EXPANDING PROTEIN SEQUENCE SPACE THROUGH INCORPORATION OF NON-CANONICAL AMINO ACIDS

Thesis by

Frank Truong

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2013

(Defended January 28th 2013)

This work is dedicated to
the family I was born into,
the family that I married into,
and the family that I have started.

ACKNOWLEDGEMENTS

I recall that, sometime around the third grade, I wanted to be a scientist when I grew up. I think it was at that point that I learned what a Nobel prize was and how amazing it would be to discover something worthy of that accolade. As a naïve young boy, I thought the job would be easy, financially rewarding, and glamorous all at the same time. Now that I am older, I realized that I was mostly wrong about all three, but I am glad that I have stuck with science. I would be remiss in saying that I accomplished all of the things I have done without the help of countless people.

I am fortunate that Professor David A. Tirrell took a chance on me when I asked many years ago to join his research group. He has allowed me great flexibility in my research endeavors; sometimes resulting in success but often ending failure. Throughout that time, he has provided me with immeasurable moral, scientific, and financial support, for which I am very grateful. I hope that one day I can pass along all of the memorable quotes and parables he has taught me throughout the years.

I would also like to thank Professors Frances Arnold, Stephen Mayo, and Rustem Ismagilov for serving on my thesis committee. I greatly appreciate all the time they have taken to provide me with valuable input and advice. Thanks are also due to Professor Michael Vicic for his support and guidance before, during, and after ChE130. I am thankful for the great people at Caltech, especially Anne Hormann, Kathy Bubash, Joe Drew, Steve Gould, and Professor Ali Hajimiri.

My gratitude goes to Professor Peter G. Vekilov, my undergraduate advisor. I would not be at Caltech without all of his support while at the University of Houston and afterwards. Thanks are also due to Dr. Luis Filobelo, Dr. Yasser Qutub, and the rest of the former Vekilov lab from the University of Houston, who were once my mentors but are now life-long friends.

Great leaders attract great talent and that is very true of the Tirrell lab. Members past and present are a collection of great scientists and people. I am thankful to have been mentored by my friends Tae Hyeon Yoo, John Ngo, and Janek Szychowski; they took the time to teach me molecular biology and synthetic chemistry. I am confident the friendships I have forged with Julie Champion, Jim Van Deventer, Caglar Tanrikulu, Eileen Fong, Beverly Lu, Alborz Mahdavi, Mark van Eldjik, Larry Dooling, JD Bagert, Kai Yuet, Brett Babin, Kat Fang, Maren Buck, Phoebe Tzou, Nicholas Ball, Wenbin Zhang, and Cole DeForest will last a lifetime. I would also like to thank my collaborator and friend Professor Adam Urbach for advice and support on research and life. Thank you to Karthik Narsimhan, Thomas Lampo, Wesley Yu, and Leigh Anna Logsdon for giving me a chance to mentor some of the finest young minds.

I am thankful for my family. My brothers, Anton and Harry, for their constant support and for always being ready to cheer me up with a dose of laughter. I am indebted to my mother Lien and stepfather Ty, who had to work long and hard hours so my brothers and I could have a chance not to live the life they had. Any success I have had in my education was made possible by them. My mother, who never got past elementary school, will always be

one of the smartest and bravest women I know; she is responsible for my humility and diligence.

I would like to thank my extended family. I am lucky to have a father-in-law and mother-in-law who treat me like their own son. Meaghan's aunt, uncle, grandmother, and cousins in San Diego have made us feel at home ever since we first moved out to California.

I would like to thank the family I have started. Meaghan was brave enough to take a chance on me ten years ago and I hope I have enough years left in my life to repay her for it. She shouldered the burden of marriage to a graduate student with dignity and grace; I share this thesis with her. Lastly, I would like to express my awe for my daughter Morgan Bailey. She has astounded me with something new every day for the past four months. She may never remember her time in Pasadena, but I hope to share with her all the things I have learned from the great people I have met at Caltech.

ABSTRACT

Protein sequence space has been augmented by researchers wanting to expand the diversity of chemical functionalities that can exist within proteins. Artificial amino acids can range from simple atom substitutions such as fluorination, to installation of reactive handles like azides and alkynes. Researchers build upon the framework of natural proteins and have developed methods of installing a wide variety of artificial amino acids into proteins. Chapter 1 discusses in detail the two methods for metabolic incorporation of unnatural amino acids: site-specific incorporation and residue-specific incorporation. Advantages and disadvantages to each method are detailed as well as applications of these methods to the examination of problems in chemical biology.

Non-canonical amino acids analogues that are structurally similar to their canonical counterparts can be recognized by the endogenous translational machinery for residue-specific incorporation. Chapter 2 describes the directed evolution of the methionyl-tRNA synthetase (MetRS) to incorporate propargylglycine, an alkyne analogue that is not recognized by the wild-type MetRS. A new MetRS variant active towards propargylglycine was identified after screening libraries of both active site mutations and error-prone PCR mutations. PraRS is capable of producing proteins where methionine is quantitatively replaced by propargylglycine. PraRS also does not recognize azidonorleucine, an azide methionine analogue for which the NLL-MetRS was evolved in order to enable cell-specific protein labeling. A method to identify cellular origins of

proteins from two different bacterial strains in co-culture was developed using the NLL-MetRS and PraRS.

Chapter 3 illustrates the effects of global incorporation of non-canonical amino acids into globular proteins. Although trifluoroleucine and homoisoleucine have shown to increase the thermostability of model proteins, incorporation into more chloramphenicol acetyltransferase (CAT) does not yield the same benefits. We find that mutations that stabilize CAT for fluorinated amino acid incorporation do not protect against homoisoleucine incorporation.

Lastly, access to new chemical reactions for protein modification requires synthesis and incorporation of new non-canonical amino acid analogues. Chapter 4 describes the design of two new artificial amino acids, S-allyl-homocysteine and 3-furanylalanine, for residue specific incorporation without expression of mutant synthetases. Also, a third amino acid, azidomethylphenylalanine, was designed for activation by a previously discovered phenylalanine-tRNA synthetase mutant. Incorporation of these three analogues provides chemical handles that are potential reagents for cross metathesis, Diels-Alder cycloaddition, and generation of a molecular epitope for binding to synthetic receptors.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

Metabolic Incorporation of Non-Canonical Amino Acids into Proteins

Introduction

Nature has utilized the twenty canonical amino acids and evolution to create a wide variety of proteins that are responsible for structure, signaling, and catalysis in all living organisms. The proteins that serve all of these functions are a small sampling of the possible protein sequence space imaginable. The need to explore all of sequence space is not a priority for nature; evolution is driven by how a protein contributes to the fitness of an organism. By removing protein evolution from a fitness-oriented boundary condition, researchers have utilized directed evolution to push proteins beyond what is possible in nature[1,2].

Natural protein sequence space is primarily composed of twenty amino acids with functional groups including amines, alcohols, carboxylic acids, and aromatic groups. Although the combination of these amino acids is capable of complex catalysis, it pales in comparison to that possible with the functional groups that are accessible to synthetic chemists. In order to introduce new chemical groups into proteins, many methods have been developed to modify canonical amino acids or append synthetic peptides to natural proteins[3-6]. However, to truly expand protein sequence space, artificial amino acids should be utilized in the production of new proteins.

The process of synthesizing a protein from its genetic blueprints is an extremely high fidelity process[7], with error rates in DNA replication greater than 10^8 (Figure 1.2). Researchers have taken advantage of the aminoacylation step to incorporate unnatural amino acids into proteins. Enzymes known as aminoacyl-tRNA synthetases (aaRS) are responsible for attaching the correct amino acids to their cognate transfer RNAs (tRNAs);

aminoacylated tRNAs are subsequently used during translation to decode mRNA into a polypeptide chain. Since steps downstream of aminoacylation are permissive to changes in amino acids, attachment of non-canonical amino acids to tRNAs is the limiting step in expanding protein sequence space. We will focus on two methods that involve the aminoacylation of non-canonical amino acids, colloquially known as site-specific incorporation and residue specific incorporation.

Site-Specific Incorporation of Non-Canonical Amino Acids

Incorporation via site-specific incorporation affords researchers the ability to install non-canonical amino acids at a preprogrammed location within a protein. In order to ensure correct placement of the non-canonical amino acid, the aaRS and tRNA used must be orthogonal to the expression host's translational machinery. The aaRS/tRNA pairs used commonly come from archebacteria, which reduces the likelihood of charging of E. coli, S. cerevisiae, and other higher organisms' tRNAs[8]. Directed evolution is performed on the aaRS to enable charging of a wide variety of non-canonical amino acids to amber suppressor tRNA[9]. Aminoacylated amber suppressor tRNA allows for "read-through" of the amber codon; full length proteins are synthesized with an artificial amino acid in place of genetically incorporated amber stop codon.

The ability to genetically encode location of the artificial amino acid without disturbing other canonical amino acids enables the protein chemist to examine numerous biological questions. Incorporation of *p*-azido-phenylalanine (Figure 1.1 [**1**]) or *p*-benzoyl-phenylalanine (Figure 1.1 [**2**]) site-specifically has been utilized for studying protein

interactions; both amino acids are capable of photo-activated crosslinking[10-14]. Through the use of bio-compatible, bio-orthogonal reactions, site specific attachment of molecules (such as sugars or PEG) to proteins can be done to improve their utility as therapeutics[15-17]. The application of site-specific incorporation has extended beyond *E.coli*, to now include yeast[18,19], mammalian cell lines[20-23], and model organisms *C. elegans*[24] and *D. melanogaster*[25]. Greater detail can be found in recent reviews about applications of site-specific incorporation of artificial amino acids[26-29].

Amber suppression does have some drawbacks, some of which have been addressed to increase the versatility of the method. The primary issue is that proteins are modified to include a stop codon at the location where the artificial amino acid is desired. Aminoacylated amber suppressor tRNA has to compete with release factor-1 for binding at the ribosomal A-site. If the amber suppressor tRNA binds first, the full length protein is synthesized with an artificial amino acid at the desired position. However, should release factor 1 bind first, a truncated polypeptide chain is the translational product[30,31]. This truncated protein can be responsible for over 80% of the total protein product[32], significantly reducing protein yields. Furthermore, incorporation of multiple amino acids using amber suppressor compounds the release factor 1 issue. Engineering of the ribosome and knockouts of release factor 1 have helped to improve protein production with amber suppression[32,33].

The amber stop codon is one of three stop codons used for translational termination by all organisms; in *E. coli*, the amber stop codon is responsible for the termination of 314 genes.

The amber suppressor tRNA cannot discriminate between naturally and genetically modified stop codons; a certain percentage of natural proteins will have unnatural amino acids appended to the ends of their sequences. The presence of these additional artificial amino acids can complicate study of biological processes; visualization of a desired protein through attachment of fluorescent probes will have background due to proteins that terminate with amber stop codons. Engineering of *E. coli* by the Church Lab aims to solve this problem by removing all natural amber stop codons[34], freeing the amber codon for use as a new sense codon. However, this work is still ongoing and extension of this method to other organisms with larger genomes will be a complex undertaking.

Residue-Specific Incorporation of Non-Canonical Amino Acids

Residue-specific incorporation is a global replacement strategy; cognate tRNAs for a specific canonical amino acid are instead aminoacylated by aaRS with a non-canonical amino acid analogue. Codons for the canonical amino acid across the transcriptome are decoded as the non-canonical amino acid. The replacement is done in a statistical manner; decoding of mRNA is based on the levels of tRNAs aminoacylated with natural and unnatural amino acids. By depleting a canonical amino acid, replacement levels can reach quantitative levels; without depletion, replacement levels are based on the activity of the aaRS to the non-canonical amino acid analogue.

Incorporation of many residue-specific amino acid analogues can be performed without any genetic modification of the host organism[35-37]; the endogenous aaRS have enough promiscuity activity towards the artificial analogues. Overexpression of aaRS has

facilitated the incorporation of a greater subset of analogues[38], but engineering of aaRSs has provided researchers with even greater options for non-canonical amino acids in protein synthesis. Tang et al. demonstrated that a single mutation made to the editing site of leucyl-tRNA synthetase enables the incorporation of many methionine analogues at leucine codons[39]. Incorporation of p-azidophenylalanine by a mutant phenylalanine-tRNA synthetase allows for the formation of photo-crosslinked protein hydrogels[40]. Fluorescence-Activated Cell Sorting (FACS) screening of methionyl-tRNA synthetase libraries has enabled the discovery of variants capable of aminoacylating azidonorleucine[41,42](Figure 1.1 [**3**]) and trifluoronorleucine[43] (Figure 1.1 [**4**]).

In contrast to site-specific methods, the residue-specific approach cannot guarantee the exact location for an artificial amino acid, only that it will appear at the codons corresponding to the canonical amino acid replaced. Protein engineering is required to change the protein sequence to remove undesired codons. However, without the need to genetically modify the host organism, the residue-specific incorporation of non-canonical amino acids has been applied to the study of proteome dynamics. By adding azidohomoalanine (Figure 1.1 [**5**]) at the same time a stimulus is given to an organism, newly synthesized proteins will be metabolically labeled with azides[44]. Utilizing the bio-orthogonal Husigen [3+2] azide-alkyne cycloaddition, the proteins made in response to the stimulus can be selectively modified with affinity tags or fluorescent probes. Bio-Orthogonal Non-Canonical Amino Acid Tagging (BONCAT) has been used to visualize new protein synthesis[45-47]; determine rates of new protein synthesis[48-51]; and identify

proteins involved in histone turnover, neuronal function, and receptor-initiated protein synthesis[52-55].

Synthesis of proteins with multiple artificial amino acids is straightforward with residue-specific incorporation. Although control over location of those positions is not possible with natural protein sequences, recombinant DNA technology allows for the modification of natural sequences or the generation of artificial DNA sequences to tailor proteins with desired artificial amino acid locations. Additionally, multiple artificial amino acids can be incorporated if they correspond to different canonical amino acid analogues[56,57]. However, global replacement of a natural amino acid with an unnatural analogue in proteins can negatively impact protein function and structure[58].

In the next few sections, we will explore three topics of non-canonical amino acid incorporation in greater detail: engineering new substrate recognition in aminoacyl-tRNA synthetases, cell-specific BONCAT, and proteins of novel amino acid compositions.

Engineering New Synthetase Activity

For both site-specific and residue-specific methods, the roadblock in the incorporation of new amino acids is the generation of aaRS activity toward a new substrate. For artificial amino acids that are structural similar to the target aaRS, such as homopropargylglycine and methionyl-tRNA synthetase[59], incorporation can be performed without any modification of the aminoacyl-tRNA synthetase. To expand chemical functionality

available to protein sequence space, novel amino acids can deviate far from their natural analogues, requiring engineering of aaRS.

Evolution of aminoacyl-tRNA synthetases for site-specific incorporation of artificial amino acids takes advantage of release factor 1 for selection purposes (Figure 1.3a). To enrich aaRS variants capable of charging a desired artificial amino acid to amber suppressor tRNA, a protein that is vital for organism fitness (ex. an antibiotic resistance gene) is modified with an internal amber codon. Since the artificial amino acid is incorporated only once, there is less concern that the protein will be destabilized. Without competition by the aminoacylated amber suppressor tRNA, release factor 1 causes the vital protein to be truncated and cells harboring that poorly active aaRS variant are selected against. However, the possibility still exists where a canonical amino acid is charged to the amber suppressor tRNA. To select against that scenario, the enriched library is transferred to a strain where a protein that is harmful to the organism (ex. a DNA or RNA nuclease) now contains the internal amber codon. When this library is expressed in this strain without the addition of the artificial amino acid, variants that charge natural amino acids are selected again. Repeated rounds of positive and negative selection allows large libraries to be distilled to variants that are specific to the desired artificial amino acid and orthogonal to the endogenous translational machinery.

Since incorporation of non-canonical amino acids in a residue-specific manner can occur at multiple locations, different high throughput screening methods were developed to circumvent the possibility of reporter gene destabilization. The incorporation of amino

acids with bio-orthogonal reactions provides a means to monitor aaRS activity. Link et al. devised a means to increase cell surface presentation of azides through overexpression of an outer membrane protein[60] (Figure 1.3b). Attachment of fluorescent probes to the cell surface through copper-catalyzed, azide-alkyne click chemistry or strain-promoted, copper-free cyclooctyne probes created a means to correlate aaRS activity to cell-surface fluorescence. Active site libraries of the methionyl-tRNA synthetase (MetRS) could be screened using FACS to acquire a variant highly active for azidonorleucine.

Screening for aaRS activity for amino acids without a bio-orthogonal reaction is not as straightforward. Yoo et al. developed a fluorescent protein reporter for MetRS activity based on green fluorescent protein (GFP)[43]. Methionine positions in GFP that were crucial to protein folding were relocated to unstructured regions of the protein, creating a new fluorescent protein GFPrmAM. Incorporation of non-canonical amino acid analogues into GFPrmAM does not knock out fluorescence, unlike the case for the wild-type GFP. Using methionine auxotrophs and media depletion of methionine, metRS activity towards non-canonical amino acid analogues controls the rate of new GFPrmAM synthesis. Since fluorescence is linked to synthetase activity, libraries of MetRS variants can now be screened, allowing for the discovery of a mutant metRS for trifluoronorleucine. However, to screen other synthetases in this manner would involve the generation of multiple fluorescent reporters similar to GFPrmAM.

Mutant aaRSs Enable BONCAT on Subpopulations of Complex Cellular Mixtures

The benefit of using azidohomoalanine or homopropargylglycine (Figure 1.1 [**6**]) for BONCAT is that the endogenous translational machinery for the organism, whether E. coli, yeast, worms, or mammalian cells lines, does not need to be perturbed. In a non-laboratory setting, organisms are rarely isolated from one another, so the need for chemical tools to identify and enrich proteins out of a single species is highly desired. Azidonorleucine can be added to complex mixtures of cells and the proteomes will remain unlabeled because azidonorleucine is not recognized by the endogenous MetRS. By constraining the expression of the NLL-MetRS[61] to a specific cell strain, changes in a particular proteome could be explored with BONCAT without interference from other proteins from other strains. Ngo et al. demonstrated that fluorescent probes could be attached specifically to the proteome of an *E. coli* strain containing the NLL-MetRS co-cultured with human macrophages[62] (Figure 1.4); this is remarkable considering the proteome of the macrophages is significantly larger than that of *E. coli*[63].

The ability to confine protein labeling with azidonorleucine in a spatial manner can also apply to organisms of the same species. Although a single strain of a bacteria has the same genotype, environmental conditions can force phenotypic changes which are represented at the proteomic level. By placing the NLL-MetRS under the control of promoters associated with a certain phenotype, the proteome of a particular cellular state can be metabolically labeled with azidonorleucine[64].

Being able to control residue-specific incorporation in a spatial and temporal manner, BONCAT is being used to identify proteomic responses to a variety of stimuli. The Tirrell Lab is currently applying the use of the NLL-MetRS and mutant PheRS to the study of bacterial secretion systems, biofilms, quorum sensing, and examining protein expression in specific cells in multicellular organisms such as *C. elegans*.

Proteins of Novel Compositions

The fluorinated region of artificial protein sequence space has been explored extensively in the hope that it will provide protein engineers the same benefits it has provided polymer chemists, especially increased hydrophobicity, thermostability, and chemical denaturant stability. The replacement of hydrophobic amino acids like leucine, valine, and isoleucine with their fluorinated non-canonical amino acid counterparts has provided mixed results.

Proteins designed by nature with well-defined interactions of leucine residues have showed remarkable increases in protein stability after replacement with fluorinated leucine analogues. Leucine zippers are alpha helical proteins with leucine residues at regular intervals, making one face of the alpha helix rich with leucine side chains. A single protein is not very stable, but the leucine zippers improve their stability by forming bundles that sequester the hydrophobic side chains. Residue-specific incorporation of trifluoroleucine (Figure 1.1 [**7**]) into various leucine zippers results in significant increase in thermostability, raising melting temperatures by $13^{o}C$[65,66] (Figure 1.5). The use of hexafluoroleucine (Figure 1.1 [**8**]) in place of leucine also shows significant enhancement of leucine zipper thermostability, with upwards of $22^{o}C$ increase in melting

temperatures[67,68]. Bilgicer et al. have shown that the leucine zippers made fluorinated amino acids preferentially segregate from leucine zippers made with leucine, showing fluorinated surfaces could be potentially used to design specific protein interactions[69,70].

Most globular proteins do not have hydrophobic residues sequestered in the same fashion as coiled-coiled proteins; the result of fluorinated amino acid incorporation depends greatly on the stability of the protein. The Budisa Lab has had success in incorporating multiple monofluroinated analogues of proline, phenylalanine, and tryptophan without knocking out lipase function[57]. Their attempts to globally incorporation trifluoroleucine in Annexin V, GFP, and Barstar (all greater than 10 kDa) resulted in significantly less than quantitative incorporation; proteins with high incorporation levels of trifluoroleucine were not observed potentially due to protein misfolding/aggregation[71]. Wang et al. showed quantitative replacement of the five isoleucine residues of murine interleukin2 with trifluoroisoleucine (Figure 1.1 [**9**]) did not affect the protein's native function[72].

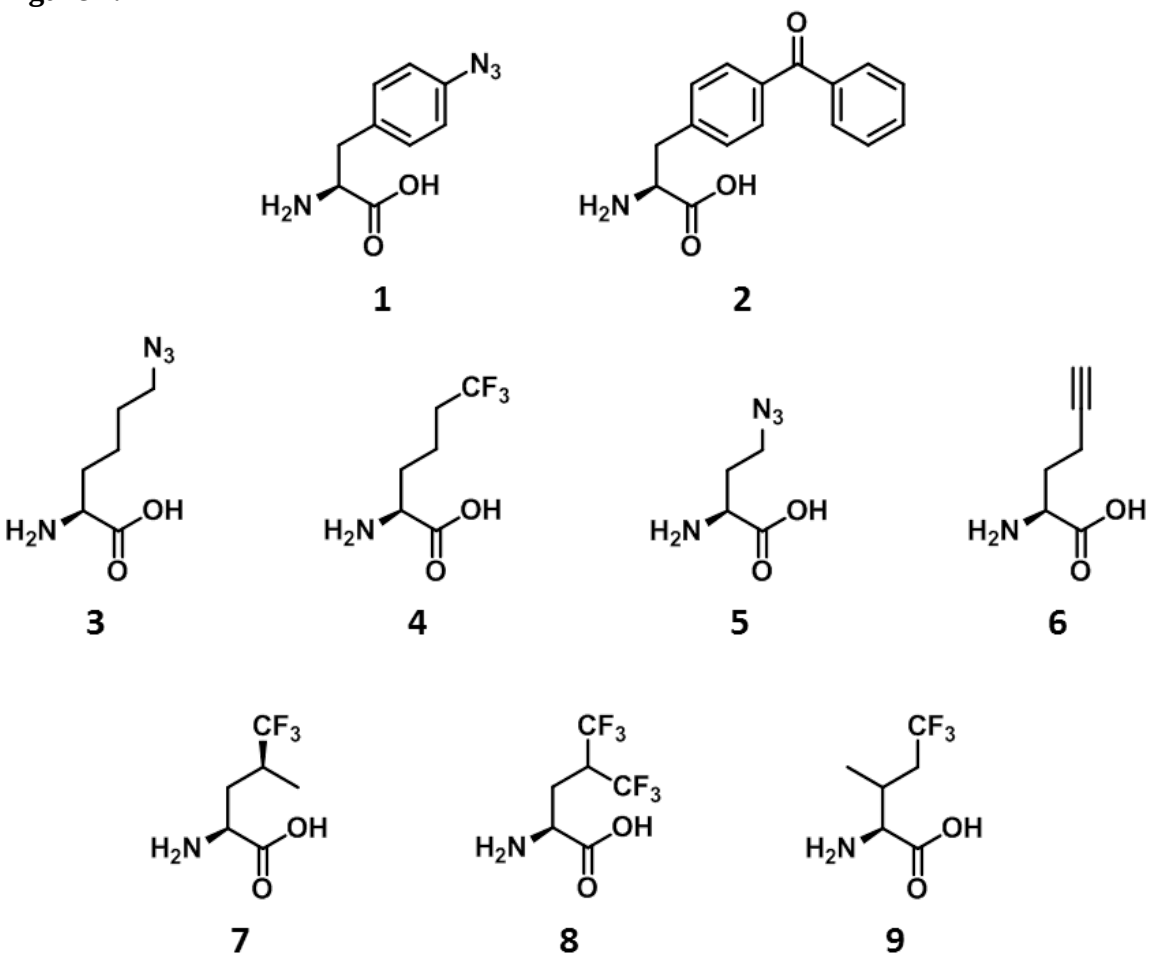
Directed evolution has been harnessed to restore protein function and stability after fluorinated amino acid incorporation. Wild-type GFP fluorescence is completely knocked out after incorporation of trifluoroleucine, requiring multiple rounds of evolution to recover lost fluorescent[73]. Chloramphenicol acetyltransferase is still active at room temperature when expressed with trifluoroleucine, but over 20 fold less thermostable at elevated temperatures[74]. Two rounds of directed evolution are required to restore thermostability back to wild-type, leucine levels. Considering the fact that most amino acid substitutions in natural protein sequence space are deleterious, it is not surprising that global incorporation

of an artificial amino acid can be significantly destabilizing. The goal of the Tirrell Lab is to continue exploring non-canonical protein sequence space in conjunction with directed evolution to create proteins that prefer artificial amino acids to their natural counterparts.

Topics Described in This Thesis

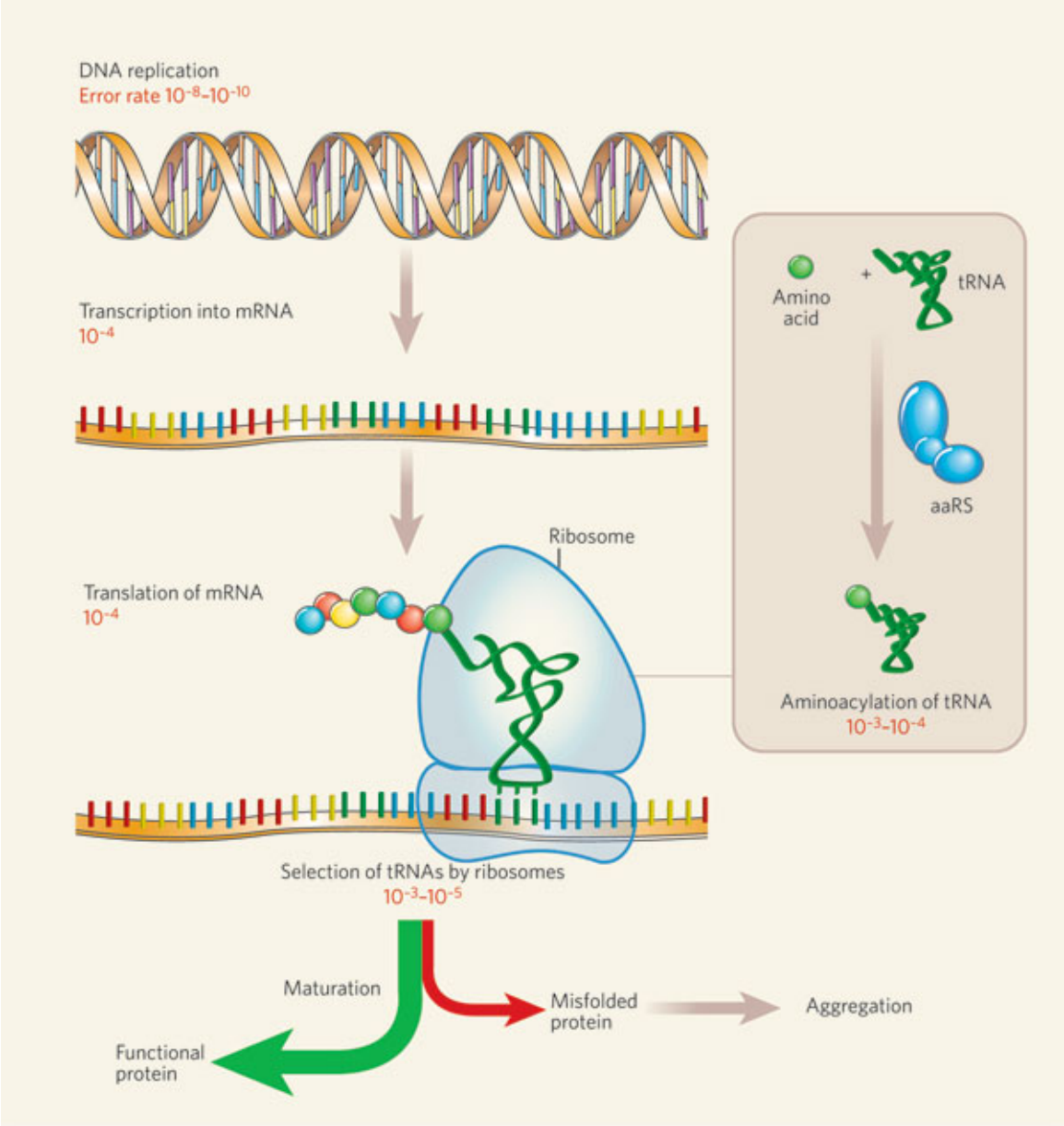
Using synthetic chemistry and engineered aminoacyl-tRNA synthetases, we described the expansion of artificial protein sequence space to include new amino acids. These amino acids show potential for introducing new chemistries to researchers using residue-specific incorporation, including cross metathesis, Diels-Alder cycloadditions, and host-guest chemistry. The directed evolution of a mutant synthetase for cell-specific incorporation of alkyne analogues of methionine lays the groundwork for examining new problems with BONCAT. In combination with the NLL-MetRS, the ability to monitor protein expression in two different bacteria strains simultaneously is possible without cross-labeling. Lastly, we examine the effects of quantitative replacement of leucine with homoisoleucine on chloramphenicol acetyltransferase.

Figure 1.1



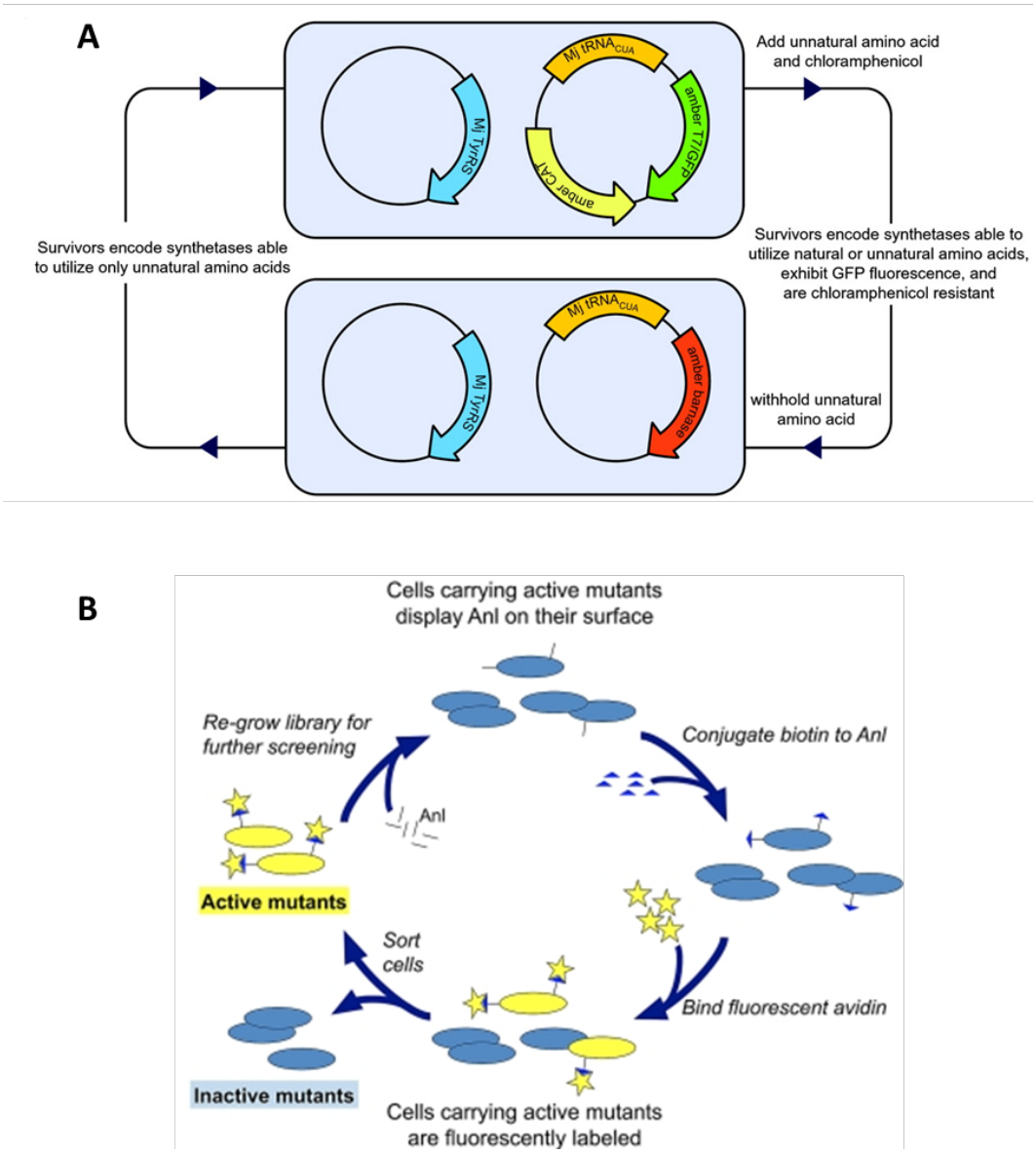
Structures of amino acids described in this chapter: 1: *p*-azidophenylalanine; **2**: *p*-benzoyl-phenylalanine; **3**: azidonorleucine; **4**: trifluoronorleucine; **5**: azidohomoalanine; **6**: homopropargylglycine; **7**: trifluoroleucine; **8**: hexafluoroleucine; **9**: 5,5,5-trifluoroisoleucine

Figure 1.2



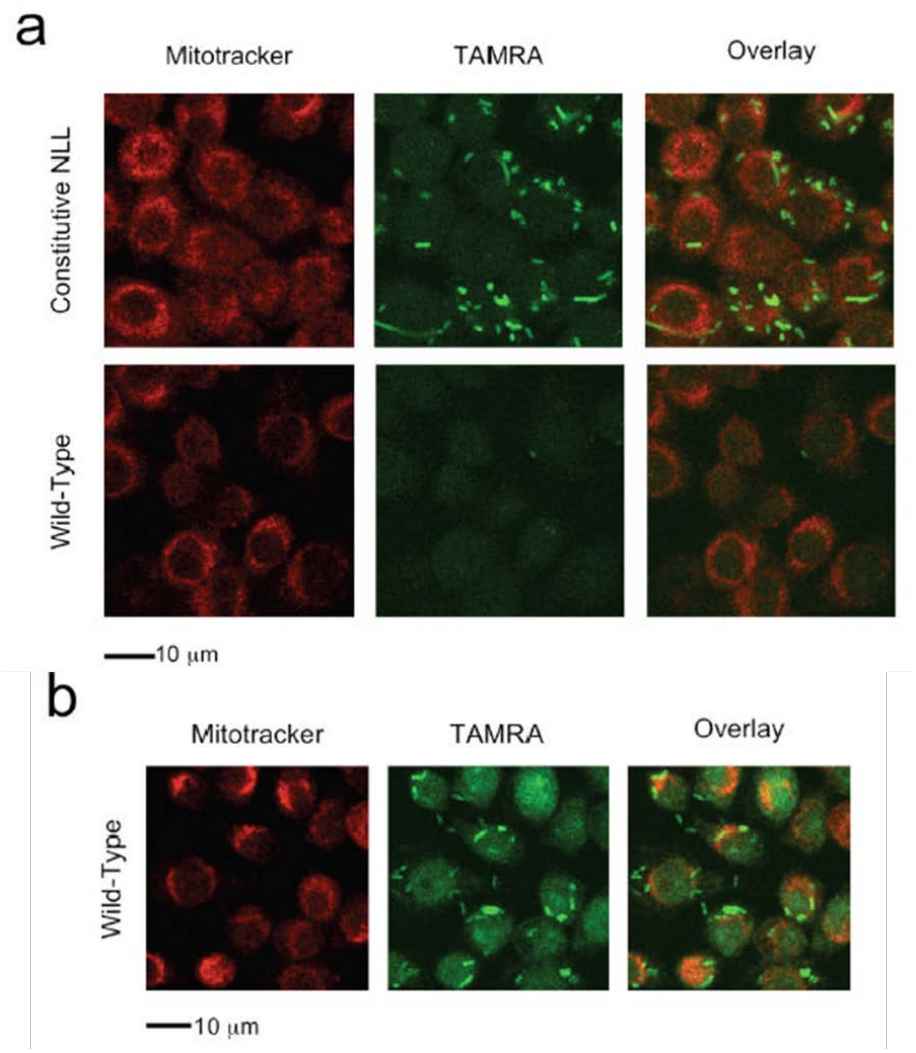
Error rates in steps involved during protein synthesis. Aminoacyl-tRNA synthetases participate in decoding of the mRNA during translation through the attachment of amino acids to their cognate tRNAs. Aminoacylation is the lowest fidelity step; incorporation of non-canonical amino acids takes advantage of promiscuous, endogenous aaRSs or use of engineered aaRSs.

Figure 1.3



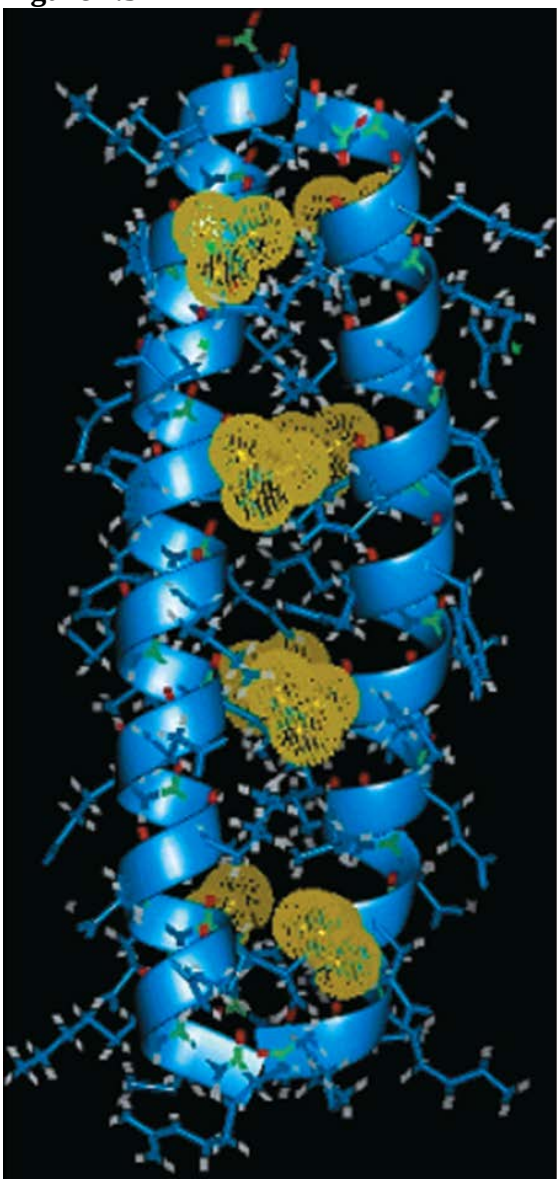
Selection/Screening of active aaRS variants toward artificial amino acids. A) Selection scheme to enrich aaRS for site-specific incorporation of artificial amino acids. B) High-throughput screening of aaRS for residue-specific incorporation of azide bearing non-canonical amino acids Adapted from from (Chin, J. W., Martin, A. B., King, D. S., Wang, L., Schultz, P. G. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020; Copyright (2002) National Academy of Sciences, U.S.A) and (Tanrikulu, I. C., Schmitt, E., Mechulam, Y., Goddard, W. A., Tirrell, D. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15285.)

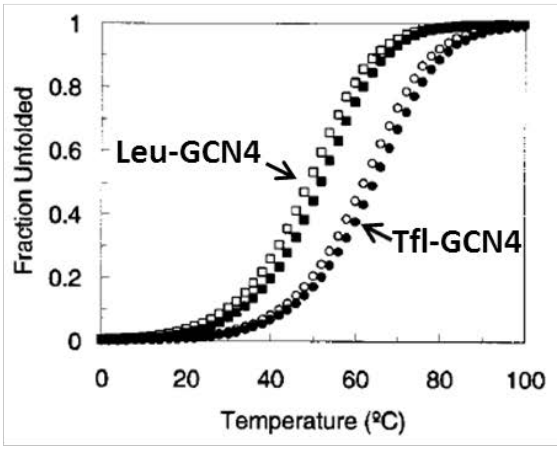
Figure 1.4



Cell-specific labeling of *E. coli* with azidonorleucine in the presence of human macrophages. A) *E. coli* and macrophages without the NLL-MetRS are incapable of utilizing azidonorleucine for protein synthesis (bottom row). *E. coli* cells expressing the NLL-MetRS metabolically incorporate azidonorleucine into cellular proteins. By performing copper-catalyzed, azide-alkyne click chemistry with TAMRA-alkyne, *E. coli* cells become fluorescent while human macrophages do not exhibit any fluorescence (top row, right panel). B) If azidohomoalanine is used in the place of azidonorleucine, *E. coli* and human macrophages metabolically incorporate azides into cellular proteins without any genetic modification. All cells exhibit fluorescence after click chemistry is performed with TAMRA (right panel).

Figure 1.5





Thermal Stabilization of leucine zipper GCN4 with trifluoroleucine. Leucine positions in the coiled-colied protein dimer are show as yellow space filling models (left panel). Incorporation of trifluoroleucine at 80% of these positions increases the melting temperature by 13°C as measured by circular dichroism spectroscopy (right panel).

Adapted with permission from (Tang, Y., Ghirlanda, G., Vaidehi, N., Kua, J., Mainz, D. T., Goddard, W. A., DeGrado, W. F., Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790). Copyright (2001) American Chemical Society.

References

1. Farinas, E. T.; Bulter, T.; Arnold, F. H. *Curr. Opin. Biotechnol.* **2001**, *12*, 545.
2. Shao, Z. X.; Arnold, F. H. *Curr. Opin. Struct. Biol.* **1996**, *6*, 513.
3. Díaz-Rodríguez, A.; Davis, B. G. *Curr. Opin. Chem. Biol.* **2011**, *15*, 211.
4. Arnold, U. *Biotechnol. Lett* **2009**, *31*, 1129.
5. Dirksen, A.; Dawson, P. E. *Curr. Opin. Chem. Biol.* **2008**, *12*, 760.
6. Tann, C.-M.; Qi, D.; Distefano, M. D. *Curr. Opin. Chem. Biol.* **2001**, *5*, 696.
7. Roy, H.; Ibba, M. *Nature* **2006**, *443*, 41.
8. Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092.
9. Young, T. S.; Schultz, P. G. *J. Biol. Chem.* **2010**, *285*, 11039.
10. Lakshmipathy, S. K.; Tomic, S.; Kaiser, C. M.; Chang, H.-C.; Genevaux, P.; Georgopoulos, C.; Barral, J. M.; Johnson, A. E.; Hartl, F. U.; Etchells, S. A. *J. Biol. Chem.* **2007**, *282*, 12186.
11. Tagami, S.; Sekine, S.-i.; Kumarevel, T.; Hino, N.; Murayama, Y.; Kamegamori, S.; Yamamoto, M.; Sakamoto, K.; Yokoyama, S. *Nature* **2010**, *468*, 978.
12. Chin, J. W.; Santoro, S. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *JACS* **2002**, *124*, 9026.
13. Chin, J. W.; Martin, A. B.; King, D. S.; Wang, L.; Schultz, P. G. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020.
14. Hino, N.; Okazaki, Y.; Kobayashi, T.; Hayashi, A.; Sakamoto, K.; Yokoyama, S. *Nat Meth* **2005**, *2*, 201.

15. Deiters, A.; Cropp, T. A.; Summerer, D.; Mukherji, M.; Schultz, P. G. *Biorg. Med. Chem. Lett.* **2004**, *14*, 5743.

16. Liu, H.; Wang, L.; Brock, A.; Wong, C.-H.; Schultz, P. G. *JACS* **2003**, *125*, 1702.

17. Cho, H.; Daniel, T.; Buechler, Y. J.; Litzinger, D. C.; Maio, Z.; Putnam, A.-M. H.; Kraynov, V. S.; Sim, B.-C.; Bussell, S.; Javahishvili, T.; Kaphle, S.; Viramontes, G.; Ong, M.; Chu, S.; GC, B.; Lieu, R.; Knudsen, N.; Castiglioni, P.; Norman, T. C.; Axelrod, D. W.; Hoffman, A. R.; Schultz, P. G.; DiMarchi, R. D.; Kimmel, B. E. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 9060.

18. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003**, *301*, 964.

19. Chin, J. W.; Cropp, T. A.; Chu, S.; Meggers, E.; Schultz, P. G. *Chem. Biol.* **2003**, *10*, 511.

20. Mukai, T.; Kobayashi, T.; Hino, N.; Yanagisawa, T.; Sakamoto, K.; Yokoyama, S. *Biochem. Biophys. Res. Commun.* **2008**, *371*, 818.

21. Gautier, A.; Nguyen, D. P.; Lusic, H.; An, W.; Deiters, A.; Chin, J. W. *JACS* **2010**, *132*, 4086.

22. Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10317.

23. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. *Nat Chem* **2012**, *4*, 298.

24. Greiss, S.; Chin, J. W. *JACS* **2011**, *133*, 14196.

25. Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W. *Nat. Chem. Biol.* **2012**, *8*, 748.

26. Liu, C. C.; Schultz, P. G. *Annu. Rev. Biochem* **2010**, *79*, 413.

27. Chin, J. W. *EMBO J.* **2011**, *30*, 2312.

28. Xie, J. M.; Schultz, P. G. *Curr. Opin. Chem. Biol.* **2005**, *9*, 548.

29. Beene, D. L.; Dougherty, D. A.; Lester, H. A. *Curr. Opin. Neurobiol.* **2003**, *13*, 264.

30. Caskey, C. T.; Tompkins, R.; Scolnick, E.; Caryk, T.; Nirenberg, M. *Science* **1968**, *162*, 135.

31. Capecchi, M. R. *Proc. Natl. Acad. Sci. USA* **1967**, *58*, 1144.

32. Wang, K.; Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat Biotech* **2007**, *25*, 770.

33. Johnson, D. B. F.; Xu, J.; Shen, Z.; Takimoto, J. K.; Schultz, M. D.; Schmitz, R. J.; Xiang, Z.; Ecker, J. R.; Briggs, S. P.; Wang, L. *Nat. Chem. Biol.* **2011**, *7*, 779.

34. Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; Emig, C. J.; Bang, D.; Hwang, S. J.; Jewett, M. C.; Jacobson, J. M.; Church, G. M. *Science* **2011**, *333*, 348.

35. Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *JACS* **1995**, *117*, 536.

36. Van Hest, J. C. M.; Tirrell, D. A. *FEBS Lett.* **1998**, *428*, 68.

37. Link, A. J.; Tirrell, D. A. *Methods* **2005**, *36*, 291.

38. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2000**, *39*, 2148.

39. Tang, Y.; Tirrell, D. A. *Biochemistry* **2002**, *41*, 10635.

40. Carrico, I. S.; Maskarinec, S. A.; Heilshorn, S. C.; Mock, M. L.; Liu, J. C.; Nowatzki, P. J.; Franck, C.; Ravichandran, G.; Tirrell, D. A. *JACS* **2007**, *129*, 4874.

41. Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10180.

42. Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15285.

43. Yoo, T. H.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2007**, *46*, 5340.

44. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9482.

45. Beatty, K. E.; Liu, J. C.; Xie, F.; Dieterich, D. C.; Schuman, E. M.; Wang, Q.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2006**, *45*, 7364.

46. Beatty, K. E.; Xie, F.; Wang, Q.; Tirrell, D. A. *JACS* **2005**, *127*, 14150.

47. Hinz, F. I.; Dieterich, D. C.; Tirrell, D. A.; Schuman, E. M. *ACS Chem. Neurosci.* **2012**, *3*, 40.

48. Melemedjian, O. K.; Asiedu, M. N.; Tillu, D. V.; Peebles, K. A.; Yan, J.; Ertz, N.; Dussor, G. O.; Price, T. J. *J. Neurosci.* **2010**, *30*, 15113.

49. Srikantan, S.; Abdelmohsen, K.; Lee, E. K.; Tominaga, K.; Subaran, S. S.; Kuwano, Y.; Kulshrestha, R.; Panchakshari, R.; Kim, H. H.; Yang, X.; Martindale, J. L.; Marasa, B. S.; Kim, M. M.; Wersto, R. P.; Indig, F. E.; Chowdhury, D.; Gorospe, M. *Mol. Cell. Biol.* **2011**, *31*, 3790.

50. Qi, D.; Huang, S.; Miao, R.; She, Z.-G.; Quinn, T.; Chang, Y.; Liu, J.; Fan, D.; Chen, Y. E.; Fu, M. *J. Biol. Chem.* **2011**, *286*, 41692.

51. Ueno, T.; Kaneko, K.; Sata, T.; Hattori, S.; Ogawa-Goto, K. *Nucleic Acids Res.* **2012**, *40*, 3006.

52. Deal, R. B.; Henikoff, J. G.; Henikoff, S. *Science* **2010**, *328*, 1161.

53. Tcherkezian, J.; Brittis, P. A.; Thomas, F.; Roux, P. P.; Flanagan, J. G. *Cell* **2010**, *141*, 632.

54. Yoon, Byung C.; Jung, H.; Dwivedy, A.; O'Hare, Catherine M.; Zivraj, Krishna H.; Holt, Christine E. *Cell* **2012**, *148*, 752.

55. Hodas, J. J. L.; Nehring, A.; Höche, N.; Sweredoski, M. J.; Pielot, R.; Hess, S.; Tirrell, D. A.; Dieterich, D. C.; Schuman, E. M. *Proteomics* **2012**, *12*, 2464.

56. Lepthien, S.; Merkel, L.; Budisa, N. *Angew. Chem. Int. Ed.* **2010**, *49*, 5446.

57. Merkel, L.; Schauer, M.; Antranikian, G.; Budisa, N. *ChemBioChem* **2010**, *11*, 1505.

58. Panchenko, T.; Zhu, W. W.; Montclare, J. K. *Biotechnol. Bioeng.* **2006**, *94*, 921.

59. Van Hest, J. C. M.; Kiick, K. L.; Tirrell, D. A. *JACS* **2000**, *122*, 1282.

60. Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *JACS* **2004**, *126*, 10598.

61. NLL-MetRS is a mutant MetRS discovered by Tanrikulu et al that is capable of charging azidonorleucine to $tRNA^{MET}$.

62. Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715.

63. Zhang, C. G.; Chromy, B. A.; McCutchen-Maloney, S. L. *Expert Rev Proteomics* **2005**, *2*, 187.

64. Ngo, J. T.; Babin, B. M.; Champion, J. A.; Schuman, E. M.; Tirrell, D. A. *ACS Chem Biol* **2012**, *7*, 1326.

65. Montclare, J. K.; Son, S.; Clark, G. A.; Kumar, K.; Tirrell, D. A. *ChemBioChem* **2009**, *10*, 84.

66. Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.

67. Lee, K. H.; Lee, H. Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. *Biochemistry* **2004**, *43*, 16277.

68. Tang, Y.; Tirrell, D. A. *JACS* **2001**, *123*, 11089.

69. Bilgiçer, B.; Kumar, K. *Tetrahedron* **2002**, *58*, 4105.

70. Bilgicer, B.; Xing, X.; Kumar, K. *J. Am. Chem. Soc.* **2001**, *123*, 11815.

71. Budisa, N.; Pipitone, O.; Siwanowicz, I.; Rubini, M.; Pal, P. P.; Holak, T. A.; Gelmi, M. L. *Chem. Biodivers.* **2004**, *1*, 1465.

72. Wang, P.; Tang, Y.; Tirrell, D. A. *JACS* **2003**, *125*, 6900.

73. Yoo, T. H.; Link, A. J.; Tirrell, D. A. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13887.

74. Montclare, J. K.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2006**, *45*, 4518.

CHAPTER 2

Two-Strain, Cell-selective Protein Labeling in Mixed Bacterial Cultures

Abstract

Cell-selective metabolic labeling of proteins with non-canonical amino acids enables the study of proteomic changes in specified subpopulations of complex multi-cellular systems. For example, azidonorleucine (Anl) and 2-aminooctynoic acid, both of which are activated by an engineered methionyl-tRNA synthetase (designated NLL-MetRS), are excluded from proteins made in wild-type cells but incorporated readily into proteins made in cells that carry NLL-MetRS. To expand the set of tools available for cell-selective metabolic labeling, we sought a MetRS variant capable of activating propargylglycine (Pra). Pra was chosen as the target amino acid because its alkynyl side chain can be selectively and efficiently conjugated to azide-functionalized fluorescence probes and affinity tags. Directed evolution, using active-site randomization and error-prone PCR, yielded a MetRS variant (designated PraRS) capable of incorporating Pra at near-quantitative levels into proteins made in a Met-auxotrophic strain of Escherichia coli cultured in Met-depleted media. Proteins made in E. coli strains expressing PraRS were labeled with Pra in Met-supplemented media as shown by in-gel fluorescence after conjugation to Cy5-azide. The combined use of NLL-MetRS and PraRS enabled differential, cell-selective labeling of marker proteins derived from two bacterial strains co-cultured in media supplemented with Met, Anl, and Pra. Treatment of the mixed marker proteins by sequential strain-promoted and copper (I)-catalyzed cycloadditions allowed straightforward identification of the cellular origin of each protein.

Introduction

The methionyl-tRNA synthetase (MetRS) of E. coli is promiscuous with respect to activation of natural and artificial amino acids under certain conditions.[1,2] For example, homopropargylglycine (Hpg) and azidohomoalanine (Aha) can be charged to tRNAMet by the wild-type MetRS and incorporated into cellular proteins; replacement of methionine (Met, **1**) occurs in statistical fashion throughout the proteome. The reactive side chains of Hpg and Aha provide sites for selective attachment of affinity tags or fluorescent probes.[3,4]

Bio-orthogonal non-canonical amino acid tagging (BONCAT) uses pulse labeling with reactive amino acids to separate newly synthesized cellular proteins from the pre-existing proteome.[5,6] Recent BONCAT experiments have enabled determination of the kinetics of nucleosome turnover,[7] analysis of localized synthesis of proteins critical to axonal maintenance,[8] visualization of localized protein synthesis regulated by transmembrane receptors,[9] in situ fluorescence imaging of new protein synthesis in rat hippocampal neurons,[10] and labeling of newly synthesized proteins in multicellular organisms.[11]

Engineering of the MetRS binding pocket has allowed further expansion of the set of Met analogs that can be incorporated into proteins.[12-15] Labeling of cellular proteins with azidonorleucine (Anl, **2**) requires expression of the L13N, Y260L, H301L variant of MetRS (NLL-MetRS), which activates Anl faster than Met.16 Ngo and coworkers recently reported the use of NLL-MetRS and Anl to effect selective, proteome-wide labeling of bacterial proteins made in mixed cultures of E. coli cells and mammalian macrophages.[16] Hang and coworkers have reported similar studies with NLL-MetRS and 2-aminooctynoic acid.[15]

In an effort to expand the set of tools available for cell-specific metabolic labeling of proteins, we chose an alkynyl amino acid, propargylglycine (Pra, **3**), an amino acid smaller than methionine , which is activated slowly or not at all by the wild-type E. coli MetRS.[17] Here we describe the engineering of a new variant of the E. coli MetRS (designated PraRS) that enables near-quantitative replacement of Met by Pra in bacterial proteins. We also show that Pra can be used as a cell-specific metabolic label in conjunction with PraRS. Finally, we report a new method that uses Anl and Pra to effect simultaneous, two-strain, cell-specific labeling of bacterial proteins.

Results and Discussion

Although Pra bears an alkyne group suitable for copper (I)- catalyzed azide-alkyne [3+2] cycloaddition, we chose a cell-based screening method that does not require exposure of host cells to the copper catalyst. The screening method used a previously evolved variant of the green fluorescent protein (GFPrm_AM) that is nearly insensitive to global replacement of Met by non-canonical amino acids.[13] Because expression of GFPrm_AM in Met-auxotrophic host cells in Met-depleted medium is limited by charging of $tRNA^{MET}$, Fluorescence-Activated Cell Sorting (FACS) can be used to identify MetRS variants that activate alternative substrates.

Our initial attempts to evolve MetRS variants for activation of Pra focused on the amino acid binding pocket, in which five positions (L13, A256, P257, Y260, and H301) were randomized using NNK codons. Thirteen rounds of positive and negative FACS screening yielded a MetRS variant (MO2b_13-2; L13P/A256G/P257T/Y260Q/H301F; Figure 2.2) capable of activating Pra. MO2b_13-2 was also found to misincorporate an unidentified amino acid at Met codons (data not shown). A new library (MO9) based on MO2b_13-2 was generated by error-prone PCR to improve

specificity for Pra. Highly active mutants obtained from screening MO9 shared a common trait: amino acid changes were localized around the KMSKS motif, a highly conserved sequence that plays an important role in stabilizing the aminoacyl adenylate intermediate in the charging of tRNA.[18,19] In addition, a truncation (E548Δ; Δ is the TGA codon) at the end of the catalytic core of MetRS was found to be reversed in some of the most active MetRS variants.[20] The C-terminal domain of MetRS has been shown to play a role in dimerization of the synthetase and contributes to increased tRNA affinity.[21] Mutations identified by screening the MO9 library were recombined to generate a MetRS variant that we designated PraRS (L13P/A256G/P257T/Y260Q/ H301F/A331V/Δ548E; Figure 2.2).

Lysates derived from E. coli strain TYJV2, which carries a plasmid-borne copy of the gene encoding PraRS, were treated with tetramethylrhodamine-azide (TAMRA-azide, **4**) to confirm incorporation of Pra into cellular proteins (Figure 2.3a). Protein labeling was detected by in-gel fluorescence within 5 min of addition of 4 mM Pra to the culture medium. Although labeling was detected in lysates treated with lower concentrations (e.g., 500 μM) of Pra, supplementation at higher concentrations increased the incorporation level, as shown by increases in the apparent mass of GFPrm_AM due to multi-site conjugation of TAMRA-azide (Figure 2.3b). Matrix-assisted laser desorption ionization mass spectrometry (MALDI-MS) of trypsinized GFPrm_AM expressed in Met-depleted M9 minimal medium supplemented with 4 mM Pra confirmed near-quantitative incorporation of Pra at positions encoding Met (Figure 2.4). Lower levels of Met replacement were accomplished without depletion of the natural amino acid. For example, expression of GFPrm_AM in M9 medium supplemented with 269 μM Met and 4 mM Pra yielded a protein containing approximately a 10:1 ratio of Met to Pra.

The use of PraRS for cell-selective proteomic labeling does not require the use of auxotrophic bacterial strains. Prototrophic DH10B E. coli strains harboring plasmid-borne, constitutively expressed genes encoding wild-type MetRS, NLL-MetRS, or PraRS were grown to mid-log phase and then treated with 1 mM Pra and 269 μM Met for 1 hr.. Cells were lysed and lysates were subjected to copper (I)-catalyzed ligation to a Cy5-azide probe (**5**). Labeling was detected only in lysates derived from cells bearing PraRS (Figure 2.5). Strains that over-expressed either wild-type MetRS or NLL-MetRS showed insignificant levels of proteomic labeling. Encouraged by the fact that Pra is not a substrate for the NLL-MetRS, we sought to develop a method for independently labeling the proteomes of two intermixed populations of cells (Figure 6a). His-tagged marker proteins DHFR and GFPm were co-expressed with NLL-MetRS and PraRS respectively, in prototrophic DH10B cells. The two strains were grown separately in M9 medium supplemented with all 20 canonical amino acids until OD600 ~ 1 and then mixed together in equal amounts. The mixed culture was then supplemented with 1 mM Anl, 4 mM Pra, and 269 μM Met. Expression of the marker proteins was induced by addition of IPTG. After 2 hr., His-tagged proteins were purified from the mixed culture.

To prevent cross-coupling of marker proteins derived from different strains, tagging reactions were run sequentially (Figure 2.6). The mixture of DHFR and GFPm was first treated with dibenzocyclooctyne-Alexa Fluor 488 (DIBO-Alexa Fluor 488, **6**) to tag Anl side chains. Without purification, the mixture was then treated with Cy5-azide under the copper-catalyzed conditions required to convert the unstrained alkyne side chains of Pra. In-gel fluorescence scanning showed that Alexa Fluor emission was confined to the marker derived from the NLL-MetRS strain, while Cy5 labeling was apparent only for the PraRS marker (Figure 2.7). Thus the use of Anl and Pra as metabolic labels, combined with the sequential labeling strategy described here, permits confident

identification of the cellular origins of proteins derived from mixed cell populations. Extension of this approach to cell-selective proteomic analysis via high-throughput mass spectrometry is straightforward.[5,15]

Conclusions

A MetRS engineered to use Pra, a small non-canonical amino acid with a reactive side chain, was developed to enable differential, cell-selective proteome-wide labeling of two intermixed bacterial populations. The mutations that differentiate PraRS from wild-type MetRS highlight the importance of error-prone PCR in the evolution of synthetase variants that activate non-canonical amino acid substrates. Previous efforts in synthetase engineering have focused for the most part on the amino acid binding pocket of the enzyme.[13,14,22-24] Here we found that mutations adjacent to the KMSKS region of MetRS can enhance charging of Pra; these mutations would not have been identified without the use of error-prone PCR.

We and others have shown previously that the use of appropriately-designed amino acids and mutant synthetases can enable time-resolved, cell-selective metabolic labeling of proteins.[15,16] By using the Anl/NLL-MetRS system in conjunction with Pra and PraRS, investigators can now interrogate the proteomic responses of two intermixed bacterial strains without crosstalk. Important cellular phenomena that would be amenable to study via dual cell-selective labeling would include bacterial quorum sensing, the development of biofilms, and interactions among commensal bacterial.[25-32]

Materials and Methods

Materials.

Restriction enzymes and T4 DNA ligase were purchased from New England Biolabs (Beverly, MA). PfuUltra II Fusion and Mutazyme II were purchased from Agilent Technologies (Santa Clara, CA). DNA oligomers were synthesized by Integrated DNA Technologies (Coralville, IA). L-Anl was prepared by previously reported methods.[33] Canonical amino acids and L-Pra were purchased from Sigma-Aldrich (St. Louis, MO). His-tag purifications were carried out using nickel-nitrilotriacetic acid resin purchased from Qiagen (Valencia, CA). Sequencing grade trypsin was purchased from Promega (Madison, WI).

Construction of MetRS Libraries.

Plasmid pMTY11, which contains a MetRS expression cassette, was described previously.[13] Active site randomization of the MetRS gene was carried out at the L13, A256, P257, Y260, and H301 positions by site-overlap extension using primers with NNK codons. PCR products were purified on 1% agarose gels. Fragments containing the randomized positions were assembled and amplified using PfuUltra II Fusion DNA polymerase to construct the MetRS library with five randomized positions. The library was re-ligated into pMTY11 using BamHI/NotI restriction sites. The ligation product was transformed into E. coli strain TYJV2 bearing the previously described plasmid pQE80L/GFPrm_AM via electroporation.[13]

Error-prone PCR mutagenesis of MO2b_13-2 was per-formed with Mutazyme II using primers Lib_Fwd (5'-TTCCGACAGCTACGTCGCGGAAC-3') and Lib_Rev (5'-GAAGGACCGTAGAAGGTCCTTTAGAG-3'), which flank the MetRS gene. The target mutation rate was two to four mutations per MetRS gene in the library. The PCR product was

purified on a 1% agarose gel and subsequently cloned into pREP4/MO2b_13-2 using the BamHI/NotI restriction sites. The ligation product was transformed into E. coli strain TYJV2 bearing pQE80L/GFPrm_AM via electroporation.

Screening of MetRS libraries via FACS.

Met auxotrophic TYJV2 cells containing pQE80L/GFPrm_AM and the MetRS libraries were grown at 37^{o}C to mid-log phase (OD600 = 0.8-1.0) in M9 minimal medium (M9 salts, 0.2% glucose, 1 mM MgSO4, 0.1 mM CaCl2, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), ampicillin (200 mg/L) and kanamycin (35 mg/L). Upon reaching mid-log phase, cells were washed twice with cold 0.9% NaCl and resuspended in M9 minimal media supplemented either with 19 amino acids (minus Met) at 40 mg/L or with 19 amino acids plus 4 mM Pra. After 30 min., expression of GFPrm_AM was induced by addition of 1 mM IPTG. After 2 hr., cells were washed and resuspended in phosphate buffered saline (pH 7.4) for FACS screening.

FACS screening was performed on a MoFLO XDP cell sorter (Beckman Coulter, Brea, CA) using an argon laser (488 nm) and 530/40 nm bandpass filter. Side scatter was utilized for event triggering; forward and side scatter gating was used to remove non-E. coli events. Event rate was maintained between 20000 and 30000 events per second. Positive sorting to enrich MetRS variants specific for Pra involved expression of GFPrm_AM in Pra-supplemented media and collecting the 0.5-1% of cells characterized by the highest levels of fluorescence. Negative sorting to remove MetRS variants that mischarge canonical amino acids was accomplished by collecting non-fluorescent cells when GFP was expressed in Met-depleted media. Cells were sorted into SOC medium (2 mL), rescued at 37^{o}C for 1 hr., and further diluted 1:10 with LB medium for overnight growth. Sorted populations were stored at -80^{o}C in 25% glycerol.

Pulse Labeling.

Met auxotrophic DH10B cells containing a constitutively expressed PraRS cassette in the pREP4 plasmid (pREP4/PraRS) were grown to mid-log phase in M9 minimal medium containing all 20 canonical amino acids. Addition of 4 mM Pra to the growth medium was performed without Met depletion. At selected time points, aliquots of cells were removed and chloramphenicol (170 mg/L) was added to inhibit further protein synthesis. Cells were pelleted, lysed, and subjected to copper (I)-catalyzed cycloaddition with 25 μM TAMRA-azide, 100 μM CuSO4, 500 μM tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), and 5000 μM sodium ascorbate as the reducing agent.[34] After 30 min, each sample was run on a 12% SDS-PAGE gel without purification. Labeled proteins were visualized using a Typhoon Trio (GE Healthcare, Piscataway, NJ) with a 532 nm laser and 580/30 nm bandpass filter.

Mass Spectrometry.

Expression of GFPrm_AM was carried out in TYJV2 cells transformed with pQE80L/GFPrm_AM and pREP4/PraRS. Cells were grown to mid-log phase and then washed twice with cold 0.9% NaCl. Cells were subsequently resuspended in M9 minimal medium supplemented with 19 amino acids (minus Met), 20 amino acids, or 19 amino acids plus 4 mM Pra. After 30 min., IPTG was added to a final concentration of 1 mM to induce expression of GFPrm_AM. Utilizing the N-terminal 6 x histidine tag of GFPrm_AM, the protein was purified under denaturing conditions according to the manufacturer's protocol. Tryptic digests of purified GFPrm_AM were subjected to MALDI time-of-flight mass spectrometry with α-cyano-4-hydroxycinnamic acid as matrix. An Applied Biosystems Voyager DE-PRO equipped with a 20-Hz nitrogen laser was used for all peptide analyses.

In-gel Fluorescence Detection of Two-strain, Cell-selective Protein Labeling.

Overnight M9 cultures of TYJV2 cells transformed either with pQE80L/GFPm and pREP4/PraRS or with the previously described pQE80L/DHFR/NLL-MetRS (pJTN5) and pREP4 were diluted 1:100 into fresh M9 minimal medium containing all 20 amino acids.[16] Cultures were grown to mid-log phase and equal numbers of cells from each of the two strains were mixed together. After mixing, the medium was supplemented with 1 mM L-Anl, 4 mM L-Pra, and 269 μM L-Met. Expression of his-tagged DHFR and his-tagged GFPm was induced by addition of 1 mM IPTG for 2 hr.. DHFR and GFPm were purified on nickel-nitrilotriacetic acid resin according to the manufacturer's protocols under denaturing conditions.

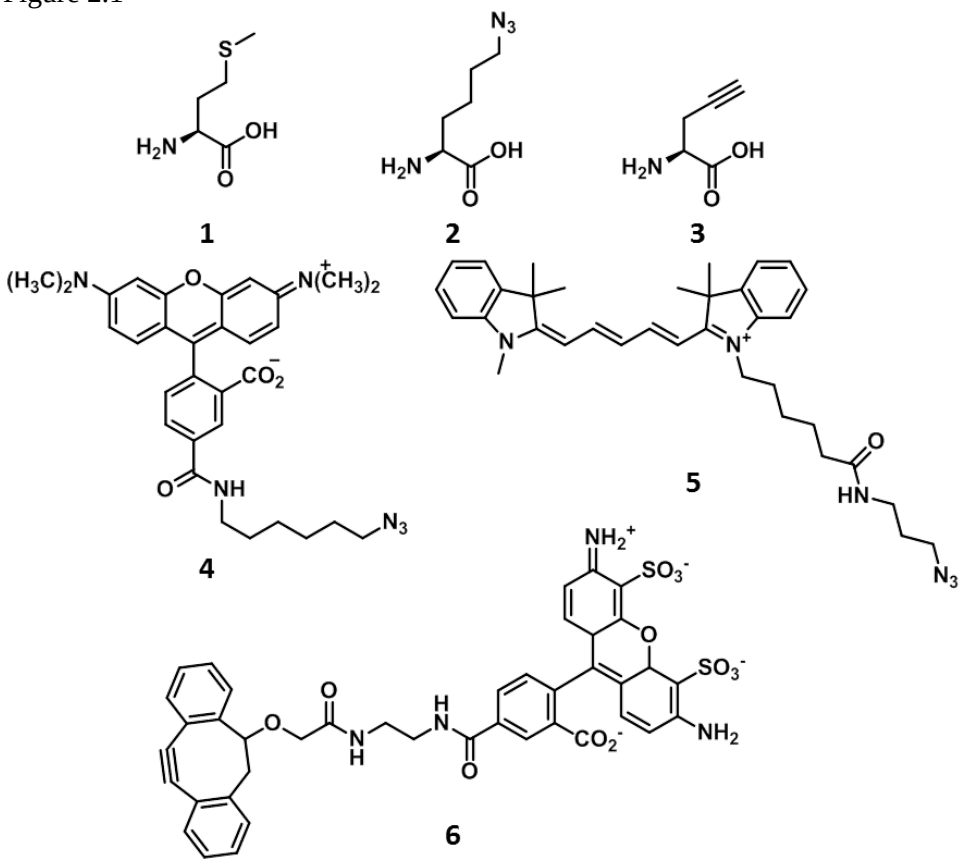
Sequential labeling of the marker protein mixture was accomplished with DIBO-Alexa Fluor 488 (Invitrogen, Carlsbad, CA), and Cy5-alkyne (Lumiprobe, Hallandale Beach, FL). The protein mixture (approximately 2 mg/ml) was first treated with 10 μM DIBO-Alexa Fluor 488 in tris-buffered saline for 1 hr.. Copper (I)-catalyzed cycloaddition was then performed without intermediate purification using 25 μM of Cy5-azide, 100 μM CuSO4, 500 μM THPTA, and 5000 μM sodium ascorbate as the reducing agent. After 30 min., samples were run on a 12% SDS-PAGE gel. Labeled marker proteins were visualized using a Typhoon Trio with a 488 nm laser and 526 nm shortpass filter as well as a 633 nm laser and 680/30 nm bandpass filter.

Acknowledgments

We thank John Ngo, Janek Szychowski, Caglar Tanrikulu, and James Van Deventer for helpful discussions, and Felicia Rusnak, Jie Zhou and Mona Shahgholi for help with spectroscopic

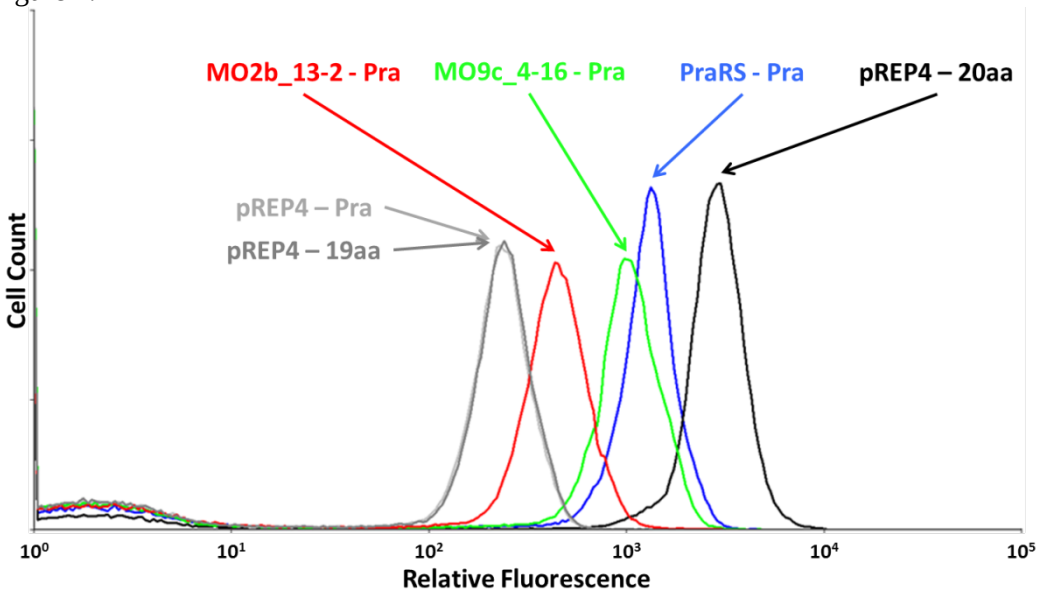
measurements. This work was supported by National Institutes of Health Grant NIH R01 GM062523 and by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office.

Figure 2.1



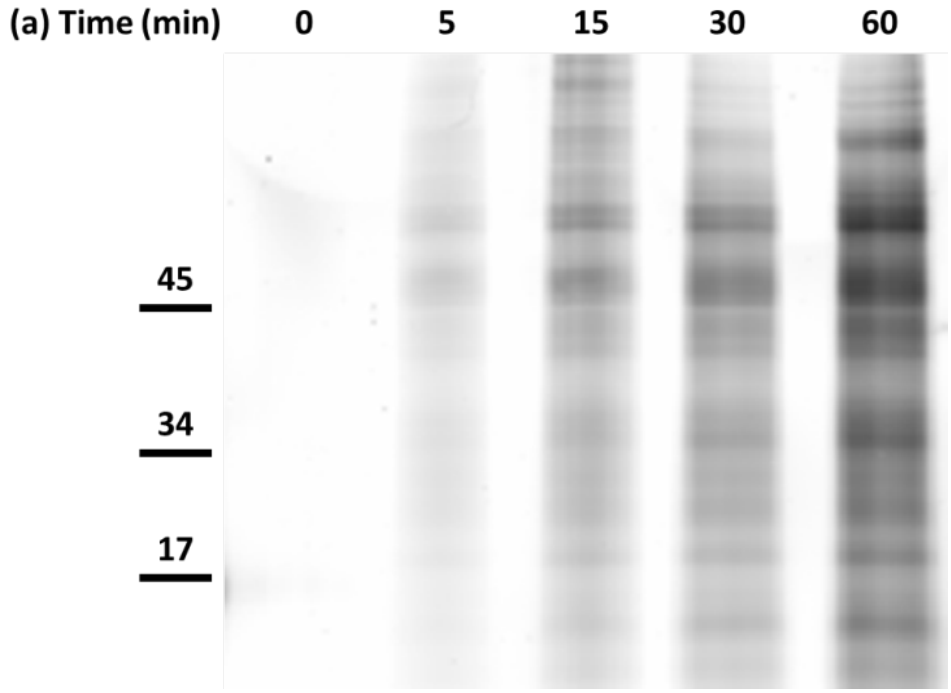
Amino acids and fluorescent probes used in this study. **1**: Met; **2**: Anl; **3**: Pra; **4**: TAMRA-azide; **5**: Cy5-azide; **6**: DIBO-AlexaFluor 488

Figure 2.2



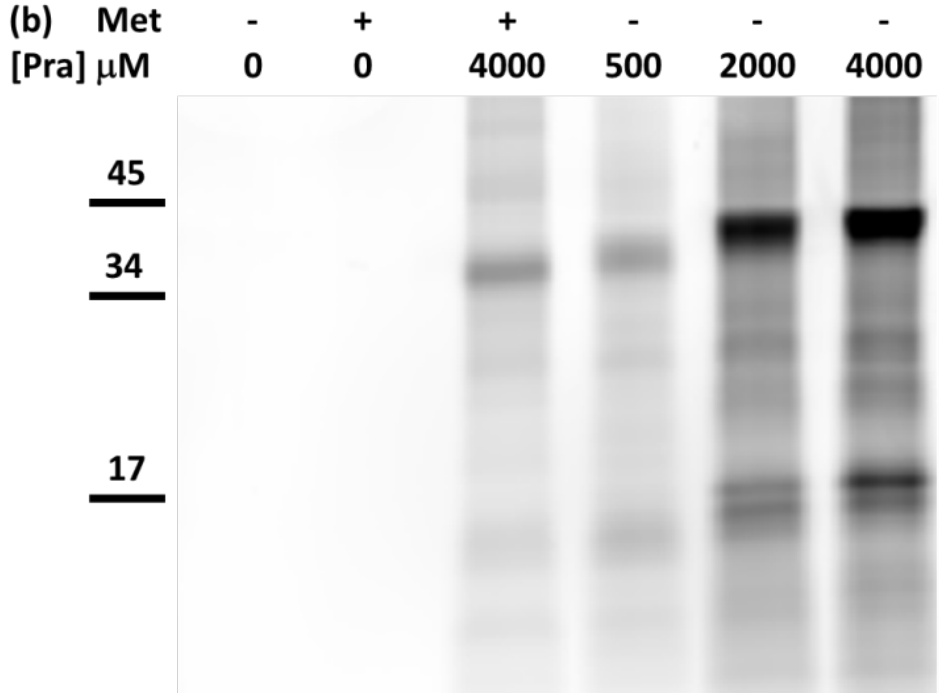
FACS histograms of *E. coli* variants obtained from evolution of MetRS for Pra activity. Addition of Pra to *E. coli* cultures in M9 medium supplemented with 19 amino acids (-Met) did not change the GFPrm_AM fluorescence of cells expressing endogenous levels of wild-type MetRS (compare pREP4-19aa and pREP4-Pra; pREP4 is the vector lacking the MetRS cassette). Variant MO2b_13-2 was isolated from the library prepared by randomizing the amino acid binding pocket. MO9c_4-16 was isolated from the error-prone PCR library generated from MO2b_13-2, and carries an additional mutation (A331V). Reverting the stop codon in MO9c_4-16 to Glu yielded PraRS, which shows the highest fluorescence signal (PraRS-Pra) for cells expressed in media supplemented with Pra.

Figure 2.3



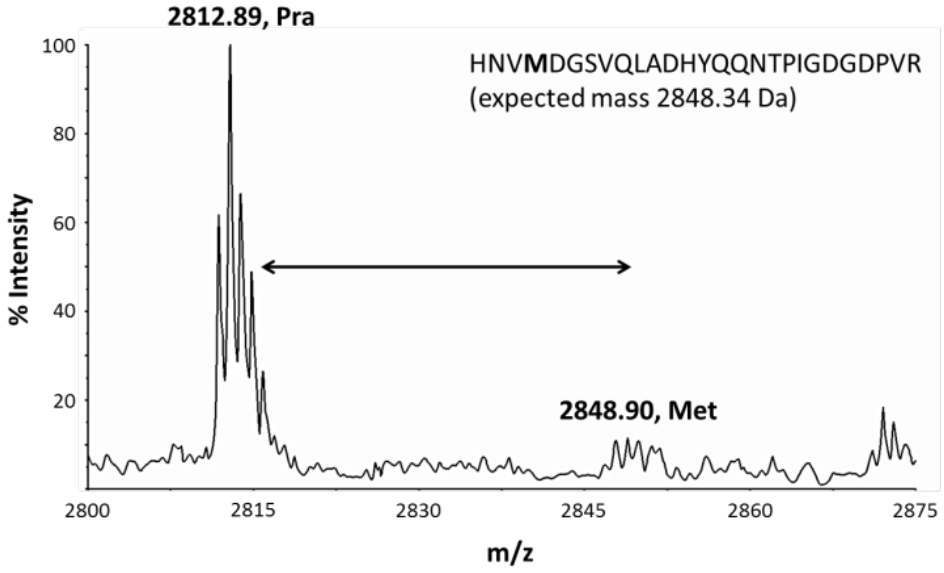
In-gel fluorescence detection of TAMRA-azide after labeling of lysates from Met-auxotrophic *E. coli* cells. a) Time-dependent, proteomic labeling after addition of Pra. At time = 0, 4 mM Pra was added to the medium without depletion of Met (269 μM). Aliquots were removed at regular time intervals and chloramphenicol was added to prevent further protein synthesis. Incorporation of Pra into cellular proteins was detected by labeling with TAMRA-azide within 5 min. of addition of Pra.

Figure 2.3 (cont.)



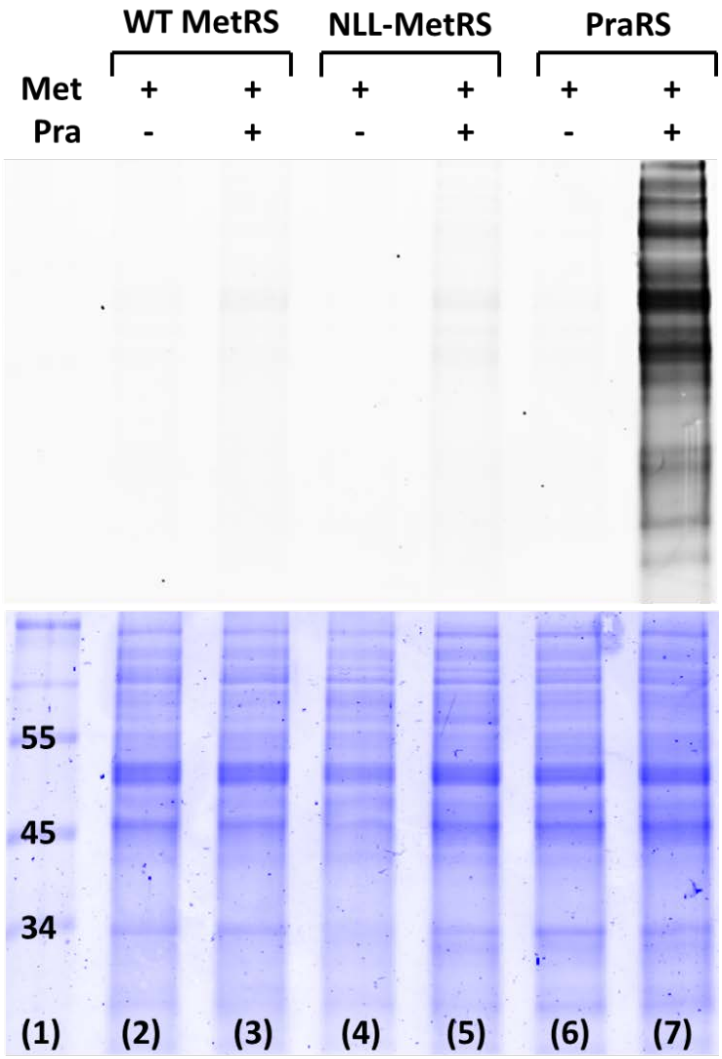
b) The extent of proteomic labeling was dependent on Pra concentration. Expression of GFPrm_AM (top band) was induced after Met depletion and resuspension in medium with and without 269 μM Met and various concentrations of Pra. TAMRA-azide was detected in all lanes where Pra was added to the culture medium. The apparent molecular weight of GFPrm_AM increased with increasing amounts of Pra due to higher levels of labeling with TAMRA-azide.

Figure 2.4



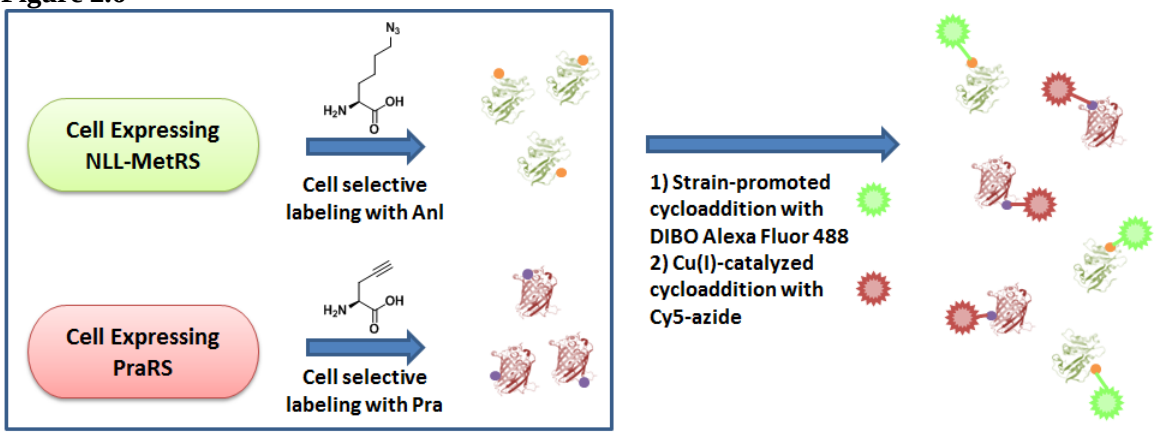
MALDI-TOF spectrum of tryptic digest of purified GFPrm_AM. His-tagged protein was expressed in a Met-auxotrophic strain of *E. coli* harboring PraRS in medium supplemented with 19 aa (-Met) plus 4 mM Pra. Replacement of Met by Pra in the peptide HNVMDGSVQLADHYQQNTPIGDGDPVR results in a decrease in mass of 36 Da.

Figure 2.5



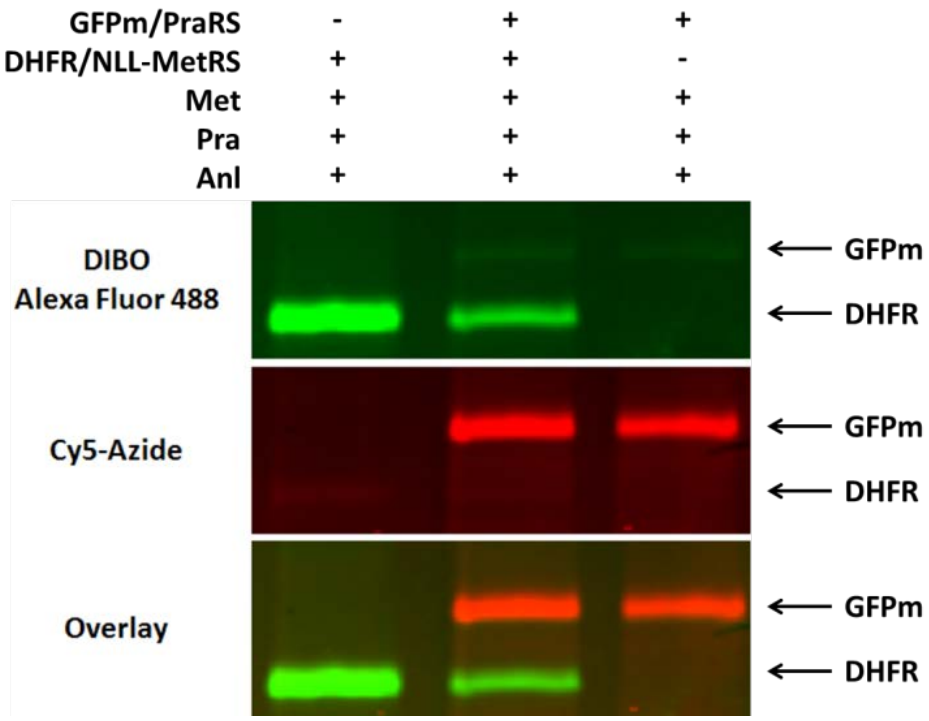
In-gel fluorescence detection of Cy5-azide (top) after labeling of *E. coli* lysates. Colloidal blue staining (bottom) confirms protein loading in all lanes. Conjugation of the fluorescent probe to cellular proteins requires both expression of PraRS and addition of 1 mM Pra to the medium (lane 7). Over-expression of the wild-type E. coli MetRS (lane 3) confirms the lack of activity toward Pra previously reported by Kiick et al.[17]. NLL-MetRS also shows little or no evidence of activation of Pra (lane 5). Incorporation of Pra into cellular proteins in the presence of Met is confirmed by MALDI-MS (supporting information).

Figure 2.6



Schematic representation of simultaneous, cell-selective metabolic labeling of two proteomes in a mixed bacterial culture.

Figure 2.7



In-gel fluorescence detection of DIBO-Alexa Fluor 488 and Cy5-azide after sequential labeling of mixed cell lysates. Dihydrofolate reductase (DHFR) metabolically labeled with Anl was conjugated to DIBO-Alexa Fluor 488 via copper-free, strain-promoted cycloaddition. GFPm metabolically labeled with Pra was conjugated to Cy5-azide by Cu(I)-catalyzed cycloaddition.

References

1. Cowie, D. B.; Cohen, G. N. *Biochim. Biophys. Acta* 1957, 26, 252-261.
2. Johnson, J. A.; Lu, Y. Y.; Van Deventer, J. A.; Tirrell, D. A. *Curr. Opin. Chem. Bio*. 2010, 14, 774-780.
3. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. *Angew. Chem. Int. Ed*. 2002, 41, 2596-2599.
4. Sletten, E. M.; Bertozzi, C. R. *Angew. Chem. Int. Ed*. 2009, 48, 6974-6998.
5. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 9482-9487.
6. Ngo, J. T.; Tirrell, D. A. *Acc. Chem. Res*. 2011, 44, 677-685.
7. Deal, R. B.; Henikoff, J. G.; Henikoff, S. *Science* 2010, 328, 1161-1164.
8. Yoon, B. C.; Jung, H.; Dwivedy, A.; O'Hare, C. M.; Zivraj, K. H.; Holt, C. E. *Cell* 2012, 148, 752-764.
9. Tcherkezian, J.; Brittis, P. A.; Thomas, F.; Roux, P. P.; Flanagan, J. G. *Cell* 2010, 141, 632-644.
10. Dieterich, D. C.; Hodas, J. J. L.; Gouzer, G.; Shadrin, I. Y.; Ngo, J. T.; Triller, A.; Tirrell, D. A.; Schuman, E. M. *Nat. Neurosci*. 2010, 13, 897-905.
11. Hinz, F. I.; Dieterich, D. C.; Tirrell, D. A.; Schuman, E. M. *ACS Chem. Neurosci*. 2012, 3, 40-49.
12. Link, A. J.; Vink, M. K. S.; Agard, N. J.; Prescher, J. A.; Bertozzi, C. R.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* 2006, 103, 10180-10185.
13. Yoo, T. H.; Tirrell, D. A. *Angew. Chem. Int. Ed*. 2007, 46, 5340-5343.
14. Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. *Proc. Natl. Acad. Sci. U.S.A.* 2009, 106, 15285-15290.
15. Grammel, M.; Zhang, M. M.; Hang, H. C. *Angew. Chem. Int. Ed*. 2010, 49, 5970-5974.

16. Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Bio*. 2009, 5, 715-717.

17. Kiick, K. L.; Weberskirch, R.; Tirrell, D. A. *FEBS Lett*. 2001, 502, 25-30.

18. Schmitt, E.; Meinnel, T.; Blanquet, S.; Mechulam, Y. *J. Mol. Bio*. 1994, 242, 566-577.

19. Mechulam, Y.; Dardel, F.; Le Corre, D.; Blanquet, S.; Fayat, G. *J. Mol. Bio*. 1991, 217, 465-475.

20. Cassio, D.; Waller, J.-P. *Eur. J. Biochem*. 1971, 20, 283-300.

21. Crepin, T.; Schmitt, E.; Blanquet, S.; Mechulam, Y. *Biochemistry* 2002, 41, 13003-13011.

22. Neumann, H.; Peak-Chew, S. Y.; Chin, J. W. *Nat. Chem. Biol*. 2008, 4, 232-234.

23. Xie, J.; Schultz, P. G. *Methods* 2005, 36, 227-238.

24. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G. *J. Am. Chem. Soc*. 2004, 126, 14306-14307.

25. Miller, M. B.; Skorupski, K.; Lenz, D. H.; Taylor, R. K.; Bassler, B. L. *Cell* 2002, 110, 303-314.

26. Zhu, J.; Miller, M. B.; Vance, R. E.; Dziejman, M.; Bassler, B. L.; Mekalanos, J. J. *Proc. Natl. Acad. Sci. U.S.A*. 2002, 99, 3129-3134.

27. Costerton, J. W.; Stewart, P. S.; Greenberg, E. P. *Science* 1999, 284, 1318-1322.

28. Davies, D. G.; Parsek, M. R.; Pearson, J. P.; Iglewski, B. H.; Costerton, J. W.; Greenberg, E. P. *Science* 1998, 280, 295-298.

29. Hall-Stoodley, L.; Costerton, J. W.; Stoodley, P. *Nat. Rev. Microbiol*. 2004, 2, 95-108.

30. Singh, P. K.; Schaefer, A. L.; Parsek, M. R.; Moninger, T. O.; Welsh, M. J.; Greenberg, E. P. *Nature* 2000, 407, 762-764.

31. Hooper, L. V.; Gordon, J. I. *Science* 2001, 292, 1115-1118.

32. Eckburg, P. B.; Bik, E. M.; Bernstein, C. N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S. R.; Nelson, K. E.; Relman, D. A. *Science* 2005, 308, 1635-1638.

33. Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *Nat. Prot*. 2007, 2, 1879-1883.

34. Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. *Angew. Chem. Int. Ed*. 2009, 48, 9879-9883.

CHAPTER 3

Effects of Quantitative Homoisoleucine Incorporation on Chloramphenicol Acetyltransferase

Abstract

Incorporation of non-canonical amino acids enables access to protein sequence space completely unknown to nature; the hope being that these new regions of sequence space will contain improved protein stability or catalytic activity. Chemists have taken advantage of the properties fluorination provides polymers: thermal and oxidative stability, hydrophobicity, and inertness to chemical reactions. Protein engineers have utilized fluorinated amino acids successfully in coiled-coiled protein interfaces lined by hydrophobic leucine residues, increasing melting temperatures up to 22^{o}C. Global incorporation of fluorinated amino acids into globular proteins like chloramphenicol acetyltransferase has had the reverse effect, reducing enzyme melting temperature by up to 9^{o}C. Fluorinated amino acids have also been limited by incorporation levels, creating proteins that are hybrids of natural and unnatural amino acids. Quantitative incorporation of homoisoleucine, a larger aliphatic analogue of leucine, into chloramphenicol acetyltransferase results in decreased thermostability. Mutations that stabilized chloramphenicol acetyltransferase for 80% trifluoroleucine incorporation did not increase protein thermostability when 100% homoisoleucine was incorporated at leucine positions. Attempts to recover lost thermostability of the homoisoleucine version chloramphenicol acetyltransferase by directed evolution are ongoing.

Introduction

Natural selection has been the driving force for protein evolution over millions of years, probing sequence space for beneficial mutations to improve organismal fitness. By removing the fitness constraint, enzymes can be evolved for a new set of criteria including enhanced thermostability[1-6], improved catalytic activity[7-9], or new substrate recognition[10-14].

Sequence space is controlled by the amino acid context available for protein synthesis. By performing codon reassignment to enable global replacement of a canonical amino acid with an artificial amino acid, sequence space is broadened beyond what was previously accessible by nature. Cirino et al. demonstrated global replacement of methionine by norleucine enabled improved peroxygenase activity by cytochrome P450, but the tradeoff involved significantly reduced thermostability[15]. Not all proteins are as permissive to global replacement; non-canonical amino acid incorporation can be similar to deleterious mutations depending on the deviation of the side chain from its natural counterpart.

For over 50 years, researchers have been exploring the use of fluorinated amino acids in organisms[16-21]. The Tirrell Lab has pioneered work in exploring proteins of fluorinated amino acid compositions through the use of residue specific incorporation[22-24]. Tang et al. has demonstrated that trifluorinated (Figure 3.1 [**2**]) and hexafluorinated (Figure 3.1 [**3**]) versions of leucine (Figure 3.1 [**1**]) provide substantial improvements to protein stability in leucine zippers[25,26]. However, attempting to enhance the stability of enzymes using trifluoroleucine is not so straightforward.

Global incorporation of trifluoroleucine into complex, folded proteins results in decreased functionality. Rather than throw away millions of years of evolution and attempt to build proteins from the ground up, directed evolution was applied to recover the lost properties. Yoo et al.

demonstrated that green fluorescent protein (GFP) required eleven rounds of directed evolution to create a trifluoroleucine GFP variant with similar spectral properties to leucine version of GFP[27]. Montclare et al. has demonstrated that incorporation of trifluoroleucine into chloramphenicol acetyltransferase (CAT) results in 20 fold loss in protein half-life at 60^{o}C. Two rounds of directed evolution resulted in a trifluoroleucine CAT variant with half-life at 60^{o}C comparable to that of wild-type CAT expressed with leucine. In either case, directed evolution did not yield a protein that had improved properties when expressed with trifluoroleucine compared to leucine.

In evolving trifluoroleucine versions of GFP and CAT, protein engineers were limited by the endogenous activity of leucyl-tRNA synthetase (LeuRS) toward trifluoroleucine. Incorporation levels of trifluoroleucine into either protein never exceeded 80%, allowing for two to four leucine codons to be decoded as leucine instead of trifluoroleucine. Residue-specific incorporation is incapable of controlling which leucine codons are decoded as trifluoroleucine or leucine; proteins evolved for trifluoroleucine are in fact evolved for tolerance to leucine or trifluoroleucine. The goal of creating proteins that prefer trifluoroleucine over leucine is not possible without quantitative replacement; another amino acid was required to explore proteins of completely novel composition.

Homoisoleucine (Figure 3.1 [**4**]) is a structural analogue of trifluoroleucine where the trifluoro group is replaced by a methyl group; replacement of leucine by homoisoleucine approaches 100% when proteins are expressed in leucine auxotrophs with overexpressed wild-type LeuRS. Van Deventer et al. demonstrated that homoisoleucine incorporation into the leucine zipper A1 increased the melting temperature by17^{o}C[28], exceeding the stabilizing effect of (2S,4R)-trifluoroleucine isomer[29]. Here we describe our attempts to evolve CAT expressed in homoisoleucine. Additionally, we look at a CAT mutant previously evolved for trifluoroleucine to determine how homoisoleucine and trifluoroleucine incorporation affect thermostability.

Results and Discussion

Sequence diversity in CAT was introduced using error-prone PCR. Starting with an initial library size on the order of 10^7, we attempted to screen CAT variants using a previously described 96 well plate assay[30]. Using a criterion of increased activity compared to wild type CAT after 1 hr. long incubation at 60°C, 1056 variants from the CAT library were screened. The top six variants from each plate were used to inoculate a plate for comparison. CAT variants with the highest activity after 60°C incubation were sequenced; only two out of twelve variants exhibited amino acid substitutions. The two variants were expressed and purified in batch conditions; neither variant exhibited any improvement in thermostability compared to wild type CAT.

A selection was performed to attempt to cull non-functioning variants as well as wild type CAT from the library. The original CAT library was grown in M9 containing all 20 amino acids; expression of the CAT variants was induced after shifting to M9 media containing 19 amino acids plus homoisoleucine. After a 15 min. expression period, aliquots were removed from the culture and spread on LB agar plates containing chloramphenicol (175 μg/L). Sequencing of 24 colonies after one round of selection resulted in no truncated variants; however, 75% of sequenced variants were still wild-type. An additional round of selection performed at higher concentrations of chloramphenicol (425 μg/L and 850 μg/L) did not yield any changes in the distribution of wild type CAT to CAT variants in the library population. Variants from the library after one round of selection were screened for activity after one hour incubation at 65°C. Of the 88 variants screened, eight of the ten highest activity variants did not contain any amino acid changes from wild-type.

Lacking success in screening CAT variants expressed in homoisoleucine for improved thermostability, we looked at the effects of homoisoleucine incorporation on a previously evolved

CAT variant[30]. Obtained through screening CAT libraries expressed in trifluoroleucine, L2A1 could potentially have acquired mutations that confer stability against homoisoleucine incorporation. Expression of wild-type CAT or L2A1 was carried out in M9 minimal media containing all 20 amino acids, 19 amino acids plus trifluoroleucine, or 19 amino acids plus homoisoleucine (Figure 3.2). As expected, expression in leucine gave the greatest yields of wild-type CAT or L2A1 (59.1 and 44.2 mg/L respectively) while trifluoroleucine expression produced the least wild-type CAT or L2A1 (3.15 and 3.43 mg/L respectively). Yields of wild-type CAT and L2A1 expressed in homoisoleucine (5.82 and 6.89 mg/L respectively) were approximately twice those of trifluoroleucine. Incorporation levels of trifluoroleucine (80%) and homoisoleucine (>95%) were consistent with previous work (Figure 3.3)[26,28,30].

L2A1 variants expressed in leucine, trifluoroleucine, and homoisoleucine were subjected to incubation at elevated temperatures to examine amino acid effects on thermostability (Figure 3.4). Consistent with previous work, the leucine form of L2A1 (Tm = 66.8^{o}C) is more stable than the trifluoroleucine form (Tm = 61.7^{o}C). Quantitative incorporation of homoisoleucine into CAT (Tm = 54.5^{o}C) results in a significant loss of thermostability compared to both leucine and trifluoroleucine. The mutations that stabilize L2A1 for trifluoroleucine do not appear to confer any stabilization with respect to homoisoleucine incorporation; L2A1 and wild-type CAT (Tm = 54.3^{o}C) expressed with homoisoleucine show similar behavior when exposed to elevated temperatures (Figure 3.5).

Similar trends are observed when looking at the half-life of L2A1 at 60^{o}C. L2A1 expressed in leucine ($t_{1/2}$ = 42.9 min) exhibited 5.7 fold greater half-life than L2A1 expressed in trifluoroleucine ($t_{1/2}$ = 7.5 min., Figure 3.7). The leucine and trifluoroleucine forms of L2A1 exhibit shorter half-life at 60^{o}C compared to previous work[30]; a 1.3 fold reduction of half-life between trifluoroleucine and

leucine forms was previously observed by Montclare et al. Negligible differences were observed between either wild-type CAT ($t_{1/2}$ = 1.2 min) or L2A1 ($t_{1/2}$ = 1.5 min) expressed in homoisoleucine (Figure 3.8).

Future Directions

Limitations in protein expression with homoisoleucine could be interfering with the screening process. Experiments with the purified, homoisoleucine form of wild-type CAT show that the protein half-life at 60^{o}C is in the order of 90 sec.. The enrichment of wild-type CAT in the screening process suggests that expression of the protein may be occurring under conditions where leucine is present. False positives could arise from leaky expression of the enzyme while in leucine-containing media or from incomplete leucine depletion before IPTG induction. We are exploring other expression vectors (such as those with arabinose inducible promoters to limit undesired protein expression), and optimization of the conditions used for leucine depletion.

Attempts to evolve a CAT variant that could tolerate quantitative incorporation of homoisoleucine may be limited by a marginally stable starting enzyme. It has been shown that thermostable enzymes are more accommodating to neutral/deleterious mutations that would benefit desired enzyme properties[31-33]. Utilizing a CAT enzyme derived from a thermophilic organism or evolving a leucine version of CAT for enhanced thermostability may create a better framework to study homoisoleucine. It may also be beneficial to introduce a preliminary step into the evolution process, in which CAT would be expressed with both homoisoleucine and leucine in the media. This would allow the accumulation of stabilizing mutations and remove leucine positions where homoisoleucine replacement is not tolerated. Later, as the protein become more stable, leucine would be removed and evolution would continue in a completely homoisoleucine context. This

technique has been previously employed for GFP where 80% trifluoroleucine incorporation reduced GFP fluorescence to background levels due to protein misfolding/aggregation[27].

Homoisoleucine incorporation provides increased thermostability to coiled-coiled proteins, but this did not translate to globular proteins. Quantitative replacement of leucine by homoisoleucine creates an enzyme of completely novel composition, but results in significant reduction of L2A1's thermostability compared to proteins expressed in leucine- or trifluoroleucine-containing media. Although evolved to recover lost thermostability upon incorporation of trifluoroleucine, L2A1 does not show any additional stability over wild-type CAT after homoisoleucine incorporation. Kwon et al. showed that hydration dynamics around homoisoleucine and trifluoroleucine residues differed[34]; mutations that stabilize for trifluoroleucine may not necessarily work for homoisoleucine. We continue to build upon this work and move toward developing proteins that have a preference for an unnatural sequence space, potentially accessing new properties not possible in nature's sequence space.

Materials and Methods

Cloning of CAT expression vector

All restriction enzymes and ligases used in this work were supplied by NEB. High fidelity polymerase PfuUltraII and error prone polymerase Mutazyme II are supplied by Agilent Technologies. In order to ensure high levels of replacement of leucine with homoisoleucine, a constitutively expressed copy of the LeuRS was cloned into the expression vector pQE-80L. LeuRS was cloned from *E. coli* genomic DNA and ligated into the NheI restriction sites of pQE-80L, creating the plasmid pQE-80L/LeuRS. The plasmid pCCCAT, constructed as described by Montclare[30], was digested with BamHI and HindIII to generate the gene for CAT; pQE-80L/LeuRS

was likewise digested with BamHI and HindIII to prepare the vector for ligation. The CAT gene was ligated into the multi-cloning site of pQE-80L/LeuRS with T4 DNA ligase, creating the plasmid pQE-80L/LeuRS/wtCAT. The same procedure was employed to generate the plasmid pQE-80L/LeuRS/L2A1; L2A1 is an evolved version of CAT stabilized for trifluoroleucine incorporation[30].

Construction of CAT gene library

Mutazyme II was employed for error prone amplification of the CAT gene. The library was prepared using pCCCAT and primers upstream of BamHI (5'-GTGAGCGGATAACAATTTCAC ACAG-3') and HindIII (5'-CAACCGAGCGTTCTGAACAAATC-3') restriction sites. The randomized CAT gene was excised with BamHI and HindIII and then ligated into pQE-80/LeuRS. The library was transformed into DH10B containing pREP4, yielding a library of $1.5x10^7$ based on the total number of transformants. Sequencing revealed a mutation rate of three to four base pair substitutions per CAT gene.

Thermostability screening in 96 well plates

CAT variants from the library were inoculated into 96 well plates containing 250 μL LB media. The first column for each 96 well plate was reserved for wild-type CAT. Cultures were grown overnight at 30°C, 70% humidity with 200 RPM shaking. Overnight 96 well cultures were used to inoculate a new 96 well plate containing 250 μL of M9 minimal media (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), ampicillin (200 μg/L), and kanamycin (35 μg/L). After 12 hr. of growth (30°C, 70% humidity with 200 RPM shaking), 96 well plates were centrifuged at 4000 RPM (6 min., 4°C)

and the medium was decanted. Cold, isotonic NaCl was added to each well and pellets were resuspended using a Beckman Multimek 96-channel pipetting robot. After repeating the wash step once, cells were resuspended in fresh M9 minimal media containing 19 amino acids minus leucine plus 500 μM homoisoleucine and 2 mM IPTG. Expression was carried out for 6 hours after which cells were pelleted at 4000 RPM for 6 min. at 4°C. Plates were stored at -80°C until assayed.

Cell pellets in 96-well plates were thawed at room temperature and resuspended into 200 μL 50 mM Tris pH 7.8, 0.5 mg/mL lysozyme, 0.5 units/ml of DNAseI, 0.25X B-PERII. Plates were incubated at 37°C for 30 min. and then lysates were clarified by centrifugation at 4000 RPM for 6 min. at 4°C. Plates were heated to 60°C in a water bath for 1 hr. and cooled back to room temperature in an ice bath. Lysates were clarified again before assays were performed. An aliquot of lysate (10 μL) was transferred to a 96 microwell plate containing 35 μL 50 mM Tris pH 7.8. The 2x reaction buffer (50 μL) comprised of 50 mM Tris pH 7.8, 2 mM 5',5'-dithiobis-(2-nitrobenzoic acid) (DTNB, Sigma), 0.8 mM acetyl-Coenzyme A, and 0.50 mg/mL BSA was added. The reaction was started by addition of 5 μL of chloramphenicol (final concentration 100 μM). Acetylation of chloramphenicol generates Coenzyme A; the free thiol of Coenzyme A reacts with DTNB creating 2-nitro-5-thiobenzoate, a yellow product. Formation of 2-nitro-5-thiobenzoate is monitored by measuring absorbance at 412 nm with a Tecan Safire2 96 well microplate reader. Variants displaying the highest activity after incubation at 60°C were pooled from all assayed plates and tested again for activity at room temperature prior to heating and after heating to 60°C.

Selection of CAT library for variants active after homoisoleucine incorporation

Cultures of *E. coli* strain DH10B containing pREP4 and pQE-80L/LeuRS/CAT library were grown at 37°C to mid-log phase (OD600 ~ 0.8-1.0) in M9 minimal media (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), ampicillin (200 μg/L), and kanamycin (35 μg/L). Upon reaching mid-log phase, cells were pelleted and washed thrice using 0.9% NaCl to deplete leucine. Cells were subsequently resuspended in M9 minimal media supplemented with 19 amino acids minus leucine plus 500 mM homoisoleucine. Addition of 1 mM IPTG to the media initiated expression of the CAT library. After 15 min., aliquots were removed from the culture and plated directly onto LB plates containing 200 μg/L ampicillin, 35 μg/L kanamycin, and 170 μg/L chloramphenicol. Colonies from plating were used directly to create new library stocks.

Batch expression of CAT variants

Expression of 6xHis-tagged CAT or 6xHis-tagged L2A1 was carried out in leucine auxotrophic DH10B/pREP4 bearing the plasmid pQE-80L/LeuRS/CAT or pQE-80L/LeuRS/L2A1. Cultures of DH10B/pREP4 containing pQE-80L/LeuRS/CAT or pQE-80L/LeuRS/L2A1 were grown at 37°C to mid-log phase (OD600 ~ 0.8-1.0) in M9 minimal media (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), ampicillin (200 μg/L), and kanamycin (35 μg/L). Upon reaching mid-log phase, cells were pelleted and washed thrice using 0.9% NaCl to deplete leucine. Cells were subsequently resuspended in M9 minimal media supplemented with all 20 amino acids, 19 amino acids minus leucine and plus 1mM trifluoroleucine, or 19 amino acids minus leucine plus 500 μM homoisoleucine. 6xHis-tagged, CAT or L2A1 expression was induced by the addition of 1 mM

IPTG and carried out for 12 hr. Cells were pelleted at 8000g at 4°C for 10 min. and frozen at -80C. Cell pellets were lysed using 0.5 mg/ml lysozyme (Sigma) in 50 mM Tris pH 7.8 supplemented with 10% glycerol, 0.5 units of DNAseI, and 0.25X B-PERII (Pierce). Cells were treated for a total of 30 sec. sonication using 5 sec. on pulses followed by 10 sec. rest. Lysates were clarified by centrifugation at 10,000g at 4°C for 30 min.. Ni-NTA resin (Qiagen) was added to the clarified lysate and allowed to bind for 2 hr. at 4°C. After binding, the resin was loaded into a chromatography column and washed twice using 50 mM Tris pH 7.8, 10% glycerol, 25 mM imidazole. CAT or L2A1, bearing a N-terminal 6xHis-tag, was eluted using 50mM Tris pH7.8, 10% glycerol, 250 mM imidazole. Samples from all steps were analyzed by SDS-PAGE to confirm purity and eluted protein concentrations were verified using BCA assay.

Measuring residual activity of CAT variants after incubation at elevated temperatures

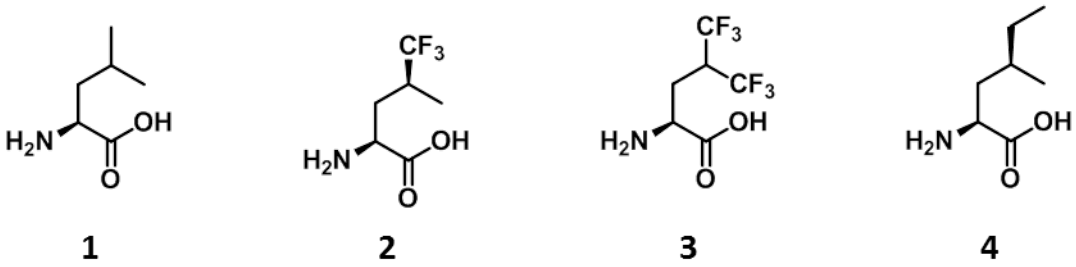
Purified 6xHis-tagged CAT or L2A1 samples containing leucine, trifluoroleucine, or homoisoleucine were incubated with 0.25 mg/ml bovine serum albumin at various temperatures using a thermocycler ranging from 30°C to 75°C for 30 min. followed by 5 min. incubation at 30°C. A 2x reaction buffer comprised of 50 mM Tris pH 7.8, 2 mM DTNB, 0.8 mM acetyl-Coenzyme A was added to aliquots of CAT or L2A1. The activity was assayed at 25°C after the addition of 100 μM chloramphenicol.

Half-life determination of CAT variants at 60°C

Purified 6xHis-tagged CAT or L2A1 containing leucine, trifluoroleucine, or homoisoleucine were incubated with 0.25 mg/ml bovine serum albumin at 60°C for various times ranging from 0 to 120 min. using a thermocycler followed by a 10 min. incubation at 25°C. A 2x reaction buffer

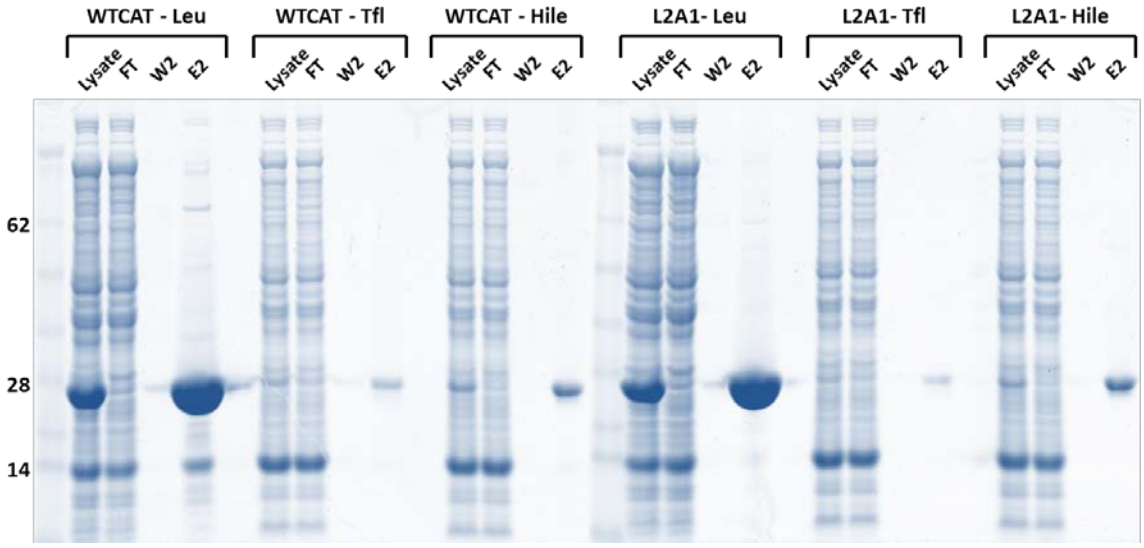
comprised of 50 mM Tris pH 7.8, 2 mM DTNB, 0.8 mM acetyl-Coenzyme A was added to aliquots of CAT or L2A1. The activity was assayed at 25^{o}C after the addition of 100 μM chloramphenicol.

Figure 3.1



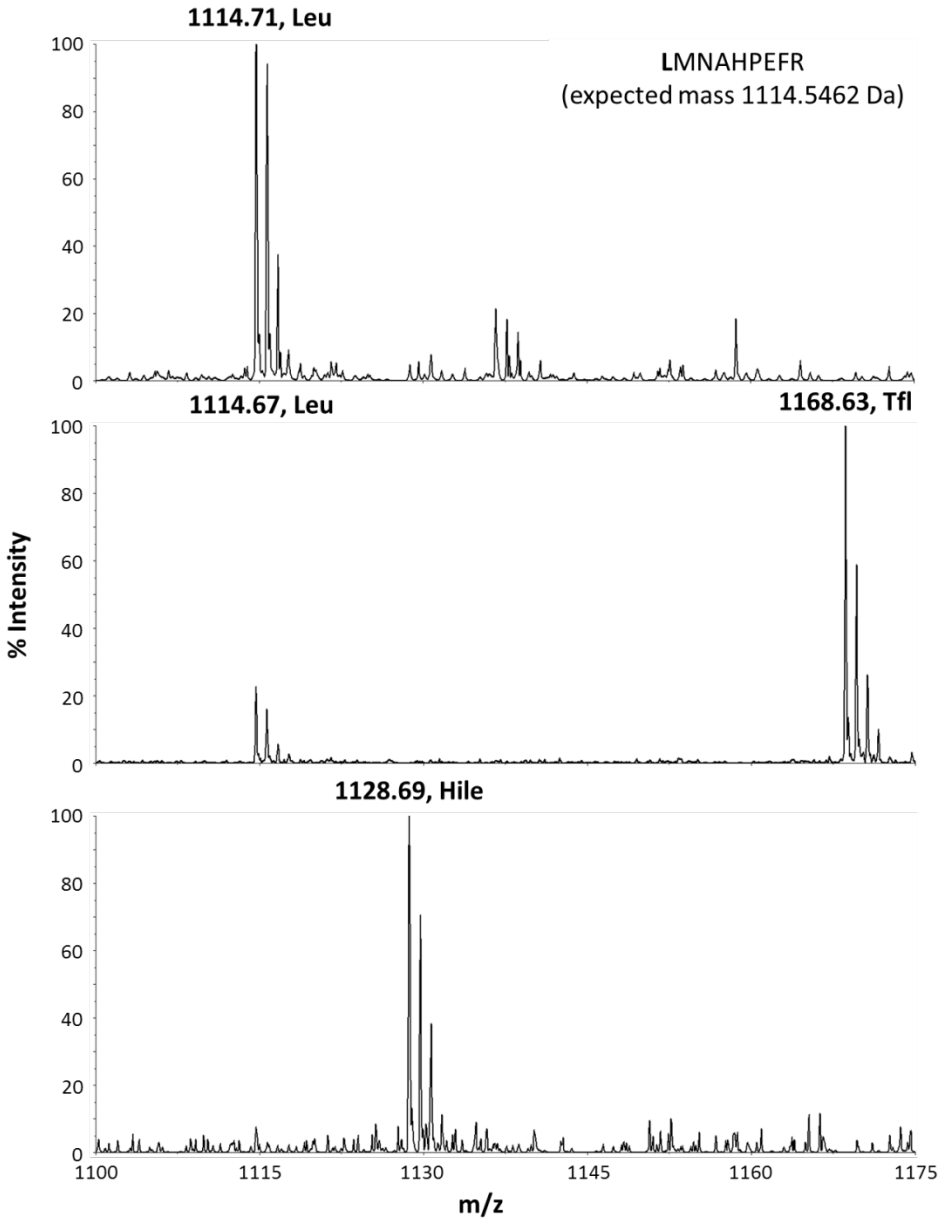
Structures of amino acids discussed in this chapter: 1: Leucine; **2**: Trifluoroleucine; **3**: Hexafluoroleucine; **4**: Homoisoleucine

Figure 3.2



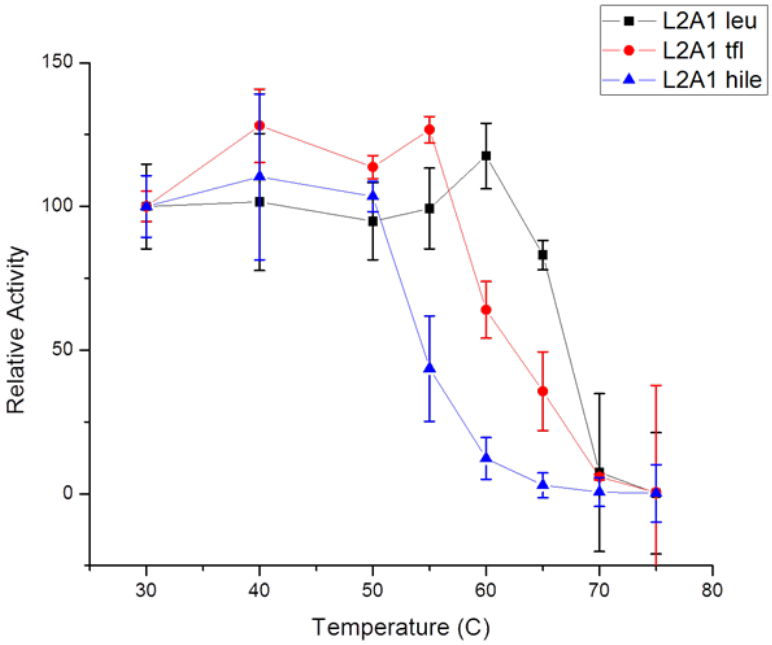
Native purification of his-tagged wild-type CAT and L2A1 expressed in media containing leucine, trifluoroleucine, or homoisoleucine. Wild-type CAT and an evolved version of CAT, L2A1, were expressed in leucine auxotrophic DH10B with overexpression of the wild-type LeuRS. Protein bands corresponding to 6xHis-CAT (expected mass: 27.417 kDa) or 6xHis-L2A1 (expected mass: 27.429 kDa) are observed in lysate lanes at the correct molecular weights. Protein is successfully loaded onto the column (absence of CAT band in lane FT) and minimal protein loss is observed with imidazole washes (lane W2). Protein eluted from the column is concentrated from the original sample and good yields are obtained for CAT or L2A1 expressed in leucine rich media. Approximately eight-fold reduction in protein yield is observed for expression carried out in homoisoleucine compared to leucine. Homoisoleucine media supports two-fold greater protein yields compared to trifluoroleucine media.

Figure 3.3

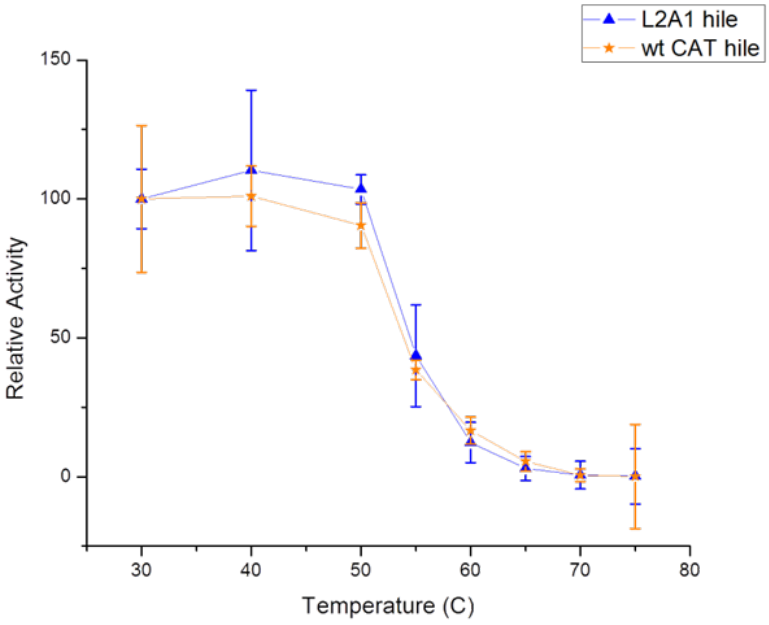


L2A1 expression with leucine, trifluoroleucine, and homoisoleucine demonstrates limits of replacement. Expression of L2A1 is carried out in a leucine auxotrophic strain with constitutively expressed wild-type LeuRS. Purified L2A1 was subjected to trypsin digestion and peptides were analyzed on MALDI-TOF MS. The peptide LMNAHPEFR was found at the expected mass when all 20 amino acids were added to the expression media (top panel). Trifluoroleucine is found to replace leucine at approximately 80%, which is consistent with previous work[30] (middle panel). Quantitative replacement of leucine by homoisoleucine is observed; complete shift of the peptide mass peak by a 14 Da increase (bottom panel).

Figure 3.4

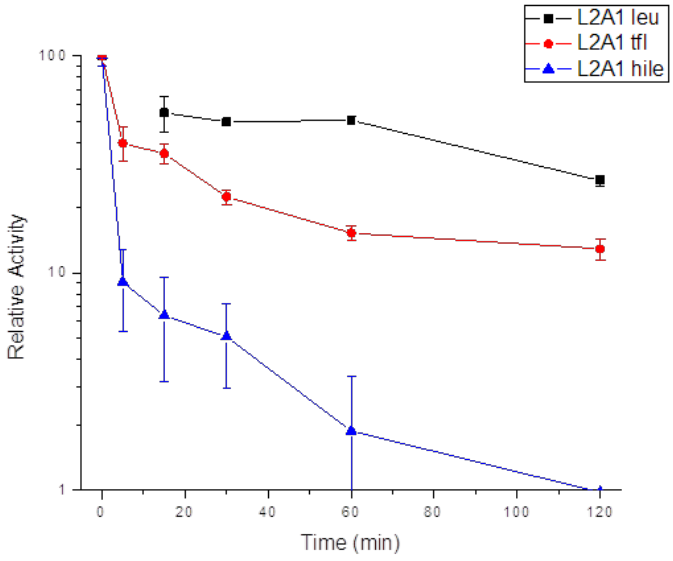


	Melting Temp
L2A1 – Leucine	66.8°C
L2A1 – Trifluoroleucine	61.7°C
L2A1 – Homoisoleucine	54.5°C
Wild type CAT – Homoisoleucine	54.3°C

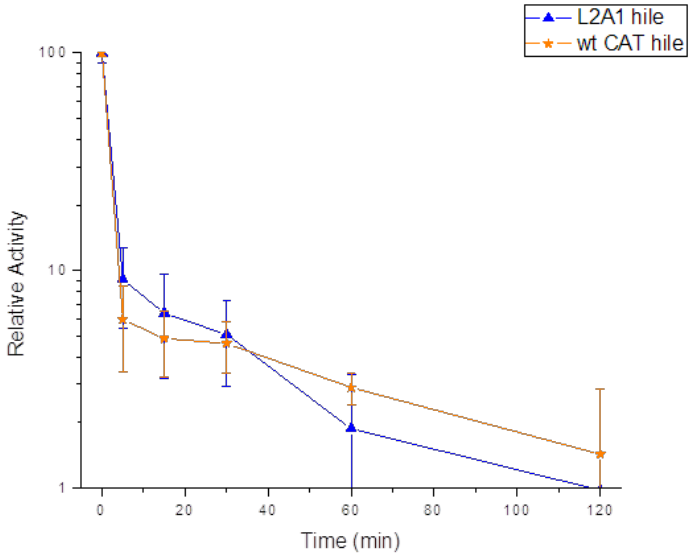


Thermostability of L2A1 is reduced by incorporation of trifluoroleucine and homoisoleucine. Purified L2A1 expressed with leucine, trifluoroleucine, or homoisoleucine was assayed for activity after 30 min. incubation at elevated temperatures (top panel). The leucine version of L2A1 exhibits the greatest thermostability followed by the trifluoroleucine version (top panel); L2A1 was evolved for increased thermostability with leucine positions decoded as trifluoroleucine. Mutations that stabilize L2A1 for trifluoroleucine do not provide any significant benefit when homoisoleucine is quantitatively incorporated (bottom panel).

Figure 3.5



	Half-life at 60°C
L2A1 – Leucine	42.9 min
L2A1 – Trifluoroleucine	7.5 min
L2A1 – Homoisoleucine	1.5 min
Wild type CAT – Homoisoleucine	1.2 min



Rapid inactivation of the homoisoleucine form of L2A1 is observed upon exposure to elevated temperatures. Purified L2A1 expressed with leucine, trifluoroleucine, or homoisoleucine was incubated at 60°C; aliquots were removed over a period of two hours and assayed for activity. The leucine form of L2A1 exhibits approximately 6-fold and 30-fold greater half-life at 60°C than trifluoroleucine and homoisoleucine versions of L2A1 respectively (top panel). Wild type CAT and L2A1 exhibit similar inactivation rates when homoisoleucine is quantitatively incorporated (bottom panel).

References

1. Gershenson, A.; Arnold, F. H. *Genet Eng (N Y)* **2000**, *22*, 55.
2. Arnold, F. H.; Giver, L.; Gershenson, A.; Zhao, H. M.; Miyazaki, K. *Molecular Strategies in Biological Evolution* **1999**, *870*, 400.
3. Giver, L.; Gershenson, A.; Freskgard, P. O.; Arnold, F. H. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12809.
4. Komor, R. S.; Romero, P. A.; Xie, C. B.; Arnold, F. H. *Protein Eng. Des. Sel.* **2012**, *25*, 827.
5. Heinzelman, P.; Snow, C. D.; Smith, M. A.; Yu, X. L.; Kannan, A.; Boulware, K.; Villalobos, A.; Govindarajan, S.; Minshull, J.; Arnold, F. H. *J. Biol. Chem.* **2009**, *284*, 26229.
6. Heinzelman, P.; Snow, C. D.; Wu, I.; Nguyen, C.; Villalobos, A.; Govindarajan, S.; Minshull, J.; Arnold, F. H. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5610.
7. Atsumi, S.; Higashide, W.; Liao, J. C. *Nat. Biotech.* **2009**, *27*, 1177.
8. Zhang, K.; Sawaya, M. R.; Eisenberg, D. S.; Liao, J. C. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 20653.
9. Gao, X.; Xie, X.; Pashkov, I.; Sawaya, M. R.; Laidman, J.; Zhang, W.; Cacho, R.; Yeates, T. O.; Tang, Y. *Chem. Biol.* **2009**, *16*, 1064.
10. Meinhold, P.; Peters, M. W.; Chen, M. M. Y.; Takahashi, K.; Arnold, F. H. *ChemBioChem* **2005**, *6*, 1765.
11. Fasan, R.; Chen, M. M.; Crook, N. C.; Arnold, F. H. *Angew. Chem., Int. Ed.* **2007**, *46*, 8414.
12. Meinhold, P.; Peters, M. W.; Hartwick, A.; Hernandez, A. R.; Arnold, F. H. *Adv. Synth. Catal.* **2006**, *348*, 763.
13. Varadarajan, N.; Gam, J.; Olsen, M. J.; Georgiou, G.; Iverson, B. L. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 6855.

14. Varadarajan, N.; Rodriguez, S.; Hwang, B. Y.; Georgiou, G.; Iverson, B. L. *Nat. Chem. Biol.* **2008**, *4*, 290.

15. Cirino, P. C.; Tang, Y.; Takahashi, K.; Tirrell, D. A.; Arnold, F. H. *Biotechnol. Bioeng.* **2003**, *83*, 729.

16. Rennert, O. M.; Anker, H. S. *Biochemistry* **1963**, *2*, 471.

17. Fenster, E. D.; Anker, H. S. *Biochemistry* **1969**, *8*, 269.

18. Marsh, E. N. G. *Chem. Biol.* **2000**, *7*, R153.

19. Yoder, N. C.; Kumar, K. *Chem. Soc. Rev.* **2002**, *31*, 335.

20. Lee, K. H.; Lee, H. Y.; Slutsky, M. M.; Anderson, J. T.; Marsh, E. N. G. *Biochemistry* **2004**, *43*, 16277.

21. Budisa, N.; Pipitone, O.; Siwanowicz, I.; Rubini, M.; Pal, P. P.; Holak, T. A.; Gelmi, M. L. *Chem. Biodivers.* **2004**, *1*, 1465.

22. Wang, P.; Tang, Y.; Tirrell, D. A. *JACS* **2003**, *125*, 6900.

23. Son, S.; Tanrikulu, I. C.; Tirrell, D. A. *ChemBioChem* **2006**, *7*, 1251.

24. Wang, P.; Fichera, A.; Kumar, K.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2004**, *43*, 3664.

25. Tang, Y.; Tirrell, D. A. *JACS* **2001**, *123*, 11089.

26. Tang, Y.; Ghirlanda, G.; Vaidehi, N.; Kua, J.; Mainz, D. T.; Goddard, W. A.; DeGrado, W. F.; Tirrell, D. A. *Biochemistry* **2001**, *40*, 2790.

27. Yoo, T. H.; Link, A. J.; Tirrell, D. A. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13887.

28. Van Deventer, J. A.; Fisk, J. D.; Tirrell, D. A. *ChemBioChem* **2011**, *12*, 700.

29. Montclare, J. K.; Son, S.; Clark, G. A.; Kumar, K.; Tirrell, D. A. *ChemBioChem* **2009**, *10*, 84.

30. Montclare, J. K.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2006**, *45*, 4518.

31. Bloom, J. D.; Arnold, F. H. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 9995.

32. Bloom, J. D.; Labthavikul, S. T.; Otey, C. R.; Arnold, F. H. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 5869.

33. Bloom, J. D.; Wilke, C. O.; Arnold, F. H.; Adami, C. *Biophys. J.* **2004**, *86*, 2758.

34. Kwon, O.-H.; Yoo, T. H.; Othon, C. M.; Van Deventer, J. A.; Tirrell, D. A.; Zewail, A. H. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 17101.

CHAPTER 4

Introduction of New Chemical Functionality into Engineered Proteins through Incorporation of Novel Non-Canonical Amino Acids

Azidomethylphenylalanine and aminomethylphenylalanine work was done in collaboration with Professor Adam Urbach and Leigh Anna Logsdon from Trinity University (San Antonio, TX).

Furanylalanine work was done in collaboration with Dr. Janek Szychowski.

Abstract

There still exists a great disparity between the chemical reactions available to synthetic chemists and those accessible to protein chemists. Recently, cross metathesis and Diels-Alder chemistry were made accessible through site-specific incorporation of artificial amino acid analogues bearing crotyl oxides and norbornene functional groups. Here we demonstrate the incorporation of three new artificial amino acids that provide access to new chemistries not previously explored via residue-specific replacement. S-allyl-homocysteine, a methionine analogue, and furanylalanine, a phenylalanine analogue, can be incorporated into auxotrophic strains without any genetic modification of the host organism. Azidomethylphenylalanine requires the expression of a mutant phenylalanine synthetase for metabolic labeling of proteins.

Introduction

Artificial protein sequence space contains a broad range of functional groups, many of which add to the repertoire of chemistries available for protein modification. In order to expand sequence space to include novel chemistries, we describe the use of synthetic chemistry to create three new non-canonical amino acids for residue-specific incorporation: S-allyl-homocysteine, furanylalanine, and azidomethylphenylalanine. This chapter describes their potential applications and the work required to utilize these new amino acids for metabolic labeling of proteins.

S-allyl-homocysteine

Olefin metathesis has provided chemists the ability to form new carbon-carbon bonds, enabling synthesis of polymers with low polydispersity from a wide range of monomers containing linear and cyclic alkenes[1-4]. Solid-phase synthesis has enabled chemists to incorporate alkene side chains into peptides; peptides stapled using cross metathesis have been designed to act as inhibitors of transcription factor assembly *in vivo*[5]. The typical solvents (nonpolar alkanes, benzene, etc) and reaction conditions used to perform cross metathesis are not favorable for protein solubility or stability; metathesis has long been considered undesirable as a means of functionalizing whole proteins. Recently, the Davis lab has demonstrated that aqueous cross metathesis is possible given the proper protein side chain[6]. Allyl sulfides were shown to be the ideal choice for cross metathesis using ruthenium-based Grubbs catalysts.

Although allylic amino acid analogues have been incorporated in a residue specific manner into proteins[7], allyl sulfides have only been incorporated through amino acid modifications and site specific incorporation. Chemical modification of cysteine on the protein surface enables a route to install S-allyl-cysteine (Figure 4.1 [**2**]); however, careful choice of protein or engineering of proteins are necessary to prevent multiple modifications or disruption of disulfide bonds[8]. The Schultz lab has demonstrated incorporation via amber suppression of O-crotylserine (Figure 4.1 [**3**]) with oxygen at the allylic position; LC-ESI-MS was used to detect the cross metathesis product between two spatially adjacent O-crotylserine residues in circularly permutated Venus protein[9]. Attempts to use S-allyl-cysteine as a methionine-specific analogue in methionine-auxotrophic *E. coli* produced low but detectable incorporation[6].

In order to utilize cross metathesis for BONCAT, we explored other allyl sulfides that could be incorporated in a residue specific manner. The inherent lack of carbon-carbon double bonds in proteins could enable selective modification using metathesis if a suitable alkene could be incorporated into proteins. We surmised that an allyl sulfide with similar heteroatom position to methionine (Figure 4.1 [**1**]) could be incorporated by the methionyl-tRNA synthetase (MetRS) or one of the known MetRS mutants. We demonstrate here the synthesis of S-allyl-homocysteine (Figure 4.1 [**4**]), and we show that S-allyl-homocysteine can be used as a methionine analogue for global incorporation into proteins at near quantitative levels.

Azidomethylphenylalanine

BONCAT has provided researchers with the ability to selectively enrich newly synthesized proteins from complex mixtures for analysis via mass spectroscopy[10]. Through the incorporation of methionine analogues bearing azide or alkyne groups, proteins can be modified with [3+2] azide-alkyne click chemistry, enabling the attachment of affinity probes for enrichment. Azidonorleucine and the mutant methionyl-tRNA synthetase NLL-MetRS have further refined BONCAT to confine labeling to a subset of organisms, enabling identification of proteins in complex cellular mixtures[11,12].

Affinity purification in BONCAT is currently mediated through biotin-streptavidin interactions; attempts to remove newly synthesized proteins during enrichment have resulted in sample contamination, with streptavidin loss from its resin support due to the harsh conditions required to separate biotin from streptavidin. Attempts to mitigate contamination have come through the development of affinity tags containing cleavable moieties that should not damage the resin-supported streptavidin[13]. We sought to explore non-protein based affinity methods that could eliminate complications arising from excess streptavidin in samples prepared for mass spectrometry.

Supramolecular chemistry provides many examples of host-guest interactions that could prove to be useful replacements for the biotin-streptavidin system. We turned to a class of water-soluble macrocycles known as cucurbiturils: annular molecules formed by glycoluril

monomers linked by methylene units. In particular cucurbit[7]uril (**Q7**, Figure 4.2), composed of 7 glycoluril monomers, has been demonstrated to bind a wide variety of targets including organic dyes, organometallics, drugs, and proteins[14]. Phenylalanine (Figure 4.1 [**5**]) has previously been shown to bind with micromolar affinity to Q7; its aromatic ring fits into the non-polar cavity of Q7 while the amino acid backbone forms favorable electrostatic interactions with the carbonyl groups decorating the circular face of Q7[15]. The Urbach group has demonstrated that non-canonical phenylalanine analogs can bind with even greater affinity to Q7; furthermore, Q7 can recognize peptides with N-terminal phenylalanine or phenylalanine analogs[14]. Peptides that bear *p*-aminomethylphenylalanine (Figure 4.1 [**8**]) at the N-terminus bind Q7 with K_D of 9.5×10^{-10} M[16].

The Tirrell lab has demonstrated the incorporation of a wide variety of phenylalanine analogues through the use of mutant phenylalanine-tRNA synthetases (PheRS). Positions A294 and T251 in the binding pocket of *E. coli* PheRS can be mutated to glycine to enable the incorporation of a wide variety of bulkier phenylalanine analogues[17]. Here we describe the synthesis and incorporation of *p*-azidomethylphenylalanine (Figure 4.1 [**7**]), a chemical precursor to *p*-aminomethylphenylalanine, into proteins.

Furanylalanine

The Diels-Alder reaction utilizes an electron-rich diene and electron-poor alkene, both of which are absent in the canonical protein side chains, to form new carbon-carbon bonds

through a [4+2] cycloaddition. Through the modification of cysteine or lysine, chemists have installed a variety of dienes and dieneophiles onto proteins for subsequent modification through Diels-Alder reactions[18]. The Ting group has recently demonstrated the incorporation of trans-cyclooctene functionality onto proteins mediated by an engineered lipoic acid ligase; an inverse-electron-demand Diels-Alder cycloaddition with a tetrazine probe that enables the attachment of fluorophores to specific proteins[19]. Recent work by the Chin Lab and others has demonstrated the installation of cyclic alkene amino acids like Nε-5-norbornene-2-yloxy-carbonyl-L-lysine, through amber suppression, for inverse electron-demand Diels–Alder cycloaddition[20,21].

To incorporate a Diels-Alder reactive amino acid in a residue specific manner, we investigated an amino acid containing the furan functional group. Among the earliest studied compounds for the Diels-Alder reaction[22], furans have a large assortment of possible Diels-Alder partners including maleimides, activated alkenes, and alkynes[23]. Although furans are considered less reactive than other dienes because of their aromaticity, performing the reaction in aqueous conditions should help to improve reaction kinetics[24,25]. The furan-maleimide reaction has been utilized extensively for polymer synthesis[26-29], taking advantage of the reversible nature of the Diels-Alder reaction[30-32]. Here we describe the incorporation of furanylalanine into proteins as a phenylalanine analogue. Proteins metabolically labeled with furanylalanine can be modified through the attachment of maleimide probes.

Results and Discussion

S-allyl-homocysteine

MetRS has been known to activate a variety of methionine analogues, including those that share methionine's sulfur position[33]. Homocysteine, a precursor to methionine, can be activated by MetRS; however, before homocysteine is transferred to $tRNA^{MET}$, it is rapidly cyclized to homocysteine thiolactone[34]. Ethionine, a methionine analogue with an ethyl group in lieu of the canonical methyl group, is also activated by the MetRS but is deacylated from $tRNA^{MET}$ [35]. We synthesized S-allyl-homocysteine (Figure 4.3), an allyl sulfide containing methionine analogue that shares the canonical position of sulfur, for test expressions in methionine auxotrophs.

Expression of GFPrmAM in media containing 19 amino acids without methionine is made possible by supplementation of the media with 1mM S-allyl-homocysteine. Replacement of methionine by S-allyl-homocysteine was determined from purified GFPrmAM to be approximately 40% by MALDI-TOF MS (Figure 4.4). Previous work by Kiick et al. had demonstrated that poorly activated methionine analogues could be incorporated with overexpression of the wild-type MetRS[36]. By including a constitutively expressed copy of the MetRS in the GFPrmAM expression plasmid, near complete replacement of methionine by S-allyl-homocysteine was observed (Figure 4.4).

Azidomethylphenylalanine

The side chain amine group of *p*-aminomethylphenylalanine, a high affinity guest for Q7, is protonated under physiological conditions, making it an unlikely substrate for activation by PheRS. The Staudinger reduction enables the conversion of azides to amines through treatment with phosphine[37]; *p*-azidomethylphenylalanine could be incorporated in the place of *p*-aminomethylphenylalanine and subsequently reduced to yield the desired amino acid. Precedent for incorporating azides at the para position of phenylalanine exists; *p*-azidophenylalanine (Figure 4.1 [**6**]) has been incorporated into proteins by the mutant PheRS-A294G to crosslink artificial protein biomaterials [17,38]. *p*-Azidomethylphenyalanine was synthesized via diazotransfer from commercially available *p*-aminomethylphenylalanine to test for activation by PheRS mutants.

Incorporation of azide analogues into proteins can be detected by treatment of cellular lysates with fluorescent alkyne probes (Figure 4.5). Overexpression of wild-type PheRS or any of the PheRS mutants in the canonical 20 amino acids does not permit the attachment of dibenyzlcyclooctyne conjugated carboxytetramethylrhodamine (DIBO-TAMRA, **11**) to cellular proteins. Metabolic labeling of proteins with *p*-azidophenylalanine or *p*-azidomethylphenylalanine is not observed in cases where the wild-type PheRS is overexpressed. Lysates from cultures with overexpressed PheRS mutants display DIBO-TAMRA modified proteins if *p*-azidomethylphenylalanine is added post-media shift. Similarly, if *p*-azidophenylalanine, a known analogue for PheRS-A294G, PheRS-T251G,

and PheRS-A294G&T251G, is added post media shift, lysates from cultures with overexpressed PheRS mutants are modified by DIBO-TAMRA.

Furanylalanine

3-Thienylalanine (**9**), a thiophene analogue of phenylalanine, has been previously incorporated into proteins by Kothakota et al[39]. We rationalized that since 3-furanylalanine (**10**) and 3-thienylalanine are structurally very similar, furanylalanine would be amenable for residue specific replacement of phenylalanine. Furthermore, if furanylalanine could not be incorporated by the wild-type PheRS, many well characterized mutant PheRSs (PheRS-A294G, PheRS-T251G, PheRS-A294G&T251G) could serve as means to incorporate furanylalanine, albeit in a cell-specific manner.

Incorporation of furanylalanine was confirmed through the overexpression of wild-type PheRS under the control of an IPTG-inducible promoter in a phenylalanine auxotrophic strain (Figure 4.6). Levels of furanylalanine incorporation in PheRS are difficult to quantify since the probability of replacement is dependent on the quantity of induced wild-type PheRS expression. Assuming wild-type PheRS is not rendered inactive by furanylalanine incorporation, PheRS synthesized at the end of the expression window should have higher incorporation levels than PheRS produced shortly after induction. However, it is worthwhile to note that trypsin digested peptides from wild-type PheRS are observed to have multiple phenylalanine positions replaced with furanylalanine via MALDI-TOF MS. Expression of green fluorescent protein in non-phenylalanine auxotrophic strains without the overexpression of the wild-type PheRS yielded

furanylalanine replacement of approximately 10 percent (Figure 4.7). Metabolic labeling of proteins with furanylalanine can occur with simple media shift, making it a potential amino acid analogue for BONCAT if a bio-orthogonal Diels-Alder reaction could be found.

The Diels-Alder reaction between maleimides and furans yields [4π+2π] cycloaddition products at room temperature, making it an attractive means of attaching molecules to proteins labeled with furanylalanine[18]. However, maleimides cannot be considered bio-orthogonal due to the competing Michael addition with cysteines, and to a lesser extent, amines. Attachment of a biotinylated maleimide tag (**12**) to purified wild-type PheRS was carried out following a reduction and iodoacetamide alkylation step (Figure 4.8). The alpha subunit of wild-type PheRS labeled with furanylalanine is observed to have a distribution of increased molecular weights, due to the attachment of **12** at multiple furanylalanine positions. Visualizing biotinylated proteins with streptavidin demonstrates the lack of bio-orthogonality of the maleimide-furan reaction; wild-type PheRS expressed without furanylalanine is nonspecifically biotinylated, as seen in middle panel of Figure 4.8. The alpha subunit of PheRS has a single cysteine codon compared to 20 phenylalanine codons; the lack of bio-orthogonality could hamper the use of furanylalanine for BONCAT where cysteine content of proteins is unknown.

Future Directions

Although S-allyl-homocysteine can be incorporated extensively at methionine positions, attempts to perform cross-metathesis to attach alkene probes on labeled proteins have thus far been unsuccessful. A combination of factors could be hindering the intermolecular reaction. Proteins and catalyst have different preferences for the liquid phases; catalysts that have increased water solubility would assist in improving metathesis for conjugation of probes to proteins. Intramolecular reactions of multiple S-allyl-homocysteine residues on a single protein or intermolecular reactions of proteins would result in undesired side products. Methionine replacement could be lowered using less S-allyl-homocysteine, thus helping to reduce intramolecular reactions. However, controlling intermolecular reactions could be difficult if the intent is to use S-allyl-homocysteine for BONCAT, unless reactions were run in extremely dilute conditions. There are conditions where intermolecular reactions between proteins could be favorable, such as the formation of protein gel networks. Additionally, thiol-ene chemistry could utilize the allyl-sulfide side chain of S-allyl-homocysteine.

Ongoing work with aminomethylphenylalanine has provided additional motivation for improving the incorporation of azidomethylphenylalanine. Recent developments in the Urbach group have shown Q7 is capable of interfering with exopeptidase activity (Figure 4.9).[40] Peptides with internal aminomethylphenylalanine residues can be treated with non-specific aminopeptidase; when N-terminal aminomethylphenylalanine is generated, Q7 binds and inhibits further exopeptidase activity. By combining proteolysis and affinity

purification in one step without the need for streptavidin, this could provide a means to simplify enrichment of newly synthesized proteins labeled with aminomethylphenylalanine. Q7 is currently being modified to enable attachment to a solid support, enabling flow purification in a column format.

Connor et al. have demonstrated another way to generate proteins with N-terminal phenylalanine analogues[41]. L,F-transferase performs the post-translational addition of leucine or phenylalanine to proteins bearing N-terminal basic residues; L,F-transferase plays a vital role in the protein degradation pathway known as the N-end rule[42]. Connor et al. showed that L,F-transferase has flexible substrate specificity; for example, it is capable of utilizing azidophenylalanine-$tRNA^{Phe}$ aminoacylated by PheRS-A294G. Extension of the L,F-transferase system would be straightforward with azidomethylphenylalanine, allowing for the generation of full length proteins as potential guests for the Q7.

Saturation mutagenesis of other binding pocket positions of PheRS could be performed to create libraries for screening improved PheRS mutants for azidomethylphenylalanine incorporation. Fluorescence activated cell sorting has already been demonstrated for screening libraries of aminoacyl-tRNA synthetases for improved activation of azide bearing amino acids[43]; a high throughput screening method could enable faster discovery of PheRS mutants specific for azidomethylphenylalanine.

Further investigations of Diels-Alder reactions specific for furanylalanine are currently on-going. Furan-maleimide chemistry provides interesting applications for the formation of hydrogels with cysteine-free proteins, including those that could be thermally sensitive to

retro Diels-Alder. However, the ideal reaction would be bio-orthogonal to enable greater flexibility for protein modification and use in the discovery of newly synthesized proteins.

Here we have demonstrated three new artificial amino acids that can be used for metabolic labeling of proteins. All three amino acids highlight new avenues for protein chemistry not afforded by the current crop of non-canonical amino acids used for residue specific incorporation. Although binary interactions between organisms can be interrogated with propargylglycine and azidonorleucine, most multi-organism systems are far more complex. The further development of new bio-orthogonal chemistries which are orthogonal to each other could provide new methods of examining complex interspecies interactions.

Materials and Methods

Synthesis of S-allyl-homocysteine

N-Boc-DL-S-allyl-homocysteine methyl ester (Figure 4.3a, **i**) was synthesized according to previously reported synthesis from the Davis Lab[6]. De-protection of **i**: 520 mg of **i** was dissolved in 10 mL of CH_2Cl_2. 1 mL of TFA was added and resulting mixture was stirred at room temperature for 4 hours. Solvent was evaporated under reduced pressure and resulting yellow oil was dissolved in 10 mL of THF. While this solution was stirred over an ice bath, 25 mL of 5 M LiOH was added. The reaction was allowed to warm up to room temperature and stirred for 2 hr.. ddH_2O (35mL) was then added to the reaction mixture followed by DOWEX® 50Wx8 (~50 g) and loaded into an empty column. The resin was washed with 250 mL of H_2O and product was eluted with 5% NH_4OH. Solvent was evaporated under reduced pressure resulting in a yellow solid which was further purified using silica gel chromatography (1-10% gradient of MeOH in CH_2Cl_2) producing S-allyl-homocysteine (**ii**) (220.4 mg, 67%.) **^{1}H NMR**: (500 MHz, D_2O) δ 5.83 – 5.73 (m, 1H), 5.16 – 5.07 (m, 2H), 3.25 (dd, J = 7.3, 5.6 Hz, 1H), 3.17 – 3.13 (m, 2H), 2.51 – 2.46 (m, 2H), 1.89 – 1.79 (m, 1H), 1.78 – 1.68 (m, 1H). **^{13}C NMR:** (126 MHz, D_2O) δ 182.72, 133.86, 117.47, 55.22, 34.26, 33.39, 26.16. **TOF MS ES$^+$:** Expected mass for $C_7H_{14}NO_2S$ is 176.0745; found 176.0709.

Determination of S-allyl-homocysteine incorporation

Expression of his-tagged GFP_rmAM was carried out in methionine auxotrophic DH10B (TYJV2) strain using the plasmids pQE-80L/GFP_rmAM as well as pMTY13. pMTY13 is pQE-80L/GFP_rmAM with a copy of the wild-type MetRS under control of its endogenous promoter[44]. Cultures of TYJV2 containing pQE-80L/GFP_rmAM or pMTY13 were grown at 37^{o}C to mid-log phase (OD_{600} ~ 0.8-1.0) in M9 minimal media (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), and ampicillin (200 μg/L). Upon reaching mid-log phase, cells were pelleted and washed twice using 0.9% NaCl to deplete methionine. Cells were subsequently resuspended in M9 minimal media supplemented with 19 amino acids minus methionine plus 2 mM S-allyl-homocysteine. GFP_rmAM expression was induced by the addition of 1 mM IPTG and carried out for 6 hr.. Cells were pelleted at 8000g and 4^{o}C for 10 min. and frozen at -80^{o}C. Cell pellet was lysed in 8 M urea with repeated freeze-thaw cycles and GFP_rmAM was purified under denaturing conditions using Ni-NTA agarose. Tryptic digests of purified GFP_rmAM were subjected to MALDI time-of-flight mass spectrometry with α-cyano-4-hydroxycinnamic acid as matrix. An Applied Biosystems Voyager DE-PRO equipped with a 20-Hz nitrogen laser was used for all peptide analyses.

Synthesis of L-azidomethylphenylalanine

An ampule of triflic anhydride (10 g, 35.5 mmol) was added dropwise to a prechilled 100 mL round bottomed flask containing sodium azide (10.6 g, 163 mmol) dissolved in 30 mL

deionized water submerged in an ice bath. The resulting mixture was stirred vigorously at room temperature for 90 min. and then extracted with three portions of dichloromethane (20 mL each). The organic layers were combined, washed twice with saturated sodium carbonate (30 mL each), and then added to a flask containing Boc-Phe(4-NH_2CH_2)-OH (1 g, 3.4 mmol), potassium carbonate (0.7 g, 5.1 mmol), and cupric sulfate (0.084 g, 0.34 mmol) in a mixture of 40 mL water and 80 mL methanol. The resulting light blue cloudy mixture was stirred overnight at room temp and then concentrated in vacuo to ~40 mL. The blue mixture was acidified with 6 M HCl to pH ~6, at which point blue clumps appeared. The clumps were removed by vacuum filtration, and the colorless filtrate was further acidified by addition of 6 N HCl. At pH 4 the mixture became cloudy, and at pH 2-3, the mixture was added to a separatory funnel and extracted with three portions of ethyl acetate (20 mL each). The organic layers were combined, dried over sodium sulfate, and concentrated in vacuo to a yellow oil with light yellow solid. This crude material was resuspended in concentrated HCl (4 mL) and mixed for 10 min. to yield a white precipitate. Water (25 mL) was added, and the yellow solution was concentrated to dryness by rotavap under high vacuum. The solid was resuspended in 20 mL water and lyophilized to dryness to yield a light tan, finely divided powder. Yield: 354 mg, 47%. **^{1}H NMR:** (400 MHz, DMSO-d6): 8.38 (br s, 3H), 7.33 (m, 4H), 4.42 (s, 2H), 4.18 (t, 1H, J = 7 Hz), 3.12 (d, J = 7 Hz).

Testing azidomethylphenylalanine incorporation via overexpression of PheRS variants

Expression of 6xHis-tagged, PheRS variants was carried out in KY14 using the plasmids pQE-80L/PheRS, pQE-80L/PheRS-A294G, pQE-80L/PheRS-T251G, and pQE-80L/PheRS-A294G&T251G. PheRS variants were cloned into the SphI and HindIII site in a similar manner to the way in which pQE-80L/PheRS was constructed, allowing for inducible expression of 6xHis-tagged PheRS variants upon the addition of IPTG. Cultures of KY14 containing pQE-80L/PheRS, pQE-80L/PheRS-A294G, pQE-80L/PheRS-T251G, and pQE-80L/PheRS-A294G&T251G were grown at 37^{o}C to mid-log phase (OD_{600} ~ 0.8-1.0) in M9 minimal media (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), and ampicillin (200 μg/L). Upon reaching mid-log phase, cells were pelleted and washed twice using 0.9% NaCl to deplete phenylalanine. Cells were subsequently resuspended in M9 minimal media supplemented with all 20 amino acids, 19 amino acids minus phenylalanine and plus 1 mM L-azidophenylalanine, a known analogue activated by the listed PheRS variants, or 19 amino acids minus phenylalanine and plus 1 mM L-azidomethylphenylalanine. Expression of 6xHis-tagged PheRS variants was induced by the addition of 1 mM IPTG and carried out for 4 hours. Cells were pelleted at 8000g and 4^{o}C for 10 minutes and frozen at -80^{o}C. The cell pellet was lysed in 8 M urea with repeated freeze-thaw cycles and 6xHis-tagged, wild-type PheRS was purified under denaturing conditions using Ni-NTA agarose. Purified 6xHis-tagged PheRS variants expressed in 20 amino acids, L-azidophenylalanine, or L-azidomethylphenylalanine were subjected to strain-promoted, copper-free azide-alkyne click chemistry using

dibenzylcyclooctyne conjugated carboxytetramethylrhodamine (DIBO-TAMRA, Invitrogen). Reaction was carried out in tris-buffered saline pH 7.4 with a 25 μM DIBO-TAMRA for 30 minutes at room temperature in the dark. Aliquots of the reaction mixture were mixed with SDS loading buffer and separated on 12% SDS-PAGE gel. Proteins conjugated with DIBO-TAMRA were visualized using Typhoon Trio with 532 nm laser and 580 nm bandpass filter.

Testing furanylalanine incorporation via overexpression of wild-type PheRS

D,L-furanylalanine was synthesized according to a previously reported procedure[45]. Expression of the 6xHis-tagged, wild-type PheRS was carried out in methionine, phenylalanine, and lysine auxotrophic DH10B (KY14), prepared by Kai Yuet using lambda red recombination knockout, using the plasmid pQE-80L/PheRS. The plasmid pQE-80L/PheRS has the wild-type PheRS cloned into the SphI and HindIII sites, allowing for inducible expression of 6xHis-tagged wild-type PheRS upon addition of IPTG. Cultures of KY14 containing pQE-80L/PheRS were grown at 37^{o}C to mid-log phase (OD_{600} ~ 0.8-1.0) in M9 minimal media (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), and ampicillin (200 μg/L). Upon reaching mid-log phase, cells were pelleted and washed twice using 0.9% NaCl to deplete phenylalanine. Cells were subsequently resuspended in M9 minimal media supplemented with 19 amino acids minus phenylalanine and plus 2 mM D,L-furanylalanine. Expression of 6xHis-tagged, wild-type PheRS was induced by addition of 1 mM IPTG and carried out for 4 hr.. Cells were pelleted at 8000g and 4^{o}C for

10 minutes and frozen at -80C. The cell pellet was lysed in 8 M urea with repeated freeze-thaw cycles and 6xHis-tagged, wild-type PheRS was purified under denaturing conditions using Ni-NTA agarose. Trypsin digests of purified 6xHis-tagged, wild-type PheRS were subjected to MALDI-TOF MS with α-cyano-4-hydroxycinnamic acid as matrix. An Applied Biosystems Voyager DE-PRO equipped with a 20-Hz nitrogen laser was used for all peptide analyses.

Determination of L-furanylalanine incorporation via expression of GFPm

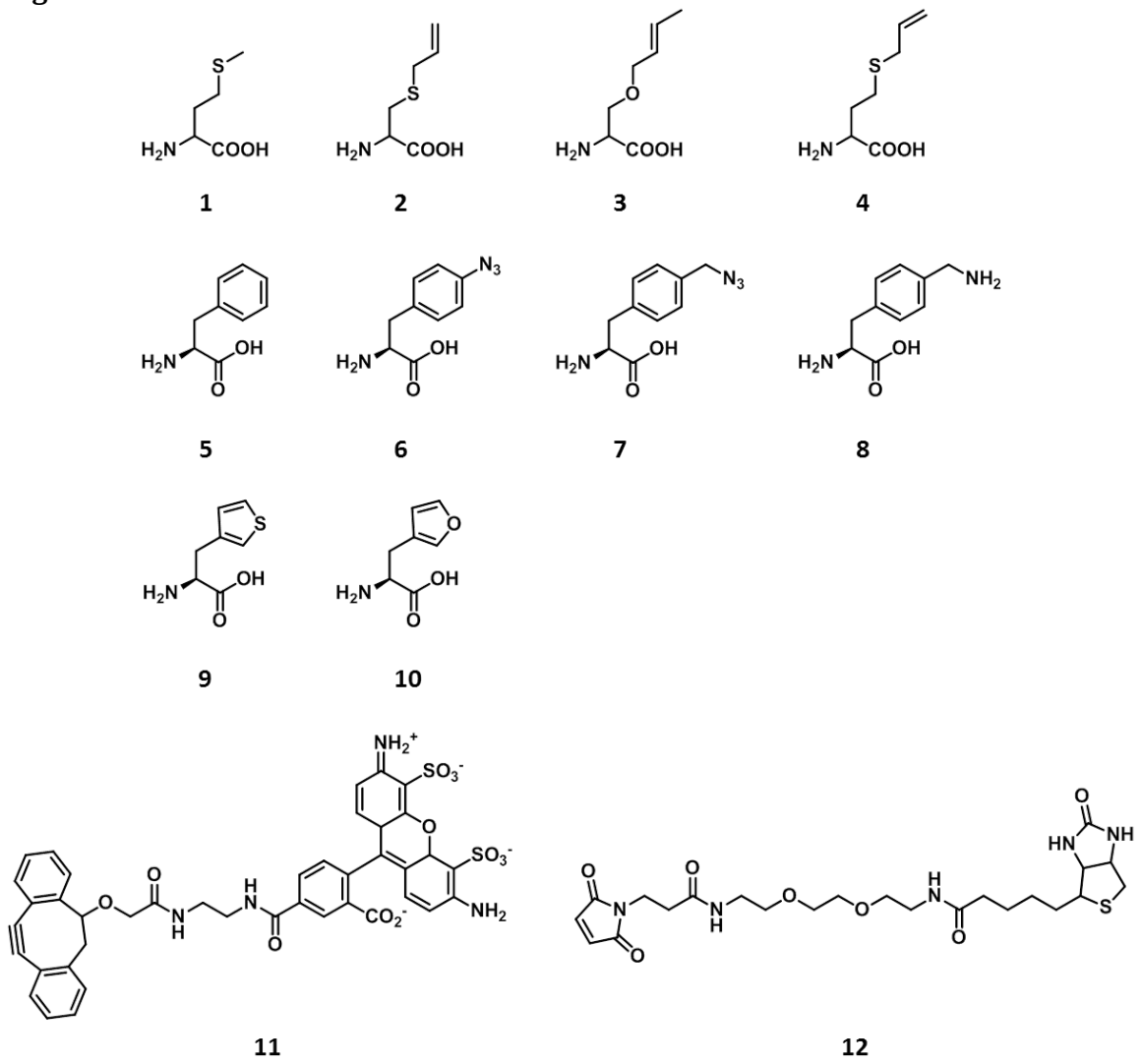
Expression of 6xHis-tagged GFPm was carried out in leucine auxotrophic DH10B (Invitrogen) transformed with pQE-80L/GFPm, which was previously constructed by Yoo et al[46]. Cultures of DH10B containing pQE-80L/GFPm were grown at 37^{o}C to mid-log phase (OD_{600} ~ 0.8-1.0) in M9 minimal media (M9 salts, 0.2% glucose, 1 mM $MgSO_4$, 0.1 mM $CaCl_2$, 35 mg/L Vitamin B) supplemented with all 20 canonical amino acids (each at 40 mg/L), and ampicillin (200 μg/L). Upon reaching mid-log phase, cells were pelleted and washed twice using 0.9% NaCl to deplete phenylalanine. Cells were subsequently resuspended in M9 minimal media supplemented with 19 amino acids minus phenylalanine and plus 2 mM D,L-furanylalanine. 6xHis-tagged GFPm expression was induced by addition of 2 mM IPTG and carried out for 6 hr.. Cells were pelleted at 8000g and 4^{o}C for 10 minutes and frozen at -80C. The cell pellet was lysed in 8 M urea with repeated freeze-thaw cycles and 6xHis-tagged GFPm was purified under denaturing conditions using Ni-NTA agarose. Trypsin digests of purified 6xHis-tagged GFPm were subjected to MALDI-

TOF MS with α-cyano-4-hydroxycinnamic acid as matrix. An Applied Biosystems Voyager DE-PRO equipped with a 20-Hz nitrogen laser was used for all peptide analyses.

Biotinylation of PheRS metabolically labeled with furanylalanine using maleimide Diels-Alder Reaction

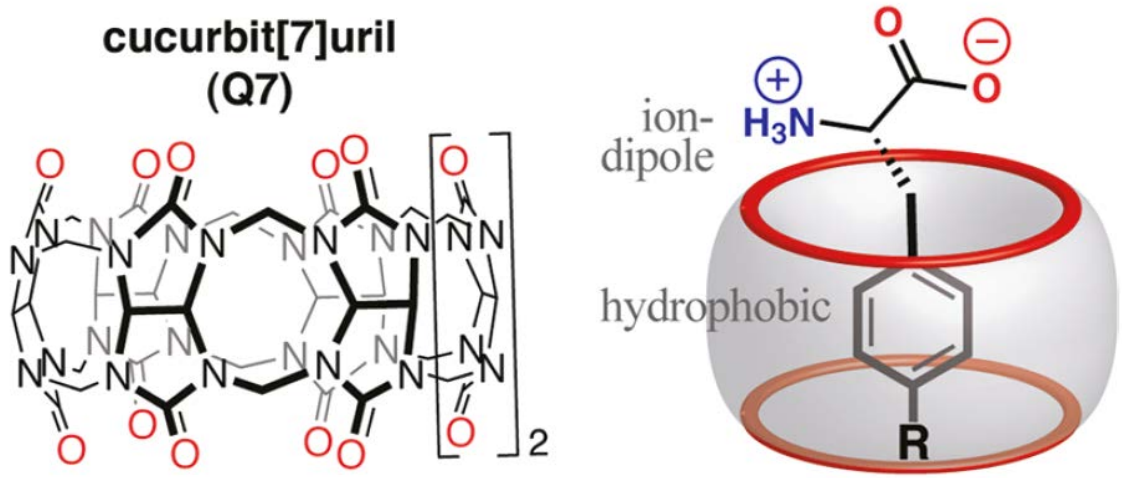
Purified 6xHis-tagged, wild-type PheRS labeled with furanylalanine was treated with 2% β-mercaptoethanol at pH 7.8 and subsequently alkylated using 100 mM iodoacetamide, pH 7.8 in the dark for 2 hr.. Acetone precipitation was performed to remove alkylating reagent and protein was re-solubilized with 1% SDS in PBS, pH 6.5. Maleimide-PEG_2-Biotin (Pierce) was added to a final concentration of 10 mM and reacted overnight at room temperature. An additional acetone precipitation was performed and protein was re-solubilized in 2% SDS loading buffer for visualization via western blot on low fluorescence nitrocellulose membrane (Amersham). Blocking was performed with 5% w/v milk in PBST while washes were performed with PBST three times for 10 minutes. Presence of 6xHis-tagged, wild-type PheRS was visualized with Penta-His-AlexaFluor 647 (1:5000 dilution, Qiagen); biotin was visualized with streptavidin conjugated AlexaFluor 488 (1:2000 dilution, Invitrogen). Visualization of fluorescently labeled antibodies used during western blotting was performed on a Typhoon Trio with appropriate laser and filter combinations.

Figure 4.1



Structures of amino acids and probes discussed in this chapter. 1: Methionine; **2**: S-allyl-cysteine; **3**: O-crotylserine; **4**: S-allyl-homocysteine; **5**: Phenylalanine; **6**: p-azidophenylalanine; **7**: p-azidomethylphenylalanine; **8**: p-aminomethylphenylalanine; **9**: 3-thienylalanine; **10**: 3-furanylalanine; **111**: DIBO-TAMRA; **12**: Maleimide-PEG_2-Biotin

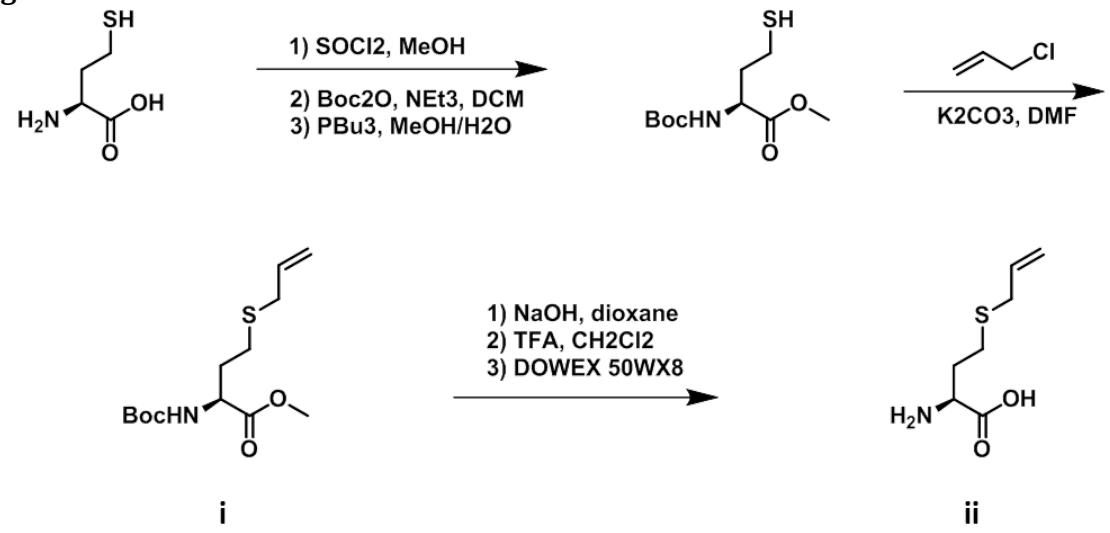
Figure 4.2



Structure of cucurbit[7]uril (Q7) and proposed binding of phenylalanine analogues. R groups at the para position of phenylalanine can range from halogens, alcohols, amines, t-butyl, carboxyl, phosphoryl, and azides. This figure was reproduced from Logsdon et al[16].

Figure 4.3

a)



i ii

b)

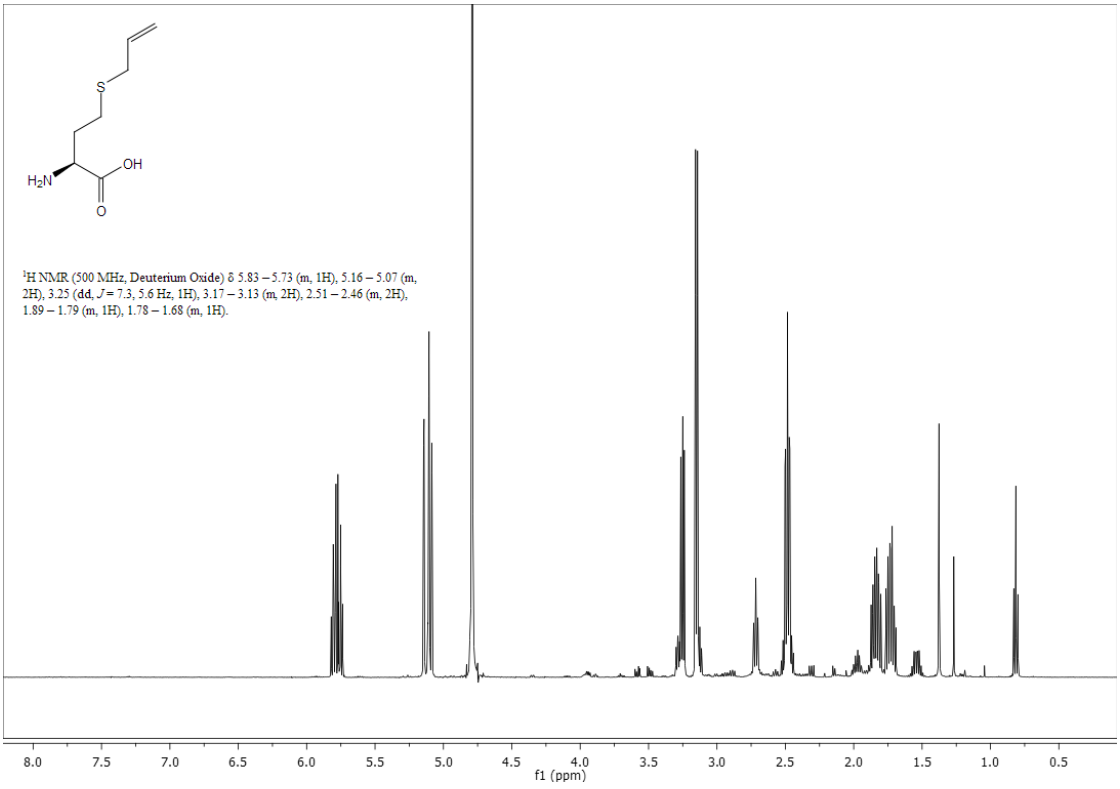
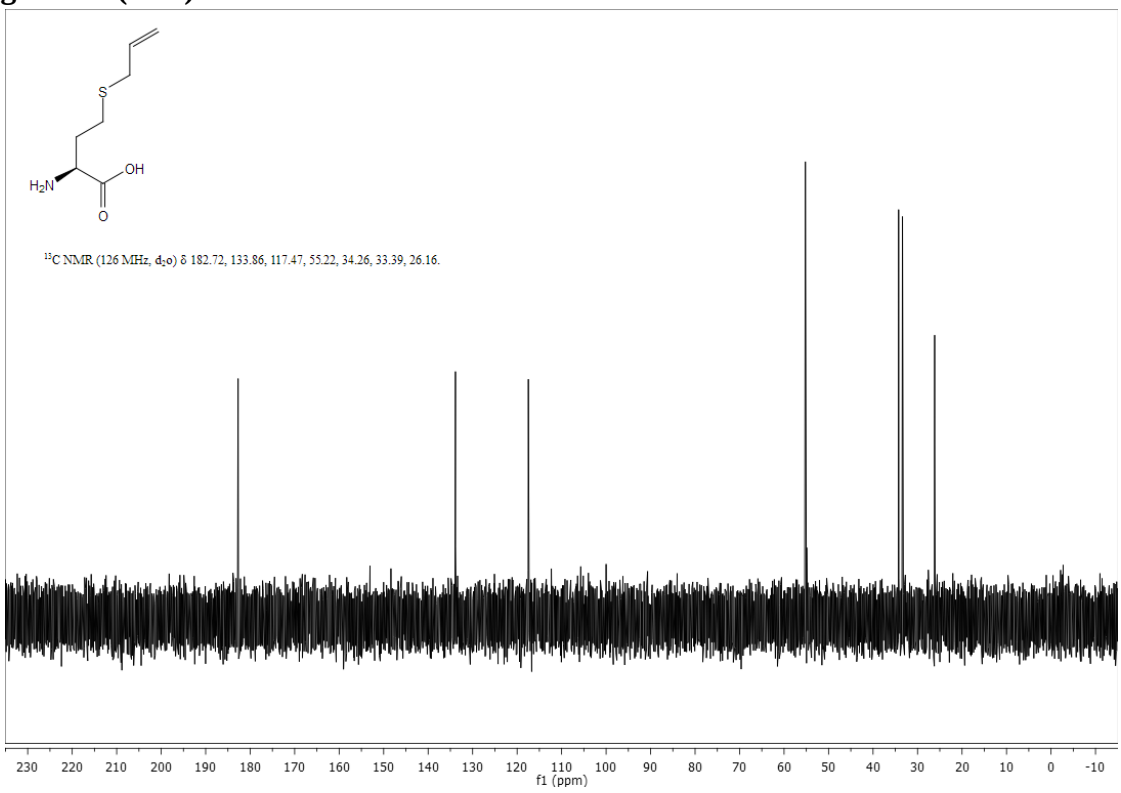


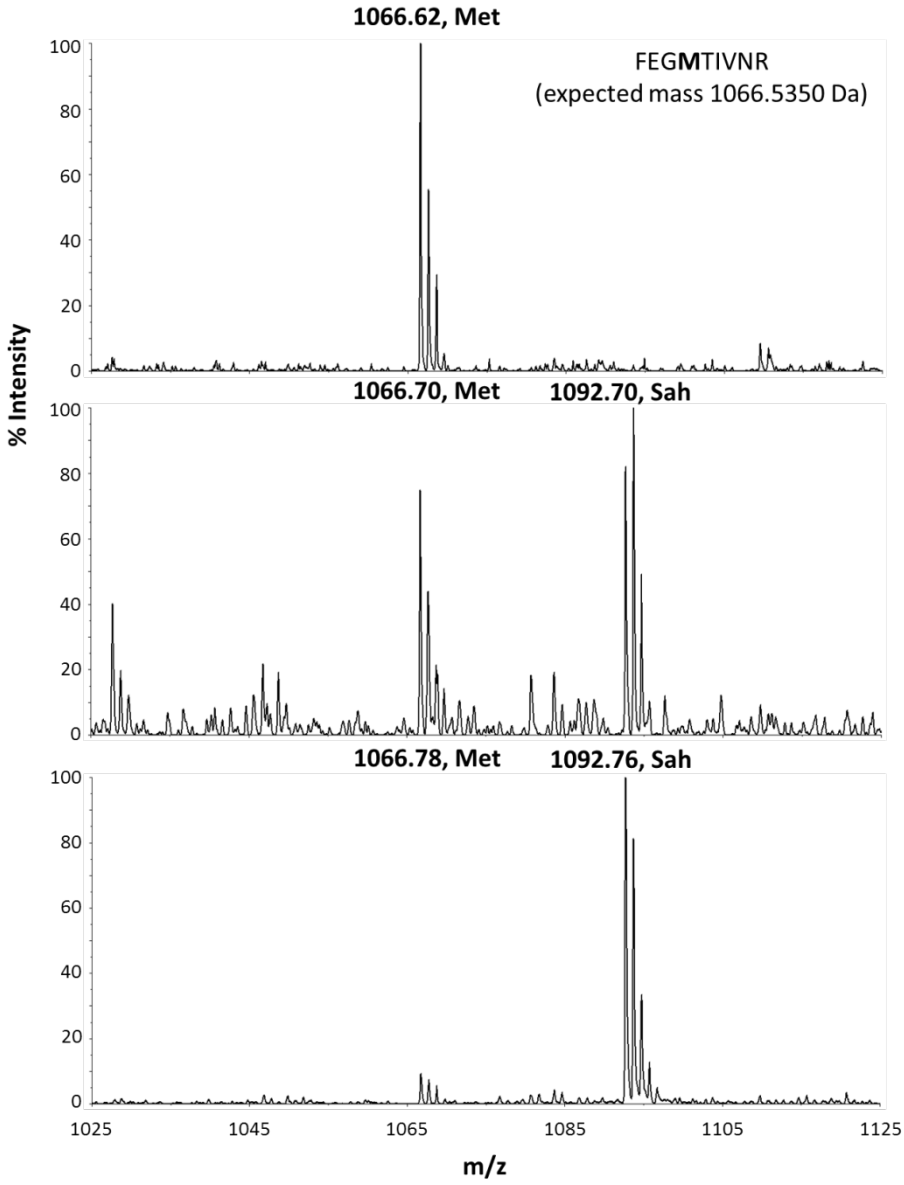
Figure 4.3 (cont)

c)



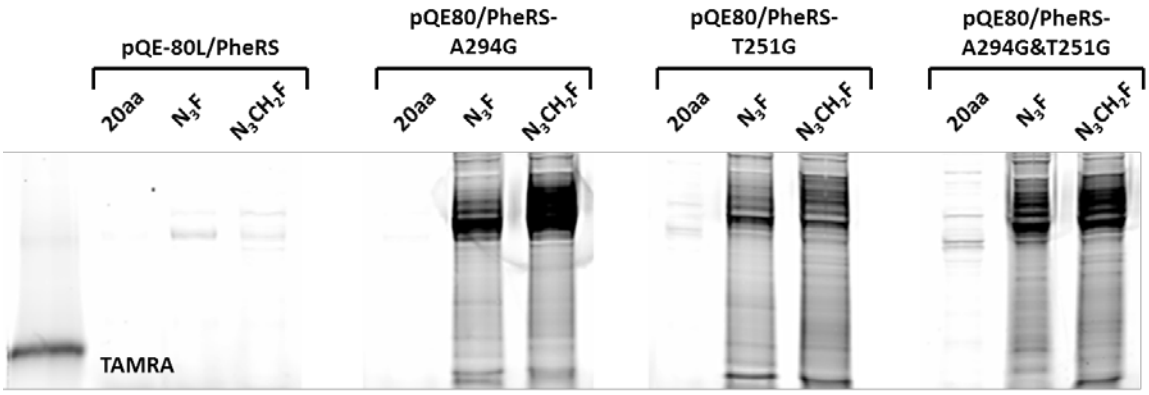
Synthesis of S-allyl-homocysteine. a) Scheme of synthetic route to produce S-allyl-homocysteine from L-homocysteine. b) ^{1}H NMR spectra of purified S-allyl-homocysteine. c) ^{13}C NMR spectra of purified S-allyl-homocysteine

Figure 4.4



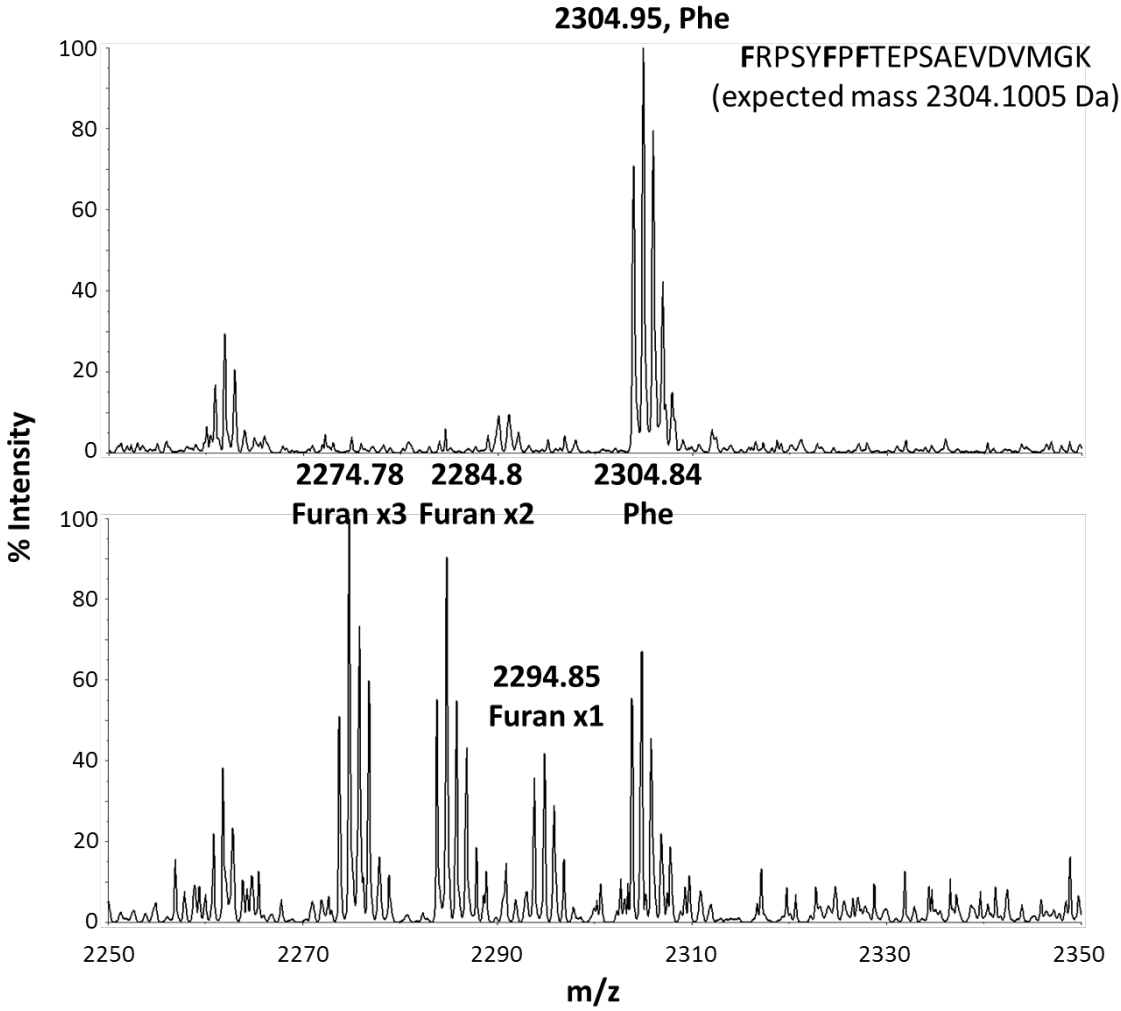
Overexpression of the wild-type MetRS enables quantitative replacement of methionine by S-allyl-homocysteine. GFPrmAM expressed in methionine auxotrophs is subjected to trypsin digestion; peptides are analyzed by MALDI-TOF MS. The peptide FEGMTIVNR is observed to have the correct expected mass when expressed in media containing all 20 canonical amino acids (top panel). Expression of GFPrmAM with endogenous levels of the MetRS in media depleted of methionine and supplemented with S-allyl-homocysteine shows the presence of a second peak corresponding to the 26 Da increase caused by methionine replacement (middle panel). Overexpression of the wild-type MetRS enables near complete replacement of methionine by S-allyl-homocysteine, as marked by the shifting of the peak from 1066 to 1092 Da.

Figure 4.5



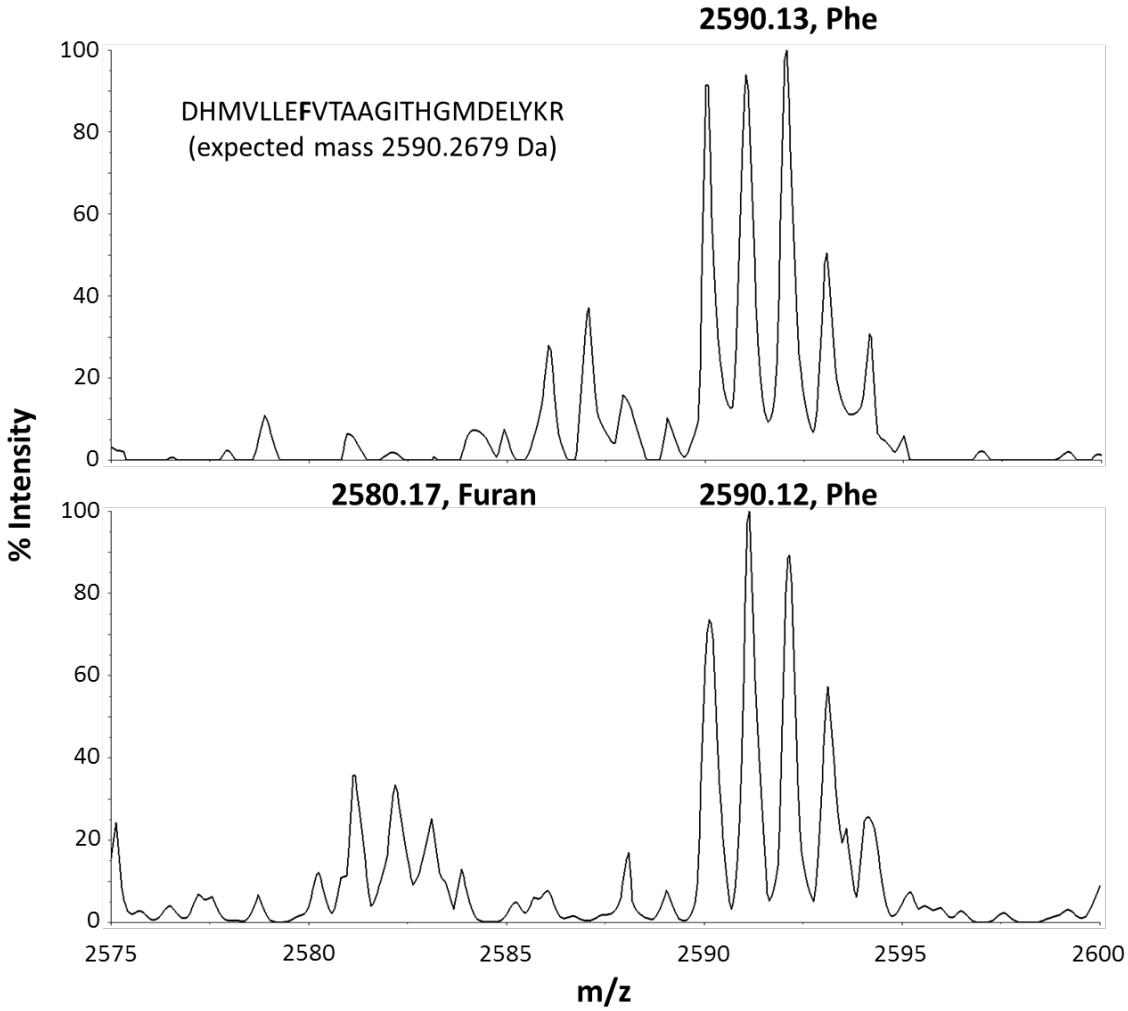
In-gel detection of cell lysates modified with DIBO-TAMRA. Induction of wild-type PheRS and PheRS mutants is performed post-media shift in media containing phenylalanine or azide bearing phenylalanine analogues in phenylalanine auxotrophic KY14 cells. Fluorescent signal in gel lanes arises from the attachment of DIBO-TAMRA to cellular proteins; DIBO-TAMRA undergoes strain-promoted azide-alkyne cycloaddition to azide analogues of phenylalanine if they are metabolically incorporated into proteins. No significant fluorescence is detected in lanes marked "**20aa**" for any overexpressed PheRS, due to the absence of metabolically incorporated azides. Proteins labeled with *p*-azidophenylalanine (**N_3F**), used as a positive control for azide incorporation, are observed in cases where PheRS-A294G, PheRS-T251G, or PheRS-A294G&T251G are overexpressed. In cases where *p*-azidomethylphenylalanine (**N_3CH_2F**) was added post-media shift, overexpression of the wild-type PheRS does not permit incorporation; however, the three overexpressed PheRS mutants tested all show attachment of DIBO-TAMRA to cellular proteins.

Figure 4.6



Furanylalanine incorporation is confirmed through MALDI-TOF MS analysis of purified 6xHis-tagged, wild-type PheRS. The correct expected mass for the peptide **F**RPSY**F**P**F**TEPSAEVDVMGK is observed when the wild-type PheRS is expressed in phenylalanine auxotrophic KY14 in media containing all 20 amino acids (top panel). Replacement of phenylalanine with furanylalanine results in a mass shift of minus 10 Da. The peptide contains three phenylalanine residues and a distribution of four different peaks is observed (lower panel).

Figure 4.7

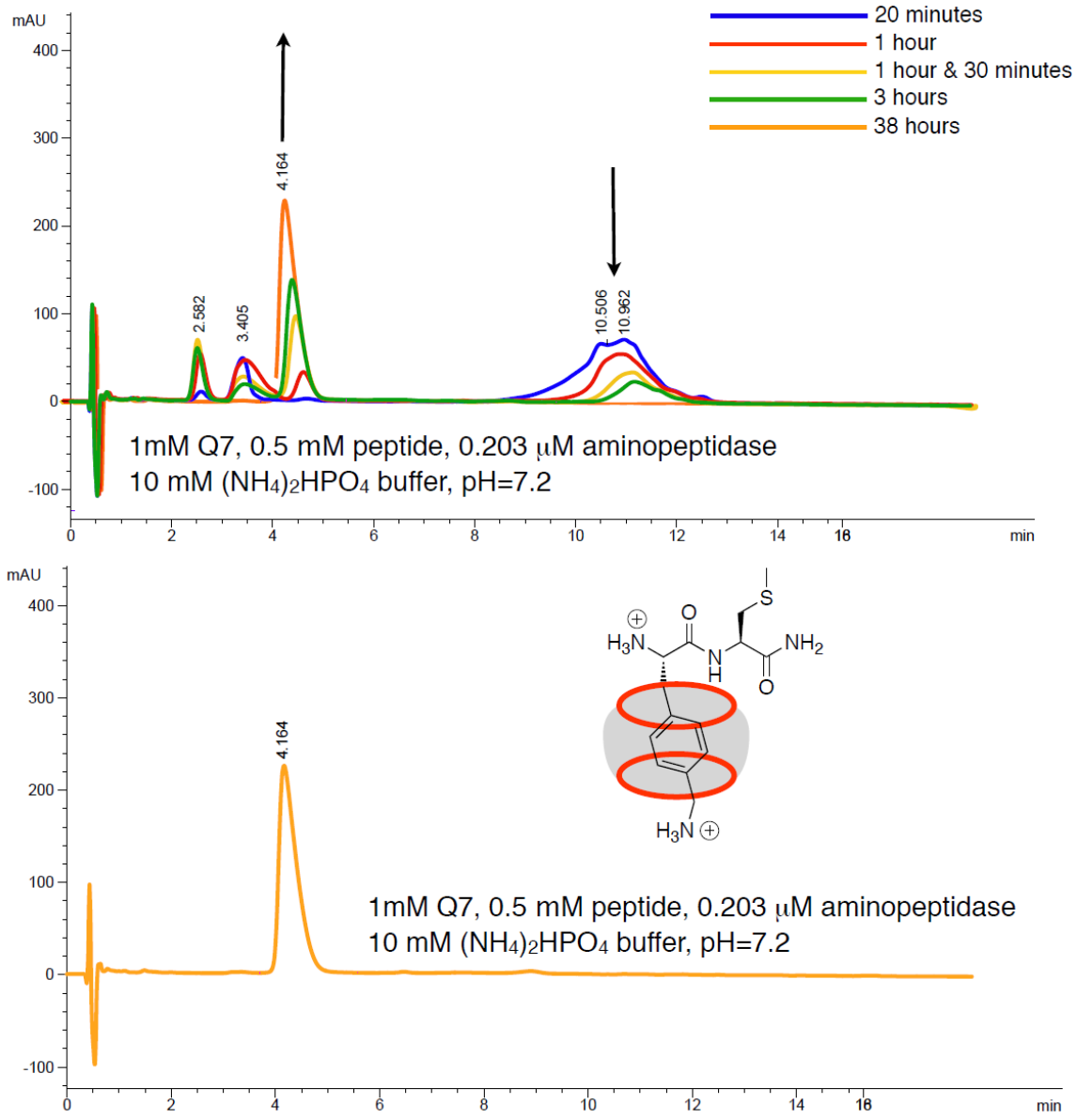


Furanylalanine incorporation is observed in 6xHis-tagged GFPm expressed without overexpressed wild-type PheRS in a non-phenylalanine auxotroph. The correct expected mass for the peptide DHMVLLE**F**VTAAGITHGMDELYKR is observed when 6xHis-tagged GFPm is expressed in DH10B in media containing all 20 amino acids (top panel). Approximately 10 percent of phenylalanine is replaced by furanylalanine, marked by the appearance of a new peak 10 Da smaller (lower panel).

Figure 4.8

Diels-Alder reaction between wild-type PheRS metabolically labeled with furanylalanine and biotinylated maleimide probes. Purified 6xHis-tagged, wild-type PheRS expressed in media containing furanylalanine was reduced and alkylated prior to overnight reaction with a biotinylated maleimide probe at room temperature. Aliquots of the reaction mixture were run on SDS-PAGE and transferred to nitrocellulose membrane for western blotting. Wild-type PheRS expressed in media containing the 20 canonical amino acids appears at correct molecular weight when probed using an antibody for penta-his (left panel). Biotin-PEG_2-Maleimide (MW: 525.62) attachment to furanylalanine labeled wild-type PheRS results in an increase in molecular weight; wild-type PheRS contains a total of 20 possible phenylalanine positions that can be replaced by furanylalanine. Incomplete alkylation enables reaction of maleimide with the lone free cysteine in wild-type PheRS resulting in biotinylation (middle panel). Biotin is detected co-localized with anti-his signal for wild-type PheRS, confirming modification of furanylalanine with the maleimide probe.

Figure 4.9



Q7 binding to newly formed N-terminal aminomethylphenylalanine protects peptides from further exopeptidase degradation. The peptide TGA(aminomethyl-phenylalanine)M is treated with leucine aminopeptidase in the presence of 1mM Q7. Aliquots from the reaction mixture are removed an analyzed on HPLC. Initial peptide undergoes a peak to peak conversion (top panel) to the Q7-stabilized dipeptide with N-terminal aminomethylphenylalanine, which is the only species present in the reaction mixture after 38 hours of digestion (bottom panel). The N-terminal aminomethylphenylalanine dipeptide bound to Q7 is confirmed through mass spectrometry. This is unpublished work from the Urbach Lab.

References

1. Grubbs, R. H. *Tetrahedron* **2004**, *60*, 7117.
2. Vougioukalakis, G. C.; Grubbs, R. H. *Chem. Rev.* **2010**, *110*, 1746.
3. Xia, Y.; Kornfield, J. A.; Grubbs, R. H. *Macromolecules* **2009**, *42*, 3761.
4. Johnson, J. A.; Lu, Y. Y.; Burts, A. O.; Lim, Y.-H.; Finn, M. G.; Koberstein, J. T.; Turro, N. J.; Tirrell, D. A.; Grubbs, R. H. *JACS* **2010**, *133*, 559.
5. Moellering, R. E.; Cornejo, M.; Davis, T. N.; Del Bianco, C.; Aster, J. C.; Blacklow, S. C.; Kung, A. L.; Gilliland, D. G.; Verdine, G. L.; Bradner, J. E. *Nature* **2009**, *462*, 182.
6. Lin, Y. A.; Chalker, J. M.; Floyd, N.; Bernardes, G. J. L.; Davis, B. G. *JACS* **2008**, *130*, 9642.
7. Van Hest, J. C. M.; Tirrell, D. A. *FEBS Lett.* **1998**, *428*, 68.
8. Chalker, J. M.; Lin, Y. A.; Boutureira, O.; Davis, B. G. *Chem. Commun.* **2009**, 3714.
9. Ai, H.-w.; Shen, W.; Brustad, E.; Schultz, P. G. *Angew. Chem. Int. Ed.* **2010**, *49*, 935.
10. Dieterich, D. C.; Link, A. J.; Graumann, J.; Tirrell, D. A.; Schuman, E. M. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 9482.
11. Tanrikulu, I. C.; Schmitt, E.; Mechulam, Y.; Goddard, W. A.; Tirrell, D. A. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15285.
12. Ngo, J. T.; Champion, J. A.; Mahdavi, A.; Tanrikulu, I. C.; Beatty, K. E.; Connor, R. E.; Yoo, T. H.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *Nat. Chem. Biol.* **2009**, *5*, 715.

13. Szychowski, J.; Mahdavi, A.; Hodas, J. J. L.; Bagert, J. D.; Ngo, J. T.; Landgraf, P.; Dieterich, D. C.; Schuman, E. M.; Tirrell, D. A. *JACS* **2010**, *132*, 18351.

14. Chinai, J. M.; Taylor, A. B.; Ryno, L. M.; Hargreaves, N. D.; Morris, C. A.; Hart, P. J.; Urbach, A. R. *JACS* **2011**, *133*, 8810.

15. Urbach, A. R.; Ramalingam, V. *Isr. J. Chem.* **2011**, *51*, 664.

16. Logsdon, L. A.; Schardon, C. L.; Ramalingam, V.; Kwee, S. K.; Urbach, A. R. *JACS* **2011**, *133*, 17087.

17. Kirshenbaum, K.; Carrico, I. S.; Tirrell, D. A. *ChemBioChem* **2002**, *3*, 235.

18. Palomo, J. M. *Eur. J. Org. Chem.* **2010**, *2010*, 6303.

19. Liu, D. S.; Tangpeerachaikul, A.; Selvaraj, R.; Taylor, M. T.; Fox, J. M.; Ting, A. Y. *JACS* **2011**, *134*, 792.

20. Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. *Nat Chem* **2012**, *4*, 298.

21. Kaya, E.; Vrabel, M.; Deiml, C.; Prill, S.; Fluxa, V. S.; Carell, T. *Angew. Chem. Int. Ed.* **2012**, *51*, 4466.

22. Diels, O.; Alder, K. *Justus Liebigs Annalen der Chemie* **1928**, *460*, 98.

23. Oliver Kappe, C.; Shaun Murphree, S.; Padwa, A. *Tetrahedron* **1997**, *53*, 14179.

24. Rideout, D. C.; Breslow, R. *JACS* **1980**, *102*, 7816.

25. Otto, S.; Engberts, J. *Pure Appl. Chem.* **2000**, *72*, 1365.

26. Canary, S. A.; Stevens, M. P. *J. Polym. Sci., Part A: Polym. Chem.* **1992**, *30*, 1755.

27. McElhanon, J. R.; Wheeler, D. R. *Org. Lett.* **2001**, *3*, 2681.

28. Chujo, Y.; Sada, K.; Saegusa, T. *Macromolecules* **1990**, *23*, 2636.

29. Gheneim, R.; Perez-Berumen, C.; Gandini, A. *Macromolecules* **2002**, *35*, 7246.

30. Kwart, H.; King, K. *Chem. Rev.* **1968**, *68*, 415.

31. Toncelli, C.; De Reus, D. C.; Picchioni, F.; Broekhuis, A. A. *Macromol. Chem. Phys.* **2012**, *213*, 157.

32. Boutelle, R. C.; Northrop, B. H. *J. Org. Chem.* **2011**, *76*, 7994.

33. Fersht, A. R.; Dingwall, C. *Biochemistry* **1979**, *18*, 1250.

34. Jakubowski, H.; R.Fersht, A. *Nucleic Acids Res.* **1981**, *9*, 3105.

35. Jakubowski, H.; Goldman, E. *Microbiological Reviews* **1992**, *56*, 412.

36. Kiick, K. L.; van Hest, J. C. M.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2000**, *39*, 2148.

37. Staudinger, H.; Meyer, J. *Helv. Chim. Acta* **1919**, *2*, 635.

38. Carrico, I. S.; Maskarinec, S. A.; Heilshorn, S. C.; Mock, M. L.; Liu, J. C.; Nowatzki, P. J.; Franck, C.; Ravichandran, G.; Tirrell, D. A. *JACS* **2007**, *129*, 4874.

39. Kothakota, S.; Mason, T. L.; Tirrell, D. A.; Fournier, M. J. *JACS* **1995**, *117*, 536.

40. Unpublished results from Urbach Lab

41. Connor, R. E.; Piatkov, K.; Varshavsky, A.; Tirrell, D. A. *ChemBioChem* **2008**, *9*, 366.

42. Varshavsky, A. *Genes Cells* **1997**, *2*, 13.

43. Link, A. J.; Vink, M. K. S.; Tirrell, D. A. *JACS* **2004**, *126*, 10598.

44. Yoo, T. H.; Tirrell, D. A. *Angew. Chem., Int. Ed.* **2007**, *46*, 5340.

45. Paizs, C.; Katona, A.; Rétey, J. *Chemistry – A European Journal* **2006**, *12*, 2739.

46. Yoo, T. H.; Link, A. J.; Tirrell, D. A. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13887.

Concluding Remarks

The amino acids described in this work have expanded artificial protein sequence space for researchers interested in doing residue-specific incorporation. We have access to new chemical reactions through the incorporation of S-allyl-homocysteine, furanylalanine, and azidomethylphenylalanine; however, further work must be done to implement the chemistry in a practical manner.

The growing complexity of unnatural amino acids often requires evolution of new aminoacyl-tRNA synthetases to enable their use in protein synthesis. The discovery of the PraRS and propargylglycine pair could help to popularize cell-specific BONCAT approaches; the commercial availability and cost of propargylglycine should make propargylglycine attractive to researchers. Another interesting aspect to propargylglycine is that it is a known natural product of various species of *Streptomyces*[1,2] as well as other fungi[3]. Although the biosynthetic pathways for propargylglycine are currently unknown, the ability to perform BONCAT without addition of any exogenous amino acids is an appealing proposition.

Exploring artificial protein sequence space requires the incorporation of artificial amino acids into proteins, which can perturb protein structure and function; directed evolution is often required to recover the lost protein properties. Expression of the homoisoleucine form of chloramphenicol acetyltransferase can be performed without any leucine remaining in the protein sequence, giving access to a new point in sequence space. Using the homoisoleucine form of chloramphenicol acetyltransferase as a starting point, we can

continue to explore paths along the protein fitness landscape. One could imagine that with sufficient directed evolution, we could arrive at a protein that can only function in the artificial protein sequence space.

1. Scannell, J. P.; Pruess, D. L.; Demny, T. C.; Weiss, F.; Williams, T. J Antibiot (Tokyo) 1971, 24, 239.
2. PIOTROWSKA, M.; PASZEWSKI, A. J. Gen. Microbiol. 1986, 132, 2753.
3. Hatanaka, S.-I. In Fortschritte der Chemie organischer Naturstoffe / Progress in the Chemistry of Organic Natural Products; Herz, W., Kirby, G. W., Moore, R. E., Steglich, W., Tamm, C., Eds.; Springer Vienna: 1992; Vol. 59, p 1.