttaaaaaaagatgaaaaatctgctctgataggttttga
accaagtggtaaaatacatttagggcattatctccaaataaaaaagatgattgatttacaaaatgctggatttgatataa
ttatagttttggctgatttacatgcctatttaaaccagaaaggagagttggatgagattcgtaaaataggagattataac
aaaaaagtttttgaagcaatggggttaaaggcaaaatatgtttatggaagtgaatggatgcttgataaggattatacac
tgaatgtctatcgcttggctttaaaaactaccttaaaacgcgcacgtcgtagtatggaacttatagcacgtgaggatgaa
aatccaaaggttgctgaagttatctatccaataatgcaggttaatggtgctcattatcttggcgttgatgttgcagttggg
gggatggagcagcgtaaaatacacatgttagcacgtgagcttttaccaaaaaaggttgtttgtattcacaaccctgtctt
aacgggtttggatggagaaggaaagatgagttcttcaaaagggaattttatagctgttgatgactctccagaagagatt
cgcgctaagataaagaaagcacgggggacgcgtggagttgttgaaggaaatccaataatggagatagctaaatacttc
cttgaatatcctttaaccataaaacgcccagaaaaatttggtggagatttgacagttaatagctatgaggagttagaga
gtttatttaaaaataaggaattgcatccaatgcagttaaaaaatgctgtagctgaagaacttataaagattttagagcca
attcgtaagcgcttataa-3'

The amino acid sequence of pCNF-RS_{AGA} gene in pAK5 is 5'-mdefemikrntseiise
eelrevlkkdeksaligfepsgkihlghylqikkmidlqnagfdiiivladlhaylnqkgeldeirkigdynkkvfeamgl
kakyvygsewmldkdytlnvyrlalkttlkrarrsmeliaredenpkvaeviypimqvngahylgvdvavggmeq
rkihmlarellpkkvvcihnpvltgldgegkmssskgnfiavddspeeirakikkargtrgvvegnpimeiakyfleyp
ltikrpekfggdltvnsyeeleslfknkelhpmqlknavaeelikilepirkrl*

Creating DH10BW strain by homologous recombination

A unique DH10BW strain was created to carry a G1 to A1 mutant within its genomic argW gene (http://www.ncbi.nlm.nih.gov/gene/6059919), in which the argW gene codes for $tRNA^{Arg}$ with an anticodon of CCU.

Overlapping PCR was used to create the DNA insert that contains G1 to A1 mutation in the argW gene. A bla(Ap) sequence which codes for beta lactamase, was fused with the homologous fragment for further selections. Competent DH10B *E.coli* cells, which were transformed by pSIM5 plasmids, were prepared as previous described . The DNA fragment which contained G1 to A1 mutant argW gene, bla (Ap) sequence, and short homologous sequences at each end, were transformed into pSIM5/DH10B competent cells and cultured on LB agar plates with 50 μg/ml carbornicillin. Individual clones were picked and the sequence was amplified by PCR . The G1 to A1 mutation of argW was confirmed by DNA sequencing and the clone that contained the correct mutation was named as DH10BW (Fig. 2.12) .

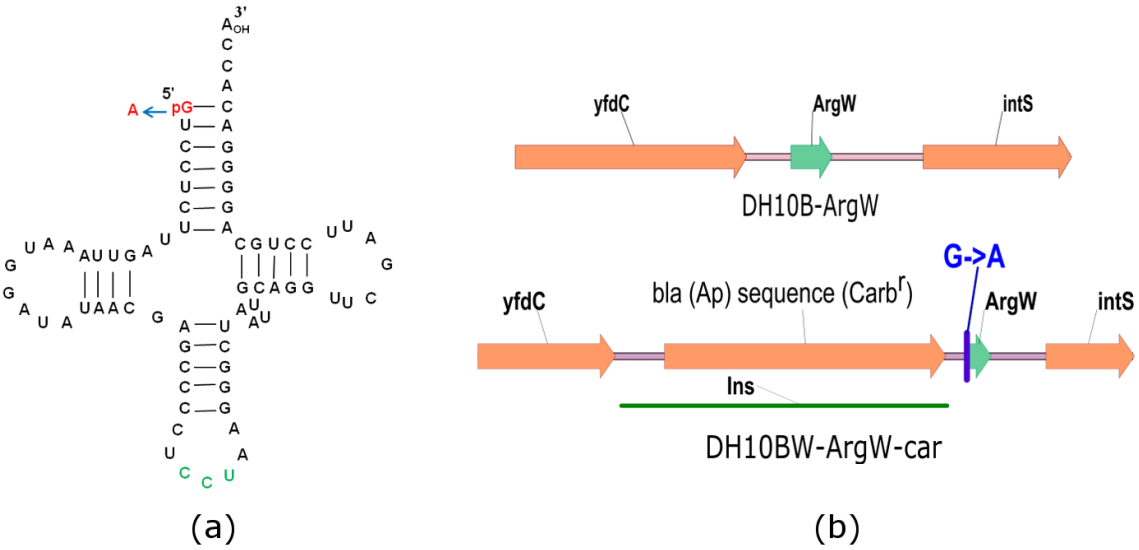


Figure 2.12: The mutation of argW gene. (a), *E. coli* $tRNA_{CCU}^{arg}$ encoded by argW, and the red residue was mutated from G1 to A1 in DH10BW strain; (b) The mutated argW gene with G1 to A1 mutation and the Carb^r gene were inserted into bacterial chromosome through homologous recombination.

The DNA sequence of argW gene in DH10B is : 5'-gtcctcttagttaaatggatataacgag cccctcctaagggctaattgcaggttcgattcctgcaggggacacca-3', and that in DH10BW is: 5'-atc ctcttagttaaatggatataacgagcccctcctaagggctaattgcaggttcgattcctgcaggggacacca-3'

To make the pSIM5/DH10B competent cells for homologous recombination [57], the pSIM5 plasmid which carries the chloramphenicol acetyltransferase marker as the

selection marker was transformed into DH10B *E. coli* first. The keeping of pSIM5 plasmids inside *E.coli* requires 32°C, therefore the transformed bacterial plate was kept at 32°C for further selection. A single colony was selected and cultured overnight. Then the overnight culture was diluted by adding 0.5 ml of the overnight to 35 ml of 2YT medium with 17 mg/mL chloramphenicol, in a 125 ml flask. The overnight at least need to be diluted 70-fold. Cells were grown in an incubator at 32°C with shaking (200 rpm) until the OD600 is from 0.4-0.5 (approximately 2 hrs). Half the culture was transferred to a 50 ml flask and that flask was shaken by hand in a 42°C H_2O bath; the other flask was kept at 32°C with shaking (200 rpm) . After shaking for 20 min. the culture at 42°C was now regarded as 'induced' for the recombination functions and the 32°C culture was the un-induced control. Both flasks would be processed identically during the rest of the protocol. After the 15 min induction, both cultures were rapidly chilled on ice for 5-10 minutes and was centrifuged for 10 min at 4000 rpm at 4°C. The supernatant was discarded and the bacteria cell pellets were re-suspended gently by 1 ml ice-cold sterile distilled H_2O with a large disposable pipet tip (do not vortex), and then another 50 ml of ice-cold distilled H_2O was added to each tube, seal. The diluted culture was centrifuged down again and washed by 50 ml of ice-cold 10% glycerol. After the third centrifuge, the cell pellets were re-suspended with 1 ml pre-chilled distilled H_2O and centrifuged down at 13000rm at 4°C for 20 seconds. The final competent cells were suspended with 200 μl 10% glycerol and kept on ice till use.

Electroporation was adopted for the transformation of homologous recombination targeting DNA fragments into DH10B. The designated DNA fragment was obtained by overlapping PCR and about 200 ng was used for each transformation. 50μl of the electro-competent cells and the DNA were mixed for one electroporation. For recovery, 1 ml SOC was added into each tube after transformation and shaked at 200 rpm at 32°C for 7 hours. 20 μL bacterial culture was innoculated onto one LB agar plate with 17 μg/ml chloramphenicol and 50 μg/ml carbornicillin and cultured at 32°C overnight or till medium-sized colonies appeared. Single colonies were selected, and direct colony PCR was performed to choose colonies with PCR products around 1.5 Kb. DNA sequencing was followed to confirm the homologous recombination.

Expression and purification of Z domain protein

Z domain vector (pAGG3-Zagg or pAGG-Zagg) and pAK-pCNF-RS_{AGG} (pAK3 or pAK5) were co-transformed into different strains of *E.coli* cells, including Top10(F'), DH10B, DH10BW, GM10, BL21(DE3). A single colony from each co-transformation was grown overnight at 37 °C in 2 ml of LB media supplemented with kanamycin and chloramphenicol to obtain the starter culture. The starter culture was then twenty times diluted to a 20 mL LB medium containing 50 μg/ml kanamycin, 34μg/mL chloramphenicol and 1.5mM or 3mM unnatural amino acids (IF, AcF, AzF, OMeY)

for further culturing. Till the culture reach OD600=0.6-0.8, 0.02 percent arabinose was used for induction, and followed by 1mM IPTG induction 10 minutes later. After expression for 14-16 h at 37 °C, 20 ml cultures were pelleted at 8,000 rpm and frozen at -80°C. In order to purify Z domain proteins with a 6His at the C-terminus under native condition, pellets were thawed and lysed for 20 minutes with 2 ml BugBuster (Novagen) supplemented with 0.5 mg/ml lysozyme (Sigma), and 2.5 U/ml bezonase (Novagen). Cell lysates were centrifuged at 7,000 rpm, and the supernatant were collected to with 0.5 ml Ni-NTA agarose (Qiagen). The mixture was gently shaded on a rotary shaker at 4°C for one hour and loaded into a polypropylene column (Qiagen). Columns were washed twice with 2 ml wash buffer (50 mM NaH2PO4 , 300 mM NaCl, 20 mM imidazole, pH 8.0) and eluted twice with 400 μl of native elution buffer (wash buffer plus 230 mM imidazole).

Protein electrophoresis

Eluted samples were analyzed on a 4-12% Tris-Gly SDS-PAGE gel (Invitrogen) and stained with Coomassie brilliant blue. Each 10 μl protein sample in native elution buffer was mixed with 10 μl 2X SDS sample loading buffer and then heated at 70°C for 10 minutes. Every 20 μl heat denatured protein sample was loaded on NuPAGE Novex Bis-Tris mini gel and run at 200 V constant for 35 minutes within 1X MES Buffer. After electrophoresis, the gel was placed in 100 ml of ultrapure water in a loosely covered container and microwaved for 1 minute until the solution almost boils. Then the gel was shaken on a rocker for 1 minute and the water was discarded. The water washing steps were repeated for two more times. 20 ml of SimplyBlue SafeStain(Coomassie brilliant blue) was added into the container and it was microwaved on High for 45 seconds until the solution almost boiling. The gel in boiled Coomassie blue) was shaked 5 minutes, and then the dye was discarded. Finally the gel was washed in 100 ml of ultrapure water until the background was destained enough on a shaker.

Calculating protein yields

Once the buffer to dissolve proteins was changed from native elution buffer, OD/tsb280 was then measured to decide the protein concentration. Based on Vector NTI protein data analysis result, wild type Z domain protein is 69aa long, and molecular weight is 7876 Da. So, 1 unit of OD/tsb280 value of Z domain protein corresponds to about 6.15mg/ml protein. The bacteria culture usually was 25ml, and total eluted protein was dissolved in 0.8 ml distilled water by using 2μl protein solution. So the equation used to estimate Z domain protein yield, if pure, is: $(OD/tsb280 \times 6.15 \times 0.800 \div 25 \times 1000$. The unit of protein concentration is mg/L.

Analyzing protein molecule weight by mass spectrometry

Protein molecular weight was analyzed by mass spectrometry. Protein samples which showed obvious bands with correct sizes on SDS-PAGE gel were sent out for mass spectrometry (ESI or MALDI-TOF). To prepare protein samples for ESI/MALDI-TOF, proteins in native elution buffer were treated by Aspire chromatography tips to eliminate salts and the buffer was changed to 70% acetonitrile/ddH2O.

labeling of unnatural amino acids

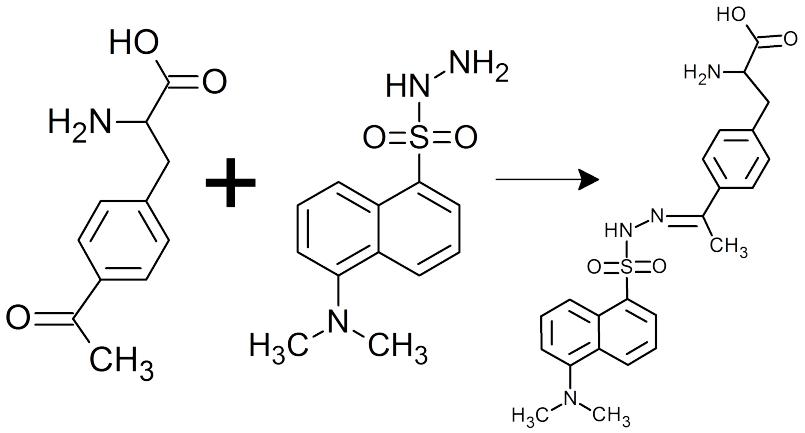


Figure 2.13: DNSH reacts with AcF.

Purified Z-7AcF was labeled by a primary aliphatic amine-reactive probe, 5-dimethylaminonaphthalene -1-sulfonyl hydrazine (Dansyl hydrazine, DNSH). The primary aliphatic amine of DNSH can be reversibly coupled to ketones of AcF to form a Schiff base, and labels the protein carrying AcF as fluorescent dyes (Fig. 2.13)[58]. The emission and excitation spectra of DNSH is 335/525nm (Fig. 2.14). The DNSH labeling reaction was carried by the following steps: mixed targets protein AcF-Z 30μL with 6.5% TFA/ACN 150μL, then added 100μL 5mM DNSH into the mixture. The mixture was shaken gently over night at room temperature (about 24°C).

Toxicity test of pCNF-RS$_{\mathbf{AGG}}$/tRNA$^{\mathbf{opt}}_{\mathbf{CCU}}$

pAK-pCNF-RS$_{\mathrm{AGG}}$ (pAK5) and pAGG3-Zagg which contains one copy of tRNA$^{\mathrm{opt}}{}_{\mathrm{CCU}}$ were co-transformed into GM10 or DH10BW. In DH10BW, one single colony was grown overnight to saturation in LB media supplemented with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol. Culture were 1:100 diluted and divided into 10 samples. Each sample was treated separately and then aliquoted into 96-well assay plates

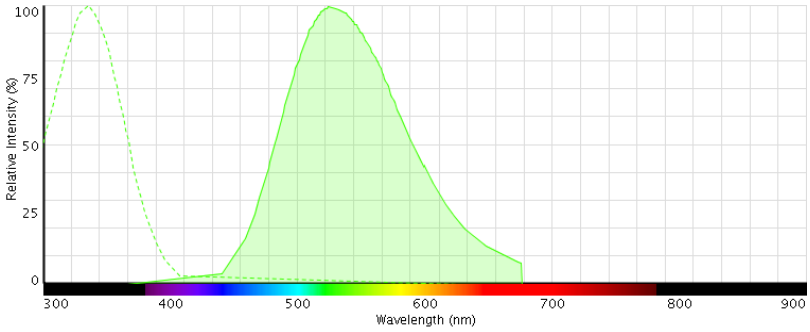


Figure 2.14: The spectra of dansyl. (http://www.invitrogen.com/site/us/en/home/Products-and-Services/Applications/Cell-Analysis/Labeling-Chemistry/Fluorescence-SpectraViewer.html#product=D100)

in octuplicate. The treatments for each sample included: the first sample, no ***UAA*** and no arabinose added; the second sample, no ***UAA*** but 0.02% arabinose added; the third sample, 3mM IF but no arabinose added; the fourth sample,3 mM IF and 0.02% arabinose added; the fifth sample, 3mM AzF but no arabinose added; the sixth sample, 3mM AzF and 0.02% arabinose added; the seventh sample, 3mM AcF but no arabinose added; the eighth sample, 3mM AcF and 0.02% arabinose added. Plates were grown in a SpectraMax M2e plate reader (Molecular Devices) and read at 600 nm every 200 seconds with mixing of 120 seconds/read at 37°C for 16 hours. Relative bacteria densities are indicated for 200 μl cultures and are an average of eight independent cultures. In GM10 cells, the cell growth rates of pAK5+pAGG3-Zagg/GM10 were measured with different concentrations of arabinose: 0%, 0.002%, 0.006%, 0.01%, 0.02%, 0.06%, 0.1%, 0.2%, 0.6% and 1%, with or without 1.5mM IF.

2.4 Results

Recoding AGG from arginine to tyrosine

pBK2-JYRS library were screened by pAGG-GFP$_{UV}$ probe vector

The pAGG-GFP$_{UV}$ probe plasmids and the pBK2-JYRS were co-transformed to Top 10 *E. coli* cells. On a bacteria plate containing transformed cells with both the pAGG-GFP vector and the PBK-2 Library vector under UV light, the cells with functional tRNA/synthetase pairs show green fluorescence. On the plate above, among hundreds of non-green-fluorescent clones, there are many clones with bright or deem green fluorescence under UV light. This result indicates that AGG codon might have been translated to tyrosine (the target amino acid) by using our new engineered aaRS/tRNA pairs. The brighter the green fluorescence the clone shows, the better fit

between the modified synthetase tRNA binding pocket and $tRNA^{Tyr}_{CCU}$ this strain possesses.

We randomly selected 7 (pBK2-9,10,12,11,14,15,16) of many individual clones which showed strong green fluorescence. Each library plasmid DNA from these 7 clones was isolated and sequenced (Table 2.1).

Clarifying the AGG ambiguity between arginine and tyrosine

Table 2.3: The amino acid sequences of pBK2-10R,pBK2-12R and pBK2-15RR.

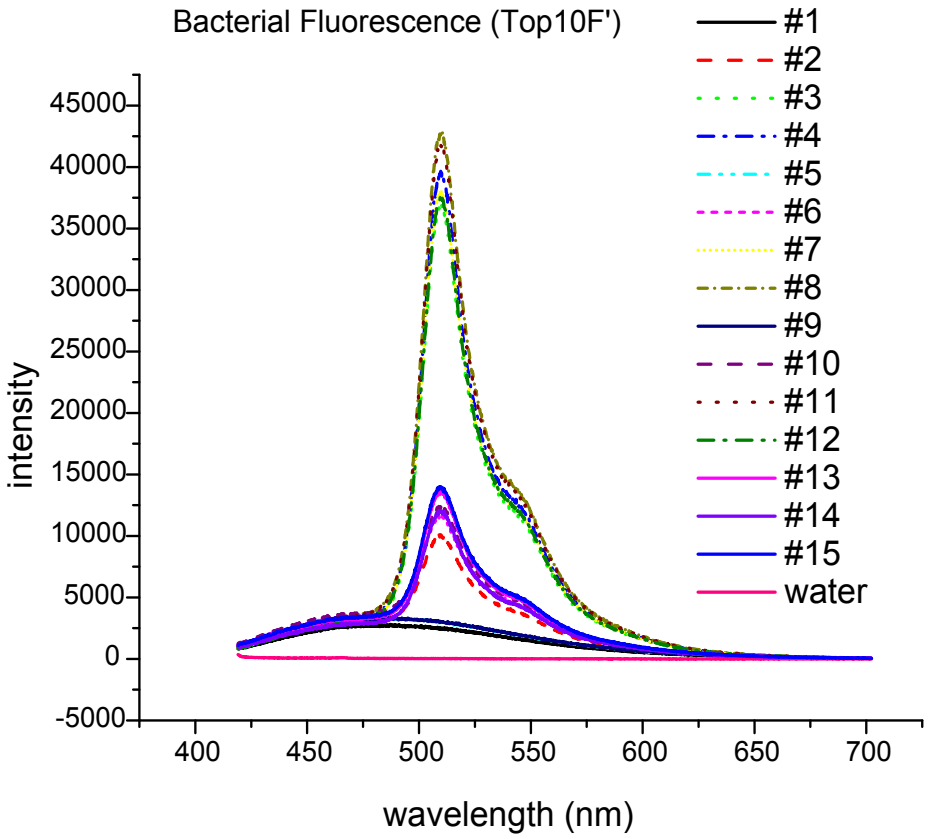
WT (M.j. TyrRS)	230Tyr	231 Cys	232 Pro	233 Ala
Strain 1 (RS-10R)	230Lys	231Arg	232Thr	233Trp
Strain 2 (RS-12R)	230Arg	231Trp	232Lys	233Arg
Strain 3 (RS-15RR)	230Arg	231Trp	232Thr	233Arg

The sequencing results of the 7 clones indicated the appearance of AGG within the designated mutation sites, which caused the ambiguity between tyrosine and arginine (Table 2.1) . In order to eliminate the ambiguity between arginine and tyrosine (R/Y), a serial vectors were constructed by overlapping PCR (Table 2.2). Among the designated vectors including pBK2-9RR, pBK2- 9RY, pBK2-9YY, pBK2-9YR, pBK2-10Y, pBK2-12R, pBK2-12Y, pBK2-14R, pBK2-14Y, pBK2-15RR, pBK2-15YR, pBK2-15YY, and pBK2-15RY, the construction of two vectors, pBK2-14R and pBK2-14Y, were not successful. Then, all available vectors and the pAGG-GFP_{UV} probe vectors were co-transformed into Top10 *E. coli* cells for GFP expression and fluorescence measurement.

Compared among all available strains, three strains containing, pBK2-10R, pBK2-12R and pBK2-15RR returned a result with strong fluorescence, and almost homogenous (Table 2.3). Therefore the AGG ambiguity was clarified: all AGG codons were translated as arginine within the synthetase genes carried by pBK2 vectors because those strains with tyrosine at the original AGG sites did not emit green fluorescence.

Green fluorescence measurement

Bacterial green fluorescence was measured in pBK2-10R+pAGG-GFP-6His/Top10F', pBK2-12R+pAGG-GFP-6His/ Top10F', pBK2-15RR+pAGG-GFP-6His/ Top10F', pAGG-GFP-6His/ Top10F' and pBK-JYRS+pTAG-GFP/ Top10F' after three different induction time (20hours, 22 hours, and 24 hours)(Fig. 2.15). Measurements at all times revealed that synthetase 12R showed similar green fluorescence intensities as the wild type Mj TyrRS did for TAG codon, indicating synthetase12R might have



Inducion Time	pGFP-AGG-his	pBK2-10R + pGFP-AGG-his	pBK2-12R + pGFP-AGG-his	*M.j.* TryRS + pGFP-AGG-his	pBK2-15RR + pGFP-AGG-his
20hours	**#1**	**#2**	**#3**	**#4**	**#13**
22hours	**#5**	**#6**	**#7**	**#8**	**#14**
24hours	**#9**	**#10**	**#11**	**#12**	**#15**

Figure 2.15: Bacterial fluorescence measurements (Top10F').

the strongest ability to incorporate arginine in response to AGG among three selected synthetases. The pBK2-12R+pAGG-GFP-6His/Top10F' displayed green fluorescence around 40,000 a.u., which was at similar level as pBK-JYRS+pTAG-GFP/Top10F' showed. While the fluorescence intensities demonstrated by pBK2-15RR+pAGG-GFP-6His/Top10F' (12,000 a.u. - 14,000 a.u.) were a little higher than those of pBK2-10R+pAGG-GFP-6His/Top10F' (10,000 a.u. - 13,000 a.u.). No significant peak was observed at 509 nm for pAGG-GFP-6His/Top10F', which meant without the help of selected synthetases, pAGG-GFP-6His/Top10F' itself did not emitted any green fluorescence.

After protein purification with Ni-NTA, the protein concentrations which were re-

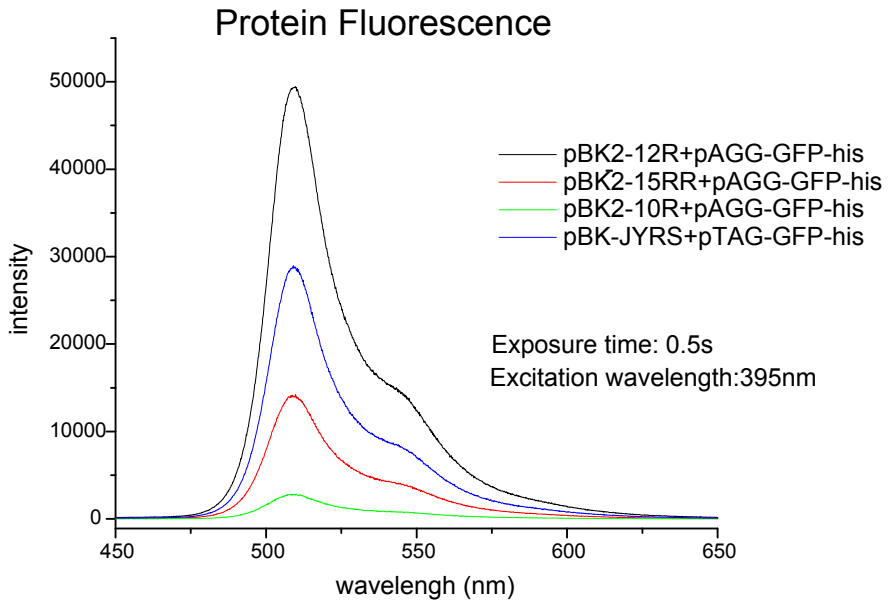


Figure 2.16: Protein green fluorescence measurements without concentration adjustment.

flected by OD_{280} varied among different strains: 0.119(10R), 1.001(12R),0.334(15RR), 0.68(JYRS+GFP-TAG-6his). On SDS-PAGE gel, all four types of GFPUV mentioned above were found closed to 30kDa, consisting with the molecular weight of GFP 28kDa. No GFP_{UV} proteins were observed from non-green-fluorescent strain, i.e., pAGG-GFP-6His/Top10F' , indicating the collapsing or mis-folding of GFP_{UV}-Arg66 proteins (Fig. 2.28). The green fluorescence intensity result of proteins purified from pBK2-10R+pAGG-GFP-6His/Top10F', pBK2-12R+pAGG-GFP-6His/Top10F', pBK2-15RR+ pAGG-GFP-6His/Top10F', and pBK-JYRS+pTAG-GFP/Top10F' were similar as their bacterial fluorescence intensity results, 12R ranking top, followed by JYRS+GFP-TAG-his, then 15RR, and 10R last (Fig. 2.16).

When adjusted all four types of proteins to OD280=0.1, SDS-PAGE still showed different concentrations for GFP band (Fig. 2.29). The non-specific bounded endogenous proteins during Ni-NTA purification might cause this phenomenon. Blended by these impurities, the green fluorescence intensity measurements of protein OD_{280}=0.1 also displayed a ranking of GFP-WT(obtained from Dr Tsao)>12R>15RR>JYRS+GFP-TAG>10R, mainly consisting with previous two fluorescence intensity measurements (Fig. 2.30). Compared the GFP protein concentration on SDS-PAGE gel and its GFP intensity, we clearly saw a correlation between these two factors– the more GFP protein the bacteria produced, the higher green fluorescence intensity would be observed. Therefore, pBK2-12R +pAGG-GFP-6His/Top10F' showed the best ability to synthesize functional GFP_{UV}, and pBK2-15RR+pAGG-GFP-6His/Top10F' fol-

lowed.

Therefore, we chose one mutant aaRS 12R with the highest green fluorescent intensity to construct pAK5 vector. The mutations in the evolved synthetase within pAK5 vector, compared to the wild-type MjTyrRS, include: Y230R, C231W, P232K, and A233R.

Recoding AGG from arginine to unnatural amino acids

Because the AGG is not a terminating codon, the ***UAA*** incorporated Z domain protein would not show any difference from arginine incorporated ones. Thus the SDS-PAGE can not lend us a powerful hand to detect the suppression of AGG codon because all expressed Z domain proteins were a band close to 8 kDa on SDS-PAGE gels. Instead, mass spectrometry was employed for this purpose. Proteins were sent to University of California Berkeley, QB3/College of Chemistry Mass Spectrometry Facility for electrospray ionization mass spectrometry (ESI-MS). The molecular weights of arginine incorporated Z domain protein at position 7 are 7921.6 Da (full length protein), 7832.4 Da (full length protein but without the first methionine; and with post translational modification acetylation, -Met). If the 7th arginine is replaced by designated unnatural amino acid, such as IF/AcF/AzF/OMeY, the molecular weights will be changed correspondingly (Table 2.4).

Determining the incorporation of unnatural amino acids

Table 2.4: Molecular weights of different Z domain proteins.

Molecular Weight	Z-7Arg	Z-7AcF	Z-7AzF	Z-7IF	Z-7NO2F	Z-7NpOH
Full length	7921.6	7954.6	7953.6	8038.5	7958	7978.6
- Met + Ac	7832.4	7865.4	7864.4	7949.3	7868	7889.5
- Met	7790.4	7823.4	7822.4	7907.3	7826	7847.4

Z-7AzF protein was listed here as an example. The SDS-PAGE result of the purified Z-7AcF protein (expressing in the presence of AcF) shows one band near 8 KDa. For proteins produced from DH10BW, the deconvoluted ESI-MS spectrum reveals two major peaks at m/z =7823.2, and 7864.9 (Fig. 2.17, left), which match the expected molecular weight for the AcF incorporated Z-domain mutant with the loss of its first methionine (Z-7AcF-Met, m/z =7823.4) and its subsequent post-translational acetylation product (Z-7AcF-Met+Ac, m/z =7865.4). There is also a minor peak at m/z =7790.2, matching the arginine incorporated Z-domain protein with the loss of

its first methionine (Z-7Arg-Met). Similarly, for protein purified from BL21(DE3), the deconvoluted ESI-MS spectrum shows two major peaks at m/z =7823 and 7865 separately matching Z-AcF-Met and Z-7AcF-Met+Ac, and a minor peak at m/z =7789 matching Z-7Arg-Met (Fig. 2.17, right). Similarly, for other three types of unnatural amino acids decorated Z domain proteins (Z-7AcF, Z-7OMeY, and Z-7IF), each one of these deconvoluted ESI-MS spectra had two major peaks matching the corresponding Z-7***UAA***-Met and Z-7***UAA***-Met+Ac, and a minor peak around m/z =7790.4 matching Z-7Arg-Met (Fig. 2.17).

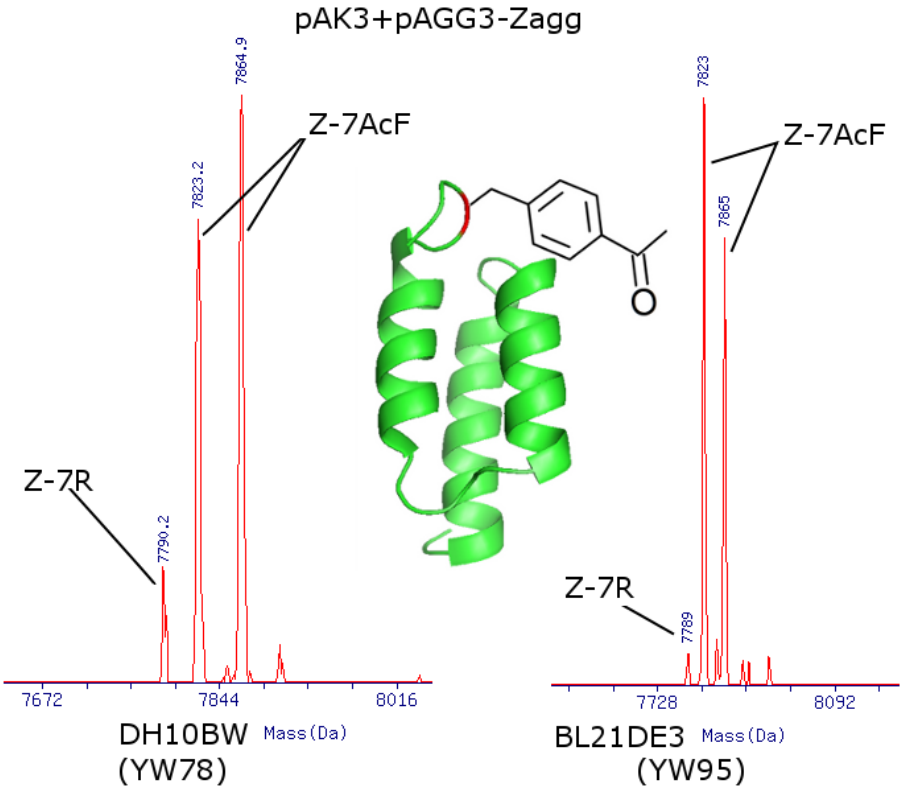


Figure 2.17: ESI-MS results of Z-7AcF from DH10BW(YW78) and BL21(DE3)(YW95).

Given the polysubstrate specificity of pCNF-RS_{AGG}, we were able to evaluate the incorporating efficiency of three ***UAA***s , AzF(Fig. 2.32), AcF(Fig. 2.31), OmeY (Fig. 2.33) and IF(Fig. 2.22), into Z domain protein by pCNF-RS_{AGG}/$tRNA^{opt}_{CCU}$. The relative ***UAA*** incorporation rate was assessed by the equation of ***UAA***s decorated proteins *(Z-7X)/total proteins (Z-7X plus Z-7R).*

Using pAK5 vector to recode AGG to unnatural amino acids in the mutant Z domain protein (Z-7AGG)

Different types of unnatural amino acids, such as IF, AzF, AcF, OMeY, were incorporated into the mutant Z domain proteins, by using pAK5 and pAGG3-Zagg co-translated *E. coli* including BL21(DE3), DH10B, DH10BW, and GM10 (Table 2.6). Top10 was not proved to be a suitable host to reassign the rare codon AGG by giving a poor IF incorporation rate 31% (YW21). Under the same or similar expression condition other bacterial strains, such as GM10, DH10BW, or BL21(DE3) usually gave over 50-80% IF incorporation rates.

Strain matters(Fig. 2.18). Given 1.5 mM as the working concentration of IF and with normal induction condition (0.02%arabinose and 1mM IPTG), the IF incorporation rates (Z-IF/(Z-IF+Z-arg)) varied among different bacteria species. In GM10 the IF incorporation rate was 80- 90% (YW 53,55,16), while in BL21(DE3) it was similar above 80%(YW22). But in DH10B the IF incorporation rate was merely about 55% (YW23) .

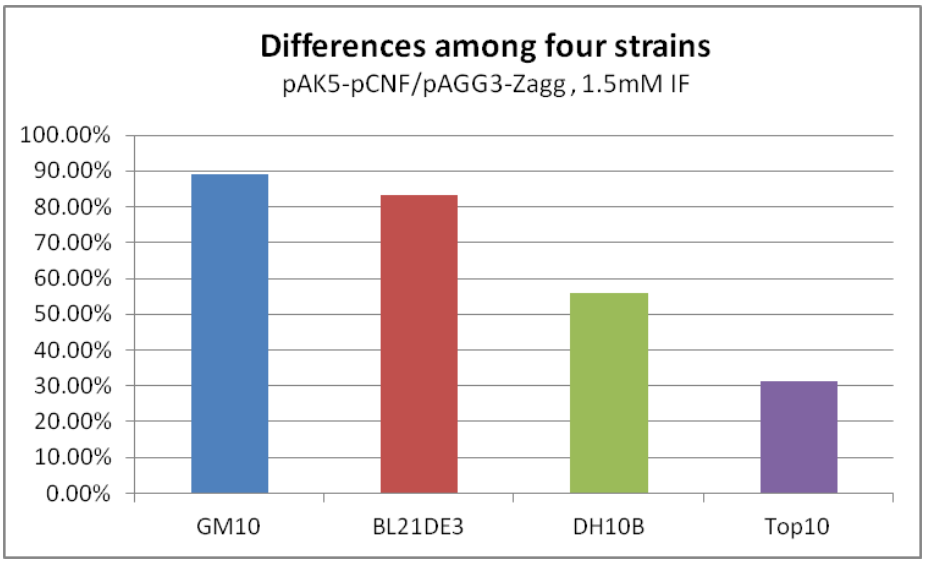


Figure 2.18: The averages of IF incorporation rates vary among different *E. coli* strains in AGG suppresion. Protein expression condition: pAK5+pAGG3-Zagg/GM10, 1.5mM IF. The y axis is the unnnatural amino acid incorporation rate.

***UAA* concentration matters(Fig. 2.19).** Interestingly, if the IF working concentration doubled to 3mM, the IF incorporation rate was then increased to over 70% (YW71), which indicated the working concentration of unnatural amino acids might be one of the important factors to win the competition with the endogenous

translation system. Therefore, we also expressed the same protein within DH10BW bacteria, which contains a mutation in endogenous $tRNA_{CCU}^{arg}$, with 3mM IF. As expected, the IF incorporation rate was pushed to 95% (YW70).

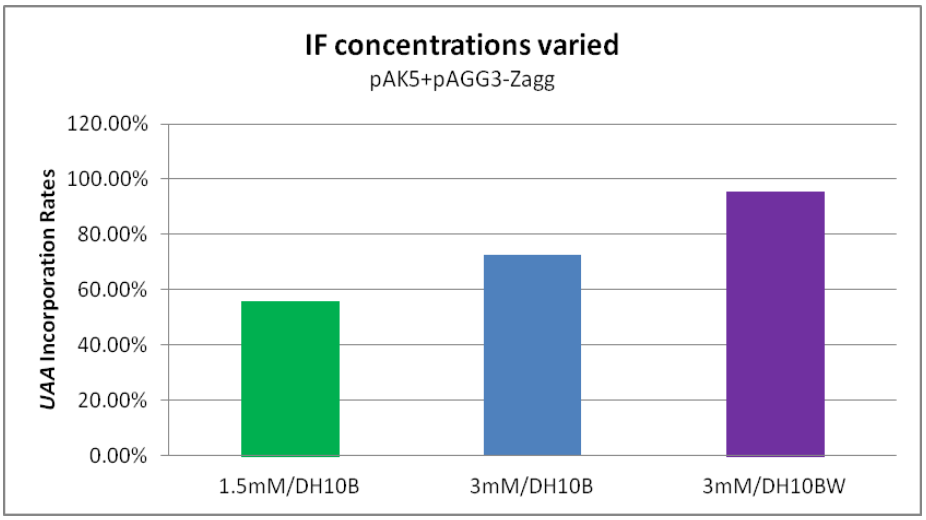


Figure 2.19: IF incorporation rates increased with the increasing of IF concentration during protein expression in AGG suppression. Protein expression condition: pAK5+pAGG3-Zagg. The y axis is the unnnatural amino acid incorporation rate.

Three other types of unnatural amino acids were also incorporated into Z domain protein in response to AGG codon under normal condition (1.5mM ***UAA***, 0.02%arabinose, and 1 mM IPTG). In GM10, the AzF incorporation rate was 57% (YW15) and the OMeY was 64% (YW26), while the AcF one was only about 10%(YW24).

Temperature matters(Fig. 2.20). Since GM10 bacteria can grow under both 37°C and 42°C, we also tested the ideal temperature for the protein expression. Both IF and AzF showed a higher unnatural amino acid incorporation rate under 37°C than 42°C. The IF incorporation rate under 37°C was over 80%(YW9) whereas the 42°C one was about 75%(YW12). More obviously, the AzF incorporation rates increased from 45%(YW13) to 68%(YW10) when temperature decreased from 42°C to 37°C. This experiment was done with pAK5 & pAGG-Z vectors since at that time we did not obtain the optimized $tRNA_{CCU}$ pAGG3-Zagg yet. However, the difference between the optimized tRNA or normal mutant tRNA was ranging from -10% to +10% regarding to ***UAA***-incorporation rates, depending on ***UAA*** types. For example, IF incorporation rates increased about 5% with the optimized tRNA (YW9 *v.s.* YW55), but the AzF one decreased about 10% (YW10 *v.s.* YW15).

Arabinose concentration matters(Fig. 2.21). In light of bacterial growth toxicity test, which will be discussed later of this chapter, we also tested the impact of various arabinose concentrations on AGG codon suppression. In this test, pAK3

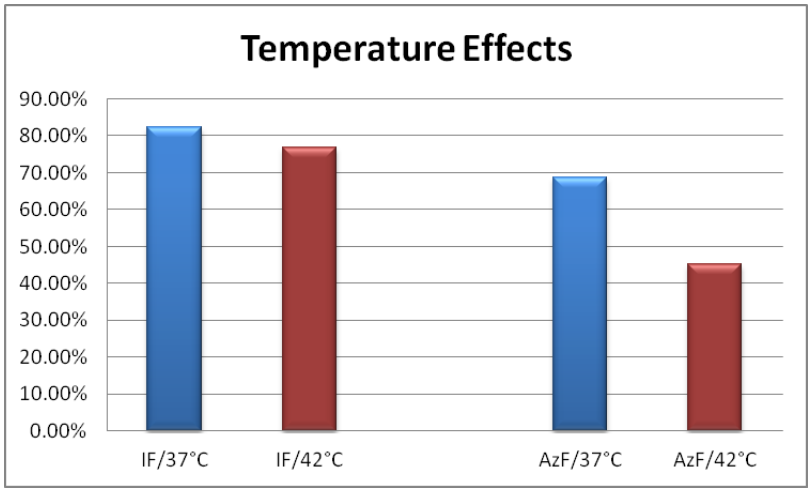


Figure 2.20: Temperature effects on ***UAA*** incorporation:37°C is better than 42°C. Expression condition: pAK5+pAGG-Zagg/GM10, 1.5mM ***UAA***s added. The y axis is the unnnatural amino acid incorporation rate.

and pAGG3-Zagg con-transformed GM10 bacteria were used for protein expression. The induction was initiated by different concentrations of arabinose (from 0.001% to 1%) and followed by 1mM IPTG. From the figure below , we found that the IF incorporation rates increased from over 10% to 92% when the arabinose concentrations went up from 0.001% to 0.1% (YW49-YW57, YW16). But when we added more arabinose(from 0.1% to 1%), the IF incorporation slightly dropped to 85% (YW58-YW60).

Using pAK3 vector to recode AGG to unnatural amino acids in the mutant Z domain protein (Z-7AGG)

The ESI results firmly supported that four ***UAA***s were successfully incorporated (Table 2.7) in three different *E. coli* strains, DH10B, DH10BW, and BL21(DE3) (Fig. 2.22, Fig. 2.32, Fig. 2.17, Fig. 2.31). Enlightened by pAK5 results, we increased the working concentration of unnatural amino acids from 1.5mM to 3mM in order to obtain the optimal unnatural amino acids incorporation rates.

Four unnatural amino acids were incorporated into Z domain proteins in response to AGG (Table 2.5). As we expected, DH10BW bacterial showed a good ability to suppress AGG codon and allow the incorporation of four ***UAA***s (YW69, YW72, YW74, YW76, YW78) . Similarly, in BL21(DE3) the incorporation rates of three types of ***UAA***s (AcF, AzF, IF) were above 90%, and particularly IF-Z and AcF-Z weighted more than 95% out of total Z proteins (YW95-YW98). But for OmeY, BL21(DE3) did not show a good ability to encode it (77%, YW97) whereas

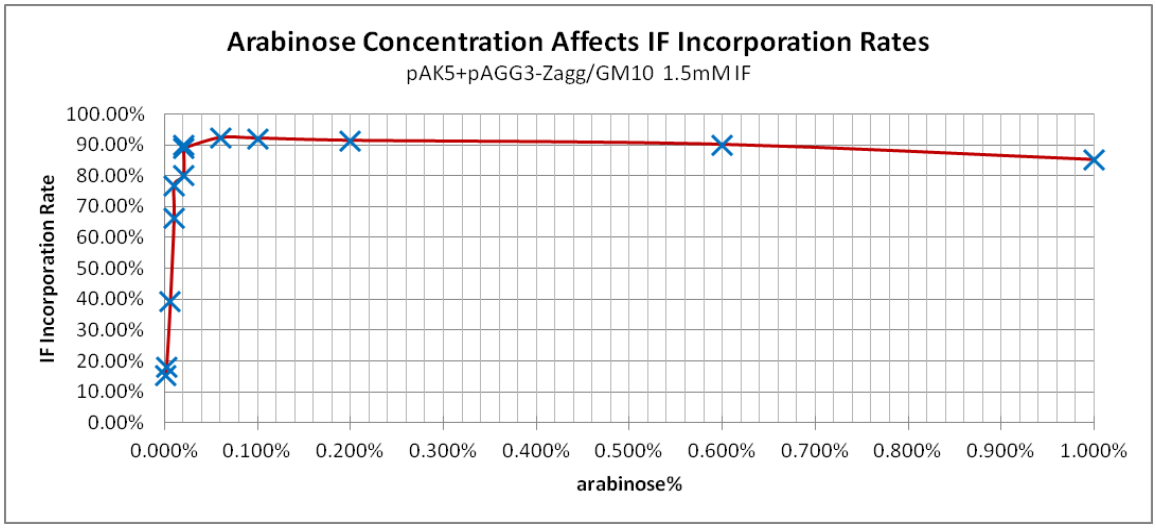


Figure 2.21: Different concentrations of arabinose impact the IF incorporation rates. Protein expression condition: pAK5+pAGG3-Zagg/GM10, 1.5mM IF. The y axis is the unnnatural amino acid incorporation rate.

DH10BW still returned OmeY incorporation rate of 98%(YW76.) Both DH10BW and BL21(DE3) can serve as good hosts for AGG reassigning.

DH10BW showed higher *UAA* incorporation rates than DH10B. The G1 to A1 mutation of endogenous $tRNA^{Arg}_{CCU}$ of host bacteria improved all four ***UAA***s incorporation rates by at least 10%, especially the in-corporation rate of AcF-Z, which was dramatically in-creased from 47% to 90% (Fig. 2.23). When compared with its deriving strain DH10B (YW73, YW75, YW77, YW79), the mutant strain DH10BW demonstrated a stronger ability to charge ***UAA***s into protein, affording ≥85% relative incorporation rates of all three ***UAA***s (YW69, YW72, YW74, YW76, YW78). These results indicate that the G1 to A1 mutation of endogenous $tRNA^{Arg}_{CCU}$ enables more ***UAA***-charged *Mj*tRNAs to out-compete with the endogenous Arg-$tRNA_{CCU}$ during translation, and then increase *UAA*s incorporation rates. Meanwhile, DH10BW still kept a moderate cell growth rate as DH10B did.

pAK3 vector works better than pAK5 vector to reassign AGG codon

When comparing the IF incorporation rates by pAK3 vector and pAK5 vector, we found that in DH10B, the one with pAK5 was only about 70% (YW71) while that with pAK3 went beyond 90% (YW75). More interestingly in DH10BW bacteria, the IF incorporation rates with pAK5 was about 95% (YW70) whereas repeated results by pAK3 were 2%-3% higher (YW69, YW74) than that (Fig. 2.24). Thus we believe that pAK3 vector which provides the synthetase is more suitable for AGG reassignment

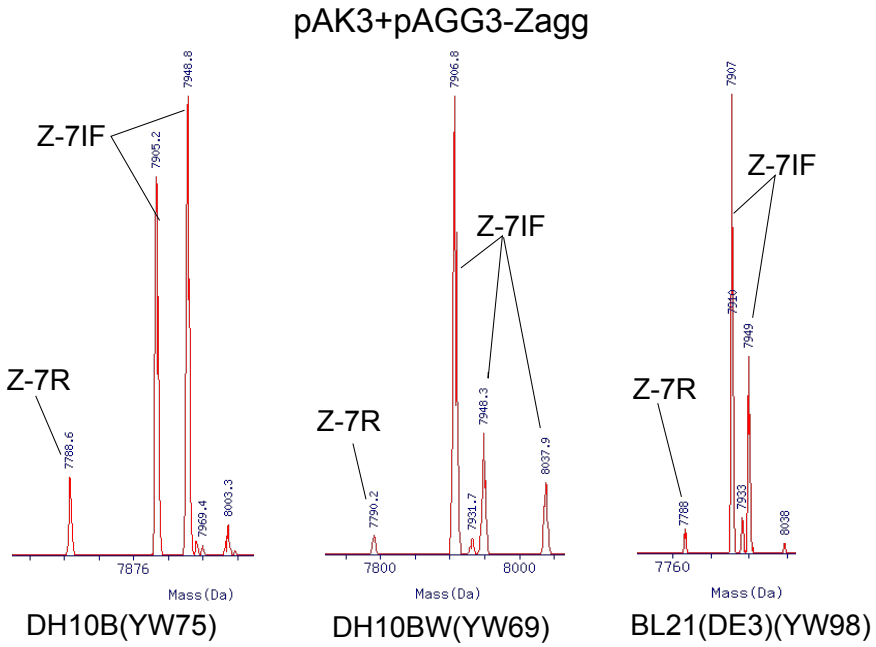


Figure 2.22: ESI-MS results of Z-7IF from DH10B(YW75), DH10BW(YW69) and BL21(DE3)(YW98).

when working with pAGG3-Zagg vector. The later carries both the optimized tRNA and the mutated Z-7AGG gene.

Using pAK7 vector to recode AGG to NpOH in the mutant Z domain protein (Z-7AGG)

Three expression tests were performed by using pAK7+pAGG3-Zagg/GM10 bacteria with 1.5mM NpOH added (Fig. 2.34). The expressions were done at 37°C and the NpOH incorporation rate was merely about 50% (YW41). The deconvoluted ESI-MS spectrum shows peaks at m/z = 7790, 7847, 7889 (YW41), which match the expected molecular weight for the full length arg-Z without the first methionine(m/z=7790.4), the full length NpOH incorporated Z-domain mutant (m/z = 7978.6) and that for this mutant protein with the loss of its first methionine (m/z = 7847.4).

Protein yields

The OD_{280} values of mutant Z domain proteins in water were usually ranging from 0.05(or less) to 0.1, depending on bacteria strains and ***UAA*** types. So, by using the equation $(OD280) \times 6.15 \times 0.800 \div 25 \times 1000$, the results were about 10 mg/L to 20 mg/L. The yields of ***UAA*** decorated Z domain proteins were about 20 mg/L

Table 2.5: Incorporation of ***UAA***s by pAK3+pAGG3-Zagg in DH10B, DH10BW and BL21(DE3).

	Ex-ms	DH10B		DH10BW		BL21DE3	
		Ob-ms	NIR	Ob-ms	NIR	Ob-ms	NIR
AcPhe-Z	7865.4[b]	7864.9[b]	47%	7864.9[b]	90%	7865[b]	97%
	7823.4[c]	7823.2[c]		7823.2[c]		7823[c]	
AzPhe-Z	7953.6[a]	n/a	77%	7954.5[a]	87%	7954[a]	93%
	7864.4[b]	7864[b]		7863.9[b]		7864[b]	
	7822.4[c]	7821.9[c]		7822[c]		7822[c]	
IPhe-Z	8038.5[a]	n/a	82% *	8037.9[a]	96% **	8038[a]	96%
	7949.3[b]	7948.8[b]		7948.3[b]		7949[b]	
	7907.3[c]	7905.2[c]		7906.8[c]		7907[c]	

Ex-ms, Expected mass (Da); Ob-ms, Observed mass (Da). NIR, NAAs incorporation rate (%), which was assessed by the equation of the height of NAAs decorated proteins (Z-R7X)/ the height of total proteins (Z-R7X plus Z-R7) from ESI-MS results. [a], Molecular mass of full length Z domain protein with N-terminal methionine (Z). [b], Molecular mass of full length Z domain protein without N-terminal methionine and with the post translational acetylation (Z-Met+Ac). [c], Molecular mass of full length Z domain protein without N-terminal methionine (Z-Met). *, The average of two replications. **, The average of three replications.

under optimized conditions in BL21(DE3), while the Z domain protein yields were roughly less than 10mg/L in DH10BW and DH10B cells. Therefore, both DH10BW and BL21(DE3) can be candidate host cells for the reassignment of AGG codon, the later has a higher protein yield though.

DHSH labeling of AzF-Z

Further, AcF-Z purified from DH10BW was labeled by a primary aliphatic amine-reactive probe, 5-dimethylaminonaphthalene-1-sulfonyl hydrazine (dansyl hydrazine, DNSH), which can react with the keto group of AcF. The DNSH conjugated AcF-Z displayed green-yellowish fluorescence under UV excitation while the Arg-Z which underwent the same reaction condition did not show any fluorescence, indicating successful incorporation of pAcF into Z domain protein (Fig. 2.25). This result also proves the successful incorporation of AcF in Z domain protein.

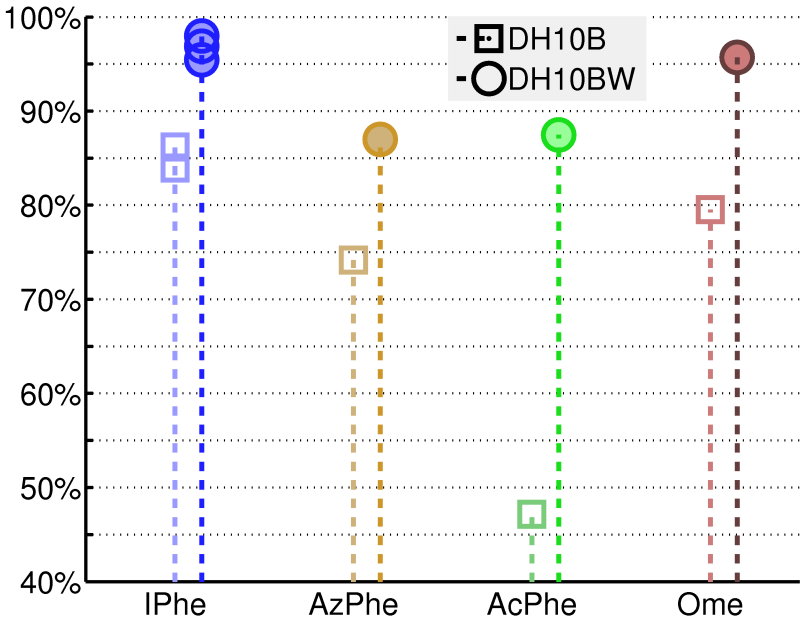


Figure 2.23: DH10BW has higher ***UAA*** incorporation rates than DH10B. The y axis is the unnatural amino acid incorporation rate.

Bacterial growth toxicity test

The cell growth rates of pAK5+pAGG3-Zagg/GM10 were measured with different concentrations of arabinose: 0%, 0.002%, 0.006%, 0.01%, 0.02%, 0.06%, 0.1%, 0.2%, 0.6% and 1%, with or without 1.5mM IF. Without IF added into culture media, the all tested bacteria showed a similar proliferating slope during the lag phase and the exponential growth phase, and in the stationary phase they reached slightly different cell densities(Fig. 2.35). These minor difference might come from the accumulation of over expressed synthetases and toxic tRNAs inside bacterial cells. However, after 1.5mM IF added into culture media, obvious difference appeared. From 0% - 0.2%, the bacterial growth curves displayed a trend that the more arabinose we added, the slower the growth rate was and the less cell density they reached (Fig. 2.26). If the working concentration of arabinose was even more than 0.2%, the bacterial growth was dramatically inhibited, and the bacterial density usually doubled and the curve was quite flat. So these results reminded us that the working concentration of arabinose might play an important role in bacterial growth and the toxicity caused by the newly introduced AGG reassigning machinery.

Furthermore, to determine the toxicity of pAK3-pCNF-RS_{AGG}/$tRNA^{opt}_{CCU}$ ex-

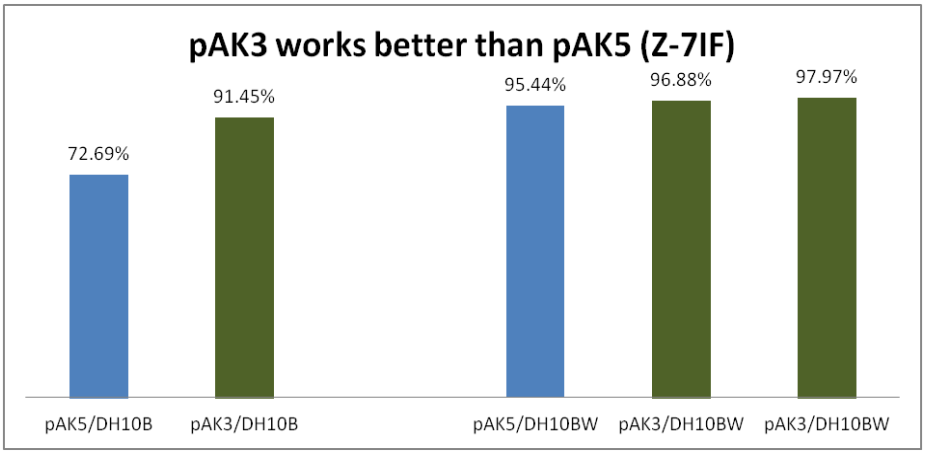


Figure 2.24: pAK3 vector works better than pAK5 vector to incorporate IF into Z domain proteins in response to AGG. Z domain gene and tRNA were carried by pAGG3-Zagg vector, and 3mM IF was added during the protein expression. The y axis is the unnatural amino acid incorporation rate.

pression system, we measured the bacterial growth curve under the presence of 3mM ***UAA***s, with and without 0.02% arabinose, which is the inducing concentration for pCNF-RS_{AGG} protein expression (Fig. 2.27). The results illustrated that without arabinose induction, bacterial growth rates were merely slightly decreased by the low basal level of NAA-charged-$tRNA^{opt}_{CCU}$. After arabinose induction, the up-regulated ***UAA***-charged-$tRNA^{opt}_{CCU}$ moderately lowered the cell saturation level, but these bacteria were still viable and stayed at stationary stage for more 10 hours (figure pAK3pAGG3ZaggDH10BW.png). Thus, this pAK-pCNF-RS_{AGG}/$tRNA^{opt}_{CCU}$ can be applied in *E. coli* recombinant protein expression system without halting bacterial growth, and the desired protein yields will be achieved under optimized conditions, such as extending protein expression time and enlarging the cell culture volume.

The BL21(DE3) bacteria carrying pAK3-pCNF-RS_{AGG}/$tRNA^{opt}_{CCU}$ expression system were also able to proliferate for four days with a 1:500 dilution each day within LB (kana+CM) culture containing 0.02% arabinose. The final cell density of each day after 24 hours culturing ranged from OD_{600}=1.7 to OD_{600}=2.1. It indicated that BL21(DE3) bacteria carrying pAK3-pCNF-RS_{AGG}/$tRNA^{opt}_{CCU}$ expression system were very capable to proliferate, with normal arabinose working concentration of 0.02%.

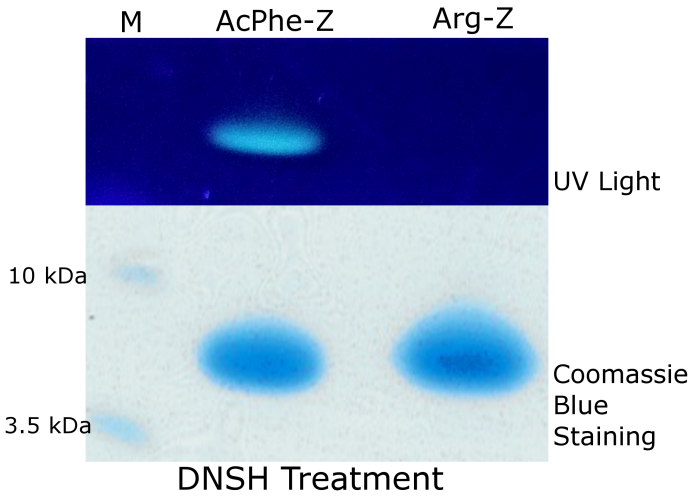


Figure 2.25: Dansyl hydrazine (DNSH) conjugation of AcF. Right: SDS-PAGE gel analysis of DNSH treated AzPhe-Z. Top panel: fluorescent imaging of the gel under UV light (365nm); bottom panel: proteins stained with Coomassie blue of the same gel. The images were captured by a regular camera and shows real colors.

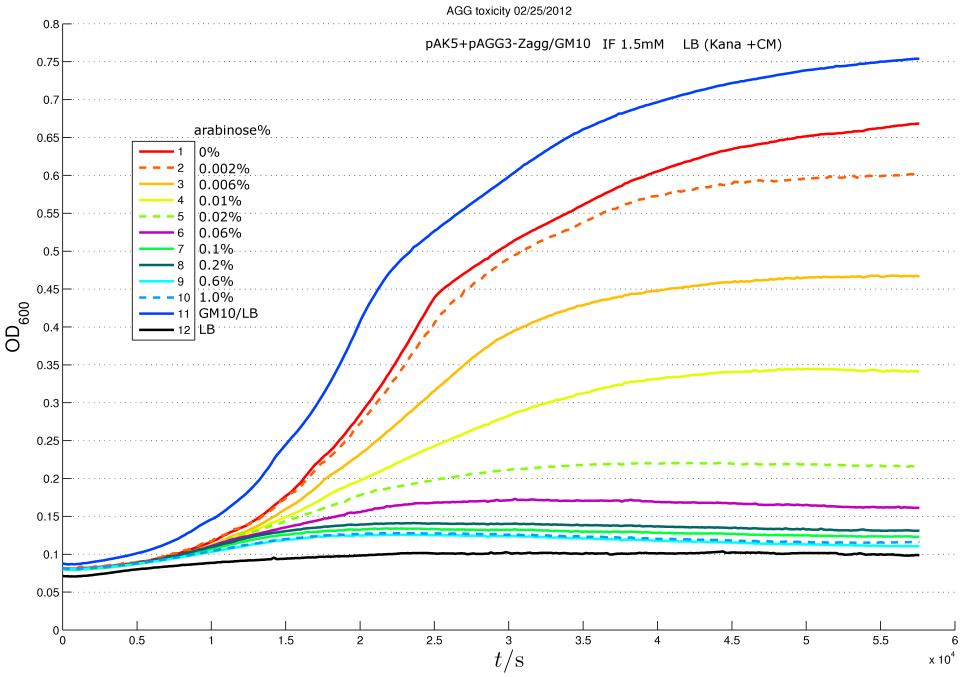


Figure 2.26: The toxicity caused by AGG global suppression during pAK5+pAGG3-Zagg/GM10 cell growth. 1.5mM IF and varied concentrations of arabinose were added at the beginning of cell growth.

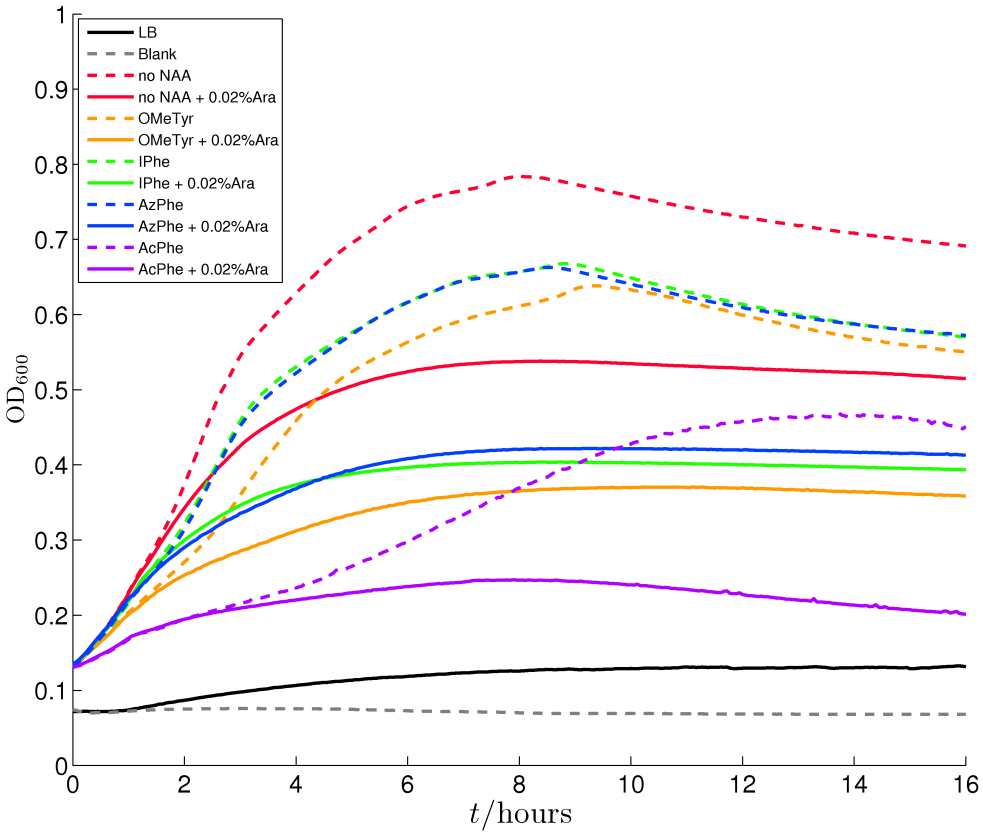


Figure 2.27: The toxicity caused by AGG global suppression during pAK3+pAGG3-Zagg/DH10BW cell growth. 1.5mM different kind of ***UAA***s and 0.02% arabinose were added at the beginning of cell growth.

Other figures and tables

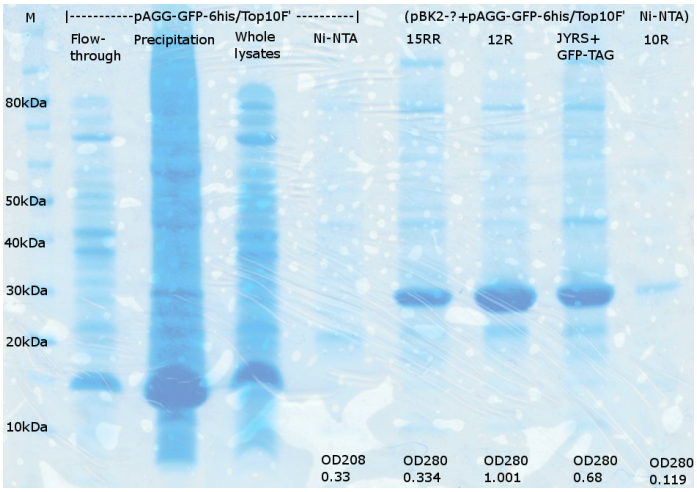


Figure 2.28: SDS-PAGE electrophoresis results of GFP proteins purified from different strains.

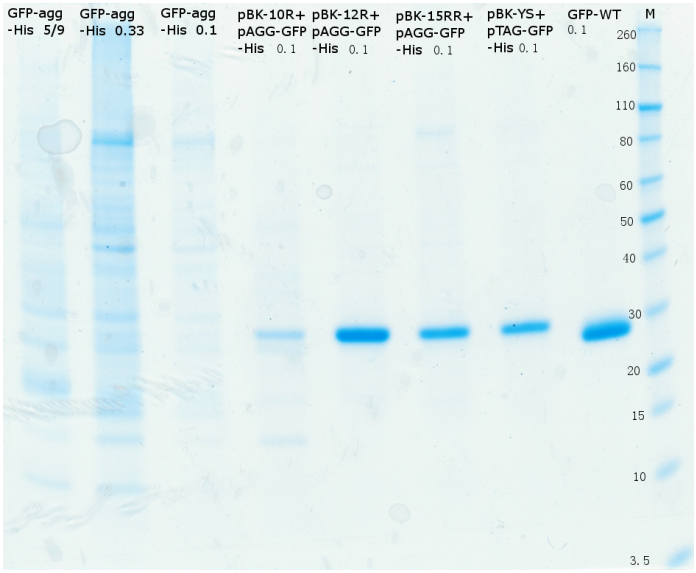


Figure 2.29: SDS-PAGE electrophoresis results of GFP proteins purified from different strains after all protein concentrations were adjusted to OD_{280}=0.1.

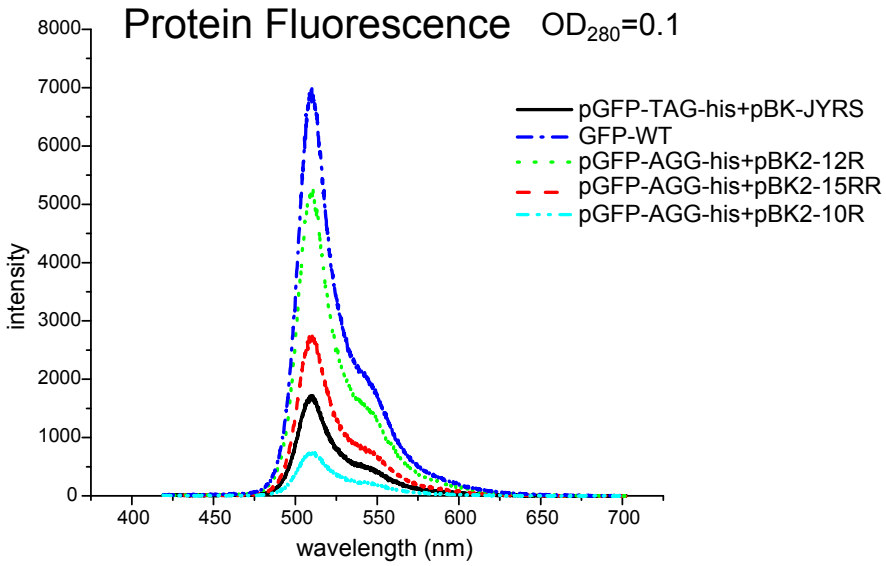


Figure 2.30: Protein green fluorescence measurements after all protein concentrations were adjusted to OD_{280}=0.1.

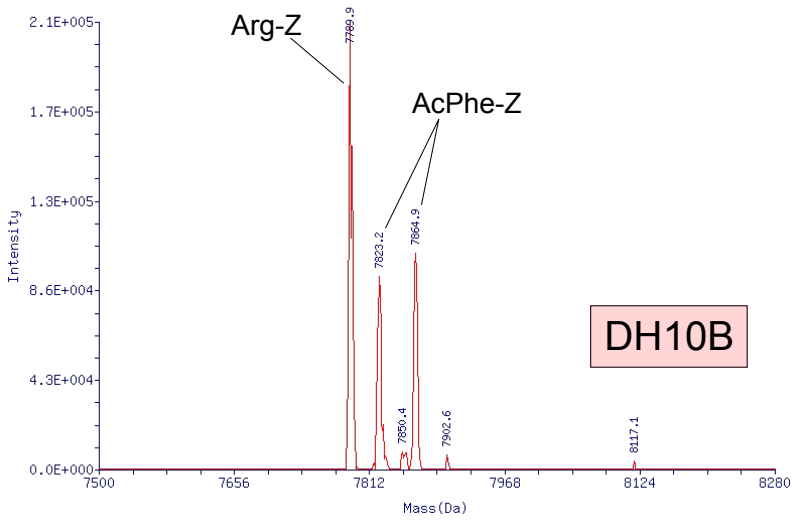


Figure 2.31: ESI-MS results of Z-7AcF from DH10B(YW79). 3mM AcF was added during protein induction.

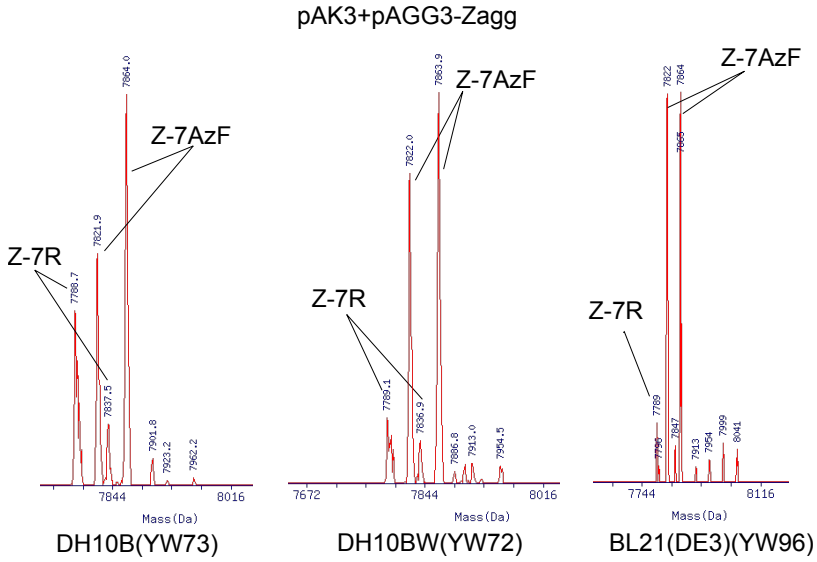


Figure 2.32: ESI-MS results of Z-7AzF from DH10B(YW73), DH10BW(YW72) and BL21(DE3)(YW96). 3mM AzF was added during protein induction.

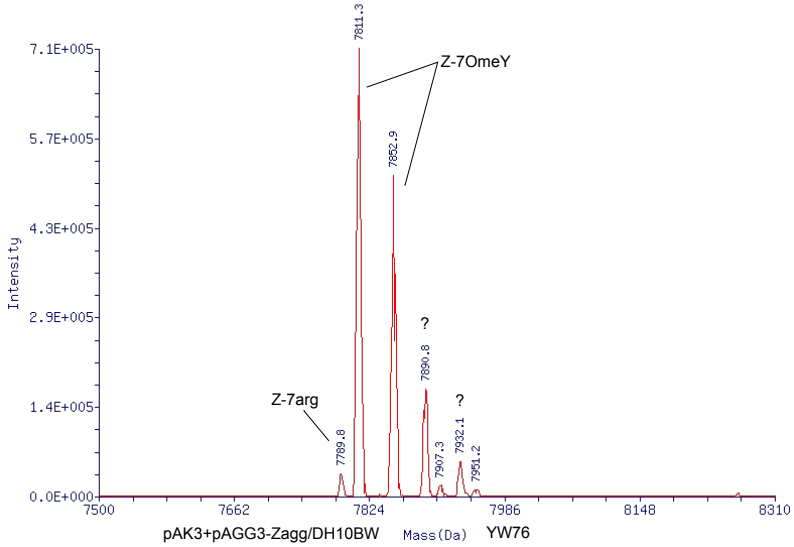


Figure 2.33: ESI-MS results of Z-7OmeY from DH10BW (YW76).3mM OmeY was added during protein induction.

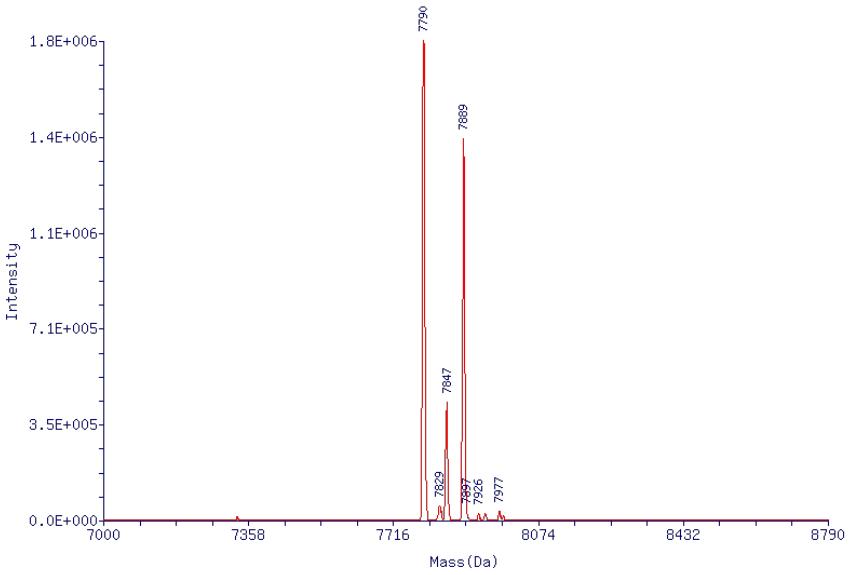


Figure 2.34: ESI-MS results of Z-NpOH from GM10 (YW41). Protein expression condition: pAK7+pAGG3-Zagg/GM10, 1.5mM NpOH .

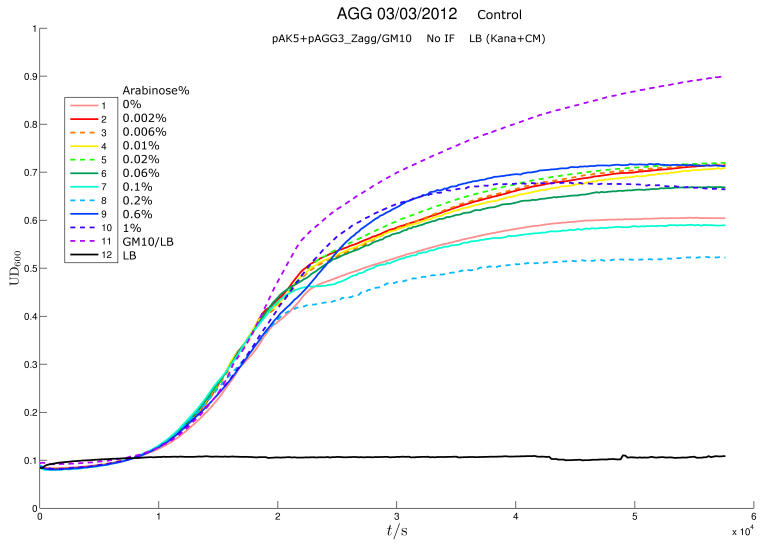


Figure 2.35: Control: the cell growth curve of pAK5+pAGG3-Zagg/GM10 cell. Only varied concentrations of arabinose were added at the beginning of cell growth.

Table 2.6: Varied ***UAA***s were incorporated into Z domain proteins in response to AGG codon by using pAK5 vector.

YW#	UAA	UAA Con.	UAA-Z%	Arabinose%	Temp.	strain	probe vector	synthetase
49	IF	1.5mM	15.38%	0.001%	37°C	GM10	pAGG3-Zagg	pAK5
50	IF	1.5mM	18.04%	0.002%	37°C	GM10	pAGG3-Zagg	pAK5
51	IF	1.5mM	39.18%	0.006%	37°C	GM10	pAGG3-Zagg	pAK5
52	IF	1.5mM	66.18%	0.010%	37°C	GM10	pAGG3-Zagg	pAK5
54	IF	1.5mM	76.90%	0.010%	37°C	GM10	pAGG3-Zagg	pAK5
53	IF	1.5mM	80.16%	0.020%	37°C	GM10	pAGG3-Zagg	pAK5
55	IF	1.5mM	89.86%	0.020%	37°C	GM10	pAGG3-Zagg	pAK5
16	IF	1.5mM	88.93%	0.020%	37°C	GM10	pAGG3-Zagg	pAK5
56	IF	1.5mM	92.30%	0.060%	37°C	GM10	pAGG3-Zagg	pAK5
57	IF	1.5mM	92.11%	0.100%	37°C	GM10	pAGG3-Zagg	pAK5
58	IF	1.5mM	91.38%	0.200%	37°C	GM10	pAGG3-Zagg	pAK5
59	IF	1.5mM	90.08%	0.600%	37°C	GM10	pAGG3-Zagg	pAK5
60	IF	1.5mM	85.17%	1.000%	37°C	GM10	pAGG3-Zagg	pAK5
15	AzF	1.5mM	57.08%	0.020%	37°C	GM10	pAGG3-Zagg	pAK5
26	OMeY	1.5mM	64.36%	0.020%	37°C	GM10	pAGG3-Zagg	pAK5
24	AcF	1.5mM	11.76%	0.020%	37°C	GM10	pAGG3-Zagg	pAK5
25	NO2F	1.5mM	0.00%	0.020%	37°C	GM10	pAGG3-Zagg	pAK5
22	IF	1.5mM	83.22%	0.020%	37°C	BL21(DE3)	pAGG3-Zagg	pAK5
23	IF	1.5mM	55.95%	0.020%	37°C	DH10B	pAGG3-Zagg	pAK5
71	IF	3mM	72.69%	0.020%	37°C	DH10B	pAGG3-Zagg	pAK5
70	IF	3mM	95.44%	0.020%	37°C	DH10BW	pAGG3-Zagg	pAK5
9	IF	1.5mM	82.38%	0.020%	37°C	GM10	pAGG-Z	pAK5
12	IF	1.5mM	76.74%	0.020%	42°C	GM10	pAGG-Z	pAK5
10	AzF	1.5mM	68.75%	0.020%	37°C	GM10	pAGG-Z	pAK5
13	AzF	1.5mM	45.05%	0.020%	42°C	GM10	pAGG-Z	pAK5
21	IF	1.5mM	31.12%	0.020%	37°C	Top10	pAGG-Z	pAK5

Table 2.7: Varied ***UAA***s were incorporated into Z domain proteins in response to AGG codon by using pAK3 vector.

YW#	UAA	UAA Con.	UAA-Z%	Arabinose%	Temp.	strain	probe vector	synthetase
75	IF	3mM	91.45%	0.020%	37°C	DH10B	pAGG3-Zagg	pAK3
73	AzPhe	3mM	77.96%	0.020%	37°C	DH10B	pAGG3-Zagg	pAK3
77	OMeY	3mM	78.26%	0.020%	37°C	DH10B	pAGG3-Zagg	pAK3
79	AcF	3mM	46.90%	0.020%	37°C	DH10B	pAGG3-Zagg	pAK3
69	IF	3mM	96.88%	0.020%	37°C	DH10BW	pAGG3-Zagg	pAK3
74	IF	3mM	97.97%	0.020%	37°C	DH10BW	pAGG3-Zagg	pAK3
72	AzPhe	3mM	86.99%	0.020%	37°C	DH10BW	pAGG3-Zagg	pAK3
76	OMeY	3mM	98.50%	0.020%	37°C	DH10BW	pAGG3-Zagg	pAK3
78	AcF	3mM	90.29%	0.020%	37°C	DH10BW	pAGG3-Zagg	pAK3
98	IF	3mM	96.52%	0.020%	37°C	BL21(DE3)	pAGG3-Zagg	pAK3
96	AzPhe	3mM	93.18%	0.020%	37°C	BL21(DE3)	pAGG3-Zagg	pAK3
97	OMeY	3mM	77.36%	0.020%	37°C	BL21(DE3)	pAGG3-Zagg	pAK3
95	AcF	3mM	96.88%	0.020%	37°C	BL21(DE3)	pAGG3-Zagg	pAK3
8/8/2011	IF	1.5mM	81.75%	0.020%	42°C	GM10	pAGG-Z(agg)	pAK3
7	AzPhe	1.5mM	66.34%	0.020%	42°C	GM10	pAGG-Z(agg)	pAK3

2.5 Discussion

The beauty of GFP fluorophore

GFP is very special among fluorescent proteins. Instead of a separately synthesized group, the fluorophore of GFP is composed of three modified amino acid residues in the middle of the peptide chain [59]. The fluorophore is a p-hydroxybenzylidene-imidazolidone, which consists of residues 65-67 (*Ser* - dehydro*Tyr* - *Gly*) of the protein[60]. After cyclization, the backbone of these residues forms the imidazolidone ring. Actually the amino acid sequence S-Y-G can be found in many other proteins[61, 59]. But this tripeptide does not form such a chromophore, because the SYG sequence from any other non-fluorescent proteins is neither cyclized, nor Tyrosine oxidized.

Tyr66, one of the fluorophore residues, is critical for GFP to emit fluorescence. First, Tyr66 itself sits in the middle of tripeptide, and is the major part of the a *p*-hydroxybenzylidene-imidazolidone [59]. Second, the oxidization of the Cα-Cβ bond of Tyr66 allows the formation of a large delocalized pi-system of the fluorophore[61]. Also, two different forms of Tyr66 decide GFP's excitation wavelength:if it's in its hydroxyl form, the fluorophore absorbs light at 395 nm; whereas in its phenolate form, 470 nm[62, 63, 64]. Furthermore, Tyr66 interacts with the side-chains of some surrounding residues via hydrogen-bonds. The hydroxyl of Thr203 forms a hydrogen-bond with Tyr66, stabilizing the phenolate form of Tyr66[64]. His148, a basic residue, also donates a hydrogen-bond to Tyr66 to stabilize the fluorophore[63]. Since Tyr 66 is the essential part of the fluorophore, the substitution of Tyr 66 with other amino acids might cause dramatic changes of GFP. Previous reports demonstrated that GFP would not show fluorescence if Tyr66 was replaced by a non-aromatic amino acid. The only two fluorescent variants are T66H and T66W [61]. T66H shows a blue shift and T66W has a red-shift, and both give only weakly fluorescent. T66F gives no detectable fluorescence [61]. It has also been demonstrated that the substitution of Tyr 66 by different unnatural amino acids caused shifts of absorbance and emission maxima of the resulting mutant GFPs[65, 66].

Given the essential role of Tyr66 for GFP's fluorophore, we designed a GFP_{UV}-Tyr66AGG mutant as a probe protein to distinguish whether AGG codon is translated to arginine or tyrosine. GFP_{UV}-Tyr66AGG mutant was developed as a biological reporter to indicate a successful recoding. When co-expressed with mutant aaRS libary and $tRNA_{CCU}$, only evolved aaRS can successfully recode AGG from arginine to tyrosine can produce functional GFP_{UV}, namely, only the clone containing such a aaRS/tRNA pair emits fluorescence. Further experimental results approved that this selection strategy was successful to choose desired clones from the synthetase library. Meanwhile, those clones did not show fluorescence must have arginine inserted at position 66 and we even could not purify GFP-6his proteins from these non-fluorescent clones. As we described above, Tyr66 is also a critical part to stabilize the barrel

shape of GFP. It is possible that replacing Tyr66 by a basic residue, such as arginine which has a very different shape and charge from tyrosine, might cause the collapse of GFP's tertiary structure, or cause other conformational change to bury its C-terminal histidine tag into the protein. So our finding is also consistent with the previous reports that GFP would not show fluorescence if Tyr66 was replaced by a non-aromatic amino acid [61].

More important, this type of GFP_{UV}-Tyr66NNN mutants (N denotes for any base) can be applied for other codon reassignments while still using tyrosyl-tRNA synthetase/$tRNA^{Tyr}$ pairs. For instance, a GFP_{UV}-Tyr66AGA was used to select evolved synthetases to recode AGA from arginine to tyrosine (Chapter 3). Other two rare codons for arginine, CGA and CGG, may also go through a similar recoding process by using GFP_{UV}-Tyr66CGA and GFP_{UV}-Tyr66CGG, separately, as selection markers. Theoretically GFP_{UV}-Tyr66NNN mutants can be used to recoding any other codons, even the ones for histidine and tryptophan as long as the blue-shift and the red-shift can be easily observed. However, the codon frequency within the host genome should be accounted because the recoding of a common codon can lead to a fatal destiny of the host. Nevertheless, the nature of cyclized tripeptide fluorophore of GFP makes it an elegant but powerful selecting tool for sense codon recoding.

A battle to fight with the natural translation system

After screened out the synthetase which can both recognize UCC anticodon and host the unnatural amino acids, a big challenge stopped our way to recode the AGG codon. The endogenous arginyl-tRNA synthetase/$tRNA^{Arg}_{CCU}$ pair is so powerful that the newly evolved ***UAA***-RS/$tRNA^{UAA}_{CCU}$ could barely compete with its natural counterpart Arg-RS/$tRNA^{Arg}_{CCU}$.

An effective synthetase vector is a must-be to recode the sense codon AGG

Experiences tell us that a powerful synthetase carrying vector, pAK-RS, is a must-be to recode AGG effectively and to boost the protein yield. The dual vectors we finally obtained, pAK3 and pAK5, showed strong abilities to suppress AGG codon. Each pAK-RS dual vector contains two copies of $pCNF\text{-}RS_{AGG}$ genes with two promoters, namely, a modified gln S (constitutive) promoter and an araBAD (arabinose-inducible) promoter. The orthogonal synthetase is normally maintained at a very low level by glnS'-RS, so that the toxicity caused by the global AGG recoding of the host bacterial cell will be also kept at the minimum level. While because of the presence of araBAD-RS, the intracellular ***UAA***-RS level will be quickly increased when arabinose is added into culture media. By using this type of dual pAK-RS vectors, we can synchronize the RS expressing time with the target protein induction time and di-

minish the toxicity brought by AGG reassignment. On the contrary, at beginning we used pBK-rs vectors which carry one RS gene copy after glnS promoter, or pSup-RS vectors (one copy of RS gene under a glnS' promoter) for protein expression. Unfortunately the ESI-MS results revealed that in response of AGG, only a very small amount of ***UAA*** was incorporated into Z domain proteins while over 95% of amino acids at 7th position were still arginine by using these single RS carrying vectors with constitutive promoters. Thus, we conclude that a single RS gene under a constitutive promoters is not good enough for AGG recoding, whereas the dual vectors (pAK-RS) are more capable to suppress AGG effectively.

argU *v.s.* argW

In another project of recoding AGA, results demonstrate that a GM10 *E. coli* strain, which has a critical mutation on its argU $tRNA_{UCU}^{Arg}$, is the best host among various *E. coli* strains (Chapter 3). In GM10 bacteria the intracellular level of endogenous $tRNAs_{UCU}^{Arg}$ is dramatically reduced because of an argU10(Ts) mutation in its argU gene[67, 68]. This argU10(TS) mutant generates tRNAs carrying a G1-to-A1 transition which results in a mismatched base pair at the terminus of the acceptor stem, and further a mal-functional tRNA.

So we wondered if a similar mechanism can be applied to the AGG decoding. Therefore we created an unique DH10BW strain, which contains a similar G1 to A1 mutation, but within its genomic argW gene. In *E. coli* AGG is mainly decoded by argW $tRNA^{Arg}_{CCU}$ ($tRNA_5^{Arg}$)[68]. Due to this G1 to A1 mutation, the phenotype of DH10BW was predicted to be similar as that of GM10: the endogenous $tRNAs_{CCU}^{Arg}$ coded by argW could be mal-functional as well. Indeed the MS results showed improvements of ***UAA*** incorporation rates (Fig. 2.23). This strategy of single site mutation within tRNA gene is easy but very powerful, so it can be applied to other sense codon recoding.

According to Wobbling effects, the products of argU gene,$tRNA_{UCU}^{Arg}$ [69] is able to recognize two rare arginine codons, AGA and AGG. However researches indicate that the argU $tRNA_{UCU}$ mainly recognizes AGA, and rarely works with AGG[67, 70]. Actually AGG codon is majorly decoded by argW $tRNA_{UCC}$, and maybe a very small amount of argU $tRNA_{UCU}$ participates in AGG decoding, if any[67, 70]. This might explain why the reassignment of AGG codon from arginine to ***UAA***s (about 90%) in DH10BW is not as good as that of AGA(about 95%) in GM10, because the orthogonal ***UAA***-tRNA has to compete with both argU $tRNA_{UCU}$ and argW $tRNA_{UCC}$. So if we want to completely recode AGG codon, it might be a direction to mutate argU and argW at the same time. However,attempts failed to create a host strain that contains simultaneous mutations in both of these genes.

Elevated amounts of cytosolic *UAA*s probably increased *UAA*-charged orthogonal tRNAs

Meanwhile we also evaluated the impacts brought by the variations of another element of the orthogonal translational machine – the working concentration of unnatural amino acids. To our surprise, the increase of ***UAA*** working concentration dramatically improved its incorporation efficiency . Simply changing from 1.5mM to 3mM, the ***UAA*** incorporation rates increased 30%. More interestingly, with 3mM ***UAA***, BL21(DE3) bacteria demonstrated an improved capability (over 90%) to incorporate various ***UAA***s into proteins in response to AGG codon. Besides, BL21(DE3), as a bio-engineered strain to produce proteins, yields four to five times more proteins than DH1OBW did. This interesting phenomenon indicates that an elevated level of free ***UAA***s inside bacterial cells can increase the chance of orthogonal tRNAs to be charged, and further improve AGG codon suppressing rate. However, this strategy is only applicable for those non-toxic ***UAA***s, such as IF. For those ***UAA***s which are toxic to bacteria growth, for example NO_2F, the ***UAA*** concentration has to be limited under certain 'safety' levels and a specific strain such as DH10BW will be necessary.

Experienced learned from the recoding of AGG

The successful AGG reassigning was accomplished through adopting the combination of two major strategies, enhancing exogenous ***UAA*** incorporation and limiting endogenous arginine charging. Among various methods to achieve these two targets, we figured out it is more powerful to work on the elements of the translational machine, such as aaRS, tRNA, ***UAA***, than the factors affects the process. Actually, at beginning before applying the pAK dual vector, we also tried many methods to improve the unnatural amino acid incorporation rates. First, we used a special GMML culture media which contains no arginine but is enriched of other amino acids plus Thiamin HCl and Biotin D[71]. Second, cyanate was added during protein expression because cyanate was reported to inhibit the arginine biosynthesis specifically[72]. Third, Colicin D was introduced into bacteria to specifically cleave cytoplasmic $tRNAs^{Arg}$ [73]. A pBAD-ColD-RS plasmid was constructed to carry both the synthetase gene and the Colicin D gene. Unfortunately all these attempts to reduce the incorporation of endogenous arginine hardly improved the ***UAA*** incorporation rates, even by combined methods.

Therefore we summarize that the optimal AGG recoding condition includes: using pAK-RS dual vectors (pAK3/pAK5) to carry two copies of synthetase and optimized orthogonal tRNAs, selecting proper host bacteria such as BL21(DE3) or DH10BW, and giving an optimized ***UAA*** concentration (for example, at least 3mM for the IF, AcF, AzF and OMeY) during protein expression.

2.6 Conclusion

In summary, we reassigned the sense codon AGG from arginine to ***UAA***s. The successful suppression of AGG codon depends on the orthogonal ***UAA***-tRNA amount in cytoplasm, the correct recognition of tRNA anticodon CCU by the ***UAA***-RS, and the interaction between ***UAA***s and the ***UAA*** binding pocket of ***UAA***-RS. The dual expression system also plays an important role to out-compete with the endogenous aaRS/tRNA. Further, for the purpose of increase ***UAA*** incorporation rate, it could be a feasible and effective way by reducing the level of endogenous functional tRNAs.

This innovative reassignment of the sense codon AGG makes it possible to site-specifically incorporate multiple ***UAA***s into individual proteins when combining AGG with other blank codons, such as amber stop codon UAG. More importantly, similar strategies can be applied to recode other sense codons to ***UAA***s and provide more powerful tools to study protein functions with novel properties.

Chapter 3

Recoding AGA codon to Unnatural Amino Acids

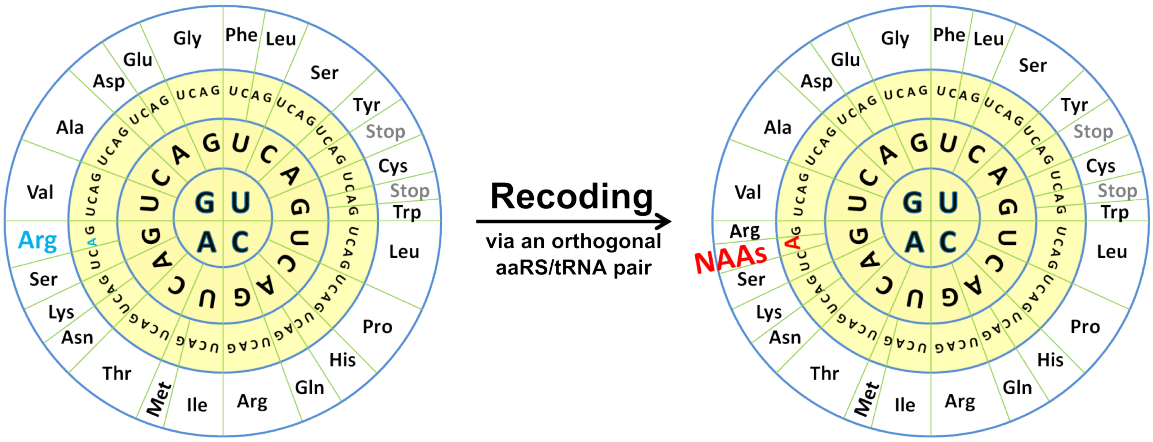


Figure 3.1: The recoding of AGA codon. ***NAA***, the abbreviation of non-canonical (unnatural) amino acid.

3.1 The two-step selection strategy of *UAA*-RS$_{\mathrm{AGA}}$/tRNA$_{\mathrm{UCU}}$

AGA, another rare codon in *E. coli*,is also a good candidate of "blank" codon for unnatural amino acids. In order to show that our approach in the reassignment of a sense codon could be generally applied to codons other than AGG, we next examined the recoding of the AGA codon. Among rare codons, the AGA codon frequency

in *Escherichia coli*(Table 6.1) is merely about 2.4 per 1000 of the entire bacterium genome and about 5% among all six codons for arginine[53].

To discover orthogonal aminoacyl-tRNA synthetase/tRNA pairs, we applied a similar two-step selection scheme. The first step was to reassign AGA from arginine to tyrosine through the screening of its anticodon recognition loop to enable it recognize UCU anticodon instead of original GUA. In this selection system, a fluorescence-based GFP reporter was used as described in Chapter 2. Then the second step was to reassign AGA to unnatural amino acids by editing the amino acids binding pocket of the aaRS. The co-translational incorporation efficiency of *UAA*s largely depends on the recognition between tRNA anticodon UCU and the evolved aaRS, how ***UAA***s can fit into the amino acid binding pockets, the ***UAA***- charged tRNA amounts inside the cell, and the compatibility between new tRNAs and the elongation factors EF. All these factors decide whether our new ***UAA***-RS/tRNA pairs can out-compete with natural-occurred aaRS/tRNAs or not.

3.2 Materials

Top10 *E. coli* (Invitrogen) was used for cloning and amplifying plasmids. DNA was isolated from E. coli. using the QIAGEN Plasmid Mini Kit (Qiagen). Expand high fidelity PCR system (Roche) was used for PCR. Restriction enzymes were obtained from New England Biolabs. BL21(DE3) and DH10B *E. coli* were purchased from Invitrogen. GM10 *E.coli* and GM241*E.coli* were obtained from Yale University Coli Genetic Stock Center. The working antibiotic concentrations were as follows: choramphenicol 68 μg/ml; kanamycin 50 μg/ml and they were purchased from Sigma-Aldrich or Fisher. All protein expression bacteria were grown in LB (Difco) media and other cultures were carried in 2YT (Difco) media. All constructed plasmids were confirmed by DNA sequencing at UC Berkeley DNA Sequencing Facility. Primers were synthesized at Integrated DNA Technology (San Diego, CA) and their sequences were listed in Table PRIMER. Unnnatural amino acids were purchased from Bachem or synthesized from our own lab. Proteins were sent to University of California Berkeley, QB3/College of Chemistry Mass Spectrometry Facility for Electrospray ionization (ESI) mass spectrometry.Liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis was done at Stanford University Mass Spectrometry (SUMS).

3.3 Methods

Reassign AGA codon from arginine to unnatural amino acids

General approach

To reassign AGA from arginine to tyrosine, an *Mj*TyrRS library, pBK4-JYRS(AGA) library, was established. The library contained the gene of MjTyrRS with five randomly mutations (Tyr230, Cys231, Pro232, Ala233, Asp286) within the tRNA anticodon recognizing region (Fig. 2.4). Meanwhile the anti-codon GUA of $Mj\text{tRNA}^{\text{Tyr}}_{\text{GUA}}$ which was coded by the gene on the same vector was altered to UCU(Fig. 3.2). Further, a GFP_{UV}-Tyr66AGA mutant was developed as a biological reporter to indicate a successful recoding. In wild type GFP_{UV} protein Tyr66 is one of the three critical fluorophore amino acid residues, while in the reporter gene GFP_{UV}-Tyr66AGA the original codon for Tyr66 was replaced by AGA (Figure 1c). The functional GFP_{UV} would be only produced when an evolved aaRS can successfully recode AGA from arginine to tyrosine.

Construction of pAGA-GFP probe vector

In order to reassign AGA codon from arginine to tyrosine, a GFP_{UV}-Tyr66AGA probe was used for tyrosyl-tRNA_{UCU} synthetase selection . In the gene of GFP_{UV}-Tyr66AGA, the original codon for tyrosine 66 was replaced by AGA. Tyrosine 66 is one of the three critical amino acid residues of GFP fluorophore(Fig. 2.6). This reporter afforded fluorescence only when the AGA codon was used as a tyrosine codon for the acylated suppressor tRNA_{UCU}, but not as an arginine codon. Further, the intensity of green fluorescence produced by a selected clone was correlated with the specificity between the mutant tyrosyl-tRNA_{UCU} synthetase and the UCU anticodon on tRNA_{UCU}.

The mutant pAGA-GFP_{UV} was derived from pAGG-GFP. GFP_{UV}-Tyr66AGA was created by PCR with the primers GFP66AGA-f and GFP66AGA-R and template pLEIG. The anti-codon GUA of *M.j.* tRNA which is coded by the gene on the same vector was altered to UCU (Fig. 3.2). The mutant GFP and wild type GFP were amplified and then cloned into pLEIZ. The pAGA-GFP was obtained from Dr Tsao, and I synthesized pAGA-GFP-6His verctor based on pAGA-GFP. The C-terminal 6His tag was inserted to GFP_{UV} mutant by the ligation of primer mixture of GFP07f/GFP07r and SacI/BglII doubled digested pAGA-GFP vectors. The probe vector carries choramphenical resistant gene.

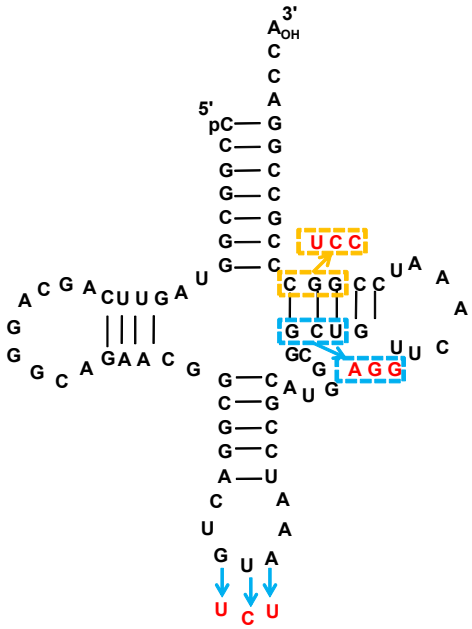


Figure 3.2: The orthogonal $tRNA_{UCU}$ from the archaebacteria *M. jannaschii*.

Construction of pBK4-JYRS(AGA) library

Based on the results of AGA recoding, Dr Tsao used pBK2-15RR, which was described in Chapter AGG, as the template to build the pBK2-JYRS library. Five randomly mutated sites (Tyr230X, Cys231X, Pro232X, Ala233X, and Asp286X) at the anticodon recognition region of *Mj*Tyr RS were introduced. The library vector also contained a kanamycin resistant gene as the selection marker.

Library selection

The pAGA-GFP probe plasmids and the pBK4-JYRS(AGA) library were co-transformed to Top 10 *E. coli* cells. Transformed cells were inoculated on LB agar plates containing kanamycin, chloramphenicol and IPTG (1mM) to culture 24-48 hours. Only the colonies with bright green fluorescence were selected by hand picking. Plasmids were extracted and then sent for DNA sequencing.

The clone which gave the brightest fluorescence, pBK-AGA (pAK1) was chosen for further experiments.

Reassign AGA to unnatural amino acids

Approaches to optimize *UAA* incorporation

To outcompete with the endogenous ArgRS/$tRNA_{UCU}$ pair and maximize unnatural amino acid incorporation rate, two approaches were employed. First, we employed dual vectors pAK-RS_{AGA} of aminoacyl-tRNA synthetase to incorporate ***UAA***s into proteins, which contains two copies of the gene for aminoacyl-tRNA synthetase under different promoters. Second, to reduce the competing capability of endogenous ArgRS/$tRNA_{UCU}$ pair, we used one special *E. coli* strain GM10 , which contains an argU10(Ts) mutation in its argU gene[68]. The mutation leads to a G-to-A transition at the 5 end of the argU tRNA and generates a mismatched base pair at the terminus of the acceptor stem (Fig. 3.3). Thus GM10 cells have reduced intracellular levels of the $tRNA_{UCU}{}^{Arg}$, which are probably caused by the defective interactions with elongation factor Tu and arginyl-tRNA synthetase[74, 75].Reports indicated that at high temperature (42-43°C) the endogenous argU10(Ts) tRNA was reduced to a very low level. Therefore we tested the protein expression under both normal condition 37°C, and 42°C .

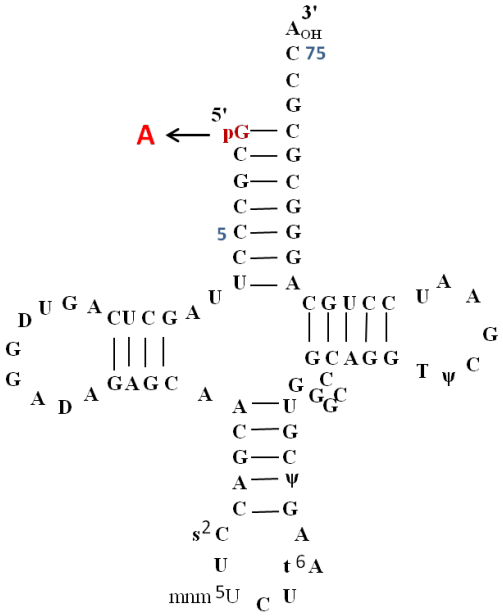


Figure 3.3: Secondary structure of argU $tRNA^{Arg}_{UCU}$ in *E. coli*. The argU10(Ts) mutation is a G-to-A transition at position 1, as indicated in red. The modified nucleotides in this tRNA include dihydrouridine (D), 2-thiocytidine (s2C), 5-methylaminomethyluridine (mnn5U), N6-threonine carbamoyl adenosine (t6A), pseudouridine (φ), and 5-methyluridine (T).

Construction of pAGA3-Zaga vector

pAGA3-Zagg vector is the pAGA3-Zaga vector which carries Z-R7X (X indicates unnatural amino acids, which is coded by AGA codon) gene and the gene for optimized $tRNA^{opt}_{UCU}$ (Fig. 3.4). A mutated Z domain protein (Z-R7X) was used as a model protein to test the unnatural amino acids incorporation rate of the evolved synthetase in response to AGA codon. Z domain protein has no di-sulfide bonds and is 69 amino acids long. To allow a rapid assessment of protein yields, an AGA rare codon was substituted at a permissive site (Lys7) in the gene for the mutant Z domain protein (Z-R7X) with a C-terminal hexameric His tag. In pAGA3-Zaga(the optimized vector) vector, the gene of Z-R7X was under the control of an isopropyl -D-1-thiogalactopyranoside (IPTG)-inducible T5 promoter. An optimized $tRNA^{opt}_{UCU}$ was also carried on the same vector with six mutations (GCU→AGG and GGC→CCU) on the T-stem of $tRNA_{UCU}$ (Fig. 3.2)[54, 56]. This vector carries the chloramphenicol acetyltransferase marker as the selection marker.

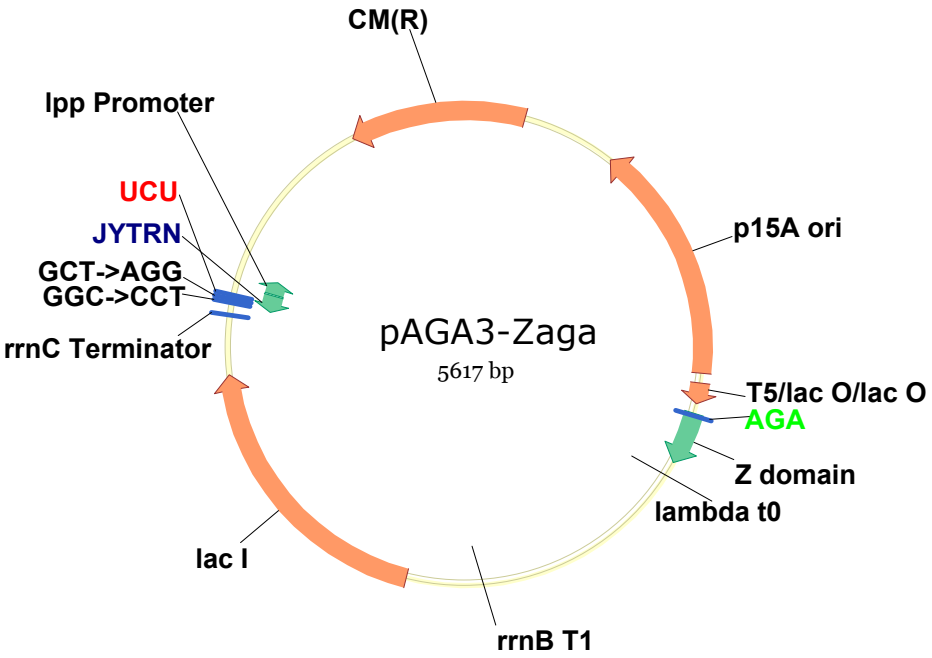


Figure 3.4: pAGA3-Zaga vector carries Z-66AGA gene and the optimized $tRNA_{UCU}^{opt}$.

The sequence of optimized $tRNA^{opt}_{UCU}$ is: 5'-tggtccggcggaggggatttgaacccctgccat gcggattagaagtccgccgttctgccctgctgaactaccgccgg-3'

The sequence of Z-7aga is 5'-atgactagtgtagacaacagaatcaacaaagaacaacaaaacgccttc tatgagatcttacatttacctaacctgaatgaggagcagcgtgatgccttcatccaaagtttaaaagatgacccaagcca

aagcgctaacctttagcagaagctaaaaagctaaatgatgctcaggcgcctaagggatctcatcaccatcaccatcact
aa-3'

Construction of pAK-RS$_{AGA}$ vectors: pAK3 and pAK7

Both pAK-RS$_{AGA}$ vectors, pAK3 and pAK7, contain two copies of aminoacyl-tRNA synthetase gene with two promoters: a modified gln S (constitutive) promoter and an araBAD (arabinose-inducible) promoter, the araC repressor gene a p15A origin of replication, and the kanamycin resistant gene marker[55].

The pAK3 (pAK-pCNF-RS$_{AGA}$) vector carries two copies of pCNF-RS$_{AGA}$ gene (Fig. 3.5). The pCNF-RS$_{AGA}$ possesses two sets of mutations. First, each pCNF-RS in pAK3 vector contains five mutations within the anticodon UCU recognition region, Tyr230Arg, Cys231Trp, Pro232Thr, Ala233Arg, and Asp286Gln[23] (Fig. 2.4). Then, the amino acid binding pocket was modified as the following mutations: Tyr32Leu, Leu65Val, Phe108Trp, Asn109Met, Asp158Gly, and Ile159Ala (Fig. 2.9)[30]. pCNF-RS$_{AGA}$ is polysubstrate specific against an array of unnnatural amino acids, including several ***UAA***s which were tested in this study, such as *p*-acetylphelalanine (AcF), *p*-azidophenaylalanine (AzF),O-methyl-Tyrosine (OMeY), and p-iodophenaylalanine (IF)(Fig. 2.10).

The DNA sequence of pCNF-RS$_{AGA}$ gene in pAK3 is 5'-atggacgaatttgaaatgataa
agcgtaacacatctgaaattatcagcgaggaagagttacgtgaggttttaaaaaaagatgaaaaatctgctctgatagg
ttttgaaccaagtggtaaaatacatttagggcattatctccaaataaaaaagatgattgatttacaaaatgctggatttg
atataattatagttttggctgatttacatgcctatttaaaccagaaaggagagttggatgagattcgtaaaataggagat
tataacaaaaaagtttttgaagcaatggggttaaaggcaaaatatgtttatggaagtgaatggatgcttgataaggatt
atacactgaatgtctatcgcttggctttaaaaactaccttaaaacgcgcacgtcgtagtatggaacttatagcacgtgag
gatgaaaatccaaaggttgctgaagttatctatccaataatgcaggttaatggtgctcattatcttggcgttgatgttgca
gttgggggatggagcagcgtaaaatacacatgttagcacgtgagcttttaccaaaaaaggttgtttgtattcacaaccc
tgtcttaacgggtttggatggagaaggaaagatgagttcttcaaaagggaattttatagctgttgatgactctccagaag
agattcgcgctaagataaagaaagcacgggggacgcgtggagttgttgaaggaaatccaataatggagatagctaaat
acttccttgaatatcctttaaccataaaacgcccagaaaaatttggtggagatttgacagttaatagctatgaggagtta
gagagtttatttaaaaataaggaattgcatccaatgcagttaaaaaatgctgtagctgaagaacttataaagattttag
agccaattcgtaagcgcttataa-3'

The amino acid sequence of pCNF-RS$_{AGA}$ gene in pAK3 is 5'-mdkkplntlisatglwm
srtgtihkikhhevsrskiyiemacgdhlvvnnsrssrtaralrhhkyrktckrcrvsdedlnkfltkanedqtsvkvkv
vsaptrtkkampksvarapkplenteaaqaqpsgskfspaipvstqesvsvpasvstsissistgatasalvkgntnpi
tsmsapvqasapaltksqtdrlevllnpkdeislnsgkpfrelesellsrrkkdlqqiyaeerenylglklereitrffvdrgfl
eikspilipleyiermgidndtelskqifrvdknfclrpmlapnlynylrkldralpdpikifeigpcyrkesdgkehleeft
mlnfcqmgsgctrenlesiitdflnhlgidfkivgdscmvfgdtldvmhgdlelssavvgpipldrewgidkpwigagf
glerllkvkhdfknikraarsesyyngistnl*

The pAK7 (pAK-NpOH-RS$_{AGA}$) vector was derived from pAK3 vector. Every synthetase in pAK7 includes two sets of mutations: one set is consist of eight mu-

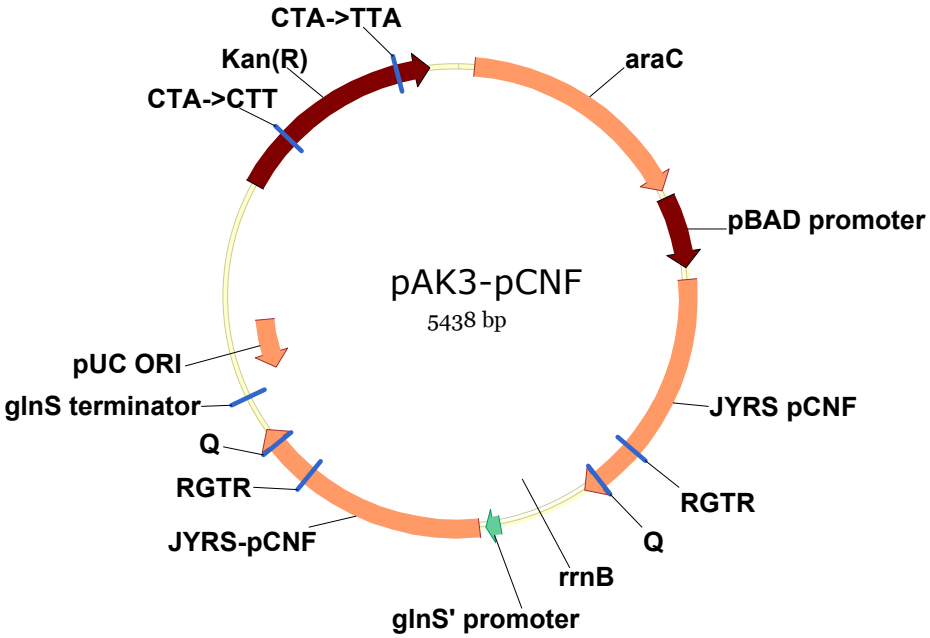


Figure 3.5: pAK3 vector.

tations (Y32E, L65T, D158S, I159A, H160P, Y161T, L162Q, A167W, compared to the wild-type *Mj*TyrRS) within the amino acid binding pocket of aminoacyl-tRNA synthetase in order to accommodate a 2-naphthol analogue of tyrosine, 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid (NpOH)(Fig. 3.6); while the other set of mutations within the anticodon recognition is the same as pCNF-RS$_{AGA}$: Tyr230Arg, Cys231Trp, Pro232Thr, Ala233Arg, and Asp286Gln.

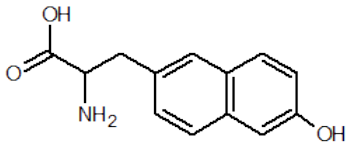


Figure 3.6: 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid (NpOH).

The amino acid sequence of 2NpOH-RS$_{AGA}$ gene in pAK7 is 5'-mdefemikrntseiis eeelrevlkkdeksaeigfepsgkihlghylqikkmidlqnagfdiiitladlhaylnqkgeldeirkigdynkkvfeamg lkakyvygsefqldkdytlnvyrlalkttlkrarrsmeliaredenpkvaeviypimqvnsaptqgvdvwvggmeqr kihmlarellpkkvvcihnpvltgldgegkmssskgnfiavddspeeirakikkaycpagvvegnpimeiakyfleyp ltikrpekfggdltvnsyeeleslfknkelhpmrlknavaeelikilepirkrl*

Overlapping PCR was used to introduce corresponding mutations for synthetases gene. Primer pairs AK1f1 and AK1r, AK2f and AK2r1, AK3f and AK3r, AK4f and

AK4r were used for over lapping PCR with pBK-NpOH and pAK3 as templates to generate four small pieces; then these small pieces were overlapped to produce long DNA fragments by using slightly different primers as follows. Primer pair AK1f2 & AK4r bring NdeI and PstI cutting site, while AK1f3 & AK4r3 have BglII and SalI sites. In pAK7 vectors, the synthetase gene after pBAD promoter carries a restriction site (BglII or SacI) at each end, whereas one restriction enzyme (PstI or NedI) cutting site was positioned at each end of the synthetase gene after glnS' promoter. These restriction enzyme cutting sites were utilized for the construction of pAK7 vector from pAK3.

Expression and purification of Z domain protein

Z domain vector (pAGA3-Zaga or pAGA-Zaga) and pAK-RS_{AGA} (pAK3 or pAK7) were co-transformed into different strains of *E.coli* cells, including GM10, Top10(F'), DH10B,and BL21(DE3). A single colony from each co-transformation was grown overnight at 37 °C in 2 ml of LB media supplemented with kanamycin and chloramphenicol to obtain the starter culture. The starter culture was then twenty times diluted to a 20 mL LB medium containing 50 μg/ml kanamycin, 34μg/mL chloramphenicol and 1.5mM or 3mM unnatural amino acids (IF, AcF, AzF, OMeY, NO2F, or NpOH) for further culturing. Till the culture reach OD600=0.6-0.8, 0.02 percent arabinose was used for induction, and followed by 1mM IPTG induction 10 minutes later. For proteins expressed under 37°C, the incubator was kept as normal 37 °C; while for proteins expressed under 42 °C, the temperature of the incubator was raised up to 4242 °C when arabinose was added. After expression for 14-16 h at 37°C, 20 ml cultures were pelleted at 8,000 rpm and frozen at -80°C. In order to purify Z domain proteins with a 6His at the C-terminus under native condition, pellets were thawed and lysed for 20 minutes with 2 ml BugBuster (Novagen) supplemented with 0.5 mg/ml lysozyme (Sigma), and 2.5 U/ml bezonase (Novagen). Cell lysates were centrifuged at 7,000 rpm, and the supernatant were collected to with 0.5 ml Ni-NTA agarose (Qiagen). The mixture was gently shaded on a rotary shaker at 4°C for one hour and loaded into a polypropylene column (Qiagen). Columns were washed twice with 2 ml wash buffer (50 mM NaH2PO4 , 300 mM NaCl, 20 mM imidazole, pH 8.0) and eluted twice with 400 μl of native elution buffer which is the wash buffer plus 230 mM imidazole.

Calculating protein yields

Once the buffer to dissolve proteins was changed from native elution buffer, OD_{280} was then measured to decide the protein concentration. Based on Vector NTI protein data analysis result, wild type Z domain protein is 69aa long, and molecular weight is 7876 Da. So, 1 unit of OD/tsb280 value of Z domain protein corresponds to about

6.15mg/ml protein. The bacteria culture usually was 25ml, and total eluted protein was dissolved in 0.8 ml distilled water by using 2μl protein solution. So the equation used to calculate Z domain protein yield is: $(OD_{280}) \times 6.15 \times 0.800 \div 25 \times 1000$. The unit of protein concentration is mg/L.

Protein electrophoresis

Eluted samples were analyzed on a 4-12% Tris-Gly SDS-PAGE gel (Invitrogen) and stained with Coomassie brilliant blue. Each 10 μl protein sample in native elution buffer was mixed with 10 μl 2X SDS sample loading buffer and then heated at 70°C for 10 minutes. Every 20 μl heat denatured protein sample was loaded on NuPAGE Novex Bis-Tris mini gel and run at 200 V constant for 35 minutes within 1X MES Buffer. After electrophoresis, the gel was placed in 100 ml of ultra-pure water in a loosely covered container and microwaved for 1 minute until the solution almost boils. Then the gel was shaken on a rocker for 1 minute and the water was discarded. The water washing steps were repeated for two more times. 20 ml of SimplyBlue SafeStain(Coomassie brilliant blue) was added into the container and it was microwaved on High for 45 seconds until the solution almost boils. The gel in boiled Coomassie blue) was shaked 5 minutes, and then the dye was discarded. Finally the gel was washed the gel in 100 ml of ultra pure water until the background was destained to almost clear on the rocker.

Analyzing protein molecule weight by mass spectrometry

Protein molecular weights were assessed by mass spectrometry to analyze the unnatural amino acids incorporation rates. Protein samples which showed obvious bands with correct sizes on SDS-PAGE gel were sent out for mass spectrometry (ESI or LC-MS/MS). To prepare protein samples for ESI, proteins in native elution buffer were treated by Aspire chromatography tips to eliminate salts and the buffer was changed to 70% acetonitrile/ddH2O. For LC-MS/MS sample, about 4 μg Z domain protein was cut from SDS-PAGE gel after electrophoresis, and then sent out for LC-MS/MS in 1% acetic acid/ddH2O.

Labeling of unnatural amino acids

Z-7AzF purified from GM10 was labeled by a primary aliphatic amine-reactive probe, 5-dimethylaminonaphthalene-1-sulfonyl hydrazine (dansyl hydrazine, DNSH), which can react with the keto group of AcF (Fig. 2.13). The DNSH labeling reaction was carried by the following steps: mixed targets protein AcF-Z 30μL with 6.5% TFA/ACN 150μL, then added 100μL 5mM DNSH into the mixture. The mixture was shaken gently over night at room temperature (about 24°C)[58].

Toxicity test of *UAA*-RS$_{AGA}$/tRNA$^{opt}_{UCU}$

pAK-pCNF-RS$_{AGA}$ (pAK3) and pAGA3-Zaga which contains one copy of tRNA$^{opt}_{UCU}$ were co-transformed into GM10. One single colony was grown overnight to saturation in LB media supplemented with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol. Culture were 1:100 diluted and divided into 10 samples. Each sample was treated separately and then aliquoted into 96-well assay plates in octuplicate. The cell growth rates of pAK3+pAGA3-Zaga/GM10 were measured with different concentrations of arabinose: 0%, 0.002%, 0.006%, 0.01%, 0.02%, 0.06%, 0.1%, 0.2%, 0.6% and 1%, with or without 1.5mM IF. Plates were grown in a SpectraMax M2e plate reader (Molecular Devices) and read at 600 nm every 200 seconds with mixing of 120 seconds/read at 37°C for 16 hours. Relative bacteria densities are indicated for 200 μl cultures and are an average of eight independent cultures.

3.4 Results

Recoding AGA from arginine to tyrosine

After screening with pAGA-GFP probe plasmid and the pBK4-JYRS(AGA) library, clones were selected and their identities were sequenced (Table 3.3). One clone which gave the brightest fluorescence, pAK1, was chosen for further experiments. DNA sequencing results indicated that it carried five mutation sites, Tyr230Arg, Cys231Gly, Pro232Thr, Ala233Arg, and Asp286Gln.

Recoding AGA from arginine to unnatural amino acids

By using pAK3 and pAK7, several ***UAA***s (AzF, AcF, IF, and NpOH) were incorporated into Z domain protein by ***UAA***-RS$_{AGA}$/tRNA$^{opt}_{UCU}$ (Table 3.1). The incorporation of each of the ***UAA***s was confirmed by ESI and LC-MS/MS. Further, the relative ***UAA*** incorporation rate was assessed by the equation of ***UAA***s decorated proteins (Z-R7X)/total proteins (Z-R7X plus Z-R7).

Determining the incorporation of unnatural amino acids

Because the AGA is not a terminating codon, the ***UAA*** incorporated Z domain protein would not show any difference from arginine incorporated ones. Thus the SDS-PAGE can not lend us a powerful hand to detect the suppression of AGA codon because all expressed Z domain proteins were a band close to 8 kDa on SDS-PAGE gels. In stead, mass spectrometry was employed for this purpose. The molecular weights of arginine incorporated Z domain protein at position 7 are 7921.6 Da (full length protein), 7832.4 Da (full length protein but without the first methionine and

Table 3.1: Sequencing results of mutation sites of 7 positive strains from pAGA-GFP and pBK4-JYRS library screening.

WT Codon / Amino Acid	763 TAC 230Tyr	766 TGC 231 Cys	769 CCA 232 Pro	772 GCT 233 Ala	931GAT 286Asp
AK1	CGG Arg	GGG Gly	ACG Thr	CGT Arg	CAG Gln
AK2	AGA Arg	TGG Trp	ACT Thr	AGA Arg	GGG Gly
AK3	AGA Arg	TGG Trp	ACT Thr	AGA Arg	TTT Phe
AK4	GGT Gly	AGG Arg	AGT Ser	AGG Arg	ATT Ile
AK5	AGG Arg	GGG Gly	AAG Lys	GTG Val	AAT Asn
AK7	AGG Arg	GTG Val	GGT Gly	ATG Met	CTT Leu
AK8	AGA Arg	TGG Trp	ACT Thr	AGA Arg	ACG Thr

with post translational modification acetylation, -Met). But, if the 7th arginine was replaced by designated unnatural amino acid, including IF, AcF, AzF and OMeY, the molecular weights will be changed correspondingly (Table 2.4).

IF-Z protein was listed here as an example. The SDS-PAGE result of the purified IF-Z protein (expressing in the presence of AcPhe) shows one band near 8 KDa. For Z domain proteins purified from GM10 bacteria, the deconvoluted **ESI-MS** spectrum reveals two major peaks at m/z =7906, and 7949 (YW44-IF), which match the expected molecular weight for the IF incorporated Z-domain mutant with the loss of its first methionine (IF-Z-Met, m/z =7907.3) and its subsequent post-translational acetylation product (IF-Z-Met+Ac, m/z =7949.3). There is also a minor peak at m/z =7791, matching the arginine incorporated Z-domain protein with the loss of its first methionine (Arg-Z-Met). Similarly, for other three types of unnatural amino acids decorated Z domain proteins (AcF-Z, OMeY-Z, and AzF-Z), each one of these deconvoluted ESI-MS spectra had two major peaks matching the corresponding ***UAA***-Z-Met and ***UAA***-Z-Met+Ac, and a minor peak around m/z =7790 matching Arg-Z-Met (YW28-AcF, YW32-OMeY, YW4-AzF) (Fig 3.7).

LC-MS/MS results *directly confirmed the incorporation of IF into Z domain protein in response to AGA codon at the 7th position.* After LC-MS/MS, the trypsin digested the protein into 8 unique peptides and the spectra covered 68 amino acids, which was the 100% coverage of the whole length Z domain protein without the first

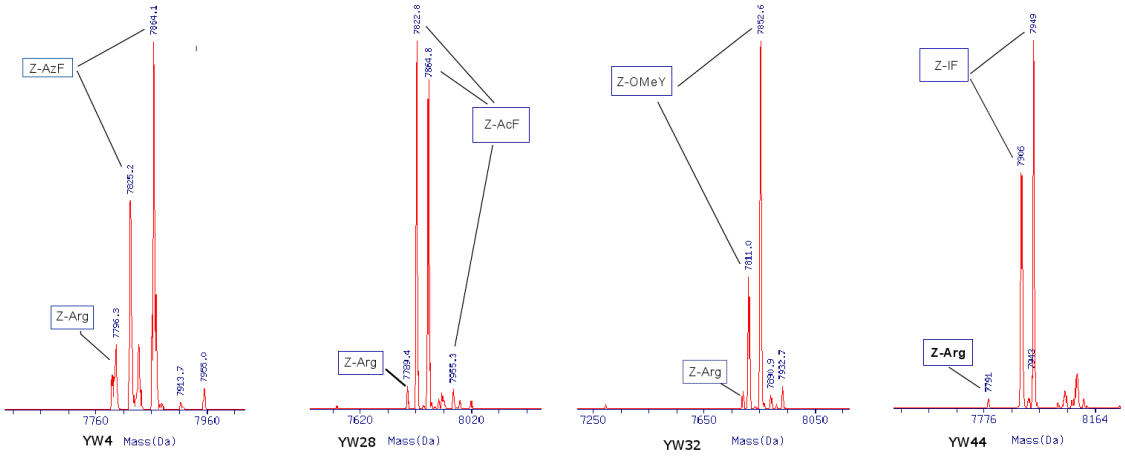


Figure 3.7: ESI results: four ***UAA***s (AzF-YW4, AcF-YW28, OMeY-YW32, IF-YW44) were incorporated into Z domain proteins at position 7.

methionine. The amino acid sequence of the first peptide fragment after trypsin digestion was (M)TSVDN**F***INKE, and **F*** was encoded by the seventh AGA and denoted *p*-iodophenaylalanine (IF). Out of total 24 spectra for the first fragment, 21 spectra demonstrated an unique molecular weight for the seventh amino acid **F***: 165(molecular weight of phenylalanine) + 127 = 292, which matches well with the molecular weight of IF 291. This was a direct evidence of AGA reassigning from arginine to IF. In the spectrum showed below (Fig 3.8) and also in the fragmentation table (Table 3.2), **B** meant counting the amino acids from N terminal while **Y** stood for counting from the C terminal, and the number 1 to 9 indicated the position of these amino acids. A closer examination revealed that the majority sequences (**B2-B8**, and **Y2-Y8**) had both matching molecular weights and peaks on the fragmentation table(Table 3.2) and the spectrum figure (Fig 3.8) .

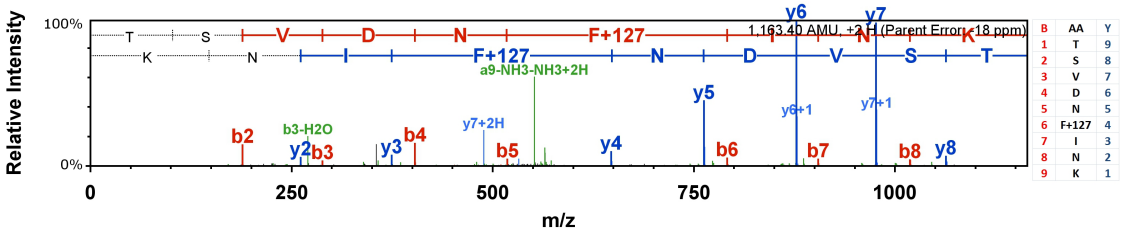


Figure 3.8: LC-MS/MS result indicates that IF was incorporated into Z domain protein at the seventh position.

Table 3.2: The LC-MS/MS fragmentation table of Z-7IF.

B Ions	B+2H	B-NH3	B-H2O	B	AA	Y	Y Ions	Y+2H	Y-NH3	Y-H2O
Missing,102.05			Missing,84.04	1	T	9	Missing,1,164.43	Missing,582.72	Missing,1,147.40	Missing,1,146.42
Found,189.09			Found,171.08	2	S	8	Found,1,063.38	Found,532.20	Found,1,046.36	Found,1,045.37
Found,288.16			Found,270.14	3	V	7	Found,976.35	Found,488.68	Found,959.32	Found,958.34
Found,403.18			Found,385.17	4	D	6	Found,877.28	Found,439.14	Found,860.26	Found,859.27
Found,517.23		Found,500.20	Found,499.21	5	N	5	Found,762.26		Found,745.23	
Found,791.20	Found,396.10	Missing,774.17	Found,773.19	6	F+127	4	Found,648.21		Missing,631.19	
Found,904.28	Found,452.64	Missing,887.26	Found,886.27	7	I	3	Found,374.24		Found,357.21	
Found,1,018.33	Found,509.67	Missing,1,001.30	Found,1,000.31	8	N	2	Found,261.16		Found,244.13	
Missing,1,164.43	Missing,582.72	Missing,1,147.40	Missing,1,146.42	9	K	1	Missing,147.11		Missing,130.09	

Using pAK3 vector to recode AGA to unnatural amino acids in the mutant Z domain protein (Z-7AGA)

The AGA suppression by various ***UAA***s, such as IF, AzF, AcF, and OMeY, was tested among five strains of *E.coli*, including GM10, GM241, DH10B, Top10, and BL21(DE3) (Table 3.3). The results indicate that in GM10, pAK3 is very effective to charge these four ***UAA***s onto $tRNA^{opt}_{AGA}$ and further recode AGA from arginine to ***UAA***s. The ***UAA*** incorporation rates, depending on ***UAA*** types, can reach over 95% for certain ***UAA***s.

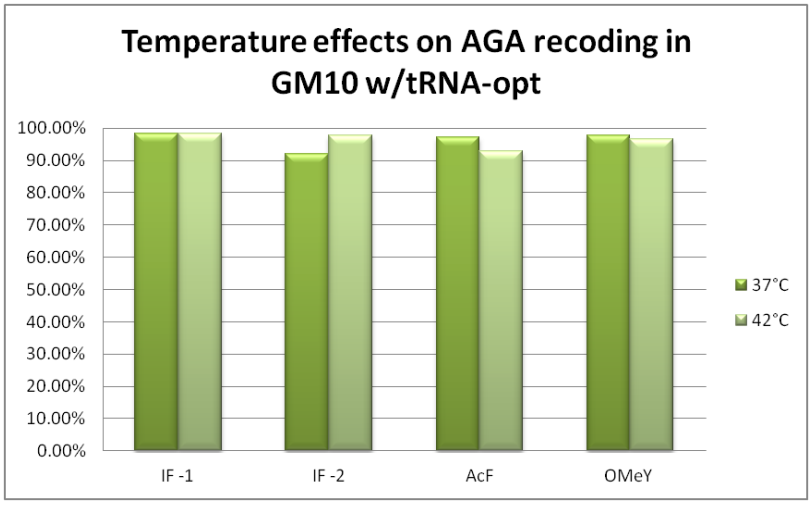


Figure 3.9: Various ***UAA***s were incorporated into Z domain proteins in respond to AGA. The y axis is the unnatural amino acid incorporation rate.

37°C *v.s.* 42°C. Because of the special trait of GM10 bacteria, we tested the

UAA's incorporation rates in response to AGA codon under two temperatures, 37°C and 42°C. IF, AcF, and OMeY were incorporated into Z domain proteins under both temperature with similar condition (pAK3+pAGA3-Zaga) within GM10 cells, and the IF incorporation tests were repeated once. We found that there were no very significant differences between different ***UAA*** incorporation rates of two temperatures among four sets (Fig 3.9). All of them varied within the range of 90% to 100%. However, although the data only show slight difference between 37°C and 42°C, more experiments at varied temperature (20°C, 25°C, 30°C) would be needed to determine the temperature effect.

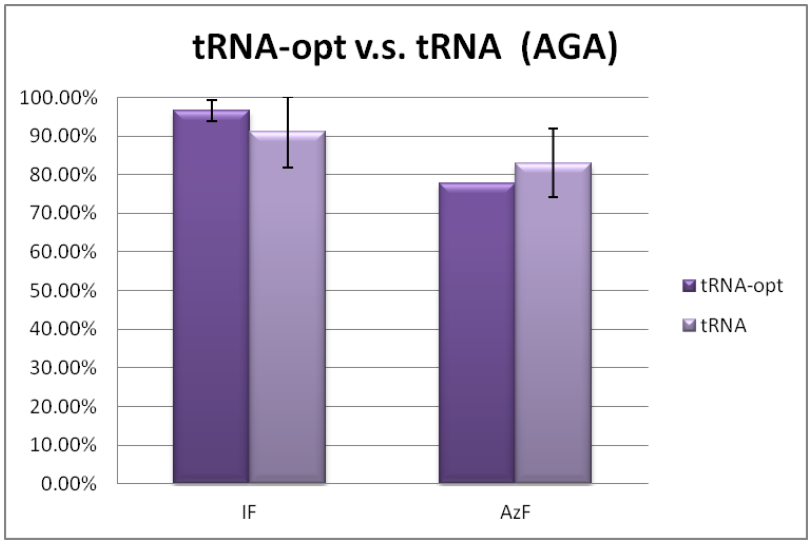


Figure 3.10: Optimized $tRNA^{opt}_{AGA}$ improved IF incorporation rates, but not for AzF. The y axis is the unnatural amino acid incorporation rate.

Optimized $tRNA^{opt}_{AGA}$ improved IF incorporation rates. While using the same pAK3 vectors, the optimized $tRNA^{opt}_{AGA}$ (carried by pAGA3-Zagg) and the normal $tRNA_{AGA}$(carried by pAGA-Zaga) were compared for their abilities to charge unnatural amino acids in GM10 cells. There were several repeats for these sets, so the average and the standard derivation were calculated (Fig 3.10). From the results we did not find dramatic differences between using $tRNA^{opt}_{AGA}$ and $tRNA_{AGA}$. The IF incorporation of $tRNA^{opt}_{AGA}$ was about 96% while that of $tRNA_{AGA}$ was around 91%. Interestingly, by using $tRNA^{opt}_{AGA}$ the AzF incorporation rate was only about 77%, but three repeats of employing normal $tRNA_{AGA}$ gave an average of 83% AzF incorporation rates. So the optimized $tRNA^{opt}_{AGA}$ might increase the ***UAA*** incorporation rates to a limited extent, and this improvement depends on ***UAA*** types.

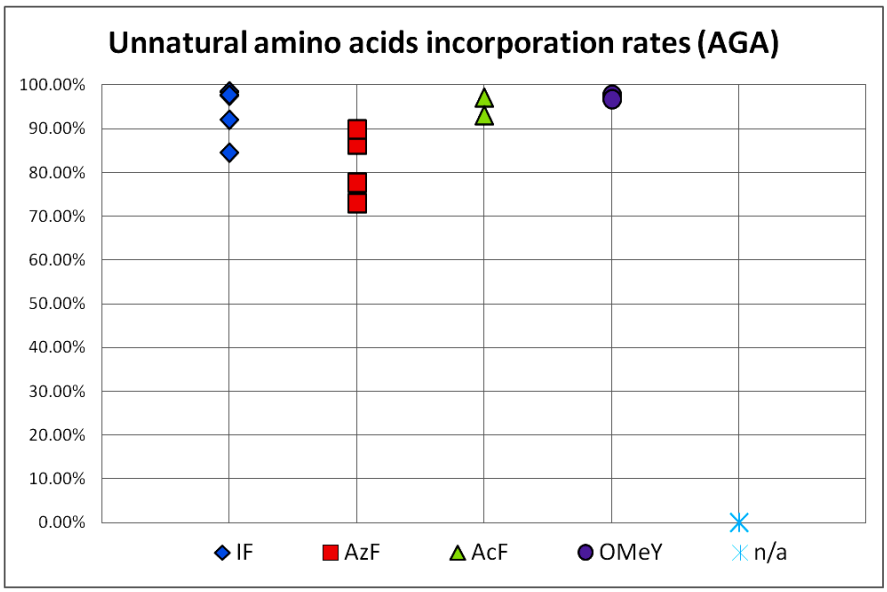


Figure 3.11: Various ***UAA***s were incorporated into Z domain proteins in respond to AGA.The y axis is the unnatural amino acid incorporation rate.

Varied *UAA*s were incorporated into Z domain proteins in response to AGA. Different types of unnatural amino acids, such as IF, AzF, AcF, and OMeY, were incorporated into the mutant Z domain proteins, by using pAK3+pAGA3-Zaga or pAK3+pAGA-Zaga co-translated *E. coli*. When considering the incorporation rates of different kinds of unnatural amino acids, all protein expressed with pAK3 vectors were considered, including both 37°C and 42°C; both $tRNA^{opt}_{AGA}$ and $tRNA_{AGA}$; since no obvious differences were observed as stated above. We found out that the IF, AcF, and OMeY were incorporated into Z domain proteins in response to AGA codon by the dual synthetase vector in GM10 bacteria, with normally over 90% incorporation rates(Fig 3.11) . The repeating results indicated that the incorporation rates of AzF were ranging from 70%-80%.

Strain matters. In order to evaluate the unnatural amino acids incorporation efficiency of pAK3-pCNF-RS_{AGA} among various *E.coli* strains, we tested its ability to insert p-IF into Z-R7X mutant proteins with the presence of $tRNA^{opt}_{AGA}$ in five bacteria strains, including GM1O, GM241, Top10, BL21DE3 and DH10B. GM241 has similar genomic mutations as GM10, except the argU10(ts) mutation which leads to the mutant tRNA. By using the same expression protocol, p-IF decorated Z-R7X protein yields varied among these five strains (Fig 3.12): first GM241 showed a similar IF incorporation as GM10, 95%; then BL21DE3 has a similar high incorporation rate (approximately >80%) as GM10; DH10B ranked third by 60%, while Top10 cells were

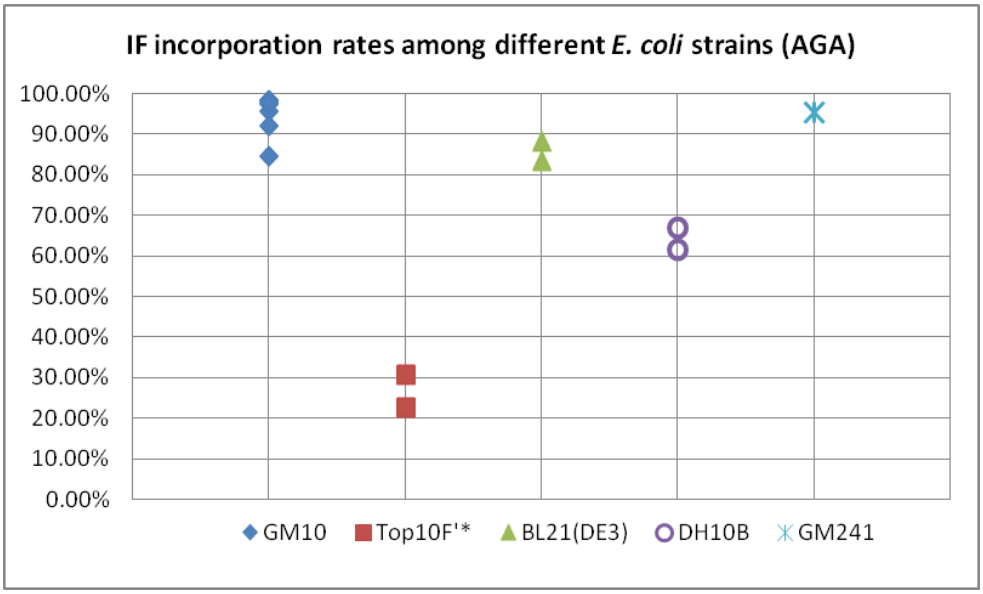


Figure 3.12: IF were incorporated into Z domain proteins in respond to AGA in different ***E. coli*** strains. The y axis is the unnnatural amino acid incorporation rate.

proven not very suitable as an unnatural amino acids incorporating host cells with a low rate of 30%. Results support that GM10 is the best candidate among these five strains to recode AGA to unnatural amino acids with outstanding IF incorporation rates above 90%. Meanwhile Top10 was not proved to be a suitable host by giving poor IF incorporation rates between 20% (YW11) and 30% (YW14).

Using pAK7 vector to recode AGA to NpOH in the mutant Z domain protein (Z-7AGA)

Three expression tests were performed by using pAK7+pAGA3-Zaga/GM10 bacteria with 1.5mM NpOH added. Two of the expressions (YW39 and YW46) were done at 37°C and the induction/expression temperature of one (YW40) expression was raised to 42°C.

In one of the 37°C expression results, the deconvoluted ESI-MS spectrum shows peaks at m/z = 7790, 7847, 7889 and 7978 (YW46) (Fig 3.13), which match the expected molecular weight for the full length arg-Z without the first methionine (m/z=7790.4), the full length NpOH incorporated Z-domain mutant (m/z = 7978.6) and that for this mutant protein with the loss of its first methionine (m/z = 7847.4) and its subsequent post-translational acetylation product (m/z = 7889.5). The NpOH incorporation rate in this expression was about 90%.

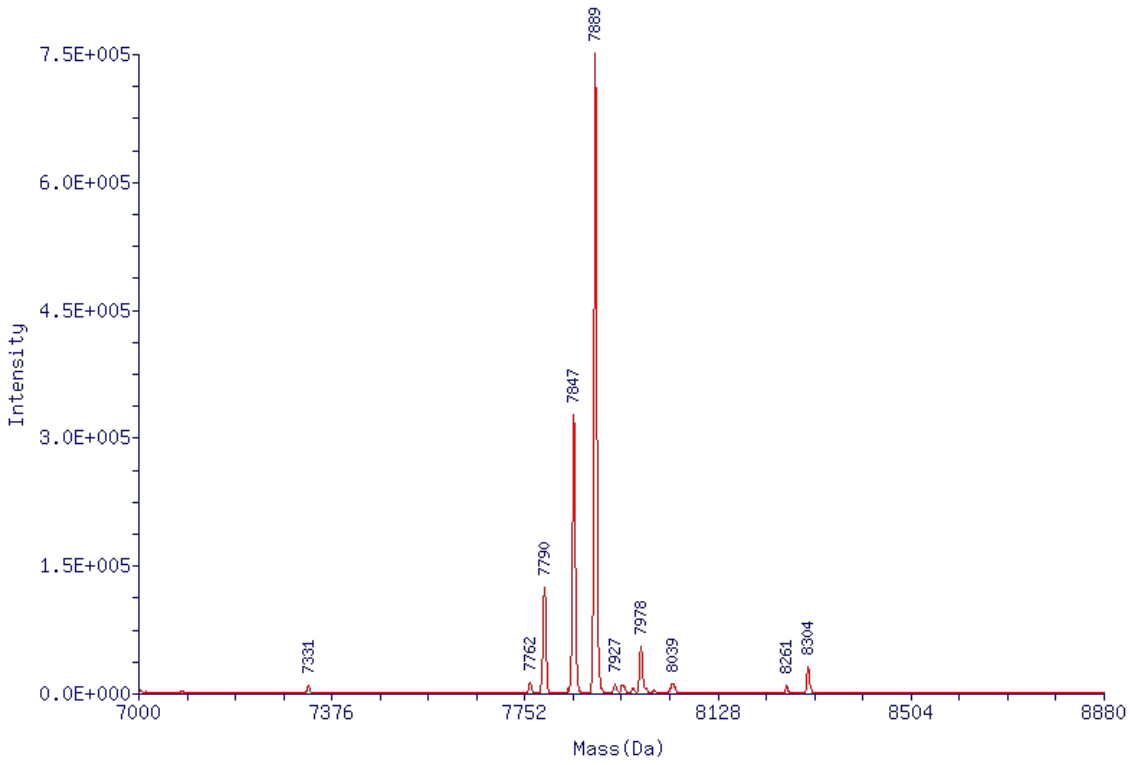


Figure 3.13: ESI result: NpOH was incorporated into Z domain proteins in respond to AGA (YW46).

The NpOH incorporation rate of another 37°C expression was close to 80%, but that of 42°C expression was merely above 60%.

Controls for unnatural amino acids incorporation

pET-Zx (carries Z-7TAG mutant gene)and pEVOL-TAG, a dual expression vector carrying two copies of MJ pCNF-RS_{TAG} and $tRNA_{GUA}$, were contransfered into BL21(DE3). AzF (1.5mM) was added into culture medium during protein expression. The ESI result (7/17/2011-5) showed two major peaks (m/z=7821.9 and m/z=7863.7) matching AzF-Z-Met and AzF-Z-Met+Ac. This test served as a positive control to test our protein expression procedure.

A serial of negative controls are now described as follows. First, without the unnatural amino acids(for example IF) added, pAGA-Zaga and pAK3 transformed GM10 bacterial could not produce any IF bearing Z domain proteins. From the

spectrum (7/17/2011-4), only peaks matching the Arg-Z (*e.g.*, m/z=7789.9) were found. Similarly, if the synthetase vector pAK3 was not added into the expression system, no IF decorated Z domain was found and only Arg-Z was seen(YW80).

Second, 1.5mM IF was presented to translate Z-7AGA gene into Z domain proteins, while different components of the translational machinery was absent: no synthetase(pAK3) and $tRNA_{AGA}{}^{opt}$ (YW84) at all (YW83), with $tRNA_{UCU}{}^{opt}$ only(YW84), with synthetase (pAK3) only (YW85). All these three negative controls returned the expected results – only peaks matching Arg-Z protein were found and no IF-Z was observed. Thus, a conclusion that either synthetase or tRNA is necessary for AGA recoding.

The specificity of the tRNA was also tested. GM10 bacteria transformed with pAK3 and pAGA3-Zagg, which carries Z-7AGG (*not Z-7AGA!*) mutant gene and $tRNA_{UCU}{}^{opt}$, was induced to express Z domain protein with the presence of 1.5mM IF (YW20) or AzF(YW19). In either case, no unnatural amino acids decorated Z domain protein were observed and only Arg-Z was generated. These interesting results indicates that a perfect match between codon AGA on mRNA and the anticodon UCU is important for AGA reassignment.

Protein yields

The OD_{280} values of mutant Z domain proteins in water were usually ranging from 0.1 to 0.2, depending on bacteria strains and ***UAA*** types. So, by using the equation $(OD/tsb280) \times 6.15 \times 0.800 \div 25 \times 1000$, the results were about 19 mg/L to 38 mg/L. The yields of ***UAA*** decorated Z domain proteins were about 20-30 mg/L under optimized conditions in GM10.

DHSH labeling of AzF-Z

To provide further chemical evidence that the ***UAA***s have been successfully incorporated, AcF-Z purified from GM10 was labeled by a primary aliphatic amine-reactive probe, 5-dimethylaminonaphthalene-1-sulfonyl hydrazine (dansyl hydrazine, DNSH), which can selectively react with the keto group of AcF. The DNSH conjugated AcF-Z displayed green-yellowish fluorescence under UV excitation while the Arg-Z which underwent the same reaction condition did not show any fluorescence, indicating successful incorporation of AcF into Z domain protein (Fig 3.14).

Bacterial growth toxicity test

Furthermore, to determine the toxicity of pCNF-RS_{AGA}/$tRNA_{UCU}{}^{opt}$ dual expression system, the cell growth rates of pAK3+pAGA3-Zaga/GM10 were measured with different concentrations of arabinose: 0%, 0.002%, 0.006%, 0.01%, 0.02%, 0.06%,

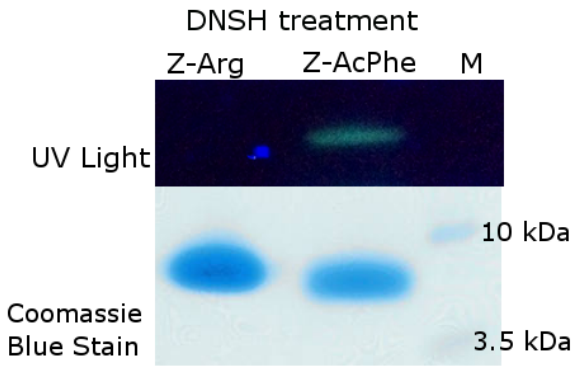


Figure 3.14: Dansyl hydrazine (DNSH) conjugation of AcF. Right: SDS-PAGE gel analysis of DNSH treated Z-AcF. Top panel: fluorescent imaging of the gel under UV light (365nm); bottom panel: proteins stained with Coomassie blue of the same gel. The images were captured by a regular camera and shows real colors. .

0.1%, 0.2%, 0.6% and 1%, with or without 1.5mM IF. Without IF added into culture media, the all tested bacteria showed a similar proliferating slope during the lag phase and the exponential growth phase, and in the stationary phase they reached slightly different cell densities(Fig 3.16). These minor difference might come from the accumulation of over expressed synthetases and toxic tRNAs inside bacterial cells. However, after 1.5mM IF added into culture media, some obvious differences were observed. From 0% - 0.1%, the bacterial growth curves displayed a trend that the more arabinose we added, the slower the growth rate was and the less cell density they reached (Fig 3.15). If the working concentration of arabinose was equal to, even more than 0.1%, the bacterial growth was dramatically inhibited, and the bacterial density did not increase a lot and the curve was quite flat. So these results reminded us that the working concentration of arabinose might play an important role in bacterial growth and the toxicity caused by the newly introduced AGA reassigning machinary. Induction with arabinose 0.02%, which is the working concentration during mutant protein expression, yielded a moderate level of growth rate of GM10 cells and a tripled final cell density at least. Also, it is not recommended to add more than 0.02% arabinose to synthesize more pCNF-RS_{AGA} because the growth rate would decrease dramatically.

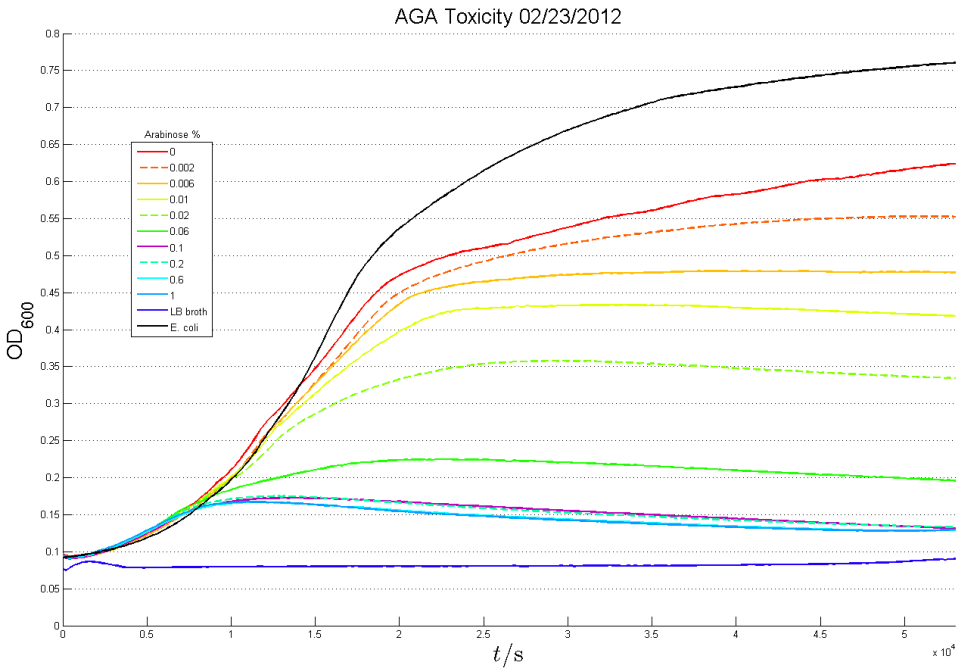


Figure 3.15: The toxicity caused by AGA global suppression during pAK3+pAGA3-Zaga/GM10 cell growth, with 1.5mM IF. The cell growth rates of pAK3+pAGA3-Zaga/GM10 were measured with different concentrations of arabinose: 0%, 0.002%, 0.006%, 0.01%, 0.02%, 0.06%, 0.1%, 0.2%, 0.6% and 1%, with 1.5mM IF. LB culture media without any bacteria, and GM10 E. coli without adding anything(black) were both used as controls.

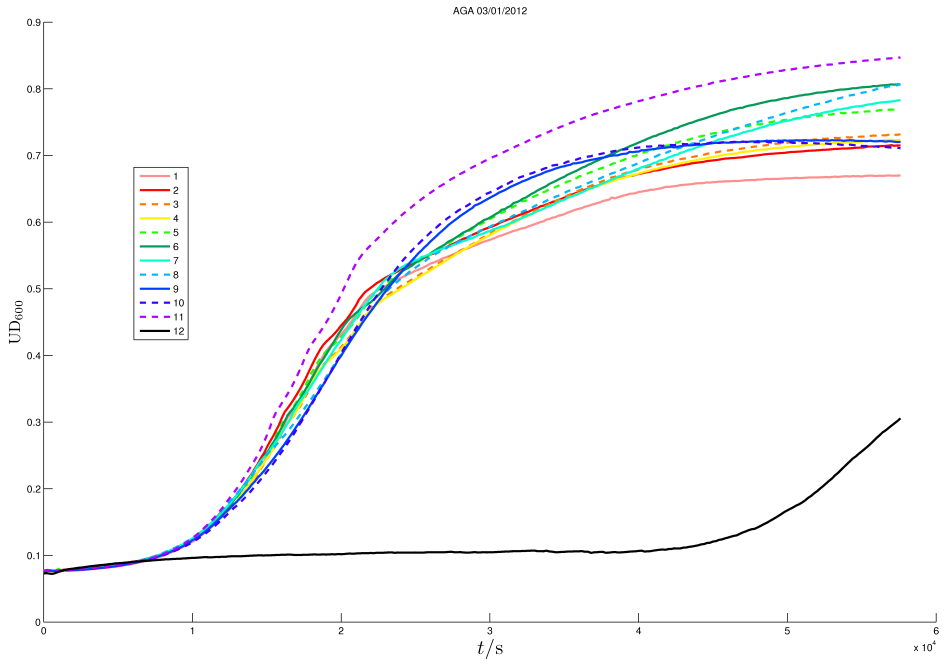


Figure 3.16: Control: the cell growth curve of pAK3+pAGA3-Zaga/GM10 cell. Only varied concentrations of arabinose were added at the beginning of cell growth, but no ***UAA*** was added.

Table 3.3: Varied ***UAA***s were incorporated into Z domain proteins in response to AGA codon.

YW#	UAA	Concerntration	UAA-Z%	Arabinose%	Temprature	E. coli	probe vector	synthetase
44	IF	1.5mM	98.40%	0.020%	37°C	GM10	pAGA3-Zaga	pAK3
37	IF	1.5mM	91.93%	0.020%	37°C	GM10	pAGA3-Zaga	pAK3
45	IF	1.5mM	98.46%	0.020%	42°C	GM10	pAGA3-Zaga	pAK3
18	IF	1.5mM	97.64%	0.020%	42°C	GM10	pAGA3-Zaga	pAK3
64	IF	1.5mM	95.30%	0.020%	37°C	GM241	pAGA3-Zaga	pAK3
28	KT	1.5mM	97.03%	0.020%	37°C	GM10	pAGA3-Zaga	pAK3
33	KT	1.5mM	92.96%	0.020%	42°C	GM10	pAGA3-Zaga	pAK3
27	OMeY	1.5mM	97.65%	0.020%	37°C	GM10	pAGA3-Zaga	pAK3
32	OMeY	1.5mM	96.68%	0.020%	42°C	GM10	pAGA3-Zaga	pAK3
17	AzF	1.5mM	77.65%	0.020%	42°C	GM10	pAGA3-Zaga	pAK3
29	NO2F	1.5mM	47.58%	0.020%	37°C	GM10	pAGA3-Zaga	pAK3
34	NO2F	1.5mM	10.00%	0.020%	42°C	GM10	pAGA3-Zaga	pAK3
5	IF	1.5mM	84.57%	0.020%	37°C	GM10	pAGA-Z(aga)	pAK3
7/17/2011 (3)	IF	1.5mM	97.47%	0.020%	42°C	GM10	pAGA-Z(aga)	pAK3
7/17/2011 (1)	IF	1.5mM	71.08%	0.020%	42°C	GM10	pAGA-Z(aga)	pAK2
3	AzF	1.5mM	86.22%	0.020%	37°C	GM10	pAGA-Z(aga)	pAK3
6	AzF	1.5mM	73.00%	0.020%	37°C	GM10	pAGA-Z(aga)	pAK3
4	AzF	1.5mM	89.77%	0.020%	42°C	GM10	pAGA-Z(aga)	pAK3
30	IF	1.5mM	83.21%	0.020%	37°C	BL21(DE3)	pAGA3-Zaga	pAK3
35	IF	1.5mM	88.21%	0.020%	42°C	BL21(DE3)	pAGA3-Zaga	pAK3
31	IF	1.5mM	67.03%	0.020%	37°C	DH10B	pAGA3-Zaga	pAK3
36	IF	1.5mM	61.62%	0.020%	42°C	DH10B	pAGA3-Zaga	pAK3
11	IF	1.5mM	22.68%	0.020%	37°C	Top10F'	pAGA-Z(aga)	pAK3
14	IF	1.5mM	30.55%	0.020%	42°C	Top10F'	pAGA-Z(aga)	pAK3
7/17/2011	AzF	1.5mM	100.00%	0.020%	37°C	BL21(DE3)	pET-Zx	pEVOL-TAG
7/17/2011 (4)	No UAA	0mM	0.00%	0.020%	42°C	GM10	pAGA-Z(aga)	pAK3
7/17/2011 (2)	No UAA	0mM	0.00%	0.020%	42°C	GM10	pAGA-Z(aga)	pAK2
83	IF	1.5mM	0.00%	0.020%	37°C	GM10	pZaga (no $tRNA_{aga}$)	n/a
84	IF	1.5mM	0.00%	0.020%	37°C	GM10	pAGA3-Zaga	n/a
85	IF	1.5mM	0.00%	0.020%	37°C	GM10	pZaga (no $tRNA_{aga}$)	pAK3
80	No UAA	0mM	0.00%	0.020%	37°C	GM10	pAGA3-Zaga	n/a
20	IF	1.5mM	0.00%	0.020%	42°C	GM10	pAGA3-Zagg	pAK3
19	AzF	1.5mM	0.00%	0.020%	42°C	GM10	pAGA3-Zagg	pAK3

3.5 Discussion

The two-step selection: an effective strategy for sense codon recoding

The recoding process of AGA or AGG demonstrates that the two-step selection was an effective strategy for sense codon reassginment. In the first step, the target rare codon, such as AGA or AGG, was reassigned from arginine to tyrosine. This completion of this step was based on the correct evolution of the anticodon recognition loop of *Mj*TyrRS to recognize the new anticodon instead of the original codon GUA. Actually a synthetase library of random-mutated four residues which direct contact with the tRNA anticodon was established to cover all possible mutations, and the anticodon of *Mj* Tyr-$tRNA_{GUA}$ was altered to UCU or UCC. Meanwhile a mutant Tyr66AGA/G-GFP_{UV} reporter was used to select out the synthetases which charge tyrosine in response to the target sense codon. As we described in Chapter 2, Tyr66 is one of the essential residue of the GFP fluorophore; and in the mutant GFP_{UV} gene, the codon for Tyr66 was mutated to the target sense codon, AGA or AGG. This mutant GFP_{UV} would not emit green fluorescence at 509nm until the evolved *Mj*TyrRS sythetase was able to charge tyrosine to the mutated *Mj*$tRNA_{UCU/CCU}$. Generally speaking, the intensity of the green fluorescence emitted by the clone reflects the ability of the new synthetase to recognize the changed anticodon. Then the second step was to reassign AGA to unnatural amino acids by editing the amino acids binding pocket of the aaRS to allow the fitting of desired ***UAA***s. By this two-step selection strategy, we can find out aaRS which can charge desired ***UAA***s onto *Mj* $tRNA_{UCU/CCU}$s. However there're still more works to improve the compatibility of the evolved aaRS/tRNAs to battle against natural-occurred aaRS/tRNAs.

The optimal AGA recoding condition

The recoding of AGA codon was more straight-forward after we learned numerous lessons from AGG reassignment. We used the pAK dual vectors to enhance ***UAA*** incorporation, and GM10 *E. coli* strain to reduce the incorporation of arginine in response of AGA.

The dual vector, pAK3/pAK7, plays an important role to recode AGA from arginine to ***UAA***s. pAK provides two copies of ***UAA***-RS gene, one after a constitutive promoter (gln S) and the other one after an arabinose-inducible promoter (araBAD). araBAD is a strong inducible promoter which boosts the cytoplasmic ***UAA***-RS level quickly after adding arabinose. The abundance of ***UAA***-RS further dramatically increases the ligation products of the synthetase, ***UAA***-charged-$tRNA_{UCU}$. Then the increased amount of $tRNA_{UCU}{}^{UAA}$s make it possible for themselves to win the competition with endogenous $tRNA_{UCU}{}^{Arg}$s during translation in response to AGA codon.

The importance of pAK dual vectors was supported by its high ***UAA*** incorporation rates (over 90%) among various ***UAA***s and bacterial strains. This is also consistent with the result that the ***UAA*** incorporation rates were improved more than 10 times when the synthetase vector was changed from pBK vectors to pAK vectors, because glnS' is presented on both pBK vectors and pAK vectors, while only pAK vectors carry araBAD.

GM10 strain is another essential factor to eliminate, to the maximal extent, the incorporation of arginine in response to AGA. GM10 bacteria have a reduced intracellular level of endogenous $tRNAs_{UCU}{}^{Arg}$ which is caused by an argU10(Ts) mutation in its argU gene[67, 68]. In fact in *E. coli* the products of argU gene,$tRNA_4{}^{Arg}$ [69] can at least recognize one rare arginine codons AGA [67, 70]. But the mutated argU tRNAs carry G1-to-A1 transition and generate a mismatched base pair at the terminus of the acceptor stem. In bacteria, the G1-C72 base pair is one of the major identity determinants[23]. Previous research shows that the arginine acceptance is impaired significantly and in GM10 due to the mismatch [67, 70, 68]. Actually the Mass-Spec results confirmed our deduction that the AGA suppression rates were pushed over 90%, or even over 95%, by switching host cells from BL21(DE3) to GM10. Besides, GM10 is a good host bacterial strain to express engineered proteins, yielding over 10 mg/mL ***UAA***-incorporated Z domain proteins regularly.

The recoding of AGA *v.s.* the recoding of AGG

Compared with the recoding of AGG codon, the recoding of AGA codon shows similar ***UAA*** incorporation rates, varying by ***UAA***s. For example, the incorporation rates of IF, AcF, and OMeY, in response to either AGA or AGG, were all higher than 90%. However, the recoding of AGA is more direct-forward and easier than that of AGG, because the endogenous $tRNA_{UCU}{}^{Arg}$ mutant bacterial strain GM10 is available from CGSG center. Enlightened by this idea, we created DH10BW strain to improve ***UAA*** incorporation rates during the recoding of AGG and this strategy did increase ***UAA*** incorporation rates, but still were not as good as those in the recoding of AGA. Finally, to our surprise,in BL21(DE3) bacteria a simple increase of ***UAA*** working concentration in the process of recoding AGG pushed ***UAA*** incorporation rates and protein yields to the level similar as that of recoding AGA. Thus, AGG and AGA were both successfully reassigned from arginine to unnatural amino acids, under different but both accessible conditions. Moreover, control experiments showed that orthogonal $tRNA_{UCU}$s did not recognize AGG codon (YW19, YW20) (Table 3.1). Therefore one future direction can be focus working on the orthogonality between these two tRNA/codon pairs: $tRNA_{UCU}$/AGA and $tRNA_{CCU}$/AGG, and reassign AGA and AGG simultaneously.

3.6 Conclusion

In summary, we reassigned the sense codon AGA from arginine to unnatural amino acids. The successful suppression of AGA codon depends on the correct recognition of tRNA anticodon UCU by the ***UAA***-RS, and the interaction between ***UAA***s and the ***UAA*** binding pocket of ***UAA***-RS. The dual expression system also plays an important role to out-compete with the endogenous aaRS/tRNA.

This innovative reassignment of the sense codon AGA makes it possible to site-specifically incorporate multiple ***UAA***s into individual proteins when combining AGA with other blank codons, such as amber stop codon UAG. More important, similar strategies can be applied to recode other sense codons to ***UAA***s and provide more powerful tools to study protein functions with novel properties.

Chapter 4

Recoding of two codons at once: AGG and UAG

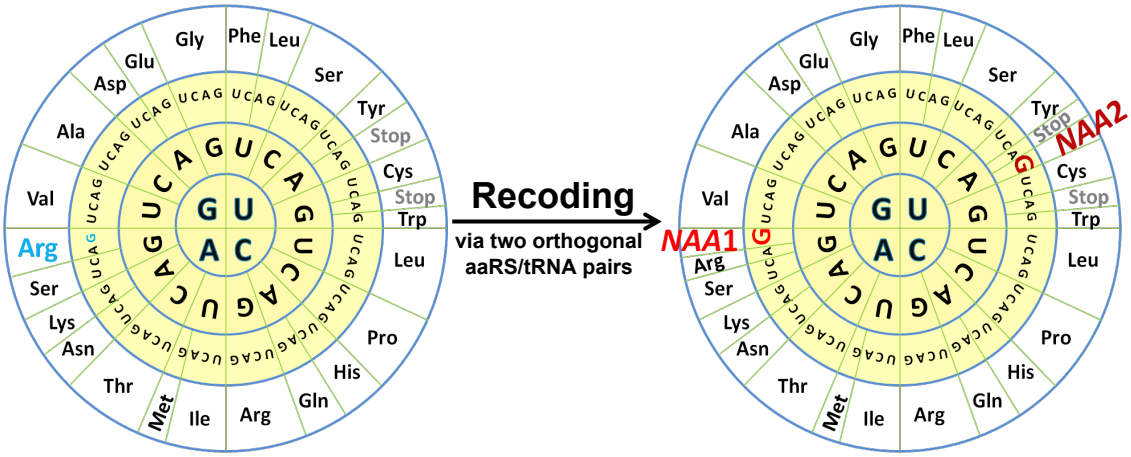


Figure 4.1: The recoding of AGG and UAG. ***NAA***, the abbreviation of non-canonical (unnatural) amino acid..

4.1 Two pairs of orthogonal *UAA*-RS/tRNA pairs were used to reassign AGG and UAG simultaneously

To expend the genetic code, the blank codon that uniquely designates the ***UAA***s was as important as the orthogonal aminoacyl-tRNA synthetase/tRNA (aaRS/tRNA)

pair. Theoretically, three groups of blank codons are possible for the ***UAA*** incorporation, including stop codons (e.g., UAG amber stop codon), frame-shift quadruplet codons (*e.g.*, AGGA), and rare codons (*e.g.*, AGG) . Also, the selection of proper orthogonal aaRS/tRNA pairs allows the correct incorporation of unnatural amino acids. There are two good candidates for this purpose: the evolved *Methanococcus jannschii* tyrosyl-tRNA synthetase (*Mj*TyrRS)/*Mj*tRNA$^{\mathrm{Tyr}}_{\mathrm{CUA}}$, and the naturally occurring wild-type or mutated *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (*Mm*PylRS$_{\mathrm{UAG}}$)/*Mm*tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$)[76]. They recognize the amber UAG codon and enabled the incorporation of more than 30 ***UAA***s into proteins in *E. coli*.

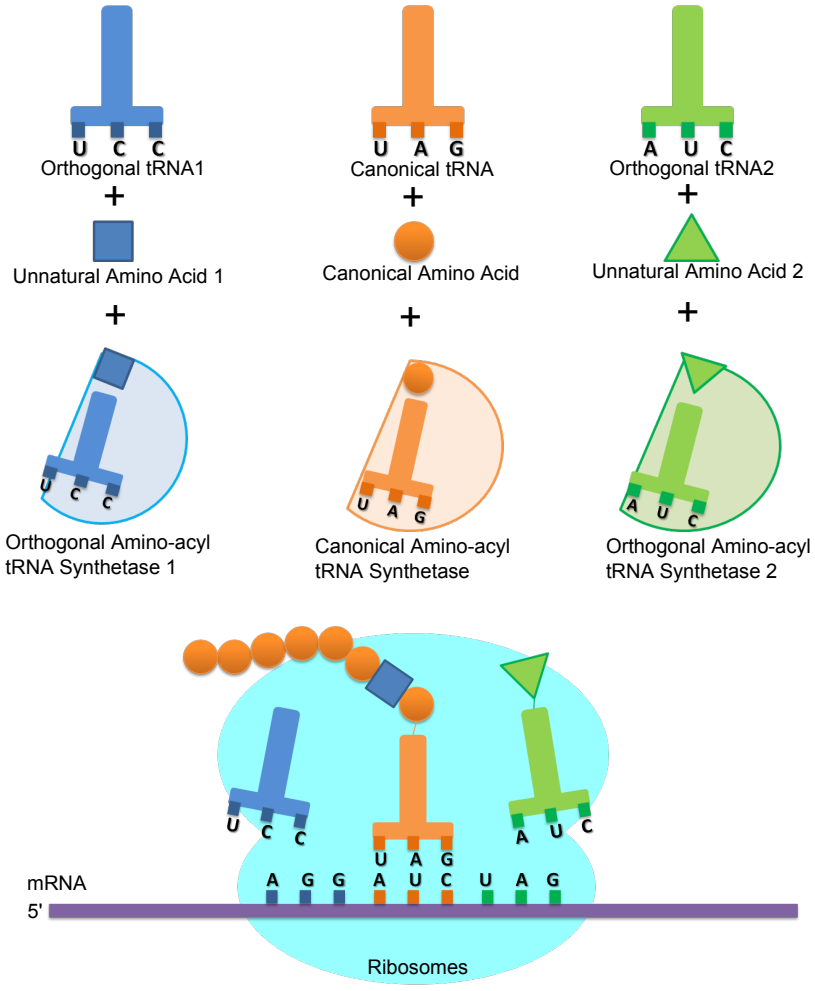


Figure 4.2: The recoding of AGG and UAG simultaneously depends on two sets of orthogonal aaRS/tRNA pairs.

In order to encode efficiently many distant unnatural amino acids into proteins, we

simultaneously reassigned AGG and UAG codons to different types of ***UAA***s in individual proteins by introducing two RS/tRNA pairs, *Mj* ***UAA***-RS_{AGG}/*Mj*$tRNA_{CCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$, into *E. coli* (Fig. 4.2). Our previous researches showed that we have successfully reassigned AGG codon from arginine to ***UAA***s in *E. coli* with the *Mj* ***UAA***-RS_{AGG}/*Mj*$tRNA_{CCU}$ pair which was evolved from *Mj*TyrRS/*Mj*$tRNA^{Tyr}_{CUA}$ (Chapter 2). Repeating results confirmed that the suppression rate of AGG codon by the evolved aaRS/tRNA pair was around 90% in general and the ***UAA***-decorated protein yield reached about 20 mg/L with an optimized expression system . Meanwhile, previous literature shows that *Mm*PylRS_{UAG}/ *Mm*$tRNA_{CUA}$ is able to take many pyrrolysine analogues as its substrate and incorporate them into proteins in response to UAG amber codon both in *E. coli* or in mammalian cells[40, 77, 78]. Furthermore, the potential mutual orthogonality between *Mj* ***UAA***-RS_{AGG}/*Mj*$tRNA_{CCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$, into *E. coli* contributed to the feasibility of this project[79]. As mentioned above, these two pairs of aaRS/tRNA were derived from different microorganisms, both are orthogonal to E. coli endogenous aaRS/tRNA pairs, and differ from each other significantly in conformation and function[3, 38, 41, 79].

4.2 Materials

The *Mm*Pyl-RS_{UAG} gene and primers was synthesized at IDTDNA. All other materials were the same as described in previous chapters.

4.3 Methods

Using *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$ to recode amber stop codon UAG

Constructing *Mm*Pyl-RS_{UAG} gene

The *Mm*Pyl-RS_{UAG} gene was synthesized to get rid of all the AGA and AGG codons within the gene. We tried to mutate the gene by overlapping PCR, however due to internal matching between primers and the gene, it was failed.

The mutant pyl-RS gene sequence without rare codons is: 5'-atggacaaaaaaccgct
gaataccctgattagcgcaaccggtctgtggatgagccgtaccggcaccattcataaaatcaaacatcatgaagtgagc
cgcagcaaaatctatattgaaatggcatgtggtgatcatctggtggtgaataatagccgtagcagtcgtaccgcacgtgc
actgcgtcatcacaaatatcgtaaaacctgtaaacgttgccgtgtgagtgatgaggatctgaacaaatttctgaccaaag
ccaatgaagatcagaccagcgttaaagttaaagttgttagcgcaccgacccgtaccaaaaaagcaatgccgaaaagcg
ttgcccgtgcaccgaaaccgctggaaaataccgaagcagcacaggcacagccgagcggtagcaaattttcaccggcaa
ttccggttagcacccaagaaagcgttagcgttccggcaagcgtgagcaccagcattagcagcatttcaaccggtgcaac

cgcaagcgcactggttaaaggtaataccaatccgattaccagcatgagcgcaccggttcaggcaagtgcaccggcactg
accaaaagtcagaccgatcgtctggaagttctgctgaatccgaaagatgaaattagcctgaatagcggtaaaccgtttc
gtgaactggaaagcgaactgctgagccgtcgtaaaaaagacctgcaacaaatttatgccgaagaacgcgaaaactatc
tgggtaaactggaacgtgaaatcacccgttttttttgtggatcgtggtttcctggaaatcaaaagcccgattctgattccgc
tggaatatattgaacgtatgggcattgataacgataccgaactgagcaaacaaatctttcgcgtggataaaaacttttgt
ctgcgtccgatgctggcaccgaatctgtataactacctgcgcaaactggatcgtgccctgcctgatccgattaaaatcttt
gaaattggtccgtgctaccgcaaagaaagtgatggtaaagaacacctggaagaattcaccatgctgaacttttgccaga
tgggtagcggttgtacccgtgaaaatctggaaagcattattaccgattttctgaaccatctgggcatcgatttcaaaattg
ttggtgatagctgcatggtgtttggtgataccctggatgttatgcatggtgatctggaactgagtagcgcagttgttggtc
cgatcccgctggatcgcgaatggggtattgataaaccgtggattggtgcaggttttggtctggaacgcctgctgaaagtt
aaacacgacttcaaaaacattaaacgtgcagcacgtagcgagagctattacaatggtattagtacgaacctgtaa-3'

The amino acid sequence of the mutant pyl-RS (pylRS-tsao) is: MDKKPLNTL
ISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTARALRH
HKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSV
ARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASA
LVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELE
SELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIER
MGIDNDTELSKQIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPC
YRKESDGKEHLEEFTMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSC
MVFGDTLDVMHGDLELSSAVVGPIPLDREWGIDKPWIGAGFGLERLLKVKH
DFKNIKRAARSESYYNGISTNL*. There is only one mutation Y384F compared to wild-type *Mm*Pyl-RS because previous reports indicated that the PylRS(Y384F) mutant exhibited higher amber suppression activities with lysine derivatives and pyrrolysine than those of the wild-type PylRS [35].

Constructing *Mm*tRNA$_{CUA}$ gene

The sequence of pylT *Mm*tRNA$_{CUA}$ is 5'-GGAAACCUGAUCAUGUAGAUCGAA
UGGACUCUAAAUCCGUUCAGCCGGGUUAGAUUCCCGGGGUUUCCGCCA-
3'. In the vector map, pylT gene has a lpp promoter and a rrnC terminator. The lpp-pylT-rrnC was placed right in front of lpp-JYTRN-rrnC, the *Mj* tRNA$^{Tyr}_{AGG}$ gene. First, a dual-tRNA sequence lpp-pylT$_{CUA}$-rrnC-lpp-JYTRN$_{UCU}$-rrnC was constructed by overlapping PCR with primers TF1, TR1, TF2, TF3, TR2, TF4 with template pAGA3-Zaga. Then in order to change the *Mj*tRNA anticodon UCU to UCC, overlap PCRs were performed by using lpp-pylT$_{CUA}$-rrnC-lpp-JYTRN$_{UCU}$-rrnC as templates and primers tRNA-AGG3-3, pAGG2-1F, pAGG-Sphl-f, and tRNA-4. The final product was lpp-pylT$_{CUA}$-rrnC-lpp-JYTRN$_{UCC}$-rrnC, a dual-tRNA sequence for *Mm* pyl-tRNA$_{CUA}$ and *Mj* tyr-tRNA$_{UCC}$

Unnatural amino acids encoded by pylRS/tRNA

In this study, two kinds of ***UAA***s were used. They are Boc-L-lysine (BocK) and propargyloxycarbonyl-L-lysine (AlK) (Fig. 4.3) .

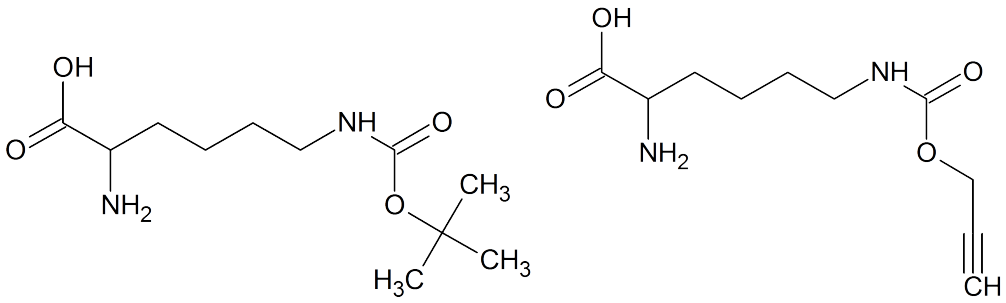


Figure 4.3: The structure of two ***UAA***s: BocK(left) and AlK(right).

Recoding both AGG and UAG simultaneously by *Mj* *UAA*-RS_{AGG}/$tRNA_{CCU}$ & *Mm*Pyl-RS_{UAG}/$tRNA_{CUA}$

Constructing pAK8 vectors to carry two synthetases

pAK8 vectors carry two different RS_{UAG} genes with two promoters – a modified gln S (constitutive) promoter and an araBAD (arabinose-inducible) promoter, the araC repressor gene a p15A origin of replication, and the kanamycin resistant gene marker. pCNF-RS_{AGA} is controlled by inducible araBAD promoter while pyl-RS_{TAG} is after the constitutive glnS' promoter. The pCNF-RS_{AGA} gene carries a restriction site (BglII or SacI) at each end, whereas one restriction enzyme (PstI or NedI) cutting site was placed at each end of the gene of the other synthetase pyl-RS_{TAG} (Fig. 4.4). These restriction enzyme cutting sites were utilized for the construction of pAK8 vector by using pAK3 as the templates.

Constructing pt2Z2GX and pt2Z238GX to carry tRNAs and the mutant Z domain gene

A restriction enzyme cutting site, either PstI or NgoMIV, was placed at each end of the dual-tRNA sequence lpp-pylT_{CUA}-rrnC-lpp-JYTRN_{UCC}-rrnC. These two enzymes were utilized to replace the single tRNA sequencelpp-JYTRN_{UCC}-rrnC in pAGG3-Zagg vector. Both pt2Z2GX and pt2Z238GX carried lpp-pylT_{CUA}-rrnC-lpp-JYTRN_{UCC}-rrnC. NgoMIV and PstI were utilized for double digestion and ligation to put the dual-tRNA sequence into vectors.

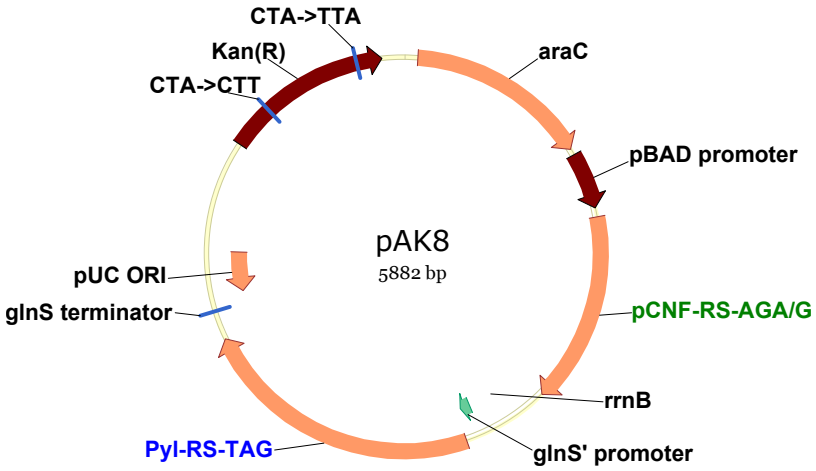


Figure 4.4: pAK8 vector

To construct pt2Z2GX vector (Fig. 4.5), two primers F-Z7agg9tag and R-Z7agg9tag were used to change the codon for 7th amino acid lysine to AGG and the codon for 9th amino acid asparagine to amber stop codon TAG. BglII and SpeI, these two restriction enzymes were used for cutting off the old DNA sequence and ligate the target DNA. T4PNK treated primers mix (F-Z7agg9tag and R-Z7agg9tag) were diluted 100 times and then ligated into the template to form the pt2Z2GX vector.

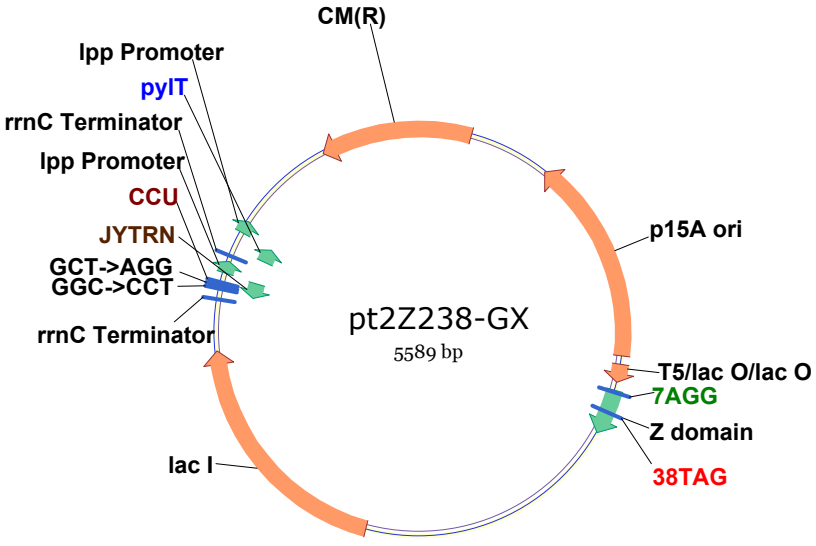


Figure 4.5: pt2Z238GX vector

In pt2Z238GX vector, there are two mutations in Z domain gene: the codon for 7th amino acid lysine was altered to AGG, and the codon for 38th amino acid lysine

was changed to TAG stop codon. To perform overlapping PCR, primers Z7aga9tag-F, Z38tag-R, Z38tag-F, and Z-Blp-R were used to introduce 38AGG mutation into Z domain gene with pAGA3-Zaga as templates, and BlpI and BglII were used for double digestion and ligation. After that, another pair of primers Z-7AGG-f(Z1f) and Z-7AGG-r(Z1r) were used to generate 7AGG mutation into Z domain gene, and then the new DNA sequence was ligated into templates to form pt2Z238GX vector with the help of SpeI and BglII.

Expressing and purifying mutant Z domain proteins decorated by two unnatural amino acids

pAK8 vectors were co-transformed with either pt2Z238GX or pt2Z2GX into BL21(DE3) or DH10BW *E. coli*. A single colony from each co-transformation was grown overnight at 37 °C in 2 ml of LB media supplemented with kanamycin and chloramphenicol to obtain the starter culture. The starter culture was then twenty times diluted to a 20 mL LB medium containing 50 μg/ml kanamycin, 34μg/mL chloramphenicol and two types of unnatural amino acids for further culturing. Three pairs of different unnatural amino acids were tested: AcF(3mM) and AlK(4.5mM), AzF(3mM) and AlK(4.5mM), IF(3mM) and BocK(1.5mM). Till the culture reach OD600=0.6-0.8, 0.02 percent arabinose was used for induction, and followed by 1mM IPTG induction 10 minutes later. After expression for 14-16 h at 37 °C, 20 ml cultures were pelleted at 8,000 rpm and frozen at -80°C. Protein purification methods are the same as mentioned in Chapter AGA and Chapter AGG.

Protein electrophoresis and calculating protein yields

Methods are the same as stated in Chapter 2 and Chapter 3.

Analyzing protein molecule weight by mass spectrometry

Protein molecular weight was analyzed by mass spectrometry. Protein samples which showed obvious bands with correct sizes on SDS-PAGE gel were sent out for mass spectrometry (ESI or MALDI-TOF) in 70% ACN.

Modifying unnatural amino acids

The unnatural amino acids AcF of Z domain proteins was labeled by 5-dimethyla minonaphthalene-1-sulfonyl hydrazine (dansyl hydrazine, DNSH)[58]. The DNSH labeling reaction was carried by the following steps: mixed targets protein AcF-Z 30μL with 6.5% TFA/ACN 150μL, then added 100μL 5mM DNSH into the mixture. The mixture was shaken gently over night at room temperature (about 24°C). It is critical to keep the mixture acidic for this test.

The unnatural amino acids AlK of Z domain protein was modified by PEGylation. The PEGylation was mediated by a click chemistry reaction, Copper(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC)[80, 81, 82, 83]. We chose a type of PEG molecules with an azide group, Methoxypolyethylene glycol azide(5Kb). The Azide group of PEG molecules could form covalent bonds with AlK by performing CuAAC reactions [28, 12, 10]. 4μl 5mM CuSO4 was mixed with 1μl 20mM bathophenanthroline; then 6μl Z domain proteins(50-100μM) and 5μl 10mM PEG(Methoxypolyethylene glycol azide,5Kb) were added and mixed well; finally 4μl 5mM ascorbic acid was added into the mixture. The mixture was shaken gently over night at room temperature (about 24°C) (Fig. 4.6). The key point of this experiment is saving the acorbic acid to the last step [84].

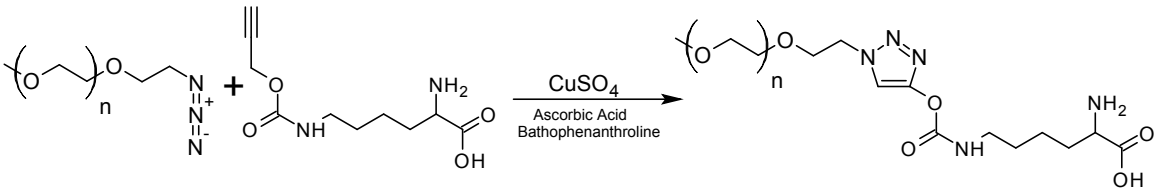


Figure 4.6: AlK reacts with methoxypolyethylene glycol azide through opper(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC).

The Z proteins with both AcF and AlK first underwent DNSH labeling, and then Amicon ultra filter (3K) was used to change buffer to pure water. After that, Z domain proteins labeled by DNSH were PEGylated by 5Kb PEG Methoxypolyethylene glycol azide.

For mutant Z domain proteins Z-7AzF-38AlK which were charged with AzF and AlK at the same time, the click chemistry reaction CuAAC was designated . In wild-type Z domain protein, the 7th lysine is very close to the 38th lysine on 3D space. When these two amino acids was replaced by AzF and AlK, the click chemistry CuAAC might be able to link AzF and AlK and form very stable convalent bonds if there is no big conformational changes. Futher, the AzF and AlK from different protein molecules might react with each other to form protein dimers, or trimers. To perform a CuAAC reaction beween AzF and AlK, 4μl 5mM CuSO4 was mixed with 1μl 20mM bathophenanthroline; then 11μl Z domain proteins(50-100μM) were added and mixed well; finally 4μl 5mM ascorbic acid was added into the mixture. The mixture was shaken gently over night at room temperature (about 24°C). The key points of this experiment were adding the acorbic acid at last, and adding enough amount of proteins to keep the protein concentration high enough.

All labeled proteins were subject to SDS-PAGE electrophoresis after the reactions finished.

Growth curves of pAK8+pt2Z238GX/DH10BW

pAK8+pt2Z238GX which contains genes for textitMj ***UAA***-RS_{AGG}/*Mj*$tRNA_{CCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$ were co-transformed into BL21(DE3). One single colony was grown overnight to saturation in LB media supplemented with 50 μg/ml kanamycin and 34 μg/ml chloramphenicol. Culture were 1:100 diluted and divided into 10 samples. Each sample was treated separately and then aliquoted into 96-well assay plates in octuplicate. The treatments for each sample included: the first sample, no NAA and no arabinose added; the second sample, no NAA but 0.02% arabinose added; the third sample, 4.5mM AlK and 3mM AcF, but no arabinose added; the fourth sample,4.5mM AlK and 3mM AcF, and 0.02% arabinose added; the fifth sample, 4.5mM AlK and 3mM AzF, but no arabinose added; the sixth sample, 4.5mM AlK and 3mM AzF, and 0.02% arabinose added; the seventh sample, 3mM AcF but no arabinose added; the eighth sample, 3mM AcF and 0.02% arabinose added; the nineth sample, 3mM AzF but no arabinose added; the tenth sample, 3mM AzF and 0.02% arabinose added; the eleventh sample, 4.5mM AlK but no arabinose added; the twelfth sample, 4.5mM AlK and 0.02% arabinose added. Plates were grown in a SpectraMax M2e plate reader (Molecular Devices) and read at 600 nm every 200 seconds with mixing of 120 seconds/read at 37°C for 16 hours. Relative bacteria densities are indicated for 200 μl cultures and are an average of eight independent cultures.

4.4 Results

AGG and UAG were reassigned to the desired unnatural amino acids simultaneously

Z-7AGG-38TAG can not be translated into Z domain proteins without the presence of AlK

pAK8+pt2Z238GX was transformed into DH10BW and the Z-7AGG-38TAG mutant was induced to express Z domain proteins with AlK only, AzF only, and AlK+AzF.SDS-PAGE electrophoresis and Commassie blue staining results showed that without the presence of AlK, no protein band was detected at 8kDa. When both AlK and AcF were added into the culture media, a thin band at 8kDa was detected; while only AlK presented during the protein expression procedure, a relatively(about 1/4 of the AlK+AcF) thinner band was seen. These results support that the AlK was inserted to the Z domain protein at the 38th position in response to TAG amber stop codon but not the 7th AGG codon (Fig. 4.7) .

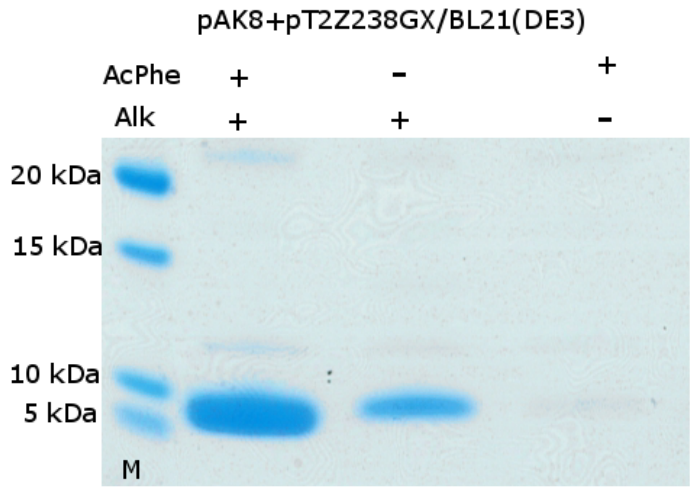


Figure 4.7: SDS-PAGE gel: the expression of Z-7AGG-38UAG proteins with the presence of different ***UAA***s.

Table 4.1: The molecular weights of Z domain proteins when its two amino acids were replaced by ***UAA***s.

Z domain protein	Z-7AzF-38Alk	Z-7AcF-38Alk	Z-7IF-38Alk	Z-7AzF-38BocK	Z-7OMeY-38BocK
Full length	8035.6	8037.6	8120.5	8053.6	8043
lost 1Met; with Ac	7946.4	7947.4	8031.3	7964.4	7954
lost 1 Met	7904.4	7905.4	7989.3	7922.4	7911
Z domain protein	Z-7arg-38Alk	Z-7arg-38BocK	Z-7IF-38BocK	Z-9BocK	Z-7IF-9BocK
Full length	8003.6	8021.6	8138.5	8035.7	8152.5
lost 1Met; with Ac	7914.4	7932.4	8049.3	7946.5	8063.3
lost 1 Met	7872.4	7890.4	8007.3	7904.7	8021.3

ESI results confirmed the incorporation of multiple *UAA*s

When normal canonical amino acids were replaced by designated unnatural amino acids, the molecule weights of the mutant Z domain proteins were changed correspondingly as well. The new mass weights are listed in the following table (Table 4.1). The relative double-*UAA* incorporation rate was assessed by the equation of two *UAA*s decorated proteins (Z-***UAA***1-*UAA*2)/total proteins (Z-***UAA***1-*UAA*2 + all other forms of Z domain proteins).

For Z-7AzF-38AlK proteins purified from pAK8+pt2Z238GX/DH10BW, the de-

convoluted ESI-MS spectrum reveals two major peaks at m/z =7904, and 7945 (Fig. 4.8) . These two peaks match the expected molecular weight for the AzF(at the 7th position) and AlK(at the 38th position) incorporated Z-domain mutant with the loss of its first methionine (m/z =7904.4), and its subsequent post-translational acetylation product (m/z =7946.4). There is also a minor peak at m/z =7872, matching the 38th AlK and 7th arginine incorporated Z-domain protein with the loss of its first methionine (m/z=7872.4). The double ***UAA***s(AzF+AlK) incorporation rate of YW92 was about 83%.

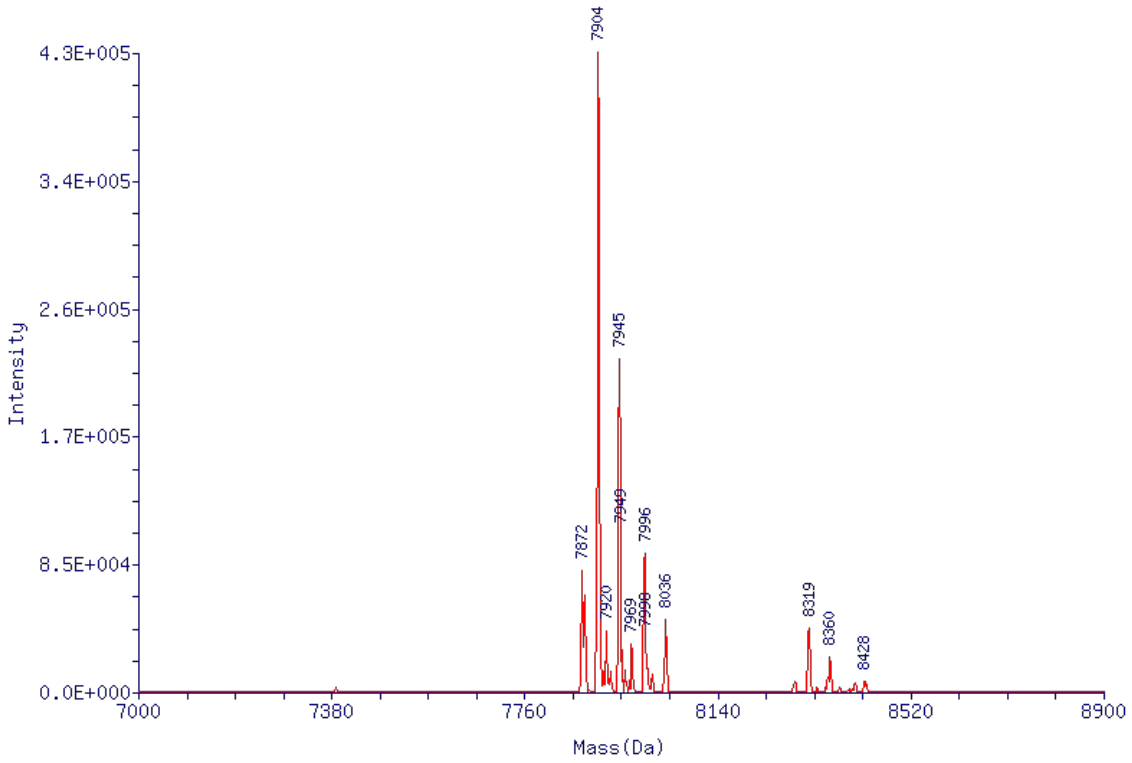


Figure 4.8: ESI results of Z-7AzF-38AlK (YW92) (DH10BW).

Similarly, for Z-7AcF-38AlK proteins purified from pAK8 and pt2Z238GX transformed *E. coli* (DH10BW and BL21(DE3)), the deconvoluted ESI-MS spectrum reveals three peaks at m/z =7905/7906, 7947 and 8039 (Fig. 4.9,. 4.10), which match the expected molecular weight for the AcF(at the 7th position) and AlK(at the 38th position) incorporated Z-domain mutant with the loss of its first methionine (m/z =7905.4), its subsequent post-translational acetylation product (m/z =7947.4), and

its full length format (m/z=8036.6). There is also a minor peak at m/z =7872, matching the 38th AlK and 7th arginine incorporated Z-domain protein with the loss of its first methionine (m/z=7872.4). The double ***UAA***s(AcF+AlK) incorporation rate of YW92 in DH10B2 was about 85%, and that of YW99 in BL21(DE3) was 87%.

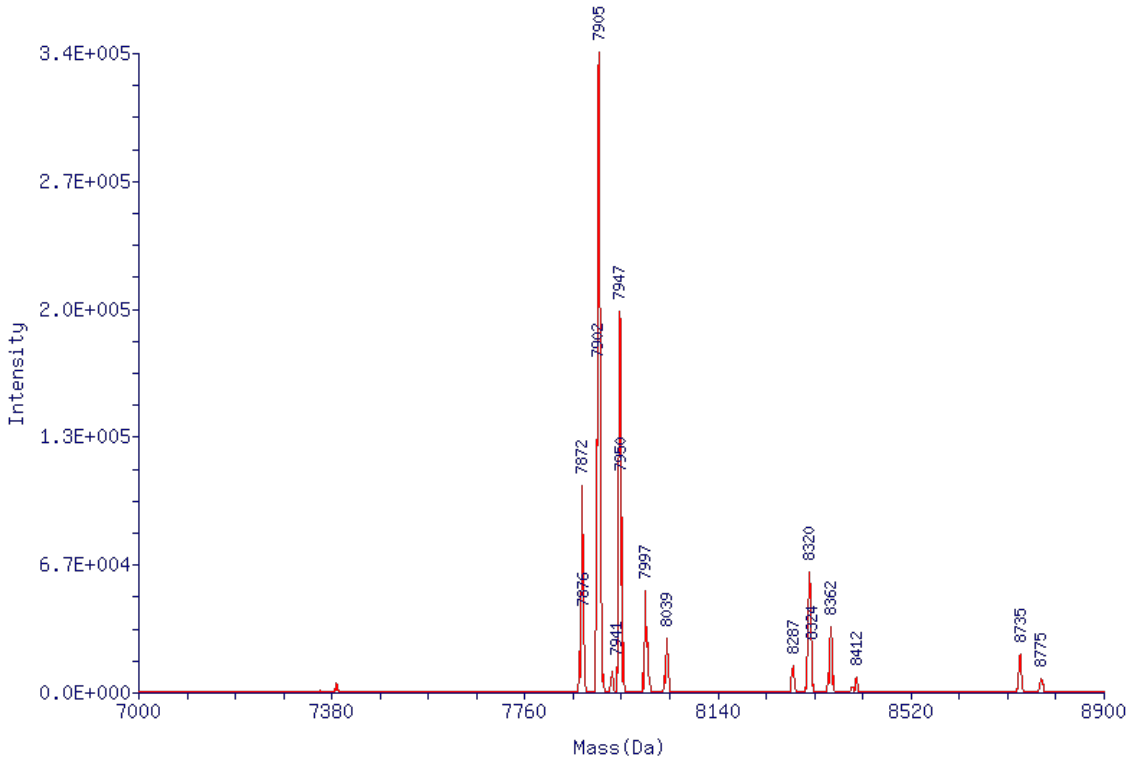


Figure 4.9: ESI results of Z-7AcF-38AlK from DH10BW (YW93).

MALDI-TOF results confirmed the incorporation of multiple *UAA*s

To analyse IF and BocK decorated Z proteins, MALDI-TOF was employed because BocK is unstable during ESI process.

For Z-7IF-9BocK proteins purified from pAK8+pt2Z2GX/DH10BW bacterial, the MALDI-TOF results show two major peaks at m/z =8021.953, and 8063.983 (Fig. 4.11) , which match the expected molecular weight for the IF(at the 7th position) and BocK(at the 9th position) incorporated Z-domain mutant with the loss of its first methionine (m/z =8021.3) and its subsequent post-translational acetylation product (m/z =8063.3). There is also a minor peak at m/z =7905.740, matching the

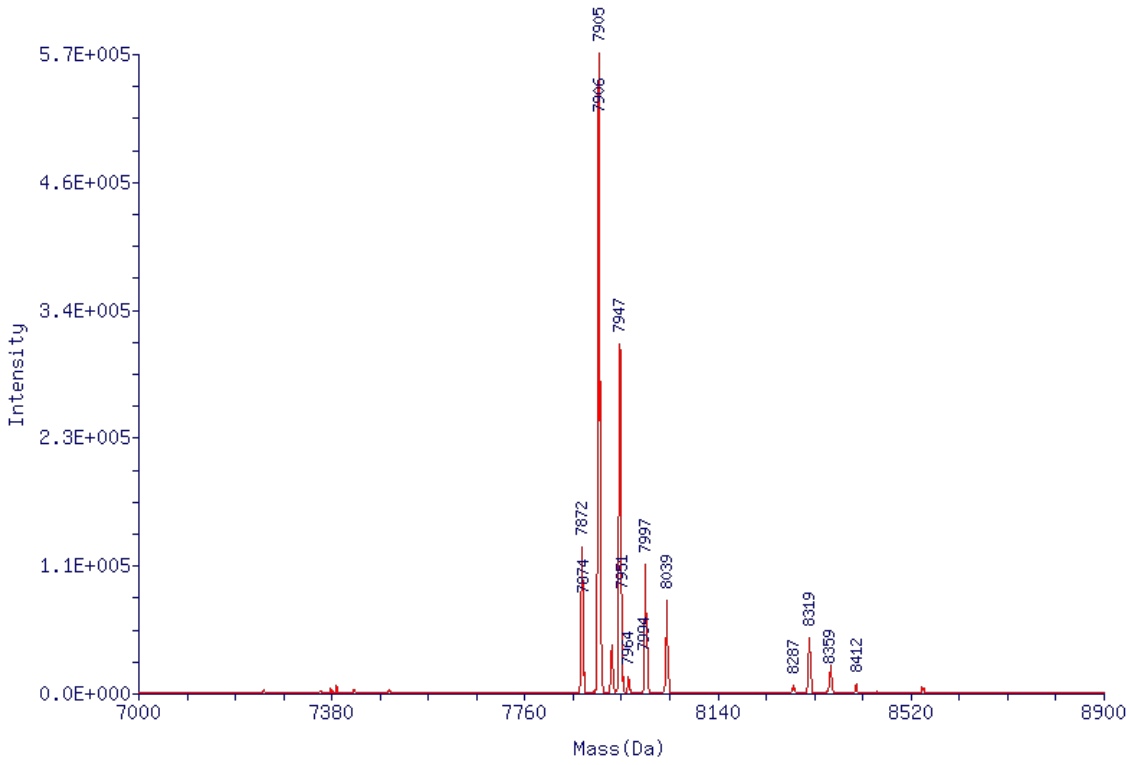


Figure 4.10: ESI results of Z-7AcF-38AlK from BL21(DE3) (YW99).

9th BocK and 7th arginine incorporated Z-domain protein with the loss of its first methionine (m/z=7904.7). The double ***UAA***s(IF+BocK) incorporation rate of this test was about 88%.

For Z-7IF-38BocK proteins purified from pAK8+pt2Z238GX/DH10BW bacterial, the MALDI-TOF results show two major peaks at m/z =8011.006, and 8053.093 (YW3M) (Fig. 4.12), which match the expected molecular weight for the IF(at the 7th position) and BocK(at the 38th position) incorporated Z-domain mutant with the loss of its first methionine (m/z =8007.3) and its subsequent post-translational acetylation product (m/z =8049.3). There is also a minor peak at m/z =7893.949, matching the 38th BocK and the 7th arginine incorporated Z-domain protein with the loss of its first methionine (m/z=7890.4). There was a m/z=4 shift in this whole MALDI-TOF spectrum. The double ***UAA***s(IF+BocK) incorporation rate of this test YW4M was about 83%. On the contrary, the ESI result of the same sample (Z-7IF-38BocK, YW91-ESI, YW4M-MADI-TOF) (Fig. 4.13)reveals multiple peaks corresponding to 38lysine instead of 38BocK at m/z=7791(matching Z-7R-38K-w/o1M),

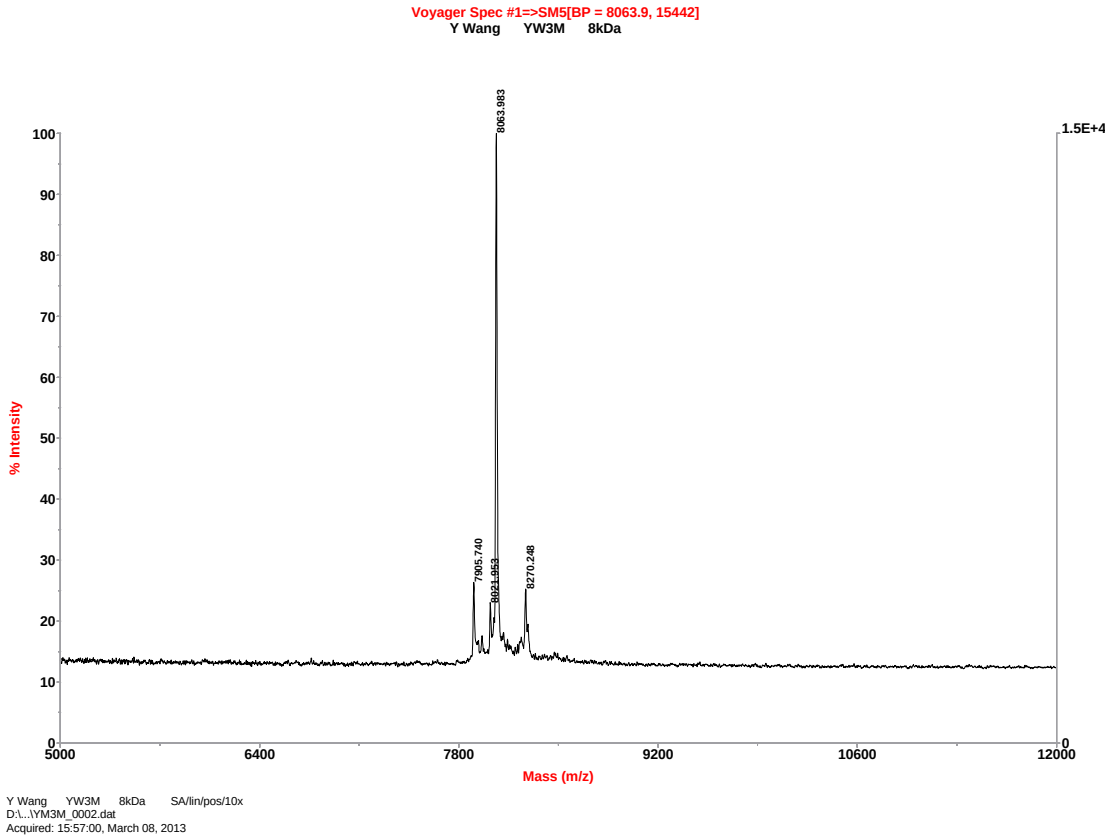


Figure 4.11: MALDI-TOF results of Z-7IF-9BocK (YW3M).

7906(matching Z-7IF-38K-w/o1M), 7949(matching Z-7IF-38K-w/o1M+Ac). YW91 also contains two peaks at m/z=8009 and 8048, matching Z-7IF-39BocK-w/o1M and Z-7IF-38BocK-w/o1M+Ac. The double ***UAA***s(IF+BocK) incorporation rate of YW91 was merely about 38%, if one simply calculates from the numbers. However, in DH10BW bacteria UAG is a strict stop condon, so it's impossible for lysine to be incorporated onto position 38 in response to UAG codon. Thus, the comparison between YW4M and YW91 clearly demonstrates that BocK is not a good substrate for ESI examination because more than half amount of BocK will lose its Boc group and become lysine.

Protein yields

The OD_{280} values of mutant Z domain proteins in water were usually ranging from 0.05(or less) to 0.1, depending on bacteria strains and ***UAA*** types. So, by using the

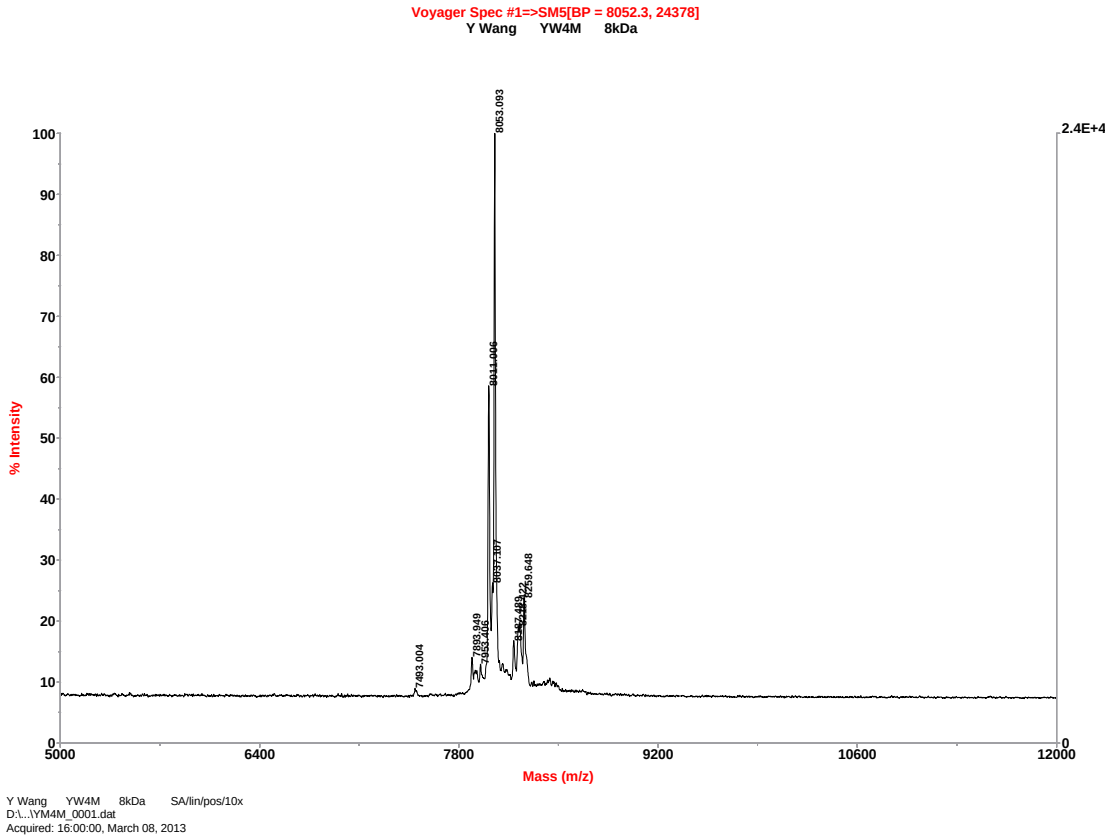


Figure 4.12: MALDI-TOF results of Z-7IF-38BocK (YW4M).

equation $(OD280) \times 6.15 \times 0.800 \div 25 \times 1000$, the results were about 10 mg/L to 20 mg/L. The yields of ***UAA***s decorated Z domain proteins were about 20 mg/L under optimized conditions in BL21(DE3), while the Z domain protein yields were roughly 5-10mg/L or less in DH10BW and DH10B cells.

Multiple unnatural amino acids were labeled

The Z-7AcF-38AlK proteins were either labeled with DNSH, or linked with PEG(5kb, with azide groups), or underwent both treatments (Fig. 4.14) . After SDS-PAGE electrophoresis, Z-7AcF-38AlK with only DNSH conjugation displayed as a green-yellowish fluorescent band under UV excitation, while Z-7AcF-38AlK with both DNSH conjugation and PEGylation appeared as a blurry and wide green-yellowish fluorescent band, at a relative higher position than that of Z-7AcF-38AlK with only DNSH conjugation. No any other fluorescence was observed other samples, untreated

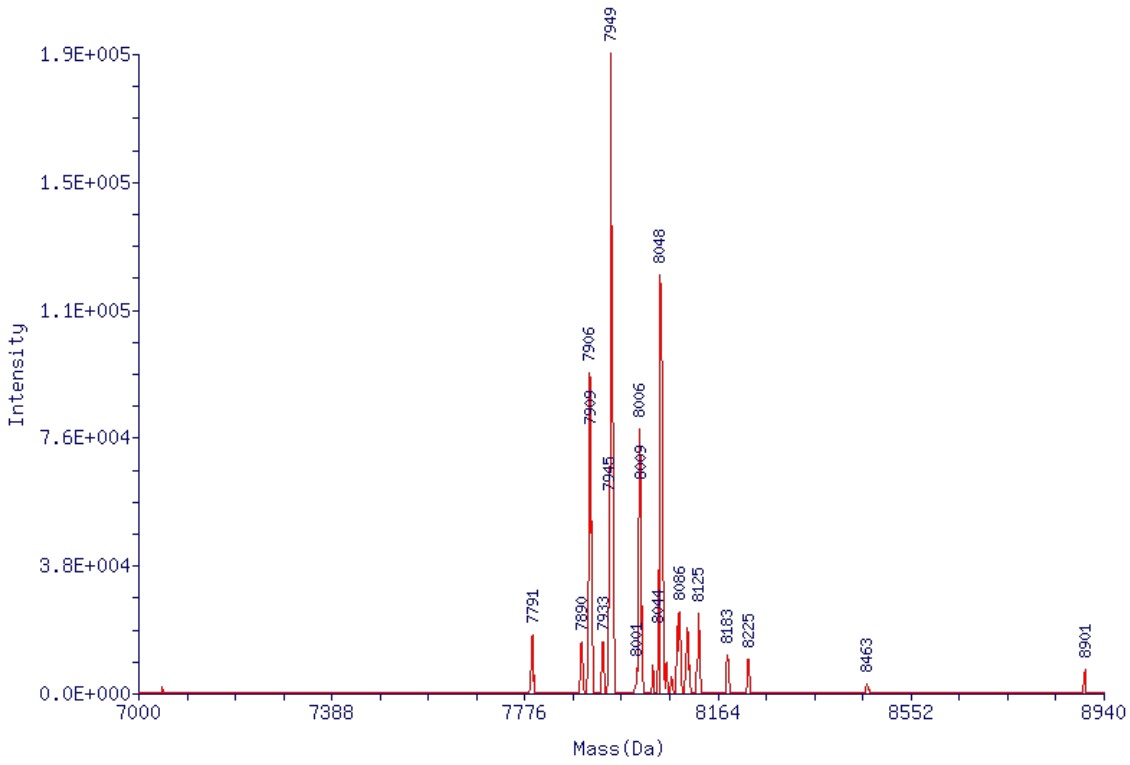


Figure 4.13: ESI results of Z-7IF-38BocK (YW91).

and PEGylated Z-7AcF-38AlK. After Commassie blue staining, we found that Z-7AcF-38AlK with only DNSH conjugation was at the same position(between 3.5Kb and 10Kb) as untreated mutant Z domain protein. But Z-7AcF-38AlK with only PEGylation and the same protein with two treatments were both at the higher position, around 20kb and blurry. So the band of Z-7AcF-38AlK with both DNSH labeling and PEGylation moved slower and showed green-yellowish fluorescence, which reconfirmed the incorporation of both unnatural amino acids. In control lanes, Z-7AcF with DNSH labeling showed fluorescence while Z-38AlK had no fluorescence; Z-38AlK with PEGylation was also shifted to an around 20kb position whereas Z-7AcF did not shifted above, but both were blurry. The blurry bands of all samples added PEG are likely to be the results of nonspecific localized changes in conductivity caused by PEG itself. The local conductibility is lower in the presence of the PEGmicelle complex than in its absence[85]. So this also contributes to the more than 5kb shift of PEGylated protein bands but the expected molecular weights of PEGylated Z domain protein should be 8kb+5kb=13kb, instead of the observed 20kb .

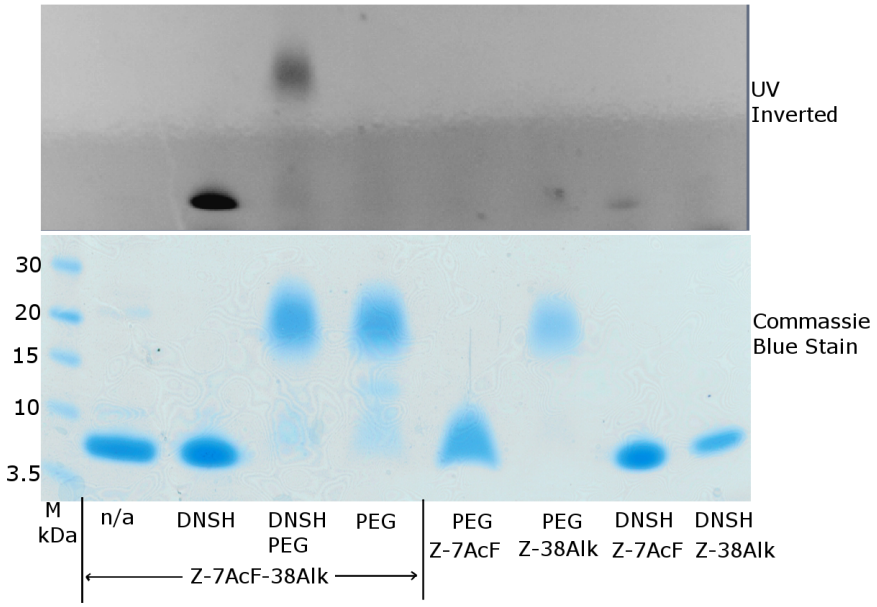


Figure 4.14: SDS-PAGE gel: DNSH labeling and PEGylation on Z-7AcF-38AlK proteins.

We also performed CuAAC on the mixture of Z-7AzF and Z-38AlK, and Z-7AzF-38AlK itself. The CuAAC reactions of Z-7AzF itself and Z-38AlK itself were used as control. On Commassie blue stained SDS-PAGE gel we found that there were two bands, one at 8KDa (the monomer of Z domain proteins) and the other one a little above 15KDa (might be the dimmer), for the lanes loaded with CuAAC of the mixture of Z-7AzF and Z-38AlK, and Z-7AzF-38AlK itself; whereas only one band at 8KDa was detectable for the two controls(CuAAC of of Z-7AzF itself and Z-38AlK itself) (Fig. 4.14,a). Further more, when the protein loading amount doubled (Fig. 4.14,b), the CuAAC of the mixture of Z-7AzF and Z-38AlK gave three bands totally, very close to 8KDa(monomer), 16KDa(dimmer), and 24KDa(trimmer).

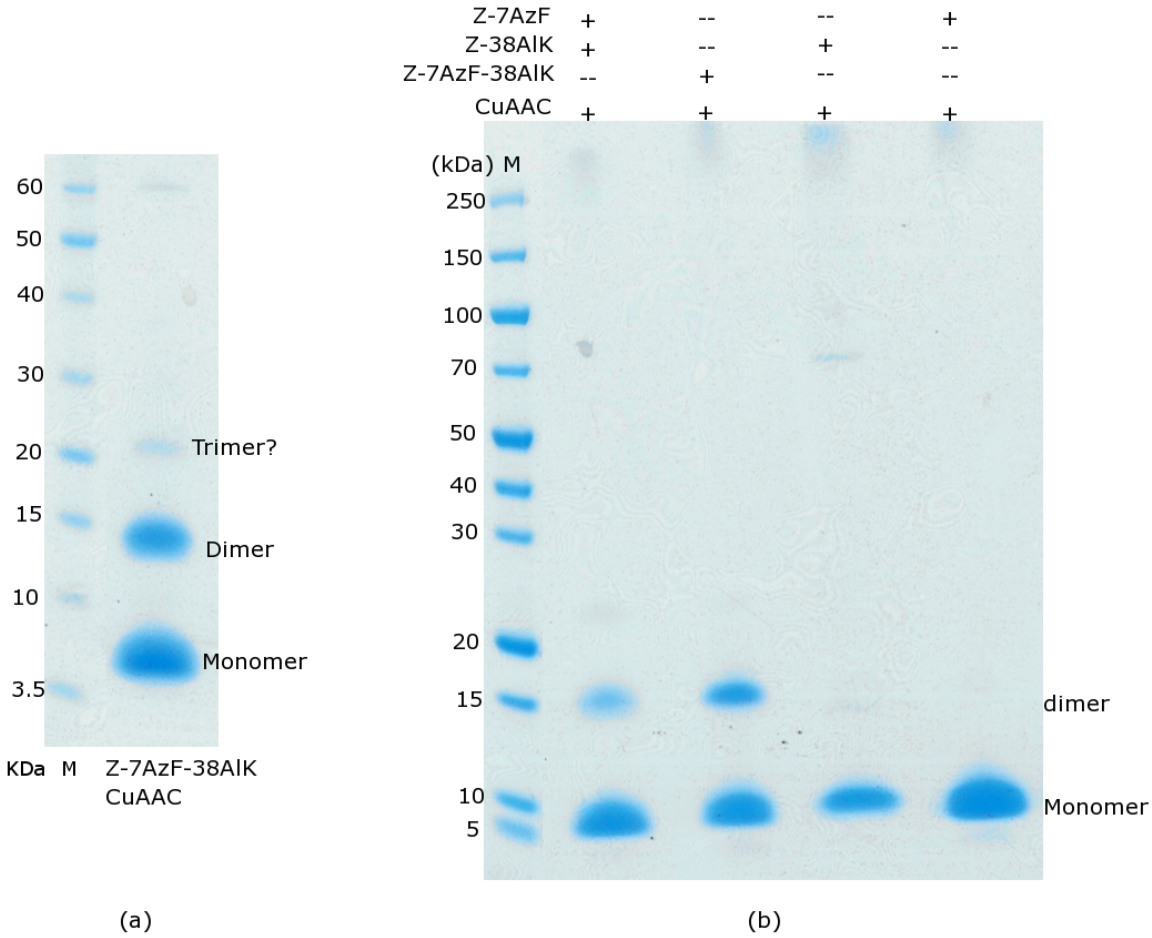


Figure 4.15: SDS-PAGE gel: CuAAC reaction of Z-7AzF-38AlK proteins.

pAK8+pt2Z238GX/DH10BW bacteria proliferate with *UAA*s

First, for none-***UAA*** test pair ('N/A' and 'ara'), the production of textitMj ***UAA***-RS_{AGG} might affect the bacterial growth to certain extent because after adding arabinose, the log phase was elongated and the bacteria population delayed the time to reach saturation(OD_{600}=0.6-0.7). We also find that AlK does not impact the DH10BW cell growth rate because the AlK-only curve was very close to the untreated bacteria, even the pyl-RS_{TAG} existed all the time since glnS' is a constitutive promoter. Similarly the 'AlK+ara' curve had a longer log phase (about 10 hours) to reach stationary phase than the one without arabinose (about 3 hours), but still very close to the 'ara' curve, which is another evidence for the impact caused by accumulation of textitMj ***UAA***-RS_{AGG} . Second, based on the results of AzF group

('AzF' and 'AzF+ara') and AlK+AzF group ('AlK+AzF' and 'AlK+AzF+ara'), we see that the adding of AzF is medium-level toxic to bacterial growth because 'AzF' curve and 'AlK+AzF' curve were similarly flatter, and the saturation concentrations (OD_{600}=0.5) were both lower than 'N/A' (OD_{600}=0.6-0.7). Further, if arabinose was added , both of them ('AzF+ara' and 'AlK+AzF+ara') became even more difficult to finish the flat log phase and the final bacterial concentrations were only near OD_{600}=0.3. Last, from the results of AcF pair ('AcF' and 'AcF+ara') and AlK+AcF pair('AlK+AcF' and 'AlK+AcF+ara'), it is very obvious that *AcF dramatically prohibit the bacteria growth*. Two curves with AcF ('AcF' and 'AlK+AcF') but without arabinose both demonstrated an extended lag phase of eight hours and the final concentrations were merely OD_{600}=0.3 after 14 hours. As expected, the adding of arabinose to these two tests ('AcF+ara'and 'AlK+AcF+ara') further halted the bacterial proliferation by returning an almost flat curve and the final concentration only increased to OD_{600}=0.17 from the original OD_{600}=0.12 (Fig. 4.16).

4.5 Discussion

The orthogonality of two exogenous RS/tRNA pairs makes them qualified for two-codon recoding

There are at least two types of orthogonality for the purpose of two-codon reassigning. First, each pair of exogenous RS/tRNA, either *Mj* ***UAA***-RS_{AGG}/*Mj*$tRNA_{CCU}$ or *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$, can not react with the endogenous aaRS/tRNA pairs of *E. coli*. Many previous studies have proved that both *Mj* ***UAA***-RS_{UAG}/$tRNA_{CUA}$ and *Mm*Pyl-RS_{UAG}/$tRNA_{CUA}$ are orthogonal within *E. coli*[3, 38, 37]. And our previous researching results also demonstrated that the evolved *Mj* ***UAA***-RS_{AGG}/$tRNA_{CCU}$ only insert ***UAA***s in response to AGG codon. Second, these two exogenous RS/tRNA pairs should not react with each other's components. Protein structure studies show that these two aaRSs belong to two different classes: *Mm*PylRS is an archaeal class II aminoacyl-tRNA synthetase[37] whereas *Mj*TyrRS is originated from class I. More important, our research results clearly showed that if only AlK, the substrate of *Mm*PylRS_{UAG}/$tRNA_{CUA}$, presents during protein expression, the target proteins whose gene carries one UAG and one AGG can be produced; but if only AcF, the substrate of *Mj* ***UAA***-RS_{AGG}/$tRNA_{CCU}$, is added into culture media, the translation of the same target proteins will be stopped. This means either *Mm*pylRS_{UAG}/$tRNA_{CUA}$ can not insert anything else, including AzF, into the polypeptide in response to UAG, or *Mj* ***A***-RS_{AGG}/$tRNA_{CCU}$ can not decode UAG as AzF. Further, the Mass Spec results also indicated the correct replacement of the 7th arginine by AzF and the 38th lysine by AlK via calculating molecular weights. We did not find peaks corresponding to the switched ***UAA***s replacements. Similar results were also observed when we

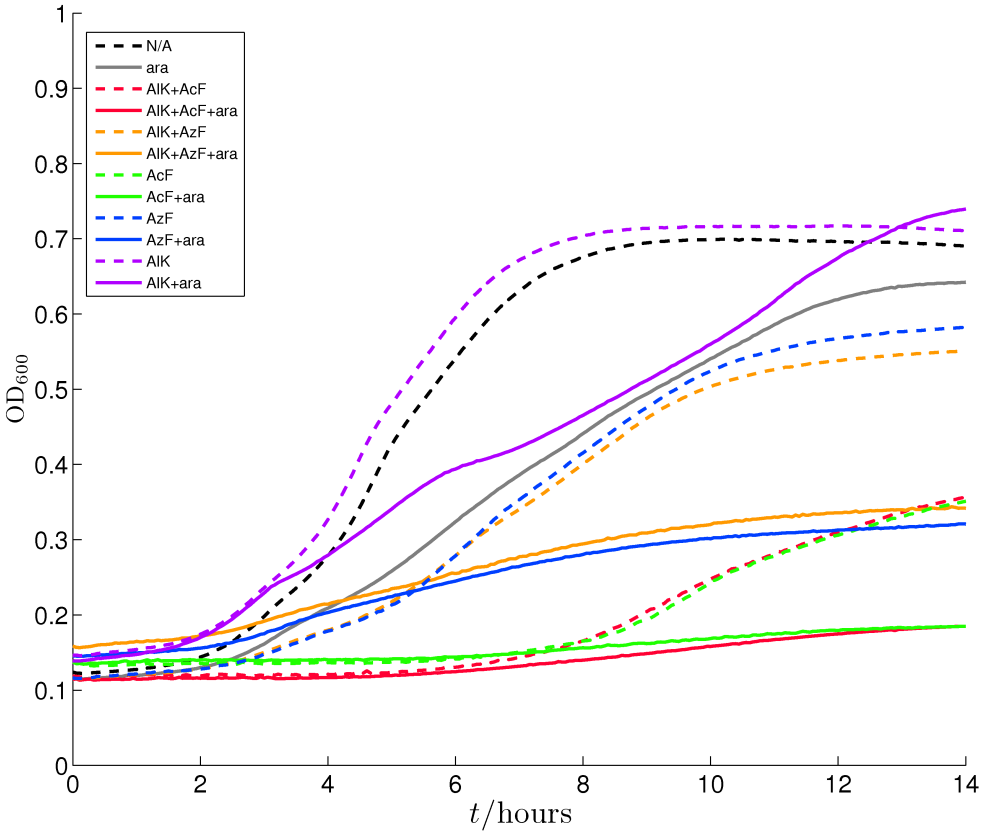


Figure 4.16: The toxicity caused by global AGG+UAG suppression during pAK8+pt2Z238/DBL21(DE3) cell growth. 1.5mM different kind of ***UAA*** pairs and 0.02% arabinose were added at the beginning of cell growth.

changed ***UAA*** types, or altered amino acid positions. Thus, we believe that these two exogenous aaRS/tRNA pairs are orthogonal to each other and are good candidates for two-codon recoding.

Double *UAA*s incorporation efficiency is determined by the recoding of AGG

The incorporation efficiency of two ***UAA***s is majorly determined by the recoding of AGG. Theoretically, the determinant is the amount of two types of ***UAA***-charged tRNAs inside bacteria, which compete with the canonical amino acid-charged endogenous tRNAs for the two codons, AGG and UAG. But actually in Mass Spec

results we only found peaks with single AlK insertion in response of UAG, or with both AzF and AlK double insertion in response of AGG and UAG, separately . That is because the competitor of *Mm*PylRS$_{\mathrm{UAG}}$/tRNA is release factor whereas endogenous ArgRS/tRNA rivals with *Mj* ***UAA***-RS/tRNA for AGG. In other words, UAG strictly constrains the incorporated amino acid as the desired ***UAA***, but AGG allows the insertion of both ***UAA*** and arginine. Thus, we used a strong promoter pBAD to increase the copy number of *Mj* ***UAA***-RS in order to charge more ***UAA***s into *Mj*tRNA$_{\mathrm{CCU}}$. Also, the ***UAA*** working concentration for AGG recoding was increased to 3mM, according to our previous results (Chapter 2). In fact, results showed that 7th arginine charged Z domian proteins were limited to 10-20% out of the total proteins by using this strategy.

The yield of double *UAA*s incorporated proteins is decided by two orthogonal aaRS/tRNA pairs and the type of *UAA*s

Who can win the contest between exogenous ***UAA***-tRNAs and endogenous aa-tRNAs is determined by their amounts inside cells. In turn, the cytoplasmic ***UAA***-tRNA level is decided by the mutual recognition between each orthogonal aaRS and its tRNAs, and the ability of the aaRS to charge ***UAA*** into its tRNAs. The *Mj* pCNF-RS/tRNA we chose to recode AGG has demonstrated good abilities to insert many types of ***UAA***s, such as IF, AzF, OMeY, and AcF, to tRNAs under the optimized conditions (Chapter 2 and [30]). On the other hand, the competition between PyltRNA$_{\mathrm{CUA}}$ and the endogenous release factors limits the total protein yield since UAG can be a sign of either stop or target ***UAA***. Besides, our experience indicates that the recoding ability of *Mm* PylRS/tRNA largely depends on the types of pyrrolysine analogues. BocK and AlK were two ***UAA***s we used for *Mm* PylRS/tRNA. Generally speaking, BocK returns a good protein yield (>10mg/L), no matter whether it was combined with IF, AzF, or other substrates of *Mj* pCNF-RS/tRNA; but the protein yield with AlK is usually quite limited (<5mg/L), even if combined with the same substrates of *Mj* pCNF-RS/tRNA. This is consistent with previous studies[41]. Also, BocK was commercially purchased while AlK was synthesized in our own lab, so the quality of ***UAA***s might also contribute to the difference.

Bring more properties to proteins

The purpose of recoding both AGG and UAG at the same time to different ***UAA***s is to bring more properties to proteins. Our results confirmed that. For example, the alykne group of AlK can conjugate with azide group of AzF via CuAAC reaction; meanwhile, the AcF can be labeled by certain aliphatic amine-reactive fluorescent dye, such as DNSH. Thus we generated special Z domain proteins whose AlK are connected with PEG(with azido group) and AcF are labeled with DNSH. These dual

function Z domain proteins are protected by PEG groups and are detectable under UV light by emitting fluorescence. Further, theoretically internal cyclization can be formed by simultaneously adding both AzF and AlK into one single protein, if the positions are correctly defined[44].

Also, recent reports showed that pylRS can be modified to host some ***UAA***s other than pyrrolysine analogues, such as N-o-Azidobenzyloxycarbonyl[36], OMeY[43]. The flexibility of pylRS provides many potential combinations of ***UAA***s by using our approach since pCNF-RS itself takes more than ten ***UAA***s as its substrate. Even more exciting, the mutual orthogonality of two aaRS/tRNA can be achieved by editing existed *Mj* ***UAA***-RS since ***Mj***TyrRS is highly tolerant for the modifications of its amino acid binding pockets. Therefore we can expect generating proteins with more dual functions. For instance, if nitrotyrosine and AcF are incorporated into a self proteins, one may observe how nitrotyrosine improve the immunogenicity of the self protein (such as its location inside or outside of cells) after labeling AcF with bright fluorescent dyes.

4.6 Conclusion

The successful reassignment of AGG and UAG at the same time was based on the collaboration of two exogenous RS/tRNA pairs, *Mj* ***UAA***-$\mathrm{RS_{AGG}}$/*Mj*$\mathrm{tRNA_{CCU}}$ and *Mm*Pyl-$\mathrm{RS_{UAG}}$/*Mm*$\mathrm{tRNA_{CUA}}$. These two pairs evolved from different bacterial species, and the results also proved that they did not react with each other. Therefore, further modifications on the amino acid hosting pockets of these synthetases will make it possible to simultaneously incorporate many different combinations of ***UAA*** pairs into single protein molecules.

Chapter 5

Other Endeavors to Expand the Genetic Code

5.1 Recoding AGA and UAG at the same time

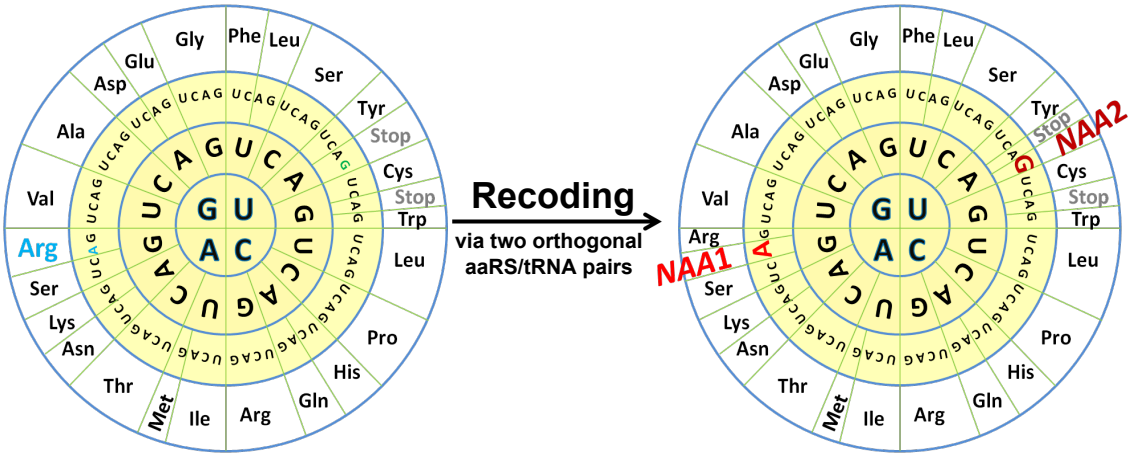


Figure 5.1: The recoding of AGA and UAG. ***NAA***, noncanonical amino acids=unnatural amino acids.

Two pairs of orthogonal *UAA*-RS/tRNA pairs were used to reassign AGA and UAG simultaneously

Our previous researches showed that we have successfully reassigned AGA codon from arginine to ***UAA***s in *E. coli* with the *Mj* ***UAA***-RS_{AGA}/*Mj*$tRNA_{UCU}$ pair which was evolved from *Mj*TyrRS/*Mj*$tRNA^{Tyr}_{CUA}$ [chapterAGA] with high ***UAA***s

incorporation rates of over 90% and a good ***UAA***-decorated protein yield of 20 mg/L . Also, the potential mutual orthogonality between *Mj* ***UAA***-RS_{AGA}/*Mj*$tRNA_{UCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$, into *E. coli* contributed to the feasibility of this project. They were derived from different microorganisms and varied significantly in conformation and function[3, 38, 41, 79]. Therefore, we tried to simultaneously reassign AGA and UAG codons to different types of ***UAA*** in individual proteins by introducing two RS/tRNA pairs, *Mj* ***UAA***-RS_{AGA}/*Mj*$tRNA_{UCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$, into *E. coli*.

Methods

Construct of pAK8 or pAK10 vectors to carry two synthetases

pAK8 vectors carry two different RS_{UAG} genes: pCNF-RS_{AGA} is controlled by inducible araBAD promoter while pyl-RS_{TAG} is after the constitutive glnS' promoter.

pAK10 vectors also carry two different RS_{UAG} genes (pCNF-RS_{AGA} and pyl-RS_{TAG}) , the araC repressor gene a p15A origin of replication, and the kanamycin resistant gene marker. Different from pAK8, both pCNF-RS_{AGA} and pyl-RS_{TAG} are controlled by inducible araBAD promoter. For the construction of pAK10 vectors, the cutting sites of restriction enzymes SacI and NedI were utilized to remove the old glnS' promoter and ligate the pBAD promoter to the cut pAK8 vectors by the same restriction enzymes.

Constructing pt2Z2 and pt2Z238 to carry tRNAs and the mutant Z domain gene

A restriction enzyme cutting site, either PstI or NgoMIV, was placed at each end of the dual-tRNA sequence lpp-pylT_{CUA}-rrnC-lpp-JYTRN_{UCU}-rrnC. These two enzymes were utilized to replace the single tRNA sequencelpp-JYTRN_{UCU}-rrnC in pAGG3-Zagg vector. Both pt2Z2GX and pt2Z238GX carried lpp-pylT_{CUA}-rrnC-lpp-JYTRN_{UCU}-rrnC. The whole sequence lpp-pylT_{CUA}-rrnC-lpp-JYTRN_{UCU}-rrnC was constructed by overlapping PCR with primers TF1, TR1, TF2, TF3, TR2, TF4. The template used was pAGA3-Zaga. NgoMIV and PstI were utilized for double digestion and ligation to put the dual-tRNA sequence into vectors.

The sequence of *Mm* pylT $tRNA_{CUA}$ is 5'-GGAAACCUGAUCAUGUAGAUCG AAUGGACUCUAAAUCCGUUCAGCCGGGUUAGAUUCCCGGGGUUUCCGCC A-3'.

To construct pt2Z2 vector, two primers Z-7aga9tag-F and Z-7aga9tag-R were used to change the codon for 7th amino acid lysine to AGA and the codon for 9th amino acid asparagine to amber stop codon TAG. BglII and SpeI, these two restriction enzymes were used for cutting off the old DNA sequence and ligate the target DNA.

T4PNK treated primers mix (F-Z7aga9tag and R-Z7aga9tag) were diluted 100 times and then ligated into the template to form he pt2Z2 vector.

In pt2Z238 vector(Fig. 5.2) , there are two mutations in Z domain gene: the codon for 7th amino acid lysine was altered to AGA, and the codon for 38th amino acid lysine was changed to TAG stop codon. While performing overlap PCR, primers Z7aga9tag-F, Z38tag-R, Z38tag-F, and Z-Blp-R were used to generate new DNA with 7AGA and 38TAG mutations by using pAGA3-Zaga as templates, and then BlpI and BglII were used for double digestion and ligation.

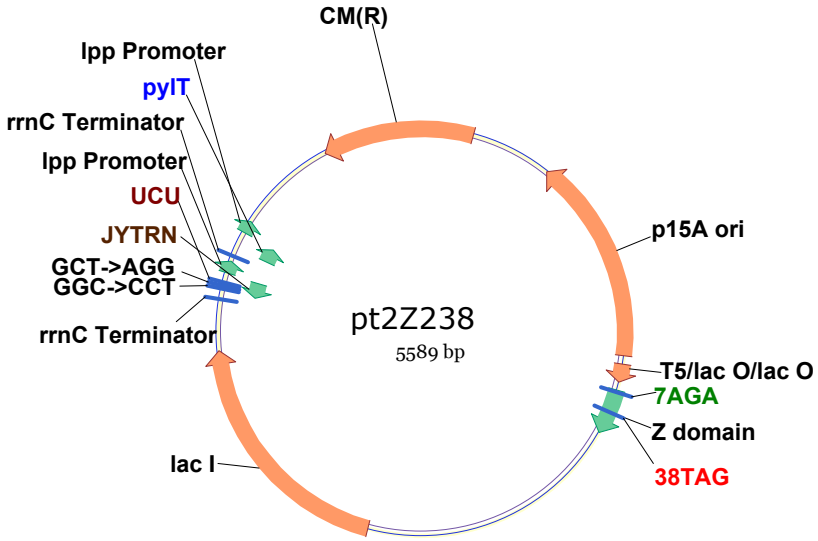


Figure 5.2: pt2Z238 vector

Expressing and purifying mutant Z domain proteins decorated by two unnatural amino acids

pAK8 or pAK10 vectors were co-transformed with either pt2Z238 or pt2Z2 into GM10 *E. coli*. Protein expression protocol was the same as previous described. IF(1.5mM) and BocK(1.5mM) were tested on pAK8+pt2Z2/ GM10, with different concentrations of initial arabinose added , 0%, 0.0002%, and 0.002%. Also,five pairs of different unnatural amino acids were tested on pAK8+pt2Z238/GM10: IF(1.5mM) and AlK(4.5mM), AzF(1.5mM) and AlK(4.5mM), IF(1.5mM) and BocK(1.5mM), AzF(1.5mM) and BocK(1.5mM), and OMeY(1.5mM) and BocK(1.5mM). One pair of unnatural amino acids, OMeY(1.5mM) and BocK(1.5mM), were tested on pAK10+ pt2Z238/GM10. All bacterial cultures were maintained at 37°C, 250rpm. When the culture reach OD600=0.6-0.8, 0.02% arabinose was added for induction, and followed by 1mM IPTG induction 10 minutes later. After expression for 14-16 h at 37 °C,

20 ml cultures were pelleted at 8,000 rpm and frozen at -80°C. Protein purification methods are the same as mentioned in previous chapters.

Protein electrophoresis

Methods are the same as stated in previous chapters (AGA/AGG)

Analyzing protein molecule weight by mass spectrometry

Protein molecular weight was analyzed by mass spectrometry. Protein samples which showed obvious bands with correct sizes on SDS-PAGE gel were sent out for mass spectrometry (ESI or MALDI-TOF) in 70% ACN.

Results and Discussion

IF and BocK were incorporated into Z domain in pAK8+pt2Z2/GM10

The proteins purified from pAK8+pt2Z2/GM10 include six different treatment: normal induction but no ***UAA***s added; normal induction with IF and BocK; normal induction with IF+BocK and 0.0002% initial arabinose added; normal induction with IF+BocK and 0.002% initial arabinose added; IPTG only induction with BocK added; IPTG only induction with IF+BocK added.

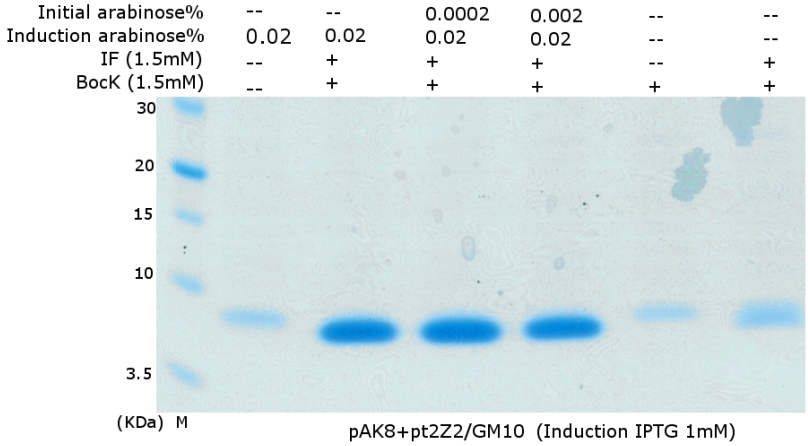


Figure 5.3: SDS-PAGE gel: Proteins purified from pAK8+pt2Z2/GM10 after protein expressed with IF, or BocK, or both, and different amount of arabinose.

On SDS-PAGE gel, the normal induction ones with IF+BocK, no matter if the initial arabinose added or not, all showed a thick band on the estimated position at 8KDa, which indicating a good protein yield around 10-20mg/mL (Fig. 5.3). However,

on the first lane (normal induction without ***UAA***s), there was a very light but still visible band at 8KDa, which was not expected because normally UAG as a stop codon should be a strict signal to halt translation if BocK or other substrates were not presented in the culture media. For the last two lanes to the right, there were two light bands, showing the low level expression of mutant Z domain proteins. Those proteins might be a few amount of Z domain with arginine charged at 7th position by endogenous translational machine since the *Mj*pCNF-RS was not expressed since the arabinose was not added to activate pBAD promoter. The low expression level might reflect the inefficiency of recoding UAG to BocK or other substrates by pyl-RS_{UAG}/$tRNA_{CUA}$. It might also be the result of the low level of functional endogenous $tRNA_{UCU}$ in GM10 bacteria.

ESI-MS results suggested that the IF and BocK were incorporated into Z domain proteins in response to AGA and UAG, separately. The sample of normal induction with IF+BocK (YW66) showed three peaks at m/z= 8063 (matching Z-7IF-9BocK, lost 1st Met and with one acetylation), 8020.9 (matching Z-7IF-9BocK, lost 1st Met), 7962.5 (matching Z-7IF-9aa, lost 1st Met and with one acetylation). Also, other two samples– normal induction with IF+BocK and 0.0002% initial arabinose (YW67) and normal induction with IF+BocK and 0.002% initial arabinose (YW68)– both displayed similar pattern of spectra that two peaks matched Z-7IF-9BocK and one peak matched Z-7IF-9aa. The IF+BocK incorporation rates of YW66, YW67 and YW68 were 88%, 91%, and 89%, which indicating that it was not very useful to turn on pCNF-RS_{AGA} gene expression at the beginning of bacterial culturing.

In ESI results there is a puzzle of the translation of the 9th UAG into certain type of amino acid (9aa) among these three samples, even only about 10% out of the total. The molecular weight of this 9aa matched that of lysine, so this 9th aa might be lysine and was disassociated from BocK during ESI. However BocK disassociation theory can not explain the band of sample 'normal induction without ***UAA***s' on SDS-PAGE gels.

***UAA*s were incorporated into Z domain in pAK8+pt2Z238/GM10**

The protein electrophoresis results on SDS-PAGE gels proved that all ***UAA*** pairs incorporated Z domain proteins showed a band close to 8KDa. If BocK was used to suppress UAG stop codon, then the band would be thick/fat, indicating a good protein yield. On the other hand, if AlK, which was synthesized in our lab, was used then the protein yield would be very low, indicated by a thin protein band. (Fig. 5.4)

The ESI/MALDI-TOF results of different ***UAA***s incorporated Z domain proteins purified from pAK8+pt2Z238/GM10 are all listed in the following table 'Mass Spectrometry results of Z domain proteins purified from pAK8+pt2Z238/GM10' (Table 5.1). From the table we can clearly see peaks matching the molecular weights of two ***UAA***s decorated Z domain proteins in all samples. As expected, there were

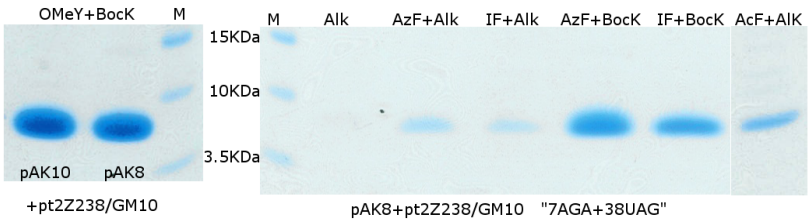


Figure 5.4: SDS-PAGE electrophoresis of different amount of proteins, which were purified from pAK8+pt2Z2/GM10 after protein expressed with an AGA-suppressing *UAA* and AlK, or BocK.

some peaks corresponding to the proteins which had arginine as the 7th amino acid since pCNF-RS_{AGA}/$tRNA_{UCU}$ can hardly completely compete out the endogenous RS_{AGA}/$tRNA_{UCU}$. But the mysterious 38th amino acid still existed. In every sample the peaks matching Z-7***UAA***-38aa were detected, and the molecular weight of the 38th amino acid matched lysine, which was almost impossible since UAG is a stop codon.

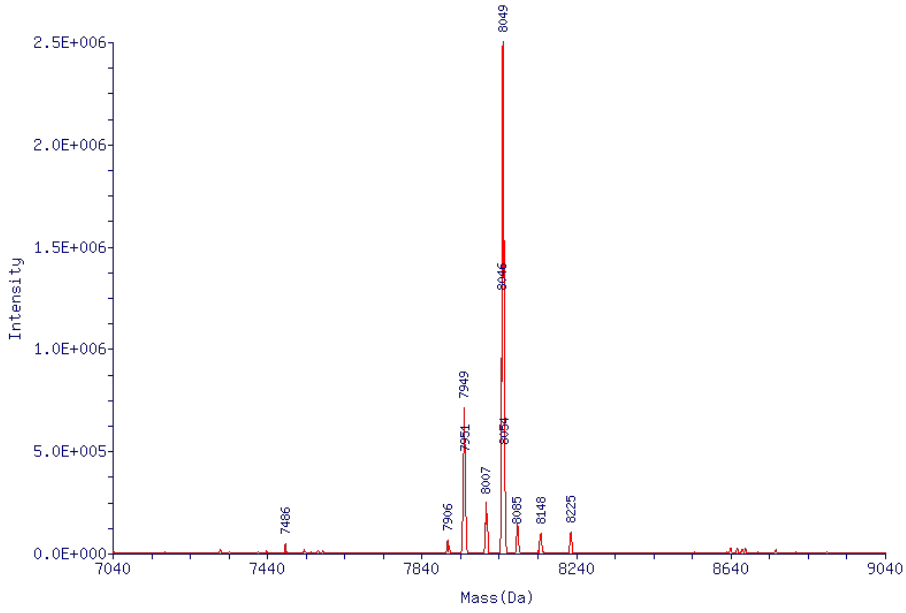


Figure 5.5: ESI-MS result of YW90: pAK8+pt2Z238/GM10 with IF+BocK.

We also measured the relative double-***UAA*** incorporation rates of these samples by the equation of two *UAA*s decorated proteins (Z-***UAA***1-*UAA*2)/total proteins (Z-

UAA1-*UAA*2 + all other forms of Z domain proteins). Generally speaking, tests with BocK as the second type *UAA* gave better double-***UAA*** incorporation rates (YW81, YW89, YW90(Fig. 5.5)) than those with AlK (YW88(Fig. 5.6), YW87, YW1M). This was consistent with previous results since pylRS showed different abilities to charge various substrates. Also the quality of ***UAA***s should also be counted.

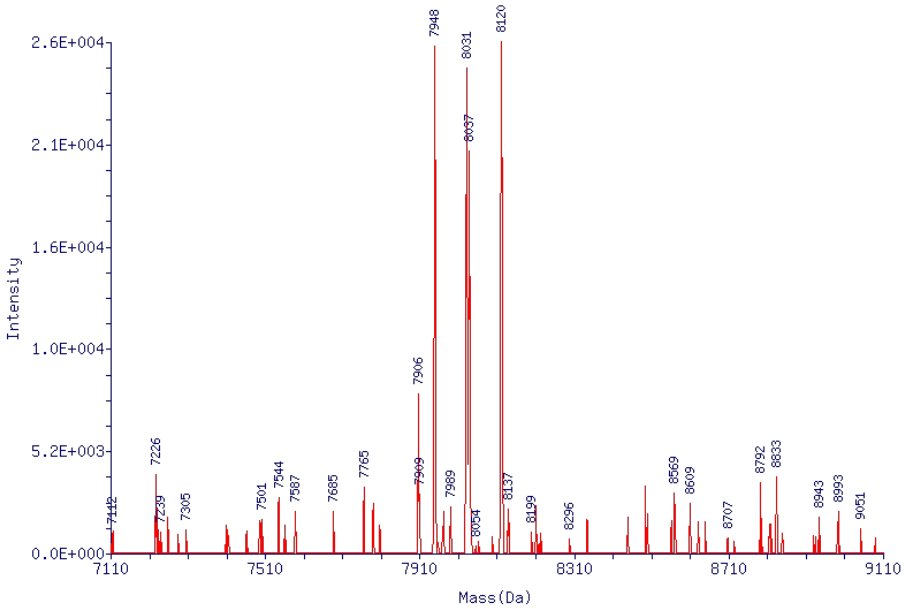


Figure 5.6: ESI-MS result of YW88: pAK8+pt2Z238/GM10 with IF+AlK.

pAK10 demonstrated similar ability to recode AGA and UAG as pAK8

OMeY and BocK were incorporated into Z domain proteins by both pAK8 and pAK10 with pt2Z238 in GM10. Proteins purified from them both showed thick bands at 8KDa on SDS-PAGE gels, and their ESI results returned very similar spectra. Both had two major peaks very close to m/z=7953, 7911 matching Z-7OMeY-38BocK(-1Met+Ac) and Z-7OMeY-38BocK(-1Met); another two peaks around m/z=7854, 7811 matching Z-7OMeY-38K/Q(-1Met+Ac) and Z-7OMeY-38K/Q(-1Met); and two tiny peaks near m/z=7932, 7890 to match Z-7R-38BocK(-1Met+Ac) and Z-7R-38BocK(-1Met). The difference of their double-***UAA*** incorporates is negligible: the one using pAK10 (YW82) (Fig. 5.7) was 67.7% where the one with pAK8 (YW81) was 64.7%.

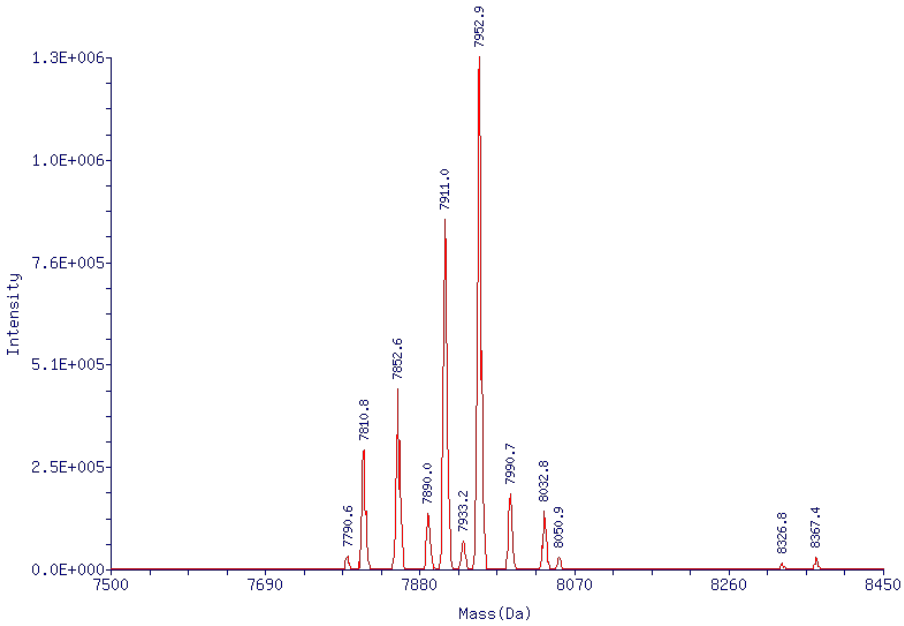


Figure 5.7: OMeY and BocK were incorporated into Z domain proteins by pAK10 with pt2Z238 in GM10. (YW82, ESI-MS).

GM10 bacteria are NOT good candidates to recode UAG

Each type of the following unnatural amino acids – AlK, BocK, AzF, AcF – was added into pAK8+pt2Z238/GM10, separately, for protein expression. Protein purified from them were used as four controls: AlK-only, BocK-only, AzF-only, and AcF-only. On SDS-PAGE gels, three(BocK-only, AzF-only, AcF-only) of the them showed protein bands at 8KDa. It is reasonable for the BocK-only sample has translational products since BocK can be used for UAG suppression. The poor quality of AlK might be responsible for the poor protein yield in the AlK-only sample. But it is profound to see bands on AcF-only and AzF-only lanes. These two unnatural amino acids were used to suppress AGA codon and it remained unclear how the UAG amber stop codon was suppressed during translation. So the AcF-only sample was sent for MALDI-TOF analysis(Fig. 5.8)

Proteins purified from pAK8+pt2Z238/GM10 with AcF only during protein expression were sent out for MALDI-TOF analysis. The MALDI-TOF result of AcF-only sample (YW2M) (Fig. 5.9) directly proved that an amino acid with molecule weight close to 146 was charged into Z domain protein at position 38. On the other hand, the molecular mass of lysine is 146.19. Three peaks at m/z=7855.8, 7865.6,

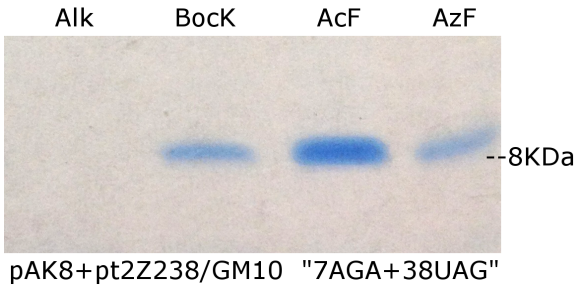


Figure 5.8: Four controls:AlK-only, BocK-only, AzF-only, and AcF-only. For each control, only one ***UAA*** was added into pAK8+pt2Z2/GM10 for protein expressed. SDS-PAGE gel.

7823.4 were found to match three forms of Z-7AcF-38K, full length, without 1st methionine and with acetylation, without 1st methionine. Also, two minor peaks m/z=7920.6, 7790.6 were also found to match Z-7R-38K. Previously we tried to explain the appearance of 38th lysine by the disassociation of BocK to lysine during mass spectrometry. But in this AcF-only control test, no BocK was added. So BocK disassociation theory can not explain this mystery.

After a close investigation, we find out that GM1O *E. coli* carries a glnV44(AS) (http://cgsc.biology.yale.edu/Strain.php?ID=10703) mutation, which is a natural suppression of amber (UAG) stop codons by insertion of glutamine. The molecular weight of glutamine is 146.14 while that of lysine is 146.19. Therefore mass spectrometry can not distinguish these two amino acids and the 38th amino acid could be either lysine(K) disassociated from BocK if it was used), or more likely glutamine(Q) inserted by the mutated endogenous GM10 translational system. So GM10 is not a good candidate for UAG reassignment.

Conclusion

The AGA and UAG codons were reassigned to different pairs of ***UAA***s in Z domain proteins by using two RS/tRNA pairs, *Mj* ***UAA***-RS_{AGA}/*Mj*$tRNA_{UCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$ in GM10 *E. coli*. However, the unnatural amino acids incorporation rates were not as high as that of AGG and UAG, because the amber UAG codon was naturally suppressed to glutamine in GM10 *E. coli* by its glnV44(AS) mutation. Other bacteria strains, such as BL21(DE3) can be further tested for this target with increased ***UAA*** working concentrations. Another strategy to co-recode AGA codon is to pair it with some other codons, such as UAA stop codon.

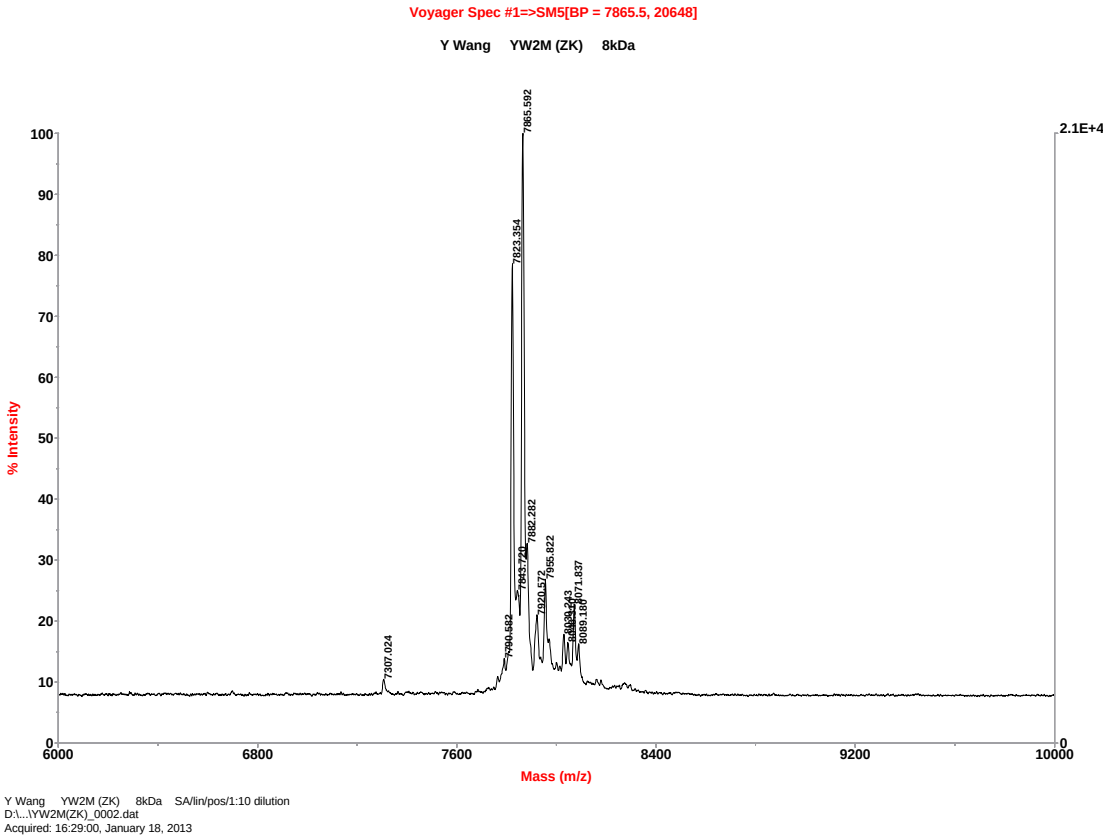


Figure 5.9: The MALDI-TOF result of AcF-only sample (YW2M).

Table 5.1: Mass Spectrometry of Z domain proteins purified from pAK8+ pt2Z238/GM10.

Table: Mass Spectrometry results of Z domain proteins purified from pAK8+pt2Z238/GM10

Sample#	***UAA*** s added	Z domain proteins (7AGA+38UAG)	*ex* MW	*ob* MW	2-***UAA*** IR
YW81 (ESI)	OMeY+BocK	Z-7OMeY-38BocK (-1Met+Ac)	7954.0	7952.9	67.4%
		Z-7OMeY-38BocK (-1Met)	7911.0	7910.9	
		Z-7OMeY-38K/Q (-1Met+Ac)	7854.0	7952.7	
		Z-7OMeY-38K/Q (-1Met)	7811.0	7811.0	
		Z-7R-38BocK (-1Met+Ac)	7932.4	7933.1	
		Z-7R-38BocK (-1Met)	7890.4	7890.6	
YW89 (ESI)	AzF+BocK	Z-7AzF-38BocK (-1Met+Ac)	7964.4	7964.0	65.2%
		Z-7AzF-38BocK (-1Met)	7922.4	7922.0	
		Z-7AzF-38K/Q (-1Met+Ac)	7864.4	7864.0	
		Z-7R-38BocK (-1Met+Ac)	7932.4	7932.0	
		Z-7R-38BocK (-1Met)	7890.4	7890.0	
YW90 (ESI)	IF+BocK	Z-7IF-38BocK (-1Met+Ac)	8049.3	8049.0	76.9%
		Z-7IF-38BocK (-1Met)	8007.3	8007.0	
		Z-7IF-38K/Q (-1Met+Ac)	7949.3	7949.0	
		Z-7IF-38K/Q (-1Me)	7907.3	7906.0	
YW88 (ESI)	IF+Alk	Z-7IF-38Alk (full)	8120.5	8120.0	49.0%
		Z-7IF-38Alk (-1Met+Ac)	8031.3	8031.0	
		Z-7IF-38Alk (-1Met)	7989.3	7989.0	
		Z-7IF-38K/Q (full)	8038.5	8037.0	
		Z-7IF-38K/Q (-1Met+Ac)	7949.3	7948.0	
		Z-7IF-38K/Q (-1Met)	7907.3	7906.0	
YW87 (ESI)	AzF+Alk	Z-7AzF-38Alk (full)	8035.6	8036.0	30.0%
		Z-7AzF-38Alk (-1Met+Ac)	7946.4	7946.0	
		Z-7AzF-38K/Q (full)	7953.6	7953.0	
		Z-7AzF-38K/Q (-1Met+Ac)	7864.4	7864.0	
		Z-7AzF-38K/Q (-1Met)	7822.4	7822.0	
		Z-7R-38K/Q (-1Met)	7790.4	7788.0	
YW1M (MALDI-TOF)	AcF+Alk	Z-7AcF-38Alk (full)	8037.6	8036.6	23.0%
		Z-7AcF-38Alk (-1Met)	7905.4	7906.5	
		Z-7AcF-38K/Q (full)	7954.6	7954.1	
		Z-7AcF-38K/Q (-1Met+Ac)	7865.4	7865.0	
		Z-7AcF-38K/Q (-1Met)	7823.4	7823.1	
		Z-7R-38K/Q (-1Met)	7790.4	7790.2	
YW2M (MALDI-TOF)	AcF only	Z-7AcF-38K/Q (full)	7954.6	7855.8	89.89%*
		Z-7AcF-38K/Q (-1Met+Ac)	7865.4	7865.6	
		Z-7AcF-38K/Q (-1Met)	7823.4	7823.4	
		Z-7R-38K/Q (full)	7921.6	7920.6	
		Z-7R-38K/Q (-1Met)	7790.4	7790.6	

ex MW = expected molecular weight *ob* MW = observed molecular weight

UAAs = unnatural amino acids IR=incorporation rate

*, the incorporation rate was only calculated based on the single unnatural amino acid AcF.

-1Met = lost the first methionine; +Ac= with acetylation; full= full length

5.2 Using modified *Mm*PylRS$_{UAG}$/tRNA to reassign sense codons

Previous reports showed that mutating the anticodon of *Mm*tRNA$^{Pyl}_{CUA}$ did not significantly affect its interaction with the catalytic domain of PylRS [41]. It's suggested that pyrrolysine is not hardwired for cotranslational insertion at UAG codons, and tRNA^{Pyl} variants may decode numerous codons[39]. So we are wondering if the altering the anticodon of tRNA^{Pyl} to UCU or CCU can make this *Mm*PylRS/tRNA pair recode AGA or AGG as well.

Methods

Constructing pAK9 vector to carry two copies of pyl-RS$_{UAG}$ gene

pAK9 vectors carry two copies of pyl-RS$_{UAG}$ gene with two promoters (a modified glnS' (constitutive) promoter and an araBAD (arabinose-inducible) promoter), the araC repressor gene a p15A origin of replication, and the kanamycin resistant gene marker. The synthetase gene after pBAD promoter carries a restriction site (BglII or SacI) at each end, whereas one restriction enzyme (PstI or NedI) cutting site was placed at each end of the synthetase gene after glnS' promoter. These restriction enzyme cutting sites were utilized for the construction of pAK9 vector by using pAK3 as the templates. '

Constructing pAGA9-Zaga and pAGG9-Zagg vectors

The pAGA9-Zaga (Fig. 5.10) vector carries two tRNA$_{UCU}$ genes: *Mj*tyr-tRNA$_{UCU}$ and *Mm*pyl-tRNA$_{UCU}$. Overlap PCRs were used to generate DNA fragment which mutated the anticodon gene of *Mm*pyl-tRNA from CTA to TCT by using the pt2Z2 as the template. The DNA fragment was ligated to pAGA3-Zaga after both of them were double digested by NgoMIV and pstI.

The pAGG9-Zagg vector carries two tRNA$_{UCC}$ genes: *Mj*tyr-tRNA$_{UCC}$ and *Mm*pyl-tRNA$_{UCC}$. Overlap PCRs were used to generate DNA fragment which mutated the anticodon gene of *Mm*pyl-tRNA from CTA to TCC by using the pt2Z2GX as the template. The DNA fragment was ligated to pAGG3-Zagg after both of them were double digested by NgoMIV and pstI.

Recoding AGA or AGG by PylRS/tRNA

The pAK9 vectors were co-transformed with pAGA9-Zaga into GM10 bacteria, or with pAGG9-Zagg into DH10BW. And then single colonies were selected for protein expression. A single colony from each co-transformation was grown overnight at 37

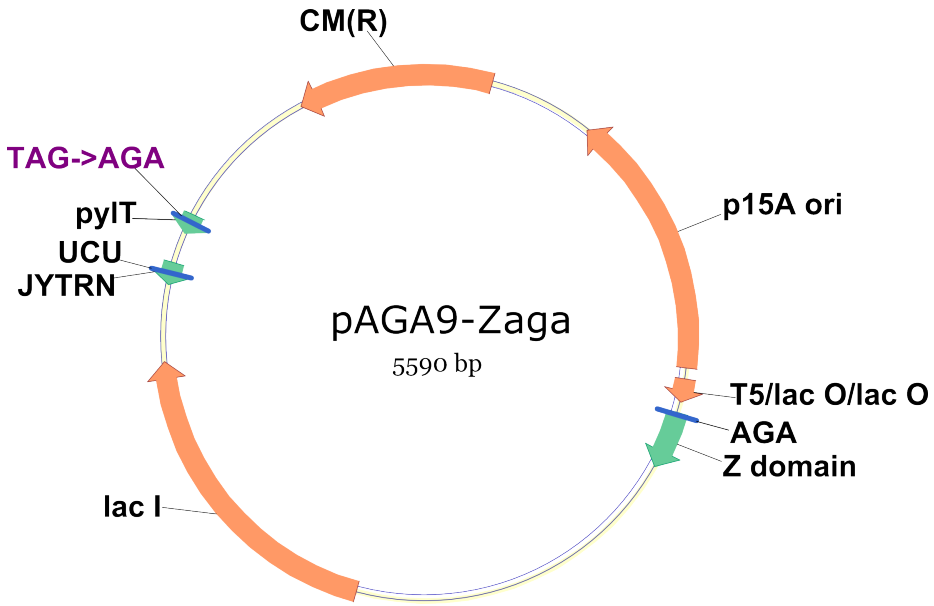


Figure 5.10: pAGA9-Zaga vector.

°C in 2 ml of LB media supplemented with kanamycin and chloramphenicol to obtain the starter culture. The starter culture was then twenty times diluted to a 20 mL LB medium containing 50 μg/ml kanamycin, 34μg/mL chloramphenicol and two types of unnatural amino acids for further culturing. 3mM BocK was used for pAK9+pAGG9-Zagg/DH10BW, and 1.5mM BocK was added for pAK9+pAGA9-Zaga/GM10 . Till the culture reach OD600=0.6-0.8, 0.02 percent arabinose was used for induction, and followed by 1mM IPTG induction 10 minutes later. After expression for 14-16 h at 37 °C, 20 ml cultures were pelleted at 8,000 rpm and frozen at -80°C. Protein purification methods are the same as mentioned in previous chapters. Purified proteins first underwent SDS-PAGE electrophoresis and then suitable samples were sent to ESI-MS at UC Berkeley.

Results: *Mm*PylRS$_{UAG}$/*Mm*tRNA$^{Pyl}_{AGA}$ can recode AGA codon, but not very efficiently.

On SDS-PAGE gel, the protein amount of Z domains purified from pAK9+pAGG9-Zagg/DH10BW was very little, so it was not sent out for ESI-MS.

But the pAK9+pAGA9-Zaga/GM10 was still able to yield regular amount of Z domain proteins (above 10mg/mL) and appeared as a thick band on SDS-PAGE gel near 8 KDa. Then it was sent for ESI-MS. The ESI result (YW86) (Fig. 5.11) , the deconvoluted **ESI-MS** spectrum reveals two peaks at m/z =7862, and 7904 which

match the expected molecular weight for the BocK incorporated Z-domain mutant with the loss of its first methionine (Z-7BocK-Met, m/z =7862) and its subsequent post-translational acetylation product (Z-7BocK-Met+Ac, m/z =7904. There is also a minor peak at m/z =7789, matching the arginine incorporated Z-domain protein with the loss of its first methionine (Z-7Arg-Met, m/z=7790.4). The BocK incorporation rate from this test is about 51.7%.

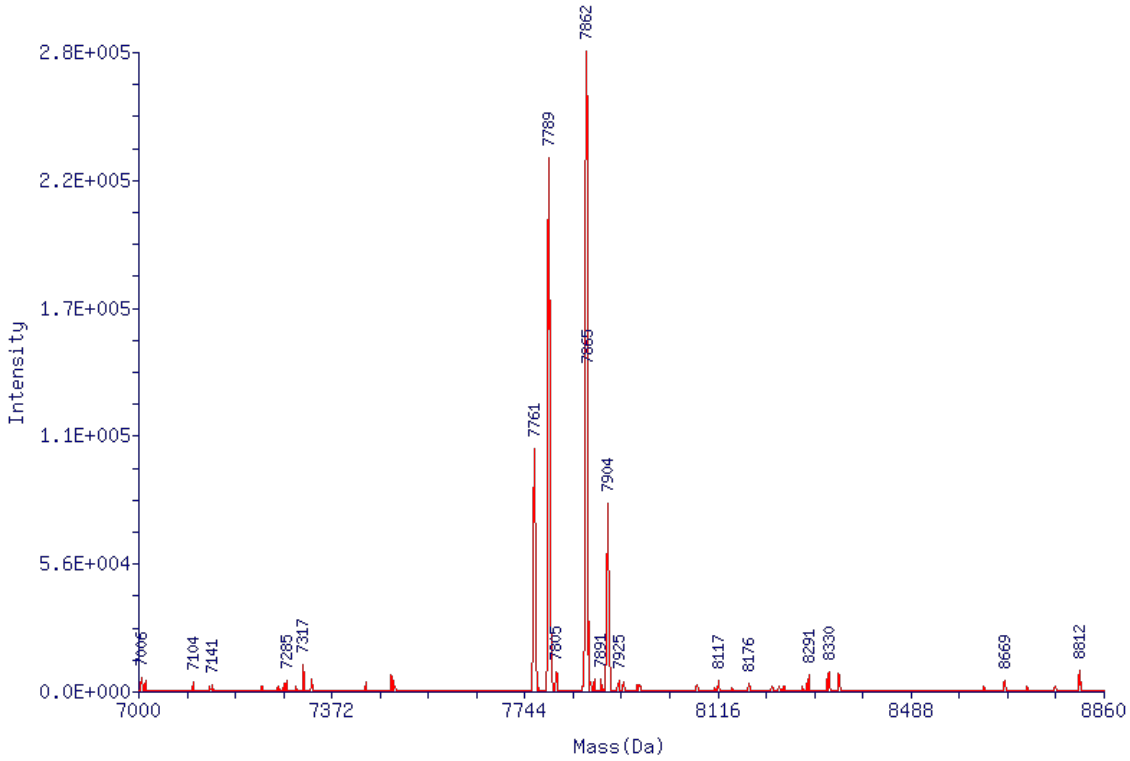


Figure 5.11: ESI-MS result of Z-9BocK from pAK9+pAGA9-Zaga/GM10 (YW86).

Conclusion

Therefore, the result supported our prediction that the pyl-RS can reassign rare codons, such as AGA, in addition to its original target UAG amber stop codon. However, more works need to be done to improve the ***UAA***s incorporation rate,for example, enhancing the competition ability of the endogenous MmPylRS$_{\mathrm{UAG}}$/MmtRNA$^{\mathrm{Pyl}}$ pair. Also, other bacterial strains, such as BL21(DE3), can be employed to recode

AGG codon by this synthetase/tRNA pari because in chapter AGG we proved that the protein yield of BL21(DE3) was four-five times higher than that of DH10BW.

5.3 Recoding multiple AGA codons in one protein

Since we can recode single AGA codon in *E.coli* with good ***UAA*** incorporation efficiency, can we recode multiple AGA codons as well?

Recoding multiple AGA codons in Z domain proteins

Methods

pAGA3-Zaga2 and pAGA3-Zaga3 vecotrs were built to carry Z domain genes with multiple AGA codons. In pAGA3-Zaga2, the codons for 6th and 7th amino acids were both replaced by AGA. In pAGA3-Zaga3, the codons for 6th, 7th, and 8th were all replaced by AGA. Two Primer pairs, Z-2AGA-f/Z-2AGA-r and Z-3AGA-F/Z-3AGA-r, were mixed well and treated by T4PNK. The primer mixtures were diluted 100 times and then ligated into the double digested pAGA3-Zaga vectors to form the pAGA3-Zaga2 or pAGA3-Zaga3 vectors. The restriction enzymes BglII and SpeI were utilized for the double digestion of vectors.

Then pAK3 plasmids were co-transformed with either pAGA3-Zaga2 or pAGA3-Zaga3 into GM10 bacteria. Protein expressions were performed with 1.5mM IF added. The protein expression and purification protocol were the same as described previous chapters. Purified proteins were analysed by SDS-PAGE electrophoresis and ESI-MS.

Results: It is difficult to recode continuous multiple AGA codons in Z domain protein.

SDS-PAGE electrophoresis results (Fig. 5.12) showed that with IF added, Z-6AGA-7AGA had a very low protein yield while without IF added, the protein band was visible. As for 3 continuous AGA codons presented in Z domain protein genes, no matter IF added or not, Z-6AGA-7AGA-8AGA could not be expressed successfully.

Repeating ESI-MS results of Z-6AGA-7AGA with IF added suggested that peaks corresponding to protein had extremely low intensity, so so spectrum deconvolution could not be done. Thus under current condition, it is very difficult to recode contiunous multiple AGA codons in Z domain proteins.

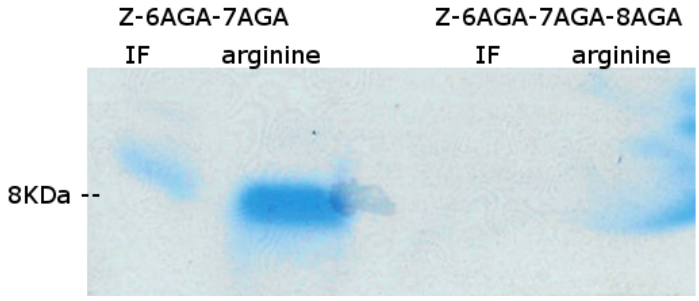


Figure 5.12: SDS-PAGE gel:trying to recode continuous multiple AGA codons in Z domain protein.

Recoding noncontinuous AGA codons in GCN4 proteins

Methods

GCN4 protein is a eukaryotic transcriptional activator protein, and binds to target DNA as a dimer. In WT GCN4 gene, five AGA codons are separately located within its DNA, and they are codons for arginines at position 12, 18, 23, 27 and 51. So we constructed pAGA3-GCN4 to carry the GCN4 gene and $Mj\text{tRNA}_{\text{AGA}}^{\boldsymbol{UAA}}$. The GCN4 gene was amplified by PCR by using GCN4-f1 and GFP06e as primers and pET-GCN4 as templates. The PCR products were double digested by BglII and SpeI and the ligated to double digested pAGA3-Zaga vectors.

Then pAK3 plasmids were co-transformed with pAGA3-GCN4 into GM10 or BL21(DE3) bacteria. Protein expressions were performed with or without 1.5mM IF added. The protein expression and purification protocol were the same as described previous chapters. Purified proteins were analysed by SDS-PAGE electrophoresis and ESI-MS.

Results: It is difficult to recode noncontinuous multiple AGA codons in GCN4 protein.

ESI results showed that samples purified from GM10 had no protein peaks detected. For samples purified from BL21(DE3), ESI results did not return any peaks corresponding to the molecular weights of GCN4, with either arginine charged or IF incorporated. So it is hard to recode noncontinuous multiple AGA codons in GCN4 proteins as well.

Discussion

The difficulty of recoding multiple AGA condons resembles a previous research which showed that translation of the tandem arginine codons AGG caused about half of the

ribosomes to shift to the +1 phase[70]. This tandem AGG translating error might be caused by $tRNA^{Arg}$ shortage. They suggest that ribosomes read the first arginine codon but pause when the second AGG codon appears again, and about 50% of the ribosomes escape by moving to the +1 reading mode while the rest can find a cognate tRNA and continue translating. Similarly, to continuously recodeanother rare codon AGA in *E. coli*, it is very possible that the amount of $tRNA^{\boldsymbol{UAA}}_{UCU}$ is even lower than that of $tRNA^{Arg}_{UCU}$.

The unsuccessful recoding multiple AGA codons at noncontinuous sites is similar as the difficulty of reassgining multiple UAG amber stop codons. The later is because the UAG suppressor tRNAs have to compete with endogenous release factors, which are used to recognize stop codons and signal for the termination of translation. The competition between suppressor tRNAs and release factors results in low ***UAA*** incorporation efficiencies when recoding UAG. Further, protein yields drop dramatically with the addition of even a second stop codon. Thus, it is also very possible that AGA suppressor tRNAs are unable to outcompete arginine $tRNA_{UCU}$; and/or extra copies of AGA exceed regular AGA occurrence in *E.coli* (0.24% out of entire bacterial genome), so there are not enough arginine $tRNA_{UCU}$ to allow the continuation of translation. To certain extent, too many extra copies of rare codons means 'stop' when expressing extended proteins.

Conclusion

Similar as the difficulty of reassgining multiple UAG amber stop codons, it is quite hard to recode multiple sense codons, such as AGA, at the same time without altering other parts of the translational matchine.

Chapter 6

Conclusion and Future Directions

6.1 The expansion of the genetic code: sense codons reassigned

To expand genetic codes, we reassigned two rare sense codons (AGG and AGA) by using engineered *Mj*TyrRS/tRNA pairs; and multiple unnatural amino acids were incorporated into one protein simultaneously in response to a sense codon and the amber codon UAG via the newly evolved ***UAA***-RS/tRNA pair and its orthogonal counterpart *Mm*PylRS/tRNA_{CUA}.

To accomplish this task, we completed the following specific aims.

Specific Aim 1. Reassigned AGG codon to encode for ***UAA***s in *E.coli*. For this purpose, a two-step selection scheme was performed to evolve *Mj*TyrRS/tRNA pairs. In the first step, AGG was recoded from arginine to tyrosine. To make this happen, the tRNA anticodon recognizing region of *Mj*TyrRS was altered to host CCU anticodon instead of the original GUA anticodon in the *Mj*-tRNA mutant. A mutant *Mj*TyrRS library was established, which contained random mutations at four residues in the anticodon recognition site; and the anticodon of *Mj*tRNA was changed to CCU. Meanwhile, a mutant GFP_{UV} was used as the reporter to detect a single amino acid mutation between Arg and Tyr. This mutant GFP_{UV}-Tyr66AGG will not show any green fluorescence if an arginine is presented at position 66. Only synthetases with compatible mutations can recognize CCU anticodon and charge the orthogonal tRNA with tyrosine but not arginine in response to 66AGG, and in turn produce functional GFP_{UV}. To screen out synthetases which are reacting to the target anticodon, the mutant *Mj*TyrRS library, *Mj*tRNA mutant, and the special GFP_{UV} reporter were all co-transformed into *E. coli*, and only colonies with intensive green fluorescence were selected for further aaRS modifications. In the second step, AGG was reassigned to ***UAA***s by genetically modifying the amino acid binding pocket of the mutated synthetases. Finally one modified orthogonal ***UAA***-$\text{RS/tRNA}_{\text{CCU}}$ pair was obtained.

To optimize expressing conditions of ***UAA*** decorated proteins, a dual expression system and a special *E. coli* strain DH10BW, which carries a mutant endogenous $tRNA_{UCU}^{Arg}$, were generated to defeat ArgRS/ $tRNA_{UCU}^{Arg}$ and to increase ***UAA*** incorporation efficiency without killing the host cells. When increase ***UAA*** concentration to 3mM, BL21(DE3) also demonstrated good ***UAA*** incorporation efficiencies.

Specific Aim 2. Reassigned AGA codon to encode for ***UAA***s in *E.coli*. We adopted a similar two-step selection scheme as that in the AGG recoding study. First, AGA was recoded from arginine to tyrosine. The anticodon of *Mj*tRNA was changed to CCU, and the tRNA anticodon recognizing region of *Mj*TyrRS was first changed to host UCU anticodon through the screening with a mutant GFP_{UV}-Tyr66AGA reporter. Second, AGA was decoded as ***UAA***s. In this step, the amino acid binding pocket of the mutated sythetases was modified to host ***UAA***s. Further, a dual expression system and a special bacterial strain GM10 *E. coli* were utilized to to suppress the endogenous ArgRS/$tRNA_{UCU}^{Arg}$ and to increase ***UAA***s incorporation efficiencies in response to AGA codon.

Specific Aim 3. Recoded multiple ***UAA***s by suppressing a sense codon (AGG or AGA) and UAG amber codon together. The evolved *Mj* ***UAA***-$RS_{AGG/AGA}$/ *Mj*$tRNA_{CCU/UCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$ were collaborated to introduce two different ***UAA***s into proteins in response to the designated sense codon and UAG correspondingly. PEGylation and fluorescent dye labeling were able to be performed on the same protein.

The recoding of sense codons, such as AGA and AGG, offers powerful tools to incorporate unnatural amino acids site specifically. The generalization of such strategy may be applied for the reassignment of other rare codons to provide many possible sources of unique blank codons. The enrichment of blank codons will make it possible to encode many more unnatural amino acid combinations. Since our reassignment of rare codons went through the creation of ambiguous state of the meaning of a codon, this could provide insights for the evolution process of the universal genetic code.

6.2 Future directions

Applications of multiple unnatural amino acids incorporation

The genetic incorporation of two different ***UAA***s into one protein has many potential applications. For instance, it may help to install a Frster resonant energy transfer (FRET) pair in a protein for conformation and dynamic studies. Still take a protein site-specifically decorated with both AcF and AlK at the same time as an example. One may use 5-(Aminomethyl)Fluorescein, Hydrochloride[86, 87], a fluorescent dye (em/ex, 497nm/516nm), to react with the keto group of AcF to form a Schiff base as DNSH did (Chapter 2,3,4). Meanwhile, another fluorophore, tetramethylrho-

damine(em/ex, 555nm/577nm) with azide is able to complete the 'click chemistry' with the AlK residue of the protein[88, 89] . These two fluorophores are a potential FRET pair since the excitation spectrum of Fluorescein overlaps with the emission spectrum of tetramethlrhodamine. So FRET makes it possible to apply these AcF and AlK dual incorporated proteins to study conformational changes of proteins by choosing particular positions for ***UAA***s.

Also, one may decorate proteins with two different unnatural amino acids for their functional investigation. For instance, if the Protective Antigen (PA) of *Bacillus anthracis* is simultaneously decorated by two ***UAA***s at specifical sites, the new properties can benefit the immunogenicity study of PA. The impetus for the development of new anthrax vaccines is to protect humans against the infection of *Bacillus anthracis*, and the first licensure of protein-based vaccine, which mainly consists of Protein Antigen (PA), is used for human-beings [90, 91, 92, 93]. But it is not ideal to employ PA in its native form as vaccine since PA may facilitates the formation of anthrax toxin for post-exposure vaccination[94]. Wang *et al* reported mutants of PA which lost its pathogenic function but still kept the immunogenicity[94]. Thus, we may insert *p*-nitrophenylalanine (NO2F) and AlK into one PA non-functional mutant, and insert NO2F and AzF into another PA mutant. Then these two PA non-functional mutants can be covalently linked to form dimers through the CuAAC between AlK and AzF. Meanwhile, the existence of NO2F might improve the immunogenicity of PA since previous researches support that this type of chemical modification results in immunogenic epitopes that induce high-titer cross reactive antibodies[95, 96, 97].

Furthermore, previous studies showed that the incorporation of unnatural amino acids (NO2F), and post-translational modifications PTM (sulfotyrosine(SO3Y) and 3-nitrotyrosine (3NO2Y)), at specific sites in a self-protein enable the generation of a strong CD4-dependent (IgG isotype) autoantibody response against the native protein[96, 97, 98]. The activation of CD4 T cells were restricted by a particular class II molecule which recognize these modifications, and this recognition does not cross-react with the endogenous epitopes in self-proteins. On the other hand, many studies suggest that PTMs, such as 3NO2Tyr and SO3Tyr, may be recognized by the immune system in some cancers because the higher levels of the post-translational modified antigens were observed in tumor tissues than the healthy ones[99, 100]. Thus, we postulate that by introducing two PTM (SO3Y and 3NO2Y) into a self-antigen (such as cancer antigens) at proper sites will break the immune tolerance and serve as potential vaccines which might activate T cells that would be directed differentially to tumors relative to non-tumor tissues.

For another example, the adding of two ***UAA***s will make it possible to construct bispecific antibodies with an extended circulation time. By inserting AcF into antibodies, PG Schultz *et al* synthesized an anti-HER2/anti-CD3 bispecific antibody, which cross-linked HER2+ cancer cells and CD3+ cell to promote *in vitro* effector-cell mediated cytotoxicity[16]. Now, if we can insert another AlK into the antibody,

the circulation time of the therapeutic bispecific antibody will be extended because AlK reacts with Azide-PEG, and the latter is known to enhance protection from proteolytic degradation and extend circulation time.

A very important advantage of inserting two ***UAA***s is to add two more blocks with new chemical properties into traditional phage-displayed peptide libraries which is limited to 20 canonical amino acids. Actually few previous researches have shown that by introducing UAG amber codon suppression system into X-E. coli, phage-displayed proteins were able to carry a corresponding ***UAA***, such as sulfotyrosine (enhancing interaction strength at protein surfaces), p-boronophenylalanine (reacting with diols as 'chemical warhead'), bipyridylalanine (BpyAla) (invovling metal ion binding)[101, 102, 103, 104, 105]. Sulfotyrosine, often found in eukaryotic proteins, form strong salt bridges and hydrogen bonds by the sulfate group to strength protein interaction[103]. Liu *et al* reported that phage-displayed antibodies carried sulfotyrosine residues at positions either within the known sequence constraints required for post-translational sulfation or beyond it, and demonstrated a higher affinity to gp120 than a sulfated antibody library did which was isolated from human serum[101, 102, 103, 104, 105]. The orthogonal translational system, ***UAA***-RS/tRNA which decodes amber UAG codon, enables the site-specifically incorporation of sulfotyrosine into antibodies in the phage library. More important, by using the phage-based system, direct evolution experiments can be performed to optimize the amino acid sequences around the ***UAA***s, or even the entire peptide. So, since now we have the ability to insert two ***UAA***s inserted into proteins by using our dual+codon (AGG+UAG) recoding aaRS/tRNA pairs, phage-display libraries bearing various ***UAA*** pairs at designated positions will be feasible to be constructed with careful optimizing phage for the unbiased display of 20 natural amino acids and 2 ***UAA***s. After panning these unnatural phage display libraries, a large variety of unnatural peptides with diversified chemical structures can be rapidly screened against cancer therapeutic targets, in order to find out new strategies for cancer treatment and prevention. Potential targets can be vascular endothelial growth factor (target of antiangiogenic drugs), the cell surface HER2/neu (erbB2) receptor (target of erbB2+ breast cancer), and CD47 (ligant of a macrophage inhibitory receptor and preventing phagocytosis in cancer)[106, 107, 108]. For example, antibodies with high affinities (enhanced by sulfotyrosine) to certain type of cancer may also carry $I^{131}F$ for therapeutic purpose, or fluorophenylalanine $F^{18}F$ for diagnose purpose through Positron Emission Tomography(PET), or Alk to mediate bio-conjugation with some anti-cancer compounds.

Recoding AGA and the ochre codon UAA

Previous reports showed that the changing of the anticodon of $Mm\text{tRNA}^{\text{Pyl}}$ from CUA to UUA did not significantly affect its interaction with the catalytic domain of PylRS, and several pyrrolysine derivatives were incorporated into proteins in response to the

ochre codon UAA[41, 46]. The pairing of two stop codons, UAG and UAA, allowed the insertion of double unnatural amino acids. Also, GM10 bacteria only carry the natural suppression of amber (UAG) stop codons glnV44(AS), but do not possess any endogenous suppression of ochre(UAA). So it will be promising to recode AGA and UAA together since the only modification is to alter the anticodon of Pyl-tRNA from CUA to UUA.

Reassigning other sense codons

The two-step selection strategy has been proved to be effective for AGG and AGA recoding. The essential part of this selection is the GFP-Tyr66NNN probe. Many other rare codons may also serve as good candidates(Table 6.1), such as other two rare codons(CGA and CGG) for arginine[109], CUA for leucine and AUA for isoleucine [53]. Take CGA as an example. During the first selecting step, we aim to choose aaRS which can recognize UCG anticodon of tRNA and incorporate tyrosine in response to CGA by using a GFP-Tyr66CGA probe. Then in the second step, the amino acid bind pockets of the selected aaRS will be mutated to host desired ***UAA***s. To improve ***UAA*** incorporation efficiency, dual aaRS vectors are capable to increase synthetase and in turn ***UAA***-tRNA copy numbers inside bacteria. Also, the G1-to-A1 mutation of arginine $tRNA_{UCG}$ may be introduced into bacterial genome and help orthogonal $tRNA^{UAA}$ win the contest.

Table 6.1: Rare codon frequencies in *Escherichia coli*.

Rare codon frequencies in ***Escherichia coli***

	Arginine				Isoleucine	Leucine	Proline	Serine	Threonine
	AGA	AGG	CGA	CGG	AUA	CUA	CCC	UCA	ACA
E. coli B	2.4(5.10)	2.1(4.12)	2.4(5.10)	5.0(10.6)	5.0(7.15)	3.4(2.81)	2.4(5.93)	6.1(10.65)	6.1(11.38)
E. coli K-12	1.4(2.39)	1.6(2.74)	4.3(7.35)	4.1(7.01)	3.7(7.06)	5.3(5.21)	6.4(13.3)	7.8(15.35)	6.4(14.00)

Codon frequencies (per 1000) in ***Escherichia coli*** genome were extracted from the Kazusa codon usage database (http://www.kazusa.or.jp/codon/).
Values in parenthesis indicate the percentage of each specific codon among the codons for the amino acid stated.

Recoding sense codons in mammalian cells

In mammalian cells the rare codons are different from those in *E. coli*. In the following table(Table 6.2), 10 rare codons were listed based on their codon frequencies (per thousand) within the genome of *Homo sapiens*. Four of them, UUA and CUA for leucine, UCG for serine, and CGU for arginine, have both low codon frequencies and low codon percentages of each specific codon among the codons for the amino

acid stated. These four codons might serve as potential 'blank codons' to encode for unnatural amino acids within mammalian cells.

Table 6.2: Rare codon frequencies in *Homo sapiens*.

Rare codon frequencies in ***Homo sapiens***

Amino Acid	Luecine		Serine	Proline	Arginine		Isoleucine	Valine	Threonine	Alanine
Codon	UUA	CUA	UCG	CCG	CGU	CGA	AUA	GUA	ACG	GCG
Codon frequencies*	7.7	7.2	4.4	6.9	4.5	6.2	7.5	7.1	6.1	7.4
Codon percentages**	7.68	5.42	5.42	11.29	7.93	12.8	16.9	11.69	11.46	10.68

*Codon frequencies (per 1000) in *Homo sapiens* genome were extracted from the Kazusa codon usage database (http://www.kazusa.or.jp/codon/).
**Codon percentages indicate the percentage of each specific codon among the codons for the amino acid stated.

In mammalian cells, there are several aaRS/tRNA pairs to incorporate ***UAA***s. One pair derived from *E. coli* TyrRS/*Bacillus stearothermophilus* tRNA^{Tyr} [110, 111] and a mutant *E. coli* TyrRS/*E. coli* tRNA^{Tyr} pair [112, 113, 114] were both proved to be orthogonal in mammalian cells and able to suppress UAG amber codon for ***UAA***s. The use of *Bst*tRNA^{Tyr} is to carry intact internal A- and B-box promoters that drive proper tRNA expression in eukaryotic cells[110]. For similar purpose, external H1 and U6 promoters and 3' flanking regions which were derived from mammalian sequences were adopted to produce functional *E. coli* tRNA^{Tyr} in mammalian cells[112]. Thus these two aaRS/pairs can serve as candidates to recode sense rare codons in mammalian cells by using a similar GFP-Tyr66NNN reporter and the two step selection strategy as mentioned in Chapter 2 and 3.

Interestingly there is another approach to modify the existing *Mj*TyrRS/tRNA to make this pair orthogonal within mammalian cells [115, 116]. The key determinant in archaeal and eukaryotic tRNA synthetase-tRNA pairs is C1:G72 base pair of tRNA, which is specifically recognized by the connective polypeptide region (CP1) domain of in the corresponding tRNA synthetase. Zhang *et al* increased the orthogonality of *Mj*TyrRS/tRNA within mammalian cells by using peptide transplantation involving CP1 domain of *E. coli* TyrRS and switching C1:G72 of *Mj*tRNA to G1:G72 [115]. By over-expressing archeabacterial tRNA in mammalian cells, they increased the ***UAA*** incorporation efficiencies [116]. So, this approach gives us another direction to recode sense codons in mammalian cells. It is feasible to introduce our evolved *Mj* ***UAA***-RS/tRNA for sense codons into the mammalian cells, after the proper modifications, such as peptide transplantation with CP1 domain of *E. coli* TyrRS and switching C1:G72 of *Mj*tRNA to G1:G72. However, it needs more investigation to improve the ***UAA*** incorporation efficiency besides using the dual RS vectors.

Summary

To sum up, the incorporation of multiple unnatural amino acids expand the scope of protein conformational and functional studies. Further expanding of the genetic code will make this tool, site-specifically incorporation of unnatural amino aicds, more powerful.

Bibliography

[1] J. Xie and P. G. Schultz. "An expanding genetic code". In: *Methods* 36.3 (2005), pp. 227–38.

[2] Caroline Kohrer and Uttam RajBhandary. *Protein engineering*. Nucleic acids and molecular biology, Berlin: Springer, 2009.

[3] C. C. Liu and P. G. Schultz. "Adding new chemistries to the genetic code". In: *Annu Rev Biochem* 79 (2010), pp. 413–44.

[4] L. Wang and P. G. Schultz. "A general approach for the generation of orthogonal tRNAs". In: *Chem Biol* 8.9 (2001), pp. 883–90.

[5] C. H. Kim, J. Y. Axup, and P. G. Schultz. "Protein conjugation with genetically encoded unnatural amino acids". In: *Curr Opin Chem Biol* 17.3 (2013), pp. 412–9.

[6] J. Kalia and R. T. Raines. "Hydrolytic stability of hydrazones and oximes". In: *Angew Chem Int Ed Engl* 47.39 (2008), pp. 7523–6.

[7] H. C. Kolb, M. G. Finn, and K. B. Sharpless. "Click Chemistry: Diverse Chemical Function from a Few Good Reactions". In: *Angew Chem Int Ed Engl* 40.11 (2001), pp. 2004–2021.

[8] M. P. Leal et al. "Copper-catalyzed azide-alkyne cycloaddition in the synthesis of polydiacetylene: "click glycoliposome" as biosensors for the specific detection of lectins". In: *Chemistry* 17.6 (2011), pp. 1828–36.

[9] S. I. Presolski, V. P. Hong, and M. G. Finn. "Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation". In: *Curr Protoc Chem Biol* 3.4 (2011), pp. 153–162.

[10] D. P. Nguyen et al. "Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA Synthetase/tRNA(CUA) pair and click chemistry". In: *J Am Chem Soc* 131.25 (2009), pp. 8720–1.

[11] J. W. Chin and P. G. Schultz. "In vivo photocrosslinking with unnatural amino Acid mutagenesis". In: *Chembiochem* 3.11 (2002), pp. 1135–7.

[12] J. W. Chin et al. "Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli". In: *J Am Chem Soc* 124.31 (2002), pp. 9026–7.

[13] S. A. Kazane et al. "Self-Assembled Antibody Multimers through Peptide Nucleic Acid Conjugation". In: *Journal of the American Chemical Society* 135.1 (2013), pp. 340–346.

[14] H. Cho et al. "Optimized clinical performance of growth hormone with an expanded genetic code". In: *Proc Natl Acad Sci U S A* 108.22 (2011), pp. 9060–5.

[15] J. Mu et al. "FGF21 analogs of sustained action enabled by orthogonal biosynthesis demonstrate enhanced antidiabetic pharmacology in rodents". In: *Diabetes* 61.2 (2012), pp. 505–12.

[16] C. H. Kim et al. "Synthesis of Bispecific Antibodies using Genetically Encoded Unnatural Amino Acids". In: *Journal of the American Chemical Society* 134.24 (2012), pp. 9918–9921.

[17] D. R. Liu and P. G. Schultz. "Progress toward the evolution of an organism with an expanded genetic code". In: *Proc Natl Acad Sci U S A* 96.9 (1999), pp. 4780–5.

[18] L. Wang and P. G. Schultz. "Expanding the genetic code". In: *Angew Chem Int Ed Engl* 44.1 (2004), pp. 34–66.

[19] J. W. Chin et al. "An expanded eukaryotic genetic code". In: *Science* 301.5635 (2003), pp. 964–7.

[20] 3rd Melancon C. E. and P. G. Schultz. "One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases". In: *Bioorg Med Chem Lett* 19.14 (2009), pp. 3845–7.

[21] S. W. Santoro et al. "An efficient system for the evolution of aminoacyl-tRNA synthetase specificity". In: *Nat Biotechnol* 20.10 (2002), pp. 1044–8.

[22] L. Wang and P. G. Schultz. "Expanding the genetic code". In: *Chem Commun (Camb)* 1 (2002), pp. 1–11.

[23] T. Kobayashi et al. "Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion". In: *Nat Struct Biol* 10.6 (2003), pp. 425–32.

[24] B. A. Steer and P. Schimmel. "Major anticodon-binding region missing from an archaebacterial tRNA synthetase". In: *J Biol Chem* 274.50 (1999), pp. 35601–6.

[25] W. Liu et al. "Structural basis for the recognition of para-benzoyl-L-phenylalanine by evolved aminoacyl-tRNA synthetases". In: *Angew Chem Int Ed Engl* 46.32 (2007), pp. 6073–5.

[26] J. M. Turner et al. "Structural plasticity of an aminoacyl-tRNA synthetase active site". In: *Proc Natl Acad Sci U S A* 103.17 (2006), pp. 6483–8.

[27] J. M. Turner et al. "Structural characterization of a p-acetylphenylalanyl aminoacyl-tRNA synthetase". In: *J Am Chem Soc* 127.43 (2005), pp. 14976–7.

[28] M. L. Tsao, F. Tian, and P. G. Schultz. "Selective Staudinger modification of proteins containing p-azidophenylalanine". In: *Chembiochem* 6.12 (2005), pp. 2147–9.

[29] L. Wang et al. "Expanding the genetic code of Escherichia coli". In: *Science* 292.5516 (2001), pp. 498–500.

[30] D. D. Young et al. "An evolved aminoacyl-tRNA synthetase with atypical polysubstrate specificity". In: *Biochemistry* 50.11 (2011), pp. 1894–900.

[31] G. Srinivasan, C. M. James, and J. A. Krzycki. "Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA". In: *Science* 296.5572 (2002), pp. 1459–62.

[32] S. K. Blight et al. "Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo". In: *Nature* 431.7006 (2004), pp. 333–5.

[33] J. A. Krzycki. "The direct genetic encoding of pyrrolysine". In: *Curr Opin Microbiol* 8.6 (2005), pp. 706–12.

[34] J. Yuan et al. "Distinct genetic code expansion strategies for selenocysteine and pyrrolysine are reflected in different aminoacyl-tRNA formation systems". In: *FEBS Lett* 584.2 (2010), pp. 342–9.

[35] T. Mukai et al. "Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases". In: *Biochem Biophys Res Commun* 371.4 (2008), pp. 818–22.

[36] T. Yanagisawa et al. "Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification". In: *Chem Biol* 15.11 (2008), pp. 1187–97.

[37] K. Nozawa et al. "Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality". In: *Nature* 457.7233 (2009), pp. 1163–7.

[38] J. M. Kavran et al. "Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation". In: *Proc Natl Acad Sci U S A* 104.27 (2007), pp. 11268–73.

[39] A. Ambrogelly et al. "Pyrrolysine is not hardwired for cotranslational insertion at UAG codons". In: *Proc Natl Acad Sci U S A* 104.9 (2007), pp. 3141–6.

[40] C. R. Polycarpo et al. "Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase". In: *FEBS Lett* 580.28-29 (2006), pp. 6695–700.

[41] W. Wan et al. "A facile system for genetic incorporation of two different noncanonical amino acids into one protein in Escherichia coli". In: *Angew Chem Int Ed Engl* 49.18 (2010), pp. 3211–4.

[42] Y. Huang et al. "A convenient method for genetic incorporation of multiple noncanonical amino acids into one protein in Escherichia coli". In: *Mol Biosyst* 6.4 (2010), pp. 683–6.

[43] J. K. Takimoto et al. "Stereochemical basis for engineered pyrrolysyl-tRNA synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids". In: *ACS Chem Biol* 6.7 (2011), pp. 733–43.

[44] H. Neumann et al. "Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome". In: *Nature* 464.7287 (2010), pp. 441–4.

[45] J. C. Anderson et al. "An expanded genetic code with a functional quadruplet codon". In: *Proc Natl Acad Sci U S A* 101.20 (2004), pp. 7566–71.

[46] K. A. Odoi et al. "Nonsense and sense suppression abilities of original and derivative Methanosarcina mazei pyrrolysyl-tRNA synthetase-tRNA(Pyl) pairs in the Escherichia coli BL21(DE3) cell strain". In: *PLoS One* 8.3 (2013), e57035.

[47] S. W. Santoro et al. "An archaebacteria-derived glutamyl-tRNA synthetase and tRNA pair for unnatural amino acid mutagenesis of proteins in Escherichia coli". In: *Nucleic Acids Res* 31.23 (2003), pp. 6700–9.

[48] A. Chatterjee, H. Xiao, and P. G. Schultz. "Evolution of multiple, mutually orthogonal prolyl-tRNA synthetase/tRNA pairs for unnatural amino acid mutagenesis in Escherichia coli". In: *Proc Natl Acad Sci U S A* 109.37 (2012), pp. 14841–6.

[49] H. Neumann, A. L. Slusarczyk, and J. W. Chin. "De novo generation of mutually orthogonal aminoacyl-tRNA synthetase/tRNA pairs". In: *J Am Chem Soc* 132.7 (2010), pp. 2142–4.

[50] A. Chatterjee et al. "A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli". In: *Biochemistry* (2013).

[51] H. S. Park et al. "Expanding the genetic code of Escherichia coli with phosphoserine". In: *Science* 333.6046 (2011), pp. 1151–4.

[52] D. B. Johnson et al. "RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites". In: *Nat Chem Biol* 7.11 (2011), pp. 779–86.

[53] G. L. Rosano and E. A. Ceccarelli. "Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain". In: *Microb Cell Fact* 8 (2009), p. 41.

[54] J. Guo et al. "Evolution of amber suppressor tRNAs for efficient bacterial production of proteins containing nonnatural amino acids". In: *Angew Chem Int Ed Engl* 48.48 (2009), pp. 9148–51.

[55] T. S. Young et al. "An enhanced system for unnatural amino acid mutagenesis in E. coli". In: *J Mol Biol* 395.2 (2010), pp. 361–74.

[56] J. M. Schrader, S. J. Chapman, and O. C. Uhlenbeck. "Understanding the sequence specificity of tRNA binding to elongation factor Tu using tRNA mutagenesis". In: *J Mol Biol* 386.5 (2009), pp. 1255–64.

[57] S. K. Sharan et al. "Recombineering: a homologous recombination-based method of genetic engineering". In: *Nat Protoc* 4.2 (2009), pp. 206–23.

[58] Angelo Palmese et al. "Dansyl labeling and bidimensional mass spectrometry to investigate protein carbonylation". In: *Rapid Commun. Mass Spectrom.* 25.1 (2011), pp. 223–231. ISSN: 1097-0231. URL: `http://dx.doi.org/10.1002/rcm.4863`.

[59] C. W. Cody et al. "Chemical structure of the hexapeptide chromophore of the Aequorea green-fluorescent protein". In: *Biochemistry* 32.5 (1993), pp. 1212–8.

[60] M. Chalfie. "Green fluorescent protein". In: *Photochem Photobiol* 62.4 (1995), pp. 651–6.

[61] R. Heim, D. C. Prasher, and R. Y. Tsien. "Wavelength mutations and post-translational autoxidation of green fluorescent protein". In: *Proc Natl Acad Sci U S A* 91.26 (1994), pp. 12501–4.

[62] D. C. Youvan and M. E. Michel-Beyerle. "Structure and fluorescence mechanism of GFP". In: *Nat Biotechnol* 14.10 (1996), pp. 1219–20.

[63] F. Yang, L. G. Moss, and Jr. Phillips G. N. "The molecular structure of green fluorescent protein". In: *Nat Biotechnol* 14.10 (1996), pp. 1246–51.

[64] M. Ormo et al. "Crystal structure of the Aequorea victoria green fluorescent protein". In: *Science* 273.5280 (1996), pp. 1392–5.

[65] L. Wang et al. "Unnatural amino acid mutagenesis of green fluorescent protein". In: *J Org Chem* 68.1 (2003), pp. 174–6.

[66] F. Wang et al. "Unnatural amino acid mutagenesis of fluorescent proteins". In: *Angew Chem Int Ed Engl* 51.40 (2012), pp. 10132–5.

[67] K. Sakamoto et al. "A modified uridine in the first position of the anticodon of a minor species of arginine tRNA, the argU gene product, from Escherichia coli". In: *Eur J Biochem* 216.2 (1993), pp. 369–75.

[68] K. Sakamoto et al. "The Escherichia coli argU10(Ts) phenotype is caused by a reduction in the cellular level of the argU tRNA for the rare codons AGA and AGG". In: *J Bacteriol* 186.17 (2004), pp. 5899–905.

[69] G. M. Garcia et al. "The E. coli dnaY gene encodes an arginine transfer RNA". In: *Cell* 45.3 (1986), pp. 453–9.

[70] R. A. Spanjaard et al. "Frameshift suppression at tandem AGA and AGG codons by cloned tRNA genes: assigning a codon to argU tRNA and T4 tRNA(Arg)". In: *Nucleic Acids Res* 18.17 (1990), pp. 5031–6.

[71] I. S. Farrell et al. "Photo-cross-linking interacting proteins with a genetically encoded benzophenone". In: *Nat Methods* 2.5 (2005), pp. 377–84.

[72] M. Guilloton and F. Karst. "Cyanate specifically inhibits arginine biosynthesis in Escherichia coli K12: a case of by-product inhibition?" In: *J Gen Microbiol* 133.3 (1987), pp. 655–65.

[73] K. Tomita et al. "A cytotoxic ribonuclease which specifically cleaves four isoaccepting arginine tRNAs at their anticodon loops". In: *Proc Natl Acad Sci U S A* 97.15 (2000), pp. 8278–83.

[74] K. S. Chen, T. C. Peters, and J. R. Walker. "A minor arginine tRNA mutant limits translation preferentially of a protein dependent on the cognate codon". In: *J Bacteriol* 172.5 (1990), pp. 2504–10.

[75] W. Fischer et al. "Interaction of methionine-specific tRNAs from Escherichia coli with immobilized elongation factor Tu". In: *FEBS Lett* 192.1 (1985), pp. 151–4.

[76] J. A. Krzycki. "Function of genetically encoded pyrrolysine in corrinoid-dependent methylamine methyltransferases". In: *Curr Opin Chem Biol* 8.5 (2004), pp. 484–91.

[77] O. Namy et al. "Adding pyrrolysine to the Escherichia coli genetic code". In: *FEBS Lett* 581.27 (2007), pp. 5282–8.

[78] W. T. Li et al. "Specificity of pyrrolysyl-tRNA synthetase for pyrrolysine and pyrrolysine analogs". In: *J Mol Biol* 385.4 (2009), pp. 1156–64.

[79] M. Tsunoda et al. "Structural basis for recognition of cognate tRNA by tyrosyl-tRNA synthetase from three kingdoms". In: *Nucleic Acids Res* 35.13 (2007), pp. 4289–300.

[80] A. Deiters et al. "Site-specific PEGylation of proteins containing unnatural amino acids". In: *Bioorg Med Chem Lett* 14.23 (2004), pp. 5743–5.

[81] J. E. Hein and V. V. Fokin. "Copper-catalyzed azide-alkyne cycloaddition (CuAAC) and beyond: new reactivity of copper(I) acetylides". In: *Chem Soc Rev* 39.4 (2010), pp. 1302–15.

[82] N. W. Nairn et al. "Development of copper-catalyzed azide-alkyne cycloaddition for increased in vivo efficacy of interferon beta-1b by site-specific PEGylation". In: *Bioconjug Chem* 23.10 (2012), pp. 2087–97.

[83] M. Gonzalez-Gonzalez, K. Mayolo-Deloisa, and M. Rito-Palomares. "PEGylation, detection and chromatographic purification of site-specific PEGylated CD133-Biotin antibody in route to stem cell separation". In: *J Chromatogr B Analyt Technol Biomed Life Sci* 893-894 (2012), pp. 182–6.

[84] V. Hong et al. "Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation". In: *Angew Chem Int Ed Engl* 48.52 (2009), pp. 9879–83.

[85] O. W. Odom et al. "An effect of polyethylene glycol 8000 on protein mobility in sodium dodecyl sulfate-polyacrylamide gel electrophoresis and a method for eliminating this effect". In: *Anal Biochem* 245.2 (1997), pp. 249–52.

[86] M. Adamczyk, R. E. Reddy, and Z. Yu. "Synthesis of a novel fluorescent probe for estrogen receptor". In: *Bioorg Med Chem Lett* 12.9 (2002), pp. 1283–5.

[87] L. Wang, Y. Wang, and A. J. Ragauskas. "A novel FRET approach for in situ investigation of cellulase-cellulose interaction". In: *Anal Bioanal Chem* 398.3 (2010), pp. 1257–62.

[88] X. Li and T. M. Kapoor. "Approach to profile proteins that recognize post-translationally modified histone "tails"". In: *J Am Chem Soc* 132.8 (2010), pp. 2504–5.

[89] C. Y. Jao and A. Salic. "Exploring RNA transcription and turnover in vivo by using click chemistry". In: *Proc Natl Acad Sci U S A* 105.41 (2008), pp. 15779–84.

[90] T. M. Koehler (ed.) *Anthrax*. Current topics in microbiology and immunology 271. Springer, c2002.

[91] St. Louis University. Center for the Study of Bioterrorism & Emerging Infections. *Clinical description and epidemiology of bioterrorism agents anthrax, smallpox, and plague*. 2002.

[92] R. John Collier and National Institutes of Health (U.S.) *Anthrax toxin structure, activity, and new inhibitors*. 2002.

[93] D. W. Zhan et al. "[Progress on the vaccine for anthrax]". In: *Wei Sheng Wu Xue Bao* 45.1 (2005), pp. 149–52.

[94] M. Yan et al. "Selection and evaluation of the immunogenicity of protective antigen mutants as anthrax vaccine candidates". In: *Vaccine* 26.7 (2008), pp. 947–55.

[95] M. L. Tsao et al. "The genetic incorporation of a distance probe into proteins in Escherichia coli". In: *J Am Chem Soc* 128.14 (2006), pp. 4572–3.

[96] J. Grunewald et al. "Immunochemical termination of self-tolerance". In: *Proc Natl Acad Sci U S A* 105.32 (2008), pp. 11276–80.

[97] J. Grunewald et al. "Mechanistic studies of the immunochemical termination of self-tolerance with unnatural amino acids". In: *Proc Natl Acad Sci U S A* 106.11 (2009), pp. 4337–42.

[98] V. Gauba et al. "Loss of CD4 T-cell-dependent tolerance to proteins with modified amino acids". In: *Proc Natl Acad Sci U S A* 108.31 (2011), pp. 12821–6.

[99] F. A. Masri et al. "Abnormalities in nitric oxide and its derivatives in lung cancer". In: *Am J Respir Crit Care Med* 172.5 (2005), pp. 597–605.

[100] B. Pignatelli et al. "Nitrated and oxidized plasma proteins in smokers and lung cancer patients". In: *Cancer Res* 61.2 (2001), pp. 778–84.

[101] F. Tian, M. L. Tsao, and P. G. Schultz. "A phage display system with unnatural amino acids". In: *J Am Chem Soc* 126.49 (2004), pp. 15962–3.

[102] C. C. Liu et al. "Protein evolution with an expanded genetic code". In: *Proc Natl Acad Sci U S A* 105.46 (2008), pp. 17688–93.

[103] C. C. Liu et al. "Mutagenesis and evolution of sulfated antibodies using an expanded genetic code". In: *Biochemistry* 48.37 (2009), pp. 8891–8.

[104] C. C. Liu et al. "Efficient expression of tyrosine-sulfated proteins in E. coli using an expanded genetic code". In: *Nat Protoc* 4.12 (2009), pp. 1784–9.

[105] J. W. Day et al. "Identification of metal ion binding peptides containing unnatural amino acids by phage display". In: *Bioorg Med Chem Lett* 23.9 (2013), pp. 2598–600.

[106] I. S. Moreira, P. A. Fernandes, and M. J. Ramos. "Vascular endothelial growth factor (VEGF) inhibition–a critical review". In: *Anticancer Agents Med Chem* 7.2 (2007), pp. 223–45.

[107] A. Batistatou et al. "Evaluation of current prognostic and predictive markers in breast cancer: a validation study of tissue microarrays". In: *Anticancer Res* 33.5 (2013), pp. 2139–45.

[108] M. H. Kershaw and M. J. Smyth. "Immunology. Making macrophages eat cancer". In: *Science* 341.6141 (2013), pp. 41–2.

[109] E. M. Gustilo, F. A. Vendeix, and P. F. Agris. "tRNA's modifications bring order to gene expression". In: *Curr Opin Microbiol* 11.2 (2008), pp. 134–40.

[110] K. Sakamoto et al. "Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells". In: *Nucleic Acids Res* 30.21 (2002), pp. 4692–9.

[111] W. Liu et al. "Genetic incorporation of unnatural amino acids into proteins in mammalian cells". In: *Nat Methods* 4.3 (2007), pp. 239–44.

[112] W. Wang et al. "Genetically encoding unnatural amino acids for cellular and neuronal studies". In: *Nat Neurosci* 10.8 (2007), pp. 1063–72.

[113] B. Shen et al. "Genetically encoding unnatural amino acids in neural stem cells and optically reporting voltage-sensitive domain changes in differentiated neurons". In: *Stem Cells* 29.8 (2011), pp. 1231–40.

[114] A. Chatterjee et al. "Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells". In: *Proc Natl Acad Sci U S A* 110.29 (2013), pp. 11803–8.

[115] G. N. Thibodeaux et al. "Transforming a pair of orthogonal tRNA-aminoacyl-tRNA synthetase from Archaea to function in mammalian cells". In: *PLoS One* 5.6 (2010), e11263.

[116] L. Xiang et al. "Crucial Optimization of Translational Components towards Efficient Incorporation of Unnatural Amino Acids into Proteins in Mammalian Cells". In: *PLoS One* 8.7 (2013), e67333.