UNIVERSITY OF CALIFORNIA, SAN DIEGO

Unnatural Amino Acid Incorporation for Genetic Code Expansion in Mammalian Cells

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

in

Biology

by

Jeffrey Kunio Takimoto

Committee in charge:

Professor Lei Wang, Chair
Professor Adah Almutairi
Professor Richard Kolodner
Professor Cornelis Murre
Professor Joseph Noel
Professor Kit Pogliano

2011

The Dissertation of Jeffrey Kunio Takimoto is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2011

DEDICATION

I would like to dedicate my PhD thesis to my mother Chikako Takimoto and Melissa Kanchanapoomi. Without them, all this would never have been possible.

EPIGRAPH

"Be great in act, as you have been in thought"

William Shakespeare

TABLE OF CONTENTS

Signature Page iii

Dedication iv

Epigraph v

Table of Contents vi

List of Figures xi

List of Scheme xiii

List of Tables xiv

Acknowledgements xv

Vita xviii

Abstract of the Dissertation xix

Chapter 1 Introduction: Unnatural Amino Acids to Investigate Biological Processes 1

1.1 Abstract 2

1.2 Introduction 2

1.3 Unnatural amino acids 3

1.4 Methodology 4

1.4.1 Chemical approaches 4

1.4.2 Biosynthetic approaches 5

1.5 Application of unnatural amino acids 7

1.5.1 Protein stability 7

1.5.2 Protein structure and function 9

1.5.3 Modification and regulation of protein activity 13

1.6 Future Directions 16

1.7 ACKNOWLEDGEMENTS 17

1.8 FIGURES 18

1.9 REFERENCES 22

Chapter 2 Genetically Encoding Unnatural Amino Acids For Cellular and Neuronal Studies 26

2.1 Abstract 27

2.2 Introduction 27

2.3 Results 29

2.3.1 Efficient expression of orthogonal tRNA in mammalian cells 29

2.3.2 Unnatural-amino-acid-specific aaRS for mammalian cells 33

2.3.3 Genetically encoding unnatural amino acids in neurons 34

2.3.4 Probing the inactivation mechanism of K^+ channel Kv1.4 35

2.4 Discussion 38

2.4.1 Unnatural amino-acid mutagenesis in mammalian cells 38

2.4.2 Investigating the fast inactivation of Kv1.4 channels 41

2.5 Methods 42

2.5.1 Chemicals 42

2.5.2 Constructs 43

2.5.3 Cell culture and transfection 44

2.5.4 Northern blot and western blot analysis 45

2.5.5 Flow cytometry 46

2.5.6 Electrophysiology 46

2.5.7 Fluorescence microscopy 47

2.6 Author contributions 47

2.7 ACKNOWLEDGEMENTS 48

2.8 FIGURES 49

2.9 TABLES 55

2.10 REFERENCES 56

Chapter 3 Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells 59

3.1 Abstract 60
3.2 Introduction 60
3.3 Results and Discussion 61
3.4 Methods 65
3.4.1 Material 65
3.4.2 Construction of plasmids 65
3.4.3 Cell culture and transfections 67
3.4.4 Flow cytometry 67
3.4.5 Photocrosslinking and Western blot analysis 68
3.5 ACKNOWLEDGEMENTS 69
3.6 FIGURES 70
3.7 REFERENCES 73
Chapter 4 Esterification of An Unnatural Amino Acid Structurally Deviating From Canonical Amino Acids Promotes Its Uptake and Incorporation Into Proteins in Mammalian Cells 75
4.1 Abstract 76
4.2 Results and discussion 76
4.3 Experimental section 80
4.3.1 Synthesis of DanAla and DanAla esters 80
4.3.2 Cell lines and culture conditions 83
4.3.3 Fluorescence imaging 83
4.3.4 HPLC analysis of intracellular concentration 84
4.3.5 Flow cytometry 84
4.3.6 Western blot 85
4.4 ACKNOWLEDGEMENTS 86
4.5 Supplemental information 87
4.5.1 Experimental Data for Compounds 87
4.6 SCHEME 90

4.7 FIGURES 91

4.8 REFERENCES 111

Chapter 5 Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient *in Vivo* Incorporation of Structurally Divergent Non-native Amino Acids 113

5.1 Abstract 114

5.2 Introduction 114

5.3 Results and discussion 118

5.3.1 Laboratory Evolution of the *Methanosarcina mazei* $tRNA_{CUA}^{Pyl}$/PylRS pair 118

5.3.2 MmOmeRS Specifically Incorporates the Selected Uaa in *E. coli* 119

5.3.3 Use of MmOmeRS for Uaa Incorporation in Mammalian Cells 120

5.3.4 Three-Dimensional Architecture of MmOmeRS 122

5.3.5 Active Site Substitutions Lead to an Unexpected Recognition of the Ome Side Chain Moiety 123

5.3.6 Recognition of the Ome Main Chain and Coordination of the Adenyl-Donor AMP-PNP-Mg^{2+} 124

5.4 Conclusions 126

5.5 Methods 129

5.5.1 Plasmid Construction 129

5.5.2 Library Construction and *in Vivo* Selection 129

5.5.3 Protein Expression and Uaa Incorporation in *E. coli* 129

5.5.4 Protein Expression and Uaa Incorporation in Mammalian Cells 130

5.5.5 Mass Spectrometry 131

5.5.6 Cloning, Expression, and Purification of MmOmeRS 131

5.5.7 Crystallization, Structure Solution, and Refinement 132

5.6 Accession Codes 133

5.7 ACKNOWLEDGEMENTS 133

5.8 Supplemental information 134

5.8.1 DNA sequences 134

5.9 FIGURES 139

5.10 TABLES 145

5.11 SUPPLEMENTARY FIGURES 146

5.10 REFERENCES 148

LIST OF FIGURES

Figure 1.1: Different forms of unnatural amino acids 18

Figure 1.2: Structures of unnatural amino acids discussed in the text 19

Figure 1.3: Backbond mutations generated by α-hydroxy acids 19

Figure 1.4: Substitution of Lys4 with 2-hydroxylethyl-cysteine in the λ-repressor changes the hydrogen-bonding pattern and DNA substrate specificity from C:G to T:A 20

Figure 1.5: Photolysis of 2-nitrobenzyl caged serine 20

Figure 1.6: Azobenzene group isomerization 21

Figure 2.1: Efficient expression of bacterial tRNA in mammalian cells using the H1 promoter 49

Figure 2.2: Unnatural-amino-acid-specific synthetases evolved in yeast are functional in mammalian cells 50

Figure 2.3: Genetically encoding unnatural amino acids in neurons 51

Figure 2.4: A model for the N-type inactivation of Kv channels 52

Figure 2.5: Probing N-type inactivation using unnatural amino acid mutagenesis with Kv1.4 channels expressed in mammalian cells 53

Figure 2.6: Diameter of the inactivation peptide affects channel inactivation owing to the size restriction of the side portal 54

Figure 3.1: Superposition of *E. coli* TyrRS on *T. thermophilus* $tRNA^{Tyr}$–TyrRS complex 70

Figure 3.2: Enhanced mutant synthetases increased the incorporation efficiency of various unnatural amino acids in mammalian cells 71

Figure 3.3: Photocrosslinking in mammalian cells using Azi 72

Figure 4.1: Analysis of the incorporation efficiency of DanAla into GFP 91

Figure 4.2: Intracellular DanAla and cell health 92

Figure 4.3: ^{1}H NMR (500 MHz, CDCl3) of Compound 6 93

Figure 4.4: ^{13}C NMR (125 MHz, CDCl3) of Compound 6 94

Figure 4.5: ^{1}H NMR (500 MHz, CDCl3) of Compound 7 95

Figure 4.6: ^{13}C NMR (125 MHz, CDCl3) of Compound 7 96

Figure 4.7: ^{1}H NMR (500 MHz, d6-DMSO) of Compound 2 97

Figure 4.8: ^{13}C NMR (125 MHz, d6-DMSO) of Compound 2 98

Figure 4.9: ^{1}H NMR (500 MHz, d6-DMSO) of Compound 3 99

Figure 4.10: ^{13}C NMR (125 MHz, d6-DMSO) of Compound 3 100

Figure 4.11: ^{1}H NMR (500 MHz, CDCl3) of Boc-DanAla-OBn 101

Figure 4.12: ^{13}C NMR (125 MHz, CDCl3) of Boc-DanAla-OBn 102

Figure 4.13: ^{1}H NMR (500 MHz, CDCl3) of Compound 9 103

Figure 4.14: ^{13}C NMR (125 MHz, $CDCl_3$) of Compound 9 104

Figure 4.15: ^{1}H NMR (500 MHz, $CDCl_3$) of Compound 10 105

Figure 4.16: ^{13}C NMR (125 MHz, $CDCl_3$) of Compound 10 106

Figure 4.17: ^{1}H NMR (500 MHz, $CDCl_3$) of Compound 11 107

Figure 4.18: ^{13}C NMR (125 MHz, $CDCl_3$) of Compound 11 108

Figure 4.19: ^{1}H NMR (500 MHz, d6-DMSO) of Compound 4 109

Figure 4.20: ^{13}C NMR (125 MHz, d6-DMSO) of Compound 4 110

Figure 5.1: Evolving *M. mazei* PylRS to aminoacylate Uaas bearing aromatic side chains 139

Figure 5.2: MmOmeRS specifically incorporates the cognate Uaa Ome into proteins in *E. coli* 140

Figure 5.3: MmOmeRS incorporates the cognate Uaa Ome in mammalian cells 141

Figure 5.4: Crystal structure of MmOmeRS reveals the stereochemical basis for Ome selectivity 142

Figure 5.5: Structural depiction of interactions between the amino acid substrate and the surrounding active site residues within MmOmeRS and PlyRS 143

Figure 5.6: Differences in coordination geometry between an ATP analogue and various amino acid substrates in PylRS and MmOmeRS 144

Supplementary Figure 5.1: Tandem mass spectrum of peptide +H_3N-HNIEDGSVQLADH**X**QQNTPI GDGPVLLPDNHYLSTQSALSK-CO_2H 146

Supplementary Figure 5.2: Superposition of *M. mazei* OmeRS complexed with Ome and AMP- PNP to *D. hafniense* PylRS complexed with tRNA (PDB ID 2ZNI) 147

LIST OF SCHEME

Scheme 4.1: Structures of DanAla and DanAla esters and synthetic approach to DanAla esters.....90

LIST OF TABLES

Table 2.1: Kinetics of current inactivation of Kv1.4 channels 55

Table 5.1: X-ray diffraction data processing and refinement statistics 145

ACKNOWLEDGEMENTS

I would first like to acknowledge my professor, Lei Wang, for his support and guidance over the past years. He has always had an open door policy and even when I would constantly barge into his office about a random idea I had, he was patient and always provided his insight. I would also like to committee through all their support and guidance throughout the years. They have provided me with great personal and professional advice.

I would like to thank my family for all the love and support they provided me throughout my life. My mother, Chikako Takimoto, thank you. She was put in a difficult position to raise three children on her own and I know that we were probably a pain to deal with but I can not thank her enough. Even though she worked tirelessly throughout the week, she found time for my extracurricular activities and never complained about them. She taught me that life it tough but you just have to deal with the cards that you are dealt and keep moving forward. To my sisters, thank you! Both of you had to deal with a younger brother that was energetic and had a difficult time listening. The two of you were always there for me and helped me throughout many hardships. Juliette, I would like to thank you for showing me about generosity and selflessness, although you were also a child, you helped raise one. Jennifer, I would like to thank you for showing the drive I needed to succeed and also for my new extended family. I would like to thank Thomas for welcoming me into your home and providing me with great professional advice. To Kyle and Jake, my two nephews, you two have brought me so much joy in my life. The two of you have given me more than you can imagine throughout my graduate career, putting things into perspective on what is important in life.

I would like to thank all of my lab mates throughout the years including those in both the Wang lab and Noel lab, especially Nikki Dellas, Kyle Merchant, Angela Parrish, David Johnson, Sam Ulin, Joseph Noel, and Marianne Bowman. Nikki Dellas was a graduate student in the Noel lab who has become an incredible friend throughout the years. During good and bad times, we were each other's support. Kyle Merchant is part of the Noel group who has grown to become of my best friends in San Diego. His love for music has reignited my passion for music. His birthday gift of 2.3 gigs of music is providing me with a journey through sound during my thesis. His moral support and comic relief has

made my graduate career and lab life extremely enjoyable. Angela and David are graduate students in the Wang lab who have been my close friends since the beginning of my graduate career. We started within a month of each other and have been each other's drive and moral support throughout the years. Sam Ulin is a graduate student in the Wang lab and has provided me with personal and professional support throughout the years. He continuously pushes me to recognize my weaknesses and helps me improve upon them. Joseph Noel is the neighboring lab professor who I would like to thank for taking the time out of his schedule to not only act as a second mentor but also a friend. He was able to transcend and balance professional lines that I look up to and hope to one day attain. Marianne Bowman is the lab manager that is the foundation of the lab and played an instrumental part in all my scientific endeavors. I would like to thank her for dealing with all my requests and questions throughout the years. She has been very understanding and a great person to work with on a personal and professional level.

I would like to thank Bari Ballew, Jimmy Ballew, and TGS. Bari and Jimmy, thank you for being such good friends during stressful times. The two of you are complete opposites and provide me the "yin and yang" balanced view to all of my problems. TGS, thank you my friends.

Greg Macias was an undergraduate student from the Noel lab who was able to transform my view of life. When he passed, the pain and self-reflection that came transformed my life forever. Thank you; you helped me realize the joy in life again.

Ultimately, I would like to thank Melissa Kanchanapoomi. During the toughest time in my life, she was by my side and never left. She had faith in my abilities even when I had lost all faith in myself. Thank you from the bottom of my heart and the only reason I am able to write this is because you were in my life.

Chapter 1 is a reprint in full of the material as it appears in the Wiley Encyclopedia of Chemical Biology, 2007. Johnson DBF, Takimoto JK, Xu J and Wang L (2007). Unnatural acids to investigate biological processes. Wiley Encyc Chem Bio. doi:10.1002/97804700486.72.wecb585. The dissertation author, D.B. Johnson, and J. Xu contributed equally to this manuscript. The dissertation author specifically contributed the methods sections in this review.

Chapter 2 is a reprint in full of the material as it appears in Nature Neuroscience, (2007), Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L., Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat Neurosci* 2007, *10* (8), 1063-72.

Chapter 3 is a reprint in full of the material as it appears in Molecular Biosystems, (2009), Takimoto, J. K.; Adams, K. L.; Xiang, Z.; Wang, L., Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol Biosyst* 2009, *5* (9), 931-4. Reproduced by permission of the Roayl Society of Chemistry (RSC)

Chapter 4 is a reprint in full of the material as it appears in Chembiochem, (2010), Takimoto, J. K.; Xiang, Z.; Kang, J. Y.; Wang, L., Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *Chembiochem 11* (16), 2268-72. The dissertation author and Xiang, Z contributed equally to this manuscript.

Chapter 5 is reprinted with permission from Takimoto, J. K.; Dellas, N.; Noel, J. P.; Wang, L., Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-native Amino Acids. *ACS Chem Biol*. Copyright 2011 American Chemical Society). The dissertation author and Dellas, N contributed equally to this manuscript

VITA

2003 – 2005	Undergraduate Research Assistant, University of California, Los Angeles
2005	Bachelor of Science, University of California, Los Angeles
2006 – 2008	Teaching Assistant, Department of Biological Sciences University of California, San Diego
2011	Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

1. Takimoto, J.K., Dellas, N., Noel, J.P. & Wang, L. Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-native Amino Acids. *ACS chemical biology*. doi: 10.1021/cb200057a (2011)

2. Takimoto, J.K., Xiang, Z., Kang, J.Y. & Wang, L. Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *Chembiochem : a European journal of chemical biology* **11**, 2268-72. (2010)

3. Takimoto, J.K., Adams, K.L., Xiang, Z. & Wang, L. Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Molecular BioSystems* **5**, 931-4 (2009).

4. Johnson, D.B., Takimoto, J.K., Xu, J., & Wang, L. Unnatural Amino Acids to Investigate Biologic Processes. *Wiley Encyclopedia of Chemical Biology* doi:10.1002/97804700486.72.wecb585. (2007)

5. Wang, W., Takimoto, J.K., Louie, G.V., Baiga, T.J., Noel, J.P., Lee, K., Slesinger, P.A. & Wang, L. Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nature Neuroscience* **10**, 1063-72 (2007).

ABSTRACT OF THE DISSERTATION

Unnatural Amino Acid Incorporation for
Genetic Code Expansion in Mammalian Cells

by

Jeffrey Kunio Takimoto

Doctor of Philosophy in Biology

University of California, San Diego, 2011

Professor Lei Wang, Chair

The genetic code of most organisms was evolved to encode 20 amino acids. Although the ability to encode 20 amino acids provides the basis to translate proteins necessary for life, researchers are also limited to these 20 amino acids for conventional site-directed mutagenesis. The ability to encode unnatural amino acids provides researchers the ability to circumvent the limitation imposed by the genetic code. Genetically encoding unnatural amino acids provides researchers the means to not only mimic naturally occurring posttranslational modifications but also the ability to encode amino acids with new physical or chemical properties to study biological processes.

The incorporation of unnatural amino acids into proteins had been developed in *Escherichia coli* and also in yeast. We have developed a methodology to genetically incorporate unnatural amino acids in mammalian cells in response to an amber codon (UAG). The incorporation of unnatural amino acids is high in *E. coli* and yeast, but the incorporation in mammalian cells is relatively low. In addition to developing the system to incorporate unnatural amino acids in mammalian cells, we have also improved suppression efficiencies by modifying the synthetase and unnatural amino acid. To incorporate unnatural amino acids in response to an amber codon, the tRNA anticodon is mutated from

a GUA to a CUA. We were able to show that engineering the anticodon-binding domain of the synthetase could enhance the recognition of the tRNA and thus increased suppression efficiencies. Furthermore, by masking the carboxyl group of the amino acid by an ester group, we were able to increase the bioavailability of an unnatural amino acid to further increase suppression efficiencies.

Most evolved synthetases aminoacylate unnatural amino acids that are structurally similar to the native substrate of the wild-type synthetase. We were able to adapt *Methanosarcina mazei* pyrrolysine synthetase (PylRS) to charge a considerable disparate amino acid from its native substrate, *O*-methyl-L-tyrosine. In addition, the X-ray crystal structure was solved for the evolved PylRS complexed with *O*-methyl-L-tyrosine at 1.75Å. This multifaceted approach provides the basis to engineer the PylRS to incorporate a significantly diverse selection of unnatural amino acids than previously anticipated.

Chapter 1

Introduction:
Unnatural Amino Acids to Investigate Biological Processes

1.1 Abstract

To circumvent the constraint imposed by the 20 canonical amino acids on the study of protein structure and function, various chemical and biosynthetic methods have been developed to incorporate unnatural amino acids into proteins. Unnatural amino acids now can be genetically encoded in living cells in a manner similar to that of common amino acids, which expands site-directed mutagenesis to diverse novel amino acids. The use of unnatural amino acids grants researchers a multitude of chemical and physical properties that cannot be found in the normal genetic repertoire, which significantly improves their ability to manipulate proteins and protein-involved biologic processes. Changes have been tailored into proteins to dissect accurately the contribution of hydrogen bonding, hydrophobic packing, cation-π interaction, and entropy to protein stability, as well as to precisely examine the structural and functional role of crucial residues. Unnatural amino acids also enable the introduction of new chemical reactivities, biophysical probes, mock posttranslational modifications, photoactive groups, and numerous other functionalities for the modification and regulation of protein activities. These studies not only reveal fundamental information of protein structure and function but also explore new means for generating novel protein properties and controlling biologic events.

1.2 Introduction

Conventional site-directed mutagenesis of specific amino acids currently is the preferred method for investigating various structural and functional characteristics of proteins. A serious limitation to this methodology is the constraint of using the 20 canonical amino acids fixed by the universal genetic code. This constraint lies in the limited chemical and physical properties of these amino acids, which hinder the ability to make precise alterations. For instance, modification of an amino acid such as glutamine is limited because only asparagine has similar characteristics. For amino acids such as proline, no analogous amino acid exists in the genetic repertoire, which makes it difficult to investigate the role of this amino acid in specific processes without abolishing it completely. Breaking this limitation would enable in-depth investigation of the principles underlying protein structure and function as well as the engineering of novel protein properties and cellular functions. In

the past decade, great progress has been made in incorporating unnatural amino acids into proteins to harness their extensive and powerful capabilities. Here, we will introduce unnatural amino acids with a brief overview of various methods for incorporating them into proteins and we will present examples that illustrate how unnatural amino acids have impacted a wide array of research that investigates biologic systems.

1.3 Unnatural amino acids

Common amino acids consist of an amino group, a carboxyl group, a hydrogen atom, and a side chain all attached to the Cα in the L configuration. Analogs with altered side chains, or those that deviate from other features, are generally called unnatural amino acids (**Figure 1.1**). The most widely used group of unnatural amino acids is the group in which the Cα side chain is changed. Variation of unnatural side chains is diverse and can range from structural analogs of canonical amino acids to those with specific chemical moieties, such as reactive functionalities and reporter groups for biophysical characterization. Modification of the amino group results in changes in the peptide backbone. For example, changing the amino group into a hydroxy or sulfhydryl group converts the endogenous amide bond between two residues into an ester and thioester link, respectively. Such changes make the resultant analog no longer an "amino" acid but an α-hydroxy acid and a thio acid. An analog with aminooxy replacing the amino group also has been incorporated into protein biosynthetically[1]. The amino group also can be alkylated with different moieties to form amino acids that contain secondary amines[2]. Another category of unnatural amino acids is α,α- disubstituted amino acids, whose α-hydrogen is replaced by an additional side chain. Moving the amino group away from the α carbon leads to extended β- or γ-amino acids, which can be compatible with protein biosynthetic machinery as well[3]. Finally, D-amino acids, which are mirror images of the L counterparts, have been introduced selectively into functional proteins to study structural characteristics[4].

1.4 Methodology

Although this review is mainly focused on the use of unnatural amino acids in the investigation of biologic systems, a brief background of the methodology of incorporation is useful in understanding the power and application of this technology. A more comprehensive coverage of various methods can be found in[5] and the references contained therein.

1.4.1 Chemical approaches

Global alterations to certain amino acids can be done *in vitro* through chemical modification of their exposed reactive side chains[6]. The selectivity of chemical modification relies on the differences in chemical reactivity of amino acid functional groups. Judiciously selected chemicals will react with specific amino acids only, which allows chemical changes to be applied to that amino acid alone. Typical modification involves the thiol group of Cys, the ε-amino group of Lys, the carboxylate group of Asp and Glu, and the N-terminal amino group. The hydroxy group of Ser and Thr can be oxidized selectively when Ser and Thr are at the N-terminus of the protein[7]. Side chains of Tyr and Trp can be modified selectively with transition metal catalysts[8, 9]. Initial applications of this method focused on determining the functional roles of certain amino acid species for biologic activity, but they have expanded to other applications, including biophysical probe tagging, chemical cross-linking, and the conjugation of various synthetic functionalities. However, site-specific alterations are difficult using this approach because the chemical will react with all accessible target amino acids if more than one such amino acid exists in the protein. In addition, chemical modification must be done *in vitro*, which affects only those side chains that are solvent-accessible.

Another method to introduce unnatural amino acids into a polypeptide chain is through complete chemical synthesis[10]. The predominantly used method, stepwise solid-phase peptide synthesis (SPPS), attaches the C-terminal amino acid to a solid support, and amino acids are added one at a time to the N-terminus. A clear advantage of chemical synthesis is that it enables the accurate introduction of unnatural amino acids at any site in a protein. The number of unnatural amino acids that can be introduced is limited only to the size of the chain, and chains of entirely unnatural amino acids can be

produced using this method. Chemical synthesis is useful particularly for the incorporation of isotopic labels and unnatural amino acids that are toxic to cells or incompatible with the translational machinery. However, construction of a polypeptide chain using even the most advanced chemical synthesis techniques is daunting when confronted with the construction of an entire protein, as these methods currently are limited to approximately 100 amino acids[10].

Semisynthetic protein ligation methods, in which two or more protein fragments of recombinant or synthetic origin are chemically ligated to make the full-length protein [11], overcome the size limitation of SPPS. Among these methods, the native chemical ligation strategy couples peptide fragments to form a native peptide link, which leaves no chemical artifacts behind[12, 13]. The desired unnatural amino acid is introduced in the synthetic fragment by using chemical synthesis and thus is incorporated into proteins after ligation. Once this unnatural protein is folded, biochemical characterization of kinetic parameters and function can be performed. Peptide ligation in living cells is also possible. A synthetic fragment can be injected into cells to react with an endogenously produced protein fragment[14]. This method, like SPPS, has the power to introduce various unnatural structures that are synthetically accessible. However, it requires appropriate sites for cleavage and ligation, and it becomes cumbersome for internal sites in large proteins. Microinjection of either the *in vitro* ligation product for *in vivo* studies or the synthetic fragment for *in vivo* ligation can be a drawback to this method as well.

1.4.2 Biosynthetic approaches

Methods that use the endogenous cellular machinery to introduce unnatural amino acids into proteins are not limited by protein size and will facilitate the investigation of biologic processes *in vivo*. A general *in vitro* biosynthetic method allows for the site-specific incorporation of unnatural amino acids into proteins[15]. In this method, a suppressor tRNA is chemically acylated with an unnatural amino acid, and the codon of interest in the target gene is mutated to the amber stop codon, TAG. When added to cell extracts that support transcription and translation, the suppressor tRNA recognizes and selectively incorporates the attached unnatural amino acid in response to the UAG in the transcribed

mRNA. Using this method, a variety of unnatural amino acids have been incorporated into proteins, regardless of position or protein size, and have been applied to a large number of problems in protein chemistry[16]. Besides the amber stop codon, rare codons and extended codons also have been used to specify the unnatural amino acid[17]. An extension of this method involves the microinjection of the chemically acylated tRNA and UAG-containing mutant mRNA into *Xenopus* oocytes[18]. The endogenous oocyte protein synthesis machinery supports translation and incorporation of the unnatural amino acid. This method enables the structure-function studies of integral membrane proteins, which are generally not amenable to *in vitro* expression systems[19]. A purified *in vitro* translation system that consisted of only ribosomes, initiation factors, elongation factors, mRNA, and tRNAs preloaded with desired amino acids was used to incorporate simultaneously several unnatural amino acids into peptides in response to sense codons[20]. By reassigning the meaning of codons, this system ultimately may allow the synthesis of peptides and proteins that contain multiple unnatural amino acids. The drawback to these methods lies in the chemical acylation of the suppressor tRNA, which is demanding technically and can exclude certain unnatural amino acid from attachment. In addition, acylated tRNA is consumed stoichiometrically and cannot be regenerated in cells or cell extracts, which leads to low expression of the target protein. Multisite incorporation of unnatural amino acids by using cellular machinery has been achieved in auxotrophic bacterial strains[21] and in mammalian cells[22]. This method relies on the idea that aminoacyl-tRNA synthetases can mischarge unnatural amino acids that are close structural analogs of the cognate amino acids, although with very high substrate specificity. An unnatural amino acid analogous to a canonical counterpart is introduced into a bacterial strain that is incapable of producing the natural amino acid or into mammalian cells that are deprived of the natural amino acid. The translational machinery then replaces the natural amino acid with its analog in all proteins. The incorporation efficiency of unnatural amino acids can be improved by increasing the expression level of the synthetase[23] and by introducing mutations that relax the substrate specificity of the aminoacylation domain[24] or attenuate the proofreading function of the editing domain of certain synthetases[25]. However, this strategy is limited because it is restricted to global replacement of one amino acid with an analog and does not allow specific single alterations with a specific protein.

It would be ideal to genetically encode an unnatural amino acid in a manner similar to that of common amino acids, which would enable site-directed mutagenesis in living cells with unnatural amino acids. A general method to expand the genetic code to include unnatural amino acids was developed. It involves the generation of a new tRNA- codon-synthetase set that is specific to the unnatural amino acid and does not crosstalk with other sets for common amino acids[26]. The new synthetase is evolved to charge specifically an unnatural amino acid onto the new tRNA. This tRNA recognizes a codon that does not encode any common amino acids (e.g., a stop codon or an extended codon). When expressed in cells, the new tRNA-synthetase pair enables the unnatural amino acid to be site-specifically incorporated into proteins at the unique codon with high fidelity and efficiency. This method allows the use of unnatural amino acids in the investigation of biologic systems in an *in vivo* setting. It may be possible to generate stable cell lines or transgenic animals capable of inheriting such alterations for long-term studies. However, toxic unnatural amino acids and those incompatible with the protein biosynthesis machinery cannot be incorporated using this approach.

1.5 Applications of unnatural amino acids

Unnatural amino acids enable the structural, chemical, and physical properties of the building blocks of proteins to be customized according to needs. Such tailored changes have contributed to our understanding of the fundamental questions of protein chemistry on the molecular and atomic level, have been used to modify and enhance protein properties, and are being exploited to control protein activities to investigate various biologic processes and to create novel biologic functions.

1.5.1 Protein stability

There are many factors contributing to protein stability, including hydrogen bonding, hydrophobicity, packing, and conformational entropy, among others. It is difficult to access individual contributions by using conventional mutagenesis because changing one common amino acid to another often alters several properties at a time. For example, mutagenesis to disrupt hydrogen bonds, usually by deleting one member of a hydrogen-bonded pair, will leave an unpaired hydrogen donor or acceptor

and/or alter local solvation and packing interactions, all of which may lead to protein destabilization. To determine the effect of side-chain hydrogen bonding on protein folding, Tyr27 in staphylococcal nuclease (SNase) was replaced with several isosteric, fluorinated tyrosine analogs (analogs 1 to 3) (**Figure 1.2**)[27]. These unnatural amino acids were designed to gradually increase the strength of the Tyr27-Glu10 hydrogen bond while minimizing the steric and electronic perturbations associated with deleting one hydrogen-bonding member. The stability constants *K*app of the corresponding mutants were found correlative with the p*K*a of the hydroxyl group in the tyrosine analogs. This result provides strong evidence that intramolecular side-chain hydrogen bonds preferentially stabilize the folded state of a protein relative to the unfolded state in water.

α-Hydroxy acids have been used to study the contribution of the backbone hydrogen bonds to protein stability (**Figure 1.3**). The replacement of a common amino acid with an α-hydroxy acid that contains the same side chain effectively substitutes a good hydrogen-bond acceptor (the amide carbonyl group) with a considerably weaker one (the ester carbonyl group) in a conservative manner and disrupts a potential backbone hydrogen bond because the ester link cannot serve as a hydrogen-bond donor as does the NH. α-Hydroxy acids were incorporated at the N-terminus, the middle, and the C- terminus of the α-helix 39-50 of T4 lysozyme[28]. At the N-terminus and the C- terminus, where only one hydrogen-bonding interaction is perturbed, the ester substitution destabilizes the protein by 0.9 kcal mol−1 and 0.7 kcal mol^{-1}, respectively. In the middle of the helix, where such substitution perturbs two hydrogen bonds, the protein is destabilized by 1.7 kcal mol^{-1}. In another study, Leu14 in an antiparallel β sheet of SNase was replaced with leucic acid[29]. This amide-to-ester change decreases the stability by 1.5-2.5 kcal mol^{-1}. Altogether, these results convincingly show that both side-chain hydrogen bonds and main-chain hydrogen bonds significantly contribute to protein stability.

To examine the importance of the packing interaction in the core of a protein, Leu133 in T4 lysozyme was replaced with a series of analogs with extended or shortened alkyl side chains (analogs 4 to 7)[30]. Leu133 lies along the edge of the largest cavity in the interior of T4 lysozyme, which makes it possible to change the bulk of the side chain with minimal concomitant strain. Incorporation of (S,S)-2-amino-4- methylhexanoic acid (analog 4) and (S)-2-amino-4-cyclopentylpropanoic acid (analog 5)

stabilizes T4 lysozyme by 0.6 kcal mol^{-1} and 1.24 kcal mol^{-1}, respectively, which indicates that the increased bulk of buried hydrophobic residues can enhance protein stability. During protein folding, the cyclic amino acid 5 will lose less conformational entropy than does 4. That the 5-containing mutant is more stable than the 4-containing mutant suggests that side-chain entropy also affects protein stability. As expected, when the side chain of Leu133 is shortened systematically, as in unnatural amino acids 6, 7, and alanine, the protein becomes increasingly less stable.

Another method for increasing hydrophobicity while minimizing structural perturbation is to replace hydrocarbons with fluorocarbons. Using solid-phase synthesis, L-5,5,5,5′,5′,5′-hexafluoroleucine 8 was substituted for seven core leucine residues in a 30-residue peptide that can form homodimeric coiled coils. Hydrophobic side chains of the core residues pack against each other in the coiled coil. Fluorination of these side chains increased the hydrophobicity and raised the melting temperature of the homodimer from 34◦C to 82◦C[31]. In addition, fluorocarbons are insoluble in hydrocarbons at room temperature and, thus, form a fluorous phase by interacting with other fluorocarbons. When a disulfide-bound heterodimer of the hexafluoroleucine core peptide and a leucine core peptide was allowed to undergo disulfide exchange, the peptides self-sorted into homodimers[31].

This fluorous effect could lead to a novel protein-protein recognition. In another report, six leucine residues in the hydrophobic core of an antiparallel 4-α-helix bundle were replaced by hexafluoroleucine 8 [32]. The free energy of the unfolding of the mutant peptide increases by 0.3 kcal mol^{-1} per residue when the two central leucines are substituted and by an additional 0.12 kcal mol^{-1} per residue when the outer leucines are replaced, which confirms that hydrophobic packing stabilizes proteins.

1.5.2 Protein structure and function

Unnatural amino acids can be designed to elucidate the functional role of a residue that is misinterpreted by or remains ambiguous to conventional mutagenesis and other methods. For example, Glu43 is important for the catalytic activity of SNase because its replacement by Asp and Gln significantly decreases the catalytic efficiency. Previous structural and mutagenesis studies suggested

that Glu43 functions as a general base to activate a water molecule for hydrolyzing the phosphodiester bond of DNA. However, substitution of Glu43 with either homoglutamate (analog 9) or (S)-4-nitro-2-aminobutyric acid (analog 10) yielded mutant enzymes with kinetic constants similar to those of wide-type SNase[33]. Because these two unnatural amino acids are isoelectronic and isosteric to glutamate but a much poorer base, such substitution would decrease SNase activity if Glu43 were a general base during catalysis. In addition, the X- ray crystal structure of the homoglutamate mutant showed that the carboxylate side chain of this residue occupies a position and orientation similar to that of Glu43 in the wildtype enzyme. Therefore, Glu43 may play a structural role instead and serve as a bidendate hydrogen-bond acceptor to fix the conformation of the neighboring loop.

Proline is unique among the natural amino acids in that its α-nitrogen is part of a pyrrolidine ring. The proline residue disrupts main-chain hydrogen bonding; it cannot serve as a hydrogen-bond donor because of the lack of a backbone NH moiety. Also, proline forms *cis*-peptide bonds at a frequency (5%) much higher than any other natural amino acids (<0.1%). In ion channels, Pro often is conserved at crucial sites, such as Pro221 in the nicotinic acetylcholine receptor (nAChR) and Pro256 in the 5- hydroxytryptamine-3A receptor (5-HT3*A*R). To probe which feature of Pro is functionally significant, α-hydroxyl acids (analogs of Gly, Val, and Leu) were incorporated at these sites, which all produced mutant receptors with properties similar to the wildtype receptor[34, 35]. In contrast, incorporation of canonical amino acids Gly, Ala, or Leu yielded nonfunctional receptors. Because α-hydroxyl acids similarly lack the NH moiety for backbone hydrogen bonding and the nature of side chains does not affect receptor activity, these results suggest that the functional importance of the conserved Pro in both receptors is to remove backbone hydrogen bonding.

Another conserved proline residue of the 5-HT3AR, Pro308, has been shown to be indispensable for channel gating using conventional mutagenesis. However, substitution of this Pro with α-hydroxy acids produced nonfunctional receptors, which suggests that the lack of backbone hydrogen bonding is not the key to the proper function of this Pro. Interestingly, proline analogs that strongly favor the *trans*-conformer (analogs 11 and 12) produced no gating response, but those that favor the *cis*-conformer (analogs 13 and 14) yielded highly sensitive channels. Moreover, a linear

energy correlation was observed between the *cis-trans* energy gap of the proline analogs and the receptor activation[36]. This study strongly suggests that the critical role of Pro308 is to provide the switch that interconverts the open and closed states of the channel through *cis-trans* isomerization.

Cation-π interaction is a noncovalent electrostatic interaction between a cation and the electrons in π orbitals, which plays an important role in protein structure, binding, and catalytic function. The energetic contribution of this interaction to proteins could not be measured accurately with conventional mutagenesis because no positively charged natural isosteres exist for common amino acids. To engineer a cation-π interaction in SNase, Val74, which occupies a hydrophobic pocket composed of one tyrosine side chain and two phenylalanine side chains, was replaced with the positively charged S- methylmethionine (analog 15). Another mutant was made by replacing Val74 with homoleucine (analog 16), which is isosteric to S-methylmethionine. Comparison of the thermodynamicstability of these two mutant proteins showed that the magnitude of cation-π interaction is about 2.6 kcal mol^{-1} [37].

Several aromatic amino acids have been identified near the agonist-binding site of the nAChR, which suggests that cation-π interactions may be involved in binding the quaternary ammonium group of the agonist acetylcholine. A series of progressively fluorinated tryptophan derivatives (analogs 17 to 20) were incorporated at αTrp149. Because fluorine is an electron-withdrawing group, substitution of H with F in the aromatic ring weakens the cation-π interaction. *Ab initio* quantum mechanics was used to predict the cation-π-binding abilities of the fluorinated tryptophans, and the calculated binding energy has a linear relationship with receptor activation by the agonist[38]. Such correlations were not observed for other aromatic residues, which suggests that the cation-π interaction indeed exists for agonist binding and pinpoints it to αTrp149. This interaction was shown later as a general binding pattern between the Cys-loop superfamily of neurotransmitter receptors, such as the 5-HT3A receptors and the γ- aminobutyric acid receptors, and their cationic ligands or substrates[39].

The site-specific introduction of biophysical probes into proteins has proven extremely powerful in revealing subtle changes of proteins with high spatial resolution. The carbon-deuterium (C-D) bond absorbs at ~2100 cm^{-1}, which is within the transparent IR window (~1800-2700 cm^{-1}) of

proteins and, therefore, makes it easily observable by IR spectroscopy. The inherently fast timescale of IR spectroscopy also provides high temporal resolution. Therefore, unnatural amino acids with C-D bonds are excellent probes of protein folding and dynamics. Absorptions at different frequencies indicate the existence of multiple intermediates, and an increased line width of the absorption shows increased flexibility of the local environment. Amino acids containing C-D bonds were incorporated at different positions throughout cytochrome *c* (cyt *c*) by using semisynthetic approaches[40]. By characterizing the absorption frequencies and line widths of the C-D bonds of these residues, it was found that no significant difference exists in the flexibilities of the oxidized and reduced states of cyt *c*. The data also show that parts of the protein exist in dynamic equilibrium with locally unfolded states and that cyt *c* is less stable than previous studies suggest.

Another infrared probe, *p*-cyano-L-phenylalanine (*p*CNPhe, 27), has been genetically encoded in *E. coli* and used to examine different ligand-bound states of the heme group in myoglobin[41]. The stretching vibration of the nitrile group of *p*CNPhe has strong absorption and a frequency (νCN) at ~2200 cm^{-1}, which falls in the transparent window of protein IR spectra. A substitution of *p*CNPhe was made for His64, which is at the distal face and close to the iron center of the heme group in myoglobin. In the ferric myoglobin, when the Fe(III) ligand was changed from water to cyanide, νCN shifted from 2248 cm^{-1} to 2236 cm^{-1}, which indicates a less polar active site. In the ferrous myoglobin, a νCN absorption at 2239 cm−1 was observed for the linear Fe(II)CO complex, and the bent Fe(II)NO and Fe(II)O2 complexes showed a νCN absorption at 2230 cm^{-1}. These results demonstrate that the nitrile group is a sensitive probe for ligand binding and for local electronic environment.

Small fluorescent probes sensitive to various environmental changes have the great potential for monitoring many biologic events as a complementary reporter for the widely used fluorescent proteins. For example, L-(7-hydroxycoumarin-4-yl)ethyl- glycine (CmrGly, 28) has been incorporated into holomyoglobin to study its local unfolding[42]. CmrGly was incorporated at position Ser4 in helix A and at position His37 in helix C, respectively. The coumarin fluorescence intensity increases with solvent polarity. When the Ser4CmrGly mutant was unfolded with 2M urea, its fluorescence increased 30%, which indicates that helix A is disordered. In contrast, the fluorescence intensity of the

His37CmrGly mutant did not change significantly until the urea concentration was raised to 3 M. These results suggest helix C and helix A unfold at different times and concentrations of the denaturing agent.

1.5.3 Modification and regulation of protein activity

Green fluorescent protein (GFP), whose chromophore is autocatalytically formed by the tripeptide Ser65-Tyr66-Gly67, has become one of the most important *in vivo* markers for biologic studies. An aromatic amino acid at position 66 is necessary for fluorescence generation. To determine how the spectral properties of GFP could be altered by this residue, tyrosine analogs bearing different substituents at the *para* position of the phenyl ring (analogs 21 to 24) were used to replace Tyr66[43]. The absorbance and fluorescence emission maxima of mutant GFPs are all blue-shifted, spanning the range from 375 to 435 nm and 428 to 498 nm, respectively. The wavelengths of the maxima increase in the order of bromo, iodo, methoxy, hydroxyl, amino, and deprotonated hydroxyl group. This shifting trend is consistent with the electron-donating ability of the substituents. In another experiment, Trp66 of the enhanced cyan fluorescent protein was replaced with L-4-aminotryptophan (analog 25)[44]. The electron-donating amino group significantly red-shifts the fluorescence emission by 69 nm, which changes the color from cyan to gold.

Comparison of the *p*-methoxy-Phe (22) mutant GFP with wildtype GFP also provides direct evidence for the peak assignment of GFP. Wildtype GFP has two absorbance maxima at 397 nm and 475 nm, which are believed to correspond to a neutral chromophore (phenol of Tyr66) and an anionic chromophore (phenolate anion of Tyr66), respectively. Excitation at either absorbance peak leads to a single fluorescence emission centered at 506nm, which corresponds to the anionic chromophore in the excited state[45]. Picosecond spectroscopy revealed that the excited neutral chromophore should emit at 460nm[46]. The absence of 460 nm emission in wildtype GFP suggests that an excited state proton transfer process is involved. Substitution of the hydroxyl group of Tyr with a methoxy group removes the possibility of deprotonation and proton transfer. Indeed, when Tyr66 is replaced with *p*-methoxy-Phe, only one absorbance maximum at 394nm is observed, which is close to the absorbance maximum of the neutral chromophore of wildtype GFP. Moreover, only one emission maximum at 460nm is

detected for this mutant, which corroborates the ultrafast spectroscopic results[43].

The specificity of nucleic acid-binding proteins relies greatly on the hydrogen bonding between protein polar atoms and nucleic acid bases. Unnatural amino acids that can change isosterically the hydrogen-bonding pattern have been exploited to alter the substrate specificity. The λ-repressor recognizes the C:G pair at position 6 in the operator site OL1, and Lys4 of the λ-repressor is crucial for this recognition. The ε-NH2 group of Lys4 forms hydrogen bonds with the carbonyl group of Asn55 and the 6-oxo group of the guanine, which function as two hydrogen bond donors. Substitution of Lys4 with isosteric *S*-(2-hydroxyethyl)-cysteine changes the ε-NH2 to the -OH group, which now should accept hydrogen bonding from the amino group of adenine while preserving hydrogen bonding with Asn55 as a donor (**Figure 1.4**). In fact, after the unnatural amino acid was introduced into the λ-repressor through site-directed mutagenesis and chemical modification, the binding specificity was switched from the C:G to T:A base pair[47].

The chirality of D-amino acids has been harnessed for pharmaceutical purposes. D-peptide ligands should be resistant to proteolytic degradation and thus are more desirable as drugs. However, large libraries of D-conformers cannot be encoded genetically and expressed for selection. A method termed mirror-image display solved this problem in an intriguing way[48]. An L-peptide library is encoded genetically and displayed on the phage surface, and peptides of this library are selected by the target protein that is synthesized using all D-amino acids. The identified L-peptide then is resynthesized using D-amino acids, which should interact with the target protein of the natural handedness for reasons of symmetry. This approach has been used successfully to identify D-peptides that bind the Src homology 3 domain of c-Src and the HIV-1 gp41 protein[48, 49].

Unnatural amino acids that mimic posttranslational modifications can control protein functions. For example, protein phosphorylation regulates many signal transduction pathways and is a reversible process catalyzed by various phosphatases and kinases. The dynamic change of the phosphorylation status of a protein makes it difficult to study the effect of this modification in detail. The generation of metabolically stable phosphoproteins would be useful to dissect the function and to direct signal transduction. Unnatural amino acid *p*-carboxymethyl-L-phenylalanine (*p*CMF[26]) is a non-

hydrolyzable analog of phosphotyrosine and was found capable of mimicking the phosphorylated state of Tyr. This capability was demonstrated in a model phosphoprotein, the human signal transducer and activator of transcription-1 (STAT1). STAT1 has only a weak affinity for DNA, but during phosphorylation of Tyr701, STAT1 forms a homodimer and strongly binds a DNA duplex that contains M67 sites. The mutant STAT1 withTyr701 substituted with *p*CMF also bound the M67-containing DNA duplex tightly, which suggests that *p*CMF could replace phosphotyrosine in the generation of constitutively active phosphoproteins[50].

The development of photoactive amino acids provides researchers with an extremely useful tool not only to probe biologic function but also to control spatially and temporally a variety of biologic processes. One strategy is to attach a suitable photoremovable protecting group to the amino acid, which renders the amino acid inactive. Photolysis releases the caging group and converts the amino acid to an active form, which generates abrupt or localized changes to the target protein. The 2- Nitrobenzyl derivative is the most prevalent form for caged compounds. For example, the conserved Ser1082 at the upstream splice junction of the self-splicing DNA polymerase of *Thermococcus litoralis* was substituted with *o*-(2-nitrobenzyl)serine (**Figure 1.5a**). The full-length precursor protein underwent protein splicing only when the unnatural residue was reverted back to wildtype Ser during photolysis[51]. In other examples, *o*-nitrobenzyltyrosine (**Figure 1.5b**) was used to replace Tyr93 or Tyr198 in the α subunit of the nAChR. These two Tyr residues are highly conserved for agonist binding. Millisecond flashes of light at 300-350nm decaged the protected tyrosines and produced abrupt increments of currents that were conducted by the ion channel[52]. Also, *o*-Nitrobenzyltyrosine has been incorporated at the essential Tyr503 site of β-galactosidase to activate its enzymatic activity by using light both *in vitro* and in *E. coli*[53]. Mutation of the active-site cysteine residue in the proapoptotic protease caspase 3 to *o*-nitrobenzylcysteine led to a catalytically inactive enzyme, whose activity could be restored by photocleavage[54]. In addition to caging the active side chains, the 2-nitrobenzyl group has been harnessed also to cleave the protein backbone photo- chemically. 2-Nitrophenyl glycine (**Figure 1.5c**) was introduced into sites of the signature disulfide loop of the nAChR. Irradiation at 360nm resulted in site-specific backbone lesion and an almost complete loss of nAChR activity[55].

Photolysis of a caged amino acid residue is an irreversible process. Reversible modulation can be achieved with the photochromic azobenzene compounds. Azobenzene undergoes a reversible *cis*-*trans* isomerization: The more stable *trans*-isomer can be converted to the *cis*-isomer during illumination at 320-340nm, and the *cis*-form can revert to *trans*-form either thermally or by irradiation at >420nm. The resultant change in geometry and/or dipole of the compound can be used for regulating protein activity. For example, a known K+ channel blocker, tetra-ethyl ammonium, was linked via an azobenzene group to a cysteine that was introduced at specific sites of a K+ ion channel (**Figure 1.6a**). When the azobenzene group isomerizes between the extended -orm and the shorter *cis*-form in response to specific wavelengths of light, the structural change moves the blocker into or out of channel-blocking position and, thus, opens and closes the ion channel, respectively[56]. Such photomodulation can be used to control neuronal activity noninvasively. The azobenzene group has been encoded genetically in the form of phenylalanine-4′-azobenzene (AzoPhe). AzoPhe was incorporated at the Ile71 site of the *E. coli* catabolite activator protein, a transcriptional activator. Its binding affinity for the promoter sequence decreased fourfold after irradiation at 334 nm (**Figure 1.6b**), which converts the predominant *trans* -AzoPhe to the *cis*-form. The isomerized *cis*-AzoPhe then was switched back to the *trans*-state by irradiation at >420nm, after which the affinity of the protein for the promoter was completely recovered[57].

1.6 Future Directions

The examples summarized here are only representative and by no means comprehensive. Many unnatural amino acids now can be incorporated, but simply have yet to be used in the investigation of biologic function. Unnatural amino acids that contain photocross-linkers, biophysical probes, chemical moieties with unique reactivities, and posttranslational modifications, among many others, have much promise in their capabilities. The use of these amino acids will expand the capabilities of probing protein structure and function as well as protein-involved biologic processes. The methodology of incorporation is advancing as well. It may be possible to encode genetically unnatural amino acids in many other cell types and organisms. The incorporation of multiple unnatural

amino acids simultaneously by using extended codons may enable more complex investigations to be performed.

Additional work using unnatural amino acids can lead to the design and synthesis of novel and diverse biologic functions. By incorporating specific chemical moieties and physical characteristics into proteins, new protein properties may be discovered and used. Such exploration can be attempted either rationally or combinatorially. Diversities of protein libraries would be increased greatly by the addition of only a few unnatural amino acids, which may enhance the probability of discovering proteins that contain novel properties and functions. It is easy to see how unnatural amino acids can be extended into the pharmaceutical industry to create more efficient therapeutics. Finally, the creation of a sustainable organism that is capable of using unnatural amino acids will enable the investigation of the evolution of the genetic code on this planet.

1.7 ACKNOWLEDGEMENTS

We thank A.R. Parrish for help editing this manuscript. J. Xu is supported in part by the Pioneer Fellowship of the Salk Institute. L. Wang is a Searle Scholar and a Beckman Young Investigator.

Chapter 1 is a reprint in full of the material as it appears in the Wiley Encyclopedia of Chemical Biology, 2007. Johnson DBF, Takimoto JK, Xu J and Wang L (2007). Unnatural acids to investigate biological processes. Wiley Encyc Chem Bio. doi:10.1002/97804700486.72.wecb585. The dissertation author, D.B. Johnson, and J. Xu contributed equally to this manuscript. The dissertation author specifically contributed the methods sections in this review.

1.8 FIGURES

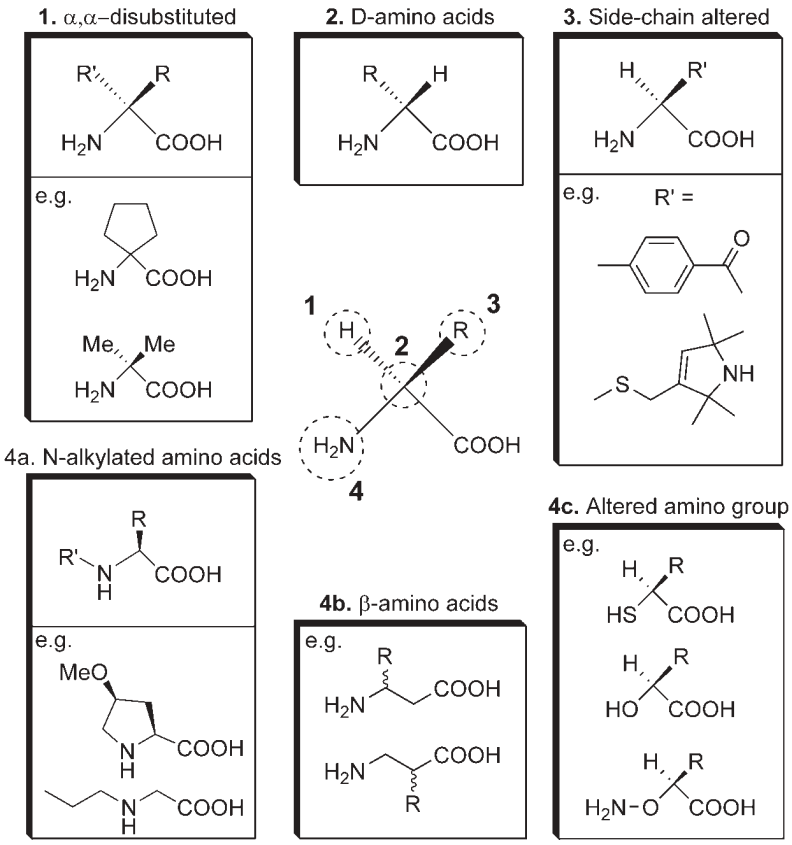


Figure 1.1: Different forms of unnatural amino acids

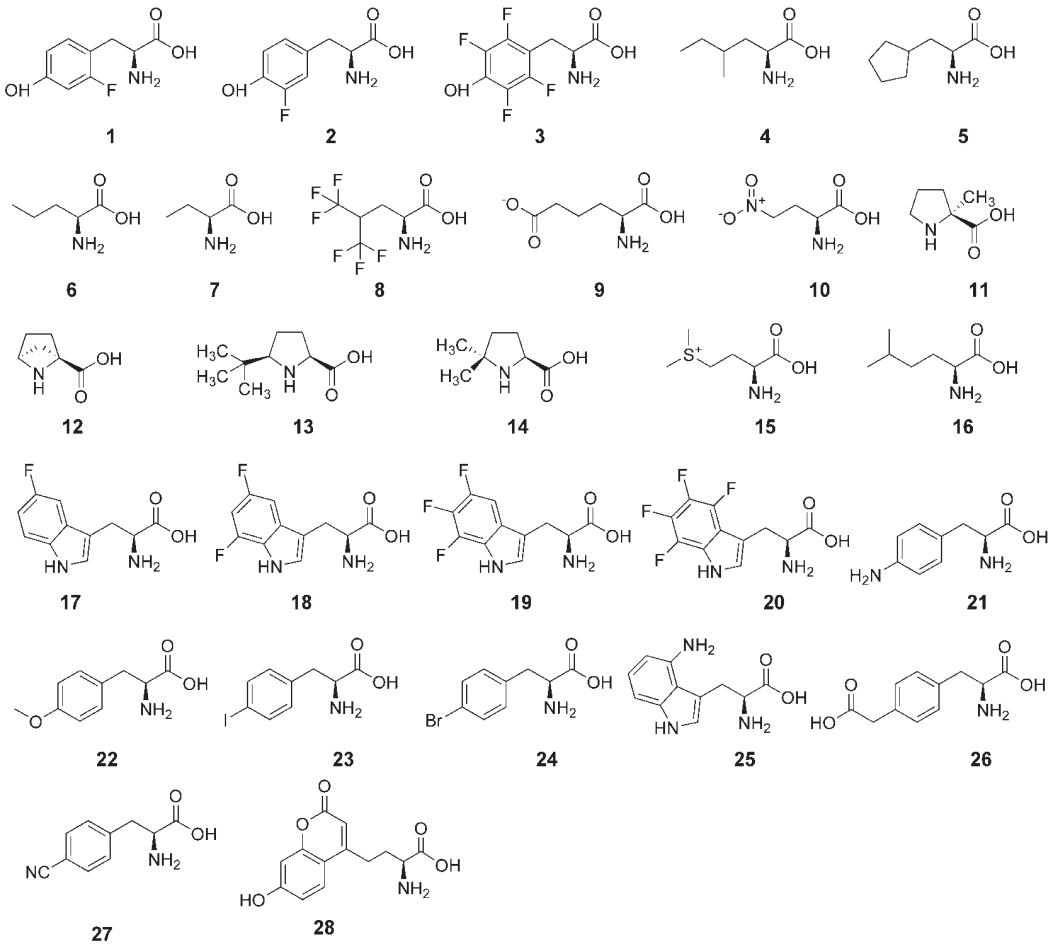


Figure 1.2: Structures of unnatural amino acids discussed in the text

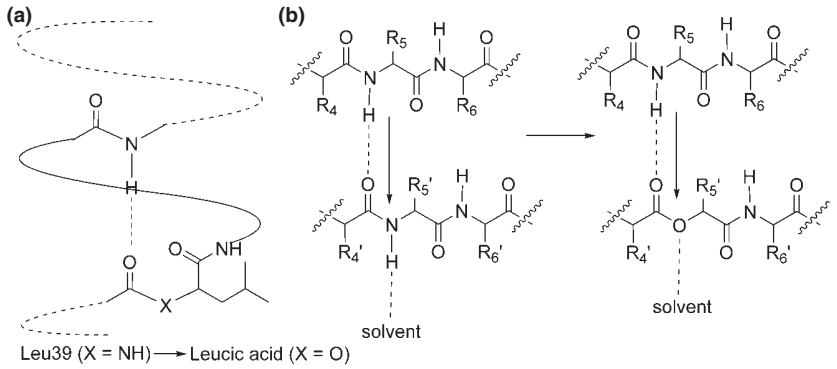


Figure 1.3: Backbond mutations generated by α-hydroxy acids. (a) N-terminal mutation of Leu39 to leucic acid in an α-helix of the T4 lysozyme. (b) Substitution of Leu14 with leucic acid in a β sheet of SNase

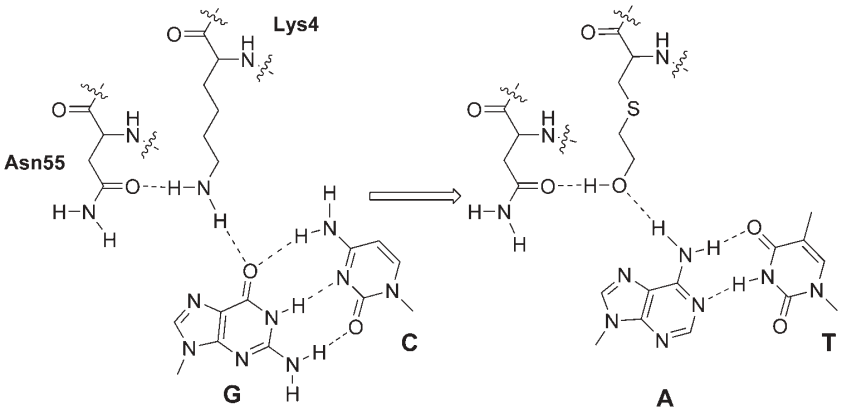


Figure 1.4: Substitution of Lys4 with 2-hydroxylethyl-cysteine in the λ-repressor changes the hydrogen-bonding pattern and DNA substrate specificity from C:G to T:A.

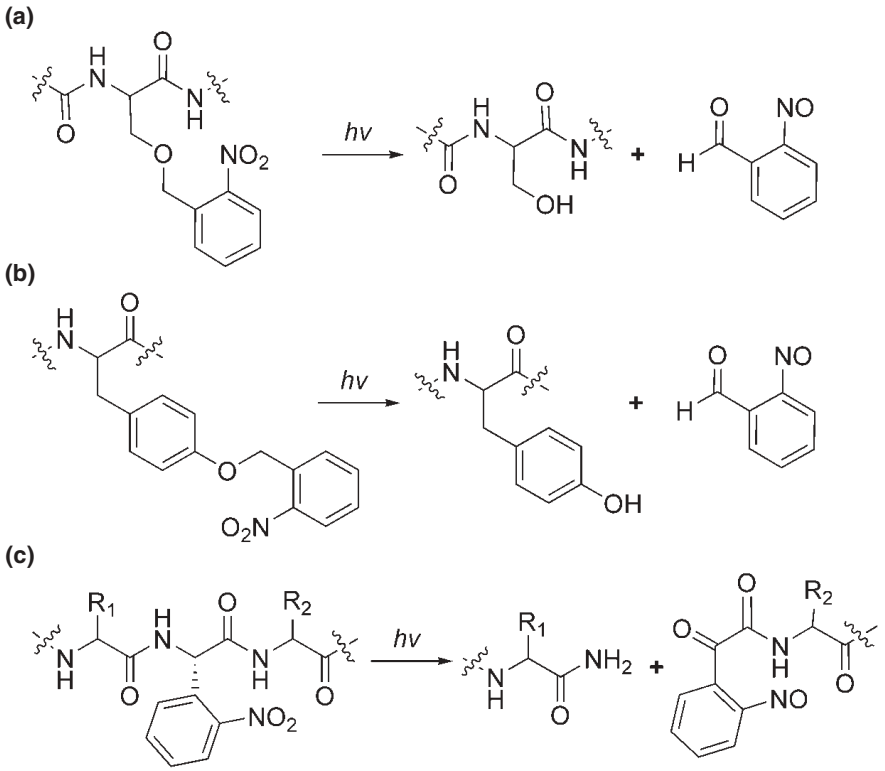


Figure 1.5: Photolysis of 2-nitrobenzyl caged serine (a) and tyrosine (b) restores the wildtype residues. Photolysis of 2-nitrophenyl glycine (c) cleaves the protein backbone.

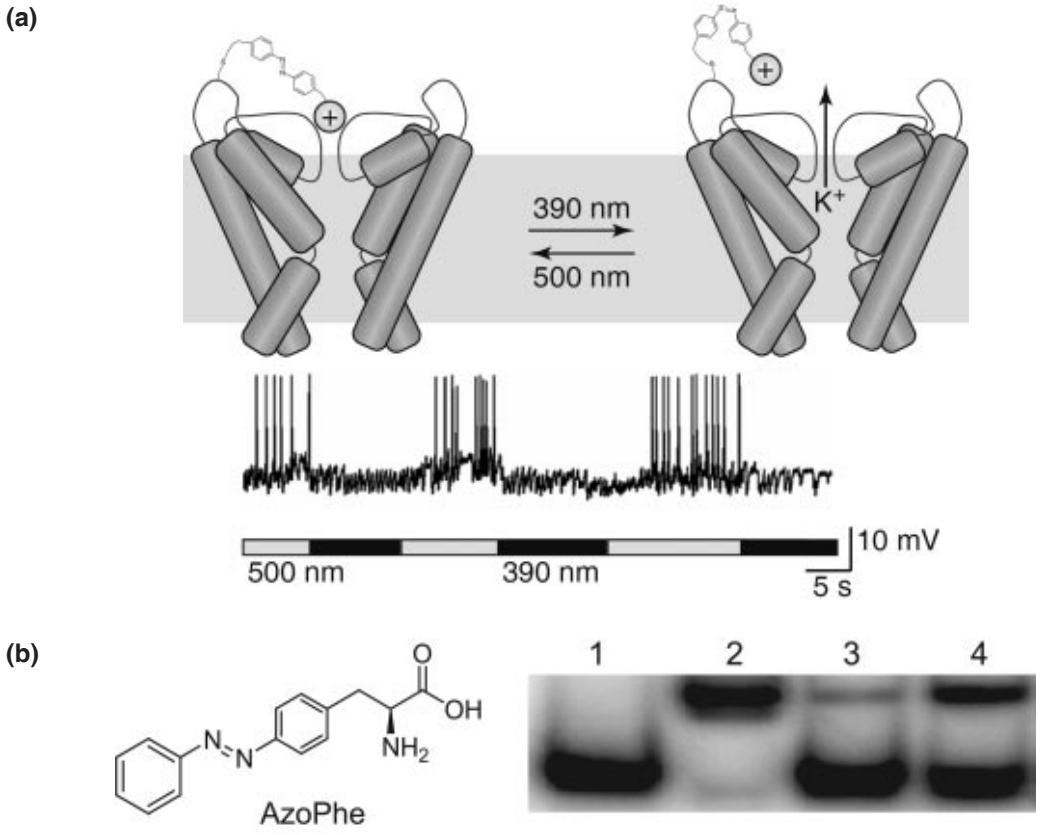


Figure 1.6: Azobenzene group isomerization (a) The geometrical change resultant from the cis-trans isomerization of azobenzene moves an ion channel blocker in and out of the ion channel to close and open the ion channel, respectively. Such activity was used to modulate the spontaneously firing hippocampal neurons. The firing frequency is significantly decreased when the azobenzene is in the cis-form after irradiation at 390 nm. Normal firing behavior is restored during irradiation at 500 nm (reprinted from[58], Copyright 2005, with permission from Elsevier). (b) Structure of phenylalanine-4′-azobenzene (AzoPhe) in trans-form and gel mobility shift assay to determine the binding affinity of the catabolite activator protein (CAP) to the lactose promoter DNA fragment (reprinted with permission from[57], Copyright 2006, American Chemical Society). Lane 1, DNA only. Lane 2, DNA+wildtype CAP. Lane 3, DNA+CAP with AzoPhe incorporated at residue 71 (after irradiation at 334 nm). Lane 4, DNA+CAP with AzoPhe incorporated at residue 71 (before irradiation at 334 nm). Substitution of Ile71 with trans-AzoPhe in CAP results in a fourfold decrease of the binding constant Kb of the CAP for its promoter sequence. Photoirradiation at 334 nm partially converts the trans-AzoPhe to the cis-form and decreases the Kb by another fourfold. The latter affinity loss can be completely recovered after irradiation at > 420 nm, which switches the cis-form back to the predominant trans-state.

1.9 REFERENCES

1. Eisenhauer, B. M.; Hecht, S. M., Site-specific incorporation of (aminooxy)acetic acid into proteins. *Biochemistry* **2002,** *41* (38), 11472-8.

2. Ellman, J. A.; Mendel, D.; Schultz, P. G., Site-specific incorporation of novel backbone structures into proteins. *Science* **1992,** *255* (5041), 197-200.

3. Hartman, M. C.; Josephson, K.; Szostak, J. W., Enzymatic aminoacylation of tRNA with unnatural amino acids. *Proc Natl Acad Sci U S A* **2006,** *103* (12), 4356-61.

4. Valiyaveetil, F. I.; Sekedat, M.; Mackinnon, R.; Muir, T. W., Glycine as a D-amino acid surrogate in the K(+)-selectivity filter. *Proc Natl Acad Sci U S A* **2004,** *101* (49), 17045-9.

5. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angew Chem Int Ed Engl* **2004,** *44* (1), 34-66.

6. Means, G. E.; Feeney, R. E., Chemical modifications of proteins: history and applications. *Bioconjug Chem* **1990,** *1* (1), 2-12.

7. Geoghegan, K. F.; Stroh, J. G., Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine. *Bioconjug Chem* **1992,** *3* (2), 138-46.

8. Antos, J. M.; Francis, M. B., Selective tryptophan modification with rhodium carbenoids in aqueous solution. *J Am Chem Soc* **2004,** *126* (33), 10256-7.

9. Tilley, S. D.; Francis, M. B., Tyrosine-selective protein alkylation using pi-allylpalladium complexes. *J Am Chem Soc* **2006,** *128* (4), 1080-1.

10. Kent, S. B., Chemical synthesis of peptides and proteins. *Annu Rev Biochem* **1988,** *57*, 957-89.

11. Wallace, C. J., Peptide ligation and semisynthesis. *Curr Opin Biotechnol* **1995,** *6* (4), 403-10.

12. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9.

13. Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci U S A* **1998,** *95* (12), 6705-10.

14. Giriat, I.; Muir, T. W., Protein semi-synthesis in living cells. *J Am Chem Soc* **2003,** *125* (24), 7180-1.

15. Noren, C. J.; Anthony-Cahill, S. J.; Griffith, M. C.; Schultz, P. G., A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989,** *244* (4901), 182-8.

16. Cornish, V. W.; Mendel, D.; Schultz, P. G., Probing protein structure and function with an expanded genetic code. *Angew Chem Int Ed Engl* **1995,** *34*, 621-633.

17. Hohsaka, T.; Ashizuka, Y.; Taira, H.; Murakami, H.; Sisido, M., Incorporation of nonnatural amino acids into proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry* **2001,** *40* (37), 11060-4.

18. Nowak, M. W.; Kearney, P. C.; Sampson, J. R.; Saks, M. E.; Labarca, C. G.; Silverman, S. K.; Zhong, W.; Thorson, J.; Abelson, J. N.; Davidson, N.; et al., Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. *Science* **1995,** *268* (5209), 439-42.

19. Dougherty, D. A., Unnatural amino acids as probes of protein structure and function. *Curr Opin Chem Biol* **2000,** *4* (6), 645-52.

20. Forster, A. C.; Tan, Z.; Nalam, M. N.; Lin, H.; Qu, H.; Cornish, V. W.; Blacklow, S. C., Programming peptidomimetic syntheses by translating genetic codes designed de novo. *Proc Natl Acad Sci U S A* **2003,** *100* (11), 6353-7.

21. Budisa, N., Prolegomena to future experimental efforts on genetic code engineering by expanding its amino acid repertoire. *Angew Chem Int Ed Engl* **2004,** *43* (47), 6426-63.

22. Suchanek, M.; Radzikowska, A.; Thiele, C., Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. *Nat Methods* **2005,** *2* (4), 261-7.

23. Tang, Y.; Tirrell, D. A., Biosynthesis of a highly stable coiled-coil protein containing hexafluoroleucine in an engineered bacterial host. *J Am Chem Soc* **2001,** *123* (44), 11089-90.

24. Ibba, M.; Kast, P.; Hennecke, H., Substrate Specificity Is Determined by Amino Acid Binding Pocket Size in Escherichia coli …. *Biochemistry* **1994**.

25. Doring, V.; Mootz, H. D.; Nangle, L. A.; Hendrickson, T. L.; de Crecy-Lagard, V.; Schimmel, P.; Marliere, P., Enlarging the amino acid set of Escherichia coli by infiltration of the valine coding pathway. *Science* **2001,** *292* (5516), 501-4.

26. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.

27. Thorson, J. S.; Chapman, E.; Murphy, E. C.; Schultz, P. G.; Judice, J. K., Linear free energy analysis of hydrogen bonding in proteins. *J Am Chem Soc* **1995,** *117*, 1157-1158.

28. Koh, J. T.; Cornish, V. W.; Schultz, P. G., An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. *Biochemistry* **1997,** *36* (38), 11314-22.

29. Chapman, E.; Thorson, J.; Schultz, P. G., Mutational analysis if backbone hydrogen bonds in staphylococcal nuclease. *J Am Chem Soc* **1997,** *119*, 7151-7152.

30. Mendel, D.; Ellman, J. A.; Chang, Z.; Veenstra, D. L.; Kollman, P. A.; Schultz, P. G., Probing protein stability with unnatural amino acids. *Science* **1992,** *256* (5065), 1798-802.

31. Bilgicer, B.; Xing, X.; Kumar, K., Programmed self-sorting of coiled coils with leucine and hexafluoroleucine cores. *J Am Chem Soc* **2001,** *123* (47), 11815-6.

32. Lee, H. Y.; Lee, K. H.; Al-Hashimi, H. M.; Marsh, E. N., Modulating protein structure with fluorous amino acids: increased stability and native-like structure conferred on a 4-helix bundle protein by hexafluoroleucine. *J Am Chem Soc* **2006,** *128* (1), 337-43.

33. Judice, J. K.; Gamble, T. R.; Murphy, E. C.; de Vos, A. M.; Schultz, P. G., Probing the mechanism of staphylococcal nuclease with unnatural amino acids: kinetic and structural studies. *Science* **1993,** *261* (5128), 1578-81.

34. England, P. M.; Zhang, Y.; Dougherty, D. A.; Lester, H. A., Backbone mutations in transmembrane domains of a ligand-gated ion channel: implications for the mechanism of gating. *Cell* **1999,** *96* (1), 89-98.

35. Dang, H.; England, P. M.; Farivar, S. S.; Dougherty, D. A.; Lester, H. A., Probing the role of a conserved M1 proline residue in 5-hydroxytryptamine(3) receptor gating. *Mol Pharmacol* **2000,** *57* (6), 1114-22.

36. Lummis, S. C.; Beene, D. L.; Lee, L. W.; Lester, H. A.; Broadhurst, R. W.; Dougherty, D. A., Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **2005,** *438* (7065), 248-52.

37. Ting, A. Y.; Shin, I.; Lucero, C.; Schultz, P. G., Energetic analysis of an engineered cation-pi interaction in staphylococcal nuclease. *J Am Chem Soc* **1998,** *120*, 7135-7136.

38. Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. A.; Dougherty, D. A., From ab initio quantum mechanics to molecular neurobiology: a cation-pi binding site in the nicotinic receptor. *Proc Natl Acad Sci U S A* **1998,** *95* (21), 12088-93.

39. Lummis, S. C.; D, L. B.; Harrison, N. J.; Lester, H. A.; Dougherty, D. A., A cation-pi binding interaction with a tyrosine in the binding site of the GABAC receptor. *Chem Biol* **2005,** *12* (9), 993-7.

40. Sagle, L. B.; Zimmermann, J.; Matsuda, S.; Dawson, P. E.; Romesberg, F. E., Redox-coupled dynamics and folding in cytochrome c. *J Am Chem Soc* **2006,** *128* (24), 7909-15.

41. Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G., A genetically encoded infrared probe. *J Am Chem Soc* **2006,** *128* (43), 13984-5.

42. Wang, J.; Xie, J.; Schultz, P. G., A genetically encoded fluorescent amino acid. *J Am Chem Soc* **2006,** *128* (27), 8738-9.

43. Wang, L.; Xie, J.; Deniz, A. A.; Schultz, P. G., Unnatural amino acid mutagenesis of green fluorescent protein. *J Org Chem* **2003,** *68* (1), 174-6.

44. Bae, J. H.; Rubini, M.; Jung, G.; Wiegand, G.; Seifert, M. H.; Azim, M. K.; Kim, J. S.; Zumbusch, A.; Holak, T. A.; Moroder, L.; Huber, R.; Budisa, N., Expansion of the genetic code enables design of a novel "gold" class of green fluorescent proteins. *J Mol Biol* **2003,** *328* (5), 1071-81.

45. Tsien, R. Y., The green fluorescent protein. *Annu Rev Biochem* **1998,** *67*, 509-44.

46. Chattoraj, M.; King, B. A.; Bublitz, G. U.; Boxer, S. G., Ultra-fast excited state dynamics in green fluorescent protein: multiple states and proton transfer. *Proc Natl Acad Sci U S A* **1996,** *93* (16), 8362-7.

47. Maiti, A.; Roy, S., Switching DNA-binding specificity by unnatural amino acid substitution. *Nucleic Acids Res* **2005,** *33* (18), 5896-903.

48. Schumacher, T. N.; Mayr, L. M.; Minor, D. L., Jr.; Milhollen, M. A.; Burgess, M. W.; Kim, P. S., Identification of D-peptide ligands through mirror-image phage display. *Science* **1996,** *271* (5257), 1854-7.

49. Eckert, D. M.; Malashkevich, V. N.; Hong, L. H.; Carr, P. A.; Kim, P. S., Inhibiting HIV-1 entry: discovery of D-peptide inhibitors that target the gp41 coiled-coil pocket. *Cell* **1999,** *99* (1), 103-15.

50. Xie, J.; Schultz, P. G., Adding amino acids to the genetic repertoire. *Curr Opin Chem Biol* **2005,** *9* (6), 548-54.

51. Cook, S. N.; Jack, W. E.; Xiong, X.; Danley, L. E.; Ellman, J. A.; Schultz, P. G.; Noren, C. J., Photochemically initiated protein splicing. *Angew Chem Int Ed Engl* **1995,** *34*, 1629-1630.

52. Miller, J. C.; Silverman, S. K.; England, P. M.; Dougherty, D. A.; Lester, H. A., Flash decaging of tyrosine sidechains in an ion channel. *Neuron* **1998,** *20* (4), 619-24.

53. Deiters, A.; Groff, D.; Ryu, Y.; Xie, J.; Schultz, P. G., A genetically encoded photocaged tyrosine. *Angew Chem Int Ed Engl* **2006,** *45* (17), 2728-31.

54. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., A genetically encoded photocaged amino acid. *J Am Chem Soc* **2004,** *126* (44), 14306-7.

55. England, P. M.; Lester, H. A.; Davidson, N.; Dougherty, D. A., Site-specific, photochemical proteolysis applied to ion channels in vivo. *Proc Natl Acad Sci U S A* **1997,** *94* (20), 11025-30.

56. Banghart, M.; Borges, K.; Isacoff, E.; Trauner, D.; Kramer, R. H., Light-activated ion channels for remote control of neuronal firing. *Nat Neurosci* **2004,** *7* (12), 1381-6.

57. Bose, M.; Groff, D.; Xie, J.; Brustad, E.; Schultz, P. G., The incorporation of a photoisomerizable amino acid into proteins in E. coli. *J Am Chem Soc* **2006,** *128* (2), 388-9.

58. Gandhi, C. S.; Isacoff, E. Y., Shedding light on membrane proteins. *Trends Neurosci* **2005,** *28* (9), 472-9.

Chapter 2

Genetically encoding unnatural amino acids for cellular and neuronal studies

2.1 Abstract

Proteins participate in various biological processes and can be harnessed to probe and control biological events selectively and reproducibly, but the genetic code limits the building block to 20 common amino acids for protein manipulation in living cells. The genetic encoding of unnatural amino acids will remove this restriction and enable new chemical and physical properties to be precisely introduced into proteins. Here we present new strategies for generating orthogonal tRNA-synthetase pairs, which made possible the genetic encoding of diverse unnatural amino acids in different mammalian cells and primary neurons. Using this new methodology, we incorporated unnatural amino acids with extended side chains into the K^+ channel Kv1.4, and found that the bulkiness of residues in the inactivation peptide is essential for fast channel inactivation, a finding that had not been possible using conventional mutagenesis. This technique will stimulate and facilitate new molecular studies using tailored unnatural amino acids for cell biology and neurobiology.

2.2 Introduction

The use of proteins as experimental tools has led to significant advances in neurobiology. These proteins can be selectively targeted to cells of interest or engineered, enabling them to make precise measurements or to perturb signaling[1,2,-]. For instance, some genetically encoded biosensors have provided new ways to 'see' the activity of ion channels[3]. Proteins derived from green fluorescent protein (GFP), such as cameleon[4] and synapto-pHluorin[5], can report on changes in calcium flux or neurotransmitter release. In addition, channelrhodopsin-2, a light-gated cation channel, allows action potentials to be noninvasively evoked using light, facilitating the study and modulation of behavior in intact animals[6]. However, a protein-based sensor that can report a single action potential is not available, and there are no ideal protein tools for the *in vivo* study of other neurobiological events such as protein misfolding, aggregation, trafficking and interactions. In addition to known proteins with unique properties, further advances would be expected if it were possible to make changes in neuronal proteins that generate new properties. Currently, we are limited to the naturally occurring 20 common amino acids as building blocks.

One solution to this limitation is to use unnatural amino acids that possess unique side chains. Incorporation of unnatural amino acids has provided a powerful method for studying the physical, chemical and biological properties of proteins[7]. Unnatural amino acids have been incorporated into ion channels and receptors by microinjecting the chemically acylated tRNA and UAG-containing mutant mRNA into *Xenopus* oocytes[8], and have greatly contributed to our understanding of receptor functions[9-11]. Unfortunately, application of this method in neurobiology is restricted. The requirement for microinjection limits the technique mainly to large *Xenopus* oocytes, and it is not suitable for studies that require large numbers of cells. A more significant obstacle is that the tRNA is chemically acylated with the unnatural amino acid *in vitro*, and the acylated tRNA is consumed as a stoichiometric reagent during translation and cannot be regenerated. Therefore, yields of mutant proteins are low and data cannot be collected over extended periods.

The genetic encoding of unnatural amino acids would overcome such restrictions and enable proteins and related biological events to be investigated directly in living cells and organisms. Previously, bacterial *Escherichia coli* have been used to expand the genetic code to include unnatural amino acids[12]. This method involves the generation of a new tRNA-aminoacyl–tRNA-synthetase pair that is specific for the unnatural amino acid and decodes a blank codon that is unused by a common amino acid (such as a stop codon or an extended codon with four or more bases). The tRNA-synthetase pair needs to be compatible with the protein biosynthesis machinery of the host cell, and crosstalk with endogenous tRNA-synthetase pairs must be avoided. Such orthogonal tRNA-synthetase pairs for use in *E. coli* have been evolved from large RNA and protein libraries ($>10^9$ members) through high-throughput selection[7]. However, mammalian cells and *E. coli* differ significantly in tRNA transcription, processing and transportation, leading to inefficient biosynthesis of orthogonal tRNAs in mammalian cells. Moreover, primary neurons have low transfection efficiency and cannot divide. It will therefore be impractical, if not impossible, to devise and implement the high-throughput selections necessary for engineering orthogonal tRNAs and unnatural-amino-acid-specific synthetases in neurons.

Here we report new strategies that enable unnatural amino acids to be efficiently encoded in various types of mammalian cell, including primary neurons. We used a type-3 pol III promoter to

efficiently express and process bacterial tRNAs, and unnatural-amino-acid-specific synthetases generated in yeast were successfully transferred for use in mammalian cells and neurons. To show that unnatural amino acids can alter the function of a neuronal protein, we chose to study the fast, N-type inactivation of voltage-gated K^+ channels[13]. Upon membrane depolarization, such as during an action potential, some voltage-gated K^+ channels open and then quickly enter a non-conducting state; this process is known as inactivation. Structural studies[14] and mutagenesis studies[15-17] have suggested that fast inactivation occurs when the amino-terminal domain of the channel feeds through a small portal in the cytoplasmic domain and binds to the inner channel cavity, preventing the efflux of K^+ ions. Here, we used unnatural amino-acid mutagenesis to investigate whether the bulkiness of an amino acid in the N-terminal inactivation peptide affects the fast inactivation of the neuronal K^+ channel Kv1.4. Our data support a model in which the overall diameter of the inactivation peptide is crucial for determining the rate of channel inactivation. We discuss these results as well as the potential impact of unnatural amino-acid mutagenesis in neurons on neurobiology.

2.3 Results

2.3.1 Efficient expression of orthogonal tRNA in mammalian cells

An efficient way of generating an orthogonal tRNA-synthetase pair is to import a tRNA-synthetase pair from species in a different kingdom[18,19], because the cross-aminoacylation between different species is often low. The $tRNA^{Tyr}$-TyrRS pair from the archaebacterium *Methanococcus jannaschii* was successfully used in *E. coli*. Similarly, the *E. coli* $tRNA^{Tyr}$-TyrRS pair has been used in yeast, and should be orthogonal in mammalian cells. To test and use *E. coli* tRNAs, one challenge lies in the expression of functional *E. coli* tRNAs in mammalian cells. *E. coli* and mammalian cells differ significantly in tRNA transcription and processing. *E. coli* tRNAs are transcribed by the sole RNA polymerase through promoters upstream of the tRNA structural gene. However, the transcription of mammalian tRNA genes depends principally on promoter elements within the tRNA known as the A and B box sequences, which are recognized by RNA polymerase III (pol III) and its associated factors[20]. Whereas all *E. coli* tRNA genes encode full tRNA sequences, mammalian tRNAs have the 3'-CCA

sequence added enzymatically by the tRNA nucleotidyltransferase after transcription. In addition, the 5'- and 3'-flanking sequences, the removal of introns, and the export from nucleus to cytoplasm also affect mammalian tRNA expression and function. Owing to these differences, *E. coli* tRNAs, especially those that diverge from the preserved eukaryotic A and B box sequences, are not efficiently biosynthesized or correctly processed in mammalian cells.

We reasoned that a pol III promoter lacking any requirement for intragenic elements could efficiently transcribe tRNAs without the preserved internal A and B boxes in mammalian cells. The H1 promoter[21] was chosen for the following reasons: (i) it drives the expression of human *H1RNA* and thus is of mammalian origin; (ii) it is a type-3 pol III promoter that has no downstream transcriptional elements; (iii) the transcription initiation site of the H1 promoter is well defined, and can be used to generate the 5' end of the tRNA without further post-transcriptional processing; and (iv) the H1 promoter has been successfully used to express short interfering RNAs in mammalian cells.

We developed a fluorescence-based functional assay in mammalian cells to identify those expression elements that can efficiently drive the transcription of *E. coli* tRNAs to generate functional tRNAs in mammalian cells (**Figure 2.1a**). The gene for the candidate *E. coli* amber suppressor tRNA (EctRNA$^{aa}_{CUA}$, whose anticodon is changed to CUA to decode the amber stop codon TAG) is co-expressed with its cognate synthetase (aaRS). A TAG stop codon is introduced at a permissive site of the *GFP* gene, and this mutant*GFP* gene is co-expressed with the EctRNA$^{aa}_{CUA}$-aaRS pair in mammalian cells. If the EctRNA$^{aa}_{CUA}$ is expressed and correctly processed to a functional tRNA, the synthetase will aminoacylate this tRNA with the cognate amino acid. The acylated EctRNA$^{aa}_{CUA}$will suppress the TAG codon in the *GFP* gene, producing full-length GFP and rendering cells fluorescent. By comparing the fluorescence intensities of cells, this method can also serve as a sensitive *in vivo* assay for the orthogonality of the EctRNA$^{aa}_{CUA}$ to endogenous synthetases of host cells when the cognate *E. coli* synthetase is not expressed, and for the activity of the orthogonal EctRNA$^{aa}_{CUA}$ towards an unnatural-amino-acid-specific mutant synthetase when the mutant synthetase is expressed in place of the cognate synthetase.

The *E. coli* tyrosyl amber suppressor tRNA ($EctRNA^{Tyr}_{CUA}$) was chosen as the candidate orthogonal tRNA because it is orthogonal to yeast synthetases and suppresses the amber stop codon efficiently in yeast when coexpressed with *E. coli* TyrRS[22]. *In vitro* aminoacylation assays indicate that *E. coli* TyrRS does not charge eukaryotic tRNAs[23]. Mammalian and yeast cells are eukaryotic, so we reasoned that the $EctRNA^{Tyr}_{CUA}$ –*E. coli* TyrRS pair should be orthogonal in mammalian cells. For 3'-end processing of the $EctRNA^{Tyr}_{CUA}$, we used the 3'-flanking sequence of the human $tRNA^{fMet}$. The 5'- and 3'-flanking sequences of the human $tRNA^{fMet}$ can drive the functional expression of *E. coli* $tRNA^{Gln}_{CUA}$ (which has the A box and B box) in mammalian cells[24]. To test the importance of the 3'-CCA trinucleotide, these nucleotides were included or removed in the tRNA gene, resulting in four expression cassettes (tRNA-1 to tRNA-4; **Figure 2.1b**). For comparison, we made a control plasmid (tRNA-5) in which the $EctRNA^{Tyr}_{CUA}$ was placed downstream of the 5'-flanking sequence of the human $tRNA^{Tyr}$.

To compare the ability of different expression cassettes to generate functional tRNAs, we established a clonal stable HeLa cell line expressing the *GFP* gene with a TAG stop codon introduced at the permissive site 182 (GFP-TAG HeLa). The tRNA-aaRS expression plasmid was transfected into the stable GFP-TAG HeLa cell line, and cells were analyzed with flow cytometry after 48 h. The total fluorescence intensity of the green fluorescent cells indicates the amount of GFP produced (**Figure 2.1c**). When $EctRNA^{Tyr}_{CUA}$-TyrRS was not expressed, the fluorescence intensity of the GFP-TAG HeLa cell line was similar to that of HeLa cells, indicating that the background readthrough of the TAG codon in GFP is negligible. Using the 5'-flanking sequence of human $tRNA^{Tyr}$ in tRNA-5, only weak amber suppression was detected, confirming that bacterial tRNAs without the preserved A and B boxes could not be functionally expressed in mammalian cells. The highest fluorescence intensity was found in cells transfected with tRNA-4, which was 71-fold higher than that of tRNA-5, indicating that the H1 promoter can drive the functional biosynthesis of $EctRNA^{Tyr}_{CUA}$ much more efficiently than the 5'-flanking sequence of the human $tRNA^{Tyr}$. This also indicates that the H1 promoter can generate the correct 5'-end of the tRNA directly from the transcription initiation site without the post-transcriptional processing that is necessary for endogenously expressed tRNAs. The fluorescence intensity of cells

transfected with tRNA-2 was 10% of that of cells transfected with tRNA-4, indicating that the 3'-flanking sequence of the human $tRNA^{fMet}$ is also needed for the efficient expression of the $EctRNA^{Tyr}_{CUA}$. It is intriguing to find that functional tRNA was produced in mammalian cells transfected with tRNA-1 (21% of tRNA-4), in which the CCA trinucleotide but no 3'-flanking sequence is included, as mammalian cells do not encode the CCA in the tRNA gene. However, when both the CCA trinucleotide and the 3'-flanking sequence were included in tRNA-3, the fluorescence intensity dropped markedly to 1.3%.

We used Northern blotting to measure the transcription levels of the $EctRNA^{Tyr}_{CUA}$expressed by different constructs in GFP-TAG HeLa cells (**Figure 2.1d**). Very low levels of $EctRNA^{Tyr}_{CUA}$ could be detected using an $EctRNA^{Tyr}_{CUA}$-specific probe in samples transfected with tRNA-5, tRNA-3 or tRNA-2. By contrast, in cells transfected with tRNA-4 and tRNA-1, the amounts of $EctRNA^{Tyr}_{CUA}$ were about 93-fold and 19-fold higher, respectively, than in cells transfected with tRNA-5. The northern blotting data confirmed that the $EctRNA^{Tyr}_{CUA}$ was transcribed in HeLa cells, and the increase in tRNA transcription is consistent with the increase in fluorescence intensity measured by cytometry in different samples.

To test the orthogonality of the $EctRNA^{Tyr}_{CUA}$ to endogenous synthetases in HeLa cells, we removed the *E. coli* TyrRS in tRNA-4 to express the $EctRNA^{Tyr}_{CUA}$ only. Transfection of the resultant plasmid into the GFP-TAG HeLa cell line did not change the fluorescence intensity of the cells, showing that the $EctRNA^{Tyr}_{CUA}$ is not aminoacylated by any synthetases in HeLa cells.

To test whether the H1 promoter, together with the 3'-flanking sequence, can be used to express other *E. coli* tRNAs, we replaced the $EctRNA^{Tyr}_{CUA}$ in construct tRNA-4 with the *E. coli* leucyl amber suppressor tRNA ($EctRNA^{Leu}_{CUA}$) and the TyrRS with the cognate leucyl-tRNA synthetase (LeuRS). When only the $EctRNA^{Leu}_{CUA}$ was expressed, we saw no fluorescence change in the GFP-TAG HeLa cells, indicating that the $EctRNA^{Leu}_{CUA}$ is also orthogonal in HeLa cells. By contrast, when the $EctRNA^{Leu}_{CUA}$-LeuRS pair was expressed, the GFP-TAG HeLa cells became very bright. The total fluorescence intensity was 104% of that of cells transfected with the $EctRNA^{Tyr}_{CUA}$-TyrRS pair. The $EctRNA^{Leu}_{CUA}$ does not have a fully matched A box, whereas the $EctRNA^{Leu}_{CUA}$ has no fully matched A

or B box sequences. Taken together, these results show that the H1 promoter can efficiently drive the expression of *E. coli* tRNAs, regardless of the internal promoter elements, in mammalian cells, and that the transcribed tRNAs are functional for amber suppression.

2.3.2 Unnatural-amino-acid-specific aaRS for mammalian cells

Synthetases that are specific for a variety of unnatural amino acids have been evolved in *E. coli* and later in yeast from large mutant synthetase libraries consisting of $>10^9$members[7]. Similar strategies cannot be practically employed in mammalian cells and neurons because the transfection efficiencies of these cells are lower by several orders of magnitude than those of *E. coli* and yeast. As both yeast and mammalian cells are of eukaryotic origin, we decided to transfer the mutant synthetases that had been evolved in yeast to mammalian cells. To test the feasibility of this approach, the *E. coli*TyrRS gene in the tRNA-aaRS expression plasmid (**Figure 2.1a**) was replaced with the gene for OmeTyrRS, a synthetase specific for the unnatural amino acid *o*-methyl-L-tyrosine (OmeTyr; **Figure 2.2a**). The resultant plasmid was transfected into the GFP-TAG HeLa cell line, and cells were grown in the presence and absence of OmeTyr. As shown in **Figure 2.2b**, in the absence of OmeTyr, these cells were virtually nonfluorescent and similar to the GFP-TAG HeLa cells, indicating that the expression of the EctRNA$^{Tyr}_{CUA}$-OmeTyrRS pair does not suppress amber codons efficiently. When OmeTyr was added, 71% of cells (normalized to the total number of fluorescent cells transfected with the EctRNA$^{Tyr}_{CUA}$ and wild-type TyrRS) became fluorescent, indicating that OmeTyr was incorporated into the GFP. The incorporation efficiency is about 41% by comparing the total fluorescence intensity of these cells to the intensity of cells transfected with the EctRNA$^{Tyr}_{CUA}$-TyrRS pair.

To determine whether the transfer strategy could be generally applied to other synthetases evolved in yeast, we next tested the BpaRS, a synthetase that is specific for *p*-benzoylphenylalanine (Bpa; (**Figure 2.2a**). As expected, when the BpaRS was coexpressed with the EctRNA$^{Tyr}_{CUA}$ in the GFP-TAG HeLa cell line, 47% of cells were fluorescent in the presence of Bpa, and virtually no fluorescent cells (≤4%) were detected in the absence of Bpa. The incorporation efficiency of this unnatural amino acid is about 13% (**Figure 2.2b**).

In addition to tRNA-aaRS pairs derived from the *E. coli* $tRNA^{Tyr}$-TyrRS, we also tested a tRNA-aaRS pair derived from *E. coli* $tRNA^{Leu}$-LeuRS. The $EctRNA^{Leu}_{CUA}$ and a mutant synthetase specific for a fluorescent unnatural amino acid 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamido)propanoic acid (DanAla)[25] (**Figure 2.2a**) were expressed in the GFP-TAG HeLa cell line. DanAla was incorporated with 13% efficiency, and 42% of the cells became fluorescent (**Figure 2.2c**).

These results confirm that unnatural-amino-acid-specific synthetases evolved in yeast can generally be transferred for use in mammalian cells. The tested unnatural amino acids were incorporated less efficiently than Tyr by the wild-type TyrRS, indicating that the activities of the evolved synthetases are not optimal. More rounds of directed evolution of these synthetases in yeast might further improve their incorporation efficiencies.

2.3.3 Genetically encoding unnatural amino acids in neurons

We first investigated whether the H1 promoter and the 3'-flanking sequence identified in HeLa cells could also generate functional amber suppressor tRNAs in neurons. Mouse hippocampal neurons were transfected with two plasmids simultaneously (**Figure 2.3a**): the reporter plasmid pCLHF-GFP-TAG encoding a mutant GFP (182TAG), and the expression plasmid encoding the *E. coli* TyrRS, the $EctRNA^{Tyr}_{CUA}$ driven by either the H1 promoter or the 5' flanking sequence of human $tRNA^{Tyr}$, and a red fluorescent protein, mCherry, as an internal marker for transfection. Fluorescence microscopy was used to look for red transfected cells, and then to image their green fluorescence. The presence of green fluorescence in transfected cells indicates that functional $EctRNA^{Tyr}_{CUA}$ is biosynthesized to incorporate Tyr at the 182TAG position of the *GFP*gene. As shown in **Figure 2.3b**, neurons transfected with the expression plasmid in which the $EctRNA^{Tyr}_{CUA}$ was driven by the H1 promoter showed intense green fluorescence, whereas no green fluorescence could be detected in neurons in which the $EctRNA^{Tyr}_{CUA}$ was driven by the 5'-flanking sequence of the human $tRNA^{Tyr}$.

Next we tested whether unnatural amino acids could be genetically encoded in neurons using the $EctRNA^{Tyr}_{CUA}$ and mutant synthetases that are specific for different unnatural amino acids. Synthetases that had been evolved in yeast and proven to be functional in HeLa cells were used. When

the OmeTyrRS was coexpressed with the EctRNA$^{Tyr}_{CUA}$, transfected neurons showed no green fluorescence in the absence of the corresponding unnatural amino acid OmeTyr (**Figure 2.3c**), indicating that the EctRNA$^{Tyr}_{CUA}$ is orthogonal to endogenous synthetases in neurons. Transfected neurons showed bright green fluorescence only when OmeTyr was fed to the growth medium. These results indicate that OmeTyr, but no common amino acid, is incorporated into GFP at the 182TAG position. The same results were obtained for the unnatural amino acid Bpa when the BpaRS was coexpressed with the EctRNA$^{Tyr}_{CUA}$ (**Figure 2.3d**). Using this approach, OmeTyr and Bpa were also genetically encoded in hippocampal and cortical neurons isolated from rats (data not shown).

2.3.4 Probing the inactivation mechanism of K^+ channel Kv1.4

The fast inactivation of voltage-dependent K^+ (Kv) channels involves ~20 amino acids on the channel's intracellular N terminus, or on the N terminus of an associated β-unit. The first ~10 amino acids are predominantly hydrophobic, and the other ~10 amino acids are hydrophilic and positively charged[26]. Recent crystal structure and mutation studies of Kv1.2 suggest that the inactivation peptide might snake through a side portal in the fully extended conformation, and allow the hydrophobic terminus to reach and plug the inner pore for inactivation[14-17] (**Figure 2.4a**). Through unnatural amino-acid mutagenesis of the inactivation peptide of Kv1.4 in live mammalian cells, we aimed to experimentally test this inactivation model by determining whether the bulkiness of the inactivation peptide would influence channel inactivation.

The N-terminal domain of Kv1.4 possesses two inactivation domains, consisting of amino acids Met1–Ala40 for the primary and Ala40–Gly60 for the secondary inactivation domains[27]. The structure of the primary inactivation domain in aqueous solution has been analyzed using NMR spectroscopy[28,29]. The N-terminal Met1–Met17 is disordered, followed by a β-turn (Pro18–Tyr21) and a well-defined α-helix (Tyr21–Ala36) (**Figure 2.4b**). In the Kv1.2 structure, a negatively charged patch of residues lies at the entrance to the side portal, and is proposed to interact with the positively charged amino acids on the inactivation peptide[14]. According to this hypothesis, several arginines (Arg26, Arg28, Arg30 and Arg32) in the α-helix of Kv1.4 will interact with the negatively charged residues at

the side portal entrance, which will place Tyr19 inside the side portal (**Figure 2.4a**). In addition, Tyr19 resides at the beginning of the α-helical portion of the inactivation domain (**Figure 2.4b**). We hypothesized that increasing the bulkiness of the amino-acid side chain at Tyr19 would alter the channel inactivation owing to the size restriction of the side portal.

We first investigated the inactivation properties of wild-type Kv1.4 channels expressed heterologously in human embryonic kidney 293 (HEK293) cells using whole-cell patch-clamp recordings (**Figure 2.5a**). A series of 1-s voltage steps from –100 mV to +40 mV elicited a rapidly inactivating macroscopic K^+ current. To avoid undesired transcriptional initiation at the downstream Met109 site and thus the generation of truncated Kv1.4, the methionine at 109 was mutated to leucine (M109L; see Methods). Kv1.4-M109L (Kv1.4*) channels displayed fast inactivation similar to wild-type Kv1.4 channels (**Figure 2.5a**). To investigate the role of bulkiness in determining fast inactivation, we first mutated Tyr19 of Kv1.4* to Phe or Trp, using conventional site-directed mutagenesis. Compared to the native Tyr, Phe is smaller (lacking the hydroxyl) whereas Trp is considerably larger. The macroscopic K^+ currents recorded from HEK293 cells expressing Kv1.4*-Y19F or Kv1.4*-Y19W channels were similar to those of Kv1.4* channels (**Figure 2.5a**). Therefore, the Y19F or Y19W mutation did not greatly alter the inactivation properties.

Next, we tested whether a functional Kv1.4 channel could be expressed in HEK293 cells through amber suppression using the orthogonal EctRNA$^{Tyr}_{CUA}$-TyrRS pair to incorporate the native Tyr. An amber stop codon (TAG) was introduced at Tyr19, and the mutant gene (*Kv1.4*-19TAG*) was cotransfected into HEK293 cells with the EctRNA$^{Tyr}_{CUA}$-TyrRS pair and the *GFP-TAG* reporter gene (**Figure 2.5b**). Transfected cells fluoresced green owing to the suppression of the amber codon in the *GFP-TAG* gene by the EctRNA$^{Tyr}_{CUA}$-TyrRS pair. GFP-positive cells also showed large voltage-dependent K^+ currents that inactivated with large depolarizations (**Figure 2.5c,d**). To investigate possible differences in the rates of fast inactivation, we fit the decay of current elicited by the 1-s voltage-step to +20 mV with a sum of two exponentials relaxing to a small steady-state current (*C*). Neither the τ_1 or τ_2 time constants nor the amplitudes of current were significantly different between Kv1.4* and Kv1.4*-19Tyr channels (**Figure 2.5e,f**; **Table 2.1**), indicating that the incorporation of

amino acids through amber suppression did not adversely affect the cell or the properties of the Kv1.4 channel. Importantly, no current could be detected in the absence of the EctRNA$^{Tyr}_{CUA}$-TyrRS pair, confirming that functional Kv1.4 is produced through amber suppression and that the tRNA-synthetase pair is required for the incorporation of Tyr.

We then tested the effect of introducing a large bulky unnatural amino acid at position 19. OmeTyr was incorporated at the 19TAG position of Kv1.4* using the EctRNA$^{Tyr}_{CUA}$-OmeTyrRS pair. When no OmeTyr was added to the growth medium, only small background levels of current could be detected in transfected cells (identified by cotransfected enhanced GFP in this case). By contrast, large voltage-gated K^+currents were seen when OmeTyr was added to the culture medium (**Figure 2.5c,d**). Interestingly, the rate and extent of inactivation were markedly slower than with the wild-type channel — note the large amount of current at the end of the 1-s voltage step (**Figure 2.5c**). To quantify this change in inactivation more accurately, we measured the macroscopic current elicited by a 4-s voltage step to +20 mV and fit the decay of current with the sum of two exponentials, as described above (**Figure 2.5e**). The time constants of the fast (τ_1) and the slow inactivation component (τ_2) for the OmeTyr mutant increased 5- and 7-fold, respectively, compared to those of the wild-type Kv1.4* channel. In addition, the amplitude of the fast component (*A*1) decreased to ~20%, whereas the amplitude of the slow component (*A*2) increased by ~1.5-fold for the OmeTyr mutant (**Figure 2.5f**, **Table 2.1**). We then tested the effect of incorporating an amino acid with an even larger side chain, DanAla, at Tyr19 (**Figure 2.2a**). As when OmeTyr was inserted, slowly inactivating, voltage-gated K^+ currents were seen in HEK293 cells transfected with Kv1.4*-19TAG, and the EctRNA$^{Leu}_{CUA}$-DanAlaRS pair, along with DanAla in the medium (**Figure 2.5c,d**). Quantification of the rate of inactivation showed that DanAla increased the fast and slow time constants for inactivation by 3.2-fold and 7-fold, respectively (**Figure 2.5a**, **Table 2.1**). Similarly, the amplitude of the slow component was larger than that of the fast component of inactivation, in contrast to the wild-type channel. Thus, both OmeTyr and DanAla can be functionally incorporated into Kv1.4 channels, leading to markedly slower inactivation.

We also carried out several important controls. First, only small background currents were evident in HEK cells when the unnatural amino acid was omitted from the medium (**Figure 2.5**). Thus, there was no inappropriate 'read-through' of the amber mutation. To exclude the possibility that OmeTyr is toxic to cells or to the Kv channel, OmeTyr was also incorporated into Kv1.4* at 557TAG. Residue 557 is on the transmembrane helix S6 and lies at the inner pore, which is between the selectivity filter and intracellular solution. Like Kv1.4*, the inactivation rate of Kv1.4*-557OmeTyr was similar to that of wild-type Kv1.4 (**Figure 2.5c,f**, **Table 2.1**). These experiments show that the abolition of fast inactivation of Kv1.4 by OmeTyr is site-dependent. To address whether the introduction of OmeTyr or DanAla rendered the N-terminal inactivation peptide subject to proteolytic cleavage, we added a FLAG tag to the C terminus of Kv1.4*-19TAG, and used western blotting to analyze channel proteins containing OmeTyr or DanAla at residue 19 expressed in HEK293 cells (**Figure 2.5g**). In both samples, one band corresponding to the full-length Kv1.4* channel was detected using FLAG-specific antibodies, indicating that there was no proteolytic cleavage accompanying the unnatural amino-acid incorporation.

2.4 Discussion

2.4.1 Unnatural amino-acid mutagenesis in mammalian cells

Genetically encoding unnatural amino acids in mammalian cells, such as neurons, has been greatly hampered by two main challenges: inefficient biosynthesis of orthogonal bacterial tRNAs and difficulty in generating unnatural-amino-acid-specific synthetases in these cells. The transcription of mammalian tRNAs relies on conserved promoter elements within the tRNA gene, which are lacking in many bacterial tRNAs. Early attempts to create such internal promoters in the EctRNA$^{Tyr}_{CUA}$ through mutation resulted in nonfunctional tRNAs[30]. As for bacterial tRNAs containing the internal promoter elements, expression of them using 5'-flanking sequences of mammalian tRNAs is inefficient even when multiple copies of the bacterial tRNA gene are used[30]. Here we developed a general strategy for efficient expression of bacterial tRNA in mammalian cells, regardless of the internal promoter elements, by using a type-3 pol III promoter, the H1 promoter. This promoter does not require downstream

transcriptional elements, and has a well-defined transcription initiation site for generating the correct 5'-end of tRNA. We showed that the H1 promoter can drive the expression of different tRNAs (EctRNA$^{Tyr}_{CUA}$ and EctRNA$^{Leu}_{CUA}$) in various cell types (HeLa, HEK293, mouse and rat primary neurons) for the incorporation of diverse natural or unnatural amino acids. Other members of the type-3 class of pol III promoter, such as the promoter for U6 snRNA[31], 7SK[32] and MRP/7-2[33] should also work in a similar manner. To evolve a synthetase specific for a desired unnatural amino acid, mutant synthetase libraries containing more than 10^9members have been made and selected in *E. coli* and later in yeast. Owing to the low transfection efficiency, it is impractical to generate such huge libraries in mammalian cells. We show here that synthetases evolved in yeast are functionally compatible with mammalian HEK293 cells, HeLa cells and neurons. This transfer strategy will facilitate the future incorporation of diverse unnatural amino acids tailored for mammalian and neuronal studies. Using these strategies, we were able to genetically encode unnatural amino acids in different mammalian cells, and in primary neurons.

A key advance in our study was that the orthogonal tRNA and synthetase are genetically expressed in mammalian cells. Genetically encoding unnatural amino acids overcomes restrictions imposed by *in vitro* semisynthetic and biosynthetic unnatural-amino-acid incorporation methods on protein type, size, quantity and location[34,35]. Importantly, genetic expression enables studies of structure and function to be carried out in living cells and possibly organisms. For studies with cultured neurons or brain slices, the cDNA for tRNA-synthetase need only be introduced via transfection or virus into the cells. Many unnatural amino acids can be added directly to the growth medium and taken up by the cell through nonspecific amino-acid and amine transporters. Highly polar amino acids can be derivatized with enzymatically labile groups (for example, esterification or acylation) to facilitate cellular uptake. A second level of specificity could be achieved by engineering cell-specific promoters that would drive the expression of the mutant protein and the orthogonal tRNA-synthetase in a subset of neurons[1]. One consideration for future studies is that the combination of transfection efficiency and the efficiency of incorporating unnatural amino acids in neurons must not be too low. The incorporation efficiency of unnatural amino acids used in this study (relative to Tyr incorporation by the wild-type TyrRS) is in the

range of 13% to 41%. Cytoplasmic GFP containing unnatural amino acids was expressed in sufficient amounts in neurons to be easily detected using fluorescence microscopy, and the integral membrane protein Kv1.4 was expressed in HEK293 cells sufficiently for patch clamping. Factors that could determine the incorporation efficiency of unnatural amino acids include the activity of the evolved synthetase, the gene context of the amber stop codon, the ability of cells to take up the unnatural amino acid, and the stability of the target mRNA and tRNA-synthetase pair. The overall efficiency should benefit from improvements in any of these aspects. For instance, a viral expression system might prove to be advantageous for expressing the tRNA-synthetase pair and the target gene. Genetically encoding unnatural amino acids in a multicellular organism will be more challenging. It is possible to engineer a transgenic mouse to express the mutant gene and tRNA-synthetase pair in a specific cell. The relevant unnatural amino acid could be added to the food or drinking water, or injected directly into the brain ventricles. The compatibility of this method with living systems is especially valuable for proteins whose function requires native complex cellular environments such as integral membrane proteins and proteins involved in signaling. Genetic stability and inheritance are well-suited for researching long-term biological processes such as developmental and evolutionary studies. In addition, this technology does not require special expertise, and is easily transferable to the scientific community in the form of plasmid DNA, stable cell lines or transgenic animals.

In the future, we anticipate that unnatural amino acids will be tailor designed and encoded to probe, and more excitingly, to control proteins and protein-related biological processes. For instance, fluorescent unnatural amino acids could be used to sense local environmental changes and serve as reporters for enzyme activity, membrane potential or neurotransmitter release; unnatural amino acids bearing photocrosslinking agents could be applied to identify protein-protein and protein-nucleic acid interactions in cells; and photocaged and photoisomerizable amino acids could be designed to switch on and off signal initiation and transduction noninvasively. Many of these unnatural amino acids have been encoded in *E. coli* and in yeast[36]. The method reported here will enable the genetic encoding of such novel amino acids in mammalian cells and neurons, and stimulate more precise and profound molecular studies of cell biology and neurobiology.

2.4.2 Investigating the fast inactivation of Kv1.4 channels

Neuronal Kv1.4 channels exhibit rapid inactivation (N-type) upon depolarization, and are important for regulating the firing rate of neurons[13]. Using the genetic expression of unnatural amino acids in mammalian cells, we were able to probe the role of a single amino acid in determining the rate of inactivation for Kv1.4 channels. We focused on the Tyr19 amino acid because of its unique position in the N-terminal domain—just downstream of the unfolded inactivation peptide and upstream of a structured α-helix (**Figure 2.4b**). One model of fast inactivation proposes that the N-terminal inactivation domain is threaded through a small portal on the cytoplasmic domain, where the unfolded region extends into the channel's inner cavity. We hypothesized that the portal could exert steric constraints on the size of amino acids that could move through it. The incorporation of OmeTyr or DanAla at the Tyr19 position generated voltage-gated K^+ channels with markedly slowed inactivation. Interestingly, the similarity in inactivation between Y19F and wild-type channels indicates that the hydroxyl group of Tyr is not required for fast inactivation. Thus it seems unlikely that OmeTyr alters fast inactivation by blocking hydrogen bonding. Furthermore, the current response of the Y19W mutant indicates that the increase in the lateral bulkiness of the side chain also does not affect channel inactivation. All three Phe, Tyr and Trp natural aromatic amino acids have a rigid phenyl ring. Incorporating the unnatural amino acids OmeTyr and DanAla enabled us to lengthen the side chain in the direction orthogonal to the peptide backbone and thus effectively to increase the diameter of the peptide cross section. The NMR structure of the inactivation peptide indicates that residues 19–21 are in a β-turn conformation[28]. The distal distance between Tyr19 and Tyr21 is ~19.9 Å (**Figure 2.6**). We modeled Trp and OmeTyr at the Tyr19 site. The distal distance remained 19.9 Å for Trp, but increased to 20.8 Å for OmeTyr. In the crystal structure of Kv1.2, the side portal has a diameter of 15–20 Å[14]. Changing Tyr19 to OmeTyr extends the length of the side chain by 0.9 Å. According to the model in which the inactivation peptide snakes through a side portal[14,15], the change in fast inactivation could occur because the larger diameter of the activation peptide relative to that of the side portal interferes with the ability of the peptide to extend into the channel pore. Such effects would not be observable

through conventional site-directed mutagenesis, as no natural amino acids can further extend the side-chain length rigidly.

Another possibility is that the introduction of unnatural amino acids alters the interaction between the N-terminal inactivation peptide and other proteins, slowing its entry into the inactivation gate. Other proteins that are known to interact with the N terminus of Kv1.4 channels, such as KChIPs, DPPs and Kv- β subunits, were not coexpressed with the channel cDNA in our experiments. If the presence of an unnatural amino acid leads to unexpected interactions between the inactivation peptide and unknown proteins, such interactions should vary with the nature of the unnatural amino acid. OmeTyr and DanAla differ significantly in structure and chemical properties. Nonetheless, the inactivation kinetics were similar for channels containing OmeTyr and channels containing DanAla, indicating that such interactions probably did not occur. In addition, proteolysis of the N-terminal domain can be ruled out because we did not observe any significant degradation of the Kv1.4*-OmeTyr and Kv1.4*-DanAla channel protein (**Figure 2.5f**). Finally, incorporation of OmeTyr or DanAla at Tyr19 could completely disrupt fast N-type inactivation, leaving only C-type inactivation. However, deletion of the N-terminal domain, which eliminates N-type inactivation, produces channels with inactivation kinetics even slower than those containing OmeTyr or DanAla[27]. Thus, we favor the conclusion that the bulky unnatural amino acids OmeTyr and DanAla prevent the N-terminal peptide from fully extending into the channel pore cavity. Future studies will explore the effects of other unnatural amino acids substituted at Tyr19 and neighboring sites.

In summary, genetically encoding unnatural amino acids with orthogonal tRNA-synthetase pairs provides neuroscientists and other researchers with new tools for probing the function of proteins in mammalian cells.

2.5 Methods

2.5.1 Chemicals

OmeTyr and Bpa were purchased from Chem-Impex. DanAla was synthesized as described[25]. All other chemicals were purchased from Sigma-Aldrich.

2.5.2 Constructs

All constructs were assembled by standard cloning methods and confirmed by DNA sequencing. Plasmid pCLHF is a derivative of pCLNCX (Imgenex), and has a hygromycin resistance gene instead of a neomycin resistance gene. The amber stop codon TAG was introduced into the *enhanced GFP* (*EGFP*) gene at position 182 through site-directed mutagenesis. The woodchuck hepatitis virus post-transcriptional regulatory element (WPRE)[37] was added to the 3'-end of the *GFP-TAG* mutant gene. The *GFP-TAG-WPRE*gene fragment was ligated into the *Hin*dIII and *Cla*I sites of *pCLHF* to afford the plasmid *pCLHF-GFP-TAG*.

The *E. coli TyrRS* gene was amplified from *E. coli* genomic DNA using the primers 5'-CCA CCA TGG AAC TCG AGA TTT TGA TGG CAA GCA GTA ACT TGA TTA AAC-3' and 5'-ACA AGA TCT GCT AGC TTA TTT CCA GCA AAT CAG ACA GTA ATT C-3'. Genes for OmeTyrRS (Y37T, D182T and F183M) and Bpa-TyrRS (Y37G, D182G and L186A)[38] were made from the *E. coli TyrRS* gene by site-directed mutagenesis using overlapping PCR. The gene for EctRNA$^{Leu}_{CUA}$ in construct tRNA2 was made using the primers 5'-GTG GGA TCC CCG GTG GGG TTC CCG AGC GGC CAA AGG GAG CAG ACT CTA AAT CTG CCG TCA TCG ACT TCG-3' and 5'-GAT AAG CTT TTC CAA AAA TGG TGG TGG GGG AAG GAT TCG AAC CTT CGA AGT CGA TGA CGG CAG ATT TAG-3' through Klenow extension. Other tRNA constructs were made by PCR using tRNA2 as the template. Genes for EctRNA$^{Tyr}_{CUA}$and the mutant synthetase specific for DanAla were amplified from plasmid pLeuRSB8T252A[25] using PCR. The *E. coli LeuRS* gene was amplified from *E. coli*genomic DNA using primers 5'-GCC TCG AGA AGA GCA ATA CCG CCC GG-3' and 5'-CGC TAG CTT AGC CAA CGA CCA GAT TGA GGA G-3'. The H1 promoter was amplified from plasmid pSUPER (OligoEngine).

To make the tRNA-aaRS expression plasmid pEYCUA-YRS, we used pBluescript II KS (Stratagene) as the backbone for construction. The PGK promoter and the SV40 polyA signal were inserted between *EcoR*I and *Not*I sites. The *E. coli TyrRS* gene was inserted between the PGK and SV40 polyA sequences using the introduced *Xho*I and *Nde*I sites. The H1 promoter containing the *Bgl*II

and *Hin*dIII sites at the 3'-end was cloned into the *EcoR*I and *Cla*I sites. The EctRNA$^{Tyr}_{CUA}$ was then inserted between the *Bgl*II and*Hin*dIII sites. Finally, a gene cassette containing the SV40 promoter followed by the neomycin resistance gene and the SV40 poly A signal was amplified from pCDNA3 (Invitrogen) and inserted into the *Cla*I and *Kpn*I sites. Other tRNA-synthetase plasmids were modified from plasmid pEYCUA-YRS by swapping the synthetase gene or the tRNA gene, or by inserting various 3'-flanking sequences after the tRNA.

We used rat Kv1.4 cDNA, which matched NM_012971 except for one difference (L42A). Mutations M109L, 19TAG, 557TAG, Y19F and Y19W were created using the QuikChange Site-Directed Mutagenesis Kit (Stratagene). When Kv1.4-19TAG was expressed in HEK293 cells, voltage-dependent K^+ currents could be detected in the absence of the orthogonal EctRNA-synthetase pair. We discovered that this was due to undesired transcription initiation at residue Met109, which generates a truncated Kv1.4 channel. To solve this problem, we mutated Met109 to Leu to make Kv1.4*. No truncated Kv1.4 channel was expressed from Kv1.4*-19TAG alone, and only background currents were detected.

2.5.3 Cell culture and transfection

HeLa, HEK293T and HEK293 cells were cultured and maintained with Dulbecco's modified Eagle's medium (DMEM, Mediatech) supplemented with 10% FBS.

To establish the GFP-TAG HeLa stable cell line, we transfected 293T cells with the retroviral vector pCLHF-GFP-TAG and the packaging vector pCL-Ampho (Imgenex) using FuGENE 6 transfection reagent (Roche). Viruses were harvested after 48 h and used to infect HeLa cells grown in 50% conditioned medium in the presence of 8 ng/ml hexadimethrine bromide (Sigma). From the next day on, cells were split to a very low confluence. Stably infected cells were selected with 200 ng/ml hygromycin (Invitrogen). Hygromycin (50 ng/ml) was always present in subsequent cell culture to ascertain plasmid DNA maintenance.

For electrophysiological recordings, HEK293 cells were transiently transfected using Lipofectamine 2000 (Invitrogen). After 8 h, cells were reseeded onto 12-mm glass cover slips (Fisher)

coated with poly-L-lysine (10 μg/ml; Sigma, MW30,000) in 24-well plates. OmeTyr or DanAla (1 mM) was added to the culture medium, and cells were incubated for 24 h at 37 °C before recording.

Hippocampi of postnatal day 0 Sprague-Dawley rats or mice were removed and treated with 2.5% trypsin (Invitrogen) for 15 min at 37 °C. The digestion was stopped with 10 ml of DMEM containing 10% heat-inactivated FBS. The tissue was triturated in a small volume of this solution with a fire-polished Pasteur pipette, and ~100,000 cells in 0.5 ml neuronal culture medium were plated per coverslip in 24-well plates. Glass coverslips were prewashed overnight in HCl followed by several rinses with 100% ethanol and flame sterilization. They were subsequently coated overnight at 37 °C with poly-L-lysine. Cells were plated and grown in Neurobasal-A (Invitrogen) containing 2% B-27 (Life Technologies), 1.8% HEPES, and 2 mM glutamine (Life Technologies). Half of the medium was replaced the next day. For imaging, cells cultured for 3 days were transfected with Lipofectamine 2000, changed into fresh medium with 1 mM OmeTyr or Bpa after 5 h, and cultured for another 24 h before the experiment.

2.5.4 Northern blot and western blot analysis

RNA was prepared from the GFP-TAG HeLa cells transfected with different tRNA-aaRS constructs using PureLink miRNA Kit (Invitrogen). The RNA was denatured, electrophoresed on 15% PAGE gel, blotted onto Hybond-N (Amersham) membrane, and crosslinked by ultraviolet fixation. ^{32}P-labeled DNA probes specific for the EctRNA$^{Tyr}_{CUA}$ were made using Klenow extension with primer 5'-AAC CTT CGA AGT CGA TGA CGG CAG ATT TAC AGT CTG C-3' and primer 5'-CCG TCT AAA TGT CAG ACG AGG GAA ACC GGC GAG-3'. After pre-hybridization for 4 h in the hybridization buffer (5x sodium chloride-sodium citrate buffer, 40 mM Na_2HPO_4 (pH 7.2), 7% sodium dodecylsulfate (SDS), 2 x Denhardt's), membranes were hybridized with ^{32}P-labeled cDNA probes (0.5–2 x 10^7 c.p.m./ml) in the same buffer plus 50 μg/ml salmon sperm DNA at 58 °C overnight. Hybridized membrane was sequentially washed with high stringency buffer (40 mM Na_2HPO_4, 1 mM EDTA, 1% SDS, 58 °C) twice and exposed to an X-ray film (Kodak) for 48 h. To control the total RNA amount

loaded in each lane, glyceraldehyde-3-phosphate dehydrogenase (GAPDH) transcript was used as an internal standard.

A FLAG tag was added to the C termini of genes for Kv1.4*, Kv1.4* with the first 19 amino acid residues deleted, and Kv1.4*-19TAG, respectively. Full-length and truncated (Δ1–19) Kv1.4* were expressed in HEK293 cells. Kv1.4*-19TAG was co-expressed with EctRNA$^{Tyr}_{CUA}$-TyrRS, with EctRNA$^{Tyr}_{CUA}$-OmeTyrRS supplied with OmeTyr (1 mM), or with EctRNA$^{Leu}_{CUA}$-DanAlaRS supplied with DanAla (1 mM) in HEK293 cells. Cells were lysed with 1% Nonidet P-40 (Calbiochem) in PBS at 4 °C for 30 min. After centrifugation, supernatants containing membrane proteins were loaded and separated by SDS-PAGE. A monoclonal antiflag antibody (Sigma) was used to detect the FLAG-containing proteins.

2.5.5 Flow cytometry

GFP-TAG HeLa cells were transfected with plasmid DNA by lipofection 2000 according to the protocol of the vendor (Invitrogen). Unnatural amino acids (1 mM) were added to the medium immediately after transfection. Cells were collected after 48 h, washed twice, and resuspended in 1 ml of PBS containing 0.05 μg/ml propidium iodide. Samples were analyzed with a FACScan (Becton & Dickinson).

2.5.6 Electrophysiology

Whole-cell patch clamp was used to record macroscopic currents from HEK293 cells. Borosilicate glass (Warner) electrodes had resistance of 3–8 MΩ. Membrane currents were recorded with an Axopatch 200B amplifier (Axon Instruments), adjusted electronically for cell capacitance and serial resistance (80–100%), filtered at 1 kHz, and digitized at 10 kHz. The intracellular solution consisted of 150 mM KCl, 5.46 mM $MgCl_2$, 10 mM HEPES (pH7.4 with 3.3 mM Na^+), 5 mM EGTA, 2.56 mM K_2ATP and 0.3 mM Li_2GTP. The external bath solution contained 155 mM NaCl, 5 mM KCl, 2 mM $MgCl_2$, 0.5 mM $CaCl_2$ and 10 mM HEPES (pH7.4). The osmolarity was 310–330 mOsm.

Macroscopic currents were elicited with 1-s voltage steps from –100 mV to +40 mV in 20-mV increments. The time interval between voltage steps was 10 s. For analysis, the current elicited by a 1-s voltage step to +20 mV was fit with a sum of two exponentials ($A1\exp(-t/\tau_1) + A2\exp(-t/\tau_2)$ and a constant (C). A 4-s voltage step to +20 mV was used for OmeTyr and DanAla mutant channels. All recordings were performed at room temperature. Clampex8 was used for recording and data were analyzed using Clampfit 8 (Axon). All values shown are mean ± s.e.m. Statistical significance ($P < 0.05$) was assessed using Kruskal-Wallis one-way ANOVA on Ranks, followed by Dunn's post-hoc test for significance (SigmaStat 3.0).

2.5.7 Fluorescence microscopy

Fluorescence images were acquired on an Olympus X81 inverted microscope using a 20 x objective. For the GFP channel, filters were 480/30 nm for excitation and 535/40 nm for emission. For the mCherry channel, filters were 580/20 nm for excitation and 675/130 nm for emission.

2.6 Author contributions

W.W. conducted experiments for tRNA expression, encoding unnatural amino acids in neurons and probing channel inactivation, and analyzed the data. J.K.T. conducted experiments for encoding unnatural amino acids in other mammalian cells, made constructs for neuronal work, and analyzed the data. G.V.L. conducted the simulation of the inactivation peptide. T.J.B. synthesized DanAla. J.P.N. provided support for computer simulation and chemical synthesis. K.L. provided support for neuronal work. P.A.S. provided support and guidance for the electrophysiology experiments, analyzed the data, and revised the manuscript. L.W. conceived and designed the experiments for unnatural amino acid incorporation and ion channel inactivation, analyzed the data, wrote the manuscript, and supervised the project.

2.7 ACKNOWLEDGEMENTS

We thank B.R. Martin and R.Y. Tsien for plasmid pCLHF, E. Cooper for Kv1.4 cDNA, and D. Summerer and P.G. Schultz for plasmid pLeuRSB8T252A before publication. We thank A.R. Parrish, K. Adams, J. Xu, F.D. Winter, R. Nassirpour, Z. Chen for technical support, and P. Aryal for helpful discussions. P.A. Slesinger acknowledges grant support from US National Institutes of Health. L. Wang acknowledges the support of the Searle Scholar Program and the Beckman Young Investigator Program.

Chapter 2 is a reprint in full of the material as it appears in Nature Neuroscience, (2007) , Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L., Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat Neurosci* 2007, *10* (8), 1063-72.

2.8 FIGURES

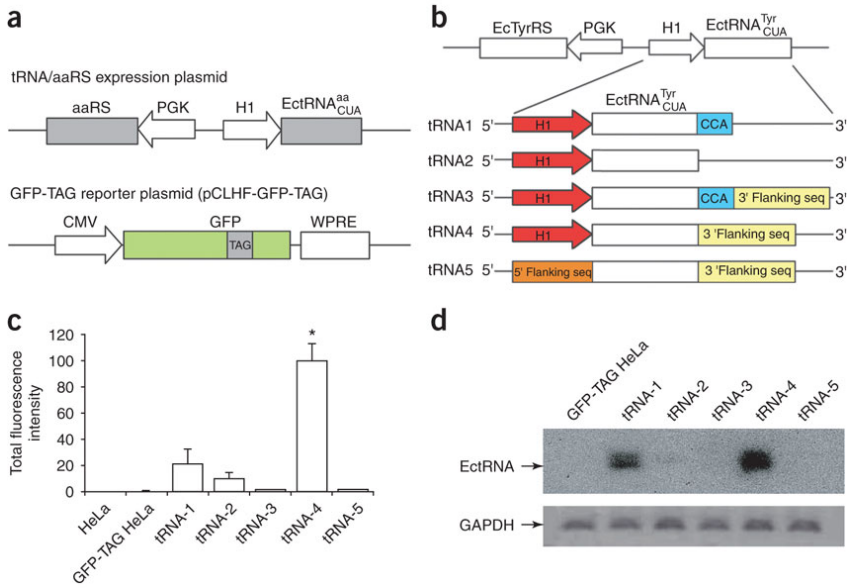


Figure 2.1: Efficient expression of bacterial tRNA in mammalian cells using the H1 promoter. (**a**) A fluorescence-based assay for the expression of functional tRNA in mammalian cells. The candidate amber suppressor tRNA and its cognate synthetase were expressed using the tRNA-aaRS expression plasmid. A reporter plasmid was used to express *GFP* with an amber stop codon at a permissive site. (**b**) Schematic illustration of tRNA-aaRS expression plasmids using different elements to drive the tRNA transcription and processing. (**c**) Total fluorescence intensity of the fluorescent GFP-TAG HeLa cells after transfection with constructs shown in **b**. The intensities were normalized to those of cells transfected with tRNA-4. The values (± s.e.m.) were: GFP-TAG HeLa 0.3 ± 0.1, tRNA-1 21 ± 11, tRNA-2 10 ± 4.7, tRNA-3 1.3 ± 0.7, tRNA-4 100 ± 12, tRNA-5 1.4 ± 0.5. For all samples, n = 5. (**d**) Northern blot analysis of the amount of transcribed $EctRNA^{Tyr}_{CUA}$ in HeLa cells. Transcript of glyceraldehyde-3-phosphate dehydrogenase (GAPDH) was used to normalize the total amount of RNA in different samples.

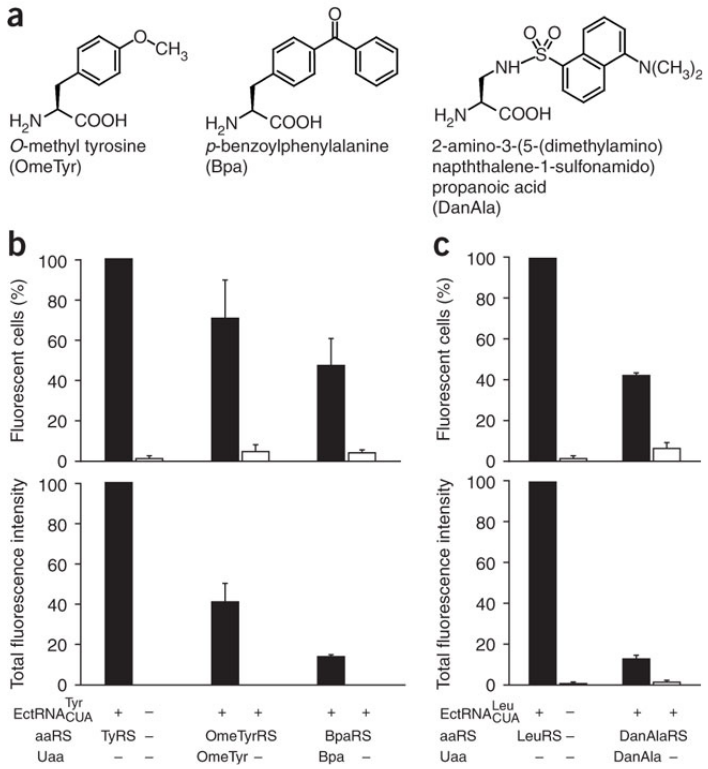


Figure 2.2: Unnatural-amino-acid-specific synthetases evolved in yeast are functional in mammalian cells. (**a**) Structures of unnatural amino acids used in this study. (**b**) Incorporation of OmeTyr and Bpa into GFP in the GFP-TAG HeLa cells using the EctRNA$^{Tyr}_{CUA}$ and corresponding synthetases evolved from *E. coli* TyrRS in yeast. All data were normalized to those obtained from GFP-TAG HeLa cells transfected with the EctRNA$^{Tyr}_{CUA}$ and wild-type *E. coli* TyrRS. Percentages (± s.e.m.) of fluorescent cells were 71 ± 19 (+ OmeTyr, n = 3); 4.8 ± 3.4 (– OmeTyr, n = 3); 47 ± 14 (+ Bpa, n = 3); and 4.2 ± 1.5 (– Bpa, n = 3). The total fluorescence intensities (±s.e.m.) were 41 ± 9.5 (+ OmeTyr, n = 3); 0.17 ± 0.02 (– OmeTyr, n = 3); 13 ± 1.4 (+ Bpa, n = 3); and 0.11 ± 0.06 (– Bpa, n = 3). (**c**) Incorporation of DanAla into GFP in the GFP-TAG HeLa cells using the EctRNA$^{Leu}_{CUA}$ and a DanAla-specific synthetase evolved from *E. coli*LeuRS. Data were normalized as in **b**. Percentages (± s.e.m) of fluorescent cells were 42 ± 1.3 (+ DanAla, n = 3) and 5.9 ± 2.6 (– DanAla,n = 3). The total fluorescence intensities (± s.e.m.) were 13 ± 2.1 (+ DanAla, n = 3) and 1.4 ± 1.0 (– DanAla, n = 3).

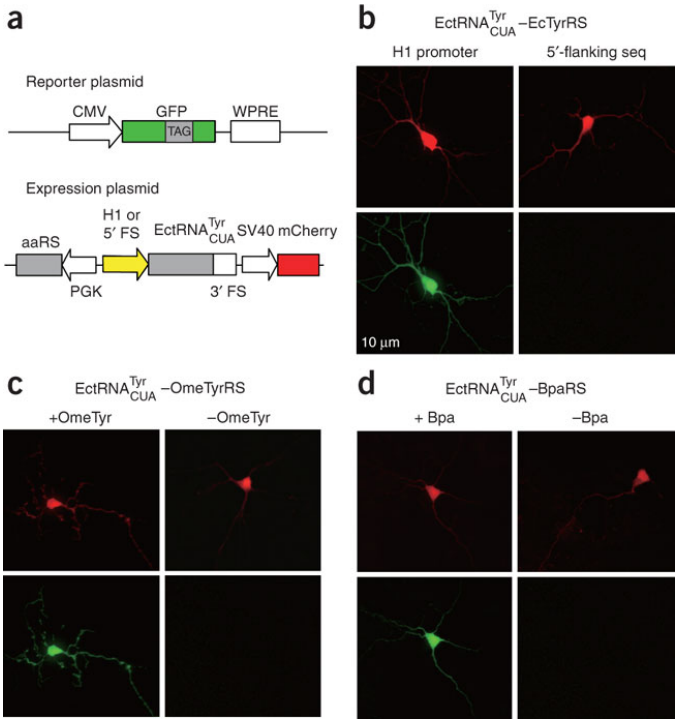


Figure 2.3: Genetically encoding unnatural amino acids in neurons. (**a**) Schematic illustration of the reporter plasmid expressing the GFP mutant gene with a TAG stop codon at site 182 and the expression plasmid encoding the EctRNA$^{Tyr}_{CUA}$, the synthetase and an internal transfection marker (mCherry). (**b**) Fluorescence images of neurons transfected with the reporter plasmid, the EctRNA$^{Tyr}_{CUA}$, and wild-type *E. coli* TyrRS. tRNA expression was driven by the H1 promoter in the left panels, and by the 5'-flanking sequence of the human tRNA^{Tyr} in the right panels. (**c**) Fluorescence images of neurons transfected with the reporter plasmid, the EctRNA$^{Tyr}_{CUA}$, and the OmeTyrRS in the presence (left panels) and absence (right panels) of OmeTyr. (**d**) Fluorescence images of neurons transfected with the reporter plasmid, the EctRNA$^{Tyr}_{CUA}$, and the BpaRS in the presence (left panels) and absence (right panels) of Bpa.

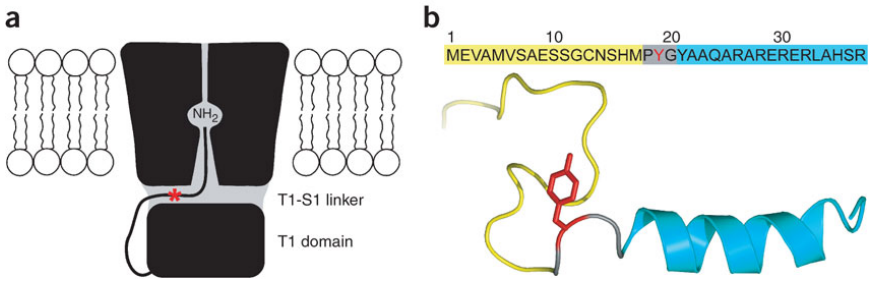


Figure 2.4: A model for the N-type inactivation of Kv channels. (**a**) The model proposes that the N-terminal inactivation peptide snakes through the side portal, and then enters and plugs the inner pore of the Kv channel[14, 15]. The side portals are formed by the transmembrane domain, the T1 domain and the T1-S1 linker. The hydrophobic region of the inactivation peptide reaches into the inner pore, and the positively charged residues in the hydrophilic region of the inactivation peptide make electrostatic interactions with negatively charged surfaces of the T1 domain and the T1-S1 linker. (**b**) The amino-acid sequences of the N-terminal inactivation peptide of Kv1.4, and its structure in aqueous solution determined by NMR[28]. Only one conformer of the random N-terminal Met1–Met17 is shown in yellow. The β-turn and α-helix are colored gray and cyan, respectively. Residue Tyr19 is colored red, and its approximate position is indicated with a red star in **a**.

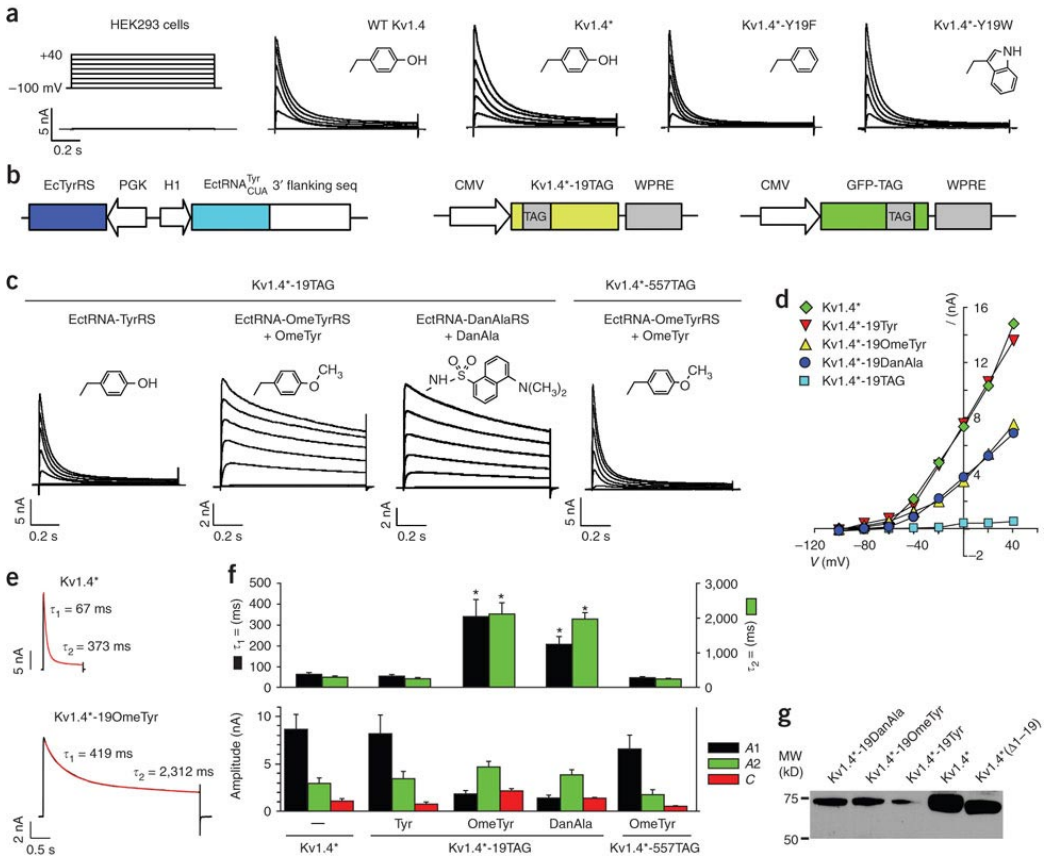


Figure 2.5: Probing N-type inactivation using unnatural amino acid mutagenesis with Kv1.4 channels expressed in mammalian cells. (**a**) Comparison of macroscopic currents recorded from HEK293 cells expressing wild-type Kv1.4, Kv1.4* (Kv1.4-M109L), Kv1.4*-Y19F or Kv1.4*-Y19W channels. Untransfected HEK293 cells displayed a small, background current (0.50 ± 0.03 nA, n = 5). Inserts in each panel show the side-chain structures of amino acids at position 19. (**b**) Constructs for expressing Kv1.4 through amber suppression. The GFP-TAG reporter plasmid was used as an indicator for transfection. (**c**) Potassium currents recorded from HEK293 cells expressing Kv1.4*-TAG channels through amber suppression using the indicated orthogonal tRNA-synthetase pair. Only background currents were recorded for Kv1.4*-19TAG alone (0.40 ± 0.05 nA, n = 5) or Kv1.4*-TAG with the EctRNA-OmeTyrRS pair in the absence of OmeTyr (0.42 ± 0.07 nA, n = 4). The Kv1.4*-19OmeTyr and Kv1.4*-19DanAla mutants exhibited significantly slower inactivation than Kv1.4* and Kv1.4*-19Tyr. Inserts in each panel are the side chain structures. (**d**) Peak current plotted as a function of voltage for the indicated Kv1.4 channels. (**e**) Current (black trace) elicited by a 1-s (Kv1.4*) or 4-s (Kv1.4*-19OmeTyr) voltage step to +20 mV were fit with a sum of two exponentials (red line) relaxing to a small steady-state current with the indicated time constants. (**f**) Quantitative analysis of inactivation kinetics. Bar graphs show the average time constants (τ_1, τ_2) for the fast and slow inactivation components, current amplitudes (A1, A2) for the exponentials and steady-state current (C). Specific values are in **Table 2.1**. Asterisks indicate statistical difference from Kv1.4* ($P < 0.05$). (**g**) Western blot analysis of Kv1.4*-19OmeTyr and Kv1.4*-19DanAla. The full-length Kv1.4*, Kv1.4*-19Tyr, and a truncated Kv1.4* with the first 19 amino acid residues deleted were run as controls.

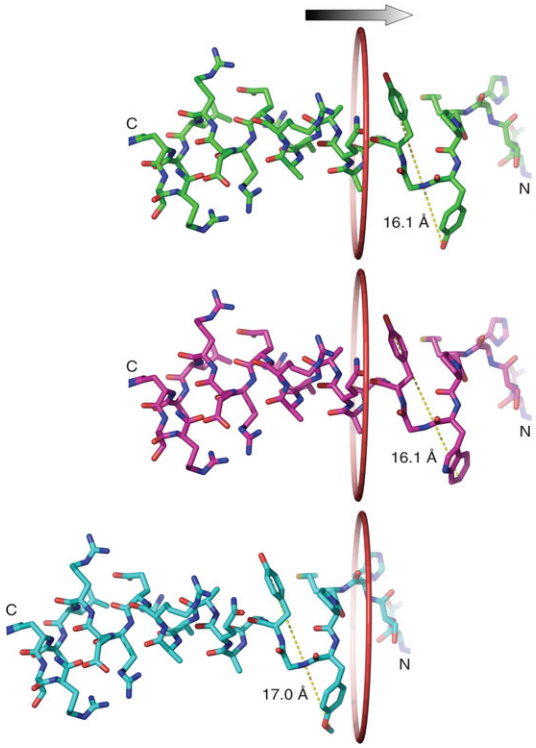


Figure 2.6: Diameter of the inactivation peptide affects channel inactivation owing to the size restriction of the side portal. Upper panel, NMR structure of the inactivation peptide from Asn14 to Arg37 of Kv1.4. The central distance between the oxygen atoms of the hydroxyl groups of Tyr19 and Tyr21 is 16.1 Å. The surface distance of these two residues is thus ~ 19.9 Å by including the Van der Waals radius of the hydrogen atom at both ends. Trp (middle panel) and OmeTyr (bottom panel) were modeled in the structure and the corresponding central distances were measured. A ring with a diameter of 20 Å is drawn to illustrate the side portal entrance. The change of Tyr19 to OmeTyr increases the distal distance by ~ 0.9 Å, which can impede the entry of residue 19 and following residues of the inactivation peptide into the portal, effectively abolishing the fast inactivation.

2.9 Tables

Table 2.1: Kinetics of current inactivation of Kv1.4 channels

	Time constant (ms)		Amplitude (nA)		Steady-state current (nA)	Cells
Kv1.4 channel	τ_1	τ_2	*A*1	*A*2	*C*	*n*
Kv1.4*	63.7 ± 5.8	300 ± 31	8.67 ± 1.53	2.99 ± 0.52	1.09 ± 0.23	9
Kv1.4*-19Tyr	54.7 ± 6.4	250 ± 34	8.14 ± 2.04	3.42 ± 0.81	0.79 ± 0.16	7
Kv1.4*-19OmeTyr	340 ± 80	2,115 ± 323	1.81 ± 0.38	4.69 ± 0.56	2.19 ± 0.18	11
Kv1.4*-19DanAla	205 ± 39	1,978 ± 174	1.35 ± 0.35	3.83 ± 0.57	1.37 ± 0.10	11
Kv1.4*-557OmeTyr	47.6 ± 3.3	234 ± 18	6.58 ± 1.44	1.77 ± 0.52	0.53 ± 0.08	8

All values are mean ± s.e.m.

2.9 REFERENCES

1. Callaway, E.M. A molecular and genetic arsenal for systems neuroscience. *Trends Neurosci.* **28**, 196-201 (2005).

2. Giepmans, B.N., Adams, S.R., Ellisman, M.H. & Tsien, R.Y. The fluorescent toolbox for assessing protein location and function. *Science* **312**, 217-224 (2006).

3. Gandhi, C.S. & Isacoff, E.Y. Shedding light on membrane proteins. *Trends Neurosci.* **28**, 472-479 (2005).

4. Miyawaki, A. Fluorescent indicators for Ca2+ based on green fluorescent proteins and calmodulin. *Nature* **388**, 882-887 (1997).

5. Miesenbock, G., De Angelis, D.A. & Rothman, J.E. Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins. *Nature* **394**, 192-195 (1998).

6. Nagel, G. Light activation of channelrhodopsin-2 in excitable cells of Caenorhabditis elegans triggers rapid behavioral responses. *Curr. Biol.* **15**, 2279-2284 (2005).

7. Wang, L. & Schultz, P.G. Expanding the genetic code. *Angew. Chem. Int. Edn. Engl.* **44**, 34-66 (2004).

8. Nowak, M.W. Nicotinic receptor binding site probed with unnatural amino acid incorporation in intact cells. *Science* **268**, 439-442 (1995).

9. Lummis, S.C. Cis-trans isomerization at a proline opens the pore of a neurotransmitter-gated ion channel. *Nature* **438**, 248-252 (2005).

10. Lummis, S.C., Beene, D.L., Harrison, N.J., Lester, H.A. & Dougherty, D.A. A cation-pi binding interaction with a tyrosine in the binding site of the GABAC receptor. *Chem. Biol.* **12**, 993-997 (2005).

11. Beene, D.L., Price, K.L., Lester, H.A., Dougherty, D.A. & Lummis, S.C. Tyrosine residues that control binding and gating in the 5-hydroxytryptamine3 receptor revealed by unnatural amino acid mutagenesis. *J. Neurosci.* **24**, 9097-9104 (2004).

12. Wang, L., Brock, A., Herberich, B. & Schultz, P.G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498-500 (2001).

13. Johnston, D. Dendritic potassium channels in hippocampal pyramidal neurons. *J. Physiol. (Lond.)* **525**, 75-81 (2000).

14. Long, S.B., Campbell, E.B. & Mackinnon, R. Crystal structure of a mammalian voltage-dependent Shaker family K+ channel. *Science* **309**, 897-903 (2005).

15. Zhou, M., Morais-Cabral, J.H., Mann, S. & MacKinnon, R. Potassium channel receptor site for the inactivation gate and quaternary amine inhibitors. *Nature* **411**, 657-661 (2001).

16. Demo, S.D. & Yellen, G. The inactivation gate of the Shaker K+ channel behaves like an open-channel blocker. *Neuron* **7**, 743-753 (1991).

17. Hoshi, T., Zagotta, W.N. & Aldrich, R.W. Biophysical and molecular mechanisms of Shaker potassium channel inactivation. *Science* **250**, 533-538 (1990).

18. Wang, L., Magliery, T.J., Liu, D.R. & Schultz, P.G. A new functional suppressor tRNA/aminoacyl-tRNA synthetase pair for the in vivo incorporation of unnatural amino acids into proteins. *J. Am. Chem. Soc.* **122**, 5010-5011 (2000).

19. Wang, L. & Schultz, P.G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8**, 883-890 (2001).

20. Galli, G., Hofstetter, H. & Birnstiel, M.L. Two conserved sequence blocks within eukaryotic tRNA genes are major promoter elements. *Nature* **294**, 626-631 (1981).

21. Myslinski, E., Ame, J.C., Krol, A. & Carbon, P. An unusually compact external promoter for RNA polymerase III transcription of the human H1RNA gene. *Nucleic Acids Res.* **29**, 2502-2509 (2001).

22. Edwards, H. & Schimmel, P. A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase. *Mol. Cell. Biol.* **10**, 1633-1641 (1990).

23. Doctor, B.P. & Mudd, J.A. Species specificity of amino acid accepter ribonucleic acid and aminoacyl soluble ribonucleic acid synthetases. *J. Biol. Chem.* **238**, 3677-3681 (1963).

24. Drabkin, H.J., Park, H.J. & Rajbhandary, U.L. Amber suppression in mammalian cells dependent upon expression of an Escherichia coli aminoacyl-tRNA synthetase gene. *Mol. Cell. Biol.* **16**, 907-913 (1996).

25. Summerer, D. A genetically encoded fluorescent amino acid. *Proc. Natl. Acad. Sci. USA* **103**, 9785-9789 (2006).

26. Murrell-Lagnado, R.D. & Aldrich, R.W. Interactions of amino terminal domains of Shaker K channels with a pore blocking site studied with synthetic peptides. *J. Gen. Physiol.* **102**, 949-975 (1993).

27. Kondoh, S., Ishii, K., Nakamura, Y. & Taira, N. A mammalian transient type K+ channel, rat Kv1.4, has two potential domains that could produce rapid inactivation. *J. Biol. Chem.* **272**, 19333-19338 (1997).

28. Antz, C. NMR structure of inactivation gates from mammalian voltage-dependent potassium channels. *Nature* **385**, 272-275 (1997).

29. Wissmann, R. Solution structure and function of the [ldquo]tandem inactivation domain[rdquo] of the neuronal A-type potassium channel Kv1.4. *J. Biol. Chem.* **278**, 16142-16150 (2003).

30. Sakamoto, K. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* **30**, 4692-4699 (2002).

31. Sakamoto, K. et al. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res* **30**, 4692-9 (2002).

32. Murphy, S., Di Liegro, C. & Melli, M. The in vitro transcription of the 7SK RNA gene by RNA polymerase III is dependent only on the presence of an upstream promoter. *Cell* **51**, 81-87 (1987).

33. Yuan, Y. & Reddy, R. 5[prime] flanking sequences of human MRP/7-2 RNA gene are required and sufficient for the transcription by RNA polymerase III. *Biochim. Biophys. Acta* **1089**, 33-39 (1991).

34. Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72**, 249-289 (2003).

35. Cornish, V.W., Mendel, D. & Schultz, P.G. Probing protein structure and function with an expanded genetic code. *Angew. Chem. Int. Edn. Engl.* **34**, 621-633 (1995).

36. Wang, L., Xie, J. & Schultz, P.G. Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 225-249 (2006).

37. Zufferey, R., Donello, J.E., Trono, D. & Hope, T.J. Woodchuck hepatitis virus post-transcriptional regulatory element enhances expression of transgenes delivered by retroviral vectors. *J. Virol.* **73**, 2886-2892 (1999).

38. Chin, J.W. An expanded eukaryotic genetic code. *Science* **301**, 964-967 (2003).

Chapter 3

Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells

3.1 Abstract

Optimizing the anticodon recognition between orthogonal tRNA and synthetase significantly increased the incorporation efficiencies of various unnatural amino acids in mammalian cells, and the enhanced incorporation enabled efficient photocrosslinking of interacting proteins in mammalian cells.

3.2 Introduction

Unnatural amino acids have been genetically encoded in *Escherichia coli*, yeast, and mammalian cells using orthogonal tRNA–synthetase pairs and unique codons[1]. This technology enables novel chemical and physical properties to be selectively introduced into proteins directly in live cells, and thus have great potential for addressing molecular and cell biological questions in the native cell settings. While the incorporation efficiency of unnatural amino acids is high in *E. coli* and yeast[2, 3], it is currently low in mammalian cells. Inefficient incorporation results in reduced yield of proteins containing the unnatural amino acid, which may not be sufficient to perform the desired function and to be detected. In addition, as stop codons are the most frequently used to encode unnatural amino acids, low incorporation efficiency often leads to an increase in truncated protein products, which may negatively interfere with the function of the full-length target protein. Therefore, efficient incorporation of unnatural amino acids is critical for their effective application in mammalian cells. Efforts to improve the efficiency have been focused on optimizing the expression of the orthogonal tRNA and synthetase[2-5]; another important aspect is to increase the affinity of the orthogonal tRNA toward the orthogonal synthetase, which has not been addressed in mammalian cells. Here we show that the recognition of the orthogonal tRNA by the synthetase could be enhanced by engineering the anticodon-binding domain of the synthetase. The mutation identified could be generally transplanted to multiple synthetases specific for different unnatural amino acids to improve their incorporation efficiencies in mammalian cells. Using the enhanced synthetase specific for *p*-azidophenylalanine (Azi), we performed site-specific photocrosslinking of interacting proteins in mammalian cells with high efficiency.

The efficiency of an orthogonal synthetase in charging the orthogonal tRNA with an unnatural amino acid depends on the affinity of the synthetase toward both the unnatural amino acid and the

tRNA. To decode the amber stop codon UAG used for specifying the unnatural amino acid, the anticodon of the orthogonal tRNA has to be changed to CUA. This change would decrease the affinity of the tRNA toward its cognate orthogonal synthetase, because the anticodon is a major recognition element of most synthetases in distinguishing tRNAs[6]. The *E. coli* $tRNA^{Tyr}_{CUA}$–tyrosyl tRNA synthetase (TyrRS) pair has been the candidate for evolving orthogonal tRNA–synthetase pairs for incorporating various unnatural amino acids in yeast and mammalian cells[2-5]. We thus decided to improve the recognition of the *E. coli* TyrRS toward its $tRNA^{Tyr}_{CUA}$.

3.3 Results and Discussion

As there is no crystal structure available for the *E. coli* $tRNA^{Tyr}$–TyrRS complex, we superposed the *E. coli* TyrRS[7] onto the homologous *Thermus thermophilus* TyrRS in the $tRNA^{Tyr}$–TyrRS complex of the latter[8] using the secondary-structure matching method[9]. The two TyrRS superposed well with an r.m.s.d. of 2.2 Å for the main chain atoms (**Figure 3.1a**). Only G34 in the anticodon of $tRNA^{Tyr}$ needs to be changed to C34 to make $tRNA^{Tyr}_{CUA}$. The G34 base of *T. thermophilus* $tRNA^{Tyr}$ is recognized by the carboxyl group of Asp259 in *T. thermophilus* TyrRS, which corresponds to Asp265 in *E. coli* TyrRS. Mutation of G34 to C34 would create a gap between the base and the Asp residue, disrupting the recognition (**Figure 3.1b**). We thus mutated Asp265 in *E. coli* TyrRS to five amino acids with side chains longer than Asp (Tyr, Arg, Gln, Phe, and Leu) in the hope to restore the interaction with C34. These five mutants were first screened in yeast to test their ability to suppress the amber codon together with the *E. coli*$tRNA^{Tyr}_{CUA}$. An amber stop codon was introduced into a permissive site (Lys5) in the HIS3 gene. If the *E. coli* $tRNA^{Tyr}_{CUA}$–mutant TyrRS suppresses the amber codon, yeast cells will be able to make full length HIS3 protein and survive in media lacking histidine. Yeast harboring mutant Asp265Phe and Asp265Leu could not grow, suggesting these two mutations could abolish the recognition for the $tRNA^{Tyr}_{CUA}$. The other three mutants grew as good as the wt TyrRS, and were further characterized in mammalian cells using an *in vivo* fluorescence assay.

In this assay, genes for the *E. coli* $tRNA^{Tyr}_{CUA}$ and synthetase mutant to be tested were transfected into a stable clonal HeLa cell line, in which the green fluorescent protein (GFP) gene containing a premature UAG stop codon at a permissive site (Tyr39) was integrated into the genome (**Figure 3.2a**)[2]. Suppression of UAG39 by the $tRNA^{Tyr}_{CUA}$–synthetase pair results in full-length fluorescent GFP. The total fluorescence intensity of cells measured by fluorescence-activated cell sorting (FACS) reports the incorporation efficiency and thus the affinity of the mutant synthetase for the $tRNA^{Tyr}_{CUA}$. Compared with the wt TyrRS, the Asp265Gln and Asp265Tyr mutants decreased the fluorescence intensity to below 10%, while the Asp265Arg mutant increased the fluorescence intensity to 186%, indicating that the Asp265Arg mutant TyrRS (EwtYRS) has higher activity toward the $tRNA^{Tyr}_{CUA}$ than wt TyrRS (**Figure 3.2c**).

To test if the Asp265Arg mutation could be transplanted to improve the incorporation efficiency of different unnatural amino acids in mammalian cells, we introduced this mutation to 4 synthetases that are derived from the *E. coli* TyrRS for the incorporation of unnatural amino acids Bzo, Azi, Ome, and Pyo, respectively (**Figure 3.2b**)[10, 11]. The resultant mutant synthetases (referred to as enhanced synthetases EBzoRS, EAziRS, EOmeRS, and EPyoRS) were individually coexpressed with the *E. coli* $tRNA^{Tyr}_{CUA}$ in the clonal GFP reporter HeLa cell line in the absence and presence of the unnatural amino acid. The incorporation efficiency was quantified with the total fluorescence intensity of cells using FACS (**Figure 3.2c**). In comparison to the original synthetases, all four enhanced mutant synthetases improved the incorporation efficiency of their respective unnatural amino acids, with the increase ranging from 1.6 fold to 5.2 fold. The EOmeRS and EPyoRS could incorporate the corresponding unnatural amino acid more efficiently than the wt *E. coli* TyrRS. Importantly, the background misincorporation in the absence of the unnatural amino acid did not increase significantly, suggesting that the Asp265Arg mutation did not affect the amino acid specificity of the wt and evolved synthetases. That all four tested mutants showed improvement suggests that the Asp265Arg mutation can generally be transplanted to more TyrRS-derived synthetases.

Prokaryotic TyrRS differ distinctly from archaeal and eukaryotic TyrRS in anticodon recognition[12]. In prokaryotic $tRNA^{Tyr}$, base ψ35 is flipped oppositely to G34 and A36 and recognized

by the extra *C*-terminal domain characteristic of prokaryotic TyrRS, and base G34 is stacked directly on base A36. In contrast, in archaeal and eukaryotic $tRNA^{Tyr}$, all three anticodon bases are in the same side, and G34 is stacked with hydrophobic residues of the TyrRS. Although G34 is recognized by the carboxyl group of Asp in both prokaryotic and archaeal/eukaryotic TyrRS, the helices in which this Asp residue locates do not correspond. By mutating the residue interacting with G34, Kobayashi *et al.* showed that the initial aminoacylation rate of the cognate $tRNA^{Tyr}_{CUA}$ by the archaeal *Methanococcus jannaschii* TyrRS can be increased[13]. Using the *in vivo* translational assay, we demonstrated here that, despite the different anticodon recognition mode, the prokaryotic TyrRS could also be optimized for its cognate $tRNA^{Tyr}_{CUA}$ that lead to higher incorporation efficiency for Tyr and various unnatural amino acids into proteins.

To test if the enhanced mutant synthetases would facilitate the application of unnatural amino acids in mammalian cells, we used the EAziRS to incorporate Azi into glutathione S-transferase (GST) and attempted photocrosslinking in HEK293T cells. *E. coli* GST is homodimeric[14]. Three sites at the dimer interface, Asn87 in a loop and Tyr92 and Val125 in α-helices were chosen for incorporating Azi (**Figure 3.3a**). GST genes with TAG mutation at these sites and a C-terminal His6 tag were cotransfected with the orthogonal $tRNA^{Tyr}_{CUA}$-EAziRS pair in HEK293 cells, and the cells were grown in the presence of 1 mM of Azi. Photocrosslinking was performed by exposing the cells under 365 nm UV light for 10 min. Cell lysates were separated on SDS-PAGE and probed with an antibody against the His6 tag. As shown in **Figure 3.3b**, GST monomers (23 kD) could be detected for all three mutants before UV exposure. After photocrosslinking, higher molecular weight bands around 46 kD emerged, accompanied by the intensity decrease of the monomer band. Based on the intensity ratio of these bands, the crosslinking efficiency was calculated to be 33%, 40%, and 98% for site 87, 92, and 125, respectively.

Several faint bands between 23 kD and 46 kD were observed on the GST-His6 blot, which can be due to proteins recognized nonspecifically by the penta-His antibody or non-GST proteins crosslinked to GST. To distinguished this, we replaced the His6 tag with the FLAG tag at the C-terminus of GST, and probed the blot with an anti-FLAG antibody after photocrosslinking (**Figure**

3.3c). These bands disappeared, indicating that they are caused by the nonspecificity of the penta-His antibody. Compared with the original TyrRS and AziRS, the ETyrRS and EAziRS increased GST protein yields to 2.2 fold and ~2.3 fold, respectively. More crosslinked GST products were detected in the EAziRS than the AziRS samples (2.1 fold for 87TAG and 1.8 fold for 125TAG). Interestingly, crosslinked GST migrated at different positions in the gel depending on where Azi was incorporated. One possible explanation is that the crosslinking of GST dimer with 2 Azis generates an interlocked polypeptide with a core and 4 arms. Different Azi incorporation sites will change the core size and arm length in the crosslinked product, thus leading to different mobility in gel. This effect may be analogous to how disulfide bonds influence protein mobility in nonreducing electrophoresis conditions. Another possibility to be excluded by mass spectrometry is that GST was crosslinked to other proteins.

These results demonstrate the feasibility of using genetically encoded Azi for detecting protein–protein interactions directly in live mammalian cells. Bzo, but no Azi, has been used to crosslink proteins in mammalian cells before[15]. Benzophenone can be photoactivated at wavelength longer than aryl azides (350–365 nm *vs.* <330 nm)[16]. However, our results showed that Azi could also be activated at the long wavelength 365 nm for crosslinking. One advantage of Azi over Bzo is its smaller size. The bulky side chain of Bzo is more likely to disrupt the protein interaction under study. In the Azi crosslinking experiment, crosslinked products were detected from cell lysate directly. The reported Bzo experiments used immunoprecipitation to enrich the target proteins for detection[15]. Although crosslinking efficiency also depends on target protein and interaction strength, our results suggest that Azi incorporated using the EAziRS enables efficient photocrosslinking in mammalian cells and should prove useful for detecting protein–protein interactions *in vivo*.

In summary, we demonstrated that the recognition of *E. coli* TyrRS for its cognate $tRNA^{Tyr}_{CUA}$ could be optimized by engineering the anticodon-binding region. This optimization could be transplanted to multiple TyrRS-derived synthetases, and the resultant enhanced synthetases improved the incorporation efficiency of various unnatural amino acids in mammalian cells. This strategy should also be generally applicable to other orthogonal tRNA–synthetase pairs, such as the *E. coli* $tRNA^{Lue}$–LeuRS pair and the *Methanosarcina mazei* $tRNA^{Pyl}$–PylRS pair. We also showed that an

enhanced synthetase allowed us to perform *in vivo* photocrosslinking in mammalian cells with high efficiency, which should help detect protein-involved interactions in native cellular surroundings. Efficient incorporation of unnatural amino acids in mammalian cells should enable the effective application of these novel amino acids in addressing a broad range of molecular and cellular biological questions.

3.4 Methods

3.4.1 Materials

For DNA preparation and cloning, DH10B *E. coli* cells were used. Polymerase chain reaction (PCR) was carried out with PhusionTM high-fidelity DNA polymerase (New England Biolabs). Bzo, Azi, and Ome was purchased from Chem-Impex. Pyo was synthesized as described[11]. Other chemicals used in experiments were purchased from Sigma- Aldrich.

3.4.2 Construction of plasmids

All plasmids were cloned using standard methods and was confirmed by DNA sequencing.

The plasmid pSNR-TyrRS expressing *E. coli* tRNATyr and wild-type TyrRS was used in yeast experiments[5].

Mutations were placed in the synthetase by digesting the plasmid with *BamH* I and *Xho* I. *BamH* I is located in the synthetase upstream from Asp265 and *Xho* I is located at the end of the gene. The region containing the mutation was amplified from pSNR-TyrRS using primers JT338 5'-CCA GTT CTG GAT CAA CAC TGC GCA AGC CGA CGT TTA CCG CTT CC-3' (for Asp265Gln), JT339 5'-CCA GTT CTG GAT CAA CAC TGC GTT CGC CGA CGT TTA CCG CTT CC-3' (for Asp265Phe), JT340 5'-CCA GTT CTG GAT CAA CAC TGC GTT GGC CGA CGT TTA CCG CTT CC-3' (for Asp265Leu), JT341 5'-CCA GTT CTG GAT CAA CAC TGC GTA CGC CGA CGT TTA CCG CTT CC-3' (for Asp265Tyr), and JT342 5'-CCA GTT CTG GAT CAA CAC TGC GAG AGC CGA CGT TTA CCG CTT CC-3' (for Asp265Arg) with primer FW22 5'-AAC TCG AGT TAT TTC CAG CAA ATC AGA CAG-3'. The products were re-amplified with primer JT313 5'-GTT GGA TCC

GAA GAA AAC CAG CCC GTA CAA ATT CTA CCA GTT CTG GAT CAA CAC TGC G-3' and FW22 to append the sequence of the synthetase to the *BamH* I site. Amplified sequences were digested with *BamH* I and *Xho* I and inserted at the *BamH* I and *Xho* I site of pSNR-TyrRS.

The BzoRS, AziRS, OmeRS, and PyoRS genes were all constructed from *E. coli* TyrRS gene through site-directed mutagenesis. They contained the following mutations: BzoRS (Y37G, D182G, and L186A); AziRS (Y37L, D182S, F183A and L186A); OmeRS (Y37T, D182T, and L183M); and PyoRS (Y37G, D182S, and F183M). To incorporate unnatural amino acids in mammalian cells, the expression plasmid pEYcua-TyrRS was digested with *BamH* I and *Nhe* I[2]. The mutated synthetase genes were amplified using primer JT313 and EYRS-C 5'-ACA AGA TCT GCT AGC TTA TTT CCA GCA AAT CAG ACA GTA ATT C-3'. The PCR products were digested with *BamH* I and *Nhe* I and inserted into the precut pEYcua-TyrRS to afford pEYcua-BzoRS, pEYcua-AziRS, pEYcua-OmeRS, and pEYcua-PyoRS.

The Asp265Arg mutation was cloned into BzoRS, AziRS, OmeRS, and PyoRS. Using pSNR-ETyrRS containing the TyrRS(Asp265Arg) as a template, the region between the internal *BamH* I site and the *Nhe* I site was amplified using primers JT313 and EYRS-C. The amplified fragment was digested with *BamH* I and *Nhe* I and ligated into the precut pEYcua-BzoRS, pEYcua-AziRS, pEYcua-OmeRS, and pEYcua-PyoRS to make the enhanced version of each synthetase pEYcua-EBzoRS, pEYcua-EAziRS, pEYcua-EOmeRS, and pEYcua-EPyoRS for use in mammalian cells.

Plasmid used for mammalian expression of Glutathion S-transferase (GST) contained a Cytomegolovirus (CMV) promoter and BGH polyA. GST was amplified from *E. coli* genomic DNA using primer GSTfor 5'-TAG TCG GTA CCA TGA AAT TGT TCT ACA AAC CGG GTG CCT GC -3' and primer GST6Hisrev 5'- AGT CCT CGA GTT AGT GGT GGT GGT GGT GGT GCT TTA AGC CTT CCG CTG AC-3'. GST mutants were amplified using overlapping PCR. Internal primers for GST N87TAG were GST87tagFor 5'-CAG TTG CTG GCA CCG GTA TAG AGT ATT TCC CGC TAT AAA ACC ATC-3' and GST87tagRev 5'-TTT ATA GCG GGA AAT ACT CTA TAC CGG TGC CAG CAA CTG GCG GTC-3', for GST Y92TAG were GST92tagFor 5'- GTA AAC AGT ATT TCC CGC TAG AAA ACC ATC GAA TGG CTG AAT TAC-3'and GST92tagRev 3'-TCA GCC ATT CGA

TGG TTT TCT AGC GGG AAA TAC TGT TTA CCG GTG-3', and for GST V125TAG were GST125tagFor 5'-GAA GAG TAC AAA CCG ACA TAG CGC GCG CAG CTG GAG AAG AAG CTG-3' and GST125tagRev 5'-TCT TCT CCA GCT GCG CGC GCT ATG TCG GTT TGT ACT CTT CCG GTG-3'. FLAG-tag GST and GST mutants were amplified from PCR products using primer JT375 5'-GAG GCT AGC GCC ACC ATG AAA TTG TTC TAC AAA CCG GGT GCC-3' and JT376 5'-GAG GGA TCC TTA TTT ATC ATC ATC ATC TTT GTA ATC CTT TAA GCC TTC CGC TGA CAG CG-3'. Amplified products were digested with *Kpn* I and *Xho* I and ligated into precut pCDNA3 (Invitrogen).

3.4.3 Cell culture and transfection

A clonal HeLa GFP-TAG reporter stable cell line[2] and HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Mediatech) supplemented with 10% fetal bovine serum (FBS, Mediatech). Cells were transfected with plasmid DNA using Lipofectamine 2000 according to the protocol of the vendor (Invitrogen).

3.4.4 Flow cytometry

Hela-GFPtag cells seeded in a 3.5 cm culture dish were transfected with 5 μg of pEYcua-aaRS plasmid DNA using Lipofectamine 2000. Hela-GFPtag cells that were not transfected with a tRNA/synthetase pair were used as a negative control. UAA was added 24 h post transfection and FACS analysis was carried out 24 h after the addition of the UAA. Cells were trypsinized and washed with PBS twice. Samples were centrifuged and re-suspended in 1 mL PBS and 10 iL of propidium iodide. Samples were analyzed with a FACScan (Becton & Dickinson). The excitation wavelength was 488 nm, and the emission filter was 530/30 nm. For each sample the total fluorescence intensity of 30,000 cells was recorded, and was normalized to the total fluorescence intensity of cells transfected with pEYcua-TyrRS.

3.4.5 Photocrosslinking and Western blot analysis

293T cells seeded in a 3.5 cm culture dish were transfected with pCDNA-GST-His (TAG mutants) and pEYcua- EAziRS or pEYcua-ETyrRS. Unnatural amino acid Azi was added to the growth media of the appropriate samples in the final concentration of 1 mM. After 48 hours, media was removed and replaced with phosphate buffer saline (PBS). The sample was kept on ice, and a 365 nm UV lamp was placed 1 cm above the sample. Samples were exposed under UV light for 10 minutes. Cells were removed by pipette, lysed on ice with 30 ìL of 1% Nonidet P-40 (NP40, Calbiochem) in PBS. After centrifugation, 5 ìL of supernatant from cells transfected with pEYcua-ETyrRS and 15 ìL of pEYcua-EAziYRS supernatant were loaded and separated by 12% SDS-PAGE. An antibody (Penta-His HRP conjugate Kit, Qiagen) against the His6 tag was used to detect GST proteins before and after photocrosslinking.

For FLAG-tag westerns, 293T cells were seeded in a 10 cm dish for transfection with pCDNA-GST-FLAG (TAG mutants) and pEYcua-TyrRS, pEYcua-ETyrRS, pEYcua-AziRS or pEYcua-EAziRS. Crosslinking was carried out under UV light for 15 minutes. Cells were removed by 1 mL cold PBS and lysed by 3 rounds of liquid nitrogen to 37°C water bath freeze-thaw. After centrifugation, the supernatant was loaded onto 15 ìL packed volume of glutathione sepharose 4B (GE Healthcare) and GST was purified according to the protocol of the vendor. After purification GST proteins were eluted and boiled in SDS loading buffer. Five microliters from pEYcua-TyrRS or pEYcua-ETyrRS transfected samples and 15 ìL from pEYcua-AziYRS or pEYcua-EAziYRS transfected samples were loaded and separated by 12% SDS-PAGE. The primary antibody (Monoclonal ANTI-FLAG M2, Sigma- Aldrich) against the FLAG tag and HRP conjugated secondary antibody (Goat Anti-Mouse IgG (H+L)-HRP Conjugate, Bio-Rad) was used to detect GST proteins.

3.5 ACKNOWLEDGEMENTS

We thank Nikki Dellas for help with the superposition of the TyrRS structures, and Dr Gordon V. Louie for help with the preparation of **Figure 3.1 and 3.3a**. L.W. thanks the support of the Searle Scholar Program, the Beckman Young Investigator Program, the Ray Thomas Edwards Foundation, and the March of Dimes Foundation.

Chapter 3 is a reprint in full of the material as it appears in Molecular Biosystems, (2009), Takimoto, J. K.; Adams, K. L.; Xiang, Z.; Wang, L., Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol Biosyst* 2009, *5* (9), 931-4.

3.6 FIGURES

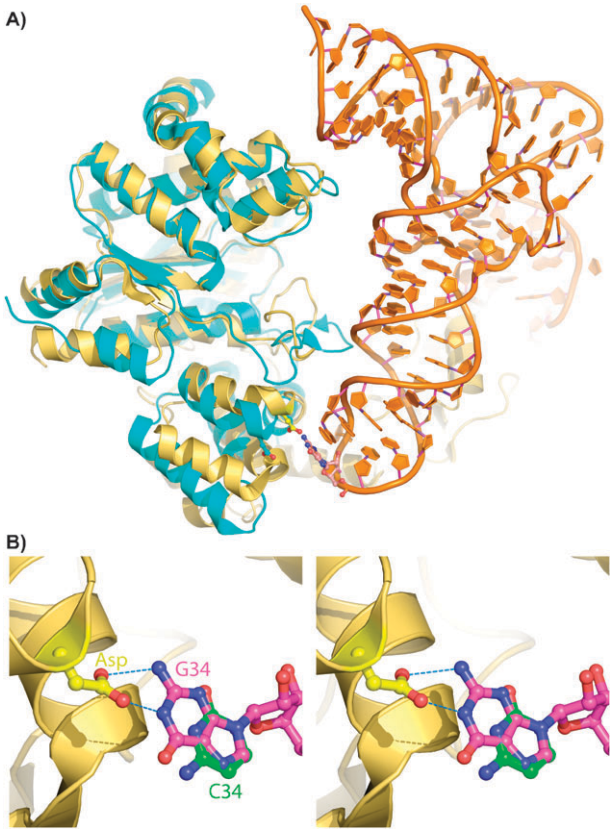


Figure 3.1: (A) Superposition of *E. coli* TyrRS (PDB ID 1X8X, cyan) on *T. thermophilus* $tRNA^{Tyr}$–TyrRS complex (PDB ID 1H3E, yellow and orange). Only one subunit of the dimeric TyrRS and one tRNA are shown. Base G34 on the tRNA and the interacting Asp on TyrRS are represented as ball-and-stick. (B) Recognition of G34 by Asp259 in the *T. thermophilus* $tRNA^{Tyr}$–TyrRS complex. C34 was modeled to show the gap after G34C change.

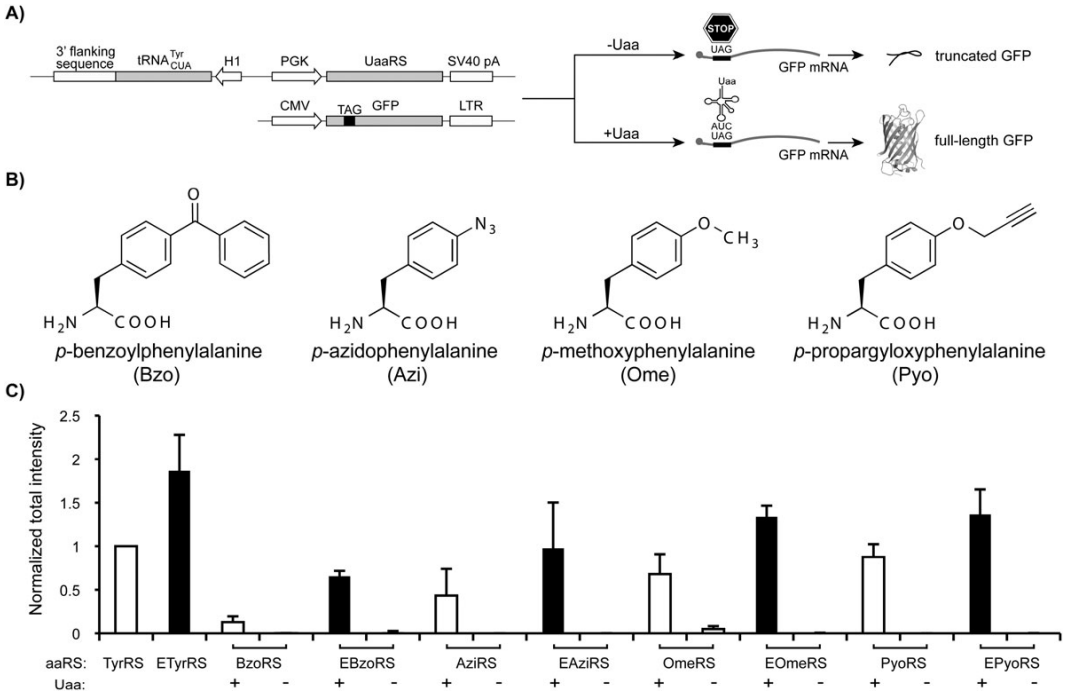


Figure 3.2: Enhanced mutant synthetases increased the incorporation efficiency of various unnatural amino acids in mammalian cells. (A) An *in vivo* fluorescence assay for measuring the amber suppression efficiency of the expressed orthogonal tRNA-synthetase. Uaa: unnatural amino acid. (B) Structures of the unnatural amino acids. (C) Fluorescence assay results for the incorporation efficiency of different synthetases. Error bars represent sem $n = 3$. Mutants used here were: EBzoRS (Y37G, D182G, L186A, and D265R), EAziRS (Y37L, D182S, F183A, L186A, and D265R), EOmeRS (Y37T, D182T, L183M, and D265R), and EPyoRS (Y37G, D182S, F183M, and D265R). All mutations are in reference to the wt *E. coli* TyrRS.

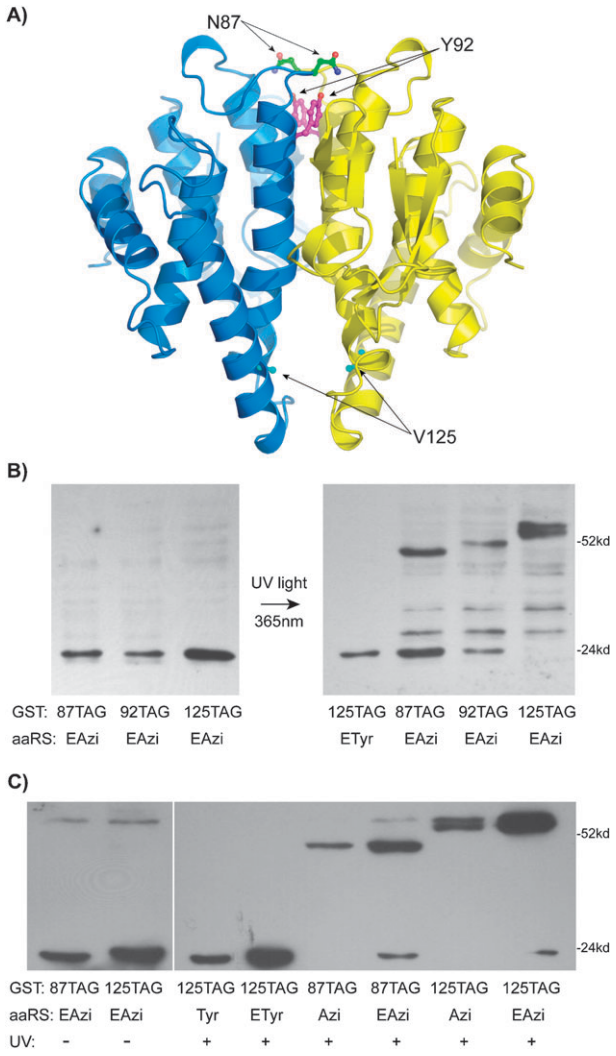


Figure 3.3: Photocrosslinking in mammalian cells using Azi. (A) Crystal structure of *E. coli* GST (PDB ID 1A0F) with three sites for Azi incorporation labeled. (B) Western blot analysis of GST mutants using a penta-His antibody before and after photocrosslinking. (C) Western blot analysis of GST mutants using an anti-FLAG antibody before and after photocrosslinking.

3.7 REFERENCES

1. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.

2. Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L., Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat Neurosci* **2007,** *10* (8), 1063-72.

3. Chen, S.; Schultz, P. G.; Brock, A., An improved system for the generation and analysis of mutant proteins containing unnatural amino acids in Saccharomyces cerevisiae. *J Mol Biol* **2007,** *371* (1), 112-22.

4. Liu, W.; Brock, A.; Chen, S.; Schultz, P. G., Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* **2007,** *4* (3), 239-44.

5. Wang, Q.; Wang, L., New methods enabling efficient incorporation of unnatural amino acids in yeast. *J Am Chem Soc* **2008,** *130* (19), 6066-7.

6. Ibba, M.; Soll, D., Aminoacyl-tRNA synthesis. *Annu Rev Biochem* **2000,** *69*, 617-50.

7. Kobayashi, T.; Takimura, T.; Sekine, R.; Kelly, V. P.; Kamata, K.; Sakamoto, K.; Nishimura, S.; Yokoyama, S., Structural snapshots of the KMSKS loop rearrangement for amino acid activation by bacterial tyrosyl-tRNA synthetase. *J Mol Biol* **2005,** *346* (1), 105-17.

8. Yaremchuk, A.; Kriklivyi, I.; Tukalo, M.; Cusack, S., Class I tyrosyl-tRNA synthetase has a class II mode of cognate tRNA recognition. *EMBO J* **2002,** *21* (14), 3829-40.

9. Krissinel, E.; Henrick, K., Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr D Biol Crystallogr* **2004,** *60* (Pt 12 Pt 1), 2256-68.

10. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G., An expanded eukaryotic genetic code. *Science* **2003,** *301* (5635), 964-7.

11. Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G., Adding amino acids with novel reactivity to the genetic code of Saccharomyces cerevisiae. *J Am Chem Soc* **2003,** *125* (39), 11782-3.

12. Tsunoda, M.; Kusakabe, Y.; Tanaka, N.; Ohno, S.; Nakamura, M.; Senda, T.; Moriguchi, T.; Asai, N.; Sekine, M.; Yokogawa, T.; Nishikawa, K.; Nakamura, K. T., Structural basis for recognition of cognate tRNA by tyrosyl-tRNA synthetase from three kingdoms. *Nucleic Acids Res* **2007,** *35* (13), 4289-300.

13. 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), 425-32.

14. Nishida, M.; Harada, S.; Noguchi, S.; Satow, Y.; Inoue, H.; Takahashi, K., Three-dimensional structure of Escherichia coli glutathione S-transferase complexed with glutathione sulfonate: catalytic roles of Cys10 and His106. *J Mol Biol* **1998,** *281* (1), 135-47.

15. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S., Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. *Nat Methods* **2005,** *2* (3), 201-6.

16. Geiger, M. W.; Elliot, M. M.; Karacostas, V. D.; Moricone, T. J.; Salmon, J. B.; Sideli, V. L.; Onge, M. A. S., Azides as protein photolabels: absorption spectral properties and quantum yields of photodissociation. *Photochemistry and Photobiology* **1984,** *40* (4), 545-548.

Chapter 4

Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells

4.1 Abstract

It's a cover up: Unnatural amino acids (UAAs) with noncanonical side chains can be efficiently transported into mammalian cells when the carboxyl group of the UAA is masked with selected ester groups. This greatly increases the UAA's uptake rate and intracellular concentration, thus reducing the amount of UAA required.

4.2 Results and discussion

Genetically encoded unnatural amino acids (UAAs) enable novel chemical and physical properties to be selectively introduced into proteins directly in live cells; this provides great potential for addressing biological questions at the molecular and cellular level in native settings[1-3]. UAAs have been genetically incorporated into proteins in mammalian cells by using orthogonal tRNA-codon–synthetase sets[4-6], yet the current low incorporation efficiency hinders their effective application. Efforts to improve efficiency have focused on optimizing the expression and activity of the orthogonal tRNA and synthetase[5-7], whereas the bioavailability of the UAA inside mammalian cells, a prerequisite for incorporation, has not been addressed. In addition, there are many UAAs, such as glycosylated and phosphorylated amino acids, that have not been genetically incorporated into proteins successfully. These amino acids can be invaluable for studying the contribution of post-translational modifications of protein function and the role of a target protein in cellular signal transduction. Among many reasons, the inability of the UAA to enter cells prevents the evolution of a mutant synthetase specific for the UAA by using cell-based selections or screens[2]. Here, we show that an UAA that deviates structurally from the canonical amino acids in side chain could not be efficiently transported into mammalian cells, but that masking the carboxyl group of the UAA as an ester greatly increased the rate of cellular uptake and intracellular concentration of the UAA. This resulted in a significant increase in the incorporation of this UAA into proteins in mammalian cells. Among three esters tested, acetoxymethyl ester (AME) yielded the highest UAA incorporation efficiency, with a concomitant reduction in the UAA required in the growth medium.

UAAs with side chains similar to canonical amino acids can be transported into cells by endogenous amino acid transporters[3, 8], which are relatively nonspecific for substrates[9]. However, UAAs that deviate significantly from canonical amino acids in side chain structure might not be recognized by these transporters. Cell membranes are more permeable to neutral than charged molecules. As zwitterions, UAAs have only a very small proportion present in the neutral form at physiological pH, thereby making it difficult for them to cross the cell membrane. Esters have been widely used in prodrugs to derivatize carboxyl, hydroxyl, and thio functionalities so as to make the parent drugs more membrane permeable[10]. In particular, Tsien et al. pioneered the use of AME to enhance the cellular uptake of various charged molecules, such as carboxylate-containing Ca^{2+} chelators and phosphate-containing second messengers[11, 12]. We reasoned that masking the carboxyl group of the UAA with an ester would convert the UAA into a protonated weak base, which has a higher percentage of neutral form and increased lipophilicity for crossing the membrane. Once inside the cell, intracellular esterases can cleave the ester to regenerate the original UAA for incorporation.

We demonstrated this strategy with the UAA 2-amino-3-(5-(dimethylamino)naphthalene-1-sulfonamide)propanoic acid (DanAla, **1**), which does not resemble and is bulkier in size than any canonical amino acid. Its fluorescence property also makes it an attractive candidate to develop optical reporters for imaging protein activities in live cells. We synthesized the methyl, ethyl, and acetoxymethyl esters of DanAla by using the methods shown in **Scheme 1b**. Compound **5** was synthesized from Boc-Dap-OH and dansyl chloride according to the known procedure[13]. Alkylation of **5** with iodomethane and iodoethane gave intermediates **6** and **7**, which afforded DanAla-OMe (**2**) and DanAla-OEt (**3**), respectively, after deprotection. However, in the presence of bromomethyl acetate and *N*,*N*-diisopropylethylamine (DIPEA), compound **5** was transformed to the undesired cyclization product **8**, instead of Boc-DanAla-OAM. Therefore, the carboxyl group and the sulfonamide group were protected with the benzyl group and the Boc group[14], respectively. After debenzylation, carboxylic acid **10** was transformed into acetoxymethyl ester **11**[15], which furnished DanAla-OAM (**4**) after removal of the Boc groups.

To determine if the DanAla esters could enhance the incorporation of DanAla into proteins in mammalian cells, we used an in cellulo fluorescence assay to evaluate the incorporation efficiency of DanAla[6, 7]. The incorporation of DanAla into green fluorescent protein (GFP) was measured by using a stable clonal HeLa cell line, into which a GFP gene containing a premature UAG stop codon at a permissive site (Tyr182) was integrated. The cells were transfected with the orthogonal tRNA–synthetase pair specific for DanAla[13], and the DanAla or DanAla ester was added to the growth medium. The incorporation of DanAla at the UAG182 position results in full-length, fluorescent GFP, whereas no incorporation results in a truncated, nonfluorescent protein. The total fluorescence intensity of cells was measured by flow cytometry analysis. As shown in **Figure 4.1a**, the cell fluorescence intensity increased when more DanAla was added to the growth medium, but reached a plateau from 0.25 to 1.00 mM. The plateau suggests that either the concentration of DanAla inside the cells has saturated the synthetase or that the cellular availability of DanAla is limiting. When DanAla-OMe was used, the cell fluorescence intensity doubled in comparison to that of cells incubated with the same concentrations of DanAla; this indicates that the cellular availability of DanAla was the limiting factor. Similar results were also obtained for DanAla-OEt. There was no significant increase in cell fluorescence intensity when the concentration of DanAla-OMe or DanAla-OEt was increased from 0.10 to 0.25 mM. In contrast, DanAla-OAM doubled the fluorescence intensity at 0.10 mM and quadrupled it at 0.25 mM. In comparison to 1.00 mM of DanAla, the amount often used in UAA incorporation, DanAla-OAM increased the cell fluorescence intensity fourfold in addition to requiring 75 % less compound in the growth medium (0.25 mM). Western blot analysis of the cell lysates confirmed the increase in the amount of GFP produced by the AME modification (**Figure 4.1b**). These results suggest that all three ester modifications were able to increase DanAla incorporation into proteins with a concomitant reduction in the extracellular supply of UAA. Furthermore, we identified DanAla-OAM as the most effective out of the three esters tested.

To understand how DanAla-OAM increases DanAla incorporation, we first monitored the uptake of DanAla and DanAla-OAM into cells by using fluorescence microscopy. We added 0.10 mM of DanAla or DanAla-OAM to the media of HEK293T cells, and measured the cytosolic

fluorescence intensity of cells at sequential time points by using confocal microscopy (**Figure 4.2a**). For cells incubated with DanAla, the cytosolic fluorescence intensity was almost flat in the range of 620–750 AU. A striking difference was observed in the DanAla-OAM incubated sample. Within 5 min, the intracellular fluorescence intensity rapidly rose to 1100 AU, a 48 % increase from that of the DanAla sample. The intensity increased for 3 h peaking at 1613 AU, which was twice the DanAla peak intensity. The intensity of DanAla-OAM treated cells then gradually dropped, possibly due to the depletion of the extracellular DanAla-OAM in combination with the equilibration of converted DanAla in and outside of cells. The AME modification thus significantly accelerates the cellular uptake rate of the compound.

Next, we quantified the intracellular concentrations of DanAla and DanAla-OAM by using HPLC. HEK293T cells were incubated with 0.10 mM of the compounds for 1 h, and cell contents were extracted. Small molecules in the cell extracts were separated by HPLC, and peaks corresponding to DanAla and DanAla-OAM were verified by using pure compounds and mass spectrometry. Surprisingly, after only 1 h of incubation, cells incubated with DanAla-OAM showed no peak for DanAla-OAM but a large peak for DanAla (**Figure 4.2b**); this indicates that all intracellular DanAla-OAM had been hydrolyzed into DanAla. The peak area for DanAla was used to determine its intracellular concentration. Cells incubated with DanAla and DanAla-OAM had 0.28 and 8.9 mM intracellular concentration of DanAla, respectively. The AME modification dramatically increased the intracellular concentration of DanAla by 31-fold.

When the concentration of DanAla-OAM was further increased to 0.50 mM, the cell fluorescence intensity decreased unexpectedly (**Figure 4.1a**). We used propidium iodide (PI) staining to assess the health of cells incubated with DanAla or DanAla-OAM (**Figure 4.2c**). For cells treated with 0.10 to 0.50 mM DanAla, the percentage of PI-positive cells was similar to that of cells without UAA. A slight increase of PI-positive cells was observed at 1.00 mM of DanAla. In contrast, cells treated with DanAla-OAM showed a significantly higher percentage of PI-positive cells, especially at the concentration of 0.50 mM. DanAla-OAM hydrolysis releases formaldehyde, which will negatively affect cell health at high concentration[11]. High intracellular concentration of DanAla could be toxic to

cells as well. Therefore, an optimal concentration of the AME ester should be used to achieve highest UAA incorporation efficiency with minimal negative effect to cell health.

In summary, we have demonstrated that masking the carboxyl group of DanAla with AME increased the uptake rate and intracellular concentration of the unnatural amino acid in mammalian cells. This resulted in a higher DanAla incorporation efficiency accompanied by the added benefit of reducing the amount of UAA in the growth medium. Efficient UAA incorporation could prove valuable for the effective application of genetically encoded UAAs to studying various biological processes in mammalian cells. Modification of UAAs with AME could be generally applicable to other UAAs that have trouble entering mammalian cells. To test this hypothesis, more UAAs need to be modified, and their incorporation efficiency determined. However, there is currently no orthogonal tRNA–synthetase available for the incorporation of these candidate UAAs because a synthetase cannot be evolved for an UAA that cannot readily enter the cell. We are applying the esterification strategy to highly polar amino acids, such as glycosylated and phosphorylated amino acids, so as to overcome this problem with the final goal of genetically incorporating these biologically important amino acids into proteins.

4.3 Experimental section

4.3.1 Synthesis of DanAla and DanAla esters

All reactions were carried out under nitrogen with dry solvents in anhydrous conditions, unless otherwise noted. Dry *N*,*N*-dimethylformamide (DMF), methanol, methylene chloride, and acetonitrile were obtained by passing commercially available predried, oxygen-free formulations through activated alumina columns. Yields refer to chromatographically homogeneous materials. Reagents were purchased from Sigma–Aldrich and used without further purification. Reactions were monitored by TLC, and carried out on 0.25 mm Silicycle silica gel plates (60F-254) by using UV light for visualization and an ethanolic solution of phosphomolybdic acid and cerium sulfate for developing under heat. Silicycle silica gel (60, particle size 0.040–0.063 mm) was used for flash column chromatography (FCC). NMR spectra were recorded on a Varian Oxford AS500 instrument and calibrated by using residual undeuterated solvent or tetramethylsilane as an internal reference. LCMS

experiments were performed on an Agilent 1100 Series LC/MSD instrument with a Phenomenex Synergi 4u Fusion-RP 80 A column (150×4.6 mm). DanAla was synthesized as described[13].

NMR (^{1}H and ^{13}C) and HRMS data for compounds are given in the Supporting Information.

Compound 5: Boc-Dap-OH (2.0 g, 9.8 mmol) was added in one portion to a stirring solution of dansyl chloride (2.4 g, 8.9 mmol) and Et_3N (2.6 mL) in CH_2Cl_2 (50 mL) at 0 °C. The reaction mixture was allowed to warm to room temperature. After being stirred for 12 h, the reaction mixture was concentrated under reduced pressure. The residue was purified by FCC ($MeOH/CH_2Cl_2$ 7:93, *v*/*v*) to give compound5 (3.675 g, 94 %) as a yellow solid. R_f=0.22 (SiO_2, $MeOH/CHCl_3$ 1:8, *v*/*v*).

Compound 6: Compound 5 (218.8 mg, 0.50 mmol) was dissolved in DMF (2 mL), and the mixture was cooled to 0 °C. DIPEA (0.096 mL, 0.55 mmol) and MeI (0.062 mL, 1.00 mmol) were added dropwise to the solution. The reaction mixture was allowed to warm to room temperature and stirred for 12 h. Water (20 mL) was added, and the solution was extracted with ethyl acetate (3×20 mL). The organic phase was washed with brine, dried with sodium sulfate, and concentrated under reduced pressure. The residue was purified by FCC (ethyl acetate/hexanes 3:7, *v*/*v*) to give compound 6 (197.4 mg, 87 %) as a yellow solid.

Compound 7: Compound 5 (218.8 mg, 0.50 mmol) was dissolved in DMF (2 mL), and the mixture was cooled to 0 °C. DIPEA (0.096 mL, 0.55 mmol) and EtI (0.080 mL, 1.00 mmol) were added dropwise to the solution. The reaction mixture was allowed to warm to room temperature and stirred for 12 h. Water (20 mL) was added, and the solution was extracted with ethyl acetate (3×20 mL). The organic phase was washed with brine, dried with sodium sulfate, and concentrated under reduced pressure. The residue was purified by FCC (ethyl acetate/hexanes 3:7, *v*/*v*) to give compound 7 (174.1 mg, 75 %) as a yellow solid.

Compound 2: Compound 6 (45.2 mg, 0.10 mmol) was dissolved in CH_2Cl_2 (0.9 mL), and the mixture was cooled to 0 °C. Trifluoroacetic acid (TFA; 0.3 mL) was added dropwise to the solution. The reaction mixture was allowed to warm to room temperature and stirred for 4 h, then concentrated under reduced pressure. After being under vacuum for 48 h, the residue was dissolved in dimethyl sulfoxide (DMSO; 1.0 mL) for use.

Compound 3: Compound 7 (46.6 mg, 0.10 mmol) was dissolved in CH_2Cl_2 (0.9 mL), and the mixture was cooled to 0 °C. TFA (0.3 mL) was added dropwise to the solution. The reaction mixture was allowed to warm to room temperature and stirred for 4 h, then concentrated under reduced pressure. After being under vacuum for 48 h, the residue was dissolved in DMSO (1.0 mL) for use.

Compound 9: Compound 5 (569.4 mg, 1.30 mmol) was dissolved in DMF (10 mL), and the mixture was cooled to 0 °C. DIPEA (0.25 mL, 1.43 mmol) and benzyl bromide (0.31 mL, 2.60 mmol) were added dropwise to the solution. The reaction mixture was allowed to warm to room temperature and stirred for 24 h. Water (70 mL) was added and the solution was extracted with ethyl acetate (50 mL). The organic phase was washed with brine, dried with sodium sulfate, and concentrated under reduced pressure. The residue was purified by FCC (ethyl acetate/hexanes 1:3–1:2, *v*/*v*) to give Boc-DanAla-OBn (522.1 mg, 76 %) as a yellow solid. R_f=0.25 (SiO_2, ethyl acetate/hexanes 1:2,*v*/*v*). Boc-DanAla-OBn (466.7 mg, 0.885 mmol) and 4-dimethylaminopyridine (DMAP; 21.6 mg, 0.177 mmol) were dissolved in CH_2Cl_2 (4 mL), and the mixture was cooled to 0 °C. Et_3N (0.136 mL, 0.974 mmol) and Boc_2O (289.8 mg, 1.328 mmol) in CH_2Cl_2 (2 mL) were added dropwise to this stirring mixture. The reaction mixture was allowed to warm to room temperature and stirred for 4 h. Saturated aqueous NH_4Cl solution was added to the mixture. The aqueous phase was separated and washed with CH_2Cl_2 (3×10 mL). The combined organic phase was washed with brine, dried with sodium sulfate, and concentrated under reduced pressure. The residue was purified by FCC (ethyl acetate/hexanes 1:3, *v*/*v*) to give compound 9 (538.6 mg, 97 %). R_f=0.34 (SiO_2, ethyl acetate/hexanes 1:2, *v*/*v*).

Compound 10: Pd/C (87.3 mg) was added to a solution of compound 9 (515.0 mg, 0.82 mmol) in methanol. The reaction mixture was stirred under hydrogen for 2 h at room temperature. The reaction mixture was passed through a short plug of celite and eluted with ethyl acetate. The filtrate was concentrated to give compound 10 (467.3 mg) as a yellow-green solid. R_f=0.22 (SiO_2, MeOH/$CHCl_3$ 1:10, *v*/*v*).

Compound 11: Compound 10 (434.9 mg, 0.809 mmol) was dissolved in CH_3CN, and the mixture was cooled to 0 °C. DIPEA (0.56 mL, 3.236 mmol) and bromomethyl acetate (0.24 mL, 2.427 mmol) were added dropwise to the solution. The reaction mixture was allowed to warm to room

temperature and stirred for 12 h, then concentrated under reduced pressure. The residue was purified by FCC (ethyl acetate/hexanes 1:2, *v*/*v*) to give compound 11 (389.5 mg, 84 % over 2 steps) as yellow-green solid. R_f=0.31 (SiO_2, ethyl acetate/hexanes 1:2, *v*/*v*).

Compound 4: Compound 11 (61.0 mg, 0.10 mmol) was dissolved in CH_2Cl_2 (0.9 mL), and the mixture was cooled to 0 °C. TFA (0.3 mL) was added dropwise to the solution. The reaction mixture was allowed to warm to room temperature and stirred for 6 h, then concentrated under reduced pressure. After being under vacuum for 48 h, the residue was dissolved in DMSO (1.0 mL) for use.

4.3.2 Cell lines and culture conditions

HEK293T cells were used for determining the UAA intracellular concentration and imaging experiments. A stable clonal HeLa cell line containing the GFP gene with Tyr182UAG mutation was previously established in this lab and was used here for assaying UAA incorporation efficiency[6]. HEK293T and HeLa-GFP(Tyr182UAG) cells were cultured in Dulbecco's modified Eagle's medium (DMEM, Mediatech) supplemented with 10 % fetal bovine serum (FBS).

4.3.3 Fluorescence imaging

HEK293T cells (3×10^5) were seeded into an imaging dish supplemented with DMEM with 10 % FBS, but without phenol red. After 18–24 h, DanAla or DanAla-OAM was added to the medium to a final concentration of 0.10 mM. Cells were immediately imaged with an Olympus IX81 spinning disc microscope with a Hamamatsu EM-CCD under the same conditions, λ_{ex}=360±30 nm,λ_{em}=535±20 nm. After each time point, cells were placed back into the incubator. Images were analyzed in Slidebook 4.2 by using the masking tool and average intensity function. Cells not treated with any compound were used as control to determine background fluorescence, which was subtracted from the measured intensities of samples to yield the final net intensities.

4.3.4 HPLC analysis of intracellular concentration

HEK293T cells (8.5×10^6) were seeded into a 10 cm tissue culture dish supplemented with DMEM containing 10 % FBS. After 18–24 h, DanAla or DanAla-OAM was added to the medium to a final concentration of 0.1 mM. The compounds were incubated with cells for 1 h. Cells were then washed with phosphate-buffered saline (PBS; 1 mL) several times. After trypsinization with trypsin (1 mL, 0.05 %; Mediatech), cells were centrifuged, and the medium was removed. Cells were washed again with water (1 mL) and centrifuged. The supernatant was removed, and cells were resuspended in water (250 μL). Toluene (50 μL) was added, and the cells were placed in a shaking incubator at 37 °C for 30 min. Cells were centrifuged at 20 000 relative centrifugal force (RCF) for 10 min. The aqueous layer was placed into a Microcon Ultracel YM10 column (Millipore) and spun for 30 min at 12 000 RCF. The flow-through was analyzed by HPLC-MS (Agilent 1100 Series LC/MSD instrument): and aliquot (5 μL) of the aqueous layer was separated on a Phenomenex Synergi 4u Fusion-RP 80 A column (150×4.6 mm) with a gradient of aqueous acetonitrile/0.1 % formic acid (75:25 to 25:75) over 10 min. DanAla and DanAla-OAM were identified by extracting their MW (+1) from the total ion mass spectrum. Pure DanAla and DanAla-OAM were also added into the flow-though prepared from cells not treated with compounds so as to verify the peak positions of DanAla (3.3 min) and DanAla-OAM (3.7 min). Calculation of DanAla concentrations in each sample was based on the extracted ion area of DanAla and a standard curve. The standard curve was made by extracting the area of DanAla in solutions with different concentrations (0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, and 1.00 mM). Cells not treated with any compound were used to measure the background area at 3.3 and 3.7 min. The intracellular concentration was given by the total molar amount of DanAla divided by the total volume of cells, by using the volume of a single cell[16].

4.3.5 Flow cytometry

Stable HeLa–GFP(Tyr182TAG) clonal cells were seeded in a 3.5 cm culture dish and transfected with pELcua-LeuRS or pELcua-DanAlaRS (5 μg) plasmid DNA by using Lipofectamine 2000[6]. HeLa–GFP(Tyr182TAG) cells that were not transfected with a tRNA–synthetase pair were used

as a negative control. DanAla and DanAla-OAM were added 24 h post transfection, and flow cytometry was carried out 24 h after the addition of the compounds. Cells were trypsinized and washed with PBS twice. Samples were centrifuged and resuspended in PBS (1.0 mL) and PI (5 μL). Samples were analyzed with a FACScan (Becton, Dickinson and Company, Franklin Lakes, USA). The excitation wavelength was 488 nm, and the emission filter was 530/30 nm. The excitation light (488 nm) excites GFP but not DanAla. For each sample, the total fluorescence intensity of 30 000 cells was recorded, and was normalized to the total fluorescence intensity of cells transfected with pELcua-LeuRS. PI (5 $\mu g\,mL^{-1}$, Sigma–Aldrich) was used to determine cell viability by flow cytometry. Cells were trypsinized and incubated with PI for 15 min at room temperature, and then analyzed by using FACScan (Becton, Dickinson and Company).

4.3.6 Western blot

HeLa–GFP(Tyr182TAG) clonal cells were transfected and incubated with the appropriate DanAla or DanAla ester, as previously described for flow cytometry analysis. Cells were trypsinized and washed with PBS, and cell number was counted by using a hemocytometer. Samples were centrifuged and resuspended with PBS (20 μL) containing ethylendiaminetetraacetate-free protease inhibitor cocktail (Roche) and DNase I (Roche). These samples were lysed by flash freezing in liquid nitrogen, and thawed by sonication. The lysis procedure was repeated three times to ensure complete lysis. Loading buffer was added, and the samples were boiled. Prepared samples were loaded onto a 12 % SDS-PAGE gel. For the pELcua-LeuRS sample, 10^4 cells were loaded, while 7.5×10^4 cells were loaded for all other samples. The primary antibody (Anti-GFP Monoclonal Antibody 7G9, BioPioneer, San Diego, USA) and horseradish peroxidase (HRP) conjugated secondary antibody (Goat Anti-Mouse IgG-HRP Conjugate, Santa Cruz Biotechnology) were used to detect GFP proteins.

4.4 ACKNOWLEDGEMENTS

We thank Dr. Michael Burkart for help with the HRMS. This work was supported by the Ray Thomas Edwards Foundation, Beckman Young Investigator Program, March of Dimes Foundation (5-FY08-110), California Institute for Regenerative Medicine (RN1-00577-1, and National Institutes of Health (1DP2OD004744-01).

Chapter 4 is a reprint in full of the material as it appears in Chembiochem, (2010), Takimoto, J. K.; Xiang, Z.; Kang, J. Y.; Wang, L., Esterification of an unnatural amino acid structurally deviating from canonical amino acids promotes its uptake and incorporation into proteins in mammalian cells. *Chembiochem 11* (16), 2268-72. The dissertation author and Xiang, Z contributed equally to this manuscript.

4.5 Supplemental information

4.5.1 Experimental Data for Compounds

Compound 6. 1H NMR (500 MHz, CDCl3): δ = 8.56 (d, *J* = 8.5 Hz, 1 H), 8.25-8.22 (m, 2 H), 7.57-7.50 (m, 2 H), 7.18 (d, *J* = 7.0 Hz, 1 H), 5.45 (t, *J* = 6.5 Hz, 1 H), 5.39 (d, *J* = 6.5 Hz, 1 H), 4.29 (brs, 1 H), 3.59 (s, 3 H), 3.35-3.26 (m, 2 H), 2.89 (s, 6 H), 1.40 ppm (s, 9 H); 13C NMR (125 MHz, CDCl3): δ = 170.7, 155.5, 152.0, 134.4, 130.8, 130.0, 129.9, 129.6, 128.8, 128.6, 128.4, 123.3, 118.8, 115.4, 80.5, 53.5, 52.8, 45.5, 44.9, 28.2 ppm. (**Figure 4.3**) (**Figure 4.4**)

Compound 7. 1H NMR (500 MHz, CDCl3): δ = 8.55 (d, *J* = 8.5 Hz, 1 H), 8.25-8.22 (m, 2 H), 7.56 (t, *J* = 8.0 Hz, 1 H), 7.52 (t, *J* = 8.0 Hz, 1 H), 7.19 (d, *J* = 7.5 Hz, 1 H), 5.39 (t, *J* = 6.5 Hz, 1 H), 4.26 (brs, 1 H), 4.14-4.05 (m. 1 H), 4.03-3.97 (m, 1 H), 3.35-3.30 (m, 1 H), 3.28-3.23 (m, 1 H), 2.89 (s, 6 H), 1.52 (s, 9 H), 1.17 ppm (t, *J* = 7.2 Hz, 3 H); 13C NMR (125 MHz, CDCl3): δ = 170.2, 155.5, 152.0, 134.4, 130.8, 130.0, 129.9, 129.6, 128.7, 123.3, 118.9, 115.5, 80.5, 62.2, 53.6, 45.6, 45.0, 28.3, 14.1 ppm. (**Figure 4.5**) (**Figure 4.6**)

Compound 2. 1H NMR (500 MHz, [D6]DMSO): δ = 8.51 (dt, *J* = 8.5, 1.0 Hz, 1 H), 8.48 (brs, 3 H), 8.37 (t, *J* = 6.5 Hz, 1 H), 8.25 (d, *J* = 8.5 Hz, 1 H), 8.11 (dd, *J* = 7.2, 1.0 Hz, 1 H), 7.67 (dd, *J* = 8.8, 7,2 Hz, 1 H), 7.63 (dd, *J* = 8.5, 7.5 Hz, 1 H), 7.31 (dd, *J* = 8.0, 1.0 Hz, 1 H), 4.09 (m, 1 H), 3.60 (s, 3 H), 3.26 (ddd, *J* = 14.5, 7.0, 5.5 Hz, 1 H), 3.18 (dt, *J* = 14.5, 6.0 Hz, 1 H), 2.85 ppm (s, 6 H); 13C NMR (125 MHz, [D6]DMSO): δ = 167.8, 158.4 (q, *J* = 36.5 Hz), 151.2, 134.9, 129.9, 129.1, 128.9, 128.7, 128.2, 123.7, 119.1, 116.8, 115.5, 114.5, 52.9, 52.3, 45.1, 42.3 ppm. HRMS (ESI-FT): calcd for C16H22N3O4S: 352.1326 [*M*+1], found 352.1328. (**Figure 4.7**) (**Figure 4.8**)

Compound 3. 1H NMR (500 MHz, [D6]DMSO): δ = 8.51 (dd, *J* = 8.0, 1.0 Hz, 1 H), 8.48 (brs, 3 H), 8.34 (t, *J* = 6.2 Hz, 1 H), 8.28 (d, *J* = 8.5 Hz, 1 H), 8.12 (dd, *J* = 7.0, 1.0 Hz, 1 H), 7.67 (dd, *J* = 8.8, 7.2 Hz, 1 H), 7.63 (dd, *J* = 8.5, 7.5 Hz, 1 H), 7.32 (d, *J* = 7.5 Hz, 1 H), 4.11 (dq, *J* = 11.0, 7.0 Hz, 1 H), 4.08 (m, 1 H), 4.00 (dq, *J* = 11.0, 7.0 Hz, 1 H), 3.27 (ddd, *J* = 14.5, 6.5, 5.5 Hz, 1 H), 3.19 (dt, *J* = 14.5, 5.5 Hz, 1 H), 2.86 (s, 6 H), 1.15 ppm (t, *J* = 7.0 Hz, 3 H); 13C NMR (125 MHz, [D6]DMSO): δ = 167.4, 158.6 (q, *J* = 36.5 Hz), 151.0, 135.0, 129.9, 129.1, 129.0, 128.8, 128.2, 123.8, 119.3, 116.8, 115.6, 62.1, 52.3, 45.2, 42.4, 13.8 ppm. HRMS (ESI-FT): calcd for C17H24N3O4S: 366.1482 [*M*+1],

found 366.1481. (**Figure 4.9**) (**Figure 4.10**)

Boc-DanAla-OBn. 1H NMR (500 MHz, CDCl3): δ = 8.53 (d, *J* = 8.5 Hz, 1 H), 8.22-8.20 (m, 2 H), 7.53 (t, *J* = 8.0 Hz, 1 H), 7.48 (d, *J* = 8.0 Hz, 1 H), 7.35-7.32 (m, 3 H), 7.27-7.26 (m, 2 H), 7.17 (d, *J* = 7.5 Hz, 1 H), 5.41 (brs, 2 H), 5.08 (d, *J* = 12.0 Hz, 1 H), 4.97 (d, *J* = 12.0 Hz, 1 H), 4.31 (brs, 1 H), 3.32-3.30 (m, 2 H), 2.86 (s, 6 H), 1.38 ppm (s, 6 H); 13C NMR (125 MHz, CDCl3): δ = 170.0, 155.3, 152.0, 134.9, 134.2, 130.7, 129.9, 129.8, 129.5, 128.6, 128.5 (2), 128.3, 123.1, 118.6, 115.3, 80.4, 77.2, 67.6, 53.6, 45.4, 28.2 ppm. (**Figure 4.11**) (**Figure 4.12**)

Compound 9. 1H NMR (500 MHz, CDCl3): δ = 8.57 (d, *J* = 8.5 Hz, 1 H), 8.37 (d, *J* = 7.5 Hz, 1 H), 7.81 (d, *J* = 9.0 Hz, 1 H), 7.53 (t, *J* = 8.0 Hz, 2 H), 7.41-7.32 (m, 5 H), 7.16 (d, *J* = 8.0 Hz, 1 H), 5.51 (d, *J* = 9.0 Hz, 1 H), 5.26 (d, *J* = 12.5 Hz, 1 H), 5.17 (d, *J* = 12.5 Hz, 1 H), 4.85 (dd, *J* = 13.2, 7.8 Hz, 1 H), 4.45 (dd, *J* = 14.0, 5.0 Hz, 1 H), 4.38 (dd, *J* = 14.2, 7.8 Hz, 1 H), 2.87 (s, 6 H), 1.47 (s, 9 H), 1.13 ppm (s, 9 H); 13C NMR (125 MHz, CDCl3): δ = 170.4, 155.2, 152.0, 150.6, 135.3, 134.8, 131.9, 131.1, 129.6, 129.3, 128.7, 128.5, 128.4, 128.3, 122.8, 117.7, 115.1, 84.8, 80.0, 77.2, 67.5, 52.8, 47.3, 45.3, 28.3, 27.6 ppm. (**Figure 4.13**) (**Figure 4.14**)

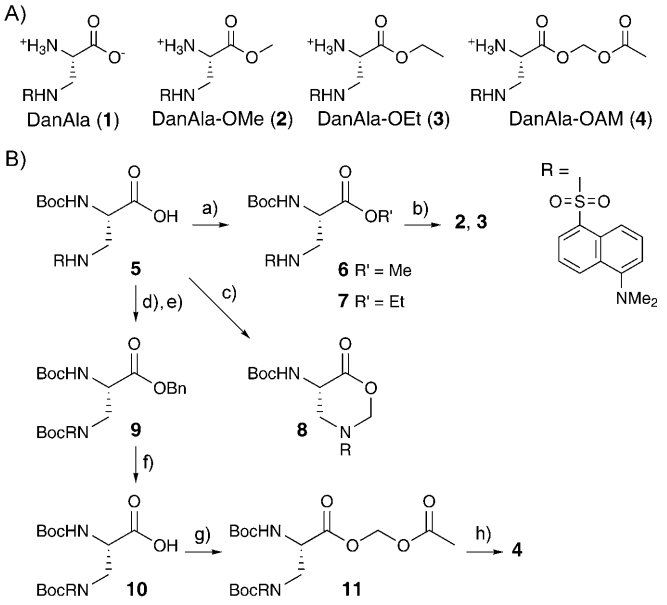
Compound 10. 1H NMR (500 MHz, CDCl3): δ = 8.57 (d, *J* = 8.5 Hz, 1 H), 8.37 (d, *J* = 7.5 Hz, 1 H), 7.83 (d, *J* = 8.5 Hz, 1 H), 7.55 (dd, *J* = 16.8, 8.2 Hz, 2 H), 7.17 (d, *J* = 8.0 Hz, 1 H), 5.59 (d, *J* = 8.5 Hz, 1 H), 4.84- 4.83 (m, 1 H), 4.44-4.41 (m, 2 H), 2.88 (s, 6 H), 1.49 (s, 9 H), 1.17 ppm (s, 9H); 13C NMR (125 MHz, CDCl3): δ = 156.0, 151.9, 150.8, 134.7, 132.0, 131.2, 129.6, 129.4, 128.8, 122.8, 117.8, 115.2, 85.1, 80.6, 77.2, 47.0, 45.4, 28.3, 27.6 ppm; (**Figure 4.15**) (**Figure 4.16**)

Compound 11. 1H NMR (500 MHz, CDCl3): δ = 8.58 (d, *J* = 8.0 Hz, 1 H), 8.36 (d, *J* = 7.0 Hz, 1 H), 7.79 (d, *J* = 9.0 Hz, 1 H), 7.55 (dd, *J* = 16.0, 8.5 Hz, 2 H), 7.17 (d, 7.5 Hz, 1 H), 5.84-5.81 (m, 2 H), 5.51 (d, *J* = 8.5 Hz, 1 H), 4.82 (dd, *J* = 12.8, 7.2 Hz, 1 H), 4.46 (dd, *J* = 14.5, 4.8 Hz, 1 H), 4.39 (dd, *J* = 14.5, 7.5 Hz, 1 H), 2.88 (s, 6 H), 2.13 (s, 3 H), 1.48 (s, 9 H), 1.14 ppm (s, 9 H); 13C NMR (125 MHz, CDCl3): δ = 169.5, 169.4, 155.1, 152.0, 150.6, 134.7, 132.0, 131.2, 129.6, 129.3, 128.8, 122.8, 117.7, 115.2, 85.0, 80.2, 79.9, 77.2, 52.7, 47.0, 45.3, 28.3, 27.5, 20.7 ppm. (**Figure 4.17**) (**Figure 4.18**)

Compound 4. 1H NMR (500 MHz, [D6]DMSO): δ = 8.51 (d, *J* = 8.5 Hz, 1 H), 8.37 (t, *J* = 6.5 Hz, 1 H), 8.26 (d, *J* = 8.5 Hz, 1 H), 8.11 (dd, *J* = 7.2, 1.2 Hz, 1 H), 7.67 (dd, *J* = 8.2, 7.8 Hz, 1 H), 7.63

(dd, *J* = 8.5, 7.5 Hz, 1 H), 7.32 (d, *J* = 8.0 Hz, 1 H), 5.73 (d, *J* = 6.0 Hz, 1 H), 5.59 (d, *J* = 5.5 Hz, 1 H), 4.20 (t, *J* = 4.8 Hz, 1 H), 3.27 (ddd, *J* = 14.5, 7.0, 5.5 Hz, 1 H), 3.17 (dt, *J* = 14.5, 5.5 Hz, 1 H), 2.85 (s, 6 H), 2.07 ppm (s, 3 H); 13C NMR (125 MHz, [D6]DMSO): δ = 169.1, 166.7, 158.5 (q, J = 36.1 Hz), 151.1, 134.8, 129.9, 129.1, 128.9, 128.8, 128.2, 123.8, 119.1, 115.6, 79.8, 52.3, 45.2, 42.2, 20.5 ppm. LC-MS (ESI): (m/z) 410.2 [M+H]+. HRMS (ESI-FT): calcd for C18H24N3O6S: 410.1380 [*M*+1], found 410.1382. (**Figure 4.19**) (**Figure 4.20**)

4.6 SCHEME



Scheme 4.1: A) Structures of DanAla and DanAla esters. B) Synthetic approach to DanAla esters. Reagents and conditions: a) MeI, DIPEA, DMF, 12 h, 87% for **6**; EtI, DIPEA, DMF, 12 h, 75% for **7**; b) TFA, CH_2Cl_2, 100%; c) bromomethyl acetate, DIPEA, CH_3CN, 81%; d) BnBr, DIPEA, DMF, 0 °C to RT, 76%; e) Boc_2O, DMAP, Et_3N, CH_2Cl_2, 97%; f) Pd/C, H_2, MeOH; g) bromomethyl acetate, DIPEA, CH_3CN, 84% over two steps; h) TFA, CH_2Cl_2, 100 %.

4.7 FIGURES

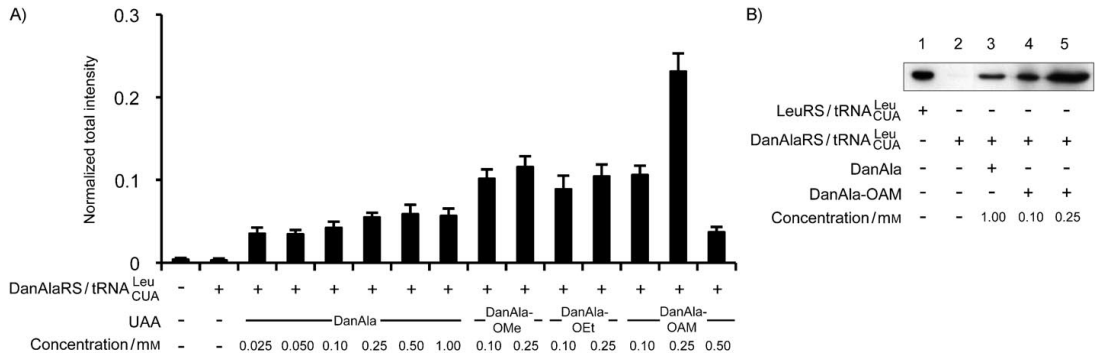


Figure 4.1: A) Flow cytometric analysis of the incorporation efficiency of DanAla into GFP with different compounds added in the growth medium. Reporter cells transfected with the orthogonal suppressor tRNA and the wild-type LeuRS were used as the positive control to normalize the total fluorescence intensity. Error bars represent standard error of mean (SEM, *n*=3). B) Western blot analysis of the GFP protein with a GFP-specific antibody. The same number of cells were used in lanes 2–5.

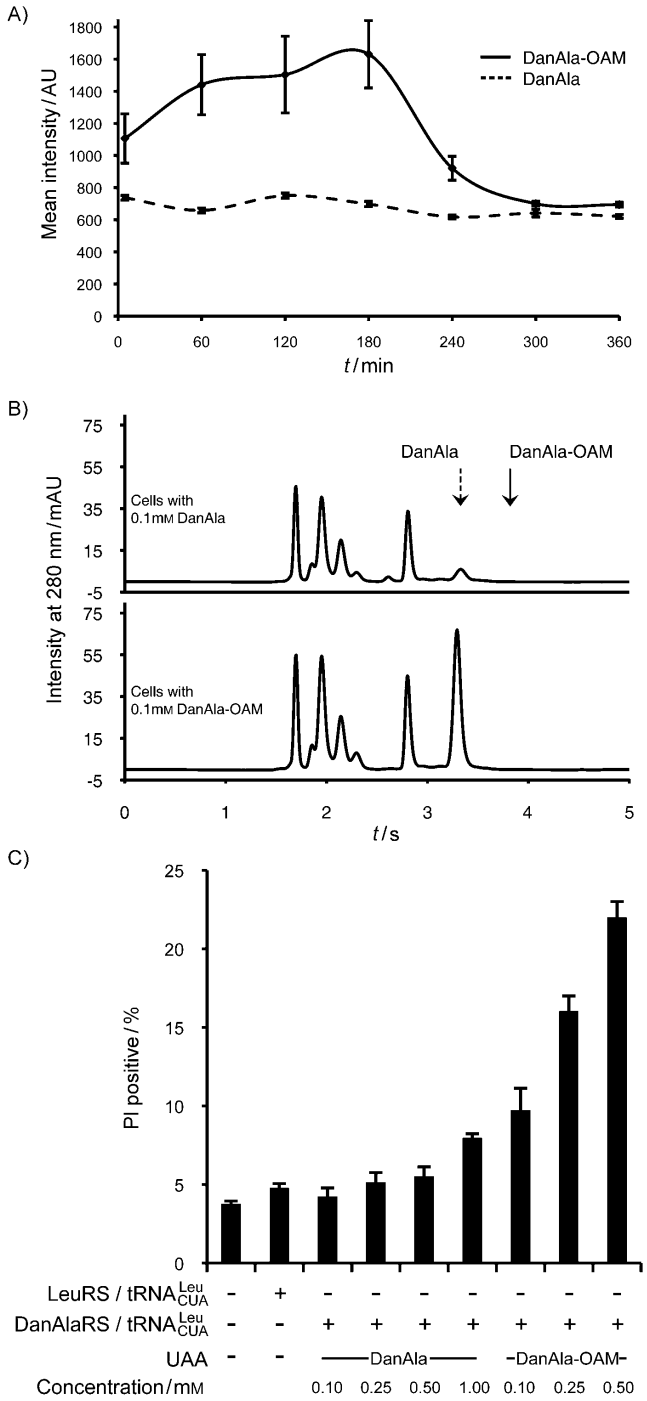


Figure 4.2: A) Cytosolic fluorescence intensity of cells incubated with 0.10 mMDanAla or DanAla-OAM. Error bars represent standard deviation (SD). B) HPLC analysis of cell extracts from cells incubated with 0.10 mM DanAla or DanAla-OAM. Arrows indicate the peak positions for DanAla and DanAla-OAM, as determined by using pure compounds. C) PI staining analysis of cell health at indicated conditions. Error bars represent SEM, *n*=3.

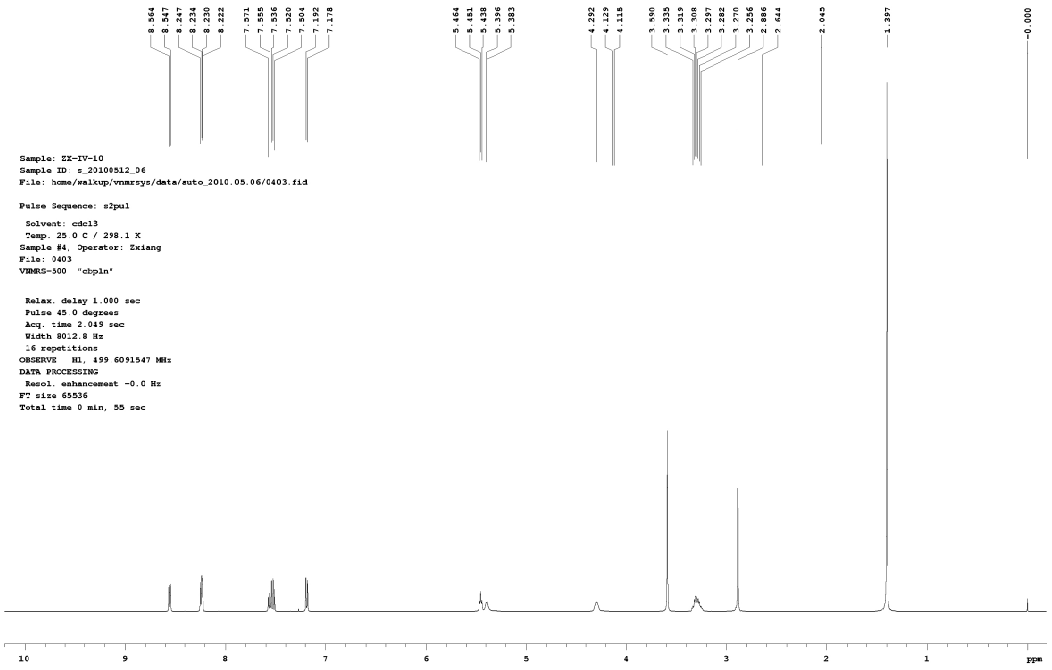


Figure 4.3: ^{1}H NMR (500 MHz, CDCl3) of Compound 6

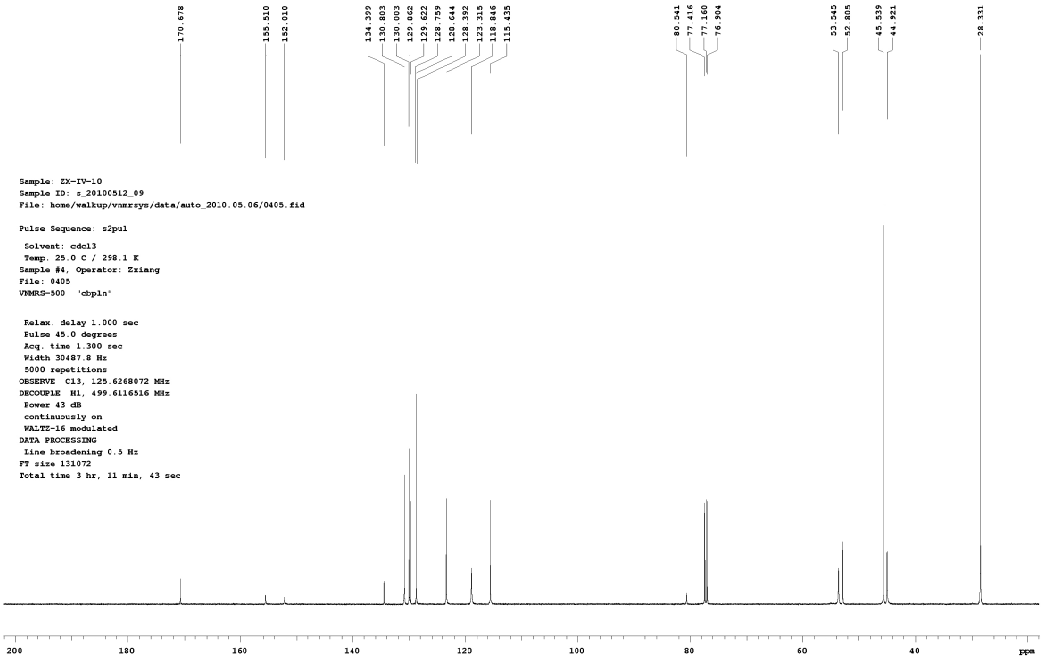


Figure 4.4: ^{13}C NMR (125 MHz, CDCl3) of Compound 6

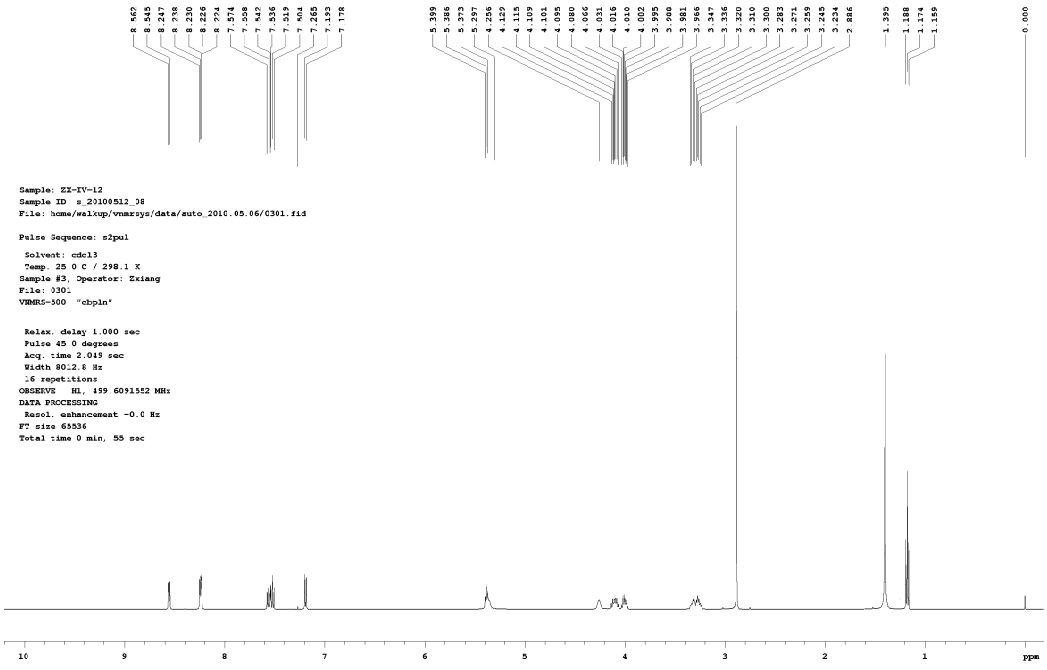


Figure 4.5: ^{1}H NMR (500 MHz, CDCl3) of Compound 7

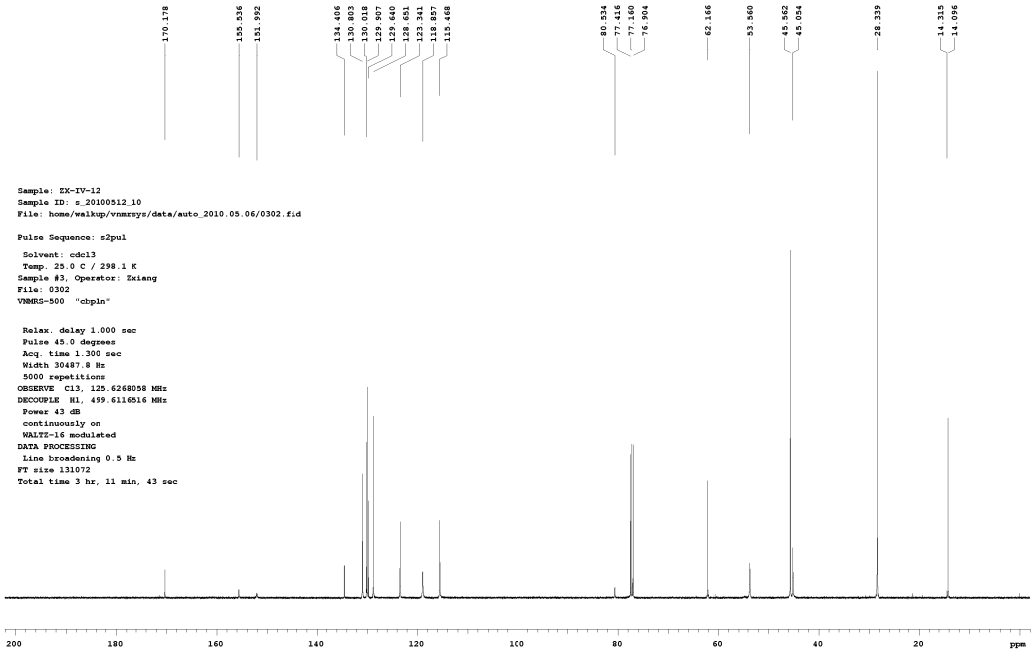


Figure 4.6: ^{13}C NMR (125 MHz, CDCl3) of Compound 7

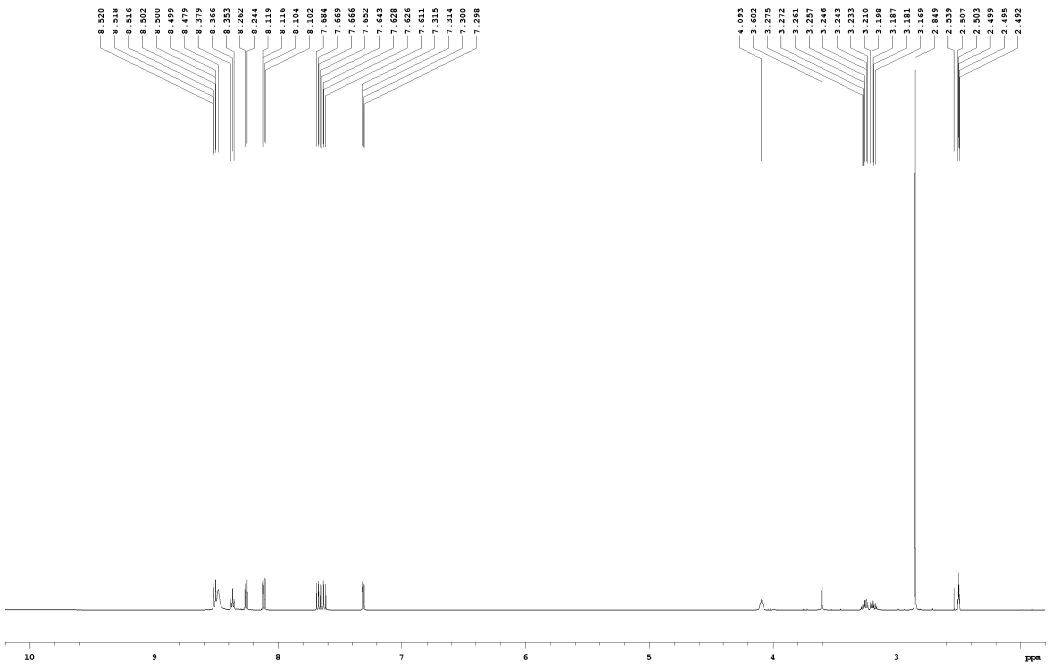


Figure 4.7: ^{1}H NMR (500 MHz, d6-DMSO) of Compound 2

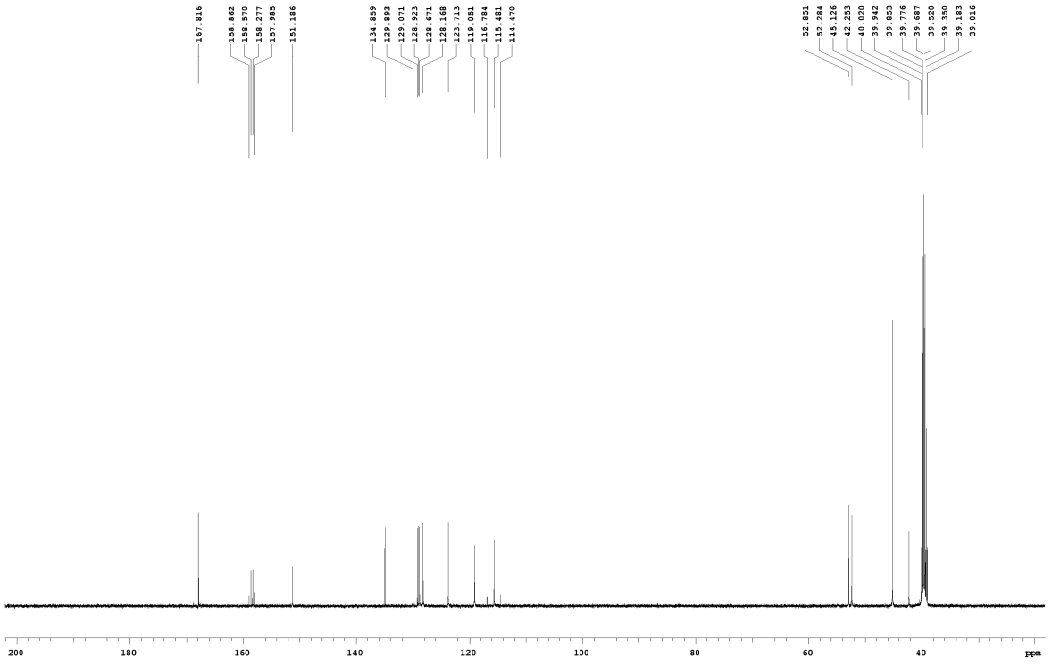


Figure 4.8: ^{13}C NMR (125 MHz, d6-DMSO) of Compound 2

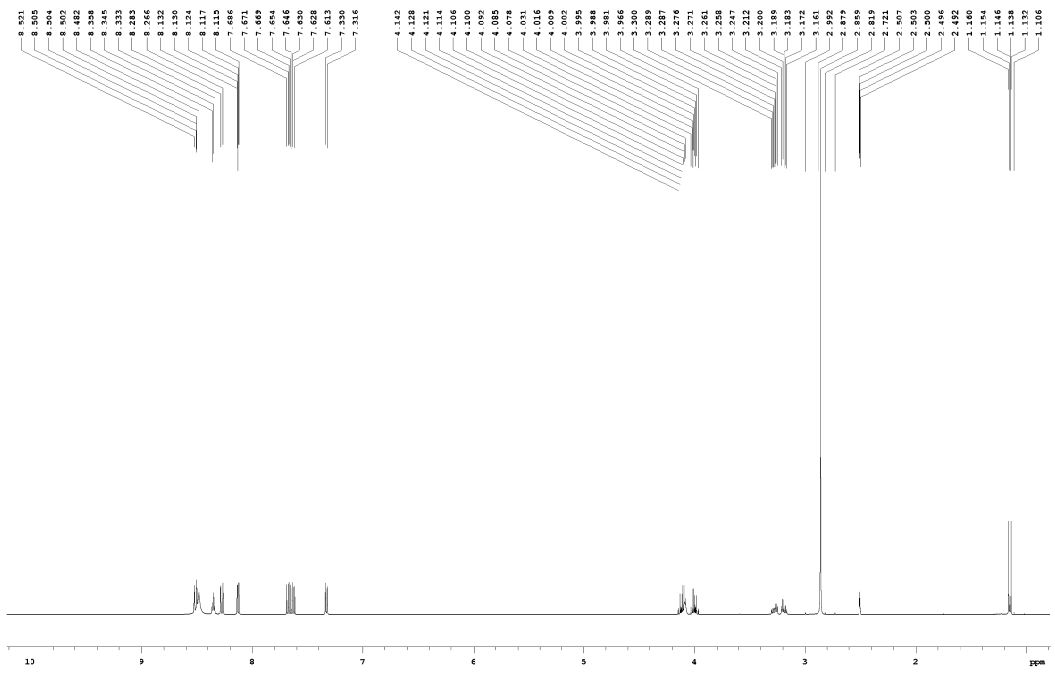


Figure 4.9: ^{1}H NMR (500 MHz, d6-DMSO) of Compound 3

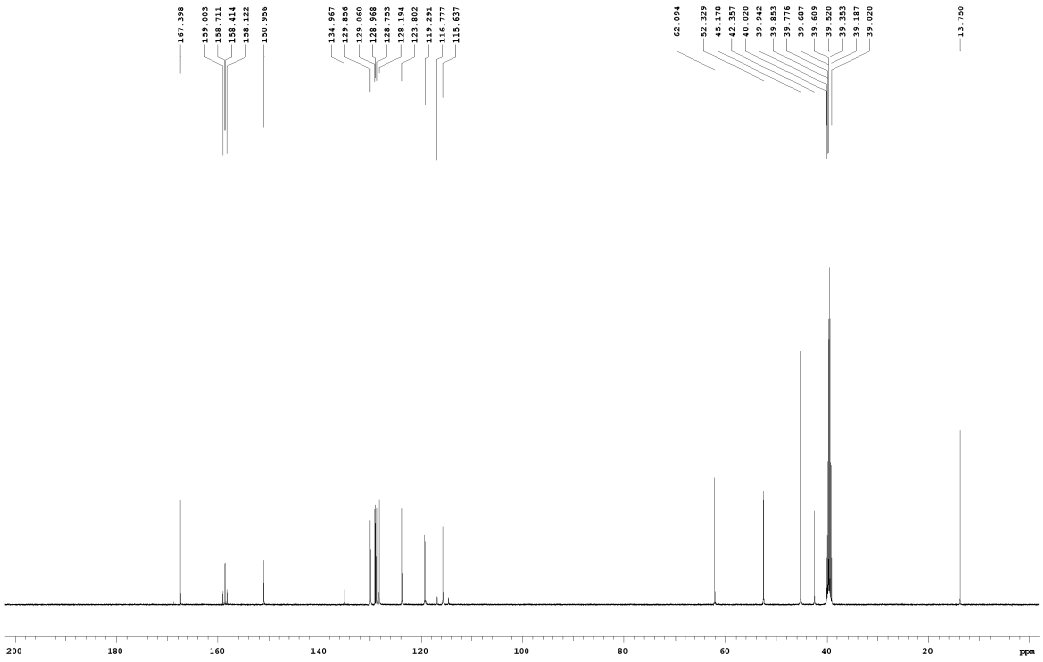


Figure 4.10: ^{13}C NMR (125 MHz, d6-DMSO) of Compound 3

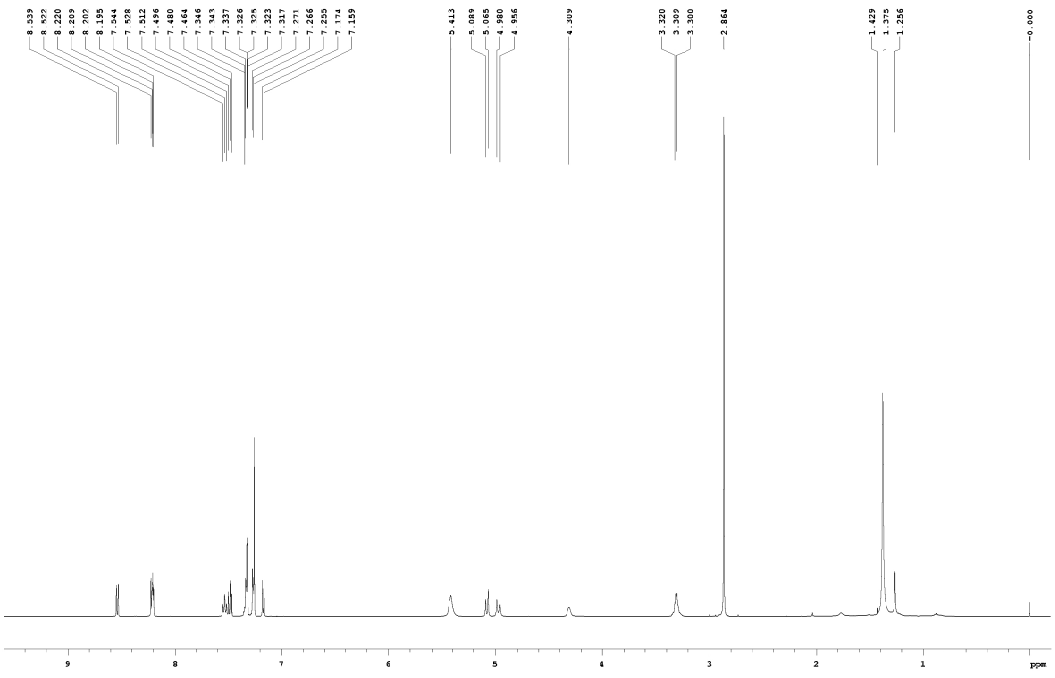


Figure 4.11: ^{1}H NMR (500 MHz, CDCl3) of Boc-DanAla-OBn

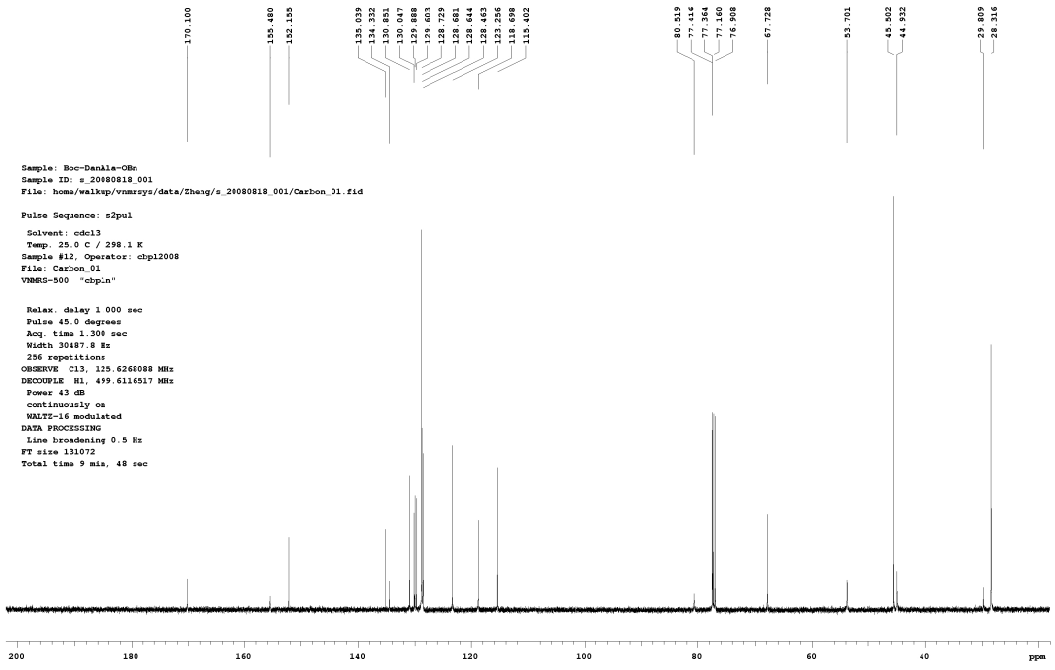


Figure 4.12: ^{13}C NMR (125 MHz, CDCl3) of Boc-DanAla-OBn

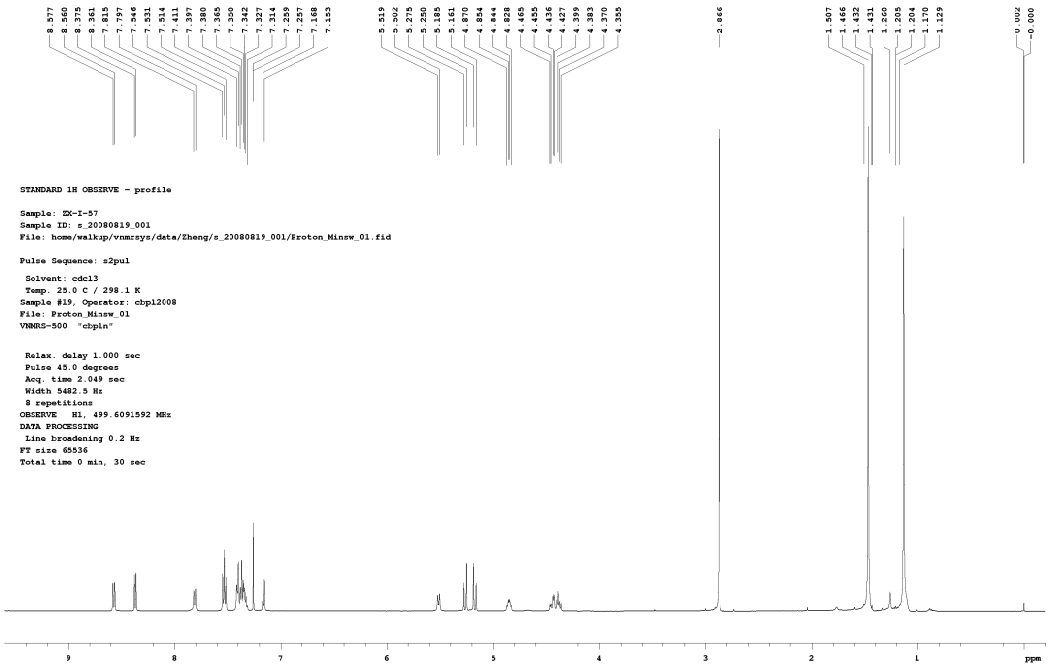


Figure 4.13: ^{1}H NMR (500 MHz, CDCl3) of Compound 9

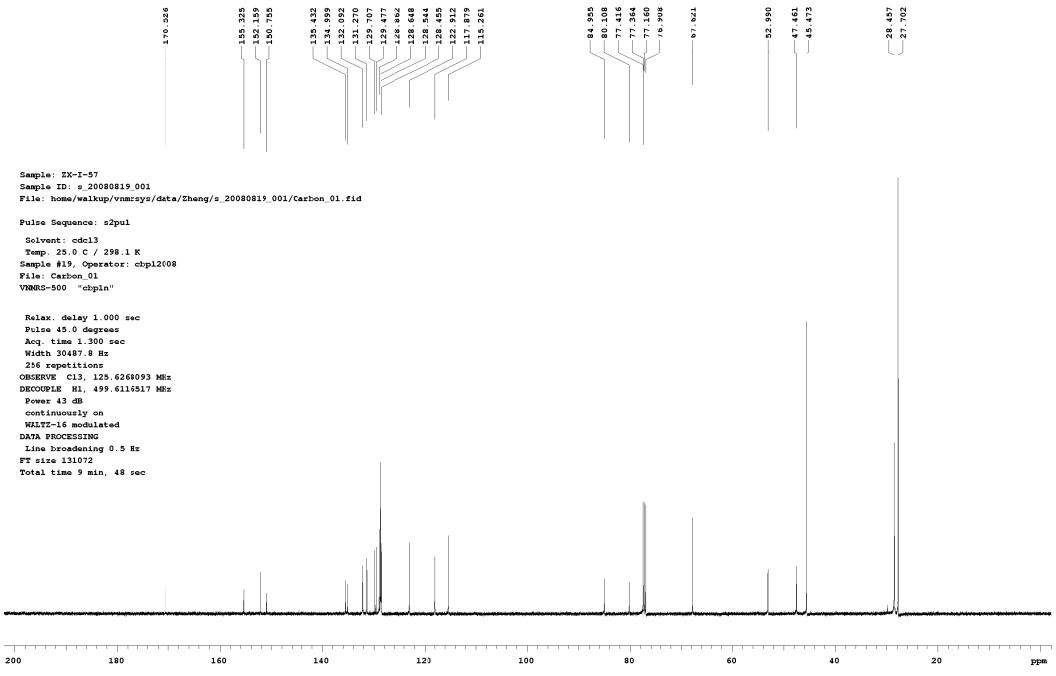


Figure 4.14: ^{13}C NMR (125 MHz, $CDCl_3$) of Compound 9

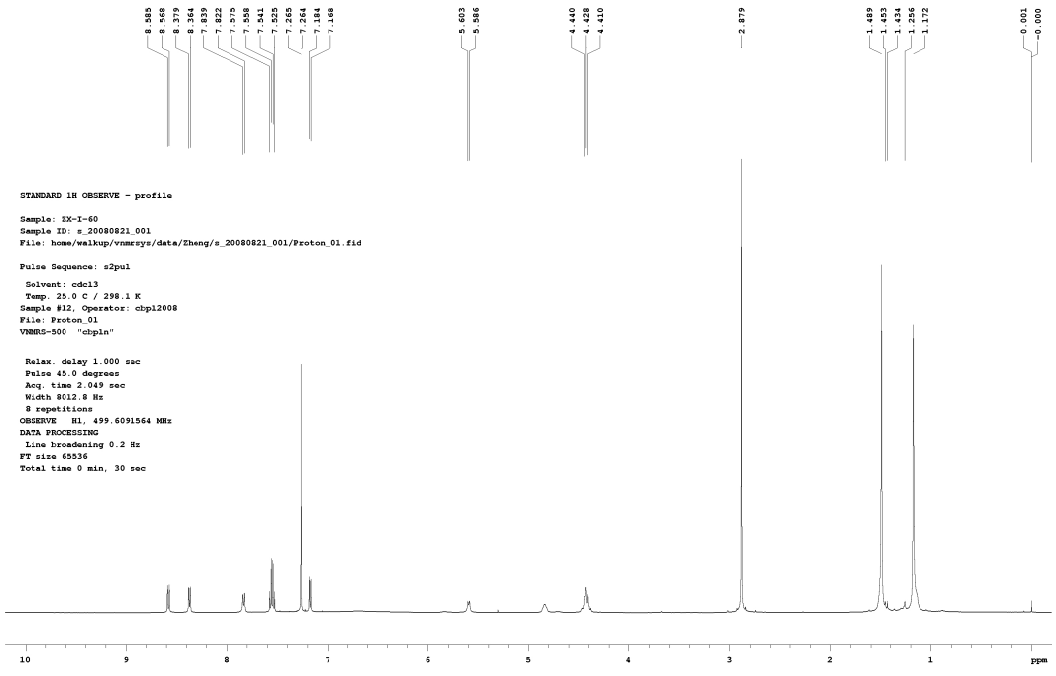


Figure 4.15: ^{1}H NMR (500 MHz, $CDCl_3$) of Compound 10

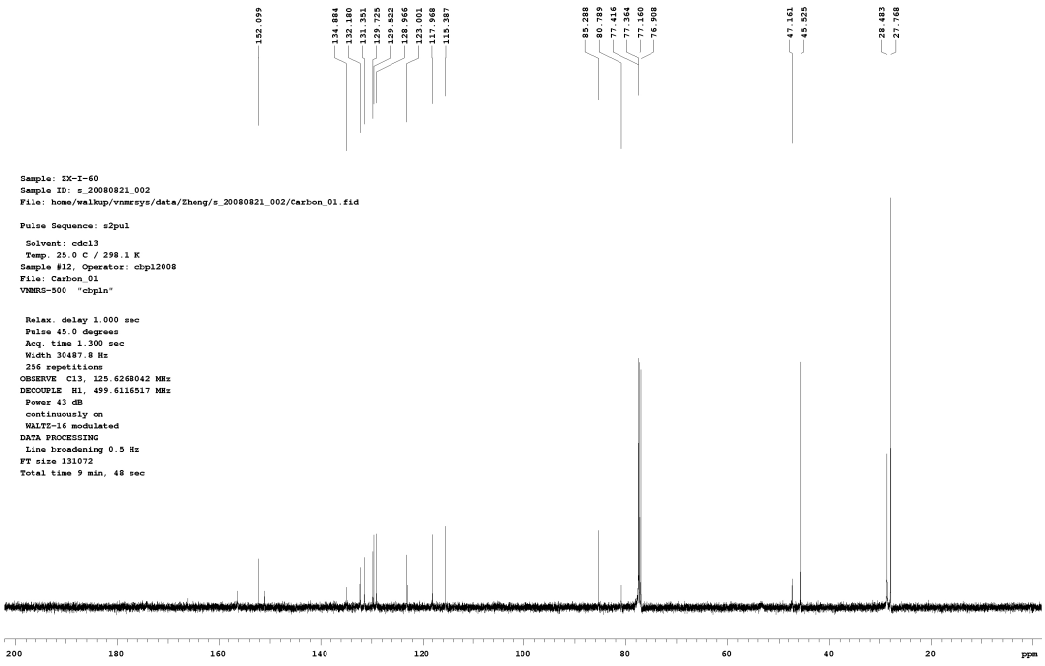


Figure 4.16: ^{13}C NMR (125 MHz, $CDCl_3$) of Compound 10

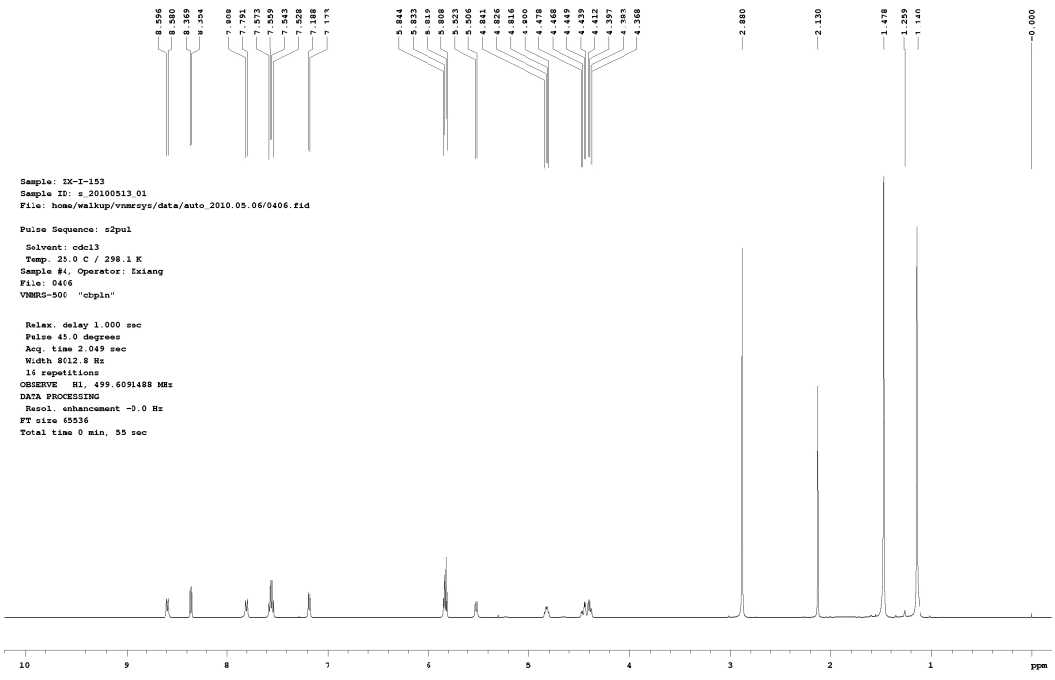


Figure 4.17: ^{1}H NMR (500 MHz, $CDCl_3$) of Compound 11

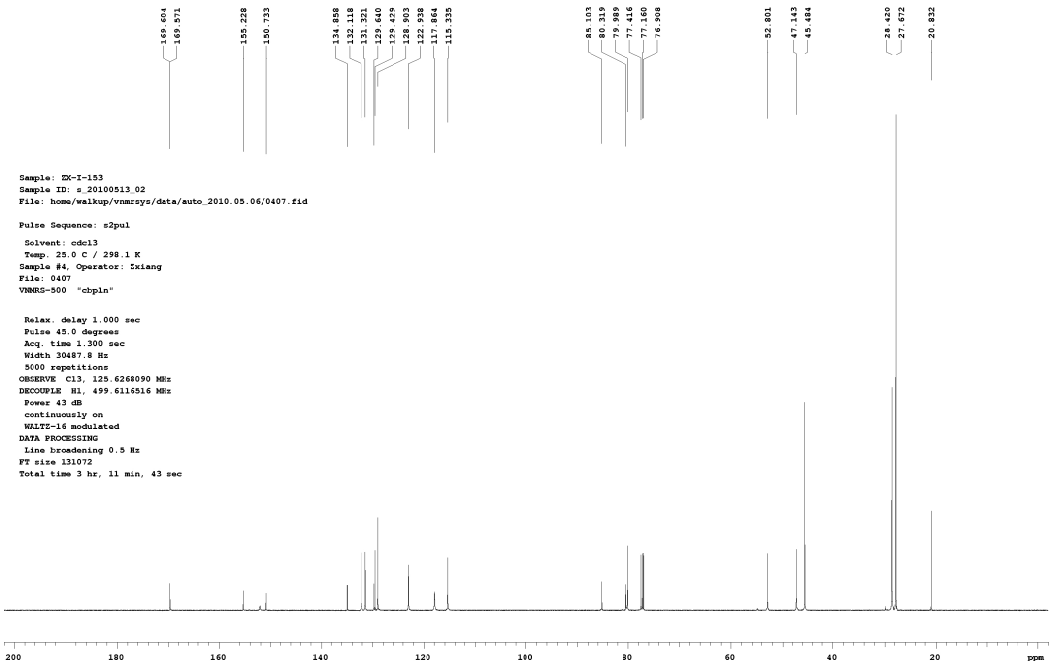


Figure 4.18: ^{13}C NMR (125 MHz, $CDCl_3$) of Compound 11

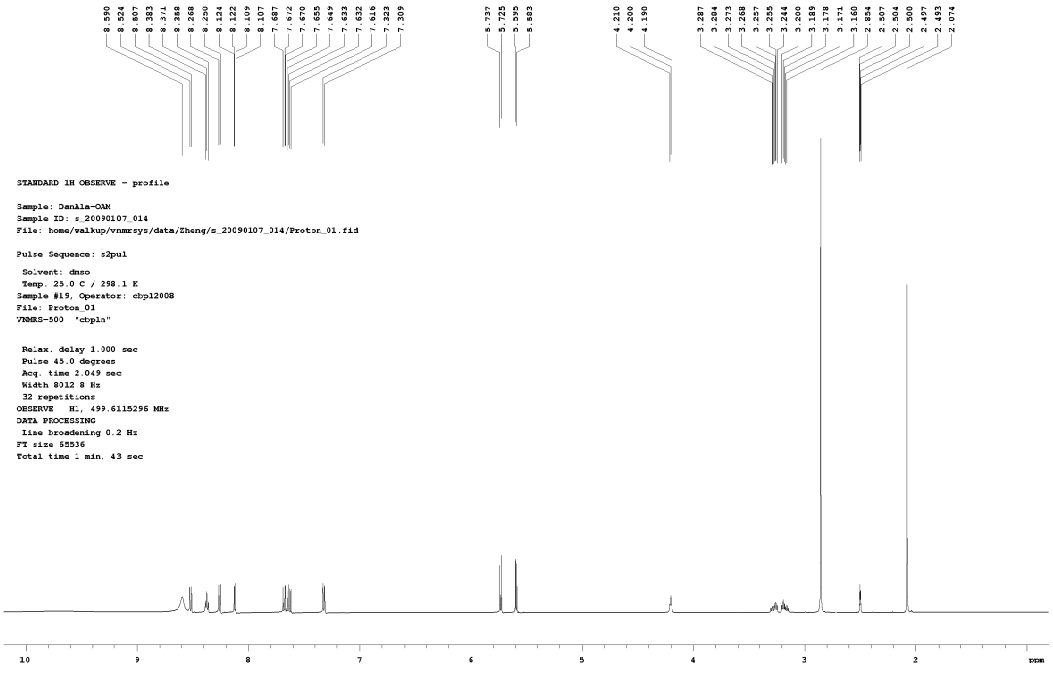


Figure 4.19: ^{1}H NMR (500 MHz, d6-DMSO) of Compound 4

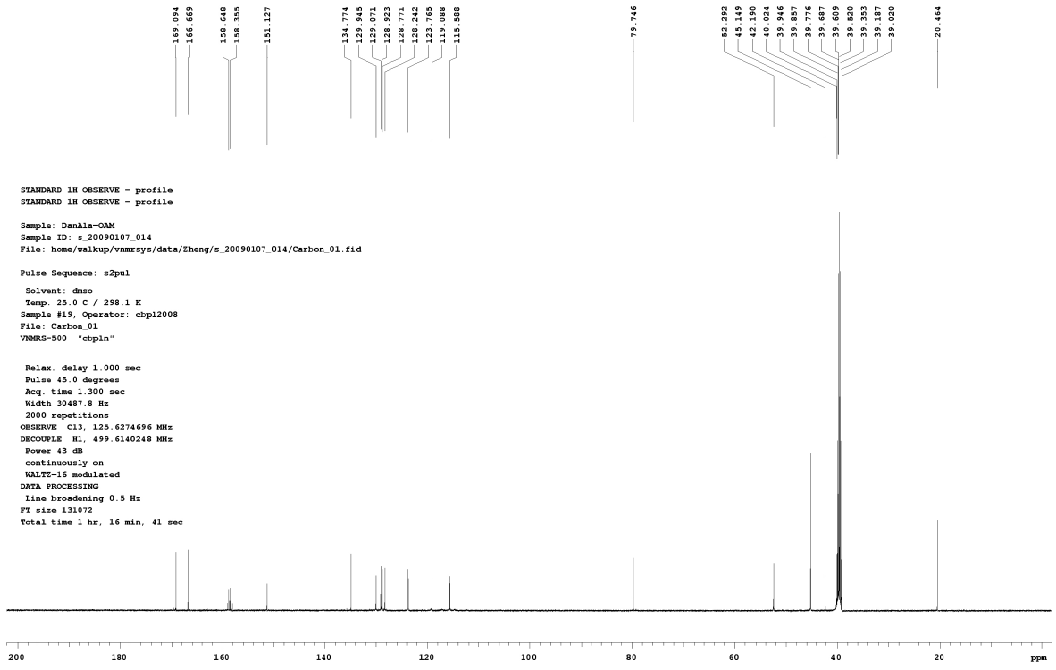


Figure 4.20: ^{13}C NMR (125 MHz, d6-DMSO) of Compound 4

4.7 REFERENCES

1. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.

2. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angew Chem Int Ed Engl* **2004,** *44* (1), 34-66.

3. Wang, Q.; Parrish, A. R.; Wang, L., Expanding the genetic code for biological studies. *Chem Biol* **2009,** *16* (3), 323-36.

4. Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S., Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res* **2002,** *30* (21), 4692-9.

5. Liu, W.; Brock, A.; Chen, S.; Schultz, P. G., Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* **2007,** *4* (3), 239-44.

6. Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L., Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat Neurosci* **2007,** *10* (8), 1063-72.

7. Takimoto, J. K.; Adams, K. L.; Xiang, Z.; Wang, L., Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol Biosyst* **2009,** *5* (9), 931-4.

8. Wang, L.; Xie, J.; Schultz, P. G., Expanding the genetic code. *Annu Rev Biophys Biomol Struct* **2006,** *35*, 225-49.

9. Malandro, M. S.; Kilberg, M. S., Molecular biology of mammalian amino acid transporters. *Annu Rev Biochem* **1996,** *65*, 305-36.

10. Rautio, J.; Kumpulainen, H.; Heimbach, T.; Oliyai, R.; Oh, D.; Jarvinen, T.; Savolainen, J., Prodrugs: design and clinical applications. *Nat Rev Drug Discov* **2008,** *7* (3), 255-70.

11. Tsien, R. Y., A non-disruptive technique for loading calcium buffers and indicators into cells. *Nature* **1981,** *290* (5806), 527-8.

12. Schultz, C.; Vajanaphanich, M.; Harootunian, A. T.; Sammak, P. J.; Barrett, K. E.; Tsien, R. Y., Acetoxymethyl esters of phosphates, enhancement of the permeability and potency of cAMP. *J Biol Chem* **1993,** *268* (9), 6316-22.

13. Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G., A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* **2006,** *103* (26), 9785-9.

14. Neustadt, B. R., Facile preparation of N-(sulfonyl)carbamates. *Tetrahedron Letters* **1994,** *35* (3), 379-380.

15. Meijler, M. M.; Arad-Yellin, R.; Cabantchik, Z. I.; Shanzer, A., Synthesis and evaluation of iron chelators with masked hydrophilic moieties. *J Am Chem Soc* **2002,** *124* (43), 12666-7.

16. Thomas, P.; Smart, T. G., HEK293 cell line: a vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods* **2005,** *51* (3), 187-200.

Chapter 5

Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient *in Vivo* Incorporation of Structurally Divergent Non-native Amino Acids

5.1 Abstract

Unnatural amino acids (Uaas) can be translationally incorporated into proteins *in vivo* using evolved tRNA/aminoacyl-tRNA synthetase (RS) pairs, affording chemistries inaccessible when restricted to the 20 natural amino acids. To date, most evolved RSs aminoacylate Uaas chemically similar to the native substrate of the wild-type RS; these conservative changes limit the scope of Uaa applications. Here, we adapt *Methanosarcina mazei* PylRS to charge a noticeably disparate Uaa, *O*-methyl-L-tyrosine (Ome). In addition, the 1.75 Å X-ray crystal structure of the evolved PylRS complexed with Ome and a non-hydrolyzable ATP analogue reveals the stereochemical determinants for substrate selection. Catalytically synergistic active site mutations remodel the substrate-binding cavity, providing a shortened but wider active site. In particular, mutation of Asn346, a residue critical for specific selection and turnover of the Pyl chemical core, accommodates different side chains while the central role of Asn346 in aminoacylation is rescued through compensatory hydrogen bonding provided by A302T. This multifaceted analysis provides a new starting point for engineering PylRS to aminoacylate a significantly more diverse selection of Uaas than previously anticipated.

5.2 Introduction

Incorporation of Uaas into proteins using a host's endogenous translation machinery opens the door to addressing questions with chemical precision that is unattainable using naturally occurring amino acids. This expanded toolset allows one to pose and answer more in-depth molecular questions without the limitations imposed by the 20 natural amino acids used in traditional mutagenic analyses[1, 2]. Aminoacyl-tRNA synthetases (RSs) obtained by structure-based engineering and directed evolution efficiently recognize and activate Uaas through ATP-dependent adenylation and subsequently catalyze transfer to their cognate tRNA. To date, more than 70 Uaas are now amenable to translational insertion into proteins in bacteria, yeast, or mammalian cells using these artificially evolved tRNA/RS pairs[3]. By choosing particular matching sets of tRNA/RSs from diverse organisms, the pairs can function *in vivo* in an orthogonal manner. In other words, there is limited if any crosstalk between the expression host's native tRNA/RS pairs and the orthogonal tRNA/RS pair; however, the orthogonal pair is still

able to functionally couple with the host's protein translational machinery. The $tRNA^{Tyr}$/TyrRS pair derived from the archaeon *Methanocaldococcus jannaschii* was the first successful pair to direct the genetically encoded incorporation of the Uaa *O*-methyl-L-tyrosine (Ome) into translated proteins *in vivo*[4, 5]. This initial result then spawned the incorporation of the majority of Tyr-based Uaas possessing distinct functional groups into recombinant proteins. Recently, another matching set, namely, the $tRNA^{Pyl}$/PylRS pair, was used to incorporate Uaas bearing functional groups similar to Pyl[6-12].

In most cases regarding $tRNA^{Tyr}$/TyrRS systems, sterically innocuous modifications of the phenyl ring, typically focused on the *para* position, are employed as unnatural substrates. Similarly, most Uaas incorporated by the $tRNA^{Pyl}$/PylRS pair all contain slight variations of the extended Pyl side chain, while the core Lys moiety and the N^{ε}-carbonyl group of Pyl remain unchanged[3]. In short, most evolved orthogonal tRNA/RS pairs demonstrate specificity for Uaas that maintain the chemical core of the native amino acid substrate, thus limiting the stereochemical diversity of Uaas accessible to the researcher. Since the RS component of the pair plays a key role in amino acid substrate selection, it is of tremendous importance to probe the extent to which it can be engineered not only to activate a chemically and structurally divergent amino acid but also its ability to transfer this new amino acid to its cognate tRNA. Given that the degree of substrate specificity exhibited by an RS is functionally coupled to the accuracy of protein translation, the RS is restrained by tremendous selective pressure; therefore, integration of structural approaches with directed evolution of RSs can expand upon our understanding of the evolutionary tenets governing substrate specificity and selection[13-15]. This unified approach applied in the form of multiple rounds of structure-based design and directed evolution can often produce RSs with altered specificity for Uaa substrates in relatively rapid fashion.

Moreover, the inability to efficiently employ mammalian cells and multicellular organisms for evolving tRNA/RS pairs further limits the scope of unnatural substrates available for probing *in vivo* protein function in organisms more complex than microbial hosts. Currently, directed approaches for evolving a Uaa-specific RS involve the generation of an initial RS mutant library minimally containing 10^8 to 10^9 members[1, 4]. These numbers, coupled with the need for efficient selection procedures, limit applicability to organisms possessing high transformation efficiency and favorable

growth characteristics[1]. The approach was first developed in *E. coli*[5, 16] and later expanded to include another unicellular organism, *Saccharomyces cerevisiae*[17, 18]. Nevertheless, maintenance of large mutant libraries and selection methods are difficult to implement in more complex systems such as mammalian cells and multicellular organisms.

One solution is to evolve an RS that possess characteristics favorable to deployment in the intended mammalian hosts in recombinant organisms such as *E. coli* or yeast. Assuming the appropriate RS is chosen, the evolved RSs should then be readily integrated into mammalian cells or a multicellular organism for further optimization once the large collection of mutants is narrowed substantially. This divide and conquer approach thus shortcuts some of the anticipated problems faced during the initial rounds of selection[19]. This kind of strategy was applied to the $tRNA^{Tyr}$/TyrRS pair from *E. coli*, wherein the TyrRS was first evolved in yeast and then transferred to mammalian cell hosts[19, 20]. Nonetheless, most reported mutant RSs evolved in yeast generally lack the *in vivo* efficiency of the original *M. jannaschii* TyrRS evolved and exploited in *E. coli*. Moreover, *M. jannaschii* TyrRS does not operate in an orthogonal manner in mammalian cells, often cross-charging mammalian tRNAs.

With these limitations in mind, $tRNA^{Pyl}$/PylRS is an attractive alternative because of the demonstrated orthogonal behavior of this particular pair in *E. coli* and mammalian cells[9, 10, 21, 22]. Mutating the active site does not impact orthogonality so the evolved pairs are highly unlikely to cross-react with the endogenous tRNA/RS pairs of their mammalian hosts. One remaining issue, however, concerns the extent to which the current assortment of orthogonal pairs can be evolved to specifically adenylate and then charge their cognate tRNAs with substrates substantially different from their natural substrates. Strikingly, to date, Uaas incorporated by mutant RSs derived from PylRS bear close chemical and steric resemblance to the natural Pyl substrate[3].

High fidelity Uaa incorporation is essential for generating chemically homogeneous proteins for investigation. An optimally evolved RS should function similarly to a wild-type RS, which achieves high incorporation fidelity under physiological growth conditions and maintains fidelity when cells are grown in nutrient-rich media. For Uaa incorporation in *E. coli*, nutrient-rich media is necessary to ensure high *in vivo* incorporation efficiency of the evolved RS; however, when an evolved RS possesses

non-optimal specificity for the Uaa, minimal media lacking certain or all natural amino acids must be used to limit the mis-incorporation of any natural amino acids into the recombinant protein[23]. Unfortunately, minimal media cannot be applied to traditional fermentations, mammalian cell cultures, and whole organism engineering; this drawback prevents the transfer and usage of non-optimally evolved RSs in these situations. Moreover, the active site of a non-optimal RS must lack the complementary stereochemical features for selection of the given Uaa substrate. Structural analyses then become critical for providing architectural guides to explain and further optimize the substrate selection and turnover. Recently, Wang *et al.* reported the incorporation of two Phe-based Uaas into proteins *via* evolved PylRSs, albeit in *E. coli* grown solely in minimal media[23, 24]. To date, efficient evolution of PylRS to specifically incorporate a Uaa significantly deviating from the native Pyl in rich media followed by transfer to mammalian hosts has not been demonstrated. Moreover, the structural transformations that occur in the active site of such a highly specific PylRS mutant necessary for accommodating a dramatic substrate change remain unclear.

Here we show that the *Methanosarcina mazei* PylRS (MmPylRS) can be evolved to efficiently charge a Uaa with a short aromatic side chain, in contrast to the long aliphatic side chain of the native substrate Pyl. The evolved RS incorporates the Uaa into proteins with high fidelity in both *E. coli* grown in rich media and mammalian cells. Additionally, we solved and refined the X-ray crystal structure of the evolved PylRS complexed with the new Uaa and an ATP analogue to a nominal resolution of 1.75 Å. The three-dimensional structure and active site architecture in the trapped substrate-bound form confirms that the mutations obtained contribute directly to the high stereochemical selectivity of our mutant PylRS for turnover of Tyr-based Uaas. Furthermore, this combined structure–function investigation provides the experimental support necessary to continue diversifying and optimizing PylRS variants in *E. coli*to further expand the variety of smaller Uaas for later use in mammalian cells.

5.3 Results and discussion

5.3.1 Laboratory Evolution of the *Methanosarcina mazei* $tRNA_{CUA}^{Pyl}$/PylRS pair

To expand the amino acid repertoire available to the orthogonal $tRNA_{CUA}^{Pyl}$/PylRS pair, we employed MmPylRS as a model for rationally engineering high specificity toward Tyr-based Uaa analogues. On the basis of the X-ray crystal structure of wild-type MmPylRS[25], six residues (Ala302, Leu309, Asn346, Cys348, Val401, and Trp417) lining the active site cavity were chosen for mutational expansion to create a genetically diverse PylRS library (Figure 5.1b). The Y384F mutation of PylRS, previously shown to increase the aminoacylation rate[26], was used as the starting PylRS sequence and remained fixed in our population of mutants. In general, RS engineering and directed evolution has subtly reshaped RS amino acid specificity by limiting mutagenic exploration to amino acid positions that encompass a portion of the extended amino acid side chain. Accordingly, mutant RS libraries generated from TyrRS and PylRS have restricted genetic diversity to deep within the amino acid binding pocket[1, 3, 4, 10]. To accommodate the dramatic substrate switch from the long aliphatic Pyl tail to a short aromatic moiety, we mutated active site residues lining the entire surface of the side chain binding pocket in the Y384F PylRS mutant library. In particular, Asn346, proposed to serve as a "gate keeper" for Pyl binding[25, 27], has until now been retained and not subjected to mutational expansion in most reported PylRS libraries. We, by contrast, specifically targeted this residue for mutational expansion to drive substrate selectivity substantially away from the natural Pyl substrate.

Our mutant library contains approximately 10^9 mutational variants, assessed by counting colony-forming units after efficient *E. coli* transformation. Cells were then subjected to positive selection in the presence of 1 mM of Ome (Figure 5.1a). Survival relies on the suppression of an amber stop codon introduced at a permissive site in the chloramphenicol acetyltransferase gene upon growth in the presence of 30 μg mL^{-1} chloramphenicol[4, 16]. Y384F PylRS mutant variants able to charge the orthogonal $tRNA_{CUA}^{Pyl}$ with either natural or the added Uaa should then suppress the amber codon, lead to chloramphenicol acetyltransferase expression, and afford chloramphenicol resistance to the host *E. coli* strain.

A total of 48 colonies survived on agar plates containing 30 μg mL^{-1} chloramphenicol and 1 mM Ome. A secondary screen to further narrow mutant selection employed a quantitative assessment of the growth of the original 48 colonies on 1 mM Ome using a matrix of chloramphenicol concentrations. One mutant, named MmOmeRS, afforded an IC_{50} of 15 μg mL^{-1}chloramphenicol in the absence of Ome and 80 μg mL^{-1} chloramphenicol in the presence of Ome. Sequencing revealed that the MmOmeRS gene contained the fixed mutation Y384F and four additional changes, A302T, N346V, C348W, and V401L.

5.3.2 MmOmeRS Specifically Incorporates the Selected Uaa in *E. coli*

To measure the *in vivo* translation efficiency and fidelity of MmOmeRS for incorporation of Ome into recombinant proteins in *E. coli*, the sperm whale myoglobin gene with an amber UAG codon at position 4 (Myo4TAGHis$_6$) was used together with the $tRNA_{CUA}^{Pyl}$/MmOmeRS pair in *E. coli* strain DH10β under control of the isopropyl β-D-1-thiogalactopyranoside(IPTG)-inducible T5 promoter. As shown in Figure 5.2a, when the nutrient-rich growth media 2xYT was supplemented with Ome (1 mM), full-length myoglobin appeared. Recombinant His6-tagged myoglobin was purified by Ni^{2+} affinity chromatography followed by gel filtration chromatography to ~98% homogeneity yielding 4.6 mg L^{-1} myoglobin. In the absence of the Uaa, small amounts (0.68 mg L^{-1}) of full-length myoglobin appeared, most likely due to low level charging of natural amino acids to $tRNA_{CUA}^{Pyl}$ by MmOmeRS. For comparison, the Ome-specific $tRNA_{CUA}^{Tyr}$/synthetase pair previously evolved from wild-type *M. jannaschii* $tRNA^{Tyr}$/TyrRS[4] was used to incorporate Ome into myoglobin, yielding 3.2 mg L^{-1} of recombinant myoglobin containing Ome.

To measure the fidelity of Uaa incorporation, the purified myoglobin was analyzed by electrospray ionization Fourier transform ion trap mass spectrometry (ESI-FTMS). For myoglobin obtained with the $tRNA_{CUA}^{Pyl}$/MmOmeRS strain, a peak with a monoisotopic mass of 18508.61 Da was observed (Figure 5.2b). This mass corresponds to intact myoglobin containing a single Ome residue at position 4 (expected $[M + H]^+$ = 18,508.77 Da). A second peak measured corresponds to Ome-containing myoglobin lacking the initiating Met (expected $[M – Met + H]^+$ = 18377.73 Da, measured

18377.58 Da). Notably, no peaks corresponding to proteins containing any other amino acids at the amber codon position were observed. The signal-to-noise ratio of >1,000 observed in the intact protein mass spectra translates to a fidelity of >99.9% for the incorporation of Ome with the $tRNA_{CUA}^{Pyl}$/MmOmeRS pair. Recombinant myoglobin produced using the $tRNA_{CUA}^{Pyl}$/MmOmeRS pair in media lacking Ome was also purified and analyzed by mass spectrometry (Figure 5.2c). The monoisotopic masses obtained indicate that Phe or Trp were incorporated at the UAG sites (expected $[M(Phe) + H]^+$ = 18477.75 Da, measured 18477.52 Da; expected $[M(Phe) – Met + H]^+$ = 18346.71 Da, measured 18347.44 Da; expected $[M(Trp) – Met + H]^+$ = 18385.72 Da, measured 18385.48 Da). While these proteins contain natural amino acids, it is notable that the incorporated amino acids are similar to the Ome Uaa and quite distinct from the cognate Pyl amino acid of the original wild-type PylRS.

To assess the generality of MmOmeRS, a GFP gene containing a UAG codon at position 182 was employed. In the absence and presence of Ome, 0.13 and 13.89 mg L^{-1} of GFP were obtained, respectively. Ome-containing GFP was digested with trypsin and analyzed by ESI-FTMS. Ions for the tryptic GFP peptides clearly indicate that Ome was incorporated at the UAG site (expected $[M + H]^+$ = 4486.19 Da, measured 4486.18 Da), and no peaks corresponding to any other amino acid incorporated at the UAG site were detected (Figure 5.2d). The precursor ion $[M + H]^+$corresponding to the peptide ^+H_3N-HNIEDGSVQLADH**X**QQNTPIGDGPVLLPDNHYLSTQSALSK-CO_2^-
(where **X** represents the UAG site) was also analyzed by tandem MS, and ion masses unambiguously confirmed the peptide sequence and assigned Ome to the UAG site (**Supplementary Figure 5.1**). In summary, these experimental observations demonstrate that while MmOmeRS will accept the sterically similar Phe or Trp natural amino acids at low levels in the absence of Ome, the $tRNA_{CUA}^{Pyl}$/MmOmeRS pair specifically incorporates Ome with high fidelity and efficiency when its cognate amino acid Ome is present.

5.3.3 Use of MmOmeRS for Uaa Incorporation in Mammalian Cells

One historical drawback associated with tRNA/RS pairs evolved in bacteria is that they are generally not orthogonal in eukaryotic cells and thus cannot be transferred to more complex systems

such as mammalian cells or whole organisms. A significant advancement is the ability to evolve pairs in easily manipulated and efficiently transformed bacterial cells and then transfer the optimized orthogonal pairs directly to mammalian cells while preserving efficient and high fidelity Uaa incorporation into expressed proteins[10, 19, 20]. MmOmeRS was tested for its ability to incorporate Uaas in HEK293 and HeLa cells. The $tRNA_{CUA}^{Pyl}$ was expressed under control of the polymerase (Pol) III H1 promoter with the 3′ flanking sequence of the human $tRNA^{Tyr}$ appended to the 3′ end of the $tRNA_{CUA}^{Pyl}$ gene for correct 3′ end processing[19]. MmOmeRS was conditionally expressed using the mammalian PGK promoter[19]. The GFP gene with a UAG stop codon introduced at a permissive site Tyr182 was cotransfected with the $tRNA_{CUA}^{Pyl}$/MmOmeRS into HEK293 cells. Cells were grown in media with or without the Uaa Ome. As shown in Figure 5.3a, green fluorescence appeared when cells were grown in the presence of Ome. In the absence of the Uaa, no bright green fluorescent cells were observed. To accurately quantify the Ome incorporation efficiency, a HeLa cell line with the GFP(182TAG) gene stably integrated into the chromosome[19] was used as a reporter. After transfecting the $tRNA_{CUA}^{Pyl}$/MmOmeRS pair, flow cytometry was then used to measure cellular GFP fluorescence. Total fluorescence intensity was normalized to the quantified fluorescence of HeLa cells transfected with the *E. coli* $tRNA_{CUA}^{Tyr}$/TyrRS pair (Figure 5.3b). Incorporation efficiency for the $tRNA_{CUA}^{Pyl}$/MmOmeRS pairs was 4.54% with a background suppression of 0.09% detected in the absence of Ome.

A remaining challenge for the incorporation of Uaas into proteins in mammalian cells is the efficient transcription of an orthogonal tRNA incapable of being charged by the host's endogenous RSs. Previously, we developed a general method to transcribe bacterial tRNAs in mammalian cells under control of type-3 Pol III promoters[19]. These promoters, including the H1 and U6 promoters, enable the efficient transcribing of tRNAs that do not possess the consensus A- and B-boxes normally required for Pol III transcription in mammalian cells. The *M. mazei* $tRNA_{CUA}^{Pyl}$ used in this current study comes from the domain Archaea, and unlike bacterial tRNAs, this archaeal tRNA contains a consensus eukaryotic B-box and an imperfect A-box sequences. The *M. mazei* $tRNA_{CUA}^{Pyl}$ can also be functionally expressed under transcriptional control by the H1 promoter in mammalian cells. The results described

here support the deployment of type-3 Pol III promoters for expressing other archaeal tRNAs in mammalian cells[9].

5.3.4 Three-Dimensional Architecture of MmOmeRS

Following structure-guided mutation and selection, MmOmeRS switched amino acid specificity from the native substrate Pyl to the structurally divergent substrate Ome. Notably, the new specificity obtained for Ome differs substantially from the wild-type specificity for the long aliphatic side chain of Pyl. To understand the molecular basis underlying this evolutionary change, we determined the high-resolution X-ray crystal structure of the catalytic domain of MmOmeRS complexed with Ome and the non-hydrolyzable ATP analogue AMP-PNP.

MmOmeRS, like its wild-type ancestor PylRS, purified as a homodimer and crystallized in space group $P6_4$ with unit cell dimensions of a = 104.81 Å, b = 104.81 Å, c = 71.67 Å and angles α = β = 90°, γ = 120° using the previously published crystallization conditions for wild-type MmPylRS[25]. MmOmeRS was also cocrystallized with Mg^{2+}, the non-hydrolyzable ATP analogue AMP-PNP, and the Uaa Ome. The crystal structure of the ternary Ome-AMP-PNP-Mg^{2+} complex was solved by molecular replacement and refined to 1.75 Å with an R_{factor} of 18.3% and an $R_{free}$of 21.2% (see **Table 5.1** for data processing and refinement statistics). The SIGMAA weighted $2F_o - F_c$ experimental electron density maps exhibited clear electron density for Ome, AMP-PNP, and two octahedrally coordinated Mg^{2+} ions. MmOmeRS crystals contained one ternary complex per asymmetric unit; two ternary complexes related by a 2-fold crystallographic symmetry axis form the likely physiological dimer of MmOmeRS.

PylRS and the MmOmeRS mutant belong to the class-II division of aminoacyl-tRNA synthetases (PF00587)[25]. This fold consists of a central antiparallel seven-stranded β-sheet decorated by several peripheral helices. The overall tertiary structure of MmOmeRS (Figure 5.4a) is consistent with all previously reported crystal structures for PylRSs, generating root-mean-square deviation (rmsd) values ranging from 0.16 to 0.23 Å upon superposition of C-α atoms with representative family members (PDB IDs 2ZCE, 2ZIN, and 2Q7G). Structural alignment of the MmOmeRS structure

reported here with the previously reported PylRS structure complexed with AMP-PNP and Pyl (PDB ID: 2ZCE)[27] illustrates conserved AMP-PNP binding with the sterically and chemically divergent amino acid Ome occupying a similar active site location as the wild-type cognate amino acid Pyl[25-27].

One notable structural difference between the MmOmeRS crystal structure reported here and other published PylRS structures encompasses residues across the β7−β8 loop (Val177–Gly185). Previous crystal structures of wild-type PylRS establish several distinct conformations of this loop, including open, intermediate, and closed conformations that are seemingly independent of the bound Pyl amino acid ligand[27]. In the crystal structure of MmOmeRS, electron density for the β7−β8 loop is noticeably absent, signifying that the loop most likely contributes in a dynamic manner to active site closure and catalysis.

5.3.5 Active Site Substitutions Lead to an Unexpected Recognition of the Ome Side Chain Moiety

MmOmeRS contains a single fixed mutation, Y384F, and four additional mutations selected for during directed evolution: A302T (located on the N-terminal end of the α5 helix), N346V and C348W (abutting residues on the C-terminal end of the β6 strand), and V401L (located on the C-terminal end of the β9 strand) (Figure 5.4b). The methyl group of Ome is clearly visible on the tip of the bound ligand, distinguishing Ome from the structurally similar natural amino acid Tyr (Figure 5.4c). In all previous studies of PylRS, the wild-type residue Asn346 serves to (1) directly bind the secondary carbonyl (N^{ε}-carbonyl) of Pyl or Pyl analogues through hydrogen bonding and (2) coordinate the α-amino group or α-carboxyl group, depending on the orientation of the main chain atoms in Pyl or Pyl analogues. This latter interaction occurs through a water-mediated hydrogen bond (Figure 5.5b,c)[25-27]. Evolved MmOmeRS possesses a non-hydrogen bond donor–acceptor at position 346 (Figure 5.4b). Curiously, previous results demonstrate that mutation of Asn346 to Ala significantly impairs aminoacylation activity in PylRS[27], suggesting some form of compensatory changes in the selected MmOmeRS.

The X-ray crystal structure of the MmOmeRS-Ome-AMP-PNP-Mg^{2+} ternary complex illustrates how the A302T mutation positions the hydroxyl-bearing Thr side chain as a bump on the

active site surface in proximity to the β-sheet defining the floor of the amino acid binding site. On this floor, the other critical change, N346V, projects up and into the active site (Figure 5.4b). The Thr hydroxyl moiety sterically and electronically compensates for the loss of the hydrogen bond contributed by Asn346 in wild-type PylRS. In particular, Thr302 in MmOmeRS assumes two distinct rotameric conformations (Figure 5.4d), both of which hydrogen bond directly to the α-carboxyl group of Ome without the intervention of a water molecule (Figure 5.5a). In addition, the N346V mutation abolishes the previously observed hydrogen-bonding interaction between Asn346 and the side chain carbonyl group of Pyl or Pyl analogues in PylRS.

In wild-type PylRS, Leu305, Tyr306, Cys348, and Trp417 form a deep hydrophobic pocket in the active site that accommodates the elongated methyl-pyrroline ring of Pyl. In MmOmeRS, the C348W mutation places the indole moiety of Trp348 directly into this elongated cavity, thereby shortening the depth of the cavity (Figure 5.4b). Consequently, the active site volume decreases from 2883 to 2174 $Å^3$ when comparing PylRS (PDB ID: 2ZCE) to MmOmeRS, respectively. In PylRS, the mutation C348A does not exert a measurable effect on activity, suggesting that this residue does not contribute significantly to stabilization of the transition state accompanying amino acid charging[27]. In contrast, MmOmeRS uses Trp348 to not only facilitate recognition of Ome but also to stabilize the *O*-methyl group on Ome through a quadrupole-dipole interaction that is a direct result of the indole plane sitting perpendicular to the *O*-methyl moiety of Ome. Finally, the V401L mutation extends the wild-type Val side chain by a methylene group allowing the Leu side chain moiety to reach and better accommodate the shorter but wider Ome side chain (Figure 5.4b). In total, these selected mutations in MmOmeRS widen part of the amino acid binding cavity while condensing its depth, ultimately providing a complementary shape to accommodate the bulky phenyl ring of Ome.

5.3.6 Recognition of the Ome Main Chain and Coordination of the Adenyl-Donor AMP-PNP-Mg^{2+}

In MmOmeRS, the α-amino and α-carboxyl groups of Ome rotate counterclockwise 103° along the Cα–Cβ bond relative to their positions in PylRS; this rotation effectively flips the positions of

the α-amino and α-carboxyl groups in Ome-bound MmOmeRS compared to Pyl-bound PylRS. This interchange results in distinctly different stabilizing interactions (Figure 5.5a,b). For example, in wild-type PylRS, the α-amino group of Pyl interacts with the side chain of Asn346*via* a water-mediated interaction. In MmOmeRS, the α-amino group of Ome hydrogen-bonds *via*a water-mediated interaction to one of two observed rotameric conformations of Ser399. Conspicuously, Ser399 resides distal to position 346 along the surface of the active site cavity. In PylRS, Arg330 interacts through electrostatic hydrogen bonds with the negatively charged α-carboxyl group of Pyl; this electrostatic role is replaced by a neutral hydrogen-bonding network in MmOmeRS triggered by the A302T mutation. Additionally, the second oxygen atom of the α-carboxyl group of Ome is further solvated by a water molecule that is unusually within hydrogen-bonding distance (2.9 Å) to the thioether sulfur atom of Met344; in contrast, for Pyl-bound PylRS (PDB ID: 2ZCE), this distance extends to 3.6 Å, which is more consistent with a van der Waals interaction rather than a more stabilizing hydrogen-bonding interaction. The α-amino-carboxyl group flip was also seen in the previously reported crystal structure of PylRS complexed with AMP-PNP and Boc-Lys (PDB ID: 2ZIN). In this latter case, the Asn346 side chain hydrogen bonds to the α-carboxyl moiety through a water molecule instead of with the α-amino group of Boc-Lys as seen in other wild-type PylRS crystal structures (Figure 5.5c)[26].

The main chain atoms of Pyl sit in a catalytically competent orientation consistent with the expected in-line nucleophilic attack of an oxygen atom from the α-carboxyl group on the α-phosphate atom of AMP-PNP (Figure 5.6b)[25, 27] By contrast, in MmOmeRS, the α-amino group of Ome sits proximal to the α-phosphate of AMP-PNP through an intervening network of water-mediated interactions (Figure 5.6a). The orientation of the main chain atoms in Ome is not uncommon among structures of PylRS complexed with various analogues; for example, Boc-Lys also has its α-amino group close to the α-phosphate of AMP-PNP (PDB ID: 2ZIN, Figure 5.6c)[26]. These Uaas could readily rotate the α-carboxyl and α-amino groups along the Cα-Cβ bond to achieve a catalytically competent orientation.

5.4 Conclusions

A challenge facing the deployment of tRNA/RS pairs in mammalian cells is the inability to streamline the directed evolution of such pairs to produce orthogonal and catalytically efficient tRNA/RS sets possessing high specificity for a particular Uaa. The current solution to this bottleneck consists of evolving *E. coli* RSs in yeast and transferring the resultant mutants to mammalian cells[19, 20]. The major drawbacks of using yeast include its lower transformation efficiency and its restriction to less efficient selection methods compared to those originally developed in *E. coli*[4, 16]. Unfortunately, *M. jannaschii* $tRNA^{Tyr}$/TyrRS, which has been extensively evolved in *E. coli*, does not act in an orthogonal manner in mammalian cells. On the other hand, the more recently employed $tRNA_{CUA}^{Pyl}$/PylRS pair does have a currently unique property of being orthogonal in *E. coli* and mammalian cells, possibly due to its special role encoding the rarer amino acid Pyl that is not used by *E. coli* or mammalian cells. Thus, PylRS mutants evolved in *E. coli* can be directly transferred to mammalian cells for efficient Uaa incorporation.

To date, nearly all Uaas incorporated by evolved PylRS mutants contain a carbonyl group at the N^{ε} position of Pyl. Our report represents a successful breakthrough regarding this limitation. The chosen amino acid for driving selection, namely, Ome, does not contain this carbonyl moiety of Pyl, and its side chain is significantly different from Pyl and reported Pyl analogues. Another striking switch in synthetase substrate specificity is the directed evolution of *E. coli*LeuRS to incorporate a fluorescent Uaa dansylalanine[28]. However, the *E. coli* LeuRS must be evolved in yeast as it is not orthogonal in *E. coli*. Although the previously reported evolution of PylRS to incorporate Phe-based Uaas into proteins in *E. coli* was somewhat promising, incorporation was achieved in *E. coli* grown in glycerol minimal media only[24]. Minimal media is used in published reports to greatly reduce the mis-incorporation of natural amino acids[23], raising concerns with regard to the specificity of these evolved PylRSs when used in rich media, a condition necessary for fermentation, mammalian cell cultures, and whole organism engineering. Here, the evolved tRNA/MmOmeRS pair efficiently and specifically incorporated Ome into proteins both in *E. coli* using rich media and in mammalian cells, and we have been able to assign this high level of specificity to interactions revealed through structural analysis.

The crystal structure of MmOmeRS reported here illustrates the stereochemical features governing Uaa recognition at the molecular level. It was previously thought that the carbonyl group of the Pyl side chain played a critical role during the aminoacylation reaction catalyzed by PylRS[26]; Asn346 in wild type PylRS interacts simultaneously with this side chain carbonyl group and the α-amino group of Pyl or a Pyl analogue (PDB IDs: 2ZCE, 2ZIN, 2Q7G)[25-27]. In contrast, we chose to subject position 346 to mutational analysis. The resulting mutation selected at this position, N346V, creates an ideal hydrophobic pocket for the encapsulation of the Ome phenyl ring. Additionally, the A302T mutation complements the loss of the Asn346 hydrogen bond by forming a network of interactions with the main chain of the Ome substrate. These surprising observations indicate that bold yet progressive approaches toward library design allowed for access to regions of sequence space not typically explored due to deleterious effects of one mutation, suggesting that context dependence *via* compensatory mechanisms encouraged the successful evolution of MmOmeRS.

A similar interaction network is observed in several other class II amino acid RSs. For example, in the crystal structure of HisRS from *E. coli* (PDB ID: 2EL9), Thr85 (A302T in MmOmeRS) is within hydrogen-bonding distance of the α-amino group of the His substrate, while Gly129 (N346V in MmOmeRS) plays no apparent role in substrate binding. Interestingly, Thr85 is a conserved residue among all HisRSs, and Gly129 is partially conserved, indicating that a naturally evolved stabilization mechanism for some RSs is comparable to the laboratory evolved substrate binding of MmOmeRS reported here. This switch in the stereochemistry of substrate recognition in MmOmeRS due to the epistatic nature of two laboratory evolved mutations, N346V and A302T, has not been reported in any engineered RS to date. Like the fixed Y384F PylRS mutation that served as a starting point for this current study, the fixation of the N346V and A302T mutations may be a necessary step forward for the design and directed evolution of PylRS-based mutant synthetases that possess an even greater expansion of selectivity for Uaas.

Apart from these two critical positions, a third mutation is C348W. The X-ray crystal structure of the ternary complex again demonstrates the stereochemistry of this selected change. The bulky side

chain indole moiety of Trp348 effectively shortens the elongated amino acid binding tunnel of wild type PylRS to accommodate the shorter Ome side chain.

In some synthetases, the tRNA binding event alters the conformation of the amino acid binding cleft and yields changes in amino acid substrate binding affinity[29, 30]. Although the crystal structure for the *M. mazei* $tRNA^{Pyl}$/PylRS complex has not been reported, the crystal structure for the *Desulfitobacterium hafniense* $tRNA^{Pyl}$/PylRS complex has previously been solved to a resolution of 3.1 Å[31]. A comparison of this $tRNA^{Pyl}$/PylRS complex to our *M. mazei*OmeRS structure reveals that the synthetase components of each structure are very similar; superposition of Cα atoms between the two structures yields an rmsd value of 1.188 Å. The only active site motif that subtly changes conformation is the β9 strand where the V401L mutation resides. However, this subtlety can be accounted for by differences in amino acid residues between the two proteins at these positions and is most likely not relevant to the tRNA binding event (**Supplementary Figure 5.2**). It is of note that the crystal structure of the $tRNA^{Pyl}$/PylRS complex models the β7-β8 loop in a solvent exposed orientation (away from the active site in an open position). This loop is absent from the MmOmeRS structure, and therefore its orientation in a complex with tRNA is unpredictable. Overall, the absence of the tRNA molecule in the crystal structure of MmOmeRS most likely does not impact library design in a directed evolution experiment. However, it is of note that structural differences between aaRS and tRNA/aaRS complex, if present, should be taken into consideration for mutant aaRS library design. In addition, rational design of amino acid specificity between natural amino acids in an aaRS has been accomplished and shown to require many substitutions throughout the catalytic domain[32]. Therefore, increase of mutant aaRS library size to target more residues for mutation may further expand the Uaa diversity and enhance incorporation efficiency.

In summary, this study presents the first stereochemical explanation for the laboratory evolved switch in PylRS substrate specificity from long aliphatic moieties to shortened aromatic side chains; this information provides both a molecular and a structural basis for engineering synthetases with expanded repertoires of Uaas that can be deployed in an orthogonal manner in complex hosts such as mammalian cells and multicellular organisms.

5.5 Method

5.5.1 Plasmid Construction

Plasmids were constructed using standard cloning procedures and confirmed by DNA sequencing. The pTak-GFPtag plasmid contained the *M. mazei* $tRNA_{CUA}^{Pyl}$ and the GFP(182TAG) gene tagged to a His_6 polypeptide appended to the C-terminus. Transcription of the $tRNA_{CUA}^{Pyl}$gene was driven by the *lpp* promoter and terminated by the *rrnC* terminator. GFP(182TAG) gene transcription was driven by the T5 promoter under regulation of the *lac* operator sequence and transcription was terminated with the λt_0 terminator. Plasmid pTak-Myo4TAG was created by replacing the GFP(182TAG) gene in pTak-GFPtag with the sperm whale myoglobin gene containing an amber codon at position 4. Plasmid pLei-Myo4TAG is identical to pTak- Myo4TAG except the *M. mazei* $tRNA_{CUA}^{Pyl}$ was replaced with the *M. jannaschii* $tRNA_{CUA}^{Tyr}$, J17[5]. Plasmid pMPcua-OmeRS was derived from pEYcua-TyrRS[19] by replacing the *E. coli*$tRNA_{CUA}^{Tyr}$ with the *M. mazei* $tRNA_{CUA}^{Pyl}$ and the *E. coli* TyrRS with MmOmeRS. DNA sequences and primers are listed in the supporting information.

5.5.2 Library Construction and *in Vivo* Selection

Three rounds of overlapping PCRs were performed on the *M. mazei* PylRS gene to randomize the codons for each selected active site residue using oligonucleotides containing NNK at each site destined for mutagenic diversification. The amplified full-length PCR product was ligated into the precut pBK5-MmPylRS plasmid to afford the mutant DNA library. The library was electroporated into DH10βT1 competent cells harboring pREP-PylT[10]. Selection was carried out on LB plates containing 12.5 μg mL^{-1} tetracycline, 50 μg mL^{-1} kanamycin, 30 μg mL^{-1}chloramphenicol and 1 mM Ome.

5.5.3 Protein Expression and Uaa Incorporation in *E. coli*

To translationally incorporate Uaas into myoglobin in *E. coli*, DH10β cells were transformed with pTak-Myo4TAG/pBK-MmOmeRS and pLei-Myo4TAG/pBK-MjOmeRS, respectively. For each sample, a colony was picked and grown overnight in 50 mL 2xYT supplemented with 30 μg mL^{-}

[1]chloramphenicol and 50 μg mL^{-1} kanamycin at 37 °C. This culture was used to inoculate 1 L of 2xYT containing antibiotics and supplemented with or without 1 mM of Ome. When OD_{600}reached 0.5, cells were induced for protein expression by adding 0.5 μM IPTG. After 3 h, cells were pelleted and sonicated to lyse in 40 mL lysis buffer (50 mM TrisHCl, pH 8.0, 500 mM NaCl, 20 mM imidazole pH 8.0, 1% (v/v) Tween 20, 10% (v/v) glycerol, 10 mM 2-mercaptoethanol, and 0.5 mg mL^{-1} lysozyme). Lysed cells were centrifuged for 10 min at 6,000 g, and clarified supernatant was passed through a 0.5 mL column of Ni^{2+}-NTA agarose resin (Qiagen). The column was washed with 10 column volumes of wash buffer (lysis buffer without Tween 20 and lysozyme). Protein was eluted with 1 mL of elution buffer (wash buffer containing 250 mM imidazole pH 8.0). Imidazole was removed from the eluent by buffer exchange using a Microcon Ultracel YM-10 spin column (Millipore). The sample was further purified by fast protein liquid chromatography using a HiLoad 26/60 Superdex 200 gel filtration column in 50 mM TrisHCl, pH 8.0 with 500 mM NaCl. To incorporate Ome into GFP, DH10β cells were transformed with pTak-GFPtag and pBK-MmOmeRS. GFP was purified using the same procedure described above.

5.5.4 Protein Expression and Uaa Incorporation in Mammalian Cells

Uaa incorporation was carried out as previously described[33]. Briefly, HEK293 cells were transfected with pCLHF-GFP(182TAG)[19] and pMPcua-OmeRS. The HeLa-GFP(182TAG) stable cells were transfected with pMPcua-OmeRS or pEYcua-TyrRS. Cells were grown in DMEM supplemented with 10% FBS; 1 mM of Ome was added to or withheld from the media for cells transfected with pMPcua-OmeRS. HEK293 cells were imaged 48 h after transfection with an Olympus IX8I microscope equipped with a Hamamatsu EM-CCD under same conditions for all samples (λ_{ex} = 480 ± 20 nm, λ_{em} = 535 ± 40 nm). Transfected HeLa cells were trypsinized and washed with PBS twice. After resuspending in 1.0 mL PBS and 5 μL of propidium iodide, cells were analyzed with a FACScan (Becton & Dickinson, excitation 488 nm, emission 530 ± 30 nm). For each sample the total fluorescence intensity of 30,000 cells was recorded and was normalized to cells transfected with pEYcua-TyrRS, which expresses the *E. coli*$tRNA_{CUA}^{Tyr}$/TyrRS[19].

5.5.5 Mass Spectrometry

For intact protein analysis, myoglobin samples were analyzed by high resolution Fourier-transform (FT) mass spectrometry (MS) on a Thermo LTQ-Orbitrap XL mass spectrometer[34] (ThermoFisher). Samples were loaded onto a capillary column with integrated spray tip, which was filled with reversed phase material (Zorbax SB C-18, particle size 5 μm). Protein was eluted using 0.1% (v/v) formic acid and a gradient of increasing concentrations of acetonitrile at a flow rate of 300 nL min^{-1}. The eluate was electrosprayed directly into the mass spectrometer. FT-MS spectra were recorded a resolution of 60,000 for a scan range of m/z = 400–1800. This was followed by a select ion scan of the most intense ion at resolution 60,000. Data were charge-deconvoluted using the Thermo Qualbrowser 2.0 Xtract program.

For GFP samples, ~1 μg of protein was first desalted using Microcon Ultracel Y-10 spin columns, and then digested with trypsin (200 ng, Sigma, Proteomics grade) in 50 mM ammonium bicarbonate buffer for 16 h. An aliquot of the resulting digest (5–10%) were dried down and redissolved in 0.1% formic acid. High resolution FT-MS analysis was performed on a Thermo LTQ-Orbitrap XL. Samples were loaded onto the same capillary column described above, eluted and electrosprayed using the same procedures. FT-MS spectra were recorded at a resolution of 60,000 in the Orbitrap FT analyzer followed by MS/MS scans of the top 5 ions in the linear ion trap. MS/MS data were searched against a custom database consisting of the protein of interest using the Mascot algorithm (Matrixscience). For the search, a variable modification at Tyr residues with a mass of 14.01564 (corresponding to the mass difference between Ome and Tyr) was allowed.

5.5.6 Cloning, Expression, and Purification of MmOmeRS

The C-terminal catalytic domain (residues 185–454) of MmOmeRS was amplified from the pBK vector containing full length MmOmeRS by PCR with the following primers: 5′-TTC CCA TGG CGC AAG CGC CCC AGC TCT G-3′ and 5′-CGC AAG CTT TTA CAG GTT AGT AGA AAT ACC ATT GTA ATA GGA CTC GG-3′. The PCR product was cloned into an in-house pET28a(+) vector (Novagen) modified to contain an N-terminal His_8-tag. Both the destination vector and insert were

digested with*Nco* I and *Hind* III and ligated using T4 DNA ligase (New England Biolabs, Inc.). The insert containing plasmid was recovered after initial transformation in *E. coli* BP5α cells (BioPioneer), confirmed by DNA sequencing and transformed into *E. coli* BL21 (DE3) cells (Novagen). One colony was grown at 37 °C in TB media to an OD_{600} of 1.0 and then induced with 1 mM IPTG overnight at 22 °C. The cells were harvested, resuspended in lysis buffer (50 mM HEPES-Na^+, pH 7.5, 500 mM NaCl, 20 mM imidazole, 1% (v/v) Tween-20, 10% (v/v) glycerol, 10 mM 2-mercaptoethanol) containing lysozyme (0.5 mg mL^{-1}), and stirred for 45 min at 4 °C. The solution was then sonicated and centrifuged to clarify the supernatant, which was subsequently passed through a 1 mL column of Ni-NTA agarose resin (Qiagen). The column was washed with ten column volumes of lysis buffer and wash buffer (lysis buffer without Tween20) prior to elution of the His_8-tagged MmOmeRS with elution buffer (wash buffer containing 250 mM imidazole). The eluent was dialyzed overnight in dialysis buffer (5 mM $MgCl_2$, 300 mM NaCl and 10 mM HEPES-Na^+, pH 7.5) and passed through a HiLoad 16/16 Superdex^{TM} 200 prep grade (GE Healthcare) gel filtration column. The purified MmOmeRS protein was concentrated to 10 mg mL^{-1} and stored at −80 °C.

5.5.7 Crystallization, Structure Solution, and Refinement

Co-crystallization of MmOmeRS with Ome and AMP-PNP was accomplished using a previously described method[25]. Briefly, 1 μL of MmOmeRS (10 mg mL^{-1}) containing 5 mM Ome and 5 mM AMP-PNP was added to 1 μL of reservoir solution (100 mM TrisHCl, pH 8.0, 11% (w/v) PEG 2000 MME) at 25 °C. Crystals were cryostabilized in the reservoir solution supplemented with 30% (v/v) ethylene glycol and frozen in liquid nitrogen.

Diffraction data were collected on ALS beamline 8.2.2 (Lawrence Berkeley National Laboratory) using an ADSC Q315 CCD detector. Images were processed and scaled with XDS[35]. A model of MmOmeRS was constructed using PylRS as a reference structure, and molecular replacement followed by rounds of refinement were performed using the ccp4 programs Molrep and Refmac, respectively[36]. The program COOT was used for model visualization and building throughout the

refinement process[37]. PROCHECK was used for quality assessment of the final MmOmeRS structure[38]. The data and refinement statistics are presented in **Table 5.1**.

5.6 Accession Codes

MmOmeRS: PDB ID 3QTC

5.7 ACKNOWLEDGEMENTS

We thank Dr. G. Louie for help with data processing of crystallographic data and Dr. W. Fisher and J. Read for help with mass spectrometry. N.D. acknowledges support of an HHMI predoctoral fellowship. This material is based in part upon work supported by the National Science Foundation under Award No. MCB-0645794 (J.P.N.) and by the Salk Institute Cancer Center Award No. P30CA014195 from the National Cancer Institute (J.P.N. and L.W.). J.P.N. is an investigator with the Howard Hughes Medical Institute. Portions of this research were conducted at the Advanced Light Source (ALS), a national user facility operated by Lawrence Berkeley National Laboratory, on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences. The Berkeley Center for Structural Biology is supported in part by the Department of Energy, Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences. L.W. acknowledges support from the Ray Thomas Edwards Foundation, March of Dimes Foundation (#5-FY08-110), California Institute for Regenerative Medicine (RN1-00577-1), and National Institutes of Health (1DP2OD004744-01).

Chapter 5 is a reprint in full of the material as it appears in ACS Chemical Biology, (2011), Takimoto, J. K.; Dellas, N.; Noel, J. P.; Wang, L., Stereochemical Basis for Engineered Pyrrolysyl-tRNA Synthetase and the Efficient in Vivo Incorporation of Structurally Divergent Non-native Amino Acids. *ACS Chem Biol*. The dissertation author and Dellas, N contributed equally to this manuscript.

5.8 Supplemental information

5.8.1 DNA sequences

Sperm whale myoglobin (Myo4TAG)

atgactagtgttctgtaggaaggtgaatggcagctggttctgcatgtttgggctaaagttgaagctgacgtcgctggtcatggtcaggacatcttgattcg
actgttcaaatctcatccggaaactctggaaaaattcgatcgtttcaaacatctgaaaactgaagctgaaatgaaagcttctgaagatctgaaaaaacatg
gtgttaccgtgttaactgccctaggtgctatccttaagaaaaaagggcatcatgaagctgagctcaaaccgcttgcacaatcgcatgctactaaacataa
gatcccgatcaaatacctggaattcatctctgaagcgatcatccatgttctgcattctagacatccaggtgacttcggtgctgacgctcagggtgctatga
acaaagctctcgagctgttccgtaaagatatcgctgctaagtacaaagaactgggttaccagggatctcatcaccatcaccatcactaa

GFP (GFPtag)

atgactagtgtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgt
gtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcg
tgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgt
ccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgcatcgagc
tgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgacaagc
agaagaacggcatcaaggccaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgccgaccactagcagcagaacaccccccat
cggcgacggccccgtgctgctgcccgacaaccactacctgagcacccagtccgccctgagcaaagaccccaacgagaagcgcgatcacatggtc
ctgctggagttcgtgaccgccgccgggatcactctcggcatggacgagctgtacaagagatctcatcaccatcaccatcactaa

MmPylRS-Y384F

atggataaaaagcctctgaacactctgatttctgcgaccggtctgtggatgtcccgcaccggcaccatccacaaaatcaaacaccatgaagttagccgt
tccaaaatctacattgaaatggcttgcggcgatcacctggttgtcaacaactcccgttcttctcgtaccgctcgcgcactgcgccaccacaaatatcgca
aaacctgcaaacgttgccgtgttagcgatgaggacctgaacaaattcctgaccaaagctaacgaggatcagacctccgtaaaagtgaaggtagtaag
cgctccgacccgtactaaaaaggctatgccaaaaagcgtggcccgtgccccgaaacctctggaaaacaccgaggcggctcaggctcaaccatccg
gttctaaattttctccggcgatcccagtgtccacccaagaatctgtttccgtaccagcaagcgtgtctaccagcattagcagcatttctaccggtgctaccg
cttctgcgctggtaaaaggtaacactaacccgattactagcatgtctgcaccggtacaggcaagcgccccagctctgactaaatcccagacggaccgt
ctggaggtgctgctgaacccaaaggatgaaatctctctgaacagcggcaagcctttccgtgagctggaaagcgagctgctgtctcgtcgtaaaaagg

atctgcaacagatctacgctgaggaacgcgagaactatctgggtaagctggagcgcgaaattactcgcttcttcgtggatcgcggtttcctggagatca

aatctccgattctgattccgctggaatacattgaacgtatgggcatcgataatgataccgaactgtctaaacagatcttccgtgtggataaaaacttctgtct

gcgtccgatgctggccccgaacctgtacaactatctgcgtaaactggaccgtgccctgccggacccgatcaaaattttcgagatcggtccttgctaccg

taaagagtccgacggtaaagagcacctggaagaattcaccatgctgaacttctgccagatgggtagcggttgcacgcgtgaaaacctggaatccatta

tcaccgacttcctgaatcacctgggtatcgatttcaaaattgttggtgacagctgtatggtgtttggcgatacgctggatgttatgcacggcgatctggagc

tgtcttccgcagtagtgggcccaatcccgctggatcgtgagtggggtatcgacaaaccttggatcggtgcgggttttggtctggagcgtctgctgaaag

taaaacacgacttcaagaacatcaaacgtgctgcacgttccgagtcctattacaatggtatttctactaacctgtaa

MmPylT

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

MmBzoRS

atggataaaaagcctctgaacactctgatttctgcgaccggtctgtggatgtcccgcaccggcaccatccacaaaatcaaacaccatgaagttagccgt

tccaaaatctacattgaaatggcttgcggcgatcacctggttgtcaacaactcccgttcttctcgtaccgctcgcgcactgcgccaccacaaatatcgca

aaacctgcaaacgttgccgtgttagcgatgaggacctgaacaaattcctgaccaaagctaacgaggatcagacctccgtaaaagtgaaggtagtaag

cgctccgacccgtactaaaaaggctatgccaaaaagcgtggcccgtgccccgaaacctctggaaaacaccgaggcggctcaggctcaaccatccg

gttctaaattttctccggcgatcccagtgtccacccaagaatctgtttccgtaccagcaagcgtgtctaccagcattagcagcatttctaccggtgctaccg

cttctgcgctggtaaaaggtaacactaacccgattactagcatgtctgcaccggtacaggcaagcgccccagctctgactaaatcccagacggaccgt

ctggaggtgctgctgaacccaaaggatgaaatctctctgaacagcggcaagcctttccgtgagctggaaagcgagctgctgtctcgtcgtaaaaagg

atctgcaacagatctacgctgaggaacgcgagaactatctgggtaagctggagcgcgaaattactcgcttcttcgtggatcgcggtttcctggagatca

aatctccgattctgattccgctggaatacattgaacgtatgggcatcgataatgataccgaactgtctaaacagatcttccgtgtggataaaaacttctgtct

gcgtccgatgctggccccgaacttttataactatctgcgtaaactggaccgtgccctgccggacccgatcaaaattttcgagatcggtccttgctaccgt

aaagagtccgacggtaaagagcacctggaagaattcaccatgctggctttcactcagatgggtagcggttgcacgcgtgaaaacctggaatccattat

caccgacttcctgaatcacctgggtatcgatttcaaaattgttggtgacagctgtatggtgtttggcgatacgctggatgttatgcacggcgatctggagct

gtcttccgcagtagtgggcccaatcccgctggatcgtgagtggggtatcgacaaacctgtgatcggtgcgggttttggtctggagcgtctgctgaaagt

aaaacacgacttcaagaacatcaaacgtgctgcacgttccgagtcctattacaatggtatttctactaacctgtaa

MmOmetRS

atggataaaaagcctctgaacactctgatttctgcgaccggtctgtggatgtcccgcaccggcaccatccacaaaatcaaacaccatgaagttagccgt
tccaaaatctacattgaaatggcttgcggcgatcacctggttgtcaacaactcccgttcttctcgtaccgctcgcgcactgcgccaccacaaatatcgca
aaacctgcaaacgttgccgtgttagcgatgaggacctgaacaaattcctgaccaaagctaacgaggatcagacctccgtaaaagtgaaggtagtaag
cgctccgacccgtactaaaaaggctatgccaaaaagcgtggcccgtgccccgaaacctctggaaaacaccgaggcggctcaggctcaaccatccg
gttctaaattttctccggcgatcccagtgtccacccaagaatctgtttccgtaccagcaagcgtgtctaccagcattagcagcatttctaccggtgctaccg
cttctgcgctggtaaaaggtaacactaacccgattactagcatgtctgcaccggtacaggcaagcgccccagctctgactaaatcccagacggaccgt
ctggaggtgctgctgaacccaaaggatgaaatctctctgaacagcggcaagcctttccgtgagctggaaagcgagctgctgtctcgtcgtaaaaagg
atctgcaacagatctacgctgaggaacgcgagaactatctgggtaagctggagcgcgaaattactcgcttcttcgtggatcgcggtttcctggagatca
aatctccgattctgattccgctggaatacattgaacgtatgggcatcgataatgataccgaactgtctaaacagatcttccgtgtggataaaaacttctgtct
gcgtccgatgctgactccgaacctgtacaactatcttcgtaaactggaccgtgccctgccggacccgatcaaaattttcgagatcggtccttgctaccgt
aaagagtccgacggtaaagagcacctggaagaattcaccatgctggttttctggcagatgggtagcggttgcacgcgtgaaaacctggaatccattat
caccgacttcctgaatcacctgggtatcgatttcaaaattgttggtgacagctgtatggtgtttggcgatacgctggatgttatgcacggcgatctggagct
gtcttccgcattggtgggcccaatcccgctggatcgtgagtggggtatcgacaaaccttggatcggtgcgggttttggtctggagcgtctgctgaaagt
aaaacacgacttcaagaacatcaaacgtgctgcacgttccgagtcctattacaatggtatttctactaacctgtaa

Primers:

GFPtag

DJ654 5’-GCA GAT CTC TTG TAC AGC TCG TCC ATG CC- 3’

DJ655 5’-ATG ACT AGT GTG AGC AAG GGC GAG G- 3’

Myo4TAG

JT762 5’-GAG ACT AGT GTT CTG TAG GAA GGT GAA TGG CAG CTG GTT C- 3’

JT763 5’ -GAG GGA TCC CTG GTA ACC CAG TTC TTT GTA CTT AGC AG- 3’

pBK5-MmPylRS

JT552 5’ -GTG TAC ACA TAT GGA TAA AAA GCC TCT GA- 3’

JT632 5’ -GCC GTG TTA GCG ATG AGG ACC TGA ACA AAT TCC TGA CC- 3’

JT633 5’ -GGT CAG GAA TTT GTT CAG GTC CTC ATC GCT AAC ACG GC- 3’

JT634 5’ -GAA ACT GCA GCC ACC CGC CTG GTT ACA GGT TAG TAG AAA TAC CAT TGT AAT AGG- 3’

MmBzoRS library

JT552 5’ -GTG TAC ACA TAT GGA TAA AAA GCC TCT GA- 3’

JT639 5’ -CAG GGC ACG GTC CAG TTT ACG MNN ATA GTT MNN MNN GTT CGG GG- 3’

JT555 5’ -AAA CTG GAC CGT GCC CTG CC- 3’

JT622 5’ -CGT GCA ACC GCT ACC CAT CTG MNN GAA MNN CAG CAT G- 3’

JT623 5’ -ATG GGT AGC GGT TGC ACG- 3’

JT640 5’ -CGA TCC AGC GGG ATT GGG CC- 3’

JT625 5’ -GGC CCA ATC CCG CTG GAT CGT GAG TGG GGT ATC GAC AAA CCT NNK ATC GGT GC- 3’

JT634 5’ -GAA ACT GCA GCC ACC CGC CTG GTT ACA GGT TAG TAG AAA TAC CAT TGT AAT AGG- 3’

MmOmeRS library

JT552 5’ -GTG TAC ACA TAT GGA TAA AAA GCC TCT GA- 3’

JT620 5’ -CAG GGC ACG GTC CAG TTT ACG MNN ATA GTT GTA CAG GTT CGG MNN CAG CAT C- 3’

JT621 5’ -AAA CTG GAC CGT GCC CTG- 3’

JT624 5’ -GAT CCA GCG GGA TTG GGC CCA CMN NTG CGG AAG- 3’

JT634 5’ -GAA ACT GCA GCC ACC CGC CTG GTT ACA GGT TAG TAG AAA TAC CAT TGT AAT AGG- 3’

JT625 5’ -GGC CCA ATC CCG CTG GAT CGT GAG TGG GGT ATC GAC AAA CCT NNK ATC GGT GC- 3’

pMPcua-OmeRS \ pMPcua-BzoRS

JT768 5’ -CAC AGA TCT CCG GAA ACC TGA TCA TGT AGA TCG AAT GGA C- 3’

JT769 5’ -GAG AAG CTT CAA AAA AAG GAC CCG GAA ACC CCG GGA ATC TAA C-3’

JT773 5’ -GAA CTC GAG GAT AAA AAG CCT CTG AAC ACT CTG ATT TCT G- 3’

JT774 5’ - TCT GCT AGC TTA CAG GTT AGT AGA AAT ACC ATT GTA ATA GGA CTC G- 3’

5.9 FIGURES

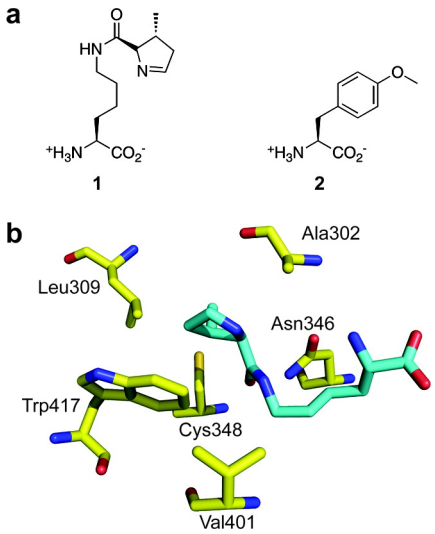


Figure 5.1: Evolving *M. mazei* PylRS to aminoacylate Uaas bearing aromatic side chains. (a) Structures of pyrrolsyine **1** and *O*-methyl-L-tyrosine (Ome, **2**). (b) Structure of the active site of *M. mazei* PylRS (PDB ID: 2ZCE). Pyl is shown in cyan. The codons of the six residues highlighted in yellow were diversified with NNK oligonucleotides by PCR to generate a mutant library. MmOmeRS was selected from this library using Uaa Ome.

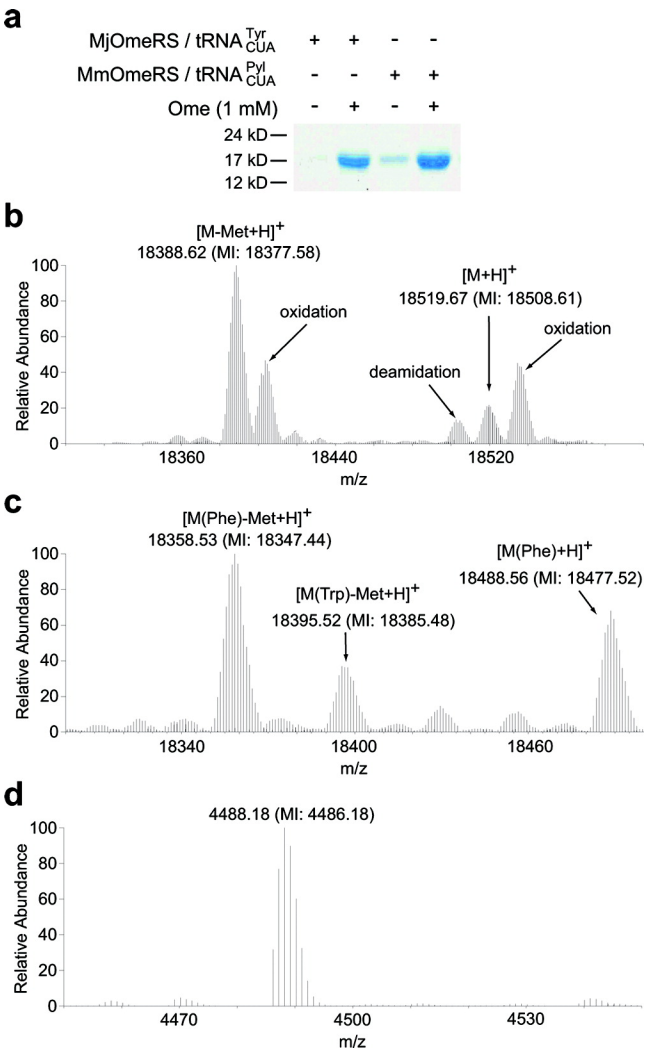


Figure 5.2: MmOmeRS specifically incorporates the cognate Uaa Ome into proteins in*E. coli*. (a) Suppression of the amber codon in sperm whale myoglobin at position 4 by *M. janaschii* OmeRS-$tRNA_{CUA}^{Tyr}$ and *M. mazei* OmeRS-$tRNA_{CUA}^{Pyl}$ with and without Ome, respectively. Myoglobin proteins were purified using Ni-NTA resin, separated by SDS-PAGE and stained with Commassie. Samples were normalized for constant cell numbers for each lane. (b, c) High resolution FT ion trap MS of the intact myoglobin protein expressed in the presence of MmOmeRS-$tRNA_{CUA}^{Pyl}$supplemented with 1 mM Ome (b) and without Ome in the growth media (c). Average and monoisotopic (indicated by MI) masses are labeled. In the absence of Ome, MmOmeRS incorporates Phe and Trp at low efficiencies. In the presence of Ome, only the Uaa was found to be incorporated at the UAG-encoded position by MmOmeRS. A peak with monoisotopic mass of 18508.61 was detected corresponding to myoglobin containing Ome at position 4; oxidation and deamidation peaks for this protein were also detected and labeled. Another peak with monoisotopic mass of 18377.58 was detected corresponding to Ome-containing myoglobin lacking the initiating Met; an oxidation for this species was also detected. (d) MS analysis of GFP expressed in the presence of MmOmeRS-$tRNA_{CUA}^{Pyl}$ supplemented with 1 mM Ome. The spectrum shows the intact precursor ions (charge deconvoluted) for one tryptic fragment of GFP ($^{+}H_3N$-HNIEDGSVQLADH**X**QQNTPIGDGPVLLPDNHYLSTQ SALSK-CO_2H, **X** representing the UAG encoded site). The labeled peak corresponds to the incorporation of Ome at the UAG encoded position. Tandem MS spectrum of this peptide is shown in **Supplementary Figure 5.1**, which unambiguously confirms the peptide sequence and Ome at position **X**.

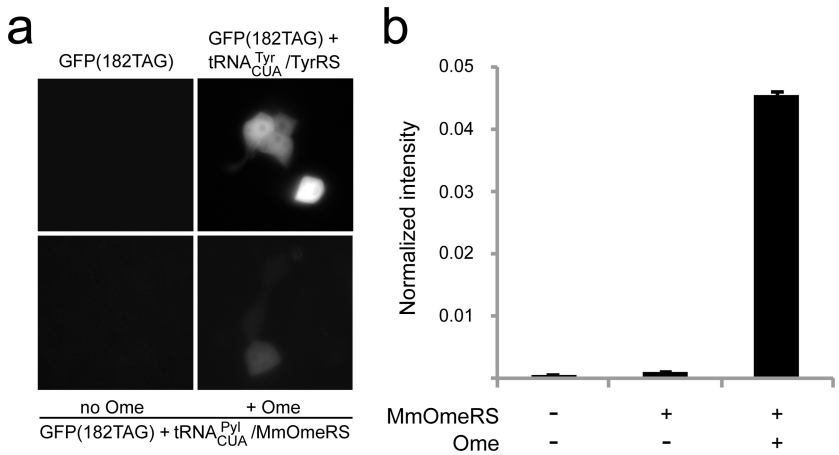


Figure 5.3: MmOmeRS incorporates the cognate Uaa Ome in mammalian cells. (a) Fluorescent imaging of HEK293 cells transfected with different genes as labeled. Cells transfected with the *E. coli* $tRNA_{CUA}^{Tyr}$/TyrRS pair, which incorporates Tyr at the UAG codon, were used as a positive control. For cells transfected with the GFP(182TAG) reporter gene and the $tRNA_{CUA}^{Pyl}$/MmOmeRS pair, green GFP fluorescence was detected in cells cultured in media containing 1 mM Ome, indicating that the Uaa is incorporated into GFP. No GFP fluorescence was detected in cells cultured in media lacking Ome. (b) Incorporation efficiency of MmOmeRS measured with flow cytometry. The stable HeLa-GFP(182TAG) reporter cell line was transfected with the $tRNA_{CUA}^{Pyl}$/MmOmeRS pair and grown in media with and without 1 mM Ome. Cellular GFP fluorescence intensities were measured by flow cytometry and normalized to cells transfected with the*E. coli* $tRNA_{CUA}^{Tyr}$/TyrRS pair. Error bars represent standard error of mean (SEM, $n = 3$).

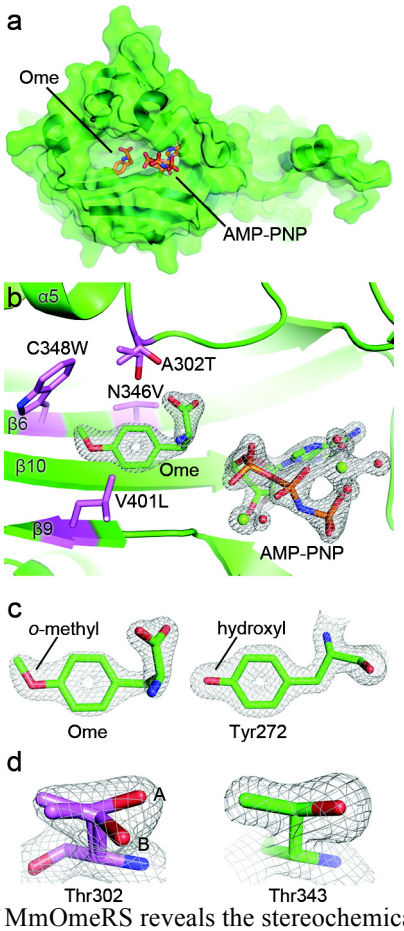


Figure 5.4: Crystal structure of MmOmeRS reveals the stereochemical basis for Ome selectivity. (a) Global view of the crystal structure of MmOmeRS expressed as a filled surface. (b) Depiction of the four mutations defining the transformation of PylRS to MmOmeRS in the crystal structure of MmOmeRS complexed with Ome, Mg^{2+}, and AMP-PNP. Both the Uaa (Ome) and the non-hydrolyzable ATP-analogue (AMP-PNP) are shown together with the calculated SIGMAA-weighted $2F_o - F_c$ electron density contoured at 2σ and 2.5σ, respectively. (c) Comparison of the calculated SIGMAA-weighted $2F_o - F_c$ electron density contoured at 2σ for both Ome (left) and a structurally related amino acid Tyr272 (right) demonstrates that the quality of electron density is excellent for both protein residues and substrates. Moreover, the *O*-methyl moiety is unequivocally present in Ome. (d) Differences in the shapes of the electron density contoured at 1.75σ for both the active site Thr302 (left) and a second Thr343 (right) support the presence of two alternate conformers for Thr302 (labeled A and B).

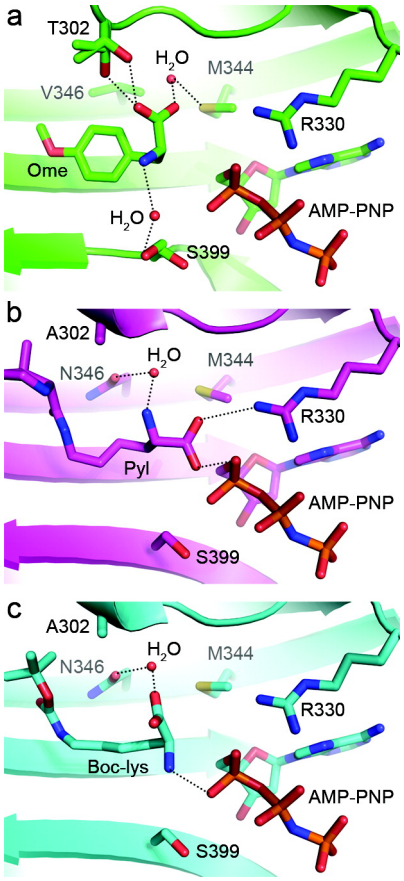


Figure 5.5: Structural depiction of interactions between the amino acid substrate and the surrounding active site residues within MmOmeRS and PlyRS. (a) Interactions between main chain atoms of the Ome substrate and surrounding protein residues in the crystal structure of MmOmeRS complexed with Ome, Mg^{2+}, and AMP-PNP. Both Thr302 and Ser399 exist in two rotameric conformations. Dashed lines between atoms indicate that they are within hydrogen bonding distance. (b) Interactions between main chain atoms of Pyl and surrounding protein residues in the crystal structure of PylRS complexed with Pyl, Mg^{2+}, and AMP-PNP (PDB ID: 2ZCE). (c) Interactions between main chain atoms of Boc-Lys and surrounding protein residues in the crystal structure of PylRS complexed with Boc-Lys and AMP-PNP (PDB ID: 2ZIN).

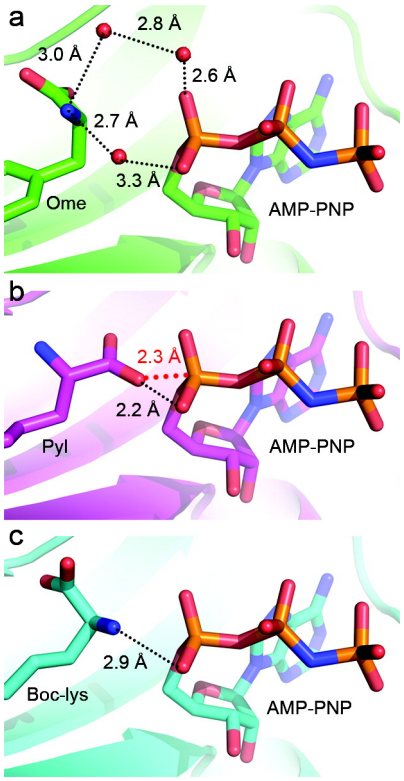


Figure 5.6: Differences in coordination geometry between an ATP analogue and various amino acid substrates in PylRS and MmOmeRS. (a) Indirect coordination of the Ome amino group to AMP-PNP is mediated by several waters in the crystal structure of MmOmeRS complexed with Ome, Mg^{2+}, and AMP-PNP. The main chain carboxyl moiety sits in an orientation that is not directly in-line for attacking the α-phosphate of AMP-PNP (ATP). (b) Direct coordination of an oxygen atom from the carboxyl group of Pyl to the α-phosphate of AMP-PNP in the crystal structure of PylRS complexed with Pyl and AMP-PNP (PDB ID: 2ZCE). Unlike in the case of MmOmeRS, the main chain carboxyl group of Pyl resides in an orientation consistent with adenylation. The red dashed line represents the distance gap between the two atoms that undergo bond formation during the aminoacylation reaction. (c) Hydrogen-bonding network of the amino group of Boc-Lys and an oxygen atom of the α-phosphate of AMP-PNP in the crystal structure of PylRS complexed with Boc-Lys and AMP-PNP (PDB ID: 2ZIN). This Uaa sits in an orientation similar to that observed in the MmOmeRS-Ome-AMP-PNP-Mg^{2+} complex.

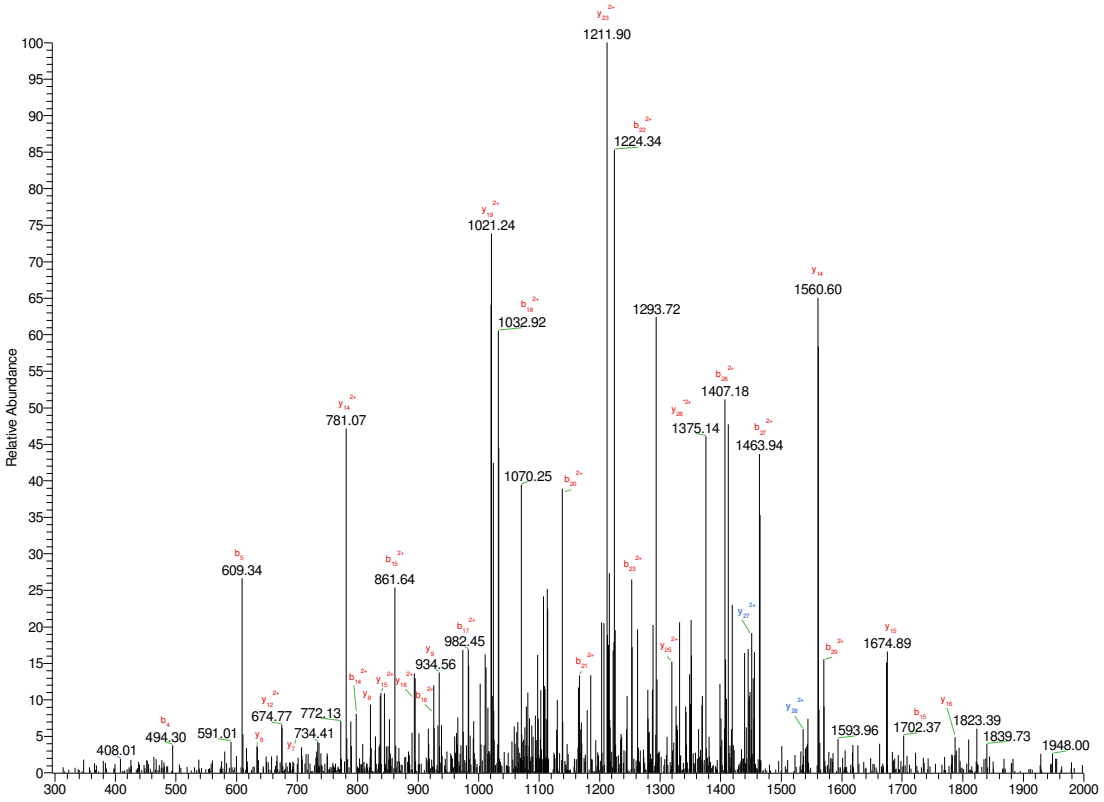
5.10 TABLES

Table 5.1: X-ray diffraction data processing and refinement statistics

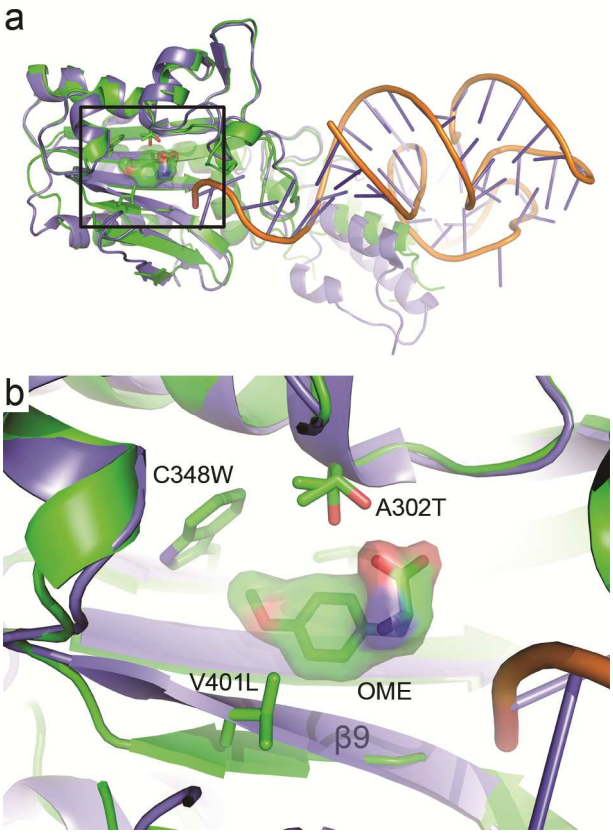
Protein	OmeRS
Ligands	Ome, AMP-PNP, Mg^{2+}
PDB ID	3QTC
Data Collection and Processing	
Space group	$P6_4$
Resolution (Å)	1.75
Cell dimensions	
a, b, c (Å)	104.81, 104.81, 71.67
α, β, γ (°)	90, 90, 120
Molecules in asymmetric unit	1
No. measured reflections[a]	458839 (45558)
No. unique reflections[a]	44965 (7161)
Redundancy	10.2 (6.4)
R_{merge} (%)[a,b]	4.0 (36.8)
Completeness (%)[a]	99.7 (98.9)
I/σ(I)[a]	35.1 (4.3)
Refinement	
Resolution range (Å)	50.0-1.75
No. reflections:	
Working set	42693
Test set	2268
R_{work}/R_{free}[c]	0.1834/0.2124
No. atoms:	
Protein	2157
Ligands	306
Water	235
Mean Overall B-factor ($Å^2$)	29.93
rmsd bond lengths (Å)[d]	0.01
rmsd bond angles (°)	1.30
Refinement program	Refmac

a values in parentheses represent data from thehighest resolution shell
b $R_{merge} = \Sigma_{hkl} \Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/\Sigma_{hkl} \Sigma_i I_i(hkl)$
c $R_{factor} = \Sigma||F_{obs}|-|F_{calc}||/\Sigma|F_{obs}|$. R_{work} is the R_{factor} calculated using all diffraction data included in the refinement. R_{free} is the R_{factor} calculated using the randomly chosen 5% of diffraction data that were not included in the refinement.
d rmsd = root mean square deviation

5.10 SUPPLEMENTARY FIGURES



Supplementary Figure 5.1: Tandem mass spectrum of peptide $+H_3N$-HNIEDGSVQLADH**X**QQ NTPIGDGPVLLPDNHYLSTQSALSK-CO_2H. **X** represents the UAG encoded site at position 182 of GFP. The sequence of the peptide containing Ome at position **X** was unambiguously assigned from the annotated b and y ion series.



Supplementary Figure 5.2: Superposition of *M. mazei* OmeRS complexed with Ome and AMP- PNP to *D. hafniense* PylRS complexed with tRNA (PDB ID 2ZNI). **a**) global view of the PylRS crystal structure (tRNA colored orange and PylRS colored purple) superposed to the OmeRS crystal structure (colored green). **b**) close-up view of the active sites for both synthetases. The Ome substrate and all four mutated residues are displayed for the OmeRS structure.

5.10 REFERENCES

1. Wang, L.; Schultz, P. G., Expanding the genetic code. *Angew Chem Int Ed Engl* **2004,** *44* (1), 34-66.

2. Wang, Q.; Parrish, A. R.; Wang, L., Expanding the genetic code for biological studies. *Chem Biol* **2009,** *16* (3), 323-36.

3. Liu, C. C.; Schultz, P. G., Adding new chemistries to the genetic code. *Annu Rev Biochem* **2010,** *79*, 413-44.

4. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G., Expanding the genetic code of Escherichia coli. *Science* **2001,** *292* (5516), 498-500.

5. Wang, L.; Schultz, P. G., A general approach for the generation of orthogonal tRNAs. *Chem Biol* **2001,** *8* (9), 883-90.

6. 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), 1459-62.

7. 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), 6695-700.

8. Namy, O.; Zhou, Y.; Gundllapalli, S.; Polycarpo, C. R.; Denise, A.; Rousset, J. P.; Soll, D.; Ambrogelly, A., Adding pyrrolysine to the Escherichia coli genetic code. *FEBS Lett* **2007,** *581* (27), 5282-8.

9. 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. *Biochem Biophys Res Commun* **2008,** *371* (4), 818-22.

10. Chen, P. R.; Groff, D.; Guo, J.; Ou, W.; Cellitti, S.; Geierstanger, B. H.; Schultz, P. G., A facile system for encoding unnatural amino acids in mammalian cells. *Angew Chem Int Ed Engl* **2009,** *48* (22), 4052-5.

11. 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), 1156-64.

12. Wang, Y. S.; Wu, B.; Wang, Z.; Huang, Y.; Wan, W.; Russell, W. K.; Pai, P. J.; Moe, Y. N.; Russell, D. H.; Liu, W. R., A genetically encoded photocaged Nepsilon-methyl-L-lysine. *Mol Biosyst* **2010,** *6* (9), 1557-60.

13. Buddha, M. R.; Crane, B. R., Structure and activity of an aminoacyl-tRNA synthetase that charges tRNA with nitro-tryptophan. *Nat Struct Mol Biol* **2005,** *12* (3), 274-5.

14. Kobayashi, T.; Sakamoto, K.; Takimura, T.; Sekine, R.; Kelly, V. P.; Kamata, K.; Nishimura, S.; Yokoyama, S., Structural basis of nonnatural amino acid recognition by an engineered aminoacyl-tRNA synthetase for genetic code expansion. *Proc Natl Acad Sci U S A* **2005,** *102* (5), 1366-71.

15. Turner, J. M.; Graziano, J.; Spraggon, G.; Schultz, P. G., Structural plasticity of an aminoacyl-tRNA synthetase active site. *Proc Natl Acad Sci U S A* **2006,** *103* (17), 6483-8.

16. Wang, L.; Zhang, Z.; Brock, A.; Schultz, P. G., Addition of the keto functional group to the genetic code of Escherichia coli. *Proc Natl Acad Sci U S A* **2003,** *100* (1), 56-61.

17. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G., An expanded eukaryotic genetic code. *Science* **2003,** *301* (5635), 964-7.

18. Wang, Q.; Wang, L., New methods enabling efficient incorporation of unnatural amino acids in yeast. *J Am Chem Soc* **2008,** *130* (19), 6066-7.

19. Wang, W.; Takimoto, J. K.; Louie, G. V.; Baiga, T. J.; Noel, J. P.; Lee, K. F.; Slesinger, P. A.; Wang, L., Genetically encoding unnatural amino acids for cellular and neuronal studies. *Nat Neurosci* **2007,** *10* (8), 1063-72.

20. Liu, W.; Brock, A.; Chen, S.; Schultz, P. G., Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nat Methods* **2007,** *4* (3), 239-44.

21. Blight, S. K.; Larue, R. C.; Mahapatra, A.; Longstaff, D. G.; Chang, E.; Zhao, G.; Kang, P. T.; Green-Church, K. B.; Chan, M. K.; Krzycki, J. A., Direct charging of tRNA(CUA) with pyrrolysine in vitro and in vivo. *Nature* **2004,** *431* (7006), 333-5.

22. Ambrogelly, A.; Gundllapalli, S.; Herring, S.; Polycarpo, C.; Frauer, C.; Soll, D., Pyrrolysine is not hardwired for cotranslational insertion at UAG codons. *Proc Natl Acad Sci U S A* **2007,** *104* (9), 3141-6.

23. Young, T. S.; Ahmad, I.; Yin, J. A.; Schultz, P. G., An enhanced system for unnatural amino acid mutagenesis in E. coli. *J Mol Biol* **2010,** *395* (2), 361-74.

24. Wang, Y. S.; Russell, W. K.; Wang, Z.; Wan, W.; Dodd, L. E.; Pai, P. J.; Russell, D. H.; Liu, W. R., The de novo engineering of pyrrolysyl-tRNA synthetase for genetic incorporation of L-phenylalanine and its derivatives. *Mol Biosyst* **2011,** *7* (3), 714-7.

25. Kavran, J. M.; Gundllapalli, S.; O'Donoghue, P.; Englert, M.; Soll, D.; Steitz, T. A., Structure of pyrrolysyl-tRNA synthetase, an archaeal enzyme for genetic code innovation. *Proc Natl Acad Sci U S A* **2007,** *104* (27), 11268-73.

26. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S., Multistep engineering of pyrrolysyl-tRNA synthetase to genetically encode N(epsilon)-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification. *Chem Biol* **2008,** *15* (11), 1187-97.

27. Yanagisawa, T.; Ishii, R.; Fukunaga, R.; Kobayashi, T.; Sakamoto, K.; Yokoyama, S., Crystallographic studies on multiple conformational states of active-site loops in pyrrolysyl-tRNA synthetase. *J Mol Biol* **2008,** *378* (3), 634-52.

28. Summerer, D.; Chen, S.; Wu, N.; Deiters, A.; Chin, J. W.; Schultz, P. G., A genetically encoded fluorescent amino acid. *Proc Natl Acad Sci U S A* **2006,** *103* (26), 9785-9.

29. Ibba, M.; Soll, D., Aminoacyl-tRNA synthesis. *Annu Rev Biochem* **2000,** *69*, 617-50.

30. Perona, J. J.; Hou, Y. M., Indirect readout of tRNA for aminoacylation. *Biochemistry* **2007,** *46* (37), 10419-32.

31. Nozawa, K.; O'Donoghue, P.; Gundllapalli, S.; Araiso, Y.; Ishitani, R.; Umehara, T.; Soll, D.; Nureki, O., Pyrrolysyl-tRNA synthetase-tRNA(Pyl) structure reveals the molecular basis of orthogonality. *Nature* **2009,** *457* (7233), 1163-7.

32. Rodriguez-Hernandez, A.; Bhaskaran, H.; Hadd, A.; Perona, J. J., Synthesis of Glu-tRNA(Gln) by engineered and natural aminoacyl-tRNA synthetases. *Biochemistry* **2010,** *49* (31), 6727-36.

33. Takimoto, J. K.; Adams, K. L.; Xiang, Z.; Wang, L., Improving orthogonal tRNA-synthetase recognition for efficient unnatural amino acid incorporation and application in mammalian cells. *Mol Biosyst* **2009,** *5* (9), 931-4.

34. Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R., The Orbitrap: a new mass spectrometer. *J Mass Spectrom* **2005,** *40* (4), 430-43.

35. Kabsch, W., Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* **1993,** *26*, 6.

36. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* **1994,** *50* (Pt 5), 760-3.

37. Brunger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J. S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L., Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* **1998,** *54* (Pt 5), 905-21.

38. Laskowski, R. A.; MacArthur, M. W.; Moss, D. S.; Thornton, J. M., PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography* **1993,** *26* (2), 283-291.