UC Merced
UC Merced Electronic Theses and Dissertations

Title
Adding new chemistry to proteins via genetic incorporation

Permalink
https://escholarship.org/uc/item/9jb4g6jb

Author
Chen, Shuo

Publication Date
2014

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, MERCED

Adding new chemistry to proteins via genetic incorporation

A Dissertation Presented

by

Shuo Chen

Submitted to the Graduate Division of the

University of California, Merced in partial fulfillment

of the requirements for the degree of

DOCTOR OF PHILOSOPHY

2014

Chemistry and Chemical Biology

Committee in Charge:

Professor Andy LiWang, Chair

Professor Patricia LiWang

Professor Matthew Meyer

Professor Jinah Choi

Professor Tao Ye, Supervisor

© Copyright by Shuo Chen 2014

All Rights Reserved

The Dissertation of Shuo Chen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Andy LiWang, Chair

_____

Patricia LiWang, Member

_____

Matthew Meyer, Member

_____

Jinah Choi, Member

_____

Tao Ye, Supervisor

_____

DEDICATION

To My Family,

Thank you for all of your love, support and sacrifice throughout my life.

To Professor Meng-Lin Tsao,

Thank you for all your kind help, guidance and the opportunity to work with you.

CURRICULUM VITAE

SHUO CHEN

220 Brookdale Dr, Apt 32 • Merced, CA 95340 • (209) 201-7756 • chenshuol@hotmail.com

EDUCATION

09/2008-present **University of California at Merced**, Merced, CA
- Ph.D. in Organic Chemistry

09/2004-06/2008 **Peking University**, Beijing, China
- B.A. in Chemistry

RESEARCH/WORKING EXPERIENCE

09/2008-present **Research Assistant**, University of California at Merced, Merced, CA
- Evolved an aminoacyl-tRNA synthetase that can be utilized to introduce an unnatural amino acid into protein *in vivo* in response to the UAG stop codon
- Developed a selective azo coupling reaction towards 2-naphthol residue for site-specific protein modification
- Developed an effective lysine-2NpOH coupling reaction

08/2007-07/2008 **Undergraduate Research Assistant**, Peking University, Beijing, China
- Synthesized a liquid crystal polymer with potential optical property

05/2006-05/2007 **Undergraduate Research Assistant**, Peking University, Beijing, China
- Synthesized natural products with potential bioactivities

09/2008-12/2013 **Teaching Assistant/Fellow**, University of California, Merced, Merced, CA
- Monitored laboratory work of General Chemistry, Organic Chemistry, Advanced Organic Chemistry

08/2012- 08/2013 **Peer Instructor**, Graduate Division, University of California, Merced, Merced, CA
- Presented teaching skills and strategies in TA orientation

PUBLICATION/PRESENTATION

- **Shuo Chen**; Meng-Lin Tsao. Genetic Incorporation of 2-Naphthol Group into Proteins for Site-Specific Azo Coupling. *(Bioconjugate Chem*. **2013**, *24*, 1645-1649).
- **Shuo Chen**; Meng-Lin Tsao. Genetic incorporation of a 2-naphthol bearing unnatural amino acid into proteins in *E. coli* and a site-specific azo formation reaction on the 2-naphthol group with aromatic diazonium salts" Poster presentation delivered at ACS National Conference, March, 2011, Anaheim, CA.

ACKNOWLEDGMENTS

Being a PhD student is not only about learning knowledge, applying scientific technology to research, but also finding the way to become a better person and realizing the strength and weakness of myself. Thus, for the past six years, I have learnt so much, and have grown both intellectually and emotionally. I will always cherish this valuable experience in my life.

My deepest and earnest gratitude is to my advisor, my mentor, Professor Meng-Lin Tsao. There is a Chinese old saying that “a teacher for a day, a father for a lifetime”. And Professor Tsao is like a kind father to me in certain way. I still remembered his emails and phone calls that helped me make the decision to attend UC Merced. I really appreciated the opportunity to be able to study abroad and conducted my graduate research in a brand new laboratory. Prof. Tsao, is quite knowledgeable in his area. Before I came, I knew only a little about chemical biology. He was patient with me, and showed me all the technique himself. He had his trust in me, although with my deepest regret, I somehow failed him. And now, he had to leave UC Merced, and could not serve as my advisor anymore. I always have my respect to him.

Secondly, I would like to thank my current advisor, Professor Tao Ye. With my fear of future and procrastination, I could almost jeopardize my path to obtain the degree. It was him, who gave me a wake-up call, pushed me out my comfort zone. He helped me to realize the future was nothing to be afraid of, but only the thing I could make by myself. With his own work and family issue, he still tried his best to get the opportunity to maintain my student’s status. Moreover, he helped me to become a more self-motivated person who can survive later in the society.

Thirdly, I would like to thank all my committee members, Prof Andy LiWang, Prof. Patricia LiWang, Prof. Matthew Meyer and Prof Jinah Choi. They had given me a lot of good advice on research and future career plan. Also their encouragement and understanding helped me finally get to this stage.

Then, I would like to thank my family. Without their support, I cannot be on my own to study abroad. They always had their faith in me, and always respected my choices. I am so glad that I was born in this caring loving and giving family. As their only son, being away from them with time difference, I not only understand but also know their feeling of missing. I would take this work as a way to show my love to them. I would also thank the optimistic way to see the world they taught me, that helped me go through some of the frustrated days.

I would like to thank my lovely girlfriend. Being away from my family, it was her that let me know that I am not alone. And it was her who gave me the will to fight for a better future. Whenever I met with some troubles, her kind warm words made me

calm down.

I would like to thank all my lab mates. As a chemistry major, I was not familiar with a lot of biological research technique when I first started at this lab. Dr. Yiyan Wang, Dr. Jingui Zhu, who were both patient with me and shared their valuable experience with me. They always were kind enough to lend me a hand, when I had questions. Also I would like to thank all the undergraduates who worked in our lab, Duc, Joza, Nova, Julie, Wyatt, and Joel. Their help and the laughter they brought into our lab are something I would remember. I would like to thank all friends for their help during my days at Merced.

Lastly but not the least, I would like to acknowledge all the funding that supported me to conduct the work in this dissertation. Also I would like to thank all the materials from Professor Peter Schultz's research group, the DNA sequencing service and Mass spectrometry service from UC Berkeley, the LC-MS equipment from Prof. Jason Hein's lab, and the NMR equipment from Prof. Matthew Meyer's lab.

ABSTRACT

Adding New Chemistry to Proteins via Genetic Incorporation

by

Shuo Chen, B.S., Peking University

Directed by: Professor Meng-Lin Tsao

Professor Tao Ye

Incorporation of unnatural amino acids into proteins provides a powerful way to add new chemistries to proteins, which would be an advantage to study protein behavior, to provide a site-specific protein modification method, and to enhance protein properties. Rather than incorporating unnatural amino acid into proteins by means of chemical methods, recent advances in genetic incorporation of unnatural amino acid into proteins might be an effective way to introduce novel chemical groups into proteins, which shows a promising future in chemical biology studies.

The unnatural amino acid, 2-amino-3- (6-hydroxy-naphthalen-2-yl)propanoic acid (2NpOH) is an analogue of tyrosine, with an orthogonal reactivity and unique fluorescence properties, which can be utilized as site-specific bio-conjugation tag or as a biochemical probe. Based on rational design and random mutation, two aminoacyl-tRNA synthetases (RS-NpOH), which can be utilized to introduce 2-amino-3- (6-hydroxy-naphthalen-2-yl) propanoic acid into protein *in vivo* in response to the UAG stop codon, were evolved. Protein expression with RS-NpOH showed that the unnatural amino acid could be successfully incorporated into proteins. Both ESI-MS analysis and the strong blue fluorescence acquired by addition of 2NpOH indicated that both aminoacyl-tRNA synthetases had high efficiency and fidelity for selective incorporation of 2NpOH.

A chemoselective azo coupling reaction between the genetically encoded 2-naphthol group and diazotized aniline derivatives was developed. The coupling reaction only required very mild condition of pH 7 at 0 °C, and possessed fast reaction rate, high efficiency and excellent selectivity. Thus, it provides us an alternate method for bio-conjugation, such as PEGylation. Furthermore, to exploit the higher reactivity of the 2-naphthol residue under a nucleophilic addition coupling, an *in vivo* Mannich type reaction was proposed. It revealed that a cross coupling reaction could achieve between a lysine residue and a 2NpOH residue

with the presence of formaldehyde at physiological conditions. Thus, new functionality can be attached to the unnatural amino acid site of proteins, and it would provide us a new tool to study the behavior of proteins and eventually the production of functionally enhanced proteins.

In addition, the fluorescent feature of this unnatural amino acid was also studied upon insertion of the 2NpOH into the GFP chromophore for protein evolution. However, no red-shift fluorescence was observed as expected

In this work, incorporation of two unnatural amino acids into single protein was also demonstrated.

TABLE OF CONTENTS

CURRICULUM VITAE v
ACKNOWLEDGMENTS vi
ABSTRACT viii
TABLE OF CONTENTS x
LIST OF TABLES xii
LIST OF FIGURES xiii
LIST OF ABBREVIATIONS xx
Chapter 1 1
Background on genetically incorporated unnatural amino acid into proteins. 1
1.1 New chemistries beyond those contained in the natural amino acid can provide advantages not only in protein function but also in protein studies. 1
1.2 Benefits from incorporation of unnatural amino acids with expanded genetic code 2
1.3 A general approach for the site-specific incorporation of unnatural amino acids into proteins *in vivo* 4
1.3.1 Components necessary for the genetically incorporation of UAAs into proteins. 4
1.3.2 Two commonly used orthogonal aminoacyl-tRNA synthetase/ tRNA pairs for *E. coli* host. 5
1.3.3 Unique codons assigned for the unnatural amino acids 8
1.4 Recent progresses on bio-conjugation reactions. 8
1.5 Summary 10
Chapter 2 12
Genetically incorporate 2NpOH into proteins. 12
2.1 Target Unnatural Amino Acid. 12
2.2 Synthesis of 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid 13
2.2.1 One-pot Synthesis 13
2.2.2 Multiple Steps Synthesis 14
2.3 Selection of tRNA Synthetase that charges the unnatural amino acid to its cognate tRNA 15
2.3 Evolution of the aaRS that charges 2NpOH 17
2.3.1 A general positive/negative selection system 17
2.3.2 An one-plasmid selection system. 19
2.3.3 Combination of two selection system. 20
2.4 Protein Expression 24
2.5 An enhanced protein expression system. 29
2.6 Conclusion 30
Chapter 3 31
Azo coupling of 2NpOH incorporated protein and reactions thereafter. 31
3.1 Diazonium reaction with tyrosine residue 31
3.2 Small molecule test for optimization of the reaction condition 32
3.3 Bio-conjugation of Z-domain proteins 36
3.4 Kinetic study of the azo-coupling reaction. 40
3.5 PEGylation of 2NpOH-Z domain protein 41
3.6 Reductive cleavage of the azo adducts. 42

3.7 Protein immobilization 44
3.8 Summary 47
Chapter 4 48
Study on a three-component Mannich-type coupling reaction towards the 2-naphthol residue. 48
4.1 A three-component Mannich-type coupling reaction with tyrosine residue. 48
4.2 Small molecule study 49
4.3 *In vivo* reaction 50
4.4 Formaldehyde-based crosslinking with lysine residue. 56
4.5 Discussion 64
Chapter 5 66
Other endeavor and future work 66
5.1 Application of the incorporation of 2-amino-3-(6-hydroxy-2-naphthyl) propanoic acid for directed protein evolution 66
5.2 Incorporation of Multiple UAAs into proteins. 72
Chapter 6 78
Conclusions and outlook 78
6.1 Conclusions 78
6.2 Outlook and future directions 79
Bibliography 83

LIST OF TABLES

Table Page

Table 1 Mutation sites in the evolved aaRS that selectively charge with 2NpOH 21

Table 2 Reaction yields for the azo coupling of *p*-cresol and 2-naphthol with diazotized aniline derivatives and the characteristic UV-vis absorption maxima for the azo products. 35

Table 3 Percent yields for the azo coupling of p-cresol and 2-naphthol with diazotized aniline derivatives. [a]..... 35

LIST OF FIGURES

Figure Page

Figure 1. 20 natural amino acids 2

Figure 2. Protein bio-conjugation via Click chemistry on *p*-azido-phenylalanine residue. 3

Figure 3. The incorporation of amino acid into proteins. 4

Figure 4. Crystal structure of *Mm* Pyrrolysyl-tRNA synthetase bound to adenylated pyrrolysine and pyrophosphate (Figure is generated with PyMOL based on the crystal structure of Pyrrolysyl-tRNA synthetase, PDB ID: 2ZIM[19]) 6

Figure 5. The comparison of *E. coli* $tRNA^{Tyr}$, *Mj* $tRNA^{Tyr}$ and an orthogonal $tRNA^{Tyr}_{CUA}$ in *E. coli, which* was derived from *Mj* $tRNA^{Tyr}$. 7

Figure 6. Classical bio-conjugation reaction towards lysine and cysteine residue 9

Figure 7. Schematic representation of the uses of genetically encoded UAA side chains, including bio-orthogonal reactive groups, spectroscopic probes, photo-crosslinkers etc. 10

Figure 8. Genetically incorporate 2NpOH into proteins 12

Figure 9. Tyrosine and its 2-naphthol analogue, 2-amino-3-(6-hydroxy-2-naphthyl) propanoic acid. 12

Figure 10. Synthesis of 6-methoxy-2-naphthaldehyde 13

Figure 11. One-pot synthesis of 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid 14

Figure 12. Multiple steps synthesis strategy 14

Figure 13. A: The crystal structure of 3-(2-naphthyl) alanine aaRS, the mesh part is indicating the amino acid binding site. B: The amino acid binding pocket, 3-(2-naphthyl) alanine is showed in cyan, the selected residues for random mutation are showed in yellow. (Figure is generated with PyMOL based on the crystal structure of L-3-(2-napthyl)alanine-tRNA synthetase, PDB ID: 1ZH0[38]) 15

Figure 14. A general alternate positive negative selection for the directed evolution of orthogonal 3-(2-naphthyl) alanine aaRS.[37] 17

Figure 15. The one-plasmid duel selection Vectors pRepCM[41] 19

Figure 16. The selection Vector pTAG-GFPx151 20

Figure 17. Vector pBK-NpOH 22

Figure 18. Vector pLEIZ 24

Figure 19. Fluorescence of the purified Z-domain mutant proteins under UV excitation. Protein samples were collected under the expression condition of (1) with NpOH-RS1 but without 2NpOH; (2) with NpOH-RS1 and with 2NpOH; (3) with NpOH-RS2 but without 2NpOH; (4) with NpOH-RS2 and with 2NpOH. 26

Figure 20. Incorporation of 2NpOH into a target protein. (a) SDS-PAGE analysis of Lys7→TAG amber mutant of Z-domain protein expressed under different conditions. Lane 1: molecular mass marker; lane 2: expression with WT *Mj*TyrRS; lane 3: expression with NpOH-RS1 in the presence of 2NpOH; lane 4: expression with NpOH-RS1 in the absence of 2NpOH. The SDS-PAGE gel was stained with SimplyBlue SafeStain reagent. 26

Figure 21. ESI-MS spectrum of 2NpOH incorporated Z-domain protein (expressed with NpOH-RS1): peak a is corresponding to the Z domain protein without the first Met residue, peak b is the acetylated product of peak a, and peak c is the full length protein..... 27

Figure 22. ESI-MS spectrum of 2NpOH incorporated Z-domain protein (expressed with NpOH-RS2): peak a is corresponding to the Z domain protein without the first Met residue, peak b is the acetylated product of peak a, and peak c is the full length protein..... 28

Figure 23. The emission spectrum of the 2NpOH-Z domain protein displayed a maximum wavelength of 423 nm when excited with 335 nm light. 28

Figure 24. Vector pAK3-NpOH 29

Figure 25. A site-specific azo coupling bio-conjugation reaction towards 2NpOH residue. (Figure is generated with PyMOL based on the crystal structure of Z domain protein, PDB ID: 1Q2N[43]) 31

Figure 26. Bio-conjugation reaction on tyrosine residue with diazonium salts. 31

Figure 27. Preparation of diazonium salts 32

Figure 28 Azo coupling with *p*-cresol versus that with 2naphthol 32

Figure 29. The formation of azo adducts between p-cresol and diazonium salts and between 2-naphthol and diazonium salts. Reaction conditions that show selectivity are highlighted in red blocks. 33

Figure 30 Azo adduct is generated only with 2-naphthol but not with any of the aromatic canonical amino acids neither for deoxyribonucleotide triphosphate (dNTPs) 34

Figure 31 Diazonium generated by 4-Aminoacetanilide derivative 34

Figure 32. Azo adduct formed only with 2NpOH but not with tyrosine 36

Figure 33 SDS-PAGE analysis. 1. The protein gel before blue staining; 2 the protein gel after blue staining. A: Tyr-Z reacts with diazonium bearing cyano group; B: Tyr-Z reacts with diazonium bearing methyl group; C: Tyr-Z; D: 2NpOH-Z reacts with diazonium bearing cyano group; E: 2NpOH-Z reacts with diazonium bearing methyl group; F: 2NpOH-Z; G: Protein standard ladder. 37

Figure 34. The ESI-MS spectrum for the reaction of 2NpOH-Z protein. The two major peaks of the spectrum are assigned to the azo adducts of 2NpOH-Z-domain protein. Mass peaks for unmodified 2NpOH-Z Protein are insignificant. 38

Figure 35. The ESI-MS spectrum for the reaction of the control Tyr-Z-domain protein. The MS peaks for the azo adduct of Tyr-Z domain were not observed 39

Figure 36. The UV-visible spectra for the Tyr-Z protein (A) and 2NpOH-Z Protein before (black line) and after (red line) the azo coupling reaction with diazotized p-toluidine. A broad band in the visible region with a λmax at about 520 nm was observed after the azo coupling, for which the 520 nm band can be assigned to the absorption for the expected azo functionality. 39

Figure 37. Kinetic study for the formation of the azo adducts. (A) The growth curves of the azo adduct produced from 140 µM of 2NpOH-Z protein and 30 mM of diazotized 4-aminoacetanilide. (B) The plots of $\ln((OD_{max}-OD_t)/OD_{max})$ versus time and their linear regression analyses. 40

Figure 38. The synthesis of a PEGylated aniline derivative 41

Figure 39. SDS-PAGE analysis for the PEGylation of Z-domain protein through azo coupling. Lane 1: molecular mass marker; lane 2: 2NpOH-Z protein before the coupling reaction; lane 3: the azo coupling product for 2NpOH-Z coupled with the PEG5000 containing reagent, the target protein band is emphasized with an arrow sign; lane 4: the azo coupling between a control Tyr-Z protein and the PEG5000 containing reagent. The SDS-PAGE gel was stained with SimplyBlue SafeStain reagent. 42

Figure 40. Reductive cleavage of the azo adduct. 43

Figure 41. The ESI-MS spectrum for the 2NpOH-Z protein after the azo coupling. 43

Figure 42. The ESI-MS spectrum for the reductive cleavage of 2NpOH-Z protein azo product. Peak a and peak b indicating the unduced azo adducts. Peak c and d indicating the fully reduced product, peak e and f indicating the partial reduced product 44

Figure 43. The fully and partial reduced product by the reductive cleavage of the azo adduct. 44

Figure 44. Proposed reversible protein immobilization strategy 45

Figure 45. Covalent binding the GFP_{UV}-X4-2NpOH to the diazotized agarose bead (Protein structure is generated with PyMOL based on the crystal structure of GFP, PDB ID 1GFL[49]) 46

Figure 46. A three-component Mannich-type coupling reaction 48

Figure 47. Mechanism of this Mannich type reaction 49

Figure 48. The ESI-MS spectrum for the Mannich reaction of 2NpOH-Z protein. Peak a and peak b are indicating the 2NpOH-Z protein with a 12 Da increment of mass. Peak c and d are impurities. 51

Figure 49. The ESI-MS spectrum for the Mannich reaction of Tyr-Z protein. No mass increase was observed. Peak a stands for the mass of unmodified Tyr-Z protein. Peak b and c are contamination. 51

Figure 50. A possible cyclic acetal type of product 52

Figure 51. The ESI-MS spectrum the Mannich reaction of Tyr-Z protein. A: Protein solution after the reaction doped with Tyr-Z protein. B: Protein solution after the reaction. Peak a matched with the unmodified Tyr-

Z protein. Peak b showed a mass increase of 174 Da than the unmodified Tyr-Z protein. 53

Figure 52. The ESI-MS spectrum the Mannich reaction of 2NpOH-Z protein. A: Protein solution after the reaction doped with 2NpOH-Z protein. B: Protein solution after the reaction. Peak a and b matched with the unmodified 2NpOH-Z protein. Peak c and d showed a mass increase of 12 Da than the unmodified 2NpOH-Z protein. Peak d showed a mass increase of 30 Da. Peak f and g were impurities. 54

Figure 53. Crystal structure of Z-domain proteins (PBD ID:1Q2N). 2NpOH[7] is only 3.43 Å away from Lys[38]. 55

Figure 54. A possible Mannich type reaction between 2NpOH residue and Lysine residue with presence of formaldehyde. 55

Figure 55. Vector pET-bZIP 56

Figure 56. Crystal structure of dimerized GCN4 leucine zipper protein(Figure generated is with PyMOL based on the crystal structure of GCN4 Basic Region of Leucine Zipper Protein, PDB ID: 1YSA[55]). 2NpOH mutant site showed in yellow (GCN4-B) and green (GCN4-C), original Lys[22] showed in red. 57

Figure 57. Vector pEVOL-NPOH 58

Figure 58. Vector pET-bZIP-GCN4 A-B 59

Figure 59. Two possible coupling product 60

Figure 60. SDS-PAGE analysis for the formaldehyde-based crosslinking reaction of 2NpOH-Z protein. Lane 1: Tyr-protein before the coupling reaction; lane 2: Tyr-protein after the coupling reaction; lane 3: 2NpOH-protein before the coupling reaction; lane 4: The coupling product of 2NpOH-Z protein; lane 5: molecular mass marker The SDS-PAGE gel was stained with SimplyBlue SafeStain reagent. 60

Figure 61. 2NpOH-Z protein (Figure is generated with PyMOL based on the crystal structure of Z domain Protein, PDB ID: 1Q2N[43]) with 5 lysine residue showing in green. 61

Figure 62. Four different aldehydes were tested in this coupling reaction 62

Figure 63. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with formaldehyde. A mass increase of 12 Da was observed as

expected. No peak for the unmodified protein was observed. 62

Figure 64. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with pyridine-2-carbaldehyde. Peak a and b matched with the mass increase of 89 as expected. Peak c showed a small portion of unreacted 2NpOH-Z protein. 63

Figure 65. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with pyruvaldehyde. Interestingly, a mass increase of 12 Da, indicating a formaldehyde coupling reaction, was observed. 63

Figure 66. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with ethyl glyoxalate. Peak a and b showed a mass increase of 102, which matched the molecular weight of ethyl glyoxalate. 64

Figure 67. Green fluorescent protein (GFP) (Figure is generated with PyMOL based on the crystal structure of an Enhanced Green Fluorescent Protein with 374-nm Absorbing Chromophore, PDB ID 1Z1Q[56]) 66

Figure 68. Mechanism for the chromophore formation in Green Fluorescent Protein.[57] 67

Figure 69. Incorporate 2NpOH to substitute tyrosine in GFP chromophore 68

Figure 70 The crystal structure of A:GFP(+36) (Figure is generated with PyMOL based on the crystal structure of a superfolder Green Fluorescent Protein mutant, PDB ID 4LQU[62]), B: GFPuv (Figure is generated with PyMOL based on the crystal structure of an Enhanced Green Fluorescent Protein with 374-nm Absorbing Chromophore, PDB ID 1Z1Q[56]). The chromophore is showed in red. The nutation site is showed in blue. 69

Figure 71 Vector pTAG-GFP+36 70

Figure 72. Vector pTAG-GFPuv 70

Figure 73. The structure of two unnatural amino acids used in this study 72

Figure 74. Synthesis of N_{ε}-propargyloxycarbonyl-L-lysine (AlK)..... 73

Figure 75. Vector pAK8..... 73

Figure 76. pt2Z238-GX Vector 74

Figure 77 SDS-PAGE gel: the expression of Z-7_{AGG}-38_{UAG} proteins with the presence of two UAAs..... 75

Figure 78. The ESI-MS spectrum for Z-AcF^{7}-AlK^{38} protein. 75

Figure 79. The hydrazone generate from the reaction between DNSH and AcF 76

Figure 80 The copper catalyzed azide-alkyne cyloaddition bwtween an azido-modified PEG and an AlK residue. 76

Figure 81. SDS-PAGE gel: Z-AcF^{7}-AlK^{38} protein was labeled with DNSH and PEG 77

Figure 82. Reversible photocontrol of an enzyme activity. 80

Figure 83. Scheme of a protein array 81

LIST OF ABBREVIATIONS

UAA	unnatural amino acid
PEG	polyethylene glycol
aaRS	aminoacyl-tRNA synthetase
E. coli	*Escherichia coli*
PylRS	pyrrolysyl-tRNA synthetase
Mb	*Methanosarcina barkeri*
Mm	*Methanosarcina mazei*
Mj	*Methanococcus Jannachii*
TyrRS	tyrosyl-tRNA synthetase
NCAA	noncanonical amino acid
2NpOH	2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid
5-FU	5-fluorouracil
cat	chloramphenicol acetyl transferase
uprt	uracil phosphoribosyl transferase
IPTG	isopropyl-β-D-1-thiogalactopyranoside
AcF	p-acetylphenylalanine
AlK	N_{ε}-propargyloxycarbonyl-L-lysine
ESI-MS	electrospray ionization mass spectrometry
DNSH	5-(dimethylamino) naphthalene-1-sulfonyl hydrazine
FRET	Förster Resonance Energy Transfer

Chapter 1

Background on genetically incorporating unnatural amino acid into proteins.

1.1 New chemistries beyond those contained in the natural amino acid can provide advantages not only in protein function enhancement but also in protein studies.

Proteins are the most versatile macromolecules of the cell, and perform a vast array of functions within living organisms. Major examples of the biochemical functions of proteins include catalyzing metabolic reactions; replicating DNA; responding to stimuli; transporting molecules and serving as structural components of cells and organisms.[1]

With the rare exceptions of selenocysteine[2] and pyrrolysine[3], the genetic codes of all known organisms specify only the 20 natural amino acids as its building blocks for protein synthesis. It is the nature's choice after millions years' evolution. Among these natural amino acids, there are several functional groups in their molecular structure, including carboxylic acid, hydroxyl, amino, thiol etc. These functional groups play important roles for protein's unique properties

However, proteins often require more chemistries besides these functional groups to actually function, including cofactors, such as metal ions, thiamine, pyridoxal; and posttranslational modifications such as methylation, glycosylation, and phosphorylation.[4] Moreover, in order to utilize proteins as the basis for new drugs and materials, chemical modification on proteins is needed, so that the enhanced physicochemical, biological or pharmacological properties can be introduced.[5] Additionally, for a better view of understanding the protein structure, functions and behaviors, biophysical probes are often required to link onto proteins for analysis or purposes of control.

From the beginning of the 1980s, scientists have tried several approaches to direct chemistry specifically at macromolecular surfaces.[6] However, chemists' ability to change the protein structure and function is still limited compared to our ability to modulate small molecules.[7] Chemists can synthesize unnatural peptides *in vitro*, by chemical methods such as solid-phase synthesis, native chemical ligation. However, due to the reaction yield, the size of the synthetic peptide is normally limited. Learnt from our mother nature, genetically encoding unnatural amino acids (UAAs)[8] may be an alternate and effective method to add new chemistries to

proteins. Therefore, the creation of organisms with expanded genetic codes will allow us to design and produce proteins with enhanced properties and activities, or to anchor a probe on protein surface to facilitate the study of protein structure, explore the behavior and function of proteins both *in vivo* and *in vitro*.[9]

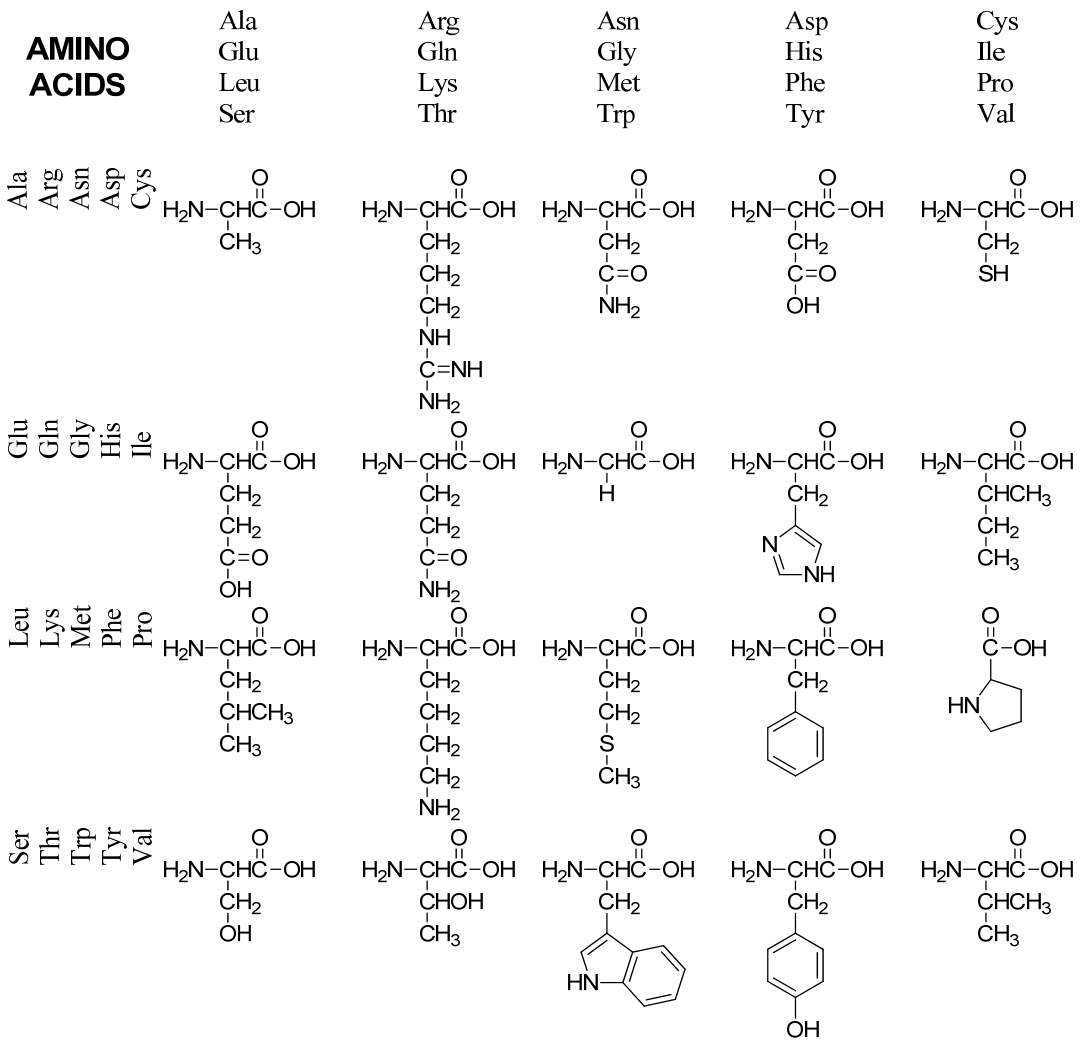


Figure 1. 20 natural amino acids

1.2 Benefits from incorporation of unnatural amino acids with an expanded genetic code

Proteins are important molecules that play essential roles for almost every biological process in all known living things. The complex biological properties of proteins result from polymeric combination of twenty common amino acids. Thus, to study protein properties and synthesize proteins with enhanced or new functions, it is our interest to introduce chemical perturbation into proteins by replacing those

common amino acids with unnatural amino acids, bearing a novel functional group, at a chosen site on the protein.

Peter G. Schultz's group has developed a general technique, which is an effective tool to genetically encode unnatural amino acids with the same high translational fidelity and efficiency characteristic of natural protein biosynthesis.[10] To date, this method has been used to incorporate over 150 unnatural amino acids into proteins in bacteria, yeast and mammalian cells.[11] These unnatural amino acids provide us a pool of candidates for selective post-translational modification, protein ligation, or simply working as a fluorescent probe.

Genetically incorporation of unnatural amino acid into proteins has one major advantage of precise controlling the chemistries at specific site of the protein. Thus, "the inserted one" can be differ from the natural ones, so that it can either provide the unique information around the surroundings at that specific site, or further react with other reagent only at the desired position on the protein surface. In a word, only the chosen one or multiple sites would be active for the latter chemical treatment or measurement, which would be very difficult to realize with traditional chemistry modification.

This methodology provides us the ability to chemically tailor and tag different groups with potential functions. For instance, alkyne or azide containing amino acid can be selectively modified under relative mild condition, through 3+2 cyclo-addition reaction ("click chemistry").[12] Shown in Figure below, with the diversity of R group, the proteins with this amino acid can be utilized to site-specifically link polyethylene glycol molecules, sugars, fluorophores, peptides and other synthetic moieties.[13] Moreover, the polyethylene glycol (PEG) conjugation could be used to extend the serum half-life of proteins, such as fibroblast growth factor 21[14] and human growth hormone[15]. Another good example of the application with unnatural amino acid is that the site-specific incorporation of ^{15}N-, ^{13}C-, or ^{19}F-labeled residues into fatty acid synthetase to use NMR to identify conformational changes that occur upon ligand binding. Thus, unnatural amino acid with novel properties can be quite useful for biochemical studies.

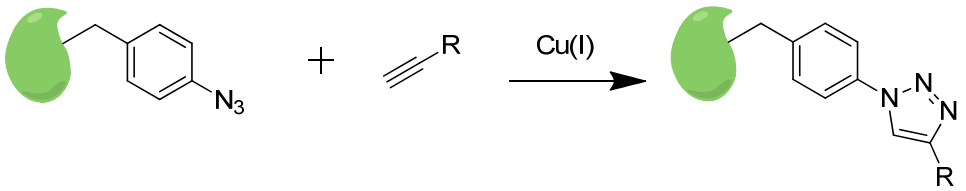


Figure 2. Protein bio-conjugation via Click chemistry on *p*-azido-phenylalanine residue.

To precisely perform selective modification of proteins, some approaches on chemically labeling native amino acid residue have been explored.[16] However, genetically incorporation of unnatural amino acid into proteins can effectively limit

the frequency of target residue in the organism, so that we can expect the selectivity of modification based on unnatural amino acid would be much more site-specific than traditional means. And it is our interest to produce proteins bearing unnatural amino acid with novel activities.

1.3 A general approach for the site-specific incorporation of unnatural amino acids into proteins *in vivo*

1.3.1 Components necessary for the genetically incorporation of UAAs into proteins.

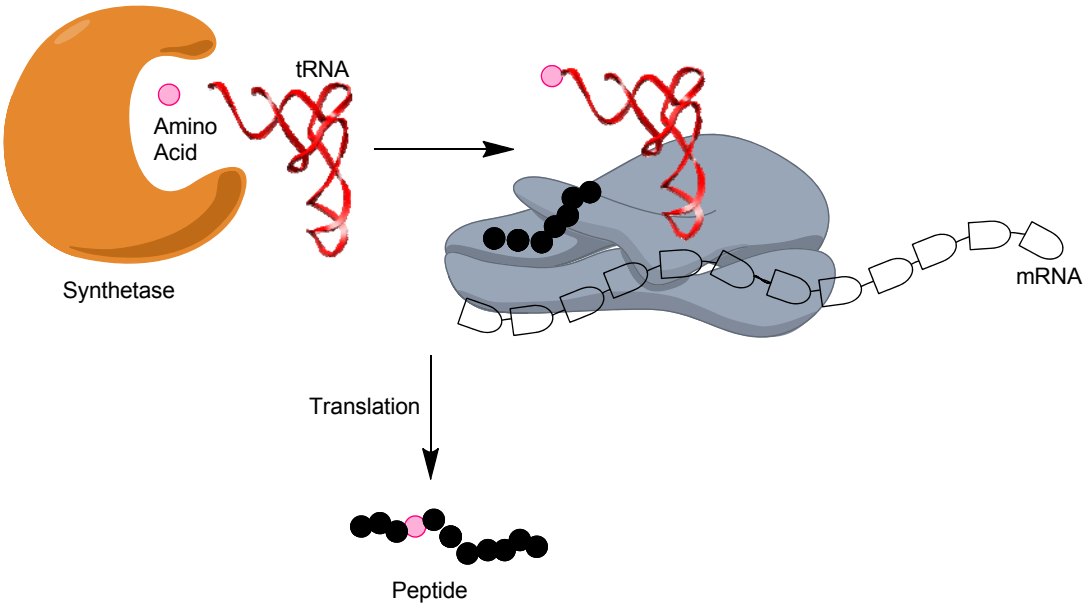


Figure 3. The incorporation of amino acid into proteins.

Similar with adding natural amino acid into peptide, we need several essential things, ribosome machinery, the amino acid, a template-mRNA, tRNA, enzyme and energy, to incorporate UAAs into proteins at specific sites. Notably, there are three important components required, which are the keys to this methodology: first, an unnatural amino acid, which should be not toxic to the cell, can be up-taken by the cell, stay stable and functional inside cells, and cannot be recognized by any endogenous aminoacyl-tRNA synthetase (aaRS); then a special codon that only designates the unnatural amino acid and would not be matched with any endogenous tRNAs; and last, a unique aminoacyl-tRNA synthetase/tRNA pair that specifically charge the unnatural amino acid and can incorporate the UAA into proteins in response to the special codon.[9] To meet with the special criteria these three components required, several rules must be followed.

First of all the UAAs, the UAAs are normally synthesized through bioorganic chemistry strategy. After dissolving into the culture media, the UAAs can be taken up by host cells. It was reported that the aminoacyl binding site of elongation factor Tu (EF-Tu) and the ribosome are highly compatible to a relative wide range of acceptable substrates including many noncanonical amino acids. Moreover, the natural aaRSs are highly specific for their cognate amino acids after the long time evolution, so that, the requirement for the UAAs can be fulfilled.

Secondly, the unique codon may be choose from the stop codons, frameshift codons, or any unused codons. For instance, in *E. coli*, the amber stop codon TAG was most commonly used to encode unnatural amino acids without perturbing the growth of the host; because the TAG codon is the least used among the three stop codons and rarely terminates essential genes, and efficiently translated by amber suppressor tRNAs *in vivo* and *in vitro*.[17] Similar principal applied with the 4-base frameshift codon and some sense codon with low frequency.

Last but not the least, the orthogonality of the aminoacyl-tRNA synthetase/ tRNA pair is needed. The tRNA cannot be recognized by the endogenous aminoacyl-tRNA synthetase of the host, but functions efficiently in translation; the tRNA synthetase only aminoacylates the tRNA with the unique anticodon, but not any of the endogenous tRNAs. Meanwhile, in order to link the unnatural amino acid to the tRNA, the amino acid binding site of aaRS need to be modified so that the unnatural amino acid can fit into it.

1.3.2 Two commonly used orthogonal aminoacyl-tRNA synthetase/ tRNA pairs for *E. coli* host.

To achieve the requirement of the orthogonality, a unique aminoacyl-tRNA synthetase/ tRNA pair is needed. The most obvious method was importing a heterologous aminoacyl-tRNA synthetase/ tRNA pair from a different species, since tRNA recognition by aaRSs is species specific. Therefore, there are two type of aminoacyl-tRNA synthetase/ tRNA pairs were identified as good candidates to serve as orthogonal aaRS/tRNA pair.

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

With the example of the nature to encode pyrrolysine using UAG amber codon, the pyrrolysyl-tRNA synthetase (PylRS) and $tRNA^{Pyl}_{CUA}$ from members of *Methanosarcinacea* was targeted as a good orthogonal aaRS/tRNA pair. PylRS is an archaeal class II aminoacyl-tRNA synthetase, which can directly aminoacylates $tRNA^{Pyl}_{CUA}$ with pyrrolysine. It was reported that these PylRS/ $tRNA^{Pyl}_{CUA}$ pairs, from *Methanosarcina barkeri (Mb)* or *Methanosarcina. mazei (Mm)*, can function well enough in either *E. coli* or mammalian cells[18], and maintain their orthogonality without any cross-reactivity with the native aaRS/tRNA pairs.

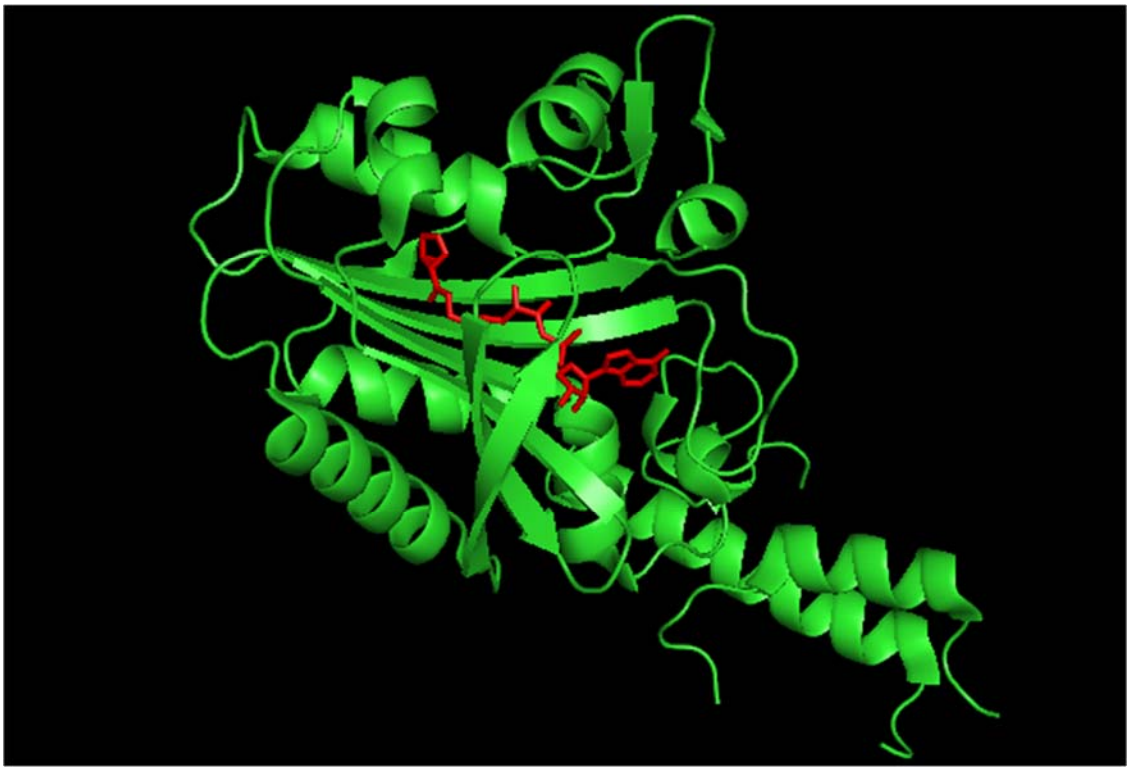


Figure 4. Crystal structure of *Mm* Pyrrolysyl-tRNA synthetase bound to adenylated pyrrolysine and pyrophosphate (Figure is generated with PyMOL based on the crystal structure of Pyrrolysyl-tRNA synthetase, PDB ID: 2ZIM[19])

Besides, this PylRS/tRNA pair is reported to be able to incorporate not only pyrrolysine, but also many of its analogues[20], and with proper modifications and screenings, the side chain binding sites of PylRS can be altered to host many other unnatural amino acids to expand the genetic code.

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

Another important orthogonal aaRS/tRNA pair, which is commonly used, is derived from *Methanococcus Jannaschii(Mj)*, the *Methanococcus Jannaschii(Mj)* tyrosyl-tRNA synthetase (TyrRS) and $tRNA^{Tyr}_{GUA}$. This pair does not interact with the endogenous syntheases of the *E. coli*.[21]

The *Mj* $tRNA^{Tyr}$ is mainly recognized by the synthetase through its discriminator base A73 and the base pair, C1-G72; while the recognition elements of *E. coli* $tRNA^{Tyr}$ are composed by A73, C1-G72, a long variable arm, and the anticodon triplet. Thus, it would allow the *Mj* TyrRS still to recognize the *Mj* $tRNA^{Tyr}$ even if the anticodon is altered to CUA, which would pair with UAG amber codon, since

the anticodon region is only weakly involved in identity determination.[22] So the orthogonal $tRNA^{Tyr}_{CUA}$ could be derived from *Mj* $tRNA^{Tyr}_{GUA}$.

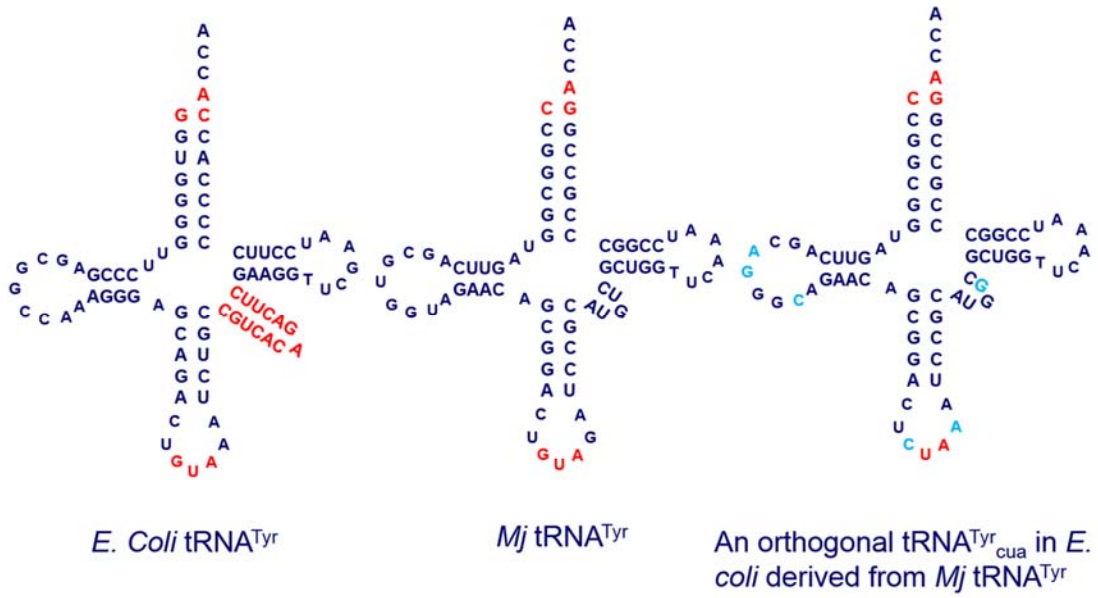


Figure 5. The comparison of *E. coli* $tRNA^{Tyr}$, *Mj* $tRNA^{Tyr}$ and an orthogonal $tRNA^{Tyr}_{CUA}$ in *E. coli, which* was derived from *Mj* $tRNA^{Tyr}$.

Besides, as a class I synthetase, *Mj* TyrRS possesses an amino acid binding pocket, which is highly tolerant for substitutions. Normally upon rational design and screening, the amino acid binding pocket could be tailored with certain mutation to accommodate the designated unnatural amino acid, while its ability to charge endogenous amino acid would be decreased dramatically.[23] Further, *Mj* TyrRS does not possess an editing mechanism and therefore, should not deacylate an unnatural amino acid ligated to the tRNA.

Since the *Mj* TyrRS/$tRNA^{Tyr}_{CUA}$ can be expressed at high levels in its functional form within *E. coli*, this pair has been used widely for unnatural amino acid incorporation. Schultz *et al.* reported a system with both the gene 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. In this system, the suppression efficiencies ranged from 25% to 50% of wild type protein and translational fidelity ≥ 99%. Due to this advantage, a large amount of unnatural amino acid tRNA synthetases have been derived, more than 50 unnatural amino acids were successfully added to the expanded genetic code using this *Mj* TyrRS/$tRNA^{Tyr}_{CUA}$ pair.

1.3.3 Unique codons assigned for the unnatural amino acids

In order to add a new amino acid into proteins genetically, a codon is needed that uniquely specifies this amino acid. As we all know, the 20 natural amino acids are encoded by degenerate 61 triplet codons generated from A, G, C and T, leaving the remaining three codons (TAG, amber; TAA, ochre; and TGA, opal)[13] to serve as termination codon. The redundancy of these codons could be used in this approach. Since the new codons are essential as the orthogonal aaRS/tRNA pair, there are mainly three types of approaches.

Naturally, the nonsense stop codons with low frequency, such as TAG the amber stop codon, TGA the opal stop codon could be borrowed for genetically incorporation of unnatural amino acid, as the nature did to incorporated pyrrolysine.

Another approach for these unique codons are the frameshift quadruplet codons. For example, the AGGA codons and its cognate suppressor tRNAs with expanded anticodon loop were reported to be able to encode unnatural amino acids into proteins.[24]

Last but not the least, reassigning some sense codons with low frequency to unnatural amino acids draws more interests nowadays. However, in order to achieve this, there will be a competition between the natural and unnatural amino acids. And it may affect the natural living system within the cells.

In summary, with all these possible choices, not only the unnatural amino acid would be added into proteins; but also, it would offer a potentially way for incorporating multiple amino acids in to proteins.

1.4 Recent progresses on bio-conjugation reactions.

Site-specific modification of proteins with synthetic probes provides us important tools to discover and engineer the features of proteins both *in vitro* and *in vivo*. For instance, with a fluorescent marker, the protein of interest could be labeled for fluorescence bio-imaging studies in live cells, which allows for real-time detection of their localization, transportation and activities.

To conduct the bio-conjugation on protein under a more mild, or even physiological condition is become an aspiration, in recent trend of chemical biology research. There are several issues, which are often addressed for the development of this type reaction: protein selectivity; labeling site selectivity; possible labeling-site number control; expansion of the labeled amino acid candidates.

Commonly, the selective protein modification methods can be roughly divided into two categories: modification of native functionality (bio-conjugation reaction with

natural amino acid residues), and modification of non-native functionality (bio-conjugation which exploiting the new chemistry of unnatural amino acid).

For the first category, there are several classical approaches to modify the surface of the proteins. Electrophilic modification towards the ε-amino group of lysines, typically by activated esters (NHS-esters), sulfonyl chlorides, isocyanates, and isothiocyanates to afford the corresponding amides, sulfonamides, ureas, and thioureas respectively, is the most common method for protein conjugation. Notably, the N termini of proteins also can be modified by these reagents. The acidic side chain of glutamate and aspartate residues can alternatively be convert into activated ester, generally by water-soluble carbodimides, which can be used for nucleophilic type of bio-conjugation subsequently. Considering the relative high nucleophilicity and low level of occurrence (only 2.3% genome-wide)[25] on the protein surface, cysteine is also a popular candidate.(Figure 6). However, the density of surface exposed these residues limits their utility for the production of well-defined bio-conjugates.

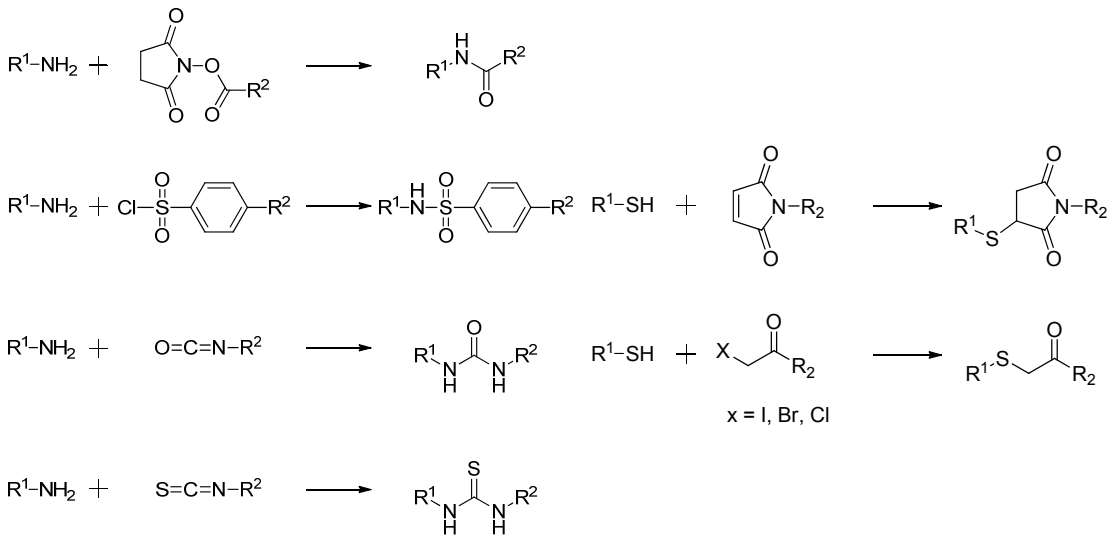


Figure 6. Classical bio-conjugation reaction towards lysine and cysteine residue

In contrast to cysteine and lysine, some aromatic amino acid residues also been exploited for novel selective modification method. Francis and co-workers has reported several reaction towards the tyrosine residue. The tyrosine can be coupling with diazonium salts resulting an azobenzene functionality effectively at pH 9 instead of physiological pH. They also succeeded in modifying the tyrosine residue through a three-component Mannich type reaction with aldehyde and anilines. However, the coupling efficiency is lower than the azo-coupling.

Secondly, genetic incorporation of noncanonical amino acids (NCAAs) into proteins is a powerful method of introducing novel chemistries into proteins at chosen sites, for which it enhances our ability to manipulate protein structures and

functions as well as to study post-translational modifications. Several chemical transformations have been utilized at the sites of NCAAs for selective bioconjugations. For instance, the addition of an azido functionality into proteins provides a perfect position for modifications through either the Staudinger ligation reaction, Cu(I) catalyzed azide-alkyne cycloaddition (CuAAC), or the strain-promoted cycloaddition to conjugate with triarylphosphines, terminal alkynes, or cyclooctynes, respectively.[26] Other examples include the hydrazone or oxime formation from a ketone or an aldehyde group,[27] the cyanobenzothiazole condensation initiated with an 1,2-aminothiol,[28] the photo-induced 1,3-dipolar cycloaddition between tetrazoles and alkenes,[29] the palladium catalyzed cross-coupling with an iodophenyl or a boronophenyl group,[30] and the cycloaddition of a norbornene moiety with tetrazines.[31]

However, these chemical transformations were sometimes hampered by limitations such as reaction speed, toxicity and synthetic accessibility.[16, 32] Besides, protein modification by the technique of genetic incorporation of unnatural amino acid sometimes is limited by the insufficient yield of the genetically engineered proteins and the restricted number of usable cell lines.

1.5 Summary

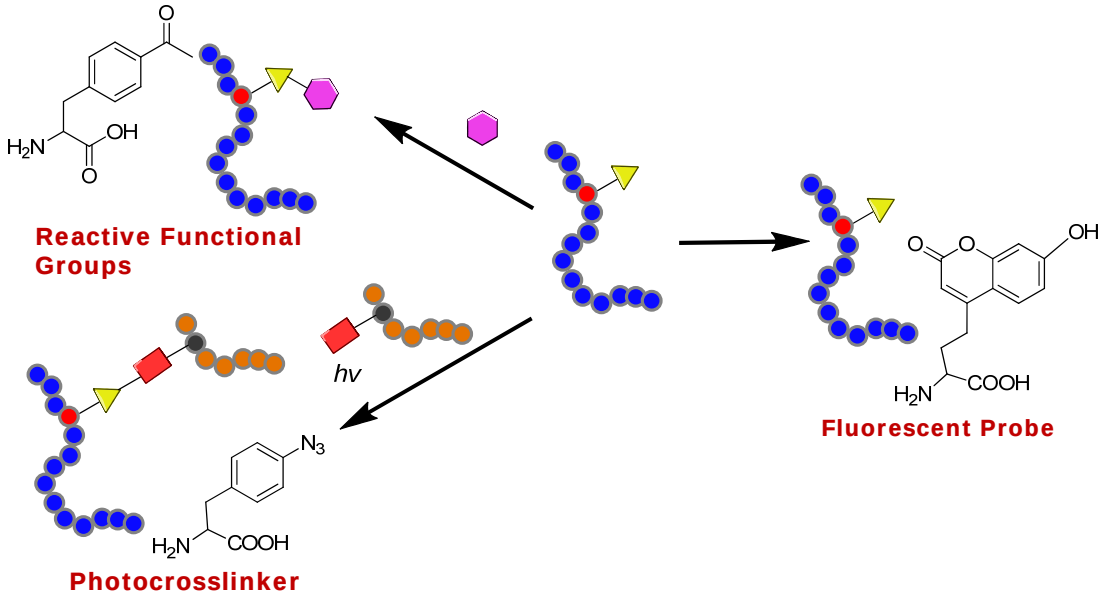


Figure 7. Schematic representation of the uses of genetically encoded UAA side chains, including bio-orthogonal reactive groups, spectroscopic probes, photo-crosslinkers etc.

As I discussed above, through genetic incorporation of unnatural amino acid, we can add new chemistries to proteins, which would be an advantage to study the

protein behavior, to provide site-specific protein modification, and to enhance the protein property.

In this dissertation, we focused on introduce the 2-naphthol chemistry into the expanded genetic code. In addition, this UAA could serve as a site-specific reaction target on the protein surface. Furthermore, the unique fluorescent property and its similar property with tyrosine could be utilized to alter the protein features.

Chapter 2

Genetically incorporate 2NpOH into proteins.

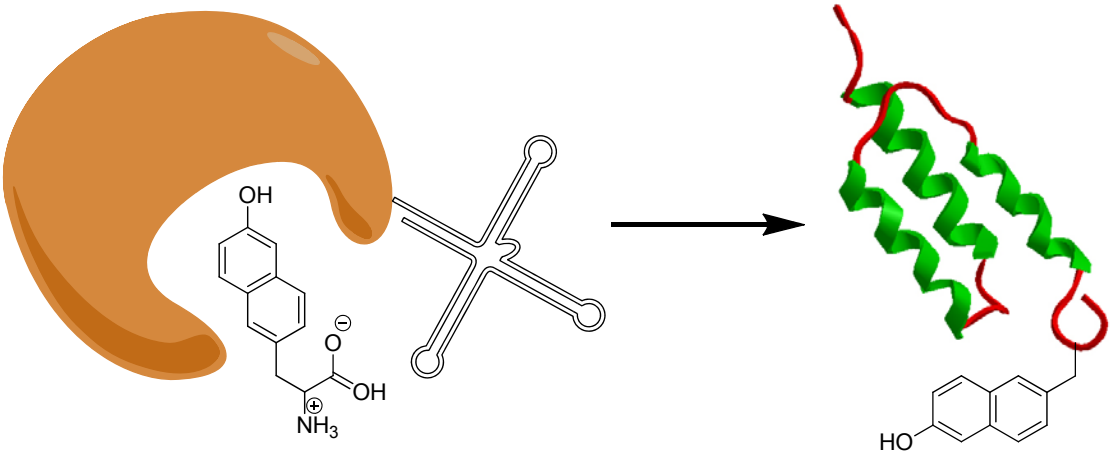


Figure 8. Genetically incorporate 2NpOH into proteins

2.1 Target Unnatural Amino Acid.

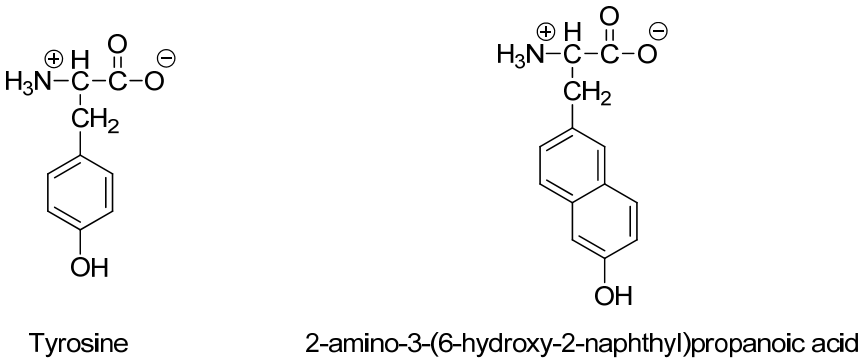


Figure 9. Tyrosine and its 2-naphthol analogue, 2-amino-3-(6-hydroxy-2-naphthyl) propanoic acid.

2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid (2NpOH) (Figure 9) is a 2-naphthol tyrosine analogue, which is a target of our research interest. Structurally it is quite similar to tyrosine, only with one extended aromatic ring. We expect that, once this unnatural amino acid has been incorporated into protein at a site where it is originally a tyrosine residue, the structure change could be limited, so that the function of the protein will be minimally perturbed. Meanwhile, since the relative maxima absorption and emission bands of naphthol are at 320 nm and 400 nm[26c];

the fluorescence behavior of this unnatural amino acid should be unique when incorporated it into peptides or proteins, considering tryptophan has the lowest energy absorption and emission among the natural amino acid at 270-300 nm and 335-350 nm depending in the polarity of local environment.[33] So that, it means this 2-naphthol tyrosine analogue can work as a potential fluorescent probe. Additionally, several selective modification methods focused on the tyrosine residue have been developed based on aromatic substitution reactions.[16] Since the aromaticity of naphthalene is lower than benzene, the electron density on α-position of naphthalene system is higher than benzene, which makes it more reactive for aromatic substitution reactions. The orthogonal reactivity can be utilized to develop new bio-conjugation methods on this 2-naphthol tyrosine analogue.

2.2 Synthesis of 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid

In order to evolve the aminoacyl tRNA synthetase that would charge 2NpOH to allow incorporation of this unnatural amino acid into proteins, synthesis of 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid, the unnatural amino acid, is needed for the screening process. Two synthetic routes were tried out.

2.2.1 One-pot Synthesis

According all the references, the synthesis starts with 6-methoxy-2-naphthaldehyde. The starting molecule was prepared as follows (Figure 10)[34]. 2-bromo-6-methoxynaphthalene was lithiated with n-BuLi and treated with DMF give the desired aldehyde with 85% yield after purification.

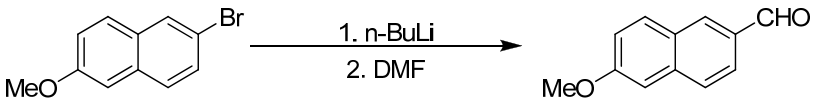


Figure 10. Synthesis of 6-methoxy-2-naphthaldehyde

Samet and co-worker have reported a one-pot synthesis route. (Figure 11)[35]. A mixture of N-acetylglycine, 6-methoxy-2-naphthaldehyde and Sodium acetate in acetic anhydride was heated for 30 min at 110 °C. Red phosphorus and more Ac_2O were added to the reaction mixture, followed by a dropwise addition of 47% HI. The resulting suspension was refluxed for 4 hours; and then followed by workup and purification. However, this one-pot synthesis did not work out well due to the fact that it needs intensive effort to purify the product.

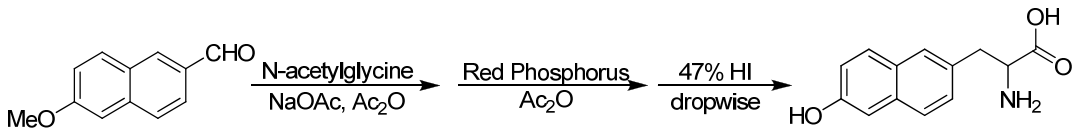


Figure 11. One-pot synthesis of 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid

2.2.2 Multiple Steps Synthesis

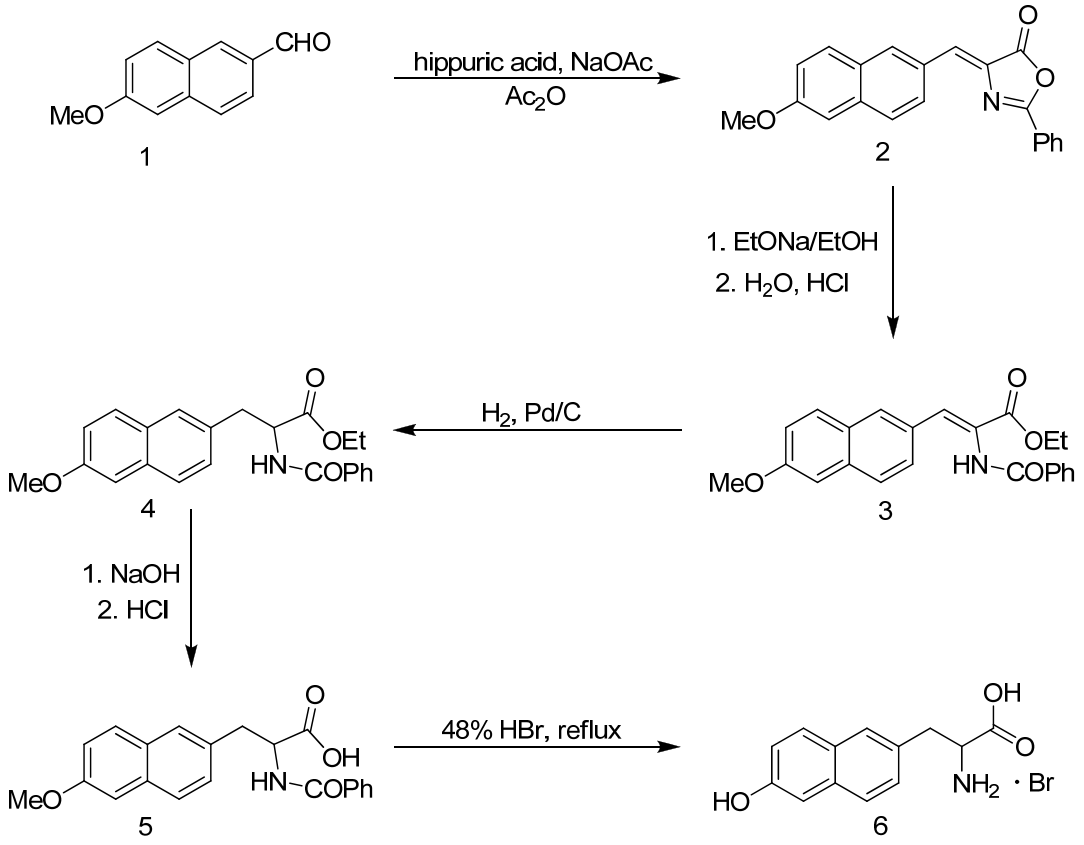


Figure 12. Multiple steps synthesis strategy

Since we cannot get the pure unnatural amino acid with limiting step of purification from the one-pot synthesis, then we tried another approach that involves a multi-step synthesis reported by Mclaughlin (Figure 12)[36]. The strategy started from 6-methoxy-2-naphth-aldehyde we prepared previously. After reacting with hippuric acid in a mixture of anhydrous sodium acetate and acetic anhydride, a crude oxazolone was obtained in 75% yield. Without any purification, the oxazolone ring was opened with sodium ethoxide, which gave crude product **3** in almost 100% yield. Under the catalysis of palladium/activated carbon, the hydrogenation of **3** gave **4** in 70% yield. With sodium hydroxide, the ester bond in **4** was hydrolysised giving bold format in 90% yield. Concomitant deprotection of the methyl ether and

benzoyl groups with refluxing 48% hydrobromic acid gave the hydrobromide of 3-(6-hydroxy- naphth-2-yl)alanine in about 40% yield. The overall yield from **1** was about 20% (racemic mixture).

2.3 Selection of tRNA Synthetase that charges the unnatural amino acid to its cognate tRNA

In our approach, *E. coli* is the host cell, which provides the protein translational machinery. The orthogonal tRNA/aminoacyl-tRNA synthetase (aaRS) in *E. coli* was originally derived from a tyrosyl-tRNA synthetase (TyrRS)/$tRNA^{Tyr}$ pair from the archaea *Methanococcus jannaschii* (*Mj*). The anti-codon loop of tRNA was altered to CUA (corresponding to the amber codon UAG).[37] To optimize the substrate specificity of the orthogonal aminoacyl-tRNA synthetase so that it recognizes the desired unnatural amino acid but not any endogenous amino acids, a library of active-site mutants was generated based on the crystal structure of the aaRS. This library was then subjected to a series of rounds of screen process to identify the synthetase with high fidelity and efficiency to incorporate UAA.[10]

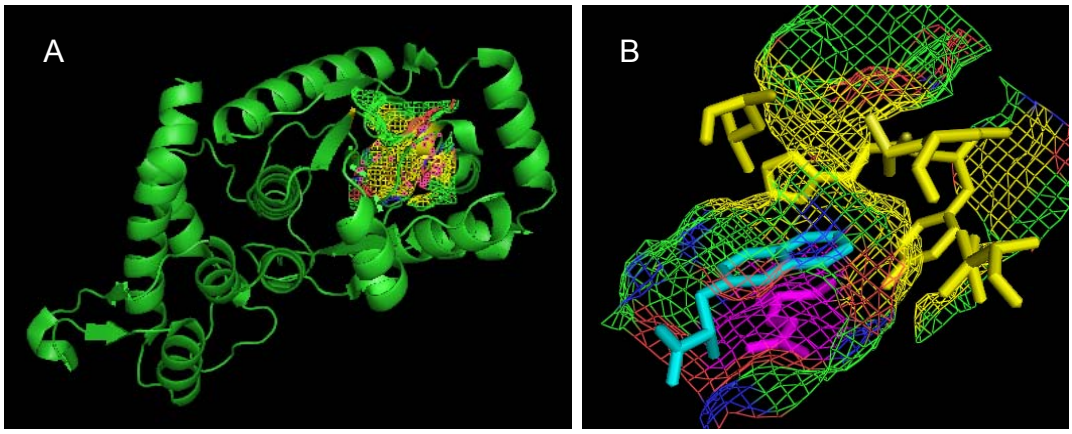


Figure 13. A: The crystal structure of 3-(2-naphthyl) alanine aaRS, the mesh part is indicating the amino acid binding site. B: The amino acid binding pocket, 3-(2-naphthyl) alanine is showed in cyan, the selected residues for random mutation are showed in yellow. (Figure is generated with PyMOL based on the crystal structure of L-3-(2-napthyl)alanine-tRNA synthetase, PDB ID: 1ZH0[38])

P. G. Schultz and coworker have successfully incorporated the unnatural amino acid 3-(2-naphthyl)alanine into proteins by this methodology.[39] Since the structural similarity between 3-(2-naphthyl)alanine and 3-(6-hydroxy- naphth-2-yl)alanine, our group have designed and created a library based on the crystal structure of an aaRS specifies 3-(2-naphthyl)alanine(Figure 13).

This *Mj*TyrRS mutant (aaRS) that selectively charges with 3-(2-naphthyl)alanine was chosen as the starting point to construct the library.[38] Random mutations were

created at the chosen sites of the synthetase gene (a gift from the Schultz's group) through regular PCR reactions and then overlapping PCR reactions with the following primers (with the mutation sites shown in bold):

RS1f: 5'-GAATCCCATATGGACGAATTTG

RS1r: 5'-GTATTTTACCACTTGGTTCAAAACCTAT**MNN**AGCAGATTTTTCATC TTTTTTTAAAAC

RS2f: 5'-TTTGAACCAAGTGGTAAAATACATTTAGG

RS2r: 5'-GGTTTAAATAGGCGTGTAAATCAGCTAA**MNN**TATAATTATATCAAAT CCAGC

RS3f: 5'-GCTGATTTACACGCCTATTTAAACC

RS3r: 5'-AACATCAACGCCCTG**MNNMNN**AGCCGGATTAACCTGCATTATTGG

RS4f: 5'-CAGGGCGTTGATGTT**NNK**GTTGGAGGGATGGAGCAGAG

RS4r: 5'-TTGAAACTGCAGTTATAATCTCTTTC

The mutations at residues Pro^{158}, Ile^{159} and Gln^{162}, compare to original *Mj* TyrRS, were preserved in this library. Five amino acid residues possibly related to the binding of the unnatural amino acid, were selected for random mutation as the followings: Leu^{32}, Leu^{65}, His^{160}, Tyr^{161} and Val^{167}, to accommodate the hydroxyl group in 2NpOH. In addition, an Arg^{286} mutation was introduced and maintained in the library to substantially increase the amber codon suppression rate.[40] To identify the synthetase that specifies 3-(6-hydroxy-naphth-2-yl) alanine, several selection methods were carried out as described in the following sections.

The final PCR product was digested with restriction enzymes Nde I and Pst I (New England Biolabs) following the recommended procedures. After purification, it was then ligated into a pBK vector that was pre-cut with the same restriction enzymes so that the final plasmid contains a kanamycin resistant gene and the Gln S promoter for the library aaRS gene. The ligation reaction was carried out at the scale of 2 g of insert DNA and the molar ratio of insert to vector is 3 to 1. T4 DNA ligase (New England Biolabs) was used as the enzyme and the ligation reaction was carried out at 16 °C for 16 hours. After ethanol precipitation purification protocol, the DNA product was dissolved in 10 μL water, and was then transformed into Top 10 competent cells through electroporation. The cells were incubated at 37 °C for 1 hour for recovery. Subsequently, 2 μL of the cell solution were saved and underwent series dilution to be applied on LB-Agar-kanamycin plate for diversity determination. The remainder of the cells (roughly 3 mL) was added into a 150 mL 2YT with kanamycin and was incubated overnight at 37 °C under shaking. The library DNA was collected using Qiagen's Maxiprep Kit

(QIAGEN™). The diversity of the library was roughly 1.2×10^7 members, which was based on the numbers of colonies (81 colonies) displayed on the LB-Agar-kanamycin plate with transformed cells that were diluted by a hundred fold and then calculated by the extrapolation from 2 L cells to 3 mL cells.

2.3 Evolution of the aaRS that charges 2NpOH

Once the aaRS library was ready, a series screening process was carried out to identify the mutant aaRS, which could link the unnatural amino acid to its corresponding tRNA with high fidelity and efficiency.

2.3.1 A general positive/negative selection system

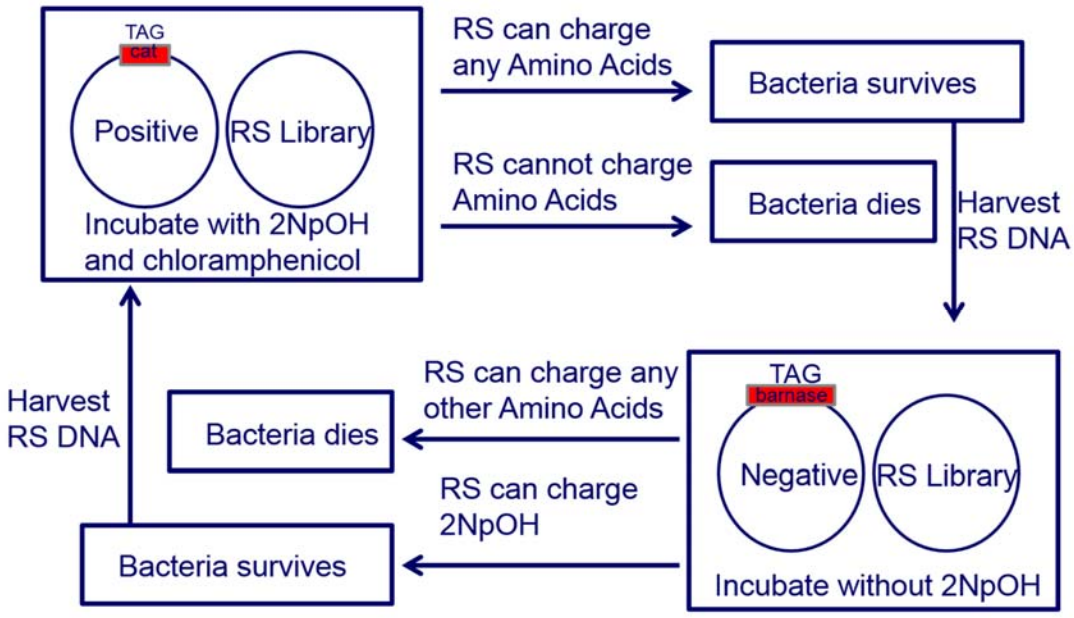


Figure 14. A general alternate positive negative selection for the directed evolution of orthogonal 3-(2-naphthyl) alanine aaRS.[37]

The general idea of this approach is shown in Figure 14. To identify synthetase variants specifying the unnatural amino acids, the libraries were first transformed into cells containing chloramphenicol acetyl transferase (cat) with an TAG codon at a permissive site, and grown in media containing chloramphenicol and the unnatural amino acid. Cells could survive only if they contain the synthetase variants that incorporate either the unnatural or any natural amino acid in response to the amber codon. Harvest the clones, and extract the synthetase plasmid, then transferred into cells containing a toxic barnase gene with three amber mutations at permissive cites, and grown in the absence of unnatural amino acid. By this means, it could rule out the synthetase that charges endogenous amino acids, with

which the cells would produce full-length barnase protein and die.

Positive and Negative competent cell were prepared separately with similar procedure. The positive or negative selection vector was transformed into DH10B *E.coli* cells. A single colony from the transformation was grown overnight at 37 °C in 2.5 mL of LB media supplemented with chloramphenicol and tetracycline (positive) or chloramphenicol (negative) to obtain the starter saturate culture. The culture then was diluted into a 200 mL LB medium containing 34 μg/mL chloramphenicol (for positive competent cell with extra 30 μg/mL tetracycline).When the culture reached OD600=0.4-0.6, the culture was let sit on ice cooling for 60 minutes, then was centrifuged for 20 min at 3500 rpm at 4 °C, the supernatant was discarded. Subsequently, the pelleted cells were re-suspended gently by shaking with 100 mL ice-cold distilled water, followed by centrifugation again. The pelleted cell was re-suspended by 50 mL ice-cold 10% glycerol and then upon centrifugation. The washing with glycerol process was repeated. Finally, the competent cell was re-suspend with 1 mL 10% glycerol. The competent cell was frozen by liquid nitrogen and store under -80 °C.

Then the library was transformed in positive competent cell, subsequently the culture was plated on minimal media plates containing 1% glycerol and 0.3 mM leucine (GMML) with 12 μg/ml tetracycline, 25 μg/ml kanamycin, 60 μg/ml of chloramphenicol, and 1 mM 2NpOH. After incubation at 37°C for 40 hours, colonies were pooled, and plasmids were extracted and purified by gel electrophoresis. Then the resulting library plasmid was transformed into negative competent cell, and was grown on LB plate containing 0.2% arabinose, 50 μg/ml kanamycin, and 34 μg/ml chloramphenicol. Cells were harvest from plate 16 hours later, the plasmid was purified for further rounds of selection.

Repeated rounds of positive and negative selections were carried out, but no desired synthetase was identified.

The general positive/negative selection needs DNA extraction and purification between each round. Unfortunately, we realize that it is not that easy to achieve high yield in these steps. It is very likely that some of the promising synthetase candidates might lose during these purification steps.

2.3.2 An one-plasmid selection system

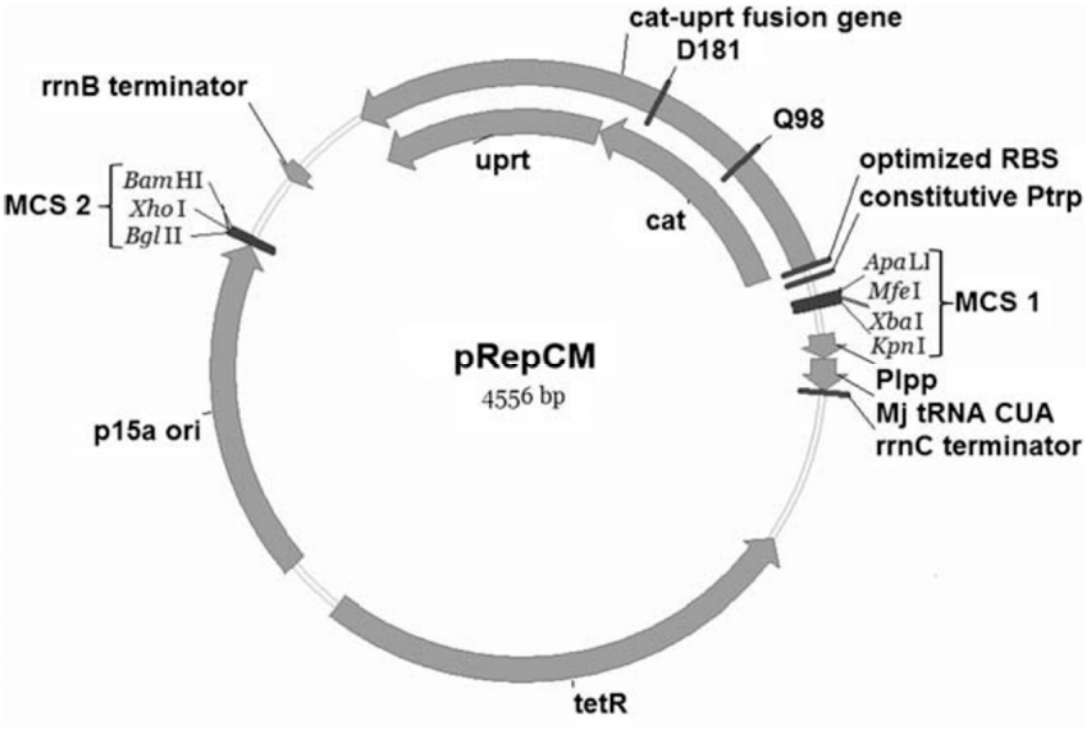


Figure 15. The one-plasmid duel selection Vectors pRepCM[41]

Peter Schultz group has reported this selection system[41]. The vector contains a chloramphenicol acetyl transferase (cat)-uracil phosphoribosyl transferase (uprt) fusion gene, which can function as a dual positive/negative selection marker. The Q^{98} site was substitute with a TAG codon. Similar to the general selection in **2.3.1**, the library DNA was transformed into the host cell with this selection vector and the cell were grown in the media with chloramphenicol and the unnatural amino acid. The survivors contain the synthetase variants that charge either the unnatural or any endogenous amino acid. To apply the negative selection, cells were harvested and simply re-plated on the media with the presence of 5-fluorouracil (5-FU) and the absence of the unnatural amino acid. Since uracil phosphoribosyl transferase converts 5-FU to 5-fluoro-dUMP, which inhibits thymidylate synthase, causing cell death. The synthetase variants incorporate canonical amino acids can be screen out by this way. Alternate positive and negative selections with increasing concentration of chloramphenicol and 5-FU were applied.

After 6 rounds of selections, 53 single colonies were picked up; however, no synthetase shows the promising specificity test by GFP protein expression with a TAG codon at the permissive site of GFP genes.

2.3.3 Combination of two selection system.

With the one-plasmid selection, it turned out that it was not easy to get single colony from the positive selection. We thought the efficiency of the negative selection is not so good, that we cannot effectively rule out the synthetase that charges endogenous amino acid. And Schultz and coworker have pointed out that, after several rounds of selection, there might be some undesired mutation happened[41]. Therefore, we also consider limiting the rounds of selections.

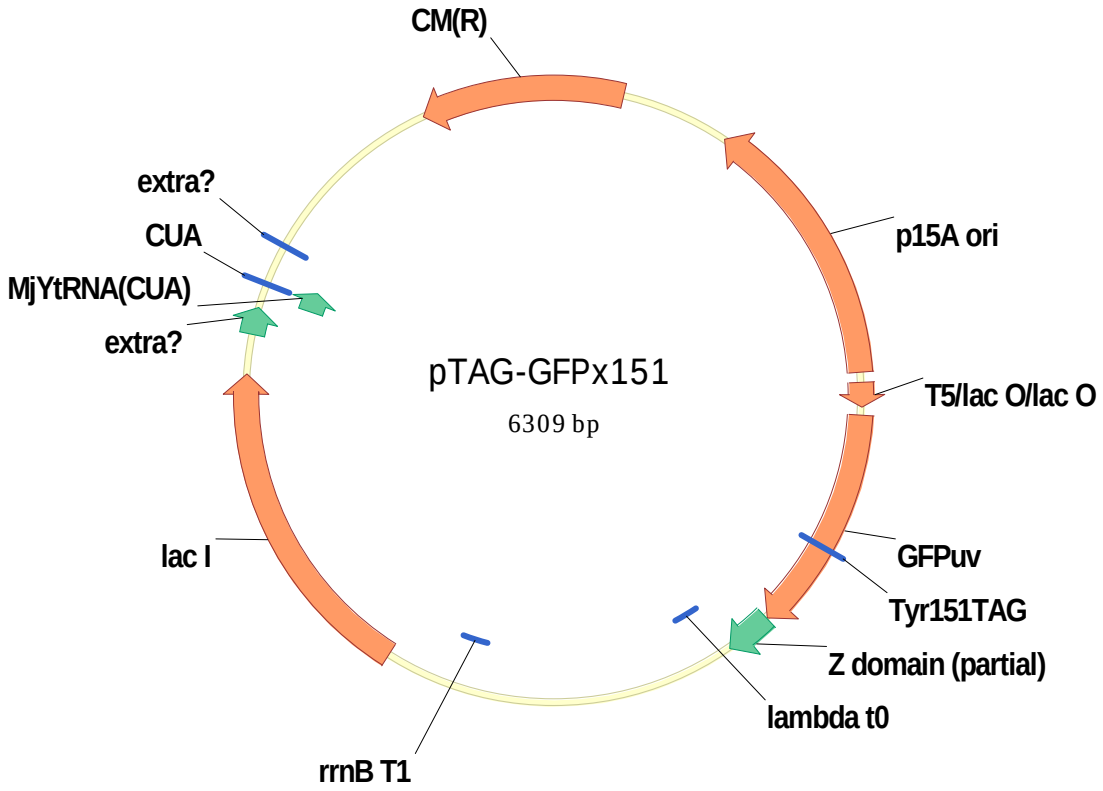


Figure 16. The selection Vector pTAG-GFPx151

The *Mj*TyrRS library DNA was transformed into *E. coli* GH371 host. The cells were grown overnight at 37°C on an LB-Agar medium with tetracycline, kanamycin, 5-fluorouracil for negative selection; then were harvested and grown for a day at 37 °C on a GMML-Agar medium with chloramphenicol, tetracycline, kanamycin and 1mM 2NpOH for positive selection; and again were harvested and grown overnight at 37 °C on an LB-Agar medium with tetracycline, kanamycin, 5-fluorouracil for another round of negative selection. After that, the cell colonies then were harvested and the tRNA/aminoacyl-tRNA synthetase (aaRS) variants DNA plasmids were extracted by using QIAprep Spin Miniprep Kit (QIAGEN™) and purified by enzyme digestion instead of electrophoresis in order to limit the loss of the desired plasmid. For a further round of negative selection, the DNA was transformed into the Top 10 host cell with vector pTAG-GFPx151, which has a TAG

codon at the Tyr[151] site in GFP gene. The cells were grown in the media with the required antibiotics but without 2NpOH. Cell colonies that didn't show any green fluorescence were picked, since the synthetases charge endogenous amino acid will produce full-length GFP and meanwhile the most synthetases cannot charge any amino acids were supposed been screened out during the one-plasmid selections. The aaRS DNA was extracted and purified, and then transformed into a DH10B host cell containing chloramphenicol acetyl transferase (cat) with a TAG codon for a final round of positive selection. The cells then were grown in GMML-Agar medium with chloramphenicol, kanamycin and 1mM 2NpOH. Subsequently 10 colonies were picked up, and inoculated, respectively, into 4mL LB medium with chloramphenicol and kanamycin. Each of the ten variants of aaRS DNA was extracted, purified, and then analyzed by DNA sequencing.

Two unique mutants, namely NpOH-RS1 and NpOH-RS2, were identified (Table 1). DNA sequencing revealed the following mutations in the evolved synthetases compared to the wild-type MjTyrRS: Y32E, L65T, D158S, I159A, H160P, Y161T, L162Q, A167W and D286R in NpOH-RS1; Y32E, L65V, K90E, I159A, H160W, Y161G, L162Q, A167I and D286R in NpOH-RS2. Interestingly, the D158S mutation in NpOH-RS1 and the K90E mutation in NpOH-RS2 were not originally included in the library. They were possibly acquired by self-mutation during the selections. Similar observations have been reported for the selection system.[41]

Table 1 Mutation sites in the evolved aaRS that selectively charge with 2NpOH

Synthetase	Frequency	Leu[32]	Leu[65]	Lys[90]	Pro[158]	Ile[159]	His[160]	Tyr[161]	Leu[162]	Val[167]	Asp[286]
NpOH-RS1	8	Glu	Thr	Lys	Ser	Ala	Pro	Thr	Gln	Trp	Arg
NpOH-RS2	2	Glu	Val	Glu	Pro	Ala	Trp	Gly	Gln	Ile	Arg

Amino acid sequence for NpOH-RS1 (mutation sites highlighted in red):

MDEFEMIKRNTSEIISEEELREVLKKDEKSA**E**IGFEPSGKIHLGHYLQIKKMIDL
QNAGFDIII**T**LADLHAYLNQKGELDEIRKIGDYNKKVFEAMGLKAKYVYGSEFQ
LDKDYTLNVYRLALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVN**SAPTQ**G
VDV**W**VGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIAV
DDSPEEIRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRPEKFGGDLTVNSY
EELESLFKNKELHPM**R**LKNAVAEELIKILEPIRKRL

Amino acid sequence for NpOH-RS2 (mutation sites highlighted in red):

MDEFEMIKRNTSEIISEEELREVLKKDEKSA**E**IGFEPSGKIHLGHYLQIKKMIDLQ
NAGFDIII**V**LADLHAYLNQKGELDEIRKIGDYN**E**KVFEAMGLKAKYVYGSEFQLDK
DYTLNVYRLALKTTLKRARRSMELIAREDENPKVAEVIYPIMQVNP**AWGQ**GVDV**I**
VGGMEQRKIHMLARELLPKKVVCIHNPVLTGLDGEGKMSSSKGNFIAVDDSPEE
IRAKIKKAYCPAGVVEGNPIMEIAKYFLEYPLTIKRPEKFGGDLTVNSYEELESLFK
NKELHPM**R**LKNAVAEELIKILEPIRKRL

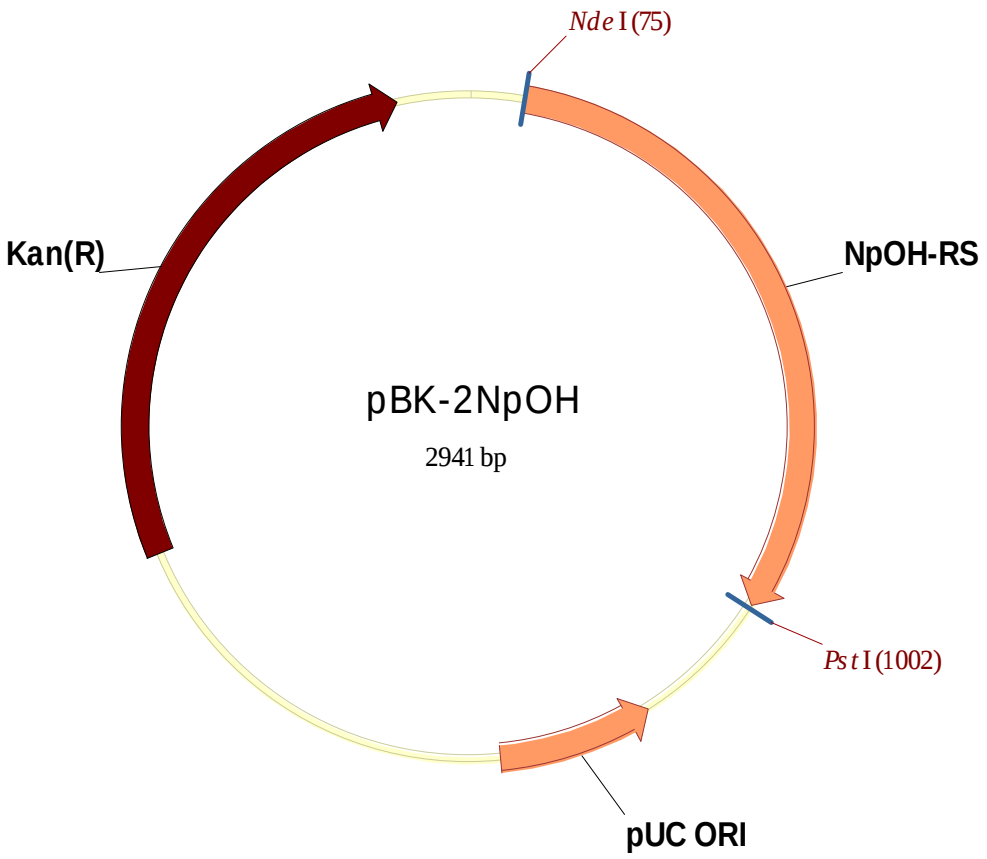


Figure 17. Vector pBK-NpOH

The DNA sequence of pBK-NpOH:

ctcgggagttgtcagcctgtcccgcttataagatcatacgccgttatacgttgtttacgctttgaggaatcccatatgg
acgaatttgaaatgataaagagaaacacatctgaaattatcagcgaggaagagttaagagaggttttaaaaaa
agatgaaaaatctgctgagataggttttgaaccaagtggtaaaatacatttagggcattatctccaaataaaaaa
gatgattgatttacaaaatgctggatttgatataattataactttagctgatttacacgcctatttaaaccagaaagga
gagttggatgagattagaaaaataggagattataacaaaaaagtttttgaagcaatggggttaaaggcaaaata
tgtttatggaagtgaattccagcttgataaggattatacactgaatgtctatagattggctttaaaaactaccttaaaa
agagcaagaaggagtatggaacttatagcaagagaggatgaaaatccaaaggttgctgaagttatctatccaa
taatgcaggttaattcggctcctactcagggcgttgatgtttgggttggagggatggagcagagaaaaatacaca
tgttagcaagggagcttttaccaaaaaaggttgtttgtattcacaaccctgtcttaacgggtttggatggagaagga
aagatgagttcttcaaaagggaattttatagctgttgatgactctccagaagagattagggctaagataaagaaa
gcatactgcccagctggagttgttgaaggaaatccaataatggagatagctaaatacttccttgaatatcctttaac
cataaaaaggccagaaaaatttggtggagatttgacagttaatagctatgaggagttggagagtttatttaaaaat
aaggaattgcatccaatgagattaaaaaatgctgtagctgaagaacttataaagattttagagccaattagaaag
agattataactgcagtttcaaacgctaaattgcctgatgcgctacgcttatcaggcctacatgatctctgcaacaca
ccgagcccgcgtgcttttgcaggccggataaggcgttcgcgccgcatccggcaagaaacagcaaacaatcca
aaacgccgcgttcggcggcgtttttctgcttttcttcgcgaattaattccgcttcgcaacatgtgagcaaaaggcca
gcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggctccgcccccctgacgagcat

cacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcgtttccccct
ggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaag
cgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgca
cgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacg
acttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagtt
cttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacct
tcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagca
gcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaa
cgaaaactcacgttaagggattttggtcatgagttgtgtctcaaaatctctgatgttacattgcacaagataaaaat
atatcatcatgaacaataaaactgtctgcttacataaacagtaatacaaggggtgttatgagccatattcaacggg
aaacgtcttgctcgaggccgcgattaaattccaacatggatgctgatttatatgggtataaatgggctcgcgataat
gtcgggcaatcaggtgcgacaatctatcgattgtatgggaagcccgatgcgccagagttgtttctgaaacatggc
aaaggtagcgttgccaatgatgttacagatgagatggtcagactaaactggctgacggaatttatgcctcttccga
ccatcaagcattttatccgtactcctgatgatgcatggttactcaccactgcgatccccgggaaaacagcattcca
ggtattagaagaatatcctgattcaggtgaaaatattgttgatgcgctggcagtgttcctgcgccggttgcattcgatt
cctgtttgtaattgtccttttaacagcgatcgcgtatttcgtctcgctcaggcgcaatcacgaatgaataacggtttggt
tgatgcgagtgattttgatgacgagcgtaatggctggcctgttgaacaagtctggaaagaaatgcataagcttttg
ccattctcaccggattcagtcgtcactcatggtgatttctcacttgataaccttatttttgacgaggggaaattaatag
gttgtattgatgttggacgagtcggaatcgcagaccgataccaggatcttgccatcctatggaactgcctcggtga
gttttctccttcattacagaaacggctttttcaaaaatatggtattgataatcctgatatgaataaattgcagtttcatttg
atgctcgatgagtttttctaatcagaattggttaattggttgtaacactggcagagcattacgctgacttgacgggac
ggcggctttgttgaataaatcgaacttttgctgagttgaaggatc

2.4 Protein Expression

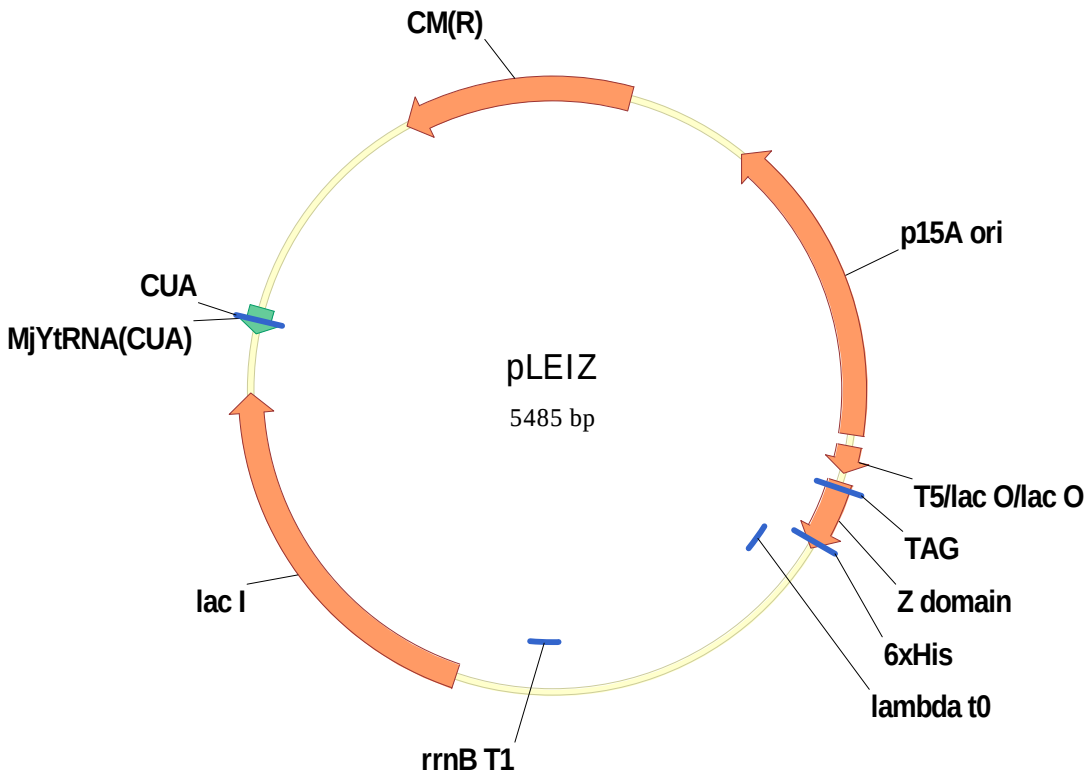


Figure 18. Vector pLEIZ

To determine the ability of the evolved synthetase to incorporate this unnatural amino acid 2NpOH into proteins with high efficiency and fidelity, an amber stop codon was substituted at a permissive site (Lys^7) in the gene for a mutant Z domain protein with a C-terminal 6×His-tag,[42] was expressed in *E. coli* in the presence of the selected tRNA/aaRS pair.

Amino acid sequences for the Z domain protein (* denotes the unnatural amino acid, 2NpOH):

Control Z-domain protein (Tyr-Z):

MTSVDN**Y**INKEQQNAF**Y**EILHLPNLNEEQRDAFIQSLKDDPSQSANLL

AEAKKLNDAQAPKGSHHHHHH

Mutant Z-domain protein (2NpOH-Z):

MTSVDN***INKEQQNAF**Y**EILHLPNLNEEQRDAFIQSLKDDPSQSANLL

AEAKKLNDAQAPKGSHHHHHH

The synthetase gene NpOH-RS1 was transformed into host *E. coli* with vector pLEIZ, which has a TAG codon at the 7th position of Z-domain protein. The transformed cells were grown overnight at 37°C on an LB-Agar plate containing chloramphenicol and kanamycin. A single colony was picked and was inoculated into 5 mL LB medium with chloramphenicol and kanamycin. The cell culture was incubated at 37°C overnight until saturation being reached. The 5mL saturated culture was subsequently inoculated into 100 mL LB medium with chloramphenicol, kanamycin and 1mM 2NpOH. The resulting culture was incubated at 37°C until the optical density at 600 nm reaching 0.5. Subsequently, 1 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG) was added to the culture to induce the protein over-expression process. The culture was kept incubating and shacking overnight. Next, the cell was harvested by centrifugation at 6,000 rpm for 15 min at 4°C. The resulting cell pellet was re-suspended in 5mL BugBuster™ Protein Extraction Reagent (Millipore) with the presence of benzonase and lysozyme. The mixture was placed on a shaking platform for 20 min at room temperature. Subsequently, the mixed solution was split into several microcentrifuge tubes and applied to centrifuge at 14,000 rpm for 10 minutes at 4°C. After that, the supernatant (cell lysate) was collected.

The 5 mL cell lysate was added with 1.25 mL of Ni-NTA resin, and gently mixed on a shaking platform at 4°C for 60 min. The lysate-Ni-NTA mixture was loaded into a column and run through. Then the agarose bead was washed twice with 5 mL of wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0). The protein solution was eluted with 2.5 mL of elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The small molecules was then removed upon centrifugation in Amicon™ Ultra centrifugal filter (>3kD) after series rounds of concentration and dilution with H_2O. Protein samples were subjected for SDS-PAGE, and Mass spectroscopy analyses.

As the negative control, Z domain expression was carried out with the same protocol except that no 2NpOH was used; while the positive control experiment, the expression of Tyr-Z domain was carried out with the same protocol except that the wild type *M*jTyrRS pair was used instead of NpOH-RS1 so that the TAG codon would be translated as tyrosine. 2NpOH was not used under this situation.

Due to the unique fluorescent behavior of the 2NpOH residue, the solution of 2NpOH incorporated protein itself would show strong fluorescence under UV excitation, which indicating the successfully incorporation of the unnatural amino acid. (Figure 19)

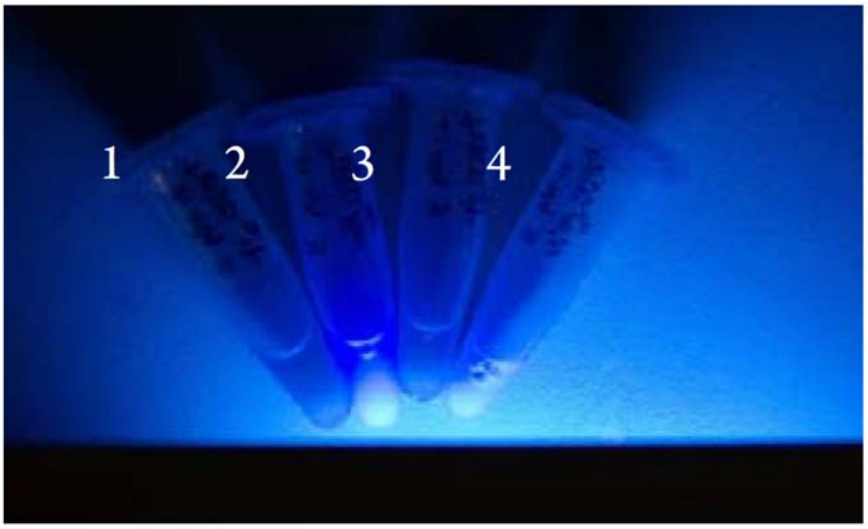


Figure 19. Fluorescence of the purified Z-domain mutant proteins under UV excitation. Protein samples were collected under the expression condition of (1) with NpOH-RS1 but without 2NpOH; (2) with NpOH-RS1 and with 2NpOH; (3) with NpOH-RS2 but without 2NpOH; (4) with NpOH-RS2 and with 2NpOH.

Then the protein solution was analyzed by SDS-PAGE electrophoresis. No Z-domain protein was expressed in absence of 2NpOH, while there was a significant band for 2NpOH-Z protein.

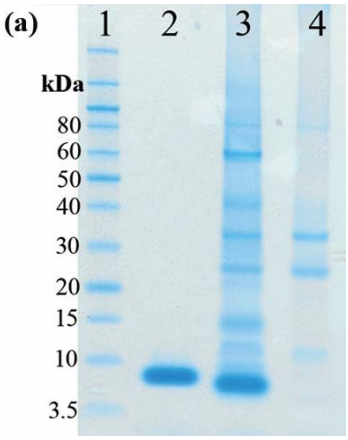


Figure 20. Incorporation of 2NpOH into a target protein. (a) SDS-PAGE analysis of Lys7→TAG amber mutant of Z-domain protein expressed under different conditions. Lane 1: molecular mass marker; lane 2: expression with WT *Mj*TyrRS; lane 3: expression with NpOH-RS1 in the presence of 2NpOH; lane 4: expression with NpOH-RS1 in the absence of 2NpOH. The SDS-PAGE gel was stained with SimplyBlue SafeStain reagent.

The yield of the Z-domain mutant was roughly 4 mg/L culture with the presence of

2NpOH, but was insignificant without 2NpOH (Fig. 1a), indicating a very high fidelity in the incorporation of the unnatural amino acid.

Moreover, the ESI-MS spectrum showed peaks at m/z = 7847.0, 7889.1 and 7978.3 (for NpOH-RS1), and 7847.0, 7888.7, 7977.6 (for NpOH-RS2), which match the expected molecular weight for the full length 2NpOH 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)[26c].

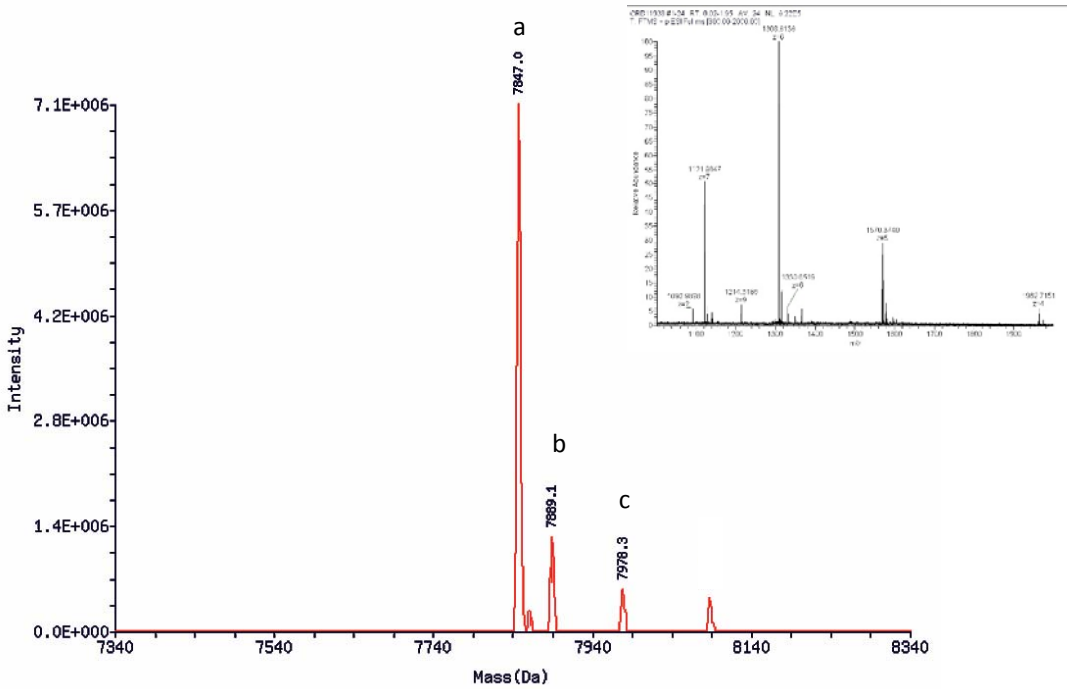


Figure 21. ESI-MS spectrum of 2NpOH incorporated Z-domain protein (expressed with NpOH-RS1): peak a is corresponding to the Z domain protein without the first Met residue, peak b is the acetylated product of peak a, and peak c is the full length protein.

No evidence showed that the wild type protein was produced during the process.

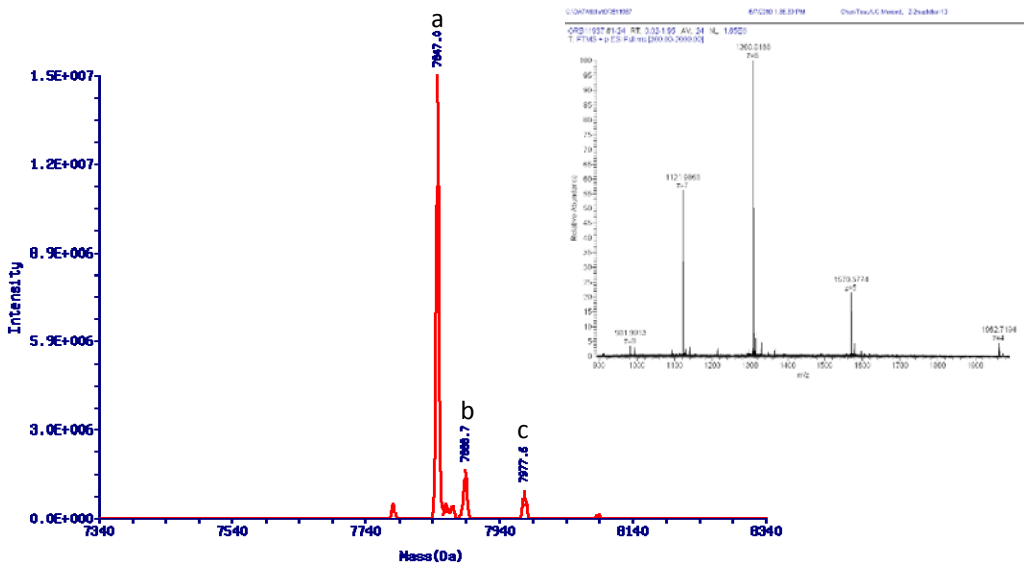


Figure 22. ESI-MS spectrum of 2NpOH incorporated Z-domain protein (expressed with NpOH-RS2): peak a is corresponding to the Z domain protein without the first Met residue, peak b is the acetylated product of peak a, and peak c is the full length protein.

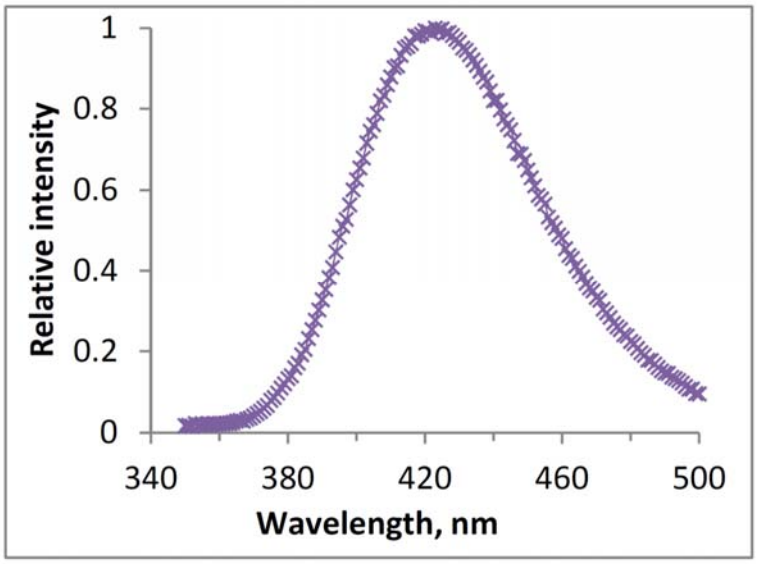


Figure 23. The emission spectrum of the 2NpOH-Z domain protein displayed a maximum wavelength of 423 nm when excited with 335 nm light.

Furthermore, under 335 nm excitation, the purified Z-domain mutant displayed strong blue fluorescence with an emission maximum at 423 nm, a new feature acquired from the successful addition of 2NpOH.

2.5 An enhanced protein expression system.

All the evidence above showed that the unnatural amino acid 2NpOH was successfully incorporated into proteins genetically. With this starting point, we could begin to exploit the unique properties of 2NpOH in proteins. One limitation of this methodology is that, sometimes the yield is relative low to produce the UAA decorated proteins.

To boost the expression of the mutant protein, a vector containing two copy of NpOH-RS gene, was construct (Figure 24). The higher yields of mutant proteins can be achieved with this dual aaRS system.

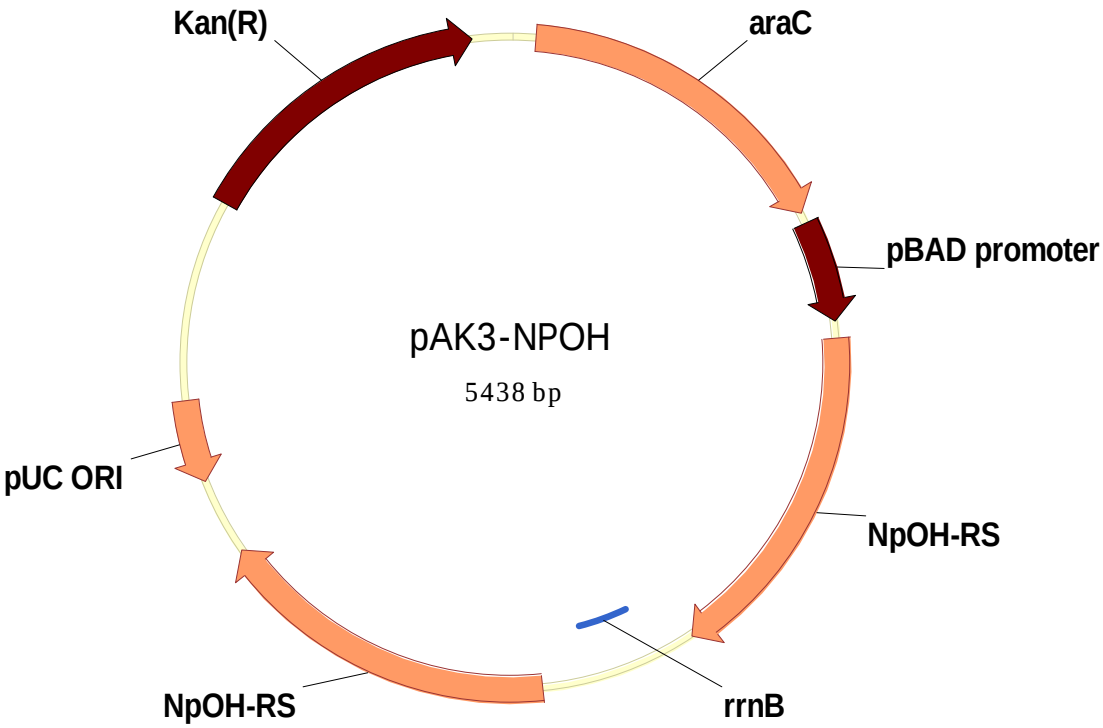


Figure 24. Vector pAK3-NpOH

The yield of the Z-domain protein was determined by measuring the UV-Vis absorbance at 280 nm. Given that, Z-domain-2NpOH mutant protein has 1 2NpOH and 1 Tyrosine. And the molar extinction coefficient at 280 nm for tyrosine is 1280; and for 2NpOH is 3783(Calculated by 1mM 2NpOH 10mm absorbance = 3.783). So the molar extinction coefficient of Z-domain-2NpOH mutant protein is 1280+3783 = 5063. 110 uL protein solution was extracted and purified from 50 mL LB culture with 0.5 mM unnatural amino acid, with absorbance of 2.00. So the concentration was 3.95×10^-4 M. The Molecular weight of Z-domain-2NpOH mutant protein was ~7900 g/mol (varied by the fraction of different post-translational modification product). So 1 L culture could produce protein 7900×3.95×10^-4×1000/50×110×10^-6 g=6.9*10^-3 g. So the Z-domain protein

was produced with yield of roughly 7 mg/L in liquid LB medium, with 0.5mM 2NpOH in presence.

2.6 Conclusion

In summary, we have evolved *Mj*TyrRS synthetase mutants to genetically encode 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid (2NpOH) in *Escherichia coli* with high fidelity and efficiency. The acquired strong blue fluorescence (Figure 19), SDS-PAGE gel(Figure 20), mass spectrometry analysis (Figure 21, Figure 22)The 2-naphtol chemistry was added into proteins successfully,

Chapter 3

Azo coupling of 2NpOH incorporated protein and reactions thereafter.

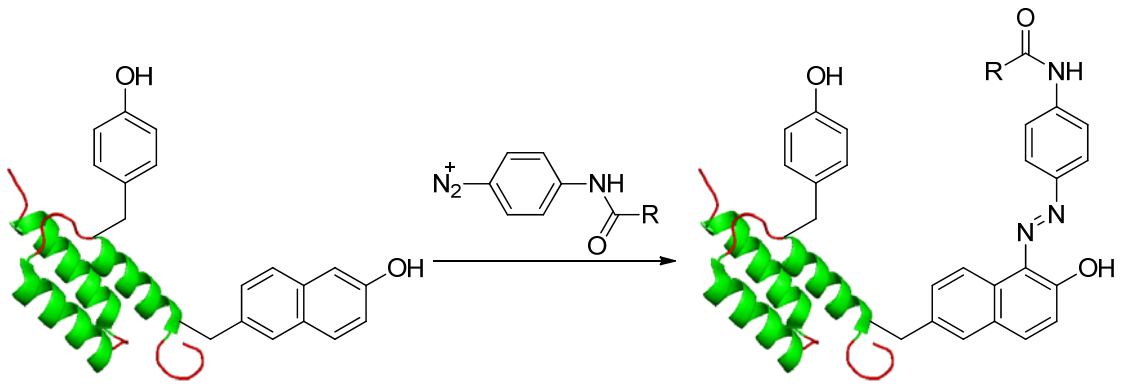


Figure 25. A site-specific azo coupling bio-conjugation reaction towards 2NpOH residue. (Figure is generated with PyMOL based on the crystal structure of Z domain protein, PDB ID: 1Q2N[43])

Once the 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid has been genetically incorporated into proteins site-specifically, we can expect that, a possible modification reaction may be developed based on this residue, because of the unique chemical properties of this unnatural amino acid.

3.1 Diazonium reaction with tyrosine residue

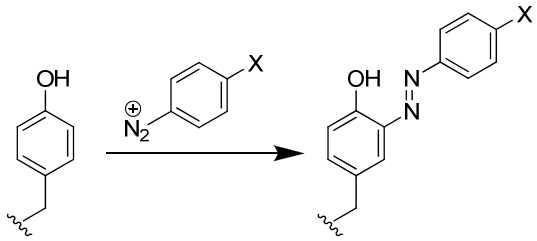


Figure 26. Bio-conjugation reaction on tyrosine residue with diazonium salts.

Prof. Matthew B. Francis has reported a specific modification method towards tyrosine residue (Figure 26)[44]. Tyrosine residue effectively reacts with diazonium

salt, derived from aniline with electron-withdraw group, under mild condition. However, all exposed tyrosine residues on the surface of the protein will react, no selectivity could achieve. This problem can be easily solved by specific inserting 2-naphthol chemistry at chosen site. Since naphthol has higher reactivity than phenol in electrophilic substitution. We expect that the 2-naphthol tyrosine analogue would react with these diazonium salts prior to tyrosine under certain conditions. Thus, we have developed a highly selective modification method based on this 2-amino-3-(6-hydroxy-2-naphthyl)propanoic acid residue.

3.2 Small molecule test for optimization of the reaction condition

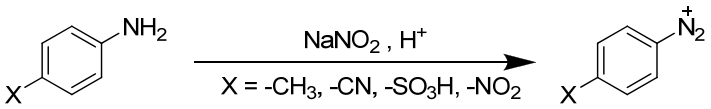


Figure 27. Preparation of diazonium salts

To uncover the proper conditions for selective azo coupling, we used *p*-cresol and 2-naphthol as the model molecules to mimic the reactions for tyrosine and 2NpOH, respectively. A series of aniline derivatives were converted to their corresponding diazonium salts after treated with aqueous *p*-toluenesulfonic acid and sodium nitrite.

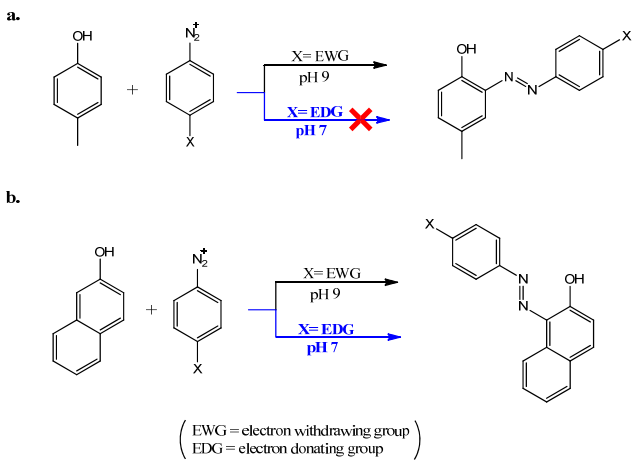


Figure 28 Azo coupling with *p*-cresol versus that with 2naphthol

The diazonium salts were reacted with either to *p*-cresol or 2-naphthol at pH 7 and pH 9, respectively (Figure 28). Since the aromatic azo compounds are normally dyes, we can just simply tracking the color to monitor the progress of the reactions.

The stock solution of a series of diazonium salts were prepared by mixing the followings: pre-cooled 100 μL of 0.84 M aqueous p-toluenesulfonic acid, 100 μL of 0.45 M aqueous sodium nitrite, and 200 μL of 0.15 M acetonitrile solution of aniline

derivatives such as sulfanilic acid, 4-nitroaniline, 4-aminobenzonitrile, aniline, p-toluidine and 4-aminoacetanilide, respectively. The reaction mixtures were kept under an ice bath for 60 min.

To the 600 µL of 0.2 M phosphate buffered solution (pH 7.0 and pH 9.0, respectively), 200 µL of either 30 mM p-cresol or 2-naphthol in acetonitrile was added at 0°C. Next, 100 µL of a fresh prepared diazonium stock solution was added. After letting the reaction mixture sit on an ice bath for 5 min, 20 µL of 1.5 M aqueous sodium azide was added to quench the reaction by deactivating the remaining diazonium compound.

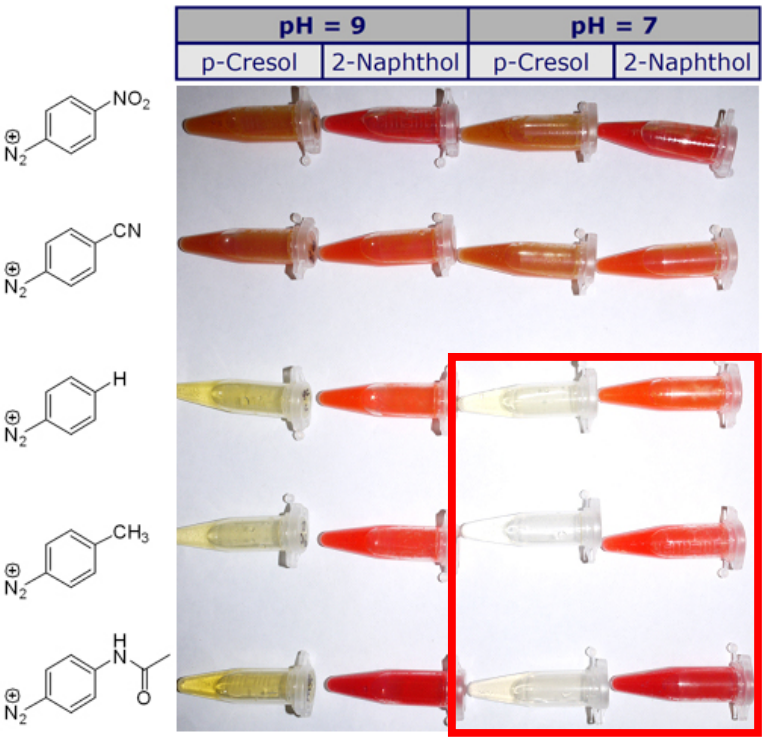


Figure 29. The formation of azo adducts between p-cresol and diazonium salts and between 2-naphthol and diazonium salts. Reaction conditions that show selectivity are highlighted in red blocks.

Analyzing the results in Figure 29, it reveals that at pH 7, the diazonium salts generated by *p*-toluidine and 4-aminoacetanilide selectively react only with the unnatural amino acid, but does not react with tyrosine. On the other hand, the diazonium salts bearing electron-withdrawing substituents were generally found not so specific, since the electron-withdrawing substituents will enhance the reactivity of the diazonium salts. Also, at high pH, the phenol or naphthol have high potential to deprotonated, thus increase the electron density of the aromatic ring to enhance the reaction. Notably, the reaction goes really fast. It usually gives the color of azo compound immediately after the addition of the diazonium salt. To prevent tyrosine residue reacting with the excess diazonium salt, the reaction can

be quenched by adding sodium azide.

In addition, we tried this type of reaction with other nature amino acids bearing aromatic ring side chain, such as Phenylalanine, Tryptophan and Histidine, under the same condition. No color was observed. Similar reactions were also carried out for dNTPs (deoxyribonucleotide triphosphate mixtures); none of them displayed the color of the azo product either.

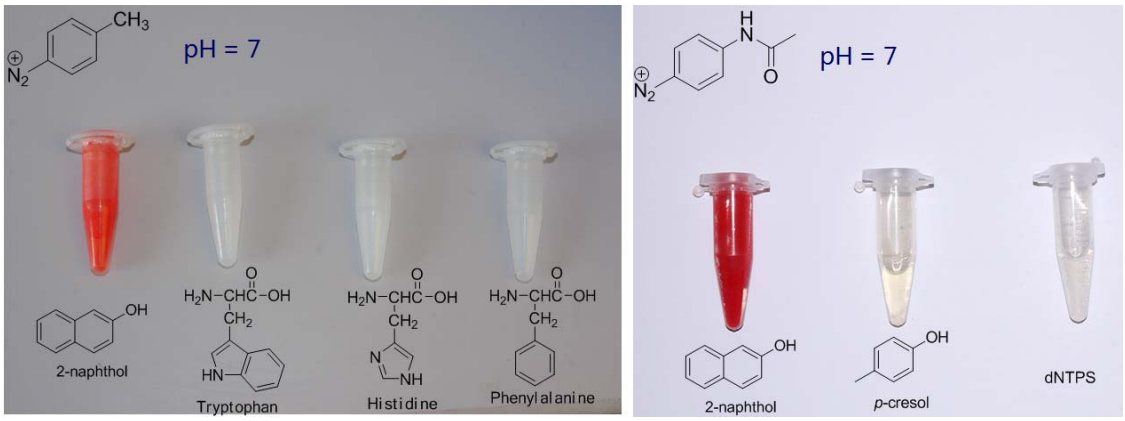


Figure 30 Azo adduct is generated only with 2-naphthol but not with any of the aromatic canonical amino acids neither for deoxyribonucleotide triphosphate (dNTPs)

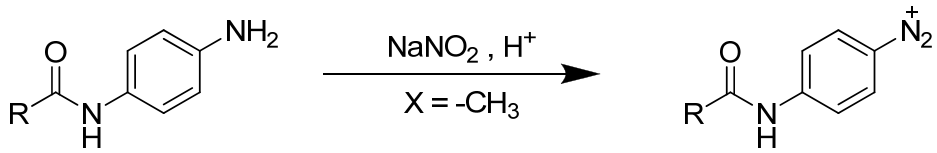


Figure 31 Diazonium generated by 4-Aminoacetanilide derivative

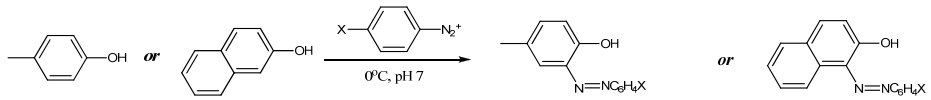
With the diversity of the R group, we can really broaden the application of this reaction to link different probe to the protein.

To obtain the reaction yields for the coupling reactions, a higher loading of the reagents were applied. The followings were used to make the stock solution of the diazonium compounds: 62.5 µL of 0.84 M p-toluenesulfonic acid, 62.5 µL of 0.48 M NaNO2, and 125 µL of 0.12 M aniline derivatives were used for the preparation of the diazonium stock solution (thus, 60 mM final concentration). The following conditions were used for the azo coupling reactions: 700 µL of phosphate buffer (0.2 M, pH 7), 200 µL of 60 mM 2-naphthol or p-cresol (12 µmol), and 250 µL of 60 mM diazonium solution (15 µmol). The following condition was used to stop the reaction: aliquots of 50 µL of 1.5 M NaN3.

Since most of the azo products (except those with a sulfonate group attached)

showed very low solubility in water and formed precipitations. These azo adducts were collected with spin columns (Quiagen) through centrifugation, washed several times with water, and then dried on a 70 °C heating block before their weights been taken. Additionally, the collected solid was partially dissolved in ethanol and then diluted with 60% ethanol$_{(aq)}$ for UV-vis spectroscopy analyses. The data is summarized in Table and Table.

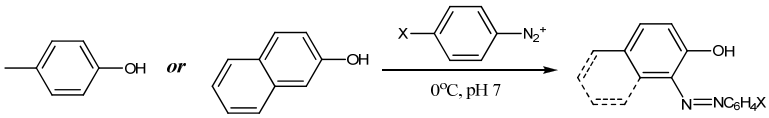
Table 2 Reaction yields for the azo coupling of *p*-cresol and 2-naphthol with diazotized aniline derivatives and the characteristic UV-vis absorption maxima for the azo products.



X	azo from *p*-cresol		azo from 2-naphthol	
	Yield (μmol)[a]	λmax (nm)	Yield (μmol) [a]	λmax (nm)
$NHCOCH_3$	0	N.D. [b]	9.2	497
CH_3	0	N.D. [b]	9.1	494
H	0	N.D. [b]	9.3	486
SO_3^-	N.D.[b]	396[c]	N.D. [b]	488[c]
CN	8.9	404	10.2	482
NO_2	9.7	406	10.6	492

[a] Based on the weight of the precipitated azo adduct; theoretical yield = 12 μmol . [b] Not determined because no azo adduct or no precipitation was formed. [c] Measured directly from the reaction mixture other than an isolated azo compound.

Table 3 Percent yields for the azo coupling of p-cresol and 2-naphthol with diazotized aniline derivatives. [a]



X	*p*-cresol (%)	2-naphthol (%)
$NHCOCH_3$	0	76
CH_3	0	76
H	0	77
SO_3^-	high[b]	high[b]
CN	74	85
NO_2	81	88

[a] Based on the weight of the precipitated azo adduct. [b] No precipitation is formed but a high yield is reported here judged by the appearance of the

intense bright color from the azo adduct.

Furthermore, the small molecule test was carried out with actual amino acid, tyrosine and the unnatural amino acid 2NpOH. Similar result were observed under the optimized condition.

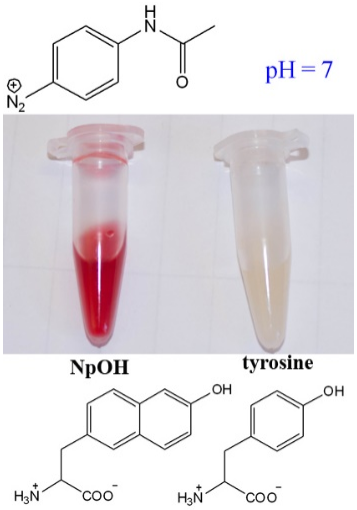


Figure 32. Azo adduct formed only with 2NpOH but not with tyrosine

3.3 Bio-conjugation of Z-domain proteins

We next used the Z-domain protein model target to test our bio-conjugation reaction at the protein level. The reaction selectivity was checked with 2NpOH-Z (Z domain incorporated with 2NpOH) and Tyr-Z (control protein with a tyrosine replacing the 2NpOH in 2NpOH-Z). Diazonium salts prepared by acetonitrile and p-toluidine were used as candidates that are electron-poor and electron-reach, respectively.

After adding the diazonium salt to the protein solution sample, we run the SDS-PAGE analysis (Figure 31). The Tyr-Z protein showed color only with diazonium salt bearing cyano group. 2NpOH-Z gave colors with both diazonium salt bearing methyl and cyano substituents, showing that diazonium salt with electron-donating group has good selectivity on the 2NpOH.

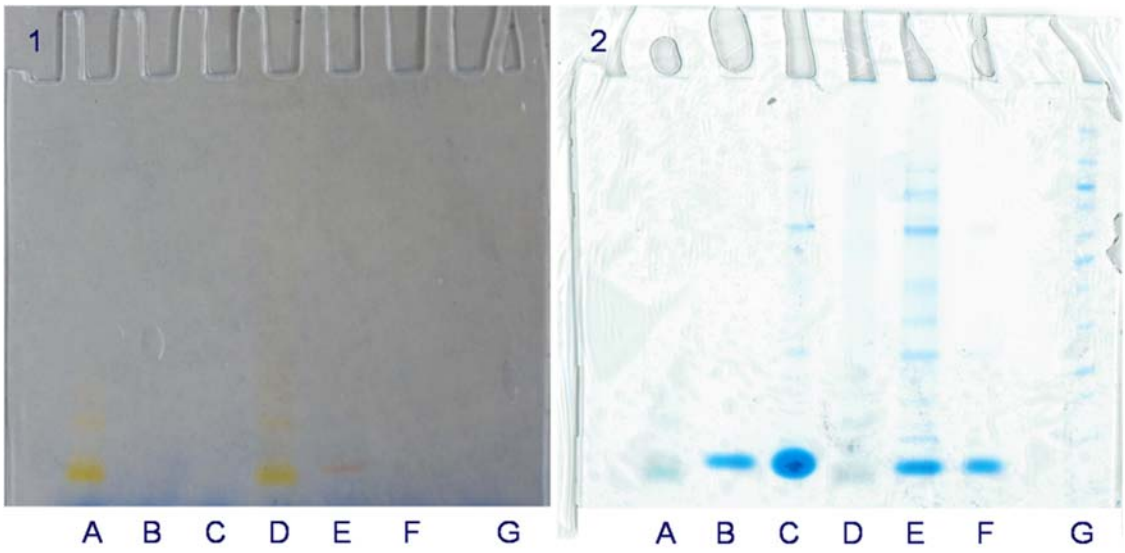


Figure 33 SDS-PAGE analysis. 1. The protein gel before blue staining; 2 the protein gel after blue staining. A: Tyr-Z reacts with diazonium bearing cyano group; B: Tyr-Z reacts with diazonium bearing methyl group; C: Tyr-Z; D: 2NpOH-Z reacts with diazonium bearing cyano group; E: 2NpOH-Z reacts with diazonium bearing methyl group; F: 2NpOH-Z; G: Protein standard ladder.

To determine the efficiency of this selective bioconjugation reaction, we plan to use the mass spectrometry analysis, which can show the ratio of the conversion of the modification.

The stock solution of the diazonium salt was prepared by mixing the followings: pre-cooled 100 µL of 0.84 M aqueous p-toluenesulfonic acid, 100 µL of 0.45 M aqueous sodium nitrite, and 200 µL of 0.15 M 4-aminoacetanilide or in acetonitrile. The reaction mixture was kept at 0 °C for 60 min.

In separate eppendorf tubes, 10 µL of 800 µM NpOH-Z domain protein and 10 µL of 1mM Tyr-Z control protein were added to 110 µL pH 7 phosphate buffer, respectively, and kept at 0°C. 20 µL of fresh diazonium stock solution (75 mM) was added to the protein solutions. After 90 min, 10 µL of 1.5 M aqueous sodium azide was added to the reaction mixture to deactivate the remaining diazonium salt. The solution was desalted by Thermo ScientificTM Aspire RP30 Desalting Tip, and the desalted protein samples were subjected for Mass-spectroscopy and UV-visible spectroscopy analyses.

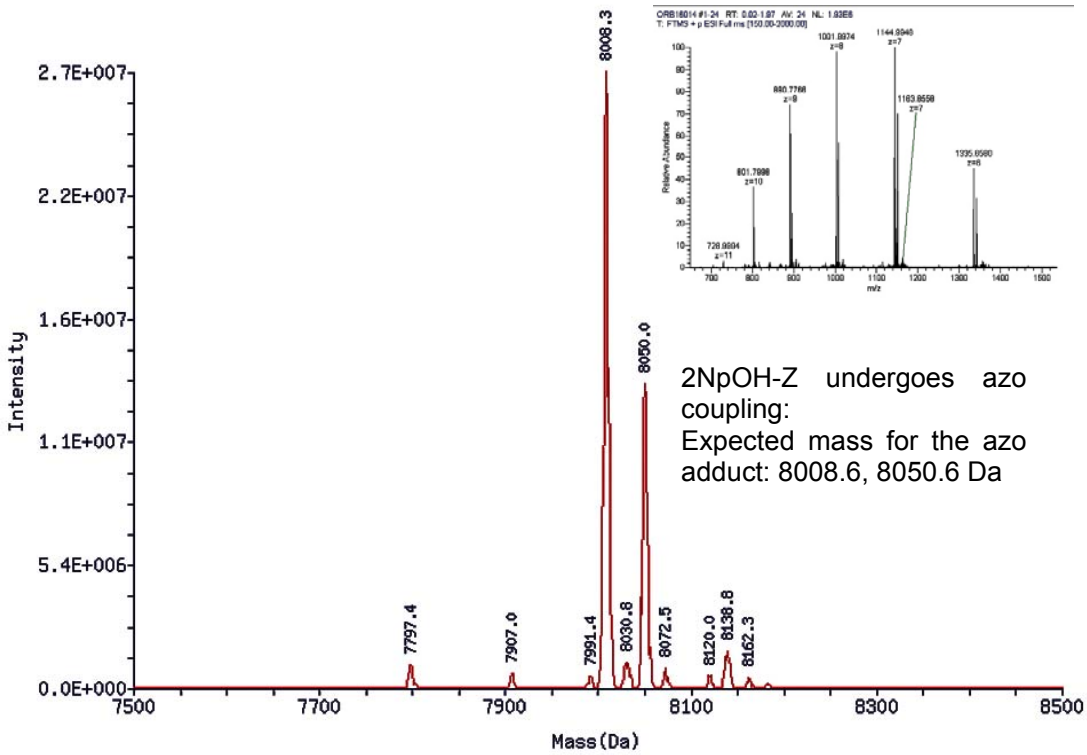


Figure 34. The ESI-MS spectrum for the reaction of 2NpOH-Z protein. The two major peaks of the spectrum are assigned to the azo adducts of 2NpOH-Z-domain protein. Mass peaks for unmodified 2NpOH-Z Protein are insignificant.

Compared to the MS peaks from the protein measured before the coupling reaction (peaks a, b and c in Figure 21), a mass increment of 161.2 Da for each of them was expected at 8008.6, 8050.6 and 8139.8 Da due to the addition of the azo moiety. Indeed, the major peaks showed in Figure 34 at 8008.3 and 8050.0 Da could be attributed unambiguously to these assignments. The mass peaks for the unmodified Z-domain mutant at 7847 and 7889 Da were insignificant in Figure 34. This suggests that the coupling reaction is very efficient and has a near quantitative conversion yield (> 96%, estimated from the heights of the MS peaks). To further demonstrate that the coupling reaction is triggered by the presence of 2NpOH, azo coupling were tested with Tyr-Z control protein (Figure 35.), in which the 2NpOH residue was replaced with a tyrosine so that the control protein contains tyrosine residues but no 2NpOH. Results from the ESI-MS analysis of the control reactions yielded only the mass peaks corresponding to the unmodified protein and that the formation of its azo adduct was negligible

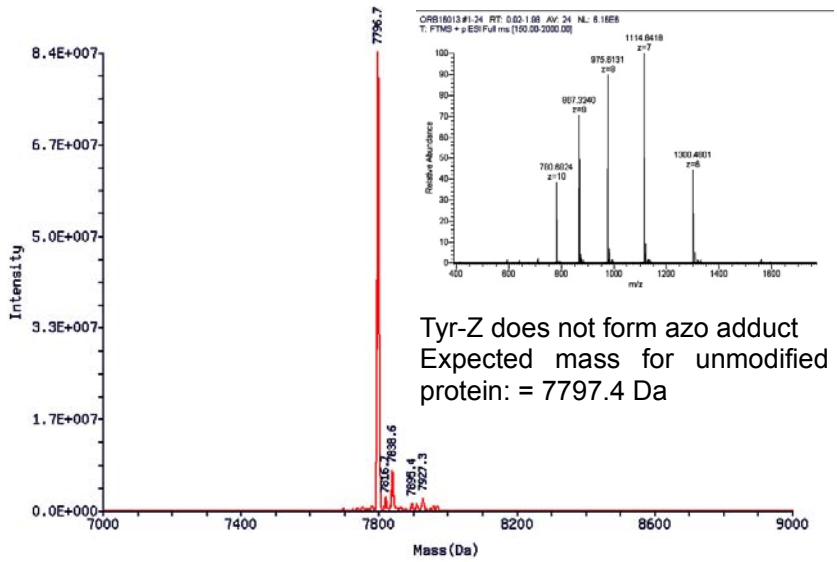


Figure 35. The ESI-MS spectrum for the reaction of the control Tyr-Z-domain protein. The MS peaks for the azo adduct of Tyr-Z domain were not observed

In addition, the modified protein displayed a distinct pink color with a broad UV-visible absorption band at 520 nm. This strongly supports that the azo coupling took place at the 2NpOH site. The UV-Vis spectrum was measured for the reaction solution with diazonium salt bearing electron-donating group. No obvious absorbance at visible light range was observed before or after the reaction with Tyr-Z protein.

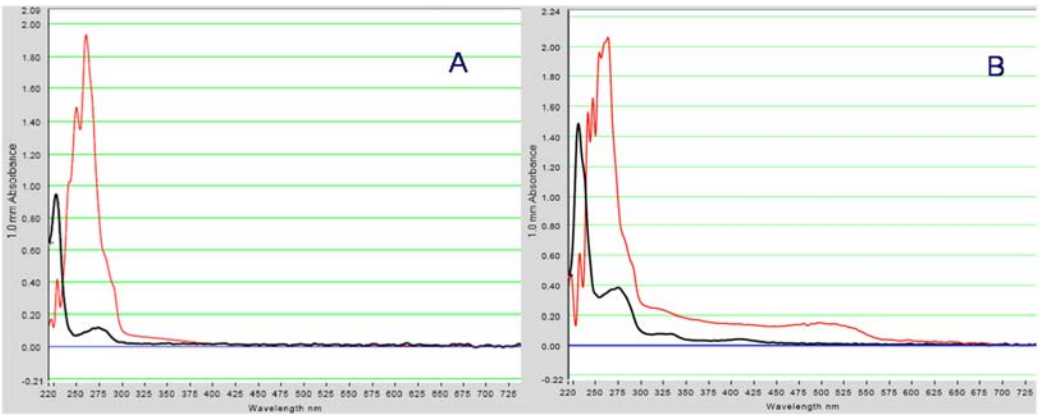


Figure 36. The UV-visible spectra for the Tyr-Z protein (A) and 2NpOH-Z Protein before (black line) and after (red line) the azo coupling reaction with diazotized p-toluidine. A broad band in the visible region with a λmax at about 520 nm was observed after the azo coupling, for which the 520 nm band can be assigned to the absorption for the expected azo functionality.

3.4 Kinetic study of the azo-coupling reaction.

With the fast appearing of the bright color during the reaction and the fact after 1.5 hours coupling most the 2NpOH-Z was modified (> 96%, estimated from the heights of the MS peaks), we can conclude that the reaction is quite efficient with a great reaction rate. To investigate the kinetic behavior of this diazonium salt coupling reaction, the following experiment was carried out.

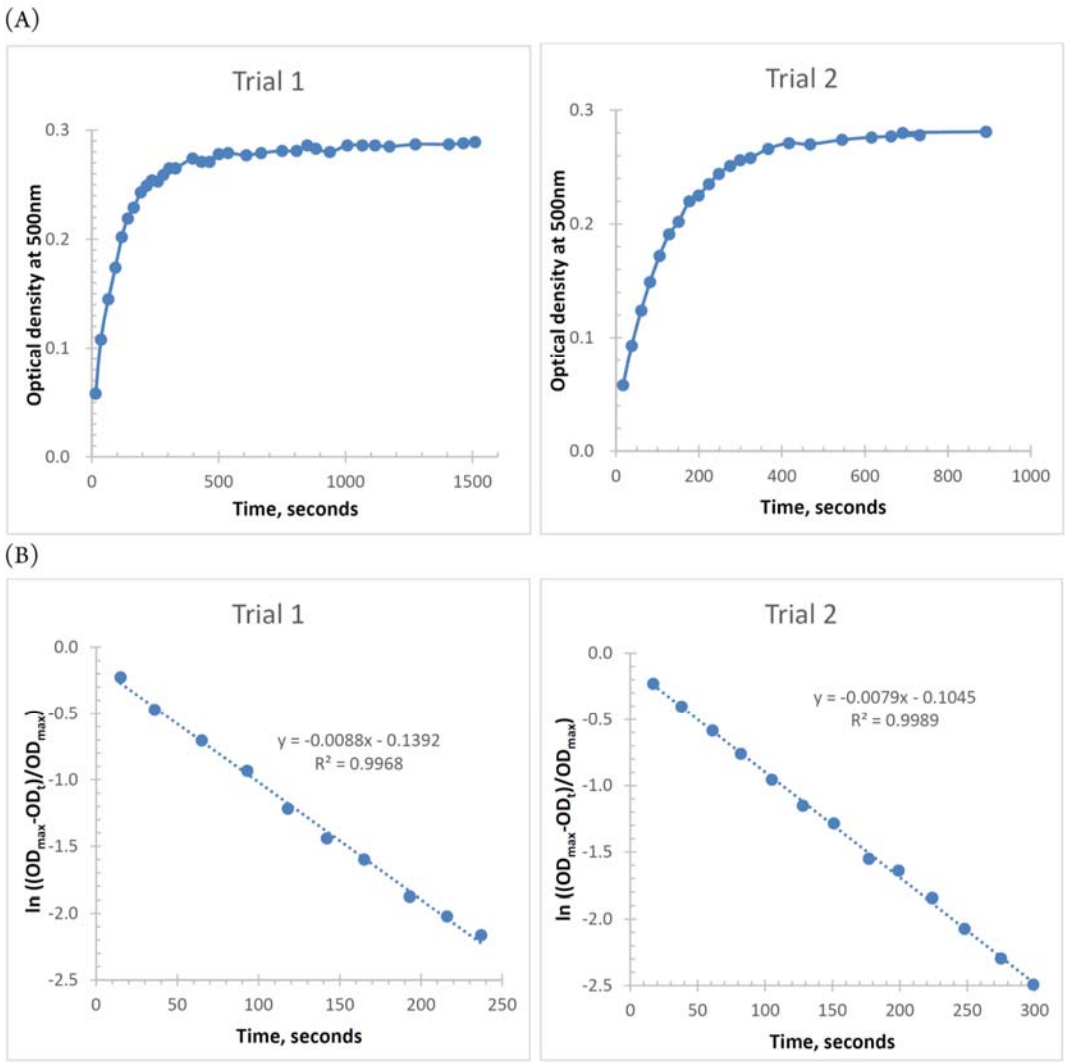


Figure 37. Kinetic study for the formation of the azo adducts. (A) The growth curves of the azo adduct produced from 140 μM of 2NpOH-Z protein and 30 mM of diazotized 4-aminoacetanilide. (B) The plots of $\ln((OD_{max}-OD_t)/OD_{max})$ versus time and their linear regression analyses.

The operations described here were conducted in a 4 °C cold room. To an eppendorf tube containing proper amount of pH 7 phosphate buffer under ice,

aliquots of 2NpOH-Z domain protein (final concentration of 140 μM) and diazotized 4-aminoacetanilide (final concentration of 30 mM) were quickly mixed in to start the coupling reaction. The progress of the reaction was monitored by following the increasing absorption at 500nm, the unique UV-vis band from the azo adduct. Portions (2 μL each) of the reaction mixture were pipette out at each of certain time points and their UV-vis spectrum was measured with a NanoDrop spectrophotometer (Thermo Scientific). According to the changes of the optical density at 500 nm as shown in Figure 36, the formation of the azo product followed an exponential growth curve, which is in agree with a pseudo first order kinetics. The maximum optical density (OD_{max}) was calculated from the average of the region where the signal leveled off. A second plot can be derived from $\ln((OD_{max}-OD_t)/OD_{max})$ versus time, where OD_t is the optical density at each time point t. As shown in Figure 37, the new plot showed a good liner trend with $R^2 > 0.99$ and the pseudo first order rate constant can be derived from the slope of the linear regression. The whole process was duplicated and the rate constants were averaged as the reported value of $(8.4 \pm 0.4) \times 10^{-3}\ s^{-1}$.

3.5 PEGylation of 2NpOH-Z domain protein

PEGylation is the process of attaching the strands of the polyethylene glycol to molecules, typically to peptides, proteins, and antibody fragments, which can used to improve the safety and efficiency of many therapeutics.[45] By increasing the molecular weight and enlarging the hydrodynamic size, the pharmacological feature of the molecule could be improved over the unmodified form, such as increasing the solubility, reducing the dosage frequency that potentially reducing the toxicity, extending the circulating life, increasing the stability and etc.

A plausible application for site-specific bio-conjugation reaction mentioned above is PEGylation via azo coupling with a diazonium reagent bearing a polyethylene glycol moiety.

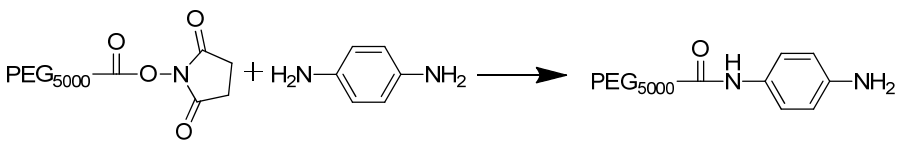


Figure 38. The synthesis of a PEGylated aniline derivative

The PEGylated aniline derivative was prepared as depicted in Scheme. The N-hydroxylsuccinimide functionalized polyethylene glycol (mPEG-NHS, PEG succinimidyl ester, MW 5000) was purchase from Nanocs Inc. 0.1688g of mPEG-NHS (33.8 μmol, 1 equivalent) and 0.0110g of 1,4-diaminobenzene (102 μmol, 3 equivalent) were added to 25 mL H_2O in a flask. The reaction mixture was kept stirring overnight at ambient temperature. Subsequently, excess amount of 1,4-diaminobenzene was removed upon centrifugation in AmiconTM Ultra centrifugal filter (>3kD). The final volume of the reagent solution was adjusted to 500 μL (67.5

mM, assuming quantitative amide bond formation) and was used directly for azo coupling without further purification. The stock solution of its diazonium salt was prepared by mixing pre-cooled 20 μL of 0.84 M aqueous p-toluenesulfonic acid, 20 μL of 0.45 M aqueous sodium nitrite and 120 μL of 67.5 mM PEGylated aniline derivative at 0°C, and then keeping the mixture at 0 °C with an ice bath for 60 min.

Next, 20 μL of 300 μM of 2NpOH-Z protein and 20 μL of 300 μM of Tyr-Z control protein were buffered, respectively, with 40 μL of pH 7 phosphate buffer at 0 °C. After that, 50 μL of the diazonium stock solution was added to each of the protein solutions. The reaction mixture was kept at 0 °C with an ice bath for 5 hours, then was added with 10 μL of 1.5 M NaN3 aqueous solution to stop the reaction, and finally was subjected for SDS-PAGE analysis.

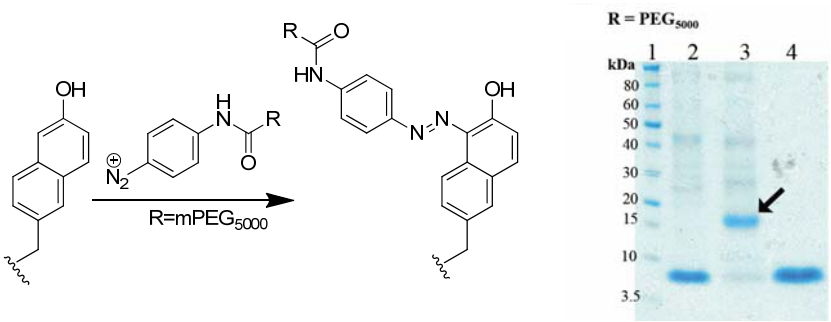


Figure 39. SDS-PAGE analysis for the PEGylation of Z-domain protein through azo coupling. Lane 1: molecular mass marker; lane 2: 2NpOH-Z protein before the coupling reaction; lane 3: the azo coupling product for 2NpOH-Z coupled with the PEG5000 containing reagent, the target protein band is emphasized with an arrow sign; lane 4: the azo coupling between a control Tyr-Z protein and the PEG5000 containing reagent. The SDS-PAGE gel was stained with SimplyBlue SafeStain reagent.

As depicted in Figure 39, PEGylated product was obtained efficiently for the conjugation of the 2NpOH incorporated protein (lane 3), while applying the same condition to the control protein didn't yield any adduct (lane 4). This indicates again that the coupling reaction is highly selective and that it could be utilized to modify proteins with not only small molecules but also macromolecules.

3.6 Reductive cleavage of the azo adducts.

There are reports showed azo adduct could be subsequently cleaved by the reducing reagent sodium dithionite to form a product with an extra amino group on the 2-naphthol residue[46] (Figure 40). So that, potentially, the synthetic tag added through this azo coupling reaction, could be removed by a following reductive cleavage.

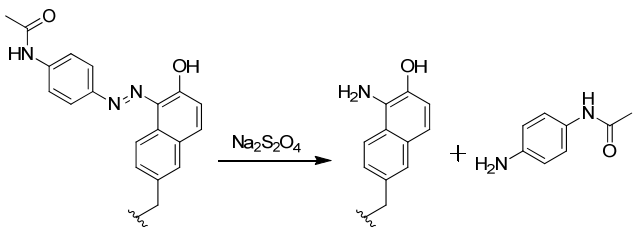


Figure 40. Reductive cleavage of the azo adduct.

At room temperature, to a solution of azo product in pH 7 buffer, excess amount of sodium dithionite (100 equivalent) was added. After 2-hour cleavage reaction, the resulting mixture was purified with Thermo ScientificTM Aspire RP30 Desalting Tip, and the desalted protein samples were subjected for Mass-spectroscopy

Compared to the MS peaks from the 2NpOH-Z protein measured after the coupling reaction (peaks a, b in Figure 41.), the reductive cleavage result (Figure 42) showed a mass d of 146.1 Da for each of them was expected at 7862.5, 7904.5 Da due to the removal of the azo moiety. Indeed, the major peaks showed in Figure at 7861.7 and 7903.7 Da could be attributed unambiguously to these assignments. The mass peaks for the un-cleaved 2NpOH-Z protein mutant at 8009 and 8051 Da were still could be found, although only around 10% (estimated from the heights of the MS peaks). Notably, there are two peaks (e, f in Figure 42) are matching the un-fully reduced form with a hydrazine residue left at the coupling site.

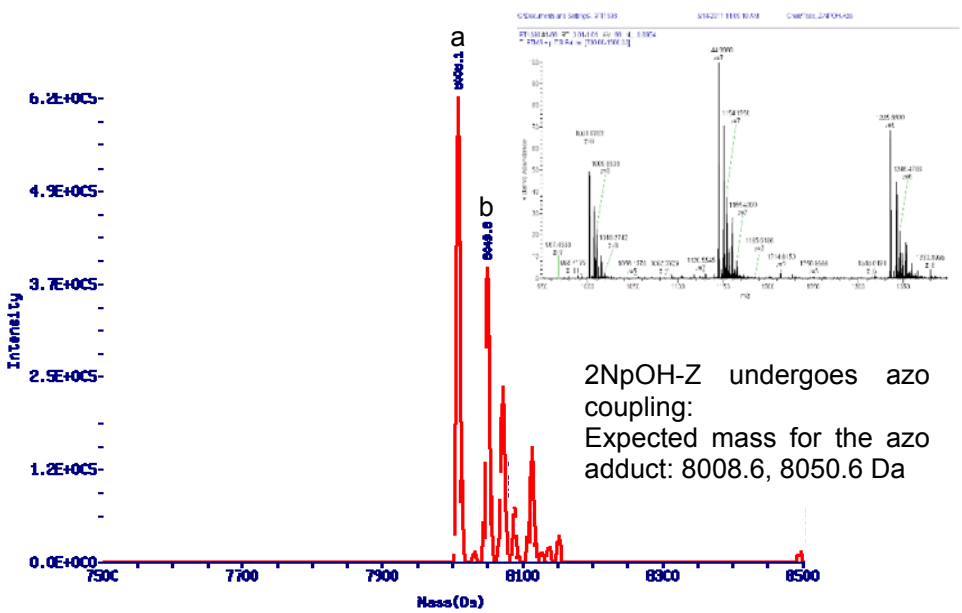


Figure 41. The ESI-MS spectrum for the 2NpOH-Z protein after the azo coupling.

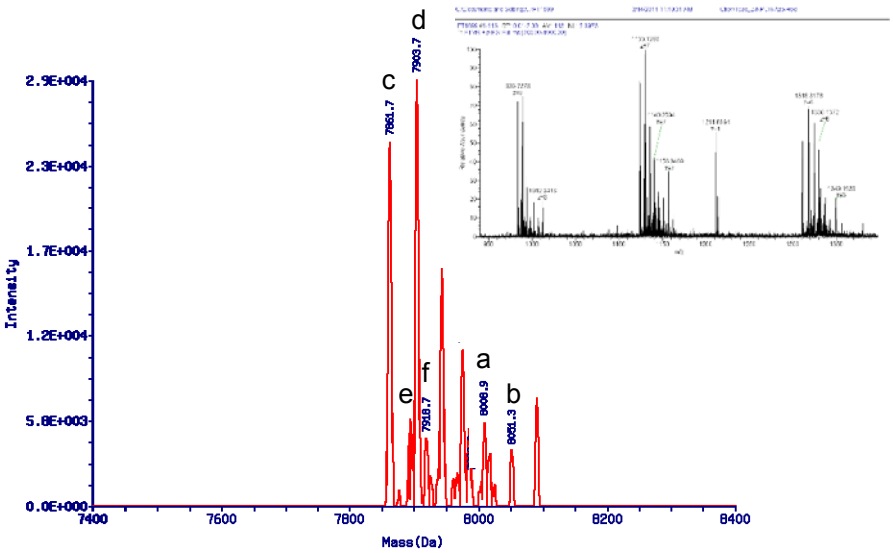


Figure 42. The ESI-MS spectrum for the reductive cleavage of 2NpOH-Z protein azo product. Peak a and peak b indicating the unduced azo adducts. Peak c and d indicating the fully reduced product, peak e and f indicating the partial reduced product

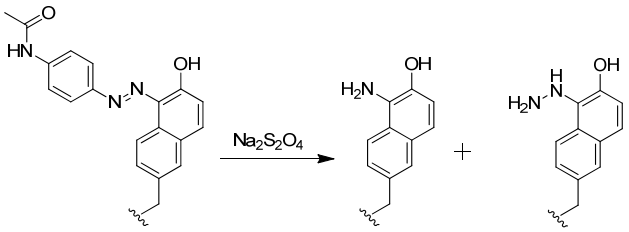


Figure 43. The fully and partial reduced product by the reductive cleavage of the azo adduct.

3.7 Protein immobilization

Protein immobilization is one important key to realize biological microarray technology[47], which can be applied to biochips, biocatalysis, immunoassays and drug selections. Diazonium salts modified solid phase can react selectively with proteins containing 2-naphthol residue to achieve the protein immobilization. Protein could be easily recovered via the reducing cleavage. Similarly, rather than the commonly used protein purification technique involved with 6xHis tagged proteins and Ni-NTA agarose[48], a new protein purification strategy can be brought up based on the same reaction. For purifying a desired protein from the crude lysate, 2NpOH can be genetically incorporated into desired proteins at permissive site. Then the proteins would be linked onto the modified resin. Unwanted proteins could be washed away and the excess diazonium salt on the bead surface could

be terminated by the wash with 2-naphthol solution. Subsequently, the reduction reaction can be carried out to obtain the desired protein, and the resin can be recovered (Figure 42).

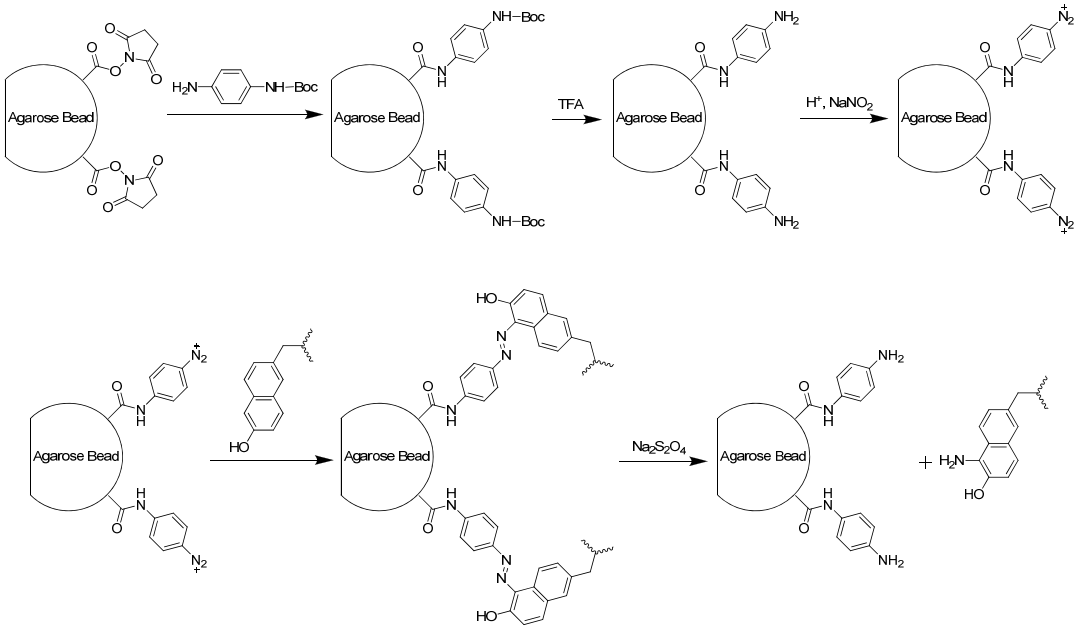


Figure 44. Proposed reversible protein immobilization strategy

The agarose bead with aniline terminal group was prepared as follow. P-pheneylenediamine mixed with Boc-ON (1.05 equivalent) under basic condition in THF/H_2O solution at 0 °C overnight to room temperature. The mono-Boc P-pheneylenediamine was purified by extraction. NHS-Activated Agarose (Thermo Scientific) was mixed with excess amount of the mono-Boc P-pheneylenediamine, reacted for 2 hours at room temperature upon gentle shaking. The modified agarose was then recovered by centrifugation of the slurry in Miniprep tube (Qiagen), followed by washing with pH 7 buffer to remove the unreacted small molecule. The resulting agarose bead was then de-protected in TFA solution. The agarose bead with aniline terminal group was recovered upon centrifuge and stored in pH 7 phosphate buffer as slurry.

A mutant GFP_{uv} protein with 2NpOH inserted at the fourth position (GFP_{UV}-X4-2NpOH) was expressed as the target for the protein immobilization.

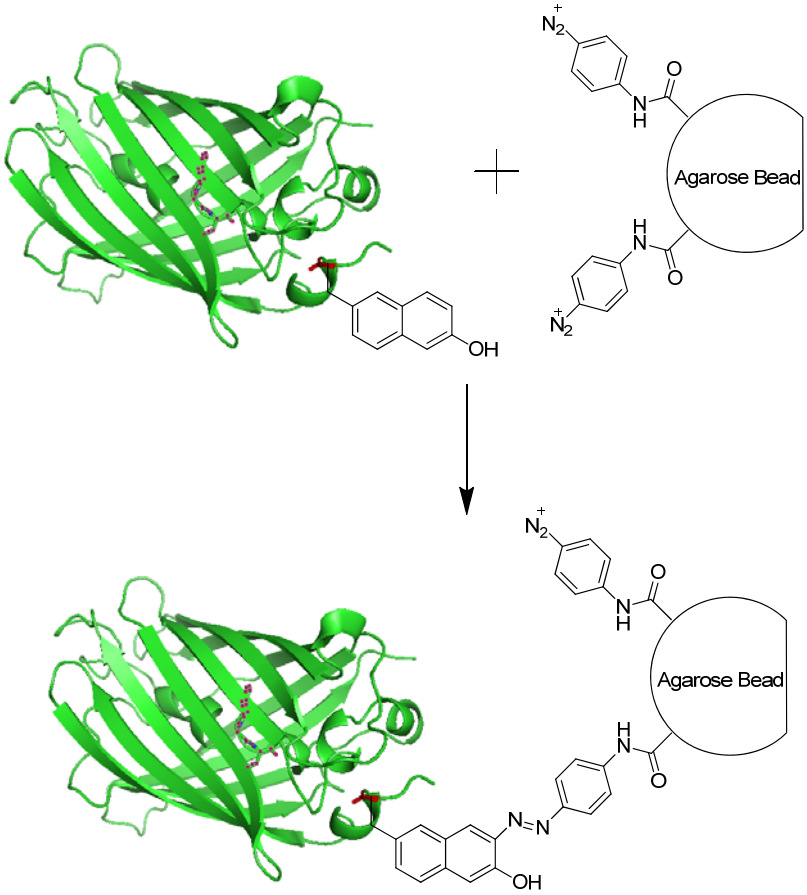


Figure 45. Covalent binding the GFP_{UV}-X4-2NpOH to the diazotized agarose bead (Protein structure is generated with PyMOL based on the crystal structure of GFP, PDB ID 1GFL[49])

50 µL of the bead slurry was mixed with 200 µL of 0.84 M aqueous p-toluenesulfonic acid, 200 µL of 0.45 M aqueous sodium nitrite at 0 °C for 1.5 h, to give the diazotized agarose bead slurry. Then 20 µL of GFP_{UV}-X4-2NpOH and the 80 µL diazotized agarose bead slurry was mixed in 400 µL pH 7 buffer. Checking with UV excitation, the green fluorescent was lost within minutes' time range. After 2 hours binding, the bead was collected by centrifuge and subsequently washed with buffer. Then the agarose bead was diluted with 500 µL buffer again, and then 0.028g sodium dithionite was added, shaking for 2 hours.

The resulting slurry was treated with centrifugation, save the supernatant and the agarose separately. The clear supernatant solution was excited with UV light, no fluorescent behavior was observed. SDS-Page Gel analysis shows no sign of protein was recovered.

2NpOH-Z protein was also tested with similar process. Both supernatant of after binding and cleavage were checked with SDS-Page gel, no protein was found,

which indicating the binding was effective, however the cleavage was no so much. Thinking of that after binding to the agarose bead, the protein molecule might cover the surface of the agarose bead, especially the covalent linkage site might be blocked, so that the access of the reducing reagent was not permitted.

3.8 Summary

We demonstrated that the 2-naphthol side chain in 2NpOH could serve as the target for site specific protein modification through the selective azo coupling reaction under neutral pH with diazonium compounds bearing an electron donating substituent. This provides an alternative way of biocompatible chemical transformation of proteins. With the reductive cleavage, the covalent-linked tag through this azo coupling reaction could be removed, thus it provided us an alternate somehow reversible bio-conjugation reaction. The selective azo coupling reported here could be potentially applied to reactions other than protein modifications and could fit to other aromatic compounds that undergo a faster electrophilic substitution than 2-naphthol does. Furthermore, the azobenzene moiety formed after the bio-conjugation reaction provides another useful property that the structure and function of a protein could be manipulated with visible light radiation.

Chapter 4

Study on a three-component Mannich-type coupling reaction towards the 2-naphthol residue.

As I mentioned above, the azo-coupling reaction is highly selective with high efficiency. However, one limitation of this reaction is that the diazonium salt needs to be prepared beforehand; and due to the relative high reactivity of the diazonium salt, it might not be an ideal method for bio-conjugation *in vivo*.

4.1 A three-component Mannich-type coupling reaction with tyrosine residue.

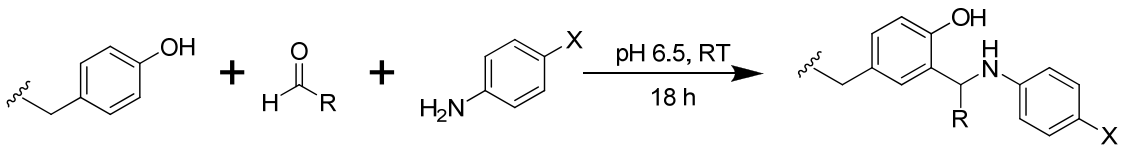


Figure 46. A three-component Mannich-type coupling reaction

Prof. Matthew B Francis has also reported a selective bio-conjugation reaction for the modification of tyrosine residues on protein substrates. This reaction utilizes imines formed in situ from aldehydes and electron-rich anilines to modify phenolic side chains through a Mannich-type electrophilic aromatic substitution pathway.[50] (Figure 46) Similar to the azo-coupling reaction he reported, despite the fact that tyrosine is only displayed with intermediate frequency on protein surface; still multiple sites might be modified by this reaction condition when more than one tyrosine residues are there. Although the different chemical condition around the tyrosine residue might affect the accessibility of the modification reagent, thus might cause yet somehow a selectivity. Based on a similar principle, we hypothesized that a three-component Mannich-type coupling reaction with 2NpOH residue might be better for the purpose of selective modification on protein surface. Additionally, since aldehyde and aniline could be taken up by cells easily, this reaction might be potentially utilized as an *in vivo* bio-conjugation method.

4.2 Small molecule study

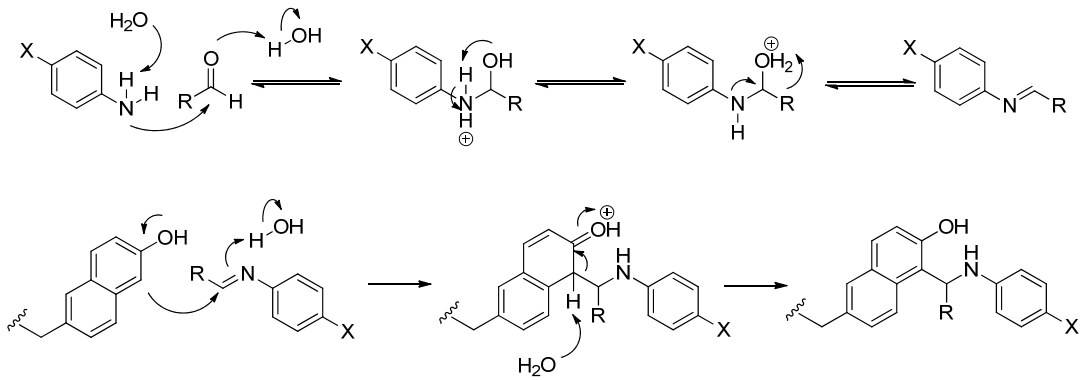


Figure 47. Mechanism of this Mannich type reaction

Based on the mechanism of this Mannich type reaction, the 2-naphthol residue is a better nucleophile than the tyrosine residue, one can assume that a selectivity could be achieved due to the faster reaction rate.

Since the original reaction condition for tyrosine coupling was at room temperature in a pH 6.5 buffer, to limit the reactivity of tyrosine residue, the optimized condition for 2NpOH residue might tune the pH of the buffer to change the electron density of reaction site; or decrease the temperature to maximally lower the reaction rate of the tyrosine, meanwhile the 2NpOH still can react; furthermore, the substituent of the aniline also can play a role in manipulate the reaction selectivity. However, the target reactant is protein, and also the bio-conjugation we want to develop is potentially an *in vivo* one, so that both the temperature and the pH can only vary from a relative narrow range.

To test our hypothesis, a small molecule study was carried out. We used p-cresol and 2-naphthol as the mimic of the tyrosine and 2NpOH. We prepared buffers at pH 5.3, 8.0, as acidic and basic condition. For the ortho-substituents of the aniline, we use 4-nitroaniline, 4-aminoacetanilide, and aniline as electron-poor and electron-reach aniline candidates. In addition, we only carried out the experiment under 37 °C in the incubator and under 0 °C on ice. The results were simply checked by the TLC.

Since the reaction mixture was fairly dilute and the UV absorbance of the product is relatively low; the TLC only show a fairly faded dot for newly formed adduct. However, it shows somehow a selectivity for the 2-naphthol prior to p-cresol. With 4-nitroaniline, at pH 8.0, under 37 °C; and with 4-aminoacetanilide, at pH 5.3, under 0 °C, both condition showed some selectivity, and can be the start for the further optimization.

4.3 *In vivo* reaction

With the preliminary small molecule study result, it could be a good start to the second stage for this bio-conjugation reaction development. To save the trouble of separating and purifying the new formed product from the dilute aqueous small molecule test mixture, we assumed that the protein scale experiment following a mass-spectrometry measurement would be a more sensitive way to have an insight into this reaction.

Since the goal was eventually to introduce a new *in vivo* protein modification method, we then directly applied the reaction condition mentioned above to the *E.coli* cell that containing protein with 2-naphthol residue instead of to the purified protein.

Z domain vector and pBK-2NpOH were co-transformed into Top10 *E.coli* cells. A single colony from the co-transformation was grown overnight at 37 °C in 2.5 mL of LB media supplemented with kanamycin and chloramphenicol to obtain the starter saturate culture. The culture then was diluted into a 25 mL LB medium containing 50 µg/mL kanamycin, 34 µg/mL chloramphenicol and 0.5 mM unnatural amino acid 2NpOH for further culturing. 1 mM IPTG was induced into the culture, when the culture reached OD600=0.4-0.6, to boost the expression of the Z domain protein. After continue culturing overnight at 37 °C, 20 mL culture was then pelleted at 5000 rpm under 4 °C, the upper supernatant was discarded. The pelleted cell was then resuspended with 20 mL ice cold 10% glycerol, then the cell was pelleted again, in order to wash away the unwanted chemical, including the free 2NpOH. The wash procedure repeated twice more. Finally, the cell pellet was resuspended with buffer solution, and then aniline and formaldehyde were added accordingly. Cell with wild type of Z-domain protein was prepared similarly. All mixture was shook overnight. Then the cell was pelleted again to harvest the product. Z domain protein with a 6His tag at the C-terminus was purified under native condition. Cell pellets were resuspended and lysed for 20 minutes with 1 ml BugBuster (Novagen) supplemented with 0.5 mg/ml lysozyme (Sigma), and 1 U/mL bezonase (Novagen). Cell lysates were centrifuged at 7,000 rpm, and the supernatant were collected to with 0.25 mL Ni-NTA agarose (Qiagen). The mixture was gently shaken on a rotary shaker at 4 °C for one hour and loaded into a polypropylene column (Qiagen). Columns were washed with 2 mL wash buffer (50 mM NaH_2PO_4, 300 mM NaCl, 20 mM imidazole, pH 8.0) and eluted with 0.25 mL of native elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein solution was desalted and concentrated with micro-centrifugal filter. The resulting sample was sent to the California Institute for Quantitative Biosciences (QB3) for mass spectrometry.

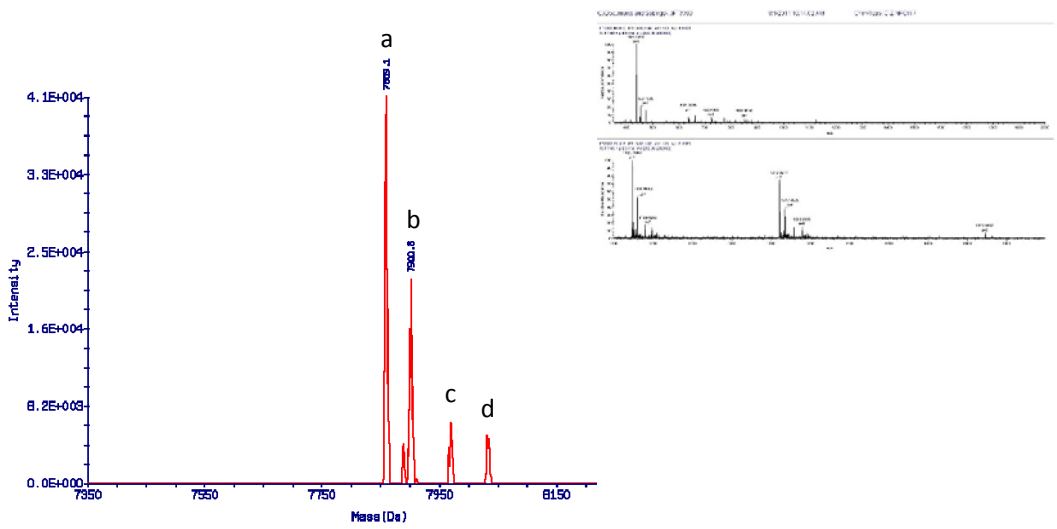


Figure 48. The ESI-MS spectrum for the Mannich reaction of 2NpOH-Z protein. Peak a and peak b are indicating the 2NpOH-Z protein with a 12 Da increment of mass. Peak c and d are impurities.

The molecular weights of 2NpOH incorporated Z domain protein at position 7 are 7978.63 Da (full length protein), 7847.43 and 7889.47 (full length protein without the first methionine and with post-translational modification acetylation). The expected mass increases after the bio-conjugation reaction were 150 for 4-nitroaniline and 162 for 4-aminoacetanilide. Interestingly, instead of a significant mass increase of 150 or 162 Da, only a mass increase of 12 Da was found.

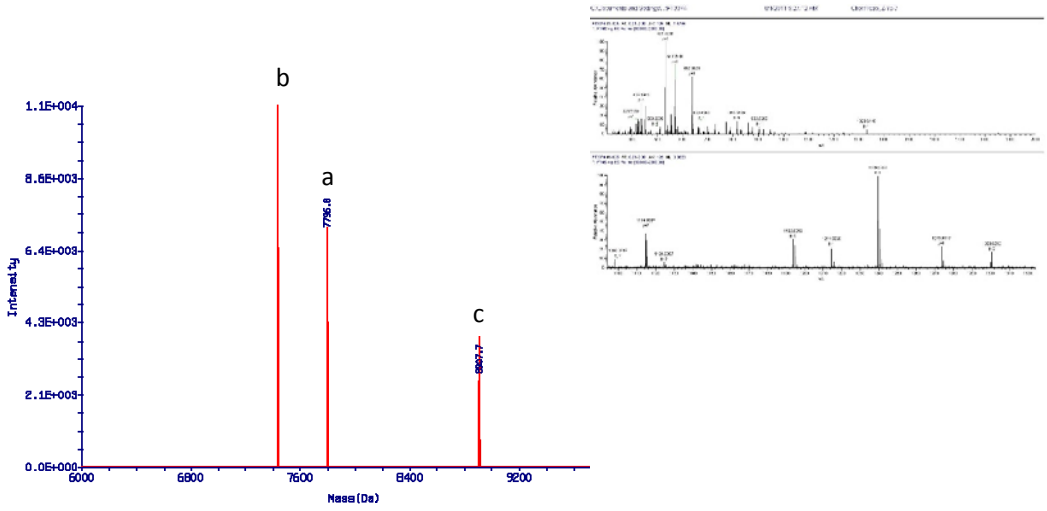


Figure 49. The ESI-MS spectrum for the Mannich reaction of Tyr-Z protein. No mass increase was observed. Peak a stands for the mass of unmodified Tyr-Z protein. Peak b and c are contamination.

As a control experiment, no such increase was observed for wild type of z-domain protein. (Figure 49) The molecular weights of wild type Z domain protein at position 7 are 7928.57 Da (full length protein), 7797.37 and 7839.41 (full length protein without the first methionine and with post-translational modification acetylation)

To further check whether this mass increase was caused by the reaction or the systemic error. A simple experiment was designed. The purified Z domain protein with 2NpOH and wild type Z domain protein were treated with aniline and formaldehyde respectively. After the reaction, the protein solution was divide into two portions, one was directly sent for Mass-spectrometry, and the other was doped with unreacted pure protein solution then sent for mass-spectrometry.

Based on the mass-spectrometry result, for 2NpOH-Z protein, not only a mass increase of 12 still could be seen, but also a mass increase of 30 was found on the mass spectrometry for Z domain protein with 2NpOH (Figure 52). However, as for wild type of Z domain protein, no such increase there was (Figure 51). Instead, there were peaks showed that the wild type of Z domain protein was partially modified with a mass increase of 174 (with 4-aminoacetanilide as reactant).

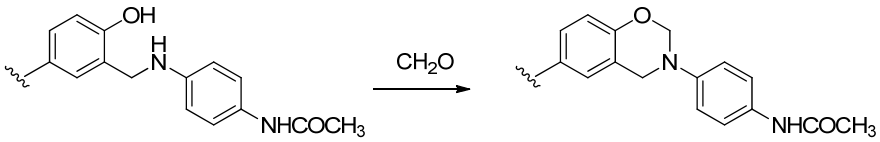


Figure 50. A possible cyclic acetal type of product

To rationalize the mass increase after the reaction, we start with the 174 Da, which are close to the expected 162 Da. Thinking of with excess of formaldehyde, an acetal type of structure could form (Figure 50), which would be a reasonable explanation for the mass increase of 174.

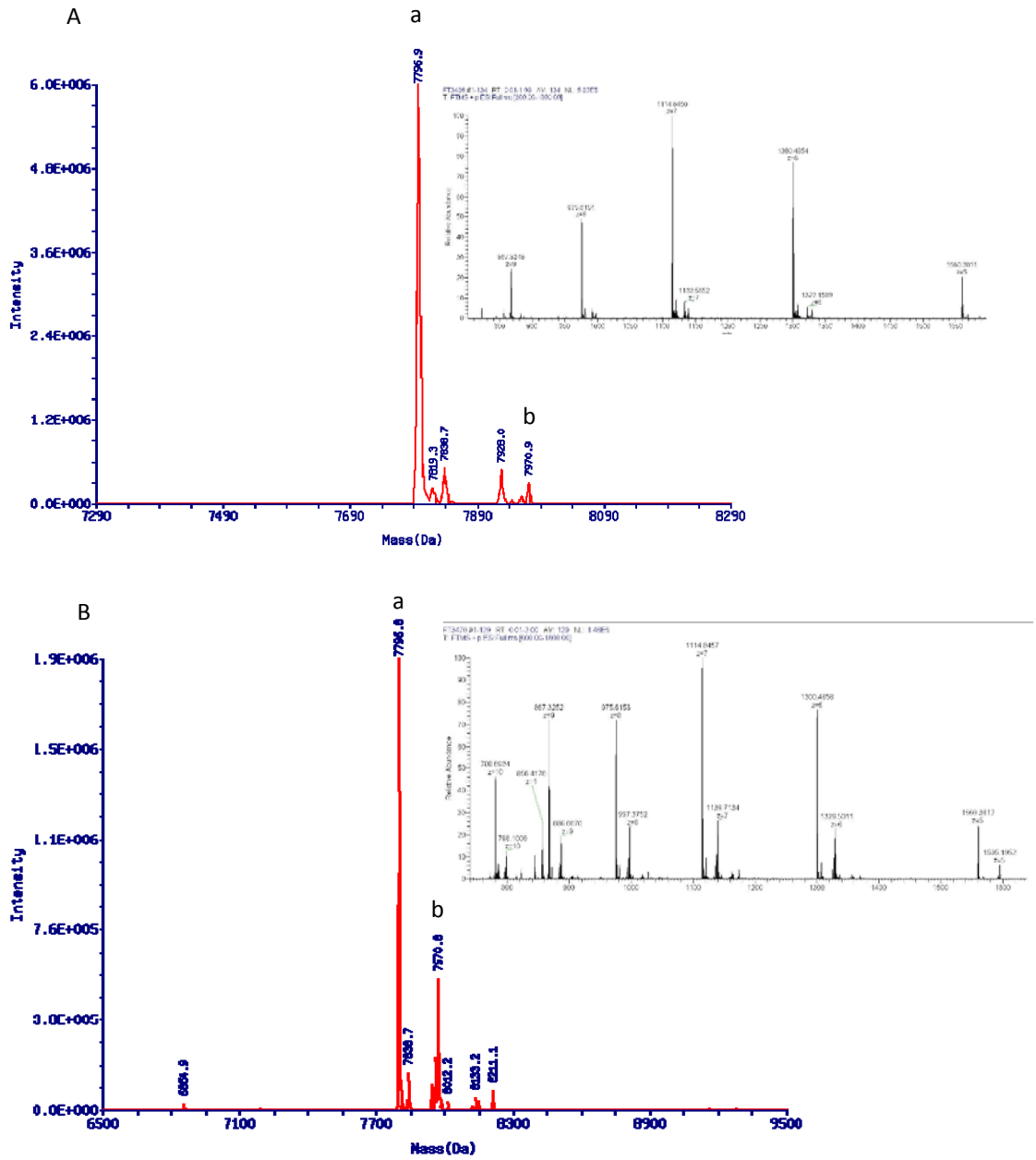


Figure 51. The ESI-MS spectrum the Mannich reaction of Tyr-Z protein. A: Protein solution after the reaction doped with Tyr-Z protein. B: Protein solution after the reaction. Peak a matched with the unmodified Tyr-Z protein. Peak b showed a mass increase of 174 Da than the unmodified Tyr-Z protein.

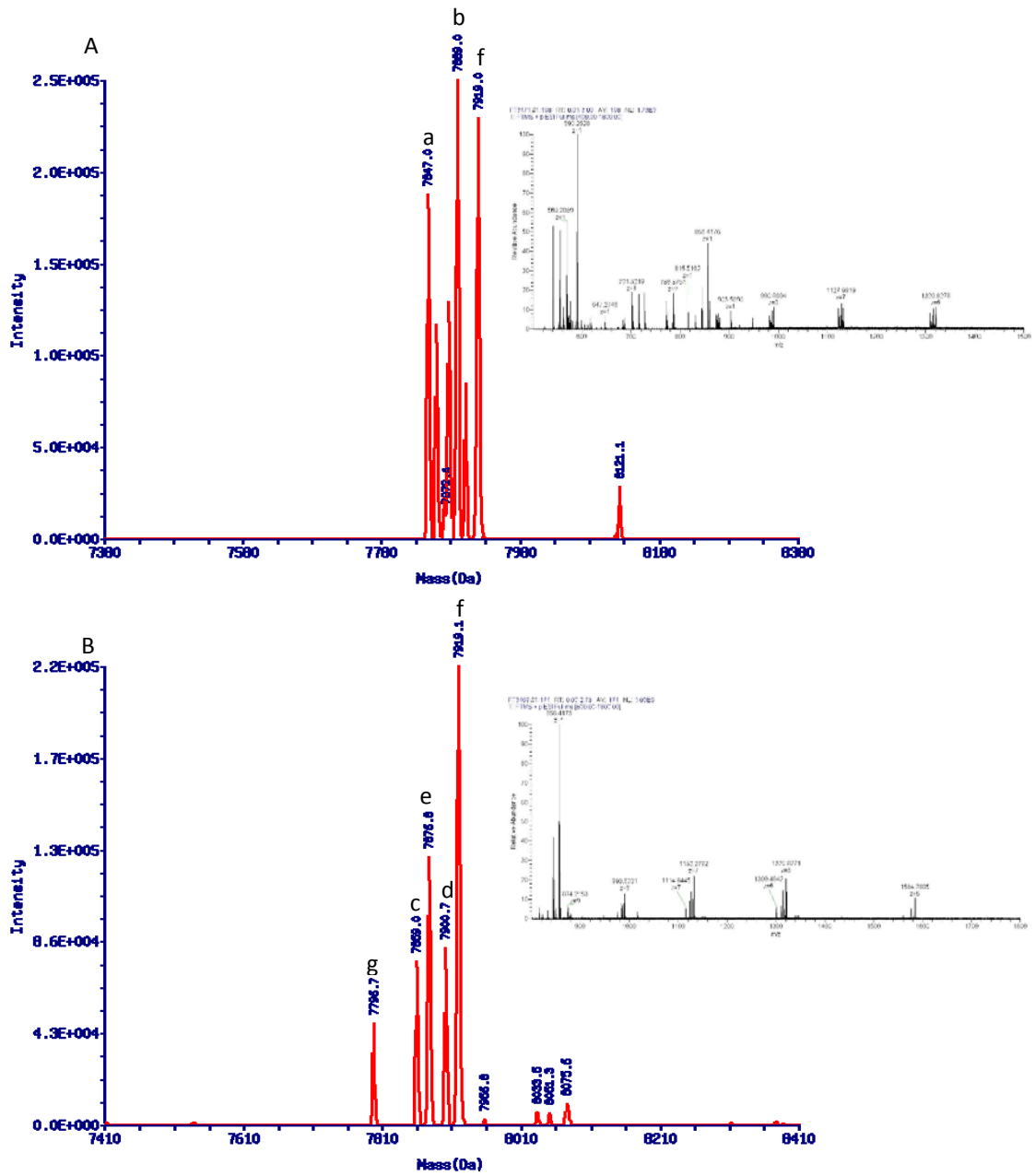


Figure 52. The ESI-MS spectrum the Mannich reaction of 2NpOH-Z protein. A: Protein solution after the reaction doped with 2NpOH-Z protein. B: Protein solution after the reaction. Peak a and b matched with the unmodified 2NpOH-Z protein. Peak c and d showed a mass increase of 12 Da than the unmodified 2NpOH-Z protein. Peak d showed a mass increase of 30 Da. Peak f and g were impurities.

However, the mass increase of 12 or 30 for the Z-domain protein with 2 NpOH was still remain unsolved. Mass increase of 30 showed that some type of conjugation with only formaldehyde there might be, most likely a hemiacetal type of product with lysine or serine. Following this thought, we checked the crystal structure of Z-domain protein. It appeared that, a lysine residue was only 3.43 Å away from the 2NpOH site. The lysine side chain could work as the role of primary amine in this Mannich type reaction.

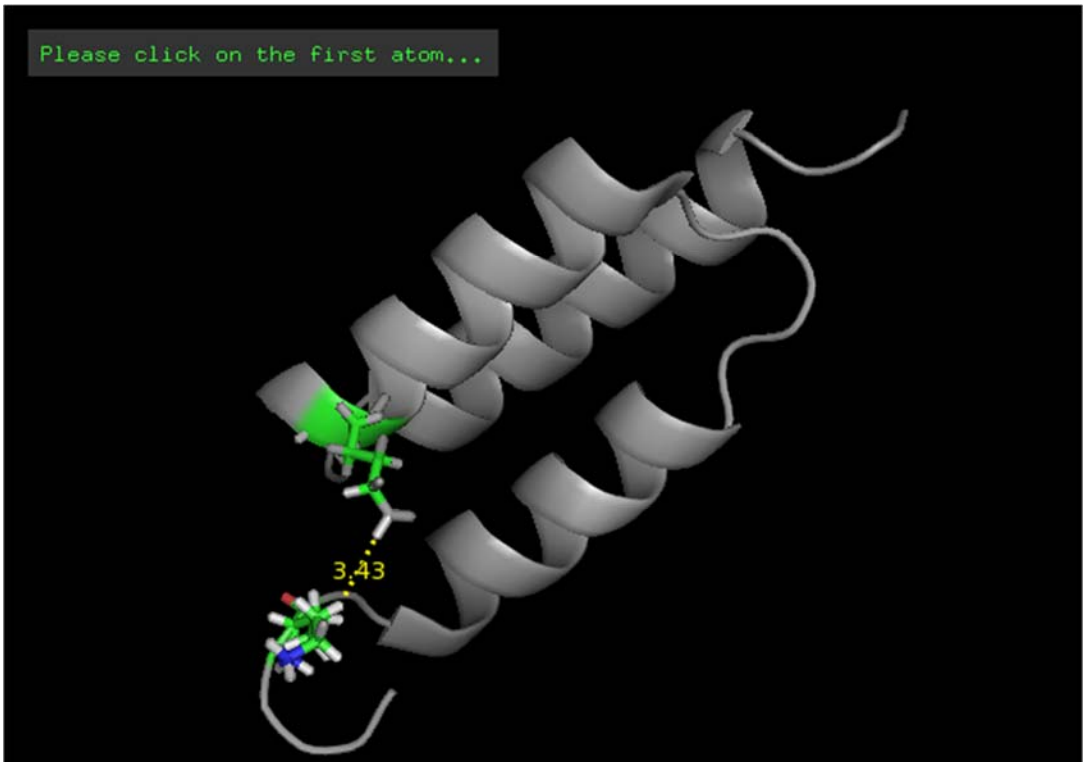


Figure 53. Crystal structure of Z-domain proteins (PBD ID:1Q2N). 2NpOH[7] is only 3.43 Å away from Lys[38].

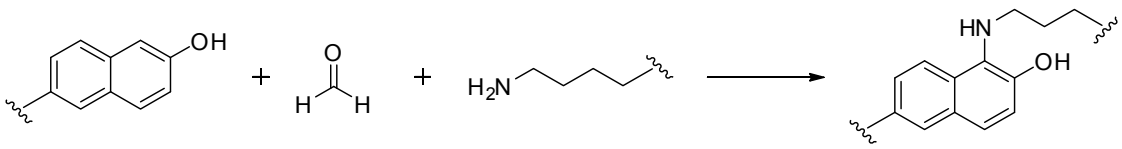


Figure 54. A possible Mannich type reaction between 2NpOH residue and Lysine residue with presence of formaldehyde.

There were reports showed that primary amine could proceed through the Mannich type reaction with tyrosine[51] or naphthol[52]. These reactions were typically carried out using high concentrations of formaldehyde or under significant high temperature with the risk of destroying the secondary and tertiary protein structure.

Thus, with relative low concentration of formaldehyde, mild temperature and proximity of the reactant, we could assume that the reactivity of tyrosine could be limited so that the selectivity for 2NpOH could be achieved.

4.4 Formaldehyde-based crosslinking with lysine residue.

This brought our interest immediately, not only because this formaldehyde based crosslinking reaction can be simply proceed just mixing the protein solution or the protein containing cell with formaldehyde, but also it could be potential method for the production of cyclic peptide. To further investigate this coupling reaction, several experiments were carried out.

First of all, we thought the proximity played an important role here. To testify this hypothesis, a model GCN4 leucine zipper protein was used. As we known, the leucine zipper protein had a super-secondary structure that functions as dimerization domain, and its presence generates adhesion forces in alpha helices,[53] which forms a parallel coiled coil homo-dimer.[54] With these properties, we could genetically design to incorporate unnatural amino acid into certain site, so that the distance to a close-by lysine could be controlled.

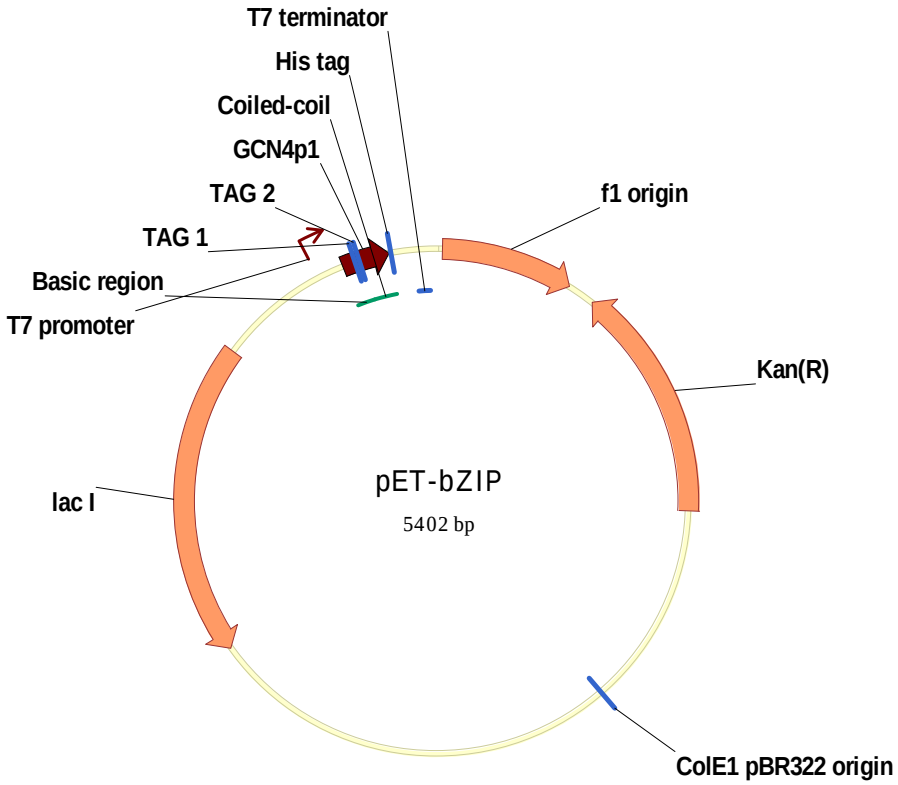


Figure 55. Vector pET-bZIP

Three different version mutation of GCN4 leucine zipper protein monomer were designed: a wild type one, two with 2NpOH substitute amino acids at selected sites. From Vector pET-bZIP (Figure 55) which contains the wild type GCN4 leucine zipper protein plus 6His tag gene, two specific mutations were introduced into permission site: TAG stop codon was inserted to substitute Lys^{22} and Arg^{17} respectively.

The amino acid sequence of GCN4-A (wild type) protein is

H_2N-mdpaalkrarnteaarrsrarklqrmkqledkveellsknyhlenevarlkklvgelehhhhhh-COOH.

The amino acid sequence of GCN4-B (Lys^{22} to 2NpOH) protein is

H_2N-mdpaalkrarnteaarrsrar*lqrmkqledkveellsknyhlenevarlkklvgelehhhhhh–COOH.

The amino acid sequence of GCN4-C (Arg^{17} to 2NpOH) protein is

H_2N-mdpaalkrarnteaar*srarklqrmkqledkveellsknyhlenevarlkklvgelehhhhhh–COOH.

* denotes the unnatural amino acid, 2NpOH

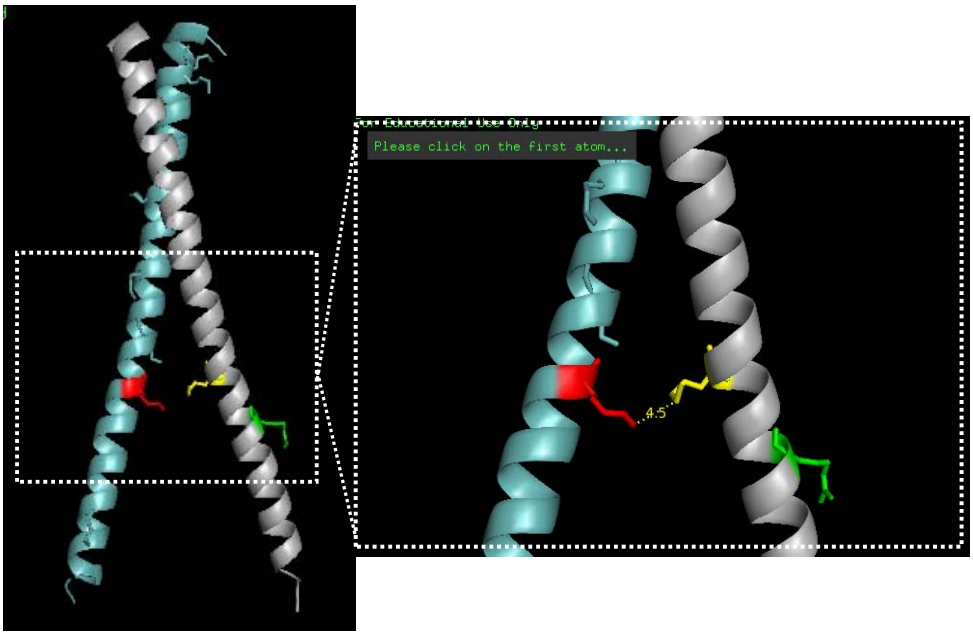


Figure 56. Crystal structure of dimerized GCN4 leucine zipper protein(Figure generated is with PyMOL based on the crystal structure of GCN4 Basic Region of Leucine Zipper Protein, PDB ID: 1YSA[55]). 2NpOH mutant site showed in yellow (GCN4-B) and green (GCN4-C), original Lys^{22} showed in red.

Based on the crystal structure of GCN4 leucine zipper protein's homo-dimer. GCN4-B has the 2NpOH at the site only 4.5 Å away from the original Lys^{22} of the paired monomer. With our hypothesis, if the coupling reaction happened, the two monomer would be covalent linked, so that the SDS-PAGE electrophoresis analysis would show a dimer band instead of monomer band after heat denaturing. On the other hand, GCN4-C has the 2NpOH at the site is away from any lysine. So that after the treatment with formaldehyde, no covalent linkage should form; and it would show a monomer pattern band in the SDS-PAGE analysis. Besides, since there is no nearby lysine, GCN4-C could also be used in the test for the three-component Mannich type reaction.

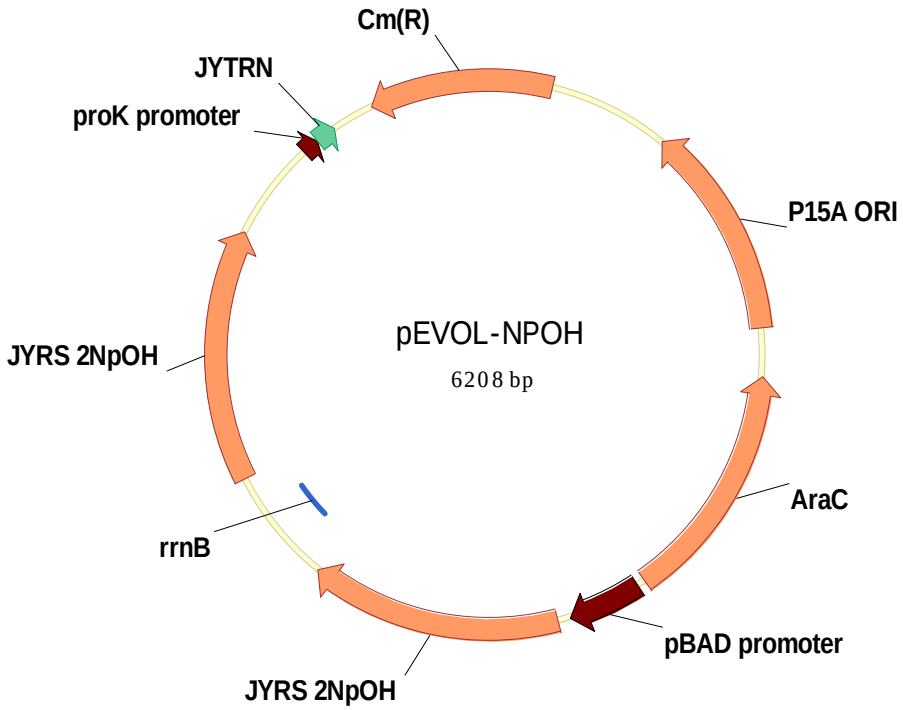


Figure 57. Vector pEVOL-NPOH

A pEVOL-NpOH verctor was constructed. The vector contains two copies JYRS-2NPOH gene, which could increase cytoplasmic the amount of tRNA synthetase. The leucine zipper protein expression vectors (GCN4-B, GCN4-C) as well as the pEVOL-NpOH were co transformed into *E. coli* BL21(DE3) cells, which were then grown in LB media with the presence of 0.5 mM 2-NpOH respectively. Wild type leucine zipper protein (GCN4-A) was also expressed separately. The three version of leucine zipper proteins were purified by Ni^{2+} affinity column.

GCN4-A was mix with GCN4-B and GCN4-C respectively, as one to one molar ratio according to the UV absorbance behavior. There should be a statistical mixture of a 2:1:1 ration of Mutant/Wild type, Mutant/Mutant, and Wild type/Wild type leucine zipper dimers. However, upon the treatment of formaldehyde, and

then electrophoresis under denatured condition, no dimer was observed.

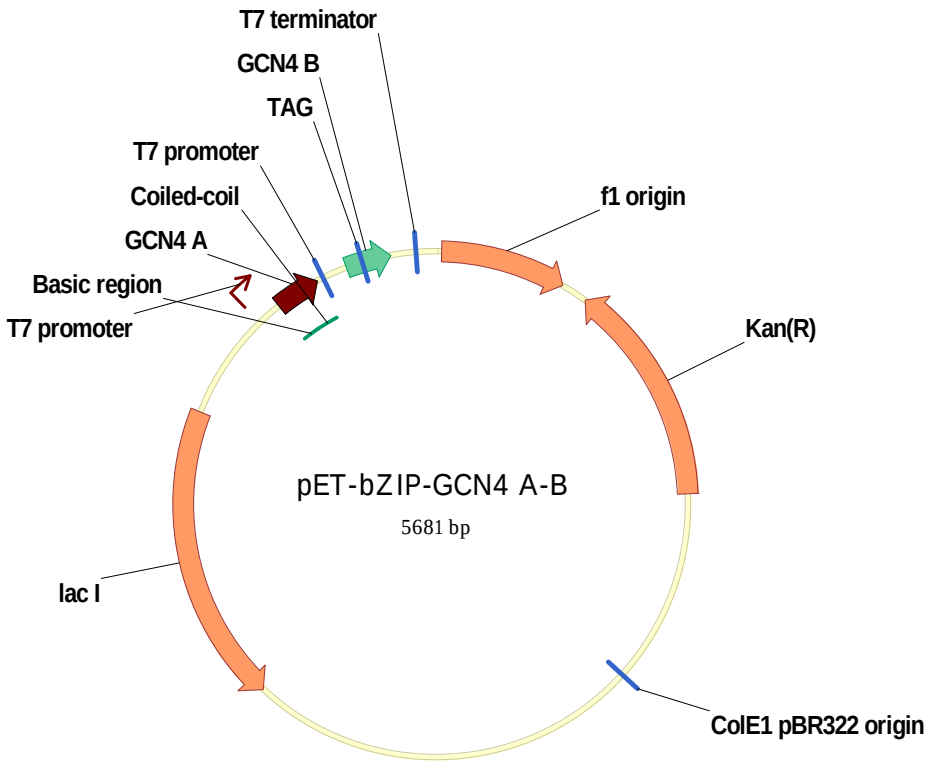


Figure 58. Vector pET-bZIP-GCN4 A-B

To avoid any salts during the purification process affecting the reactivity, and thinking of that the reaction could be able to carry out *in vivo*, another vector containing both GCN4-A and GCN4-B gene was constructed and co-transformed with pEVOL-NpOH in to *E. coli* cells, to co-express both wild type and mutant leucine zipper proteins (GCN4-B). Notably, only GCN4-B protein gene was fused with a 6His tag, so that any excess of GCN4 A would be washed away during the purification process.

Then the cell was pelleted and washed, then resuspended with pH 7 phosphate buffer with formaldehyde. After 20 hours, the cells were lysed and the protein was collected, then purified and desalted. However after electrophoresis analysis, no proteins, both monomer and dimer were identified.

Since the protein scale test was not successful. We wanted to confirm the reactivity. A small molecule test was designed. 100 µL 0.5 M of p-cresol or 2-naphthol was dissolve in the 800µL 0.2 M pH 7 phosphate buffer, then 100 µL 0.5 M of n-butyl amine and 3.7 µL of 37% formaldehyde was added. The mixtures were set on shaker over night at room temperature. The resulting reaction mixture was directly injected to LC-MS. With the LC-MS result, two new product was found for 2-naphthol, with the molecular mass of 229 and 241; meanwhile, no significant

product was identified for p-cresol. And the molecular mass matched the coupling product as we expected.

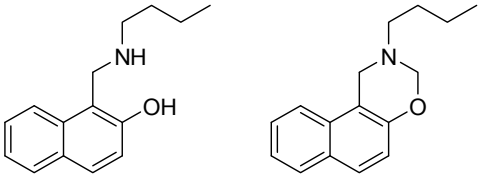


Figure 59. Two possible coupling product

Based on this result, the reaction seemed to be promising. The problem might be the protein candidate. The distance and the orientation of the 2-naphthol residue to the lysine would both be critical here. Since the Z domain protein gave a positive hit; and based on our hypothesis, after the coupling reaction, the Z domain protein should form an internal covalent linkage. Thinking of the different behavior of cyclic peptide and linear peptide, we could assume that after the reaction, the Z-domain protein should show at the different band in the electrophoresis SDS-PAGE analysis.

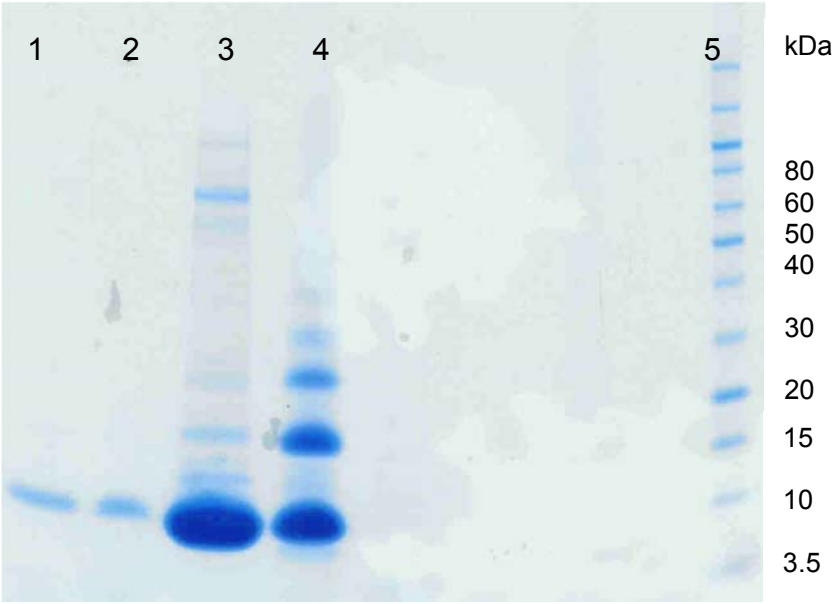


Figure 60. SDS-PAGE analysis for the formaldehyde-based crosslinking reaction of 2NpOH-Z protein. Lane 1: Tyr-protein before the coupling reaction; lane 2: Tyr-protein after the coupling reaction; lane 3: 2NpOH-protein before the coupling reaction; lane 4: The coupling product of 2NpOH-Z protein; lane 5: molecular mass marker The SDS-PAGE gel was stained with SimplyBlue SafeStain reagent.

Both concentrated wild type Z domain protein and Z domain protein with 2NpOH were treated with 0.1 M formaldehyde in pH 7 buffer under room temperature. After

20 hours; then the reaction mixtures were purified with Aspire chromatography tips. Then electrophoresis analysis applied.

Interestingly, the mutant Z domain protein with 2NpOH showed a dimer band, a trimer band and even a fade tetramer band (Figure 60); no changes were observed for the wild type protein. By checking the protein sequence of Z-domain, there were five lysine on the surface. Possible cross coupling reaction happened among the proteins due to high concentration of the protein molecule.

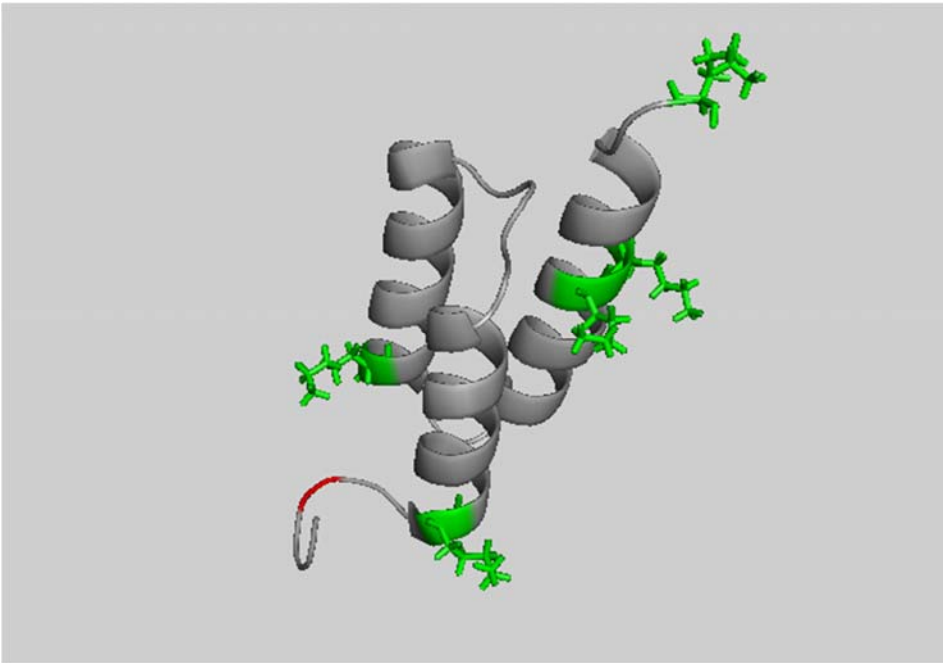
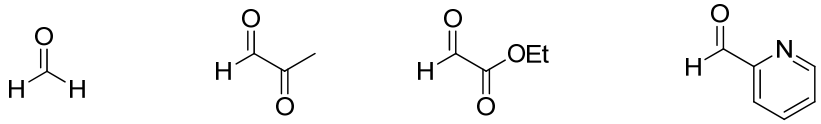


Figure 61. 2NpOH-Z protein (Figure is generated with PyMOL based on the crystal structure of Z domain Protein, PDB ID: 1Q2N[43]) with 5 lysine residue showing in green.

Besides this, another experiment was carried out to test the formaldehyde-based coupling reaction. A series of aldehydes were selected to treat the dilute solution of Z domain proteins with 2NpOH, and then send the resulting protein to QB3 for mass spectrometry analysis.



formaldehyde MW=30.03 | pyruvaldehyde MW=72.06 | Ethyl glyoxalate MW=102.09 | pyridine-2-carbaldehyde MW=107.11

Figure 62. Four different aldehydes were tested in this coupling reaction

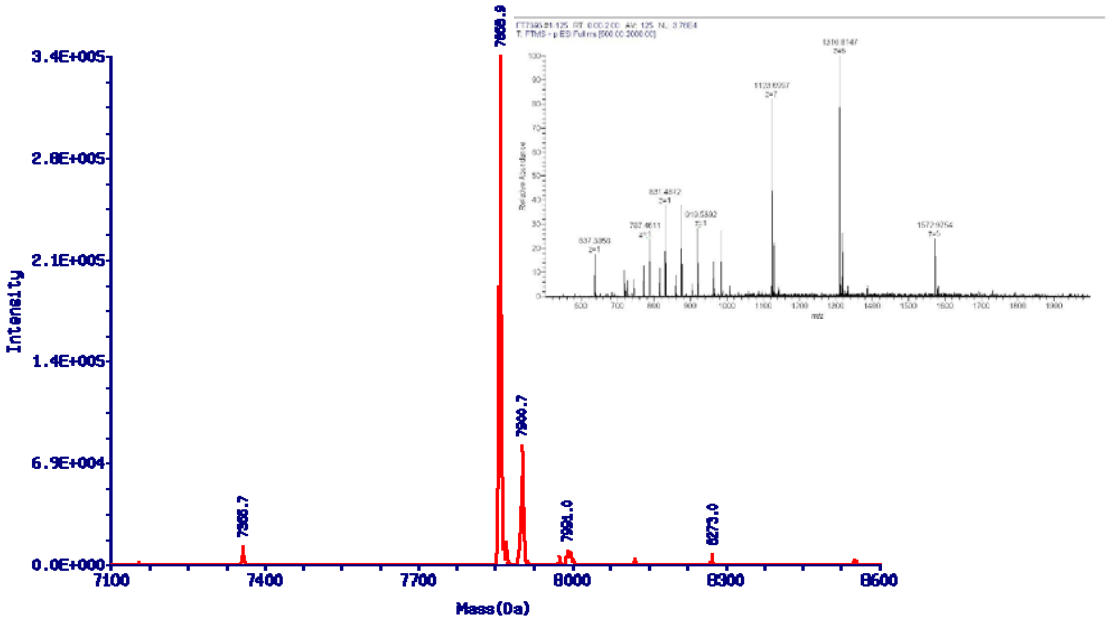


Figure 63. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with formaldehyde. A mass increase of 12 Da was observed as expected. No peak for the unmodified protein was observed.

Both formaldehyde and pyridine-2-carbaldehyde were showed a mass increase as expected, 12 and 89 respectively. However, pyruvaldehyde showed an increase of 12 also, maybe suggesting a contamination of formaldehyde. As for ethyl glyoxalate, a mass increase of 102 was observed, which match the molecular weight of the aldehyde, suggesting a simple addition reaction was involved here.

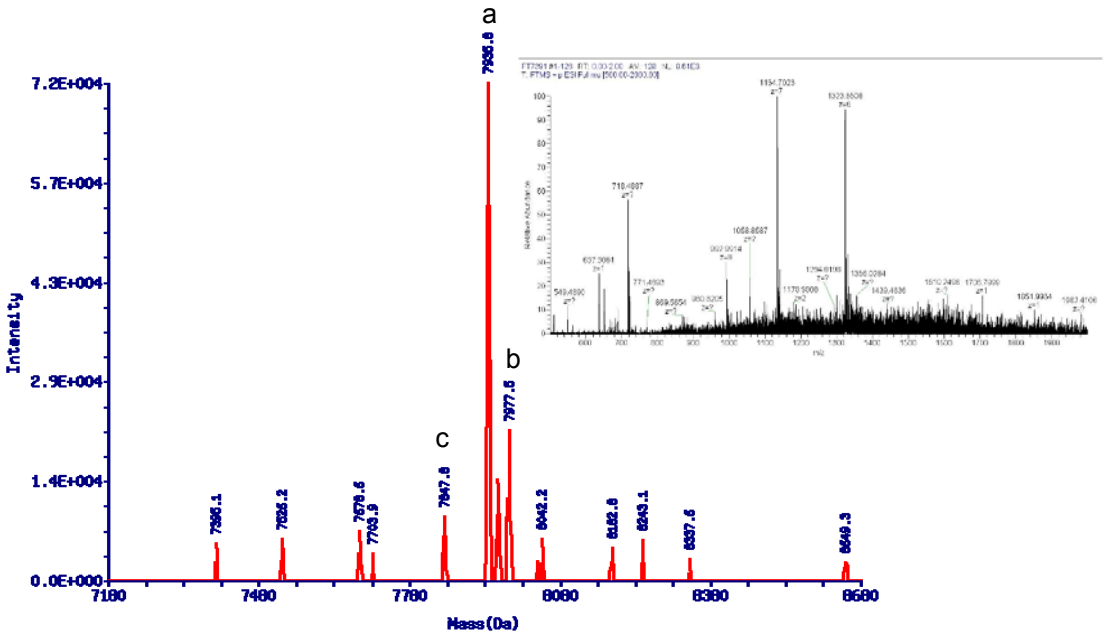


Figure 64. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with pyridine-2-carbaldehyde. Peak a and b matched with the mass increase of 89 as expected. Peak c showed a small portion of unreacted 2NpOH-Z protein.

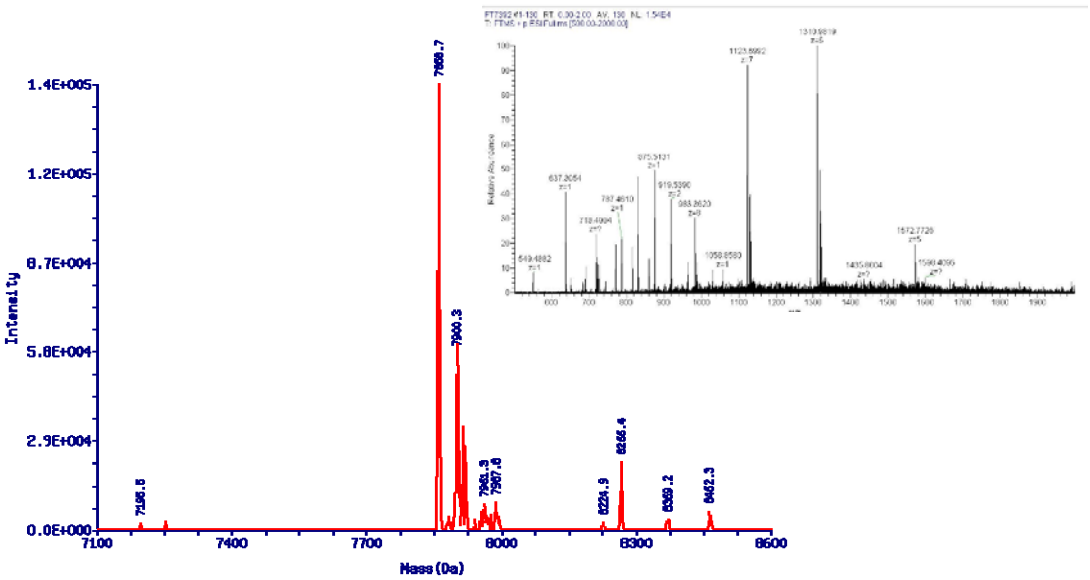


Figure 65. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with pyruvaldehyde. Interestingly, a mass increase of 12 Da, indicating a formaldehyde coupling reaction, was observed.

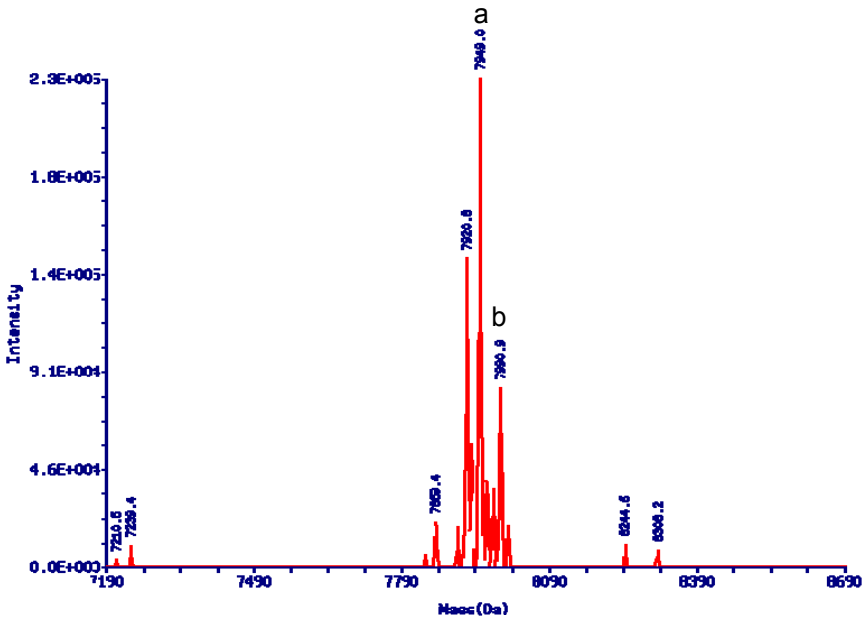


Figure 66. The ESI-MS spectrum for the coupling reaction of 2NpOH-Z protein with ethyl glyoxalate. Peak a and b showed a mass increase of 102, which matched the molecular weight of ethyl glyoxalate.

4.5 Discussion

The small molecule experiment data so far showed that this three-component Mannich type coupling reaction had some selectivity to 2NpOH residue prior to tyrosine residue. However, under physiological condition, which was the ideal condition for *in vivo* bio-conjugation reaction, the electron-rich alanine might not be a good candidate for this reaction. Our experiment (Figure 51) showed that the tyrosine could be partially modified, which also matched with Professor Matthew Francis's work.[50] Further study of this type reaction could stay focus on the electron-deficient anilines.

As for the lysine participated coupling reaction, our results showed significant difference between 2NpOH and tyrosine residues. Here we hypothesis that a nearby lysine could be ligated to the 2NpOH with the presence of certain aldehyde. The protein scale ligation with leucine zipper protein suggested that the orientation and the distance between lysine and 2NpOH might critical. However, the recent study (Figure 60) showed that the reactivity might be considerable, and might not be limited by the proximity at higher protein concentration.

With all the evidence we had so far, we can assume, this aldehyde-based lysine-2NpOH coupling reaction could be quite useful considering the high conversion achieved *in vivo* (Figure 48 and Figure 63). Even with the possibility of cross-linking

reaction, it still might give us a powerful tool to generate cyclic peptides and cyclic proteins.

Chapter 5

Other endeavor and future work

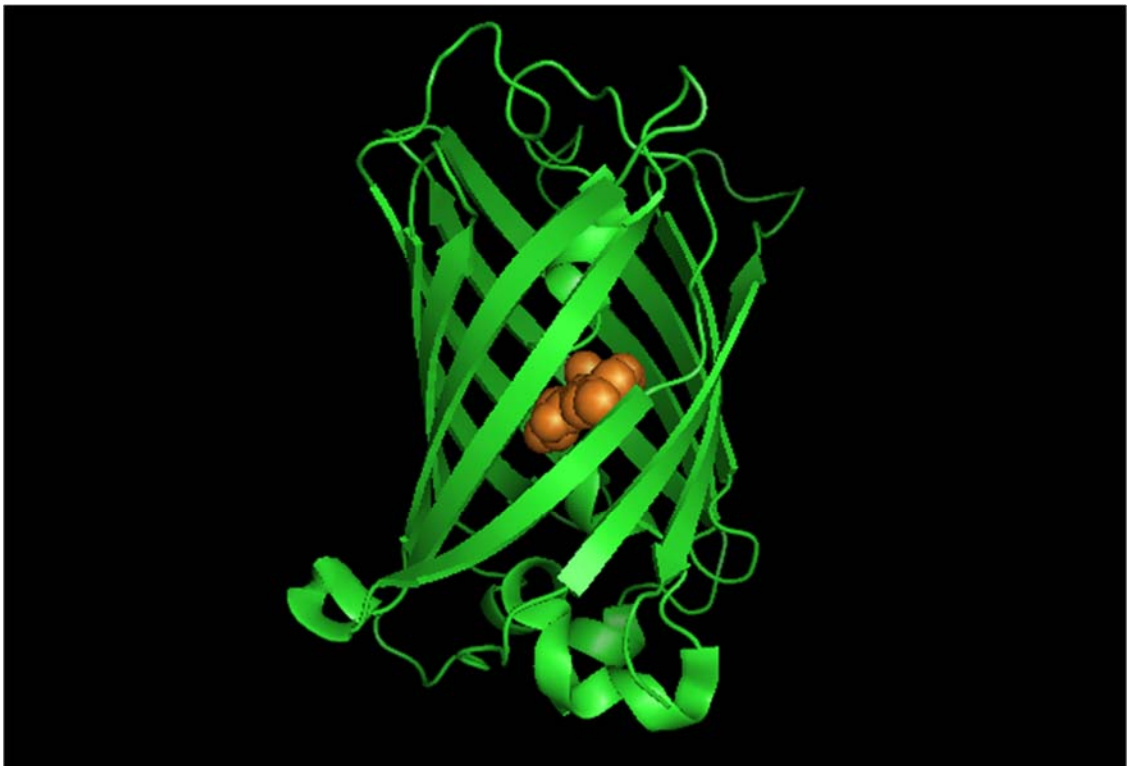


Figure 67. Green fluorescent protein (GFP) (Figure is generated with PyMOL based on the crystal structure of an Enhanced Green Fluorescent Protein with 374-nm Absorbing Chromophore, PDB ID 1Z1Q[56])

5.1 Application of the incorporation of 2-amino-3-(6-hydroxy-2-naphthyl) propanoic acid for directed protein evolution

One possible advantage brought by genetic incorporation of an unnatural amino acid is that it could enhance properties of proteins. Previously we have mentioned that the fluorescent property of 2NpOH is unique compared to any other natural amino acid, we want to utilize this to alter the feature of some protein molecules.

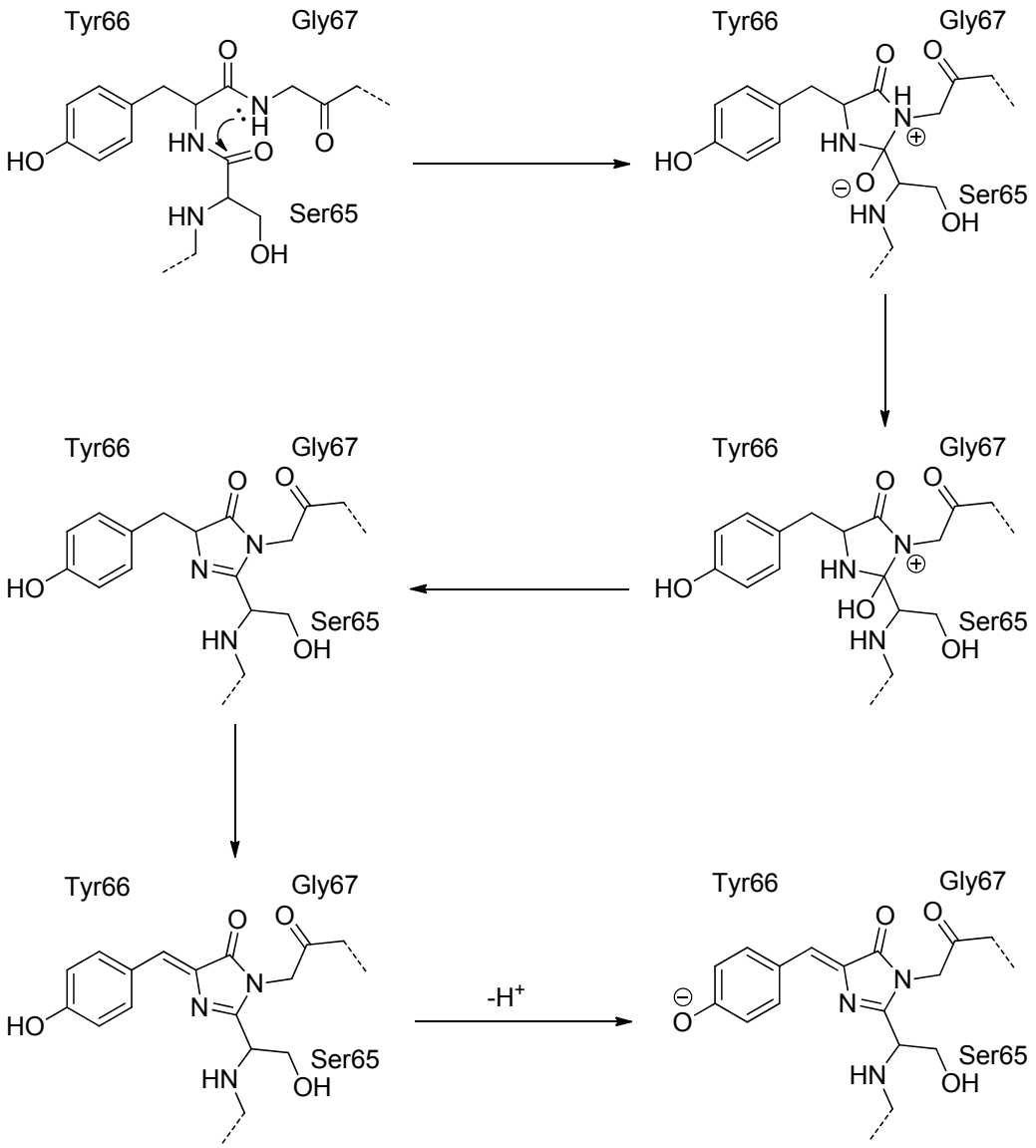


Figure 68. Mechanism for the chromophore formation in Green Fluorescent Protein.[57]

The GFP is a protein that exhibits bright green fluorescence when exposed to light in the blue to ultraviolet range[58]. The fluorescent property is coming from a unique chromophore, which is formed by the covalent modification of three adjacent amino acids: Ser^{65}-Tyr^{66}-Gly^{67}.[58-59]

Genetic incorporation of unnatural amino acid can enhance the property of proteins. Thinking of the similarity and unique fluorescent properties, one may expect that the incorporation of this 2NpOH to replace the Tyr66 residue of GFP would change the fluorescence to a longer wavelength region (Figure 69), due to the extension of π-conjugation system in the chromophore structure.

To achieve this, a library was designed and created, based on the crystal structure of GFP. GFP has a unique barrel-like structure formed by *β*-sheets[57] (Figure 67), with a rigid size and shape. To minimize the loss of fluorescence from unfolding, a super-charged super-folded GFP variant, GFP(+36)[60] was used as the starting point. Also, a library based on GFP variant GFPuv[61] was also used to create more diversity. We assumed that, after protein expression with 2NpOH, a GFP mutant with a red-shift fluorescent wavelength might be obtained.

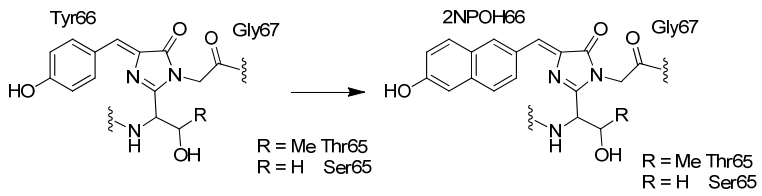


Figure 69. Incorporate 2NpOH to substitute tyrosine in GFP chromophore

Amino acid sequence for this GFP mutant GFP(+36) fused with a 6His tag: (* denotes the unnatural amino acid, 2NpOH)

MTSKGERLFRGKVPILVELKGDVNGHKFSVRGKGKGDATRGKLTLKFICTTGKL
PVPWPTLVTTLT*GVQCFSRYPKHMKRHDFFKSAMPKGYVQERTISFKKDGKY
KTRAEVKFEGRTLVNRIKLKGRDFKEKGNILGHKLRYNFNSHKVYITADKRKNGI
KAKFKIRHNVKDGSVQLADHYQQNTPIGRGPVLLPRNHYLSTRSKLSKDPKEKR
DHMVLLEFVTAAGIKHGRDERYKLEHHHHHH

Amino acid sequence for GFPuv mutant with a 6His tag: (* denotes the unnatural amino acid, 2NpOH)

MTSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKL
PVPWPTLVTTFS*GVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIK
ANFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRD
HMVLLEFVTAAGITHGMDELYKLEHHHHHH

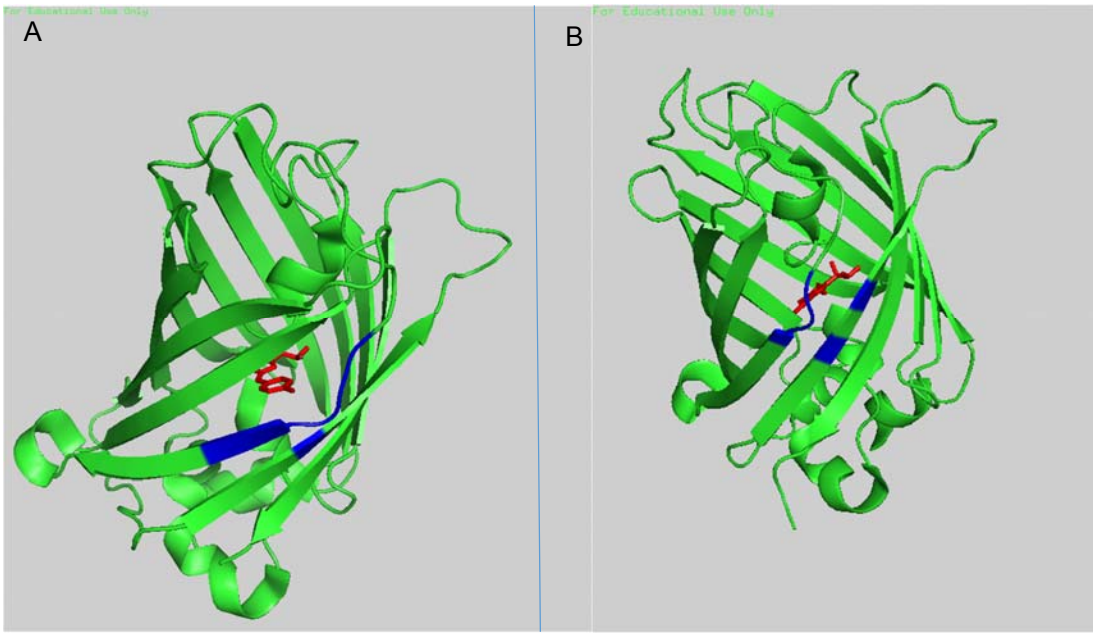


Figure 70 The crystal structure of A:GFP(+36) (Figure is generated with PyMOL based on the crystal structure of a superfolder Green Fluorescent Protein mutant, PDB ID 4LQU[62]), B: GFPuv (Figure is generated with PyMOL based on the crystal structure of an Enhanced Green Fluorescent Protein with 374-nm Absorbing Chromophore, PDB ID 1Z1Q[56]). The chromophore is showed in red. The nutation site is showed in blue.

In order to accommodate the larger size of 2NpOH, the region close to the phenol residue of the chromophore was introduced random mutation. Three different libraries were constructed. GFP-Library1, derived from GFP(+36), was designed as this: seven amino acid residues possibly related to the tolerance of the unnatural amino acid were selected as the followings: Phe^{146}, Asn^{147}, Ser^{148}, His^{149}, Lys^{150}, Val^{151} and Thr^{204}. GFP-Library2 and GFP-Library3, derived from GFPuv, were designed as these: 5 amino acid residues were selected as the followings: Asn^{147}, Ser^{148}, His^{149}, Thr^{204} and Ser^{206}. The difference between GFP-Library2 and GFP-Library3 is that, at site of Asn^{147}-Ser^{148}-His^{149}, GFP-Library2 introduced 3 amino acid mutations, while GFP-Library3 introduced 4 amino acid mutations, in order to create more space for 2NpOH incorporation.

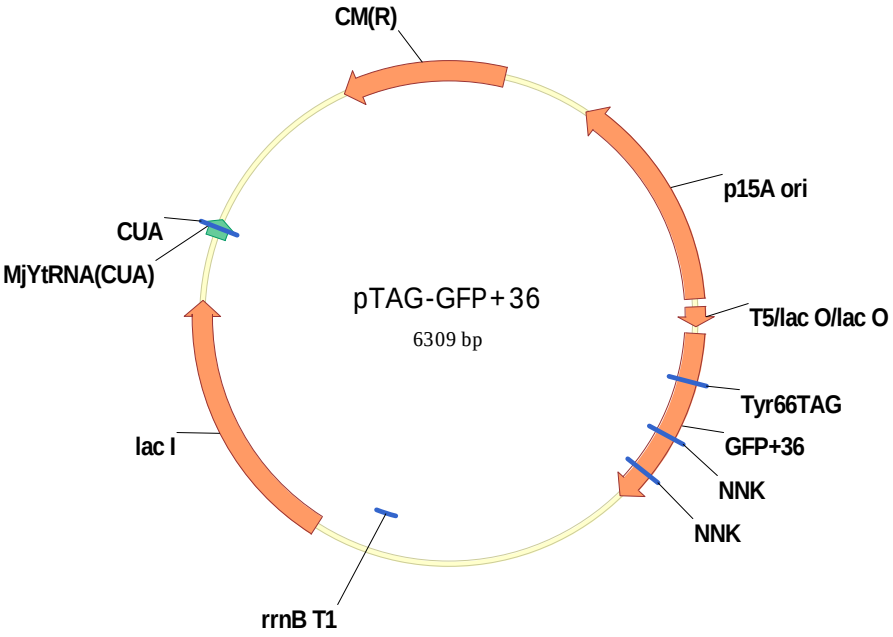


Figure 71 Vector pTAG-GFP+36

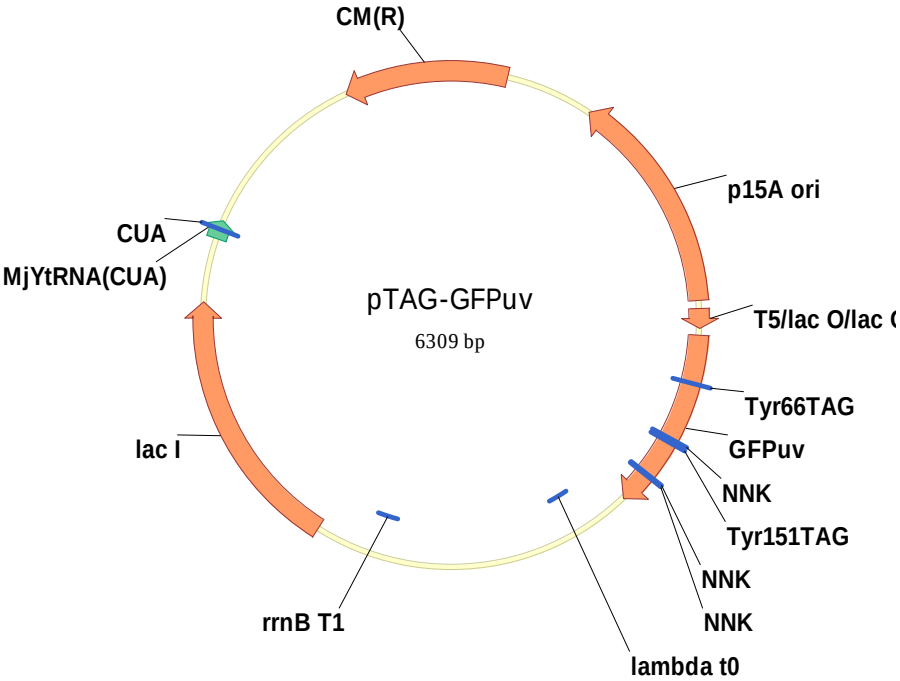


Figure 72. Vector pTAG-GFPuv

To construct the library, random mutations were created at the chosen sites of the

GFP gene through regular PCR reactions and then overlapping PCR reactions with the following primers.

GFP20a: 5'- ACTATGACTAGTAAAGGAGAACGACTTTTC

GFP20b: 5'- CGTTTGTCTGCCGTGATGTA**MNNMNNMNNMNNMNNMNN**GTTG TATCGAAGTTTGTGTCCG

GFP20c: 5'- TACATCACGGCAGACAAACG

GFP20d: 5'- CTTGGGATCTTTCGAAAGTTTAGATCG**MNN**CGACAGGTAATGGT TGCGAGG

GFP20e: 5'- CTAAACTTTCGAAAGATCCCAAG

GFP06e: 5'- GTTGTTAGATCTTTATCAGTGGTGGTGGTGGTGG

GFP01b: 5'- TTAACTATGACTAGTAAAGGAGAAGAACTTTTCACTGG

GFP02b: 5'- GTTTGTCTGCCGTGATGTATACATT**MNNMNNMNN**ATAGTTGTATT CGAGTTTGTGTCCG

GFP02c: 5'- TTTGTCTGCCGTGATGTATACATT**MNNMNNMNNMNN**ATAGTTGT ATTCGAGTTTGTGTCC

GFP03: 5'- AATGTATACATCACGGCAGACAAAC

GFP04b: 5'- GGGATCTTTCGAAAGGGC**MNN**TTG**MNN**CGACAGGTAATGGTTG TCTGG

GFP05: 5'- GCCCTTTCGAAAGATCCCAACG

The diversity of the library was roughly 5×10^6 members for each.

The GFP mutant library DNA was co-transformed into *E. coli* Top 10 host cell with vector pAK3-NpOH. The cells were grown at 37°C on an LB-Agar medium with kanamycin, chloramphenicol, IPTG, arabinose, and 1mM 2NpOH for 2 days. The colonies were checked upon UV light excitation. However, no fluorescent behavior was observed.

There were three possible reason might cause the loss of the fluorescence. First of all, the simple incorporation of the 2NpOH site, might perturb the process of the maturation of the chromophore, which involved with proximal residue catalyzed post-translational modification. Secondly, the size of the 2NpOH residue might cause the wrongly folding of the protein, so that it would not afford the following post-translational modification. Lastly, after the incorporation the unnatural amino

acid, the quantum yield might decrease, due to the chemical environment changing, so that, the fluorescent might not be able to observe by naked eyes.

5.2 Incorporation of Multiple UAAs into proteins.

Genetic incorporation of unnatural amino acid is a powerful tool to introduce new chemistries into protein. Normally, this technique in general only allows the incorporation of a single UAA into single proteins, since the amber codon is the only one available for the incorporation of NAAs and the nonsense suppression rate in living cells is low.[63] However, there is need for incorporation two or even multiple different UAAs into a single protein. Thus, the use of an additional codon is required. Our co-worker, Dr.Yiyan Wang, has reported successfully incorporation of unnatural amino acid into protein using the AGG sense codon. Therefore, it is possible to incorporated two amino acid into a single protein, with combining both TAG and AGG system.

In order to efficiently encode multiple unnatural amino acids into proteins, we simultaneously reassigned AGG and UAG codons to different UAAs by introducing two aaRS/tRNA pairs, *Mj*RS_{AGG}/*Mj*$tRNA_{CCU}$ and *Mm*Pyl-RS_{UAG}/*Mm*$tRNA_{CUA}$, into *E. coli*. Dr. Yiyan Wang, reported 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, 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* and in mammalian cells.[64]

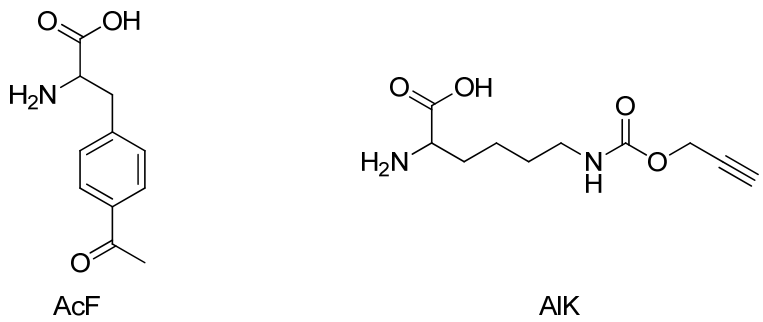


Figure 73. The structure of two unnatural amino acids used in this study

In this experiment, two unnatural amino acids were used, p-acetylphenylalanine (AcF) and N_{ε}-propargyloxycarbonyl-L-lysine (AlK). (Figure 73) AcF was purchased from commercial resource, AlK was synthesized as following. (Figure 74)

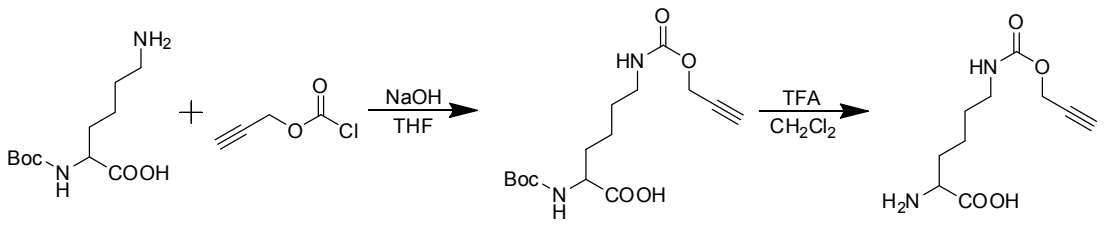


Figure 74. Synthesis of N_{ε}-propargyloxycarbonyl-L-lysine (AlK)

0.4926 g Boc-Lys-OH (2 mmol) was dissolved in the mixture of 5 mL 1 M NaOH (aqueous) solution and 5 mL THF, the resulting solution was cooled to 0 °C. Propargyl chloroformate (0.176 mL, 0.2133 g, 1.8 mmol) was added dropwise over 5 minutes and the reaction was allowed to stir for 10 hours at room temperature. The product, (S)-2-(*tert*-Butoxycarbonylamino-6-((prop-2-ynyloxy)carbonylamino) hexanoic acid was separated upon extraction. Then it was dissolved in 4 mL dry dichloromethane. TFA (4 mL) was added dropwise and the reaction was allowed to stir for 1h.The final product was separated by evaporating the solvent and following recrystallization. 0.3082 g of the final product was recovered.

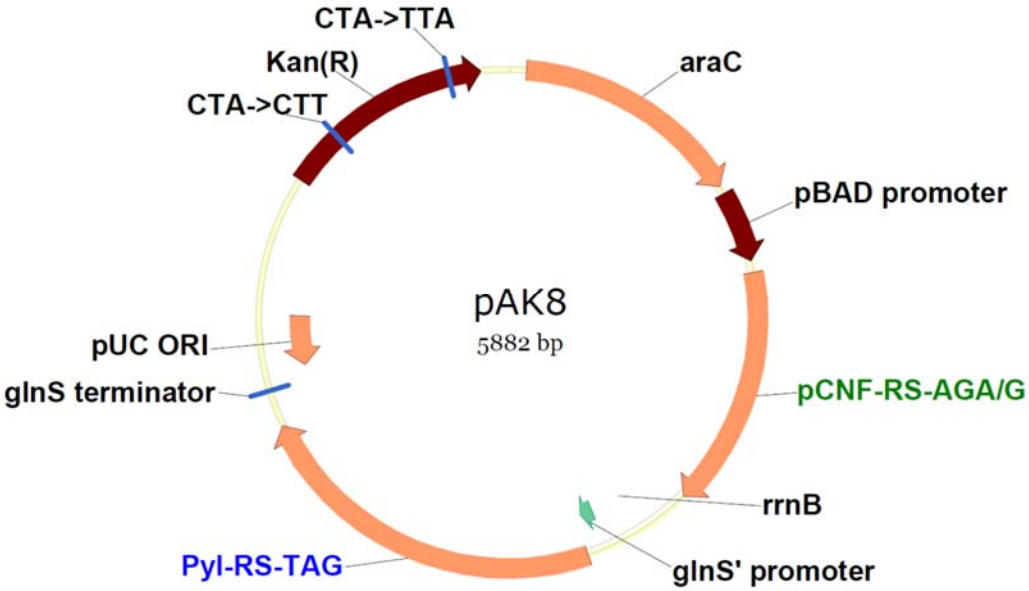


Figure 75. Vector pAK8

In order to incorporate two UAAs into protein, a vector carried two different RS genes was constructed. (Figure 75) And a mutant Z domain protein was designed as following: 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.

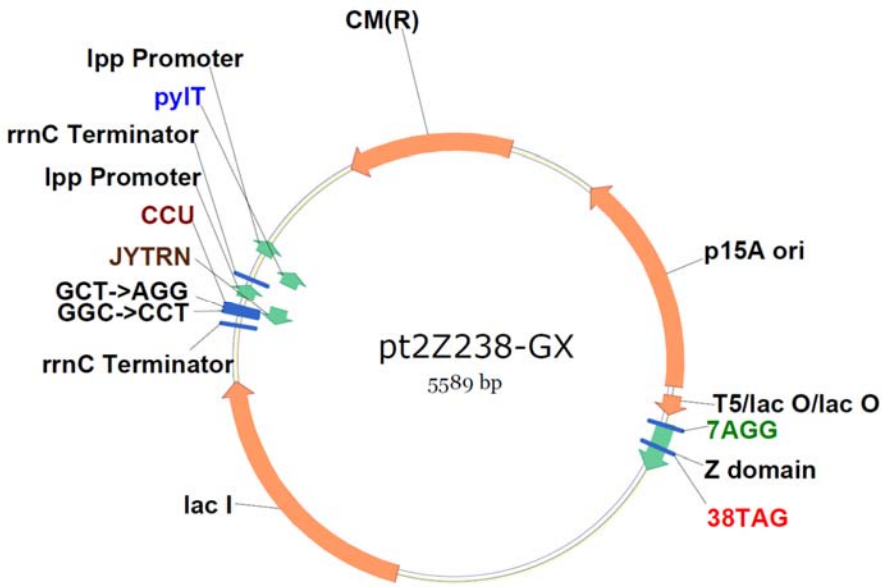


Figure 76. pt2Z238-GX Vector

Vector pAK8 was co-transformed with pt2Z238GX into BL21(DE3) *E. coli* host cell. 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, 34 µg/mL chloramphenicol and two types of unnatural amino acids for further culturing. Until the culture reach OD600=0.6-0.8, 0.02% arabinose and 1mM (working concentration) IPTG were used for induction. After expression for 14-16 h at 37 °C, 20 ml cultures were pelleted. The protein was extracted and purified.

As a control experiment, the mutant Z domain proteins were expressed with following condition: with AlK only; with AcF only, with both AlK and AcF.

The resulting protein solution was treated with SDS-PAGE electrophoresis analysis. It appeared that (Figure 77), without the presence of AlK, no protein band showed at ~8000 Da. In addition, with both UAAs, a strong band was observed at the position we expected; while with presence of only AcF, a faded band was also observed meaning a Z-domain protein with 7th Arg and 38th AlK was produced.

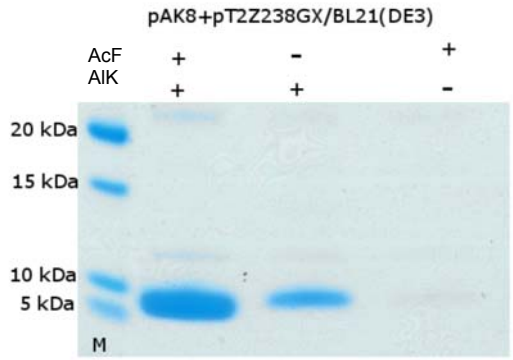


Figure 77 SDS-PAGE gel: the expression of Z-7_{AGG}-38_{UAG} proteins with the presence of two UAAs.

Similar results were also observed from mass spectrometry analysis. The mass for different post-translational product of Z-AcF^{7}-AlK^{38} protein would be 7905.4, 7947.4 and 8037.6. And there were three peaks matched expected molecular weight. (Peak a, b, c, in Figure 78) Notably there was also a minor peak d at m/z =7872, which matching the Z-Arg^{7}-AlK^{38}. The double unnatural amino acid (AcF and AlK) incorporation rate was about 85%, estimated by the intensity of the peak.

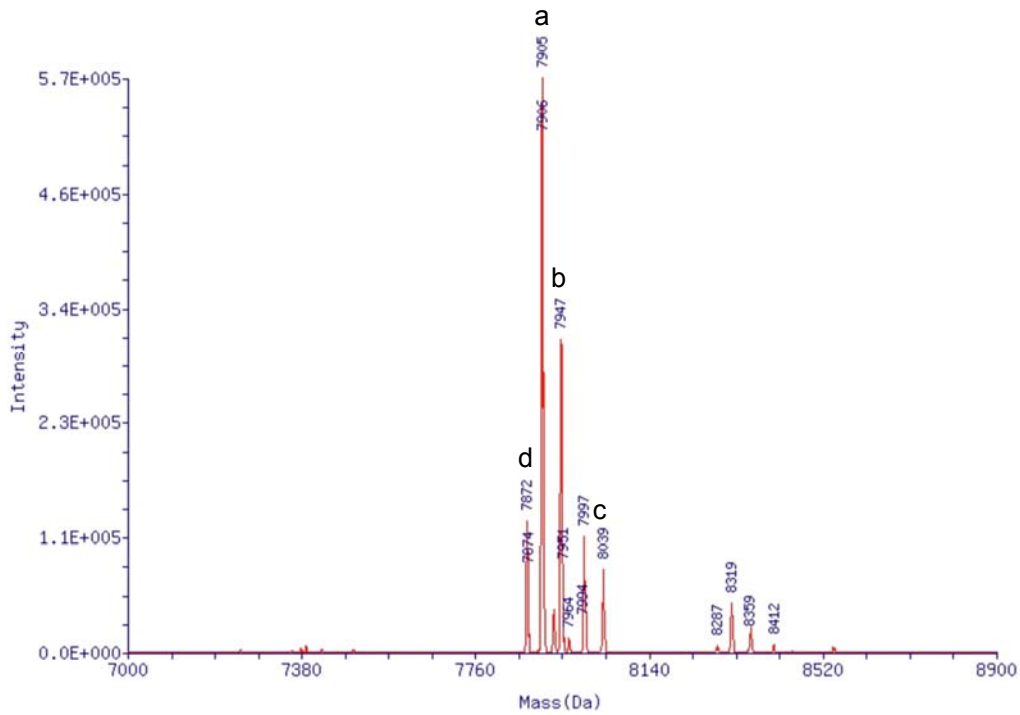


Figure 78. The ESI-MS spectrum for Z-AcF^{7}-AlK^{38} protein.

The functionalities were revealed by chemical label. The acetyl moiety of AcF could be labeled by the hydrazine residue of DNSH (Figure 79), and the alkyne moiety of the AlK could be labeled by the azido-modified PEG. (Figure 80)

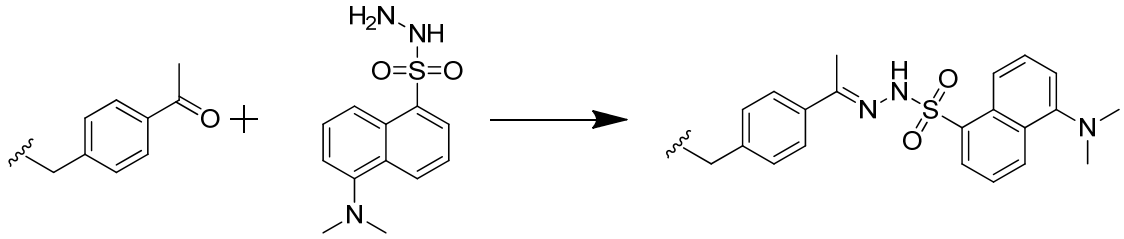


Figure 79. The hydrazone generate from the reaction between DNSH and AcF

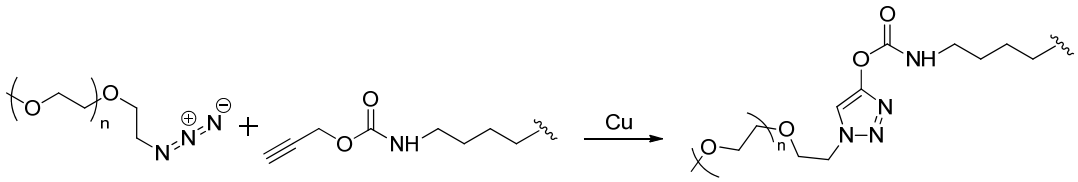


Figure 80 The copper catalyzed azide-alkyne cyloaddition bwtween an azido-modified PEG and an AlK residue.

The Z-AcF^{7}-AlK^{38} protein was treated with DNSH, or azido-modified PEG (5 kDa), or both DNSH and azido-modified PEG (5 kDa). After the reaction, the DNSH labeled protein should displayed a green-yellowish fluorescence upon UV excitation, while the PEG labeled protein should revealed a molecular weight shift.

The following SDS-PAGE electrophoresis analysis (Figure 81) showed that, the Z-AcF^{7}-AlK^{38} protein only with DNSH conjugation displayed as a green-yellowish fluorescent band with UV excitation; while it only with PEGylation reaction appeared as a wide band with significant mass increase. The protein labeled with both DNSH and PEG showed both feature. Single UAA incorporated Z domain protein with AcF and AlK was tested by DNSH and azido-modified PEG as control experiment, showing that Z domain protein with AcF can only be labeled with DNSH, while Z domain protein containing AlK can only be labeled by azido modified PEG.

Thus, with two different system, *Mj*RS$_{AGG}$/*Mj*tRNA$_{CCU}$ and *Mm*Pyl-RS$_{UAG}$/ *Mm*tRNA$_{CUA}$ co-transformed into one host cell, two UAA could be incorporated into single proteins simultaneously, which would significantly broaden the application of the new chemistry introduced by the unnatural amino acid.

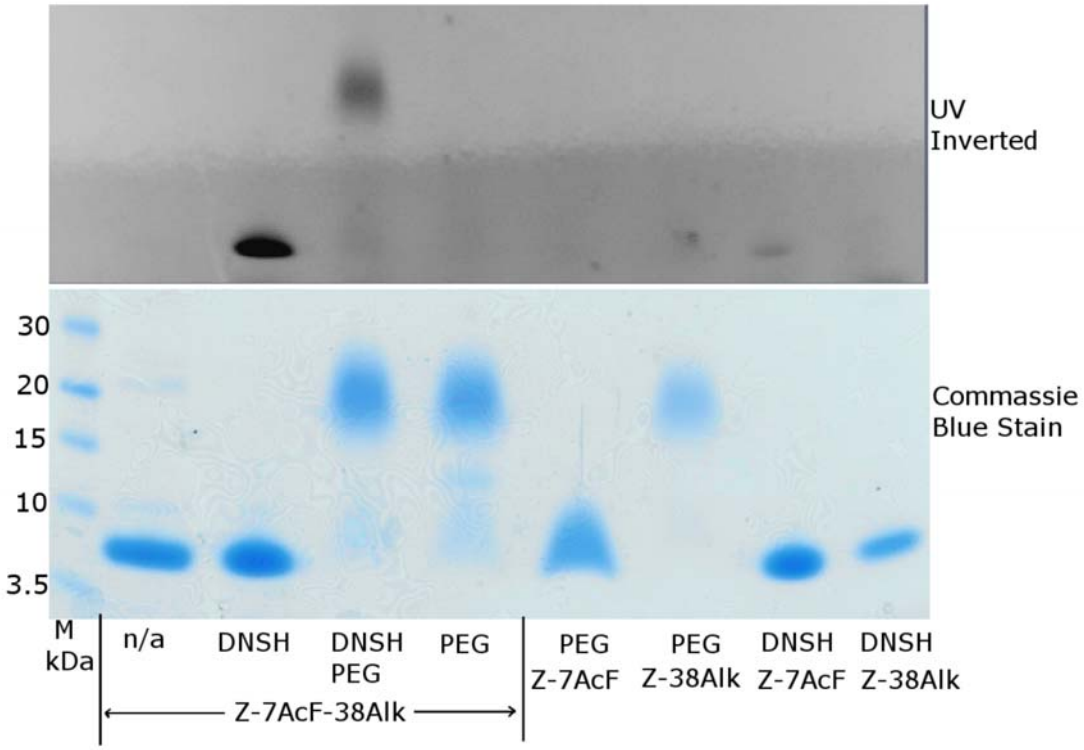


Figure 81. SDS-PAGE gel: Z-AcF^{7}-AlK^{38} protein was labeled with DNSH and PEG

Chapter 6

Conclusions and outlook

6.1 Conclusions

Genetic incorporation of unnatural amino acids into proteins provides a powerful way to add new chemistries to proteins, which would be advantageous in the studies of protein behavior by providing site-specific protein modification that can enhance the protein properties.

This dissertation has revealed a method to genetically incorporate a 2-napthol analogue, 2-amino-3-(6-hydroxy-2-naphthyl) propanoic acid (2NpOH) into proteins, and explored the potential application that utilizing the 2-naphthol chemistry added to the proteins.

To genetically incorporate this unnatural amino acid into proteins, a special aminoacyl-tRNA synthetase was required. Based on the crystal structure of an aaRS that charge 3-(2-naphthyl)alanine, an aaRS DNA library was rationally designed for random mutation. After multiple rounds of positive / negative selection and screening, two mutants of the aaRS were identified. One significant thing I would like to point out is that, rather than using the traditional positive-negative screening system based on antibiotics, we used a screen plasmid containing the GFP gene with a UAA incorporated site in middle, which can be used to rule out the undesired mutant based on the fluorescence. This screening system turned out to be more effective to get rid of the synthetase that charge natural amino acid. To test the ability of the evolved synthetases for selective incorporation of 2NpOH into proteins, an amber stop codon was substituted at a permissive site (Lys7) in the gene for a mutant Z domain protein. The yield of Z-domain protein was roughly 7 mg/L culture with the presence of 0.5 M 2NpOH, but was insignificant without it, indicating a very high fidelity in the incorporation of the unnatural amino acid. Moreover, this was also confirmed by the ESI-MS analysis and the strong unique blue fluorescence under UV excitation. Although sometimes the application of the newly added chemistry is limited due to the relative low yield of the UAA incorporated protein. Our research showed that utilizing a vector containing more than one copy of the orthogonal synthetase gene would significantly increase the production of the unnatural amino acid decorated proteins.

Secondly, a chemoselective azo coupling reaction was developed based on the higher electophilicity of the 2-naphthol residue. To uncover the proper condition for selective azo coupling, a small molecule experiment was carried out to optimize the reaction condition. We used p-cresol and 2-naphthol as the model molecules

to mimic the reactions for tyrosine and NpOH, respectively; and a series of aniline derivatives were converted to their corresponding diazonium salts after treatment with aqueous p-toluenesulfonic acid and sodium nitrite. It revealed that the coupling reaction required very mild condition of pH 7 at 0 °C with diazonium salts derived from aniline bearing electron-donating group. Ultimately, the selective azo coupling was proven to be effective at the protein level as well, with fast reaction rate, high efficiency and excellent selectivity. A plausible application of this azo coupling is demonstrated as site-specific PEGylation with a diazonium reagent bearing a polyethylene glycol moiety. Besides, we had also reported that the chemical labeled tag could be removed upon a reductive cleavage. Thus, it provides us an alternate bio-conjugation method.

Thirdly, to further utilized the orthogonal reactivity of 2NpOH, a Mannich type reaction was tested towards this UAA, which might lead to an *in vivo* bio-conjugation reaction. The preliminary data showed that there was a possibility that lysine residue could selectively react 2NpOH in the presence of formaldehyde *in vivo*. It would be potentially utilized as a reaction to produce cyclic peptides within host cells. However, the reaction is somehow limited by the proximity and the orientation of both residues.

Last, we had showed some effort to incorporate this unnatural amino acid, 2NpOH, for directed protein evolution. We hypothesized that with the substitution of the tyrosine with 2NpOH in the GFP chromophore, we might notice a red-shift fluorescent property, due to an expanded electron conjugate system. However, we only observed the loss of fluorescence rather than a change in fluorescence, the use of flow cytometry might be allow us to identify problems. Besides, in this work, we also demonstrated a method for the incorporation of two UAAs into a single protein, which would be quite helpful for biological studies with two chemical probes within one protein.

6.2 Outlook and future directions

As I mentioned above, our research showed that we have successfully added the naphthol chemistry into proteins by genetic incorporation of 2NpOH, which might lead to many applications.

Azo coupling based on the electrophilicity of the 2naphthol residue with diazotized aniline derivatives was proven to be effective, with fast reaction rate, high efficiency and excellent selectivity. The product protein would be installed with an azobenzene moiety after the treatment with corresponding diazonium salt. One of the most interesting properties of azobenzene is the photoisomerization of trans and cis isomers. Rather than chemically linking azobenzene tag to surface accessible cysteine or lysine residue, which normally need to substitute all other reactive cysteine or lysine with unreactive amino acid to reduce the false modification; incorporating the UAA 2NpOH into the active site of enzyme,

following with the azo coupling reaction, the enzyme activity would be altered by the external light stimulus. Thus, the enzyme-catalyzed reaction could be regulated. Utilizing the photocontrol of enzyme activity would be promising for the development of biotechnological devices, such as light-triggered sequential bioprocessing systems, as well as diagnostic bio-chip which can be temporally controlled with high spatial resolution.[65]

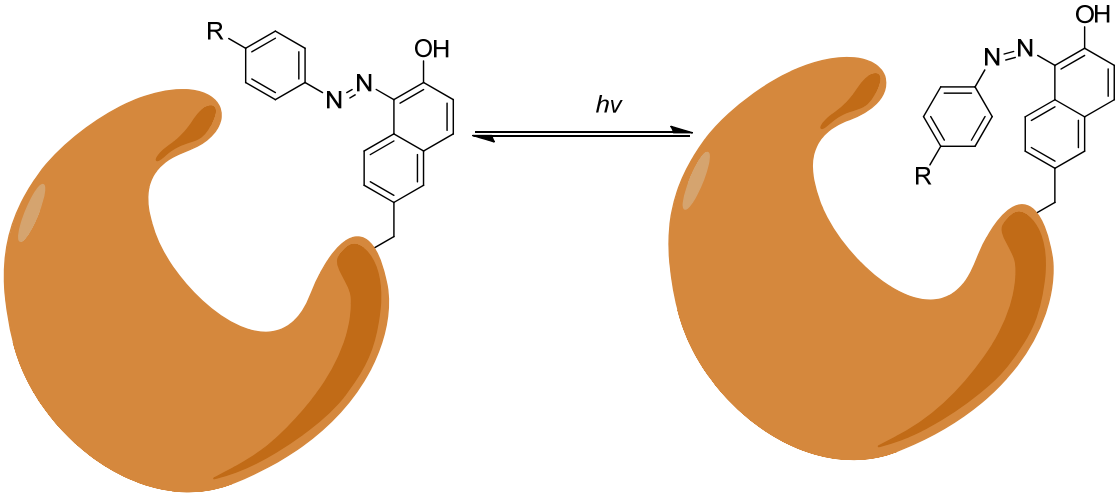


Figure 82. Reversible photocontrol of an enzyme activity.

We have demonstrated the utility of the reaction via site-specific PEGlyation of proteins with 2naphthol. This indicates that it could be utilized to modify proteins with not only small molecules but also macromolecules. We anticipate that the reaction we developed to be a powerful alternative bio-conjugation methods in drug delivery as simple as modified the protein-based drug with PEG[66] or other polymers. Phage nanoparticles could be utilized for target therapy. This application would benefit from 2NpOH incorporation, since it can serve as a highly reactive site for linking the drug, or the host-specificity-conferring ligand.[67]

As promising proteomic and diagnostic tools, protein biochips have become a booming field of research in the past few years. In order to achieve the success of microarray technology, protein immobilization is one key factor to be considered. To fully realize the potential of protein biochips, proteins need to be immobilized onto surfaces with high density; meanwhile the conformation and activity of the protein need to be maintained.[47] Covalent immobilization is often used to decorate the surface through bio-conjugation reaction with accessible functional groups of exposed amino acids. When pinning down to the surface, the protein should be oriented in a way that the binding sites are exposed to the sample solution. In most existing immobilization methods, proteins are immobilized in random orientations because on the protein surface there tend to be more than one residual that can form covalent bonds with the surface. With the site specific incorporation of 2NpOH and surface functionalization of the diazotized function groups, immobilization of

proteins with predetermined orientations will be much easier to achieve. Additionally, the covalent linkage could be cleaved by a reduction reaction, so that the binding product can be harvested.

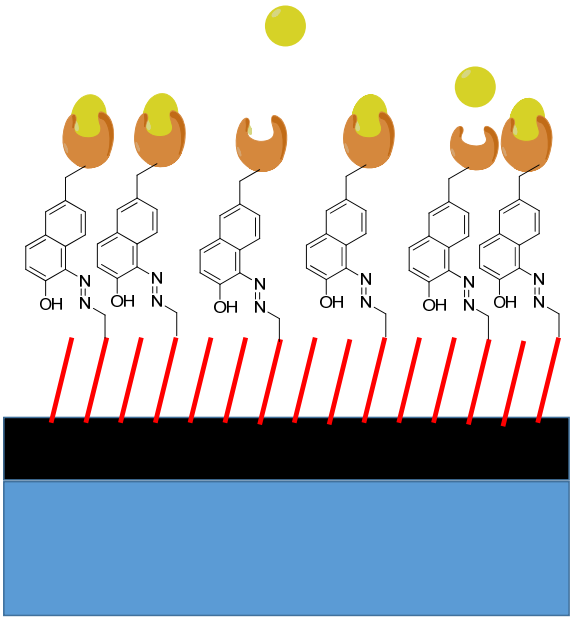


Figure 83. Scheme of a protein array

Similarly, rather than the commonly used protein purification technique involved with 6xHis tagged proteins and Ni-NTA agarose[48], a new protein purification strategy can be developed based on the same reaction. The 6xHis tag normally needs to be added to the C or N terminus of the protein sequence to increase the binding affinity. However, with the azo-coupling reaction and following reductive cleavage, 2NpOH is not necessarily a terminal residue, Therefore the terminus can be left intact, which is needed in some experiments.

The advantage of formaldehyde-based cross coupling reaction between lysine and 2NpOH, is that not only this reaction is highly selective towards the UAA residue, but also it can be carried out *in vivo.* With the recent research interest of circular proteins, people show some efforts to synthesize circular peptides through intramolecular native chemical ligation, or expressed protein ligation.[68] This reaction can provide an alternate method to produce circular proteins without extra solid-phase synthesis or peptide fusion reactions

The incorporation of multiple unnatural amino acids makes it possible to add more chemistries into protein. Thus, it provides us a powerful tool for us to alter, control and observe the behavior and properties of the protein.

The two different functionalities can be used for generating protein oligomers; producing phage-displayed libraries with increase diversity and high affinity;

developing protein based nanomaterial etc.

Additionally, it may help to install a Förster Resonance Energy Transfer (FRET) pair. Traditionally, people have showed that the FRET donor and acceptor could be mounted to the protein surface either by chemical reaction towards the lysine or cysteine residue or by the special binding motif[69]. However, sometimes this method is limited by the low selectivity and need to redesign the protein with special motif. Genetically introducing two UAAs into a single protein makes the production of protein with a FRET donor and an acceptor at specific sites possible. Alternatively, it would allow adding two reactive centers for selective bio-conjugation to link the fluorophores onto proteins. Thus, it would provide us a powerful tool to utilize FRET for biological study of the molecular interaction with highly precise control.

The incorporation of two UAAs into a single protein, may also be an enabling tool in single molecule force spectroscopy, which is a new and powerful technique to study protein folding. DNA handles are used to specifically attach the protein to polystyrene beads. The protein can be mechanically manipulated by those handles with optical tweezers. Thus, it would allow us to monitor refolding events and fluctuations of the molecular structures in a quasi-equilibrium condition directly.[70] Commonly, the handles could be linked to any exposed pair of cysteine residues. However, with two unique UAAs on surface, the handles can be selectively bonded to the specific sites of the protein, the refolding process of the protein or domain can be unveiled in a more precise manner.

Bibliography

1. Petsko, G. A.; Ringe, D., *Protein structure and function*. New Science Press ;
Sinauer Associates ;
Blackwell Pub.: London
Sunderland, MA
Oxford, 2004; p xxii, 195 p.
2. Leinfelder, W.; Zehelein, E.; Mandrandberthelot, M. A.; Bock, A., Gene for a Novel Transfer-Rna Species That Accepts L-Serine and Cotranslationally Inserts Selenocysteine. *Nature* 1988, *331* (6158), 723-725.
3. Srinivasan, G.; James, C. M.; Krzycki, J. A., Pyrrolysine encoded by UAG in Archaea: Charging of a UAG-decoding specialized tRNA. *Science* 2002, *296* (5572), 1459-1462.
4. Liu, C. C.; Schultz, P. G., Adding New Chemistries to the Genetic Code. *Annu Rev Biochem* 2010, *79*, 413-444.
5. Stephanopoulos, N.; Francis, M. B., Choosing an effective protein bioconjugation strategy. *Nat Chem Biol* 2011, *7* (12), 876-884.
6. Brannigan, J. A.; Wilkinson, A. J., Protein engineering 20 years on. *Nature Reviews Molecular Cell Biology* 2002, *3* (12), 964-970.
7. Xie, J. M.; Schultz, P. G., An expanding genetic code. *Methods* 2005, *36* (3), 227-238.
8. Ambrogelly, A.; Palioura, S.; Soll, D., Natural expansion of the genetic code. *Nat Chem Biol* 2007, *3* (1), 29-35.
9. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z. W.; Schultz, P. G., An expanded eukaryotic genetic code. *Science* 2003, *301* (5635), 964-967.
10. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* 2001, *292* (5516), 498-500.
11. Kim, C. H.; Axup, J. Y.; Schultz, P. G., Protein conjugation with genetically encoded unnatural amino acids. *Current opinion in chemical biology* 2013, *17* (3), 412-9.
12. (a) Presolski, S. I.; Hong, V. P.; Finn, M. G., Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Current protocols in chemical biology* 2011, *3* (4), 153-162; (b) Kolb, H. C.; Finn, M. G.; Sharpless, K. B., Click chemistry: Diverse chemical function from a few good reactions. *Angew Chem Int Edit* 2001, *40* (11), 2004-+.
13. Young, T. S.; Schultz, P. G., Beyond the Canonical 20 Amino Acids: Expanding the Genetic Lexicon. *Journal of Biological Chemistry* 2010, *285*

(15), 11039-11044.
14. Mu, J.; Pinkstaff, J.; Li, Z.; Skidmore, L.; Li, N.; Myler, H.; Dallas-Yang, Q.; Putnam, A. M.; Yao, J.; Bussell, S.; Wu, M.; Norman, T. C.; Rodriguez, C. G.; Kimmel, B.; Metzger, J. M.; Manibusan, A.; Lee, D.; Zaller, D. M.; Zhang, B. B.; DiMarchi, R. D.; Berger, J. P.; Axelrod, D. W., FGF21 analogs of sustained action enabled by orthogonal biosynthesis demonstrate enhanced antidiabetic pharmacology in rodents. *Diabetes* 2012, *61* (2), 505-12.
15. Cho, H.; Daniel, T.; Buechler, Y. J.; Litzinger, D. C.; Maio, Z.; Putnam, A. M.; Kraynov, V. S.; Sim, B. C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; Becky, G. C.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E., Optimized clinical performance of growth hormone with an expanded genetic code. *Proc Natl Acad Sci U S A* 2011, *108* (22), 9060-5.
16. Carrico, I. S., Chemoselective modification of proteins: hitting the target. *Chem Soc Rev* 2008, *37* (7), 1423-1431.
17. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angew Chem Int Edit* 2005, *44* (1), 34-66.
18. Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S., Adding l-lysine derivatives to the genetic code of mammalian cells with engineered pyrrolysyl-tRNA synthetases. *Biochemical and biophysical research communications* 2008, *371* (4), 818-22.
19. Kavran, J. M.; Gundliapalli, S.; O'Donoghue, P.; Englert, M.; Soell, D.; Steitz, T. A., Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *P Natl Acad Sci USA* 2007, *104* (27), 11268-11273.
20. Polycarpo, C. R.; Herring, S.; Berube, A.; Wood, J. L.; Soll, D.; Ambrogelly, A., Pyrrolysine analogues as substrates for pyrrolysyl-tRNA synthetase. *Febs Lett* 2006, *580* (28-29), 6695-6700.
21. Wang, L.; Schultz, P. G., A general approach for the generation of orthogonal tRNAs. *Chem Biol* 2001, *8* (9), 883-890.
22. Steer, B. A.; Schimmel, P., Major anticodon-binding region missing from an archaebacterial tRNA synthetase. *Journal of Biological Chemistry* 1999, *274* (50), 35601-35606.
23. Kobayashi, T.; Nureki, O.; Ishitani, R.; Yaremchuk, A.; Tukalo, M.; Cusack, S.; Sakamoto, K.; Yokoyama, S., Structural basis for orthogonal tRNA specificities of tyrosyl-tRNA synthetases for genetic code expansion. *Nat Struct Biol* 2003, *10* (6), 425-432.
24. Anderson, J. C.; Wu, N.; Santoro, S. W.; Lakshman, V.; King, D. S.; Schultz, P. G., An expanded genetic code with a functional quadruplet

codon. *P Natl Acad Sci USA* 2004, *101* (20), 7566-7571.
25. Takaoka, Y.; Ojida, A.; Hamachi, I., Protein organic chemistry and applications for labeling and engineering in live-cell systems. *Angewandte Chemie* 2013, *52* (15), 4088-106.
26. (a) Saxon, E.; Bertozzi, C. R., Cell surface engineering by a modified Staudinger reaction. *Science* 2000, *287* (5460), 2007-2010; (b) Kiick, K. L.; Saxon, E.; Tirrell, D. A.; Bertozzi, C. R., Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation. *P Natl Acad Sci USA* 2002, *99* (1), 19-24; (c) Tsao, M. L.; Tian, F.; Schultz, P. G., Selective staudinger modification of proteins containing p-azidophenylalanine. *Chembiochem* 2005, *6* (12), 2147-2149; (d) Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A., Live-Cell Imaging of Cellular Proteins by a Strain-Promoted Azide-Alkyne Cycloaddition. *Chembiochem* 2010, *11* (15), 2092-2095; (e) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R., A strain-promoted [3+2] azide-alkyne cycloaddition for covalent modification of blomolecules in living systems. *J Am Chem Soc* 2004, *126* (46), 15046-15047.
27. (a) Hudak, J. E.; Yu, H. H.; Bertozzi, C. R., Protein Glycoengineering Enabled by the Versatile Synthesis of Aminooxy Glycans and the Genetically Encoded Aldehyde Tag. *J Am Chem Soc* 2011, *133* (40), 16127-16135; (b) Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Wang, Z. Y.; Russell, D. H.; Liu, W. S., Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications. *Bioorg Med Chem Lett* 2010, *20* (3), 878-880; (c) Zhang, Z. W.; Smith, B. A. C.; Wang, L.; Brock, A.; Cho, C.; Schultz, P. G., A new strategy for the site-specific modification of proteins in vivo. *Biochemistry-Us* 2003, *42* (22), 6735-6746.
28. Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. W.; Chin, J. W., Genetically Encoded 1,2-Aminothiols Facilitate Rapid and Site-Specific Protein Labeling via a Bio-orthogonal Cyanobenzothiazole Condensation. *J Am Chem Soc* 2011, *133* (30), 11418-11421.
29. Song, W.; Wang, Y.; Qu, J.; Madden, M. M.; Lin, Q., A photoinducible 1,3-dipolar cycloaddition reaction for rapid, selective modification of tetrazole-containing proteins. *Angew Chem Int Edit* 2008, *47* (15), 2832-2835.
30. (a) Kodama, K.; Fukuzawa, S.; Nakayama, H.; Sakamoto, K.; Kigawa, T.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Yokoyama, S.; Tachibana, K., Site-specific functionalization of proteins by organopalladium reactions. *Chembiochem* 2007, *8* (2), 232-238; (b) Brustad, E.; Bushey, M. L.; Lee, J. W.; Groff, D.; Liu, W.; Schultz, P. G., A Genetically Encoded

Boronate-Containing Amino Acid. *Angew Chem Int Edit* 2008, *47* (43), 8220-8223; (c) Spicer, C. D.; Davis, B. G., Palladium-mediated site-selective Suzuki-Miyaura protein modification at genetically encoded aryl halides. *Chem Commun* 2011, *47* (6), 1698-1700; (d) Li, N.; Lim, R. K. V.; Edwardraja, S.; Lin, Q., Copper-Free Sonogashira Cross-Coupling for Functionalization of Alkyne-Encoded Proteins in Aqueous Medium and in Bacterial Cells. *J Am Chem Soc* 2011, *133* (39), 15316-15319.
31. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C. J.; Deiters, A.; Chin, J. W., Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nat Chem* 2012, *4* (4), 298-304.
32. (a) Lim, R. K. V.; Lin, Q., Bioorthogonal chemistry: recent progress and future directions. *Chem Commun* 2010, *46* (10), 1589-1600; (b) Sletten, E. M.; Bertozzi, C. R., From Mechanism to Mouse: A Tale of Two Bioorthogonal Reactions. *Accounts Chem Res* 2011, *44* (9), 666-676.
33. (a) Vivian, J. T.; Callis, P. R., Mechanisms of tryptophan fluorescence shifts in proteins. *Biophysical Journal* 2001, *80* (5), 2093-2109; (b) Beechem, J. M.; Brand, L., Time-Resolved Fluorescence of Proteins. *Annual Review of Biochemistry* 1985, *54*, 43-71.
34. Matsunaga, N.; Kaku, T.; Ojida, A.; Tanaka, T.; Hara, T.; Yamaoka, M.; Kusaka, M.; Tasaka, A., C-17,C-20-lyase inhibitors. Part 2: Design, synthesis and structure-activity relationships of (2-naphthylmethyl)-1H-imidazoles as novel C-17,C-20-lyase inhibitors. *Bioorganic & Medicinal Chemistry* 2004, *12* (16), 4313-4336.
35. Samet, A. V.; Coughlin, D. J.; Buchanan, A. C.; Gakh, A. A., An improved "one-pot" procedure for synthesis of fluorinated DL-phenylalanines. *Synthetic Communications* 2002, *32* (6), 941-946.
36. Vela, M. A.; Fronczek, F. R.; Horn, G. W.; Mclaughlin, M. L., Syntheses of 1-Naphthol and 2-Naphthol Analogs of Dl-Tyrosine - Potential Fluorescent-Probes of Peptide Structure and Dynamics in Complex Environments. *J Org Chem* 1990, *55* (9), 2913-2918.
37. Xie, J. M.; Schultz, P. G., Adding amino acids to the genetic repertoire. *Current opinion in chemical biology* 2005, *9* (6), 548-554.
38. Turner, J. M.; Graziano, J.; Spraggon, G.; Schultz, P. G., Structural plasticity of an aminoacyl-tRNA synthetase active site. *P Natl Acad Sci USA* 2006, *103* (17), 6483-6488.
39. Wang, L.; Brock, A.; Schultz, P. G., Adding L-3-(2-naphthyl)alanine to the genetic code of E-coli. *J Am Chem Soc* 2002, *124* (9), 1836-1837.
40. Ryu, Y. H.; Schultz, P. G., Efficient incorporation of unnatural amino acids into proteins in Escherichia coli. *Nat Methods* 2006, *3* (4), 263-265.

41. Melancon, C. E.; Schultz, P. G., One plasmid selection system for the rapid evolution of aminoacyl-tRNA synthetases. *Bioorg Med Chem Lett* 2009, *19* (14), 3845-3847.
42. Wang, L.; Zhang, Z. W.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of Escherichia coli. *P Natl Acad Sci USA* 2003, *100* (1), 56-61.
43. Zheng, D. Y.; Aramini, J. M.; Montelione, G. T., Validation of helical tilt angles in the solution NMR structure of the Z domain of Staphylococcal protein A by combined analysis of residual dipolar coupling and NOE data. *Protein Sci* 2004, *13* (2), 549-554.
44. Schlick, T. L.; Ding, Z. B.; Kovacs, E. W.; Francis, M. B., Dual-surface modification of the tobacco mosaic virus. *J Am Chem Soc* 2005, *127* (11), 3718-3723.
45. Veronese, F. M.; Harris, J. M., Preface - Introduction and overview of peptide and protein pegylation. *Adv Drug Deliver Rev* 2002, *54* (4), 453-456.
46. Dinica, R.; Charmantray, F.; Demeunynck, M.; Dumy, P., Regioselective synthesis of angular nitrogen polyheterocycles: dipyrido[3,2-a : 2 ',3 '-c]quinolino[2,3-h]phenazines. *Tetrahedron Lett* 2002, *43* (44), 7883-7885.
47. Rusmini, F.; Zhong, Z. Y.; Feijen, J., Protein immobilization strategies for protein biochips. *Biomacromolecules* 2007, *8* (6), 1775-1789.
48. Roggero, M. A.; Servis, C.; Corradin, G., A simple and rapid procedure for the purification of synthetic polypeptides by a combination of affinity chromatography and methionine chemistry. *Febs Lett* 1997, *408* (3), 285-288.
49. Yang, F.; Moss, L. G.; Phillips, G. N., The molecular structure of green fluorescent protein. *Nat Biotechnol* 1996, *14* (10), 1246-1251.
50. Joshi, N. S.; Whitaker, L. R.; Francis, M. B., A three-component Mannich-type reaction for selective tyrosine bioconjugation. *J Am Chem Soc* 2004, *126* (49), 15942-15943.
51. Fraenkel-Conrat, H.; Olcott, H. S., The reaction of formaldehyde with proteins; cross-linking between amino and primary amide or guanidyl groups. *J Am Chem Soc* 1948, *70* (8), 2673-84.
52. (a) Mathew, B. P.; Nath, M., One-Pot Three-Component Synthesis of Dihydrobenzo- and Naphtho[e]-1,3-oxazines in Water. *J Heterocyclic Chem* 2009, *46* (5), 1003-1006; (b) Shen, A. Y.; Tsai, C. T.; Chen, C. L., Synthesis and cardiovascular evaluation of N-substituted 1-aminomethyl-2-naphthols. *Eur J Med Chem* 1999, *34* (10), 877-882.
53. Landschulz, W. H.; Johnson, P. F.; McKnight, S. L., The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 1988, *240* (4860), 1759-64.

54. Oshea, E. K.; Klemm, J. D.; Kim, P. S.; Alber, T., X-Ray Structure of the Gcn4 Leucine Zipper, a 2-Stranded, Parallel Coiled Coil. *Science* 1991, *254* (5031), 539-544.
55. Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C., The Gcn4 Basic Region Leucine Zipper Binds DNA as a Dimer of Uninterrupted Alpha-Helices - Crystal-Structure of the Protein-DNA Complex. *Cell* 1992, *71* (7), 1223-1237.
56. Rosenow, M. A.; Patel, H. N.; Wachter, R. M., Oxidative chemistry in the GFP active site leads to covalent cross-linking of a modified leucine side chain with a histidine imidazole: Implications for the mechanism of chromophore formation. *Biochemistry-Us* 2005, *44* (23), 8303-8311.
57. Zimmer, M., Green fluorescent protein (GFP): applications, structure, and related photophysical behavior. *Chemical reviews* 2002, *102* (3), 759-81.
58. Tsien, R. Y., The green fluorescent protein. *Annual Review of Biochemistry* 1998, *67*, 509-544.
59. Reid, B. G.; Flynn, G. C., Chromophore formation in green fluorescent protein. *Biochemistry-Us* 1997, *36* (22), 6786-6791.
60. Lawrence, M. S.; Phillips, K. J.; Liu, D. R., Supercharging proteins can impart unusual resilience. *J Am Chem Soc* 2007, *129* (33), 10110-+.
61. Crameri, A.; Whitehorn, E. A.; Tate, E.; Stemmer, W. P. C., Improved green fluorescent protein by molecular evolution using DNA shuffling. *Nat Biotechnol* 1996, *14* (3), 315-319.
62. Xia, Y.; DiPrimio, N.; Keppel, T. R.; Vo, B.; Fraser, K.; Battaiei, K. P.; Egan, C.; Bystroff, C.; Lovell, S.; Weis, D. D.; Anderson, J. C.; Karanicolas, J., The Designability of Protein Switches by Chemical Rescue of Structure: Mechanisms of Inactivation and Reactivation. *J Am Chem Soc* 2013, *135* (50), 18840-18849.
63. Wan, W.; Huang, Y.; Wang, Z. Y.; Russell, W. K.; Pai, P. J.; Russell, D. H.; Liu, W. R., A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli. *Angew Chem Int Edit* 2010, *49* (18), 3211-3214.
64. Li, W. T.; Mahapatra, A.; Longstaff, D. G.; Bechtel, J.; Zhao, G.; Kang, P. T.; Chan, M. K.; Krzycki, J. A., Specificity of Pyrrolysyl-tRNA Synthetase for Pyrrolysine and Pyrrolysine Analogs. *J Mol Biol* 2009, *385* (4), 1156-1164.
65. Korbus, M.; Balasubramanian, G.; Muller-Plathe, F.; Kolmar, H.; Meyer-Almes, F. J., Azobenzene switch with a long-lived cis-state to photocontrol the enzyme activity of a histone deacetylase-like amidohydrolase. *Biol Chem* 2014, *395* (4), 401-412.
66. Veronese, F. M.; Pasut, G., PEGylation, successful approach to drug

delivery. *Drug Discov Today* 2005, *10* (21), 1451-1458.
67. Bar, H.; Yacoby, I.; Benhar, I., Killing cancer cells by targeted drug-carrying phage nanomedicines. *Bmc Biotechnol* 2008, *8*.
68. Sancheti, H.; Camarero, J. A., "Splicing up" drug discovery. Cell-based expression and screening of genetically-encoded libraries of backbone-cyclized polypeptides. *Adv Drug Deliver Rev* 2009, *61* (11), 908-917.
69. Hoffmann, C.; Gaietta, G.; Bunemann, M.; Adams, S. R.; Oberdorff-Maass, S.; Behr, B.; Vilardaga, J. P.; Tsien, R. Y.; Eisman, M. H.; Lohse, M. J., A FlAsH-based FRET approach to determine G protein - coupled receptor activation in living cells. *Nat Methods* 2005, *2* (3), 171-176.
70. Cecconi, C.; Shank, E. A.; Dahlquist, F. W.; Marqusee, S.; Bustamante, C., Protein-DNA chimeras for single molecule mechanical folding studies with the optical tweezers. *Eur Biophys J Biophy* 2008, *37* (6), 729-738.