EXPANDING THE GENETIC CODE FOR SYNTHESIS OF PROTEINS WITH NATIVE BIOLOGICAL MODIFICATIONS AND NOVEL CHEMICAL, BIOPHYSICAL PROBES

A Dissertation

by

YAN-JIUN LEE

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe R. Liu
Committee Members,	Tadhg P. Begley
	Coran M. H. Watanabe
	Jean-Philippe Pellois
Head of Department,	François P. Gabbaï

December 2015

Major Subject: Chemistry

ABSTRACT

Engineering of protein with new modifications essentially expands the protein functions which provide powerful tools for investigating significant biological questions. New modifications also offer scientists the ability to handle protein activity at will. Therefore, development of facile and robust protein modification protocol is highly desirable in almost every aspect of biological research. In combination with chemistry and biology principles, the genetic code expansion technology has emerged as an effective strategy for manipulating proteins and interrogating biological systems. In this regard, the overall objective of the presented research is to expand the chemical biology toolbox based on noncanonical amino acid incorporation and bioorthogonal chemistry.

During my graduate study, I was focused on methodology development to genetically incorporate defined, homogeneous native protein posttranslational modifications and new chemical/biophysical probes into proteins for biological investigations, particularly for epigenetic applications. I have established the protocol to incorporate various native lysine acylations into proteins including crotonylation, propionylation and butyrylation. In addition, I also developed a chemical biology approach to study epigenetics, coined “chemical antibody”. This approach allows high throughput fluorescent detection of posttranslational modifications without the use of antibodies. I genetically encoded native lysine posttranslational modification surrogates, such as acryloyllysine, into proteins. Coupled with the bioorthogonal chemistry development, the site-specific detection of the posttranslational modifications was

accomplished, essentially offering an alternative route for understanding epigenetic enzyme network. Following the same line of approach, in order to study protein long chain fatty acid acylation, I incorporated long chain unstrained alkene into proteins as chemical reports that are able to be selectively labeled *via* tetrazine-based bioorthogonal reaction. This development also generates a new direction of research in studying the long chain fatty acid metabolic acylome using chemical biology approach. Lastly, the site-specific incorporation of fluorine-containing noncanonical amino acid probes into proteins was also developed. Fluorine is referred to as "magic bullet" in medicinal organic chemistry and biochemistry. Introducing fluorine into protein allows generating an unrivalled unique probe in a protein without any structural perturbation. The site-specific incorporation of perfluorinated probes is an invaluable tool for the study of noncovalent protein interactions as well as the application on protein designs.

Overall, the collective work presented has extensively improved our ability of handling site-specific protein modifications with native biological PTMs and handy chemical/ biophysical probes for diverse medical and biological applications.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Wenshe Liu, and my committee members, Dr. Tadhg Begley, Dr. Coran Watanabe, and Dr. Jean-Phillippe Pellois, for their guidance and support throughout the course of this research.

Thanks also go to my lab colleagues for their help with both my research and life. Dr. Zhiyong Wang was a great mentor and scientist at early time of my research. Dr. Yan-Shih Wang was an outstanding lab senior and friend for very helpful advices on both research ideal and career development. Other postdocs and students also help me to overcome obstacles at certain point of my research. I also want to extend my gratitude to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my mother and father for their encouragement, their patience and love.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF SCHEMES

LIST OF TABLES

TABLE Page

CHAPTER I

INTRODUCTION

Advances in Site-Specific Protein Modifications

Protein is nature's most versatile and fascinating class of biomolecules due to their diverse properties of function and structure. While proteins are comprised of only twenty canonical amino acids with relatively simple functional groups, they sustain the entire normal life of a living organism where proteins are responsibility for almost all aspect of biological activities, such as catalysis, structure, signal transduction, gene regulation etc. [1-5] Engineering of protein with desirable handles essentially provides an exclusive ability to manipulation of protein behavior, in turn to probe, regulate and understand the entire living system. Accordingly, development of methodology for specific protein modification has emerged as an invaluable tool for investigation of protein associated biological questions including protein functions, architectures, as well as metabolism and interactive network. The chemical modification of amino acid side chains can be backtracked to 100 years ago, whereas today a variety of bioconjugation reactions are developed and applied routinely to alter the proteins properties as dictated by the application or analysis method at hand.[6, 7] Each of these approaches allows new protein targets to be addressed, leading to advances in biological understanding, new protein-drug conjugates, targeted medical imaging agents and hybrid materials with complex functions.

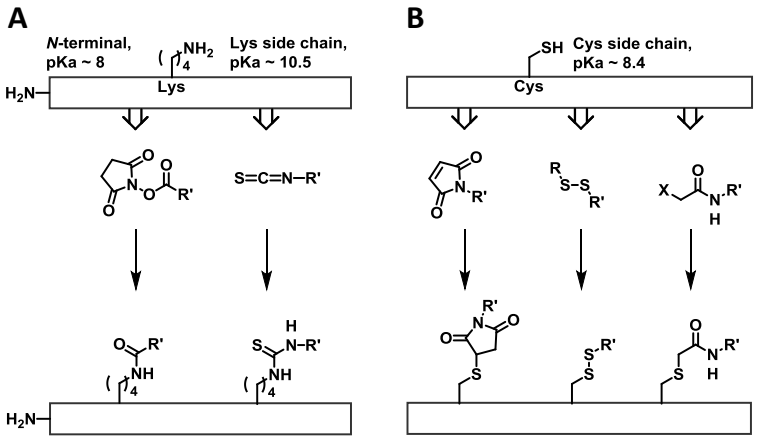


Figure 1-1. Classic methods for protein modification

Early methods of protein functionalization exploited the reactivity of endogenous functional groups. All functional groups of a protein have different pKa values and therefore their reactivity can be controlled by adjustment of pH in a buffer. This fact allows a selective reaction on one type of amino acid residue as mostly applied for modification strategies to the amine group of the lysines or the *N*-terminus of a protein and the thiol group of cysteines (**Figure 1-1**). Lysine has a side chain amine with pKa ~ 10.5 and is a common modification target due to a variety of reactions are accessible for selective primary amines modification.[8] By treating the protein with excess amine-reactive reagents such as *N*-hydroxysuccinimidyl esters or isothiocyanates, selective protein labeling is achieved (**Figure 1-1A**). The *N*-terminus of a protein has a unique amine which is slightly more acidic than that of lysine side chain with the pKa around 8.[8] Therefore, the N-terminus of protein offers the opportunity for single-site protein modification. Cystein is a rare residue in whole proteome and makes it a precious native

chemical handle for protein derivatization. The modification of cysteine can be accomplished via disulfide linkage or through thiol nucleophilic addition to the thiol-reactive reagents such as maleimides probes or haloacetamides (**Figure 1-1B**).[9, 10] In addition to lysine and cysteine modifications, methods have also been developed to chemically modify tyrosine and aspartate or glutamate residues.[9] These classic methods are convenient and actively been used for quick qualitative protein modifications. However, their chemoselectivity become the hurdle when site-specific modification is desired.

To achieve site-specific, engineering or external assistance from the non-native reagents are required. A traditional approach has relied on the creation of fusion proteins of the protein of interest with green fluorescent proteins (GFP). With the advances of molecular biology technologies, this approach can conveniently introduce a GPF tag at the genetic level. However, the GFP tagging is limited to the two terminal ends of a protein. Moreover, the size of GFP (26 kDa) may risk in perturbing the native state of the protein of interest.[11]

The complementary use of both genetic and chemical methods has provided a large toolbox that allows the preparation of almost unlimited protein constructs with either natural or synthetically modified residues. Numerous milestones were established in the past twenty years in contributing to the advances of site-specific protein modification. These milestones include click chemistry, bioorthogonal reaction, native chemical ligation, expressed protein ligations as well as genetic code expansion and will be summarized extensively but not exhaustively in the following sections.

Click Chemistry and Bioorthogonal Reactions

Given the high environmental constrains in physiological and cellular setting, it requires unique chemistry in order to handle and control manipulate the behavior of biomolecule of interest. Therefore, the pressing need promotes the development of click chemistry and bioorthogonal reactions in the past decades.[12] Described and termed by Sharpless and coworkers in 1999, the "click chemistry" concept defines a new group of chemical reactions that are modular, highly selective and efficient and can be performed at mild ambient condition in physiological solutions.[13, 14] Since it was first introduced, in past decade the development of click chemistry has had an enormous impact on substantially every fields of chemical science and technology, ranging from medicine, material science, polymer chemistry to chemical biology, biomolecule engineering and bionanotechnology. Due to their robustness, superb selectivity and versatility, click reactions allow the assembly of fundamental fragments into sophisticated but valuable targets under a benign condition with decent reaction rate. Broad utilization of click chemistry in various subjects enables scientist to achieve multidisciplinary research, effectively bridging discrete worlds into one.

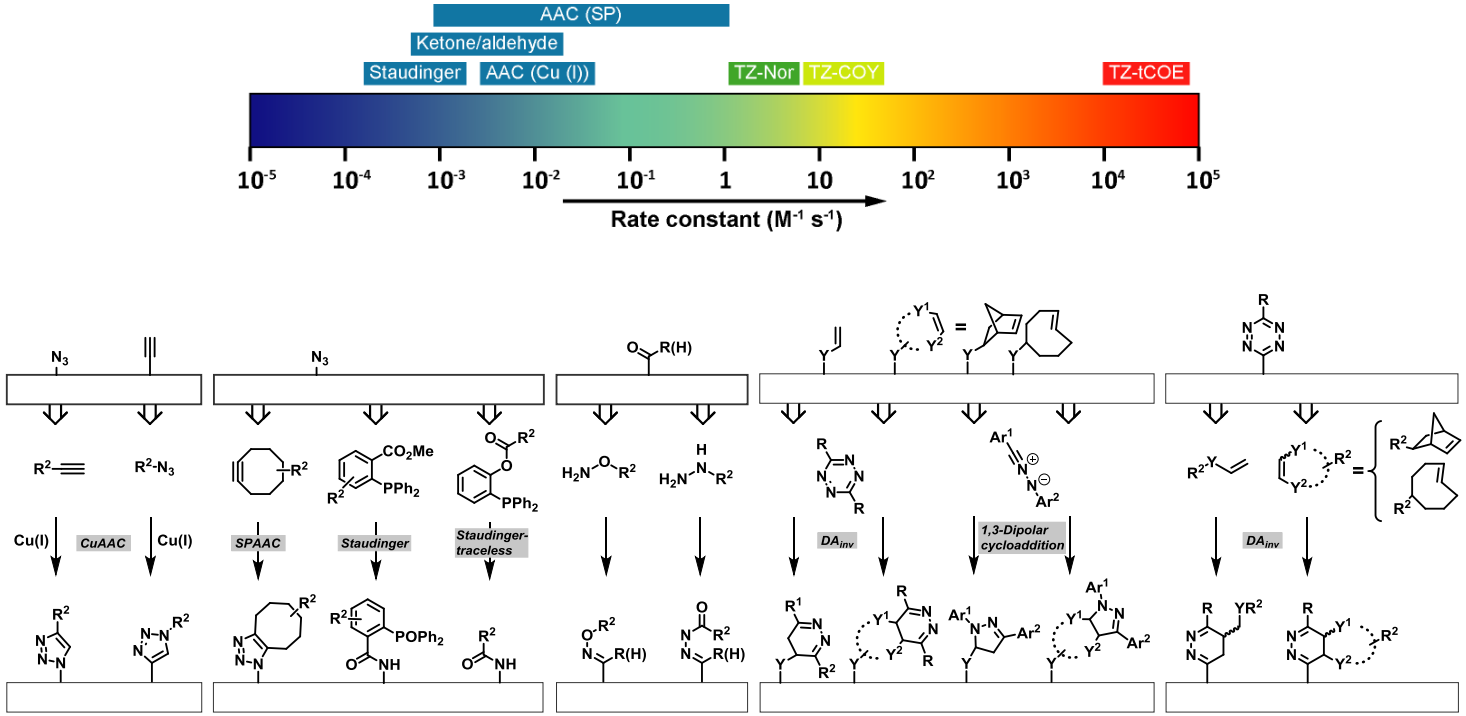


Figure 1-2. Click chemistry and bioorthogonal reactions.

Early focus of click chemistry centered on the Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) reaction (**Figure 1-2**). Due to the biologically inert nature, azide and alkyne have become unrivaled chemical tools, broadly incorporated into biomacromolecules and applied in *in vitro* biochemical studies. Applications of CuAAC in chemical biology include activity-based protein profiling and selective biomolecule labeling. However, one pitfall of using CuAAC in living systems is the detrimental effect of copper ions and reducing agent formulation which raise concerns about protein damage or precipitation as well as the reactivity oxygen-induced cytotoxicity.[15] Thereby, considering the complexity of living system that is composed of interacting biopolymers, ions and metabolites, further methodology development on the basis of click chemistry in complying to the cellular constrains are in pressing need and prevailing in chemical biology filed. To this end, the pursuit of chemical reactions with outstanding bioorthogonality falls into the development of bioorthogonal reactions. [1]

Bioorthogonal reactions defined a new category of chemical reactions that can occur in living systems without interfering with native biochemical processes and have emerged as highly specific tools that can be used for investigation the dynamics and function of biomolecules in living system. Typical bioorthogonal reactions include Straudinger ligation, Strained-promoted azide-alkyne cycloaddition, and hydrazone/oxime conjugates from aldehyde/ketone-hydrazide/alkoxylamine pair. The Staudinger ligation is based on the classic Staudinger reaction between azide and triarylphosphine, developed by Bertozzi and coworkers in 2000 for the selective glycoprotein labeling.[16] The reaction was later been used in studying protein

glycosylation both with live cells and live mice.[17] Shortly after the Staudinger ligation, a new variant was described referred as "traceless" Staudinger ligation, where the final amide-linked product is freed from the phosphine oxide moiety and gives a native amine linkage.[18, 19] Staudinger launched the field of bioorthogonal chemistry as the first reaction with completely abiotic functional groups although it is less widely used due to its relatively slow reaction rate (~ 10^{-3} M^{-1} s^{-1}). [20]

Aldehyde and ketone are two of the most versatile groups in synthetic organic chemistry and were among the first functionalities to be explored as bioorthogonal chemical reporters. Under acidic conditions (pH 4-6), the carbonyl groups are preferentially form a reversible Schiff base with the equilibrium favoring the free carbonyl form. The use of strong electron donating group at α-substituent of amine, such as hydrazide or alkoxyamines, shifts the equilibrium in favor of the hydrazone and oxime product, though the reaction kinetics is slow (10^{-4}-10^{-3} M^{-1} s^{-1}). The use of nucleophilic catalyst aniline will facilitate the condensation and boost up the reaction rate (170 M^{-1} s^{-1} for hydrazone and 8.2 M^{-1} s^{-1} for oxime). Application of aldehyde/ketone-hydrazide/alkoxyamine includes glycoprotein and cell surface receptor labeling[20].

Azide and alkyne are ideal bioorthogonal chemical reports. Although the use of copper limits CuAAC, the development of copper-free methods will circumvent the limitation. Cu(I) catalyst can be avoided by strain-promoted azide-cyclooctyne cycloaddition (SPAAC) (**Figure 1-2**). Driven by the ring strain, the cyclooctyne triple bond has an elevated energy state and therefore reacts spontaneously with an organic

azide under mild, aqueous conditions, elegantly sidestepping the use of Cu(I) catalyst. Although originally described as a relative slow reaction, cyclooctyne derivatives that show improved SPAAC reaction kinetics developed. The pursuit of other fast catalyst-free click reactions that can be applied in living systems for labeling biomolecules under biologically relevant conditions has recently revitalized one celebrated reaction, namely the Diels-Alder [4+2] cycloaddition (DA). Not only is DA rapid, selective and chemically accessible, but also DA is highly aqueous friendly, giving augment reaction kinetics up to hundred-fold in aqueous media in comparison to in organic solvents. In particular, Diels-Alder reaction with inverse electron demand (DAinv) of 1,2,4,5-tetrazines with strained cyclic dienophiles is exceptionally fast in aqueous buffer (**Figure 1-2**).[21, 22] 1,2,4,5-tetrazines react with dienophiles such as including norbornene, trans-cyclooctene, cyclooctyne, cyclobutene, and cyclopropene very selectively, irreversibly giving dihydropyridazine products and release very benign dinitrogen byproduct. Another copper-free click reaction that has been recently explored but not yet highly appreciated is the 1,3-dipolar cycloaddition (1,3-DC) between nitrile imine (**Figure 1-2**).[23] Compared to tetrazine and cyclooctyne that stably exist in aqueous conditions, nitrilimine reacts with water and therefore needs to be formed transiently. Two methods are generally used to transiently form nitrile imine. One is the photolysis of tetrazole and the other is the ionization of hydrazonyl halide. Lin et al. have extended the first approach for photoclick protein labeling in living cells.[24, 25] The second approach was recently utilized for the fluorescent turn-on labeling of proteins incorporated with norbornene, cyclopropene, and acrylamide moieties.[26-28] Both reactions display rapid

kinetics when a strained alkene/alkyne is involved. Applications of these two reactions include labeling biomolecules in cells and cancer diagnostics.[26, 29-32]

Bioorthogonal chemistry directed biomolecule conjugation has shown widespread applications attributed to their unique features of selectivity, biocompatibility and reaction kinetics in physiological conditions. The use of bioorthogonal chemistry typically proceeds with two steps. First, a cellular substrate is modified with a bioorthogonal functional group (chemical reporter) and introduced to the cell; substrates include metabolites, enzyme inhibitors, etc. The chemical reporter must not alter the structure of the substrate dramatically to avoid affecting its bioactivity. Secondly, a probe containing the complementary functional group is introduced to react and label the substrate. To introduce the chemical reporter into the proteins, current advances involve solid phase synthesis, native chemical ligation, expressed protein ligation, as well as enzymatic modification and genetic code expansion for ncAA incorporation.

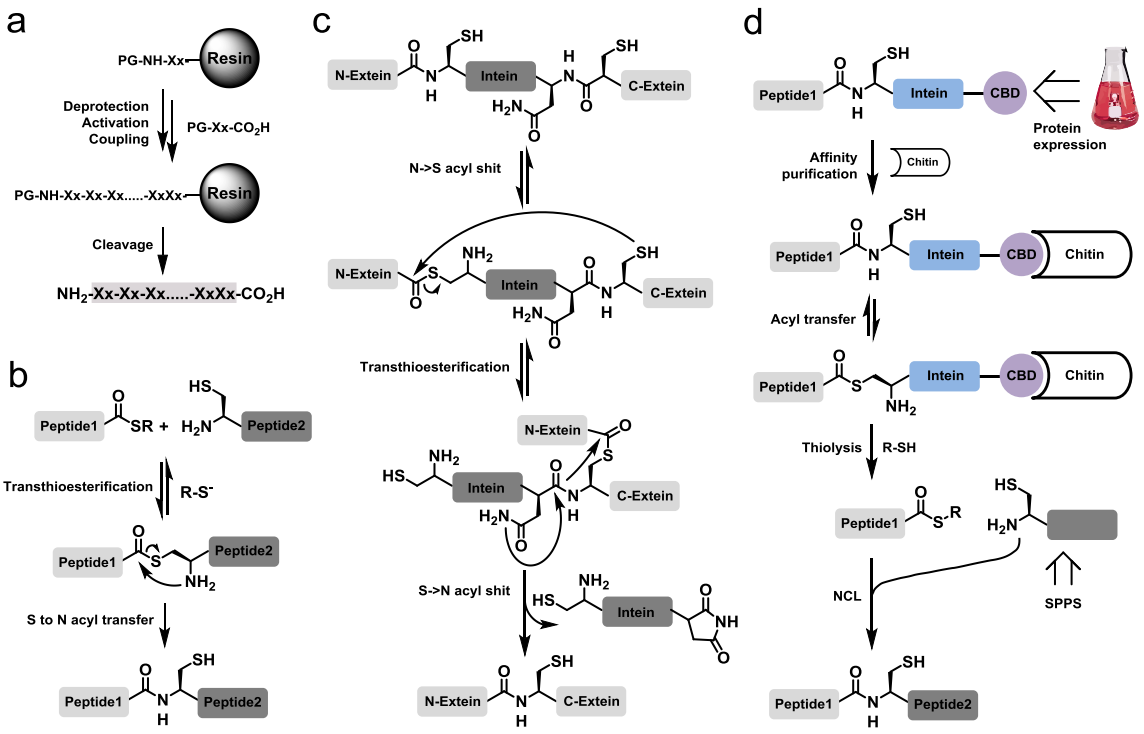


Figure 1-3. Protein semisynthesis approaches. Schematic representation of A) solid phase peptide synthesis; B) native chemical ligation; C) the mechanism of intein chemistry; D) expressed protein ligation.

Protein Semisynthesis, Native Chemical Ligation and Expressed Protein Ligation

Chemical synthesis has long term been used to manipulate functionality transformation and allows custom implementation of any chemical reporter on the target molecules. In addition, synthetic chemicals are easily scalable and economical. Therefore, to achieve the production of medical significant or biological relevant biomacromelecules, organic chemist have attempted to develop means in efficient assembly of basic biological elements such as the synthesis from amino acids to peptides and functional proteins. To this end, peptide synthesis was developed at this regard. One

critical milestone in the development history is solid phase peptide synthesis (SPPS) which greatly simplifies the tedious purification steps and boosts up the synthesis productivity.[33] SPPS has established routine protocols generally involving repeatedly deprotection, activation and coupling three steps to bring together each amino acids into peptide chain. Followed by one step deprotection and cleavage from resin, the synthesis of peptide with chemical reporter or PTM incorporated can be accomplished (**Figure 1-3A**). SPPS is straightforward and can be applied to synthesize proteins/peptides with diverse chemical reporters and PTMs. However, given the fact that this approach lies on the amide bond formation with chemical reactions, it suffers from low yield when the target polypeptide is larger than 50 amino acid residues.[34] For instance, if each amino acid addition has a 90% yield then the overall yield of the synthesis of a 50 amino acids polypeptide is merely 0.5%. Since a functional protein is normally greater than 100 amino acids, the low yield of SPPS is a conspicuous pitfall limiting the generally application in synthesis of large proteins. Although modern SPPS instrumentation has greatly optimize the coupling and deprotection yields to greater than 99%, leading to an overall yield of greater than 90% for a 50 amino acids polypeptide. To overcome the limitations, the methods to glue polypeptide from SPPS were therefore developed, which was referred as chemical ligation.[35]

There were numerous chemical ligation protocols developed and intended to assemble SPPS peptide fragments, whereas the created non-native linkages were concerns when native protein is required.[35] Not until the development of native chemical ligation (NCL), chemical ligation became routine protocol in protein semisynthesis.

Since it was reported by Kent and coworkers in 1994, NCL has been emerged as one of the major strategy to the semisynthesis of proteins with desired chemical reporters and PTMs modifications at programmed site.[34] It allows unprotected peptide fragments to be chemoselectively ligated together with native amide bond which greatly extends the size limitations of SPPS that chemist can reach. NCL relies on the reaction between the C-terminus thioester of one synthesized peptide fragment and the N-terminus cysteine residue of the other. The thiol group in the cysteine reacts to the thioester by means of reversible thiol/thioester exchange to afford the new thioester-linked conjugated peptide. Followed by irreversible intramolecular S- to N-acyl transfer rearrangement, full length native protein with more than 100 residues can be afforded in single NCL step (**Figure 1-3B**).[34] Since two peptide fragments are synthesized separately, NCL not only promotes the efficiency and extend the size limitation of SPPS but also allows the synthesis chemical reporter functionalized protein at desirable site of a protein. Specifically, when the ncAA locates in the N-terminal peptide fragment, the C-terminal peptide can be synthesized by recombinant DNA-based protein expression. This method offers a solution to synthesize larger proteins with ncAAs incorporated close to N-terminus.[35, 36]

Another notable milestone of semisynthesis is the development of expressed protein ligation (EPL). While the installation of ncAAs at N-terminus of protein was realized readily through NCL, the synthesis of protein with ncAAs incorporated at C-terminus became the remaining challenge. Unlike the C-terminal end of a peptide fragment, the N-terminal fragment contains a thioester tail which cannot be recombinantly produced in simple protocol. Therefore, a method to produce a peptide

fragment with C-terminal end thiolester through perhaps recombinant protein expression was in high demand. Protein splicing is a posttranslational process, involving a series of intramolecular rearrangement to precisely remove an internal sequence (intein) and ligate the two flanking segments (extein) via a reversible transesterification and a S- to N-acyl shift mediated by an adjacent asparagine, resulting in the formation of an native amide linkage (**Figure 1-3C**).[37] Mutation of the C-terminal asparagine residue within the intein to an alanine blocks the final step in protein splicing. Proteins expressed as in-frame N-terminal fusions to such mutant inteins can be cleaved by excessive thiols, such as dithiotheitol, β-mercatoethanol or thiolphenol, to give the corresponding protein α-thioester derivatives (**Figure 1-3D**). Commercial expression vectors are available that allow the generation of recombinant intein fusion proteins containing chitin binding domain (CBD) affinity purification tags where the cleaved intein corpses are removed from α-thioester protein by chitin-immobilized beads. The obtained recombinant protein α-thioester can then react with a N-terminal cysteine containing peptide (either synthetic or recombinant) via NCL to afford a semisynthetic target protein (**Figure 1-3D**). Expressed protein ligation refers to the combination of the intein thiolysis and native chemical ligation steps.[36, 38, 39] EPL successfully offers a solution for the semisynthesis of proteins with C-termius ncAAs incorporated. SPPS is essentially able to synthesize the peptides with any desired chemical reporter or PTM. The advances of NCL and EPL allow the semisynthesisi of theoretically any proteins functionalized with chemical reporter or biological significant PTMs for *in vivo* and *in vitro* studies.

SPPS, NCL in combined with EPL have provided a powerful semisynthetic toolbox and have greatly expanded the realm of available proteins that can be synthesized. Nonetheless, “the big three” are not invincible due to their intrinsic limitations. In order to perform the transthioesterification steps, the introduction of terminal cysteine is required. A cysteine scar remains within the ligated product and the cysteine mutation may not be desirable in following studies. Although current advances have developed the desulfurization methodologies, the compatibility of the extra step remains a concern. In addition, the biophysical parameters of the protein-intein-CBD fusion need to be taken into account at protein expression. The three component chimeric protein may cause unexpected solubility and stability during recombinant protein expression and expression processes.[36]

Enzymatic Installation of Chemical Reporter

The nature have evolved enzyme to be highly specific to its native substrates in order to ensure the catalytic function and tight regulation within the complex biological network, thereby to sustain normal life of an organism. The outstanding selectivity of enzymes has inspired the development of enzymatic protein labeling approach. Site-specific protein labeling is essentially artificial modification on protein with new chemical entities at the posttranslational stage. Based on the analogy between protein labeling and protein posttranslational modification, enzymatic tools are with high potential in developing efficient protocol for programmable labeling of target proteins with chemical probes of diverse structures and functionalities.[40, 41] Enzymatic reactions

are selective and catalytic within the cellular environment, therefore protein labeling use this approach are typically low background, efficient and biocompatible. Generally, this method relies on first installation of an enzyme-specific "tag" on the target where the tag can be introduced at genetic level, followed by enzymatically transforming the tag to or with desired chemical reporters (**Figure 1-4A**). De novo synthesis of substrates or cofactors with bioorthogonal functionality is normally required in order to incorporate chemical reporters into target proteins. Also, engineering of the cognate enzyme against unnatural substrates to assure the specificity is inevitable prior to the practical utilization. The engineering effort is tedious and cumbersome therefore at present there are only handful protocols available. Recent developed and more generally adapted are summarized as follows.

Transglutaminase (TGase) catalyzes protein crosslinking in the cell by the formation of an isopeptide bond between glutamine and lysine residues. Gunea pig liver TGase (gpTGase) demonstrates high specificity for Gln containing proteins as the acyl doners and low specificity for Lys containing substrate as acyl acceptors. This feature makes it possible to use gpTGase in protein labeling with amine functionalized synthetic probes. A short XXQXX peptide tag (Q-tag, X denotes any residues) is fused to the target protein and expressed as a fusion protein. Followed by TGase-catalyzed conjugation, the probes are covalently tagged the target protein via isopeptide linkage.[40]

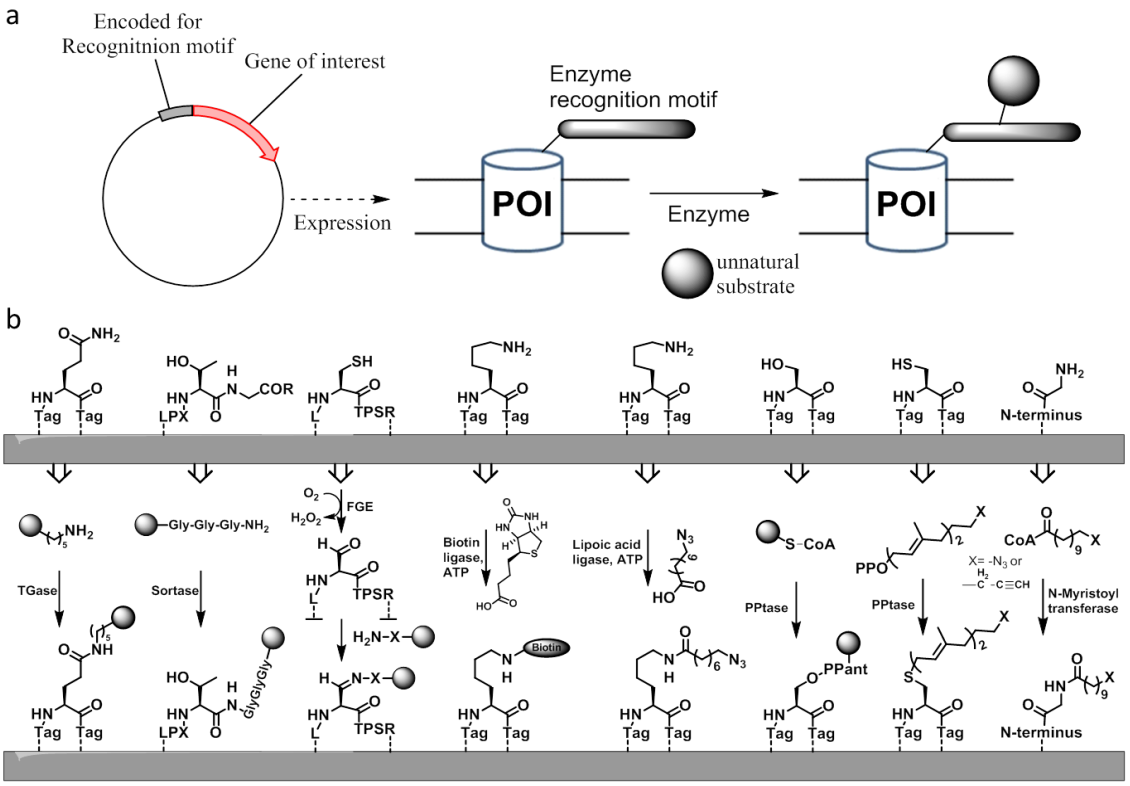


Figure 1-4. Enzymatic protein modifications. A) General scheme for introduction of chemical reporters via engineered enzyme. B) Available enzymatic tool sets for site-specific protein modification.

Sortase is a transpeptidase which catalyzes a transpeptidase reaction between a specific internal sequence of a protein and an amino group present on the N-terminus of polyglycines. The *S. aureus* sortase, SrtA, recognizes LPXTG motif and cleave the peptide bond between threonine and glycien and form a new peptide bone with pentaglycine-containing peptide. Site-specific labeling of target protein can be achieved by genetically introducing the LPXTG motif onto the C-terminus of target protein of interest. Followed by SrtA-mediated conjugation with oligoglycine-functionalized probe, the target protein is modified with desired chemical reporoters.[41, 42] The advances of

sortase method have been utilized for cell surface protein imaging as well as the development of N-terminus oligoglycine tagging protocol.[40, 41]

Formylglycine-generating enzyme (FGE) is involved in the maturation of sulfatase and catalyzes a molecular oxygen-mediated oxidation of cysteine in the LCTPSR peptide, generating formylglycine as product and hydrogen peroxide as byproduct. The consensus 6-mer LCTPSR tag can be genetically encoded next to the target protein either at N- or C-terminus. After enzymatic oxidation, the afforded aldehyde can easily be functionalized with hydrazide or alkoxyamine probe. Engineering of the cell to overexpress FGE has allowed the in situ generation of aldehyde on the target protein and enabled in vivo cell imaging applications.[40, 43, 44]

Biotin ligase catalyzes the protein posttranslational biotinylation. *E. coli* biotin ligase BirA, together with ATP, modifies the lysine residue of a15-mer consensus motif (AP tag) to biotinylated lysine. The biotinylated tagged-protein can be further enriched or detected using streptavidin resin or probes. Moreover, the engineering of BirA has allowed the incorporation of biotin analogues with bioorthogonal functionalities implemented including the ketone, azide and alkyne, which can be further functionalized via bioorthogonal reactions on the target protein.[40, 45, 46]

Similar to biotin ligase, lipoic acid ligase catalyzes ATP-dependent protein lysine posttranslational modification with lipoic acid. *E. coli* lipoic acid ligase LplA recognizes a 22-residue peptide (LAP tag) and engineered LplA is able to utilize azide-containing lipoic acid analogue. Expression of LAP-protein fusion followed by LplA-mediated

modification, the target protein is modified with azide and allows further functionalization with probe.[40, 47]

Phosphopantetheinyl transferase (PPTase) has been recently developed as a useful tool for protein labeling due to their broad substrate specificity with CoA conjugated small molecules. PPTase transfers 4'-phosphopantetheinyl group (PPant) of CoA to a conserved serine residue to an acyl carrier protein (ACP) or peptidyl carrier protein (PCP). Because PPTase only recognize the 3'-5'-ADP moiety of CoA at the active site, various CoA analogue probes have been reported incorporation by PPTase. To adopt PPTase catalyzed protein labeling, ACP or PCP domains are expressed as a fusion to the target protein, thereby to incorporate chemical reporters. Performed PPase directed evolution and the evolved mutant is able to pair with simplified CP with only 12-residues, which greatly promotes the practical utilization of this protocol. [40]

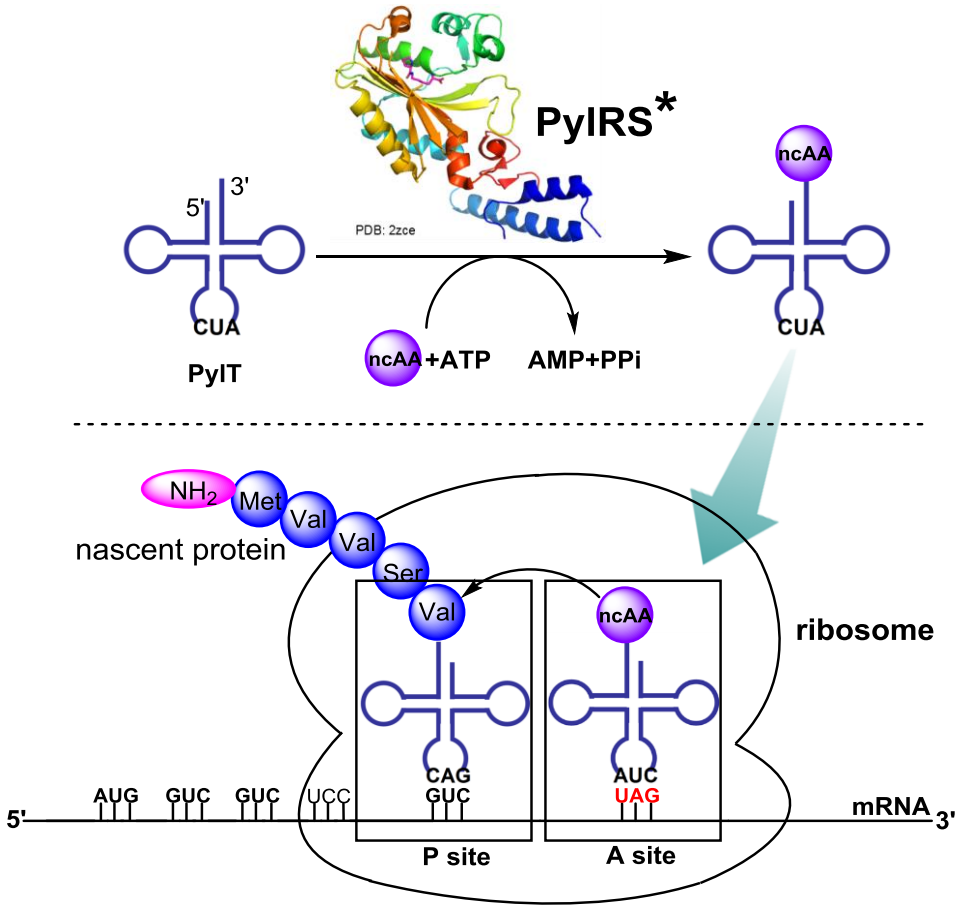


Figure 1-5. The ncAA incorporation strategy derived from the pyrrolysine incorporation machinery.

Genetic Incorporation of ncAAs

Pioneered by Schultz and coworkers, the side-specific incorporation of non-canonical amino acid (ncAA) into a protein was first accomplished in 1989.[48] This methodology makes use of an orthogonal tRNA, namely the tRNA does not have cross-reactivity with any of the endogenous aminoacyl-tRNA synthetase (aaRS). Additionally, the original anticodon of the tRNA is substituted to CUA so as to match up with UAG

codon on mRNA, translationally read through the mRNA amber stop codon. Because the tRNA is capable of reading through mRNA amber stop codon, it is also coined as amber suppressor tRNA. To charge desired ncAA, the tRNA is manipulated in vitro and ligated with a chemically synthesized deoxyribonucleotide-ncAA hybrid using DNA ligase, generating an full-length aminoacylated amber suppressor tRNA. Then the suppressor tRNA is subjected to a cell-free translation system in response to programmed mRNA amber codon site, in turn to biosynthetically produce ncAA-incorporated protein of interest. Though technically challenging, this approach has been very useful in the study of protein structure and function. Examples include the incorporation of ncAAs with modified side chines, fluorophores, native PTMs, photo-cross-linkers as well as chemically reactivity functionalities.[49, 50] In coupled with microinjection or electrotransfection technology, the chemically synthesized suppressor can be transferred into cells and undergo protein translation *in vivo*. Nevertheless, this approach is limited by poor protein yielding for two major reasons, first the accessibility and stability of the aminoacyl-tRNA adducts and second the bad molecule economy due to the stoichiometric use of the aminoacylated tRNAs that cannot be constantly delivered.[3]

An alternative route was further developed by Schultz and coworkers in 2001 to circumvent aforementioned limitations of using a chemically aminoacylated suppressor tRNA.[51, 52] The idea is to create an *in vivo* biosynthetic machinery in host cell in which the designated aminoacylated suppressor tRNA is continuously supplied within the expression organism during protein translation. This can be achieved by importing a foreign orthogonal tRNA/synthetase pair from a different organism if cross-species

aminoacylation is inefficient and the anticodon loop is not a key determinant of synthetase recognition. An eligible orthogonal $tRNA_{CUA}$/aaRS pair *Mj*TyrRS-*Mj*$tRNA^{Tyr}_{CUA}$ derived from *Methanococcus jannaschii* was introduced into *E. coli* that enabled the genetic incorporation of O-methyltyrosine at programmed amber stop codon site. The orthogonal synthetase was engineered to specifically recognize the designated ncAA, preventing aaRS cross-interaction with endogenous amino acids. By engineering the *Mj*TyrRS active site through rational design or directed evolution, the evolved *Mj*TyrRS mutants have greatly expanded ncAA substrate scope, containing mostly tyrosine or phenylalanine derivatives with diverse chemical or biophysical properties. However, in spite of the success of incorporation of ncAAs into proteins in *E. coli*, this method is not applicable to eukaryotes since the *Mj*TyrRS/$tRNA^{Tyr}_{CUA}$ fails to achieve orthogonality in eukaryotic systems. Therefore, to further extend this method in other organisms, in particular in mammalian cells, development of new aaRS/tRNA pair is highly desirable.

The opportunity was later been found after the discovery of pyrrolysine (Pyl), the 22nd genetic encoded amino acid.[53, 54] Derived from certain methanogenic archaea and Gram positive bacterium, Pyl's natural cognate $tRNA^{pyl}$ (PylT) uses CAU as anticodon in response to UAG amber codon on mRNA to incorporate Pyl during protein translation perhaps for its methyltransferase activity.[53, 54] This natural incorporation mechanism is equivalent to the genetic code expansion methodology using engineered *Mj*TyrRS/*Mj*$tRNA^{Tyr}_{CUA}$ pair. Moreover, both Pyl's cognate tRNA and aaRS, pyrrolysyl-tRNA synthetase (PylRS), are orthogonal to the bacteria and eukaryote systems. Pyl

system essentially fulfills the requirement of genetic ncAA incorporation methodology, consequently, PylRS/$tRNA^{pyl}_{CUA}$ pair is optimal to incorporation ncAA in both bacteria and eukaryotic cells. Interestingly, unlike the *Mj*TyrRS/*Mj*$tRNA^{Tyr}_{CUA}$pair, the anticodon region of $tRNA^{pyl}_{CUA}$barely affects its binding to PylRS, which allows mutations on the anticodon to make an alternative suppressor other than amber codon. It not only enables an opportunity to use a sense codon to encode ncAAs, which can boost up the suppression efficiency, but also makes it possible to incorporation two different ncAAs into proteins with the combination of the *Mj*TyrRS/*Mj* $tRNA^{Tyr}_{CUA}$ pair.[55] Structurally, PylRS has a large cavity at the catalytic site which allows PylRS to accommodate diverse lysine analogues substrates without any engineering. Along the same line of thought, further expand the substrate diversity can be achieved by active site engineering through rational design or directed evolution. At present, there are more than 50 ncAAs including lysine and phenylalanine derivatives have been incorporated.[50, 56]

CHAPTER II

GENETICALLY ENCODED ACRYLAMIDE FUNCTIONALITY*

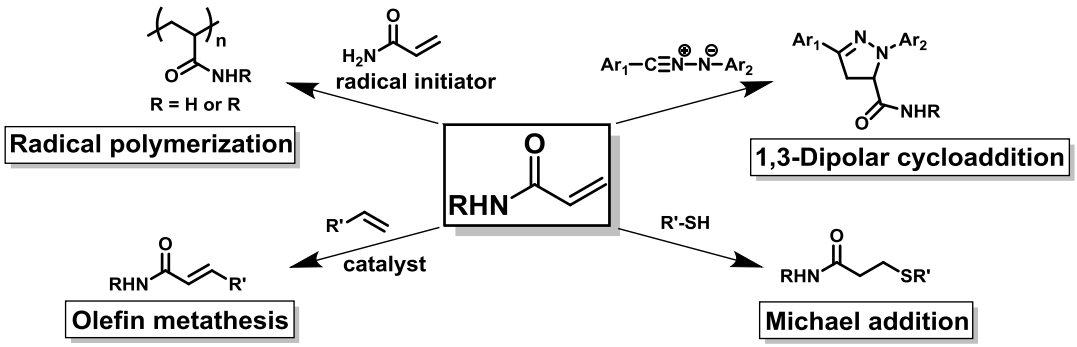
Introduction

Selective labeling has its great utility as a tool to interrogate physiological functions of proteins and develop biotechnological applications. To label a protein of interest site-specifically, the genetic noncanonical amino acid (ncAA) incorporation approach has drawn great attention due to its potentially broad applications.[51,57] Many ncAAs have been genetically incorporated into proteins in bacteria, yeast, mammalian cells and even animals.[50, 51, 58-61] Of these ncAAs, a few bioorthogonal functionalities are available, including ketone, alkyne, and azide that undergo biocompatible reactions such as hydrazone/oxime formation and azide-alkyne cycloaddition reactions for site-selective protein modifications.[62-71] Although ketone, alkyne, and azide are involved in diverse organic reaction types beyond the aforementioned bioorthogonal reactions, most of these reactions either cannot be realized in physiological conditions or lack selectivity in the biological system. A genetically encoded and simple chemical moiety that undergoes multiple types of biocompatible and selective reactions could have broad implications for selective labeling strategies *in vitro* and *in vivo*. One functional group that shows many facets of organic reactivity is the carbon-carbon double bond. Given the versatility in organic transformation, olefins have found extensive application in

*Reprinted with permission from Lee, Y. J., Wu, B., Raymond, J. E., Zeng, Y., Fang, X., Wooley, K. L., and Liu, W. R. (2013) A genetically encoded acrylamide functionality, *ACS Chem. Biol. 8*, 1664-1670. Copyright 2013 American Chemical Society.

synthetic organic chemistry. Various biocompatible olefin reactions have been demonstrated elsewhere. These include 1,4-addition, olefin metathesis, thiol-ene click, Diels-Alder, and 1,3-dipolar cycloaddition reactions.[10, 72] The olefin functionality has been genetically incorporated into proteins. Using an evolved amber-suppressing aminoacyl-tRNA synthetase-tRNA pair that was derived from *Methanocaldococcus jannaschii* tyrosyl-tRNA synthetase-$tRNA^{Tyr}$ pair, Zhang *et al.* were able to incorporate *O*-allyl-L-tyrosine into proteins in *E. coli*.[73] The same ncAA and two other aliphatic olefin-containing tyrosine analogs were also genetically encoded in *E. coli* using a mutant pyrrolysyl-tRNA synthetase (PylRS)-$tRNA^{Pyl}_{CUA}$ pair.[74] A photoclick reaction that exploits the nitrile imine-alkene cycloaddition reaction was applied to fluorescently label proteins with *O*-allyl-L-tyrosine incorporated in *E. coli*, albeit the labeling efficiency was low.[29] Five other aliphatic olefin-containing NAAs have also been encoded in *Saccharomyces cerevisiae* and used to undergo olefin metathesis on proteins.[75] In general, these aliphatic olefin-containing NAAs have low reactivity toward nitrile imine and are not capable of performing 1,4-addition under physiological conditions. Recently, several groups have used wild-type and evolved PylRS- $tRNA^{Pyl}_{CUA}$ pairs to site-specifically install ncAAs with a strain-promoted alkene into proteins.[27, 28, 76-79] Reacting rapidly with tetrazine or a tetrazole photo-cleavage product, these ncAAs contain a norbornene, transcyclooctene, or cyclopropene moiety that is a relatively large functional group. These NAAs may not be optimal options when minimal structural perturbation of proteins is necessary. Acrylamide is a small electron deficient olefin that undergoes several potentially biocompatible reactions, e.g., those shown in **Scheme 2-1**.

Here, we wish to report its genetic incorporation and its mediated diverse reactions for protein modifications. We envision that the acrylamide moiety would serve as a user-defined multifunctional chemical handle for different purposes of applications.



Scheme 2-1. Potential biocompatible reactions for acrylamide.

Results and Discussion

Due to the electrophilic nature of acrylamide, one may suspect it undergoes 1,4-addition with thiol nucleophiles such as intracellular glutathione. The kinetics of 1,4-addition of β-mercaptoethanol (β-ME) to acrylamide was investigated. β-ME instead of glutathione was used in our study due to the simple NMR spectrum of its 1,4-addition product with acrylamide. At pH 7.4, the second order rate constant determined was 0.0040 ± 0.0003 $M^{-1}s^{-1}$ (**Figure 2-1**) that is twice a reported rate constant (0.002 $M^{-1}s^{-1}$) of the reaction between glutathione and acrylamide at the same pH.[80] Providing the intracellular reduced glutathione concentration at 5 mM,[81] acrylamide will have at least a half-life of 10 h that is ample for protein expression. A parallel kinetic analysis at pH 8.0 inferred a second order rate constant of 0.013 ± 0.001 M^{-1} s^{-1} (**Figure 2-1**), indicating a linear dependence of the logarithm of the rate constant on pH. Therefore, an acrylamide-containing protein could be produced in cells and then modified at slightly alkali conditions *in vitro*. Encouraged by the initial kinetic studies, we synthesized N^{ϵ}-acryloyl-L-lysine (AcrK) and N^{ϵ}-crotonyl-L-lysine (CrtK) whose structures are shown in **Figure 2-3A** and then searched mutant PylRS-$tRNA^{Pyl}_{CUA}$ pairs for their genetic encoding in *E. coli*. Two more stable analogous NAAs, N^{ϵ}-propionyl-L-lysine (PrK) and N^{ϵ}-butyryl-L-lysine (BuK), were also synthesized and used during the initial search of mutant PylRS-$tRNA^{Pyl}_{CUA}$ pairs. Due to the structural similarity between N^{ϵ}-acetyl-L-lysine (AcK), PrK and AcrK, MmAcKRS1 that was evolved from *Methanosarcina mazei* PylRS for the AcK incorporation also recognizes PrK and AcrK.[82] To improve its efficiency, we

constructed a single-site MmAcKRS1 library that randomized Y384. Mutations at Y384 have been shown in mutant PylRS variants for other lysine derivatives.[83, 84] Clones from this library were screened against PrK using a protocol described previously.[85] Most of the positive clones converged to a same MmAcKRS1 variant that has the Y384W mutation and is coined as PrKRS. The following testing with AcrK indicated that the PrKRS-tRNA$^{\mathrm{Pyl}}_{\mathrm{CUA}}$ pair also mediates a high-level incorporation of AcrK at an amber codon (**Figure 2-2**). It was demonstrated previously that wild-type PylRS recognizes CrtK albeit at a low level. [86, 87] To improve the binding, a single site PylRS library with randomization at Y384 was constructed. Clones of this library were screened against BuK. Most positive clones converged to a single PylRS variant that has the Y384W mutation and is defined as BuKRS. BuKRS also recognizes CrtK (**Figure 2-2**).

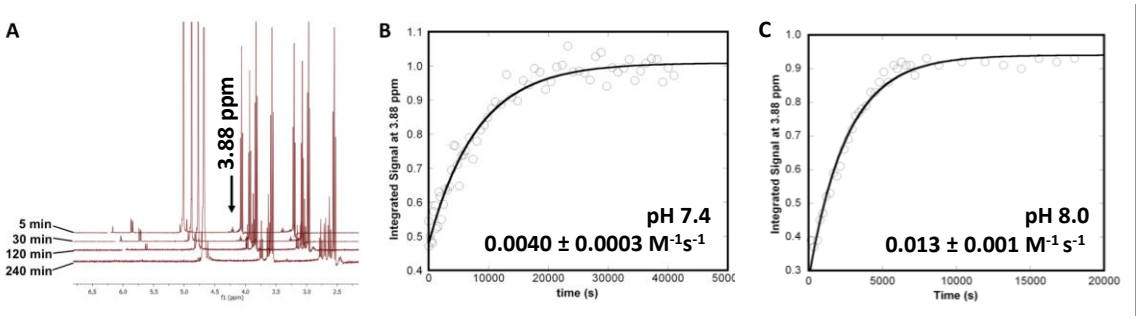


Figure 2-1. (**A**) The Michael addition reaction between acrylamide and β-ME characterized by NMR. The indicated chemical shift at 3.88 ppm is from protons on the β carbon of the product. Signals for this chemical shift were integrated and drawn against time shown in (**B**) and (C). (**B**) The time dependence of the Michael addition product formation at pH 7.4. (C) at pH 8.0. The data were fitted to the equation $p = A_1 - A_2 e^{-k \times t}$, where p represents the integrated chemical shift at 3.88 ppm, A_1 the final integrated chemical shift at 3.88 ppm, A_1-A_2 the initial integrated chemical shift at 3.88 ppm, and *k* the apparent pseudo first order reaction constant.

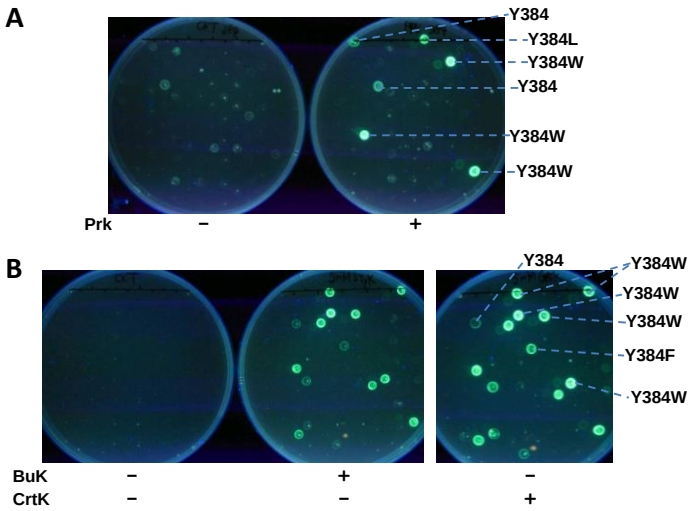


Figure 2-2. (**A**) Screening of the pBK-MmAcKRS1 plasmid library. The PrK Plate contained GMML agar/CKT, and PrK (2 mM). The control plate on the left had no PrK supplement. (**B**) Screening results of the pBK-PylRS plasmid library. The BuK Plate contained GMML agar/CKT, and BuK (2 mM). The CrtK plate contained CrtK (2 mM) instead of BuK. The control plate on the left had no NNA supplement. CKT denotes Cam (68 ug/mL), Kan (50 ug/mL), Tet (25 ug/mL).

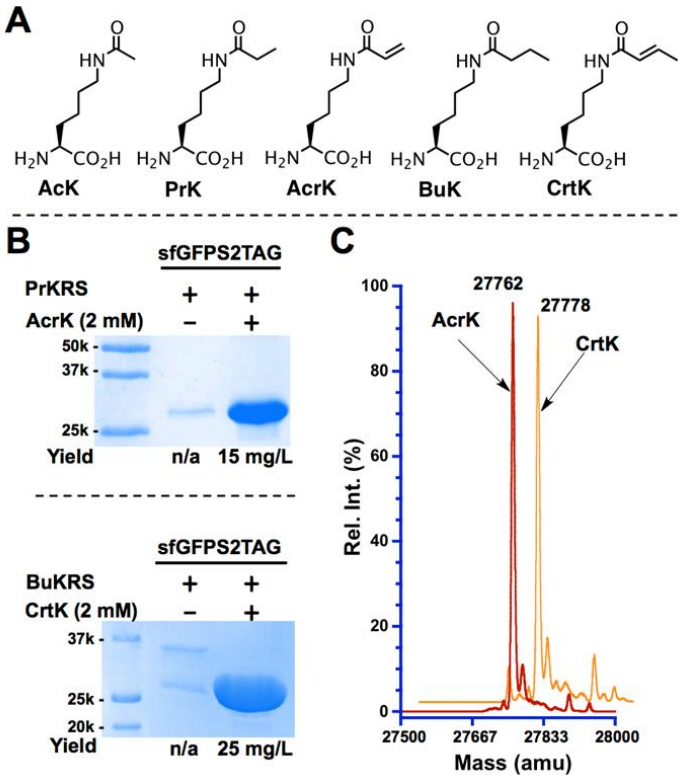


Figure 2-3: (**A**) Structures of AcK, PrK, AcrK, BuK, and CrtK. (**B**) The site-selective incorporation of AcrK and CrtK at the S2 position of sfGFP. (**C**) ESI-MS of the purified proteins.

For further characterization of the genetic incorporation of AcrK and CrtK, PrKRS and BuKRS genes were cloned separately into an optimized pEVOL vector under control of an *araBAD* promoter to afford plasmids pEVOL-PrKRS and pEVOL-BuKRS that also contain a $tRNA^{Pyl}_{CUA}$ gene under control of the *proK* promoter. Transforming BL21(DE3) cells that contained a reporter plasmid pET-sfGFP2TAG with pEVOL-PrKRS or pEVOL-BuKRS afforded cell strains for further protein expression. PET-sfGFP2TAG has a His-tagged superfolder green fluorescent protein (sfGFP) gene that contains an amber mutation at its S2 position and is under control of a IPTG-

inducible T7 promoter.[74] Growing cells and inducing with the addition of 0.2% arabinose, 1 mM IPTG, and 2 mM AcrK or CrtK resulted in the overexpression of sfGFP. When AcrK or CrtK is absent, only a very small amount of sfGFP could be observed (**Figure 2-3B**). Purified proteins were then subjected to electrospray ionization mass spectrometry (ESI-MS) analysis that showed the expected molecular weights (For AcrK, calculated: 27,764 Da; detected: 27,762 Da. For CrtK, calculated: 27,778 Da; detected: 27,778 Da). Although proteins were expressed for 8 h at 37 ^{o}C, there was no apparent 1,4-addition product that could be detected by ESI-MS, indicating intracellular stability for both AcrK and CrtK (**Figure 2-3C**).

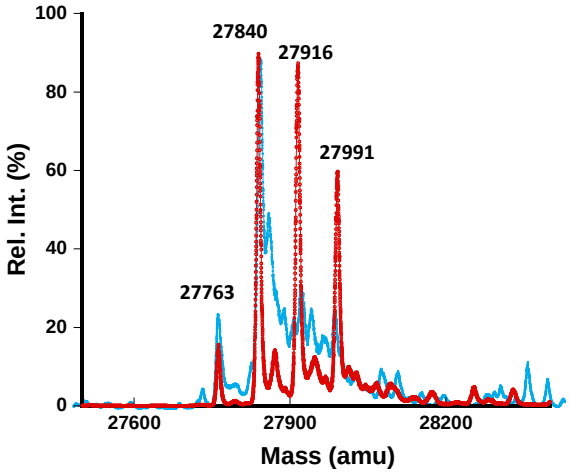


Figure 2-4. ESI-MS analysis of the β-ME modified sfGFP-AcrK. The red spectrum is for the product without further DTT treatment; the blue spectrum is for the product with further DTT treatment. The theoretical molecular weight of β-ME modified sfGFP-AcrK is 27,842 Da.

To demonstrate 1,4-addition for protein labeling, we chose to work with sfGFP that had AcrK incorporated at S2 (sfGFP-S2AcrK), due to the known steric effect of the methyl group to hinder the addition of a thiol nucleophile to CrtK. sfGFP-S2AcrK (39 μM) was incubated with *β*-ME (40 mM) at 37 ^{o}C overnight. pH 8.8 was used to elevate the reaction rate. The reaction proceeded smoothly with little original sfGFP-S2AcrK left after an overnight incubation (**Figure 2-4**). Besides the expected *β*-ME addition adduct, the ESI-MS spectrum also displayed two additional peaks that are 76 and 151 Da greater than the expected product, indicating that one and two additional *β*-ME molecules were covalently linked to sfGFP-S2AcrK. Since the *β*-ME adducts of proteins through the disulfide bond linkage were observed previously[82] and sfGFP-S2AcrK has two cysteine residues, these two additional peaks clearly resulted from the disulfide covalent linkage of *β*-ME to sfGFP-S2AcrK. These extra *β*-ME additions could be removed by further treatment with 5 mM dithiothreitol (DTT) for 1 h (**Figure 2-4**). A similar reaction was carried out to covalent link sfGFP-S2AcrK to methoxypolyethylene glycol thiol 5 kDa (mPEGSH5k). sfGFP-S2AcrK (50 μM) was incubated with mPEGSH5k (40 mM) at pH 8.8, 37 ^{o}C for 8 h and then treated with DTT (5 mM) for an additional hour. The following SDS-PAGE analysis showed two protein bands, with one around 5 kDa larger than the other. The relative intensities of the two protein bands indicated roughly 50% conversion (**Figure 2-5**). The same reaction with wild-type (WT) sfGFP did not lead to a modified protein with a significant signal that could be detected by SDS-PAGE.

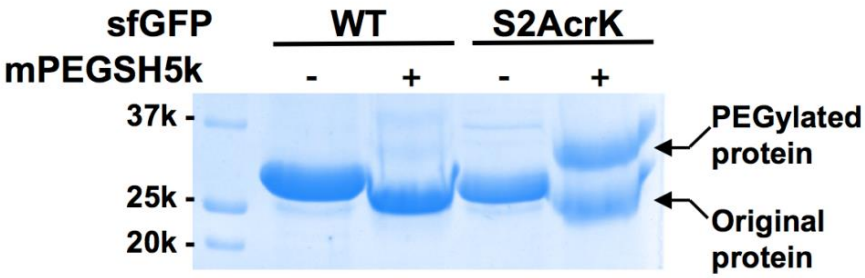


Figure 2-5: SDS-PAGE analysis of WT sfGFP and sfGFP-S2AcrK before and after their reactions with mPEGSH5k.

Radical polymerization has been applied to undergo selective protein modifications in physiological conditions.[88, 89] One typical chemical moiety for radical polymerization is acrylamide. To demonstrate that the acrylamide moiety in sfGFP-S2AcrK can serve as a monomer for radical polymerization, a copolymer hydrogel was prepared by mixing sfGFP-S2AcrK (110 μM), acrylamide (15%), and bis-acrylamide (0.5%) and initiating polymerization with the addition of tetramethylethylenediamine (0.3%) and ammonium persulfate (3%). The reaction solution was loaded to a well of a native PAGE gel and allowed to solidify. By this process, sfGFP-S2AcrK was covalently immobilized in the hydrogel and did not migrate during the following electrophoresis process (**Figure 2-6**). However, a similar test with WT sfGFP showed that the majority of the protein did not form covalent linkages with hydrogel. Moreover, as indicated by the fluorescence maintenance of the immobilized sfGFP-S2AcrK, this radical copolymerization reaction is compatible with preservation of the native protein state of sfGFP.

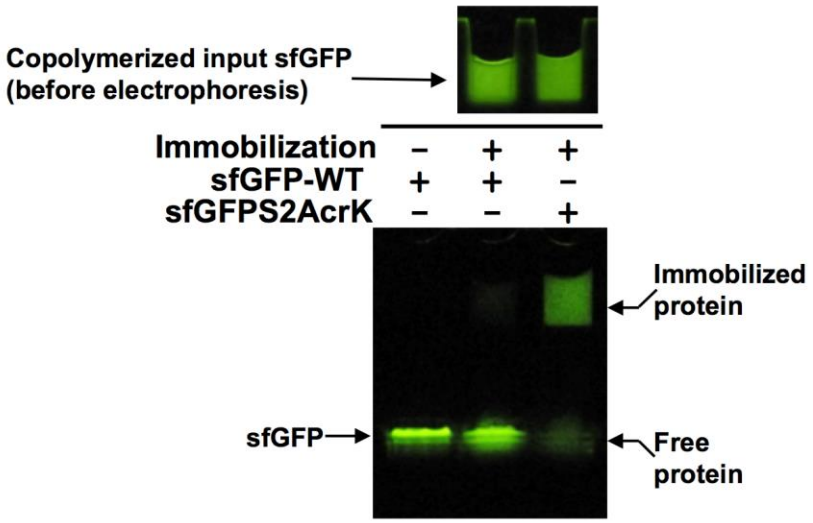


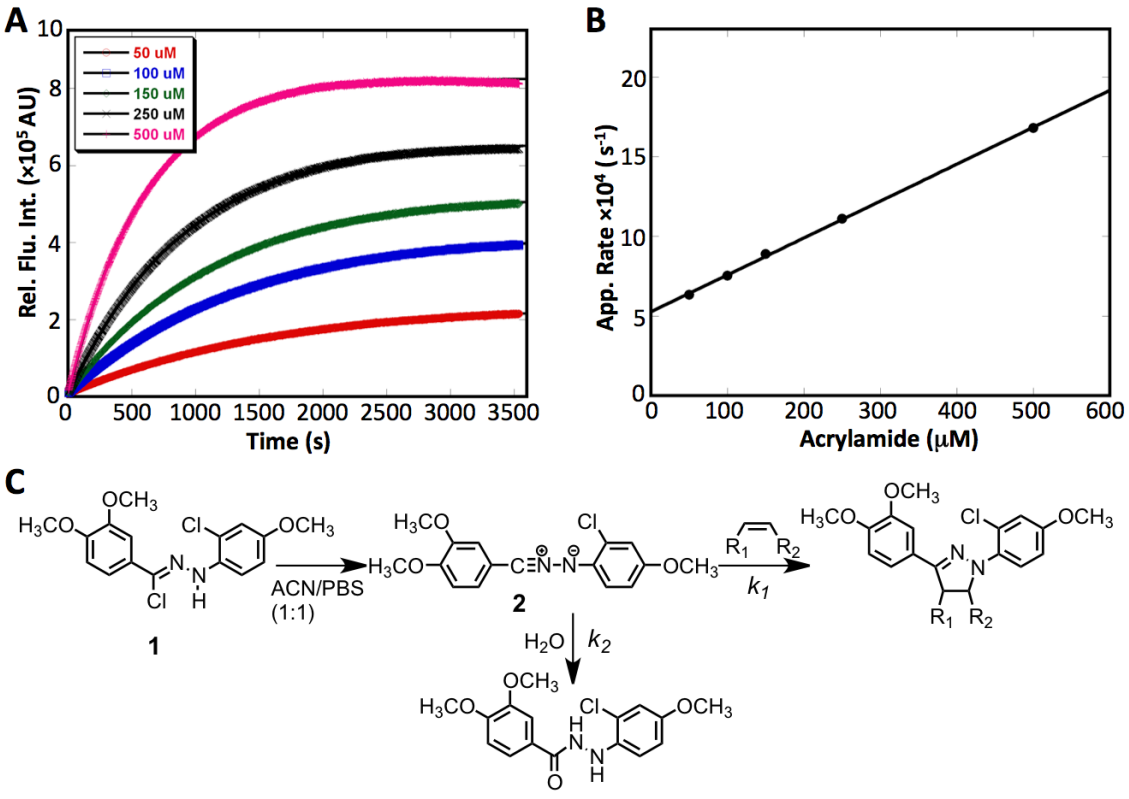
Figure 2-6: A native PAGE analysis of coimmobilized sfGFP-S2AcrK in an acrylamide hydrogel. WT sfGFP was loaded to the first lane to serve as the standard.

It was demonstrated previously that olefin reacts with diaryl nitrile imine to generate a fluorescent pyrazoline at an appreciable rate in mild aqueous conditions.[25] Using a hydrazonoyl chloride **1** as a precursor that undergoes rapid dissociation in water to form a diaryl nitrile imine (**2**), Kaya *et al.* demonstrated a genetically encoded norbornene moiety could be used to fluorescently label proteins.[27] However, a systematic kinetic investigation of the olefin reaction with diaryl nitrile imine is absent. **1** was synthesized according to the literature procedures. Reactions of **1** (5 μM) with acrylamide were examined by detecting the fluorescent emission from the product. The time-dependent fluorescent emission increase at different acrylamide concentrations is presented in **Figure 2-7A**. They all show single exponential curvatures. The determined apparent reaction rate constants have a linear dependence on the acrylamide concentrations with a significant y-axis intercept (**Figure 2-7B**). The reactions between

1 and other olefins shown in **Figure 2-7C** all display equivalent kinetic features (**Figures 2-5 - 2-11**). **Figure 2-7A** also shows very different fluorescent intensities for the final products at different acrylamide concentrations. All these observations indicate that **2** underwent two parallel reactions in an aqueous buffer, one with olefin and the other with water (**Figure 2-7C**).[25] Based on the mechanism in **Figure 2-7C**, the pyrazoline formation follows the kinetic **Equation 2-1**, where k_1 is the second order rate constant of the reaction between **2** and an olefin, k_2 the hydrolysis rate constant of **2**, [O] the olefin concentration, and [NI] the concentration of **2**. Using this equation to analyze all the data resulted in k_1 and k_2 for various olefins shown in the table in **Figure 2-7C**. All determined k_2 values are similar, supporting the parallel reaction mechanism. The results also indicated that acrylamide reacted with diaryl nitrile imine much faster than with all the other tested olefins, specifically 3 times faster than norbornene. This observation is consistent with a previous report.[90] The hydrolysis of **2** was also analyzed by examining the UV absorbance decay of **2** at 315 nm in a buffer that contained acetonitrile and PBS with a ratio of 1:1 (**Figure 2-13**) and gave a first order rate constant of $0.00034 \pm 0.00001\ s^{-1}$, similar to the k_2 values shown in **Figure 2-7C**.

Equation 2-1

$$p = \frac{k_1 \times [O] \times [N1]}{k_1 \times [O] + k_2} \times \left(1 - e^{-(k_1 \times [O] + k_2) \times t}\right)$$



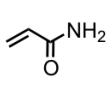
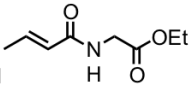
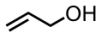
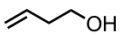
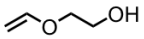
Alkene						
$k_1/M^{-1}s^{-1}$	2.30±0.03	0.75 ± 0.03	0.15±0.01	0.075±0.007	0.038±0.002	0.039 ±0.004
k_2 x 10^4 /s^{-1}	5.3±0.1	4.4±0.1	4.6±0.1	6.0±0.2	4.2±0.2	5.2±0.2

Figure 2-7: (**A**) The pyrazoline formation in reactions between 5 μM **1** and varying concentrations of acrylamide. Reactions were monitored by fluorescence emission at 480 nm, with an excitation light at 318 nm. (**B**) The linear dependence of apparent rates of the reaction between **1** and acrylamide on the acrylamide concentrations. (**C**) Mechanism of the reaction between **1** and olefin in aqueous conditions and the finally determined rate constants for different olefins.

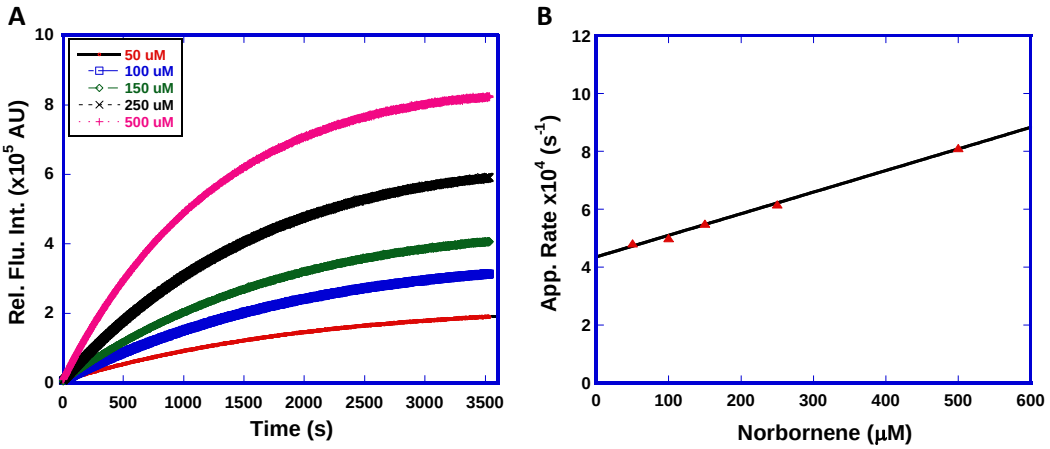


Figure 2-8. (**A**) The formation of pyrazoline from the reaction between 5 μM **1** and varying concentrations of norbornene monitored by the fluorescence increase at 480 nm. Samples were excited at 318 nm. (**B**) The linear dependence of apparent rates of the reaction between **1** and norbornene on the norbornene concentrations.

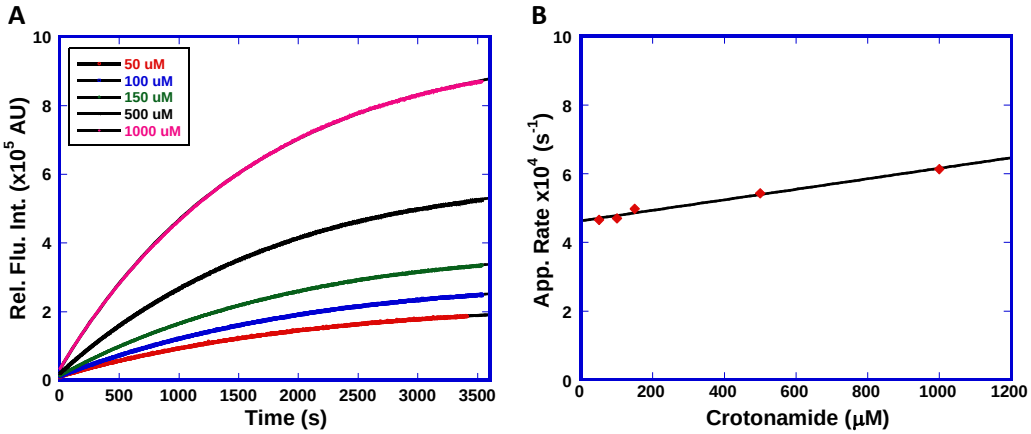


Figure 2-9. (**A**) The formation of pyrazoline from the reaction between 5 μM **1** and varying concentrations of crotonamide monitored by the fluorescence increase at 480 nm. Samples were excited at 318 nm. (**B**) The linear dependence of apparent rates of the reaction between **1** and crotonamide on the norbornene concentrations.

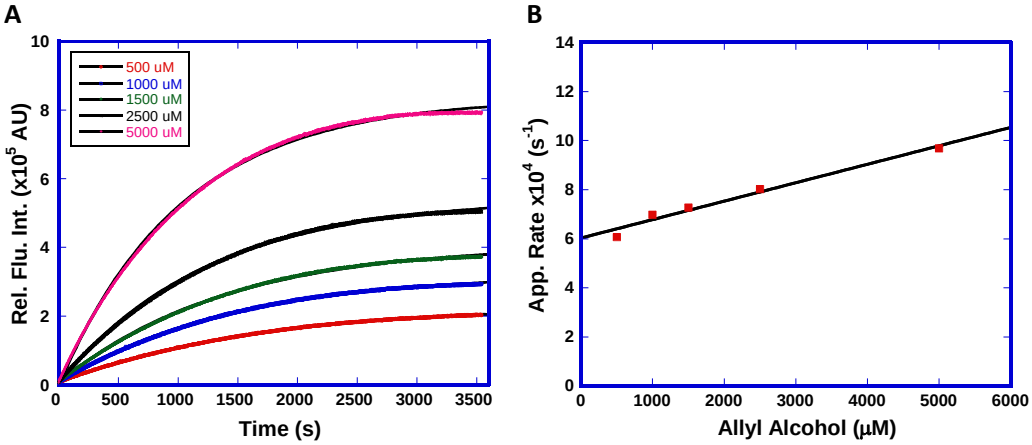


Figure 2-10. (**A**) The formation of pyrazoline from the reaction between 5 μM **1** and varying concentrations of allyl alcohol monitored by the fluorescence increase at 480 nm. Samples were excited at 318 nm. (**B**) The linear dependence of apparent rates of the reaction between **1** and allyl alcohol on the allyl alcohol concentrations.

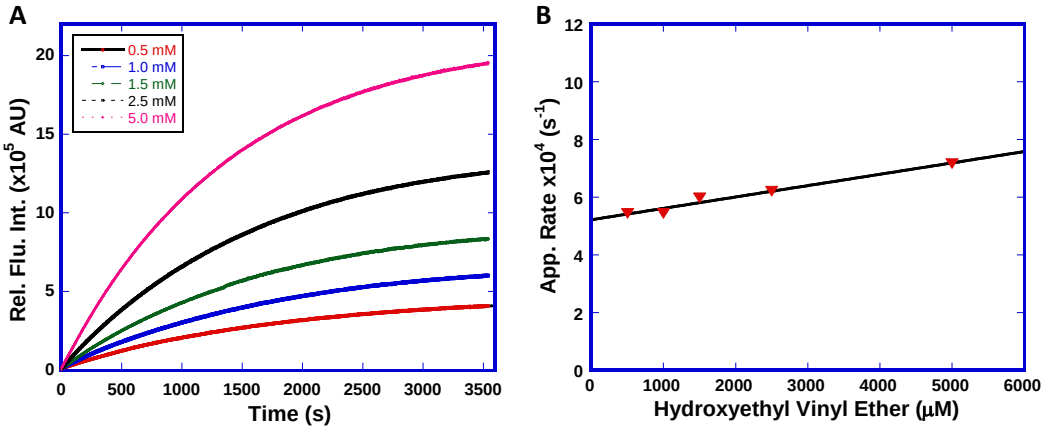


Figure 2-11. (**A**) The formation of pyrazoline from the reaction between 5 μM **1** and varying concentrations of hydroxyethyl vinyl ether monitored by the fluorescence increase at 480 nm. Samples were excited at 318 nm. (**B**) The linear dependence of apparent rates of the reaction between **1** and hydroxyethyl vinyl ether on the hydroxyethyl vinyl ether concentrations.

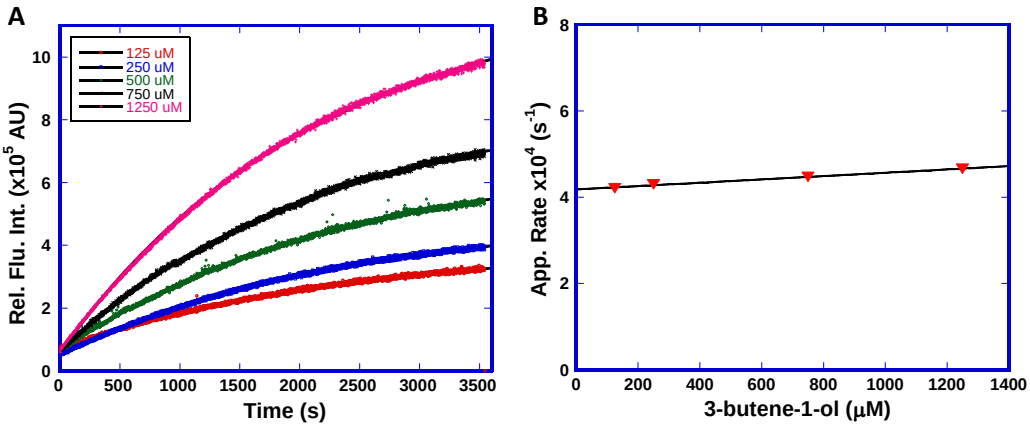


Figure 2-12. (**A**) The formation of pyrazoline from the reaction between 5 μM **1** and varying concentrations of 3-buten-1-ol monitored by the fluorescence increase at 480 nm. Samples were excited at 318 nm. (**B**) The linear dependence of apparent rates of the reaction between **1** and 3-buten-1-ol on the 3-buten-1-ol concentrations.

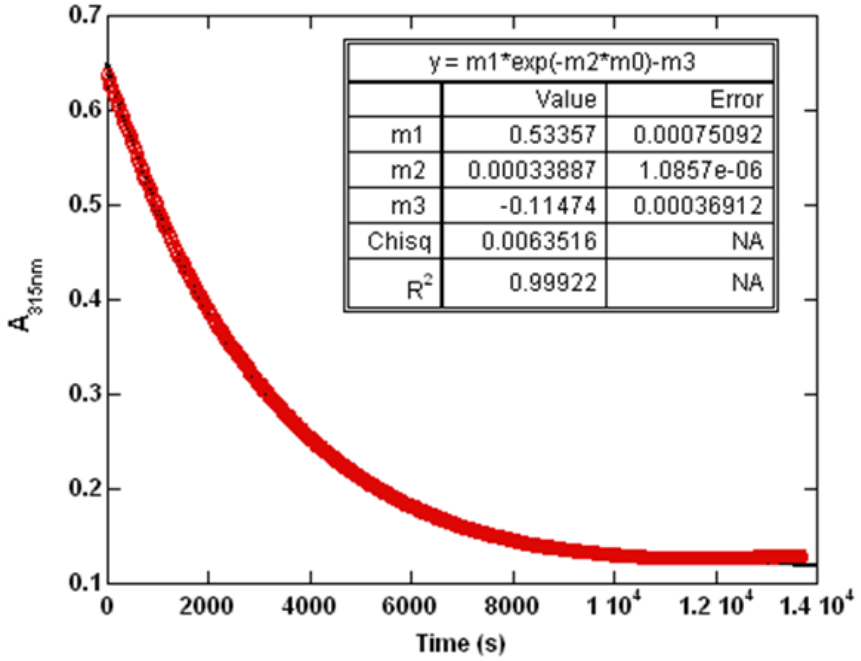


Figure 2-13. Hydrolysis of nitrile imine in the ACN/PBS (1:1) buffer monitored by UV absorption decrease at 315 nm.

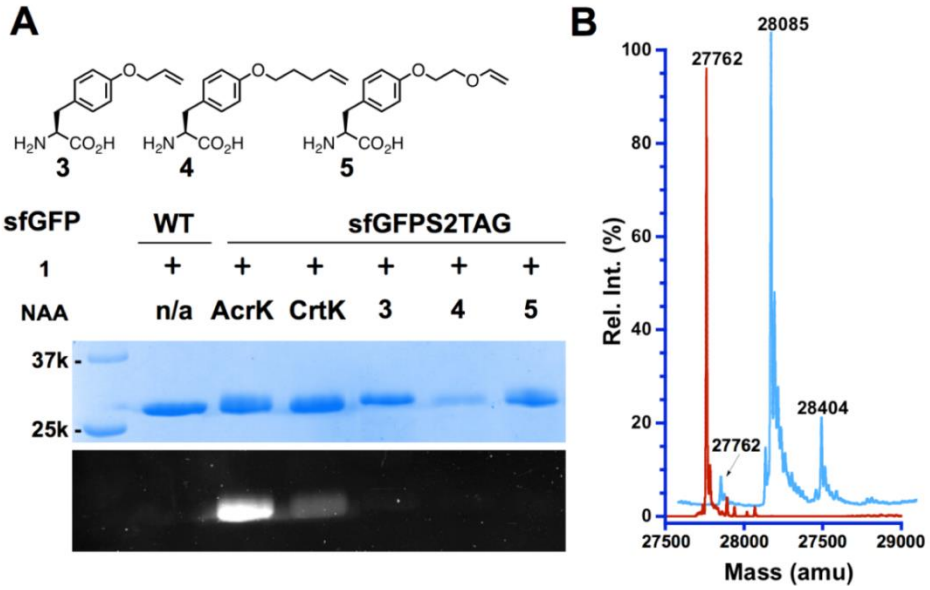


Figure 2-14: (**A**) Fluorescent labeling of sfGFP variants with **1**. The image at the top panel was the Coomassie blue stained gel; the bottom panel was the fluorescent image of the same gel before Coomassie blue staining. The image was detected by BioRad ChemiDoc XRS+ system under UV irradiation. (**B**) The ESI-MS spectra of the original (in red) and labeled (in blue) sfGFP-S2AcrK proteins.

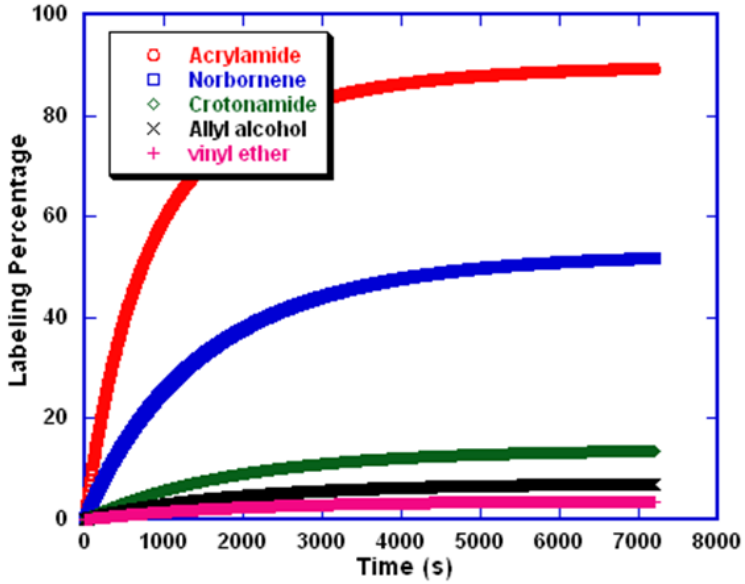


Figure 2-15. Simulated fluorescent labeling of 1 μM proteins with different olefins incorporated with 500 μM **1**.

To demonstrate the nitrile imine-olefin cycloaddition for the fluorescent protein labeling, we expressed sfGFP variants that were incorporated with AcrK, CrtK, and three other olefin-containing tyrosine analogs, illustrated in **Figure 2-14A** as **3**-**5**, at the sfGFP S2 position. The genetic incorporation of **3** and **4** was demonstrated previously using a rationally designed PylRS mutant.[74] This same mutant also recognizes **5**. sfGFP variants (14 μM) were incubated with **1** (70 μM) in a Tris-HCl buffer (pH 7.4) at RT for 1 h, fully denatured, and then analyzed by SDS-PAGE. The cyclization product of sfGFP-S2AcrK and sfGFP with CrtK incorporated at S2 (sfGFP-S2CrtK) displayed detectable fluorescent emission when irradiated by UV light (**Figure 2-14A**). The signal from sfGFP-S2CrtK is relatively weak and was estimated as 17% of that from sfGFP-S2AcrK. This was consistent with the simulation results for reactions of **1** with acrylamide and crotonamide (**Figure 2-15**). However, sfGFP variants incorporated with **3-5** did not show significant fluorescent signals after their reactions with **1**. Prolonging the reaction time did not appreciably improve the labeling efficiency. Giving the low cyclization reactivity of **3-5** with **1**, **1** preferentially underwent hydrolysis, leading to low labeling efficiency (**Figure 2-15**). To verify the cyclization product of sfGFP-S2AcrK, it was subjected to ESI-MS analysis. The major detected peak (28,085 Da) agrees well with the calculated labeling product (28,083 Da) (**Figure 2-14B**). The original protein could be barely detected.

To explore the possibility of labeling the acrylamide moiety with **1** in living cells, plasmid pETDuet-OmpXTAG was constructed for expressing the *E. coli* outer membrane protein OmpX incorporated with AcrK (OmpX-AcrK). The OmpX gene in

pETDuet-OmpXTAG has an AAAAXAA (A denotes alanine; X denotes an amber mutation) insertion between its two extracellular OmpX residues S53 and S54. Transforming BL21(DE3) cells with pEVOL-PrKRS and pETDuet-OmpXTAG and growing the transformed cells in LB medium supplemented with 1 mM IPTG and 5 mM AcrK afforded OmpX-AcrK overexpression (**Figure 2-16A**). Incubating the lysate of cells expressing OmpX-AcrK with **1** (0.5 mM) at RT for 1 h led to a detectable fluorescent band in a SDS-PAGE gel that corresponded well with the Coomassie blue stained OmpX-AcrK in the same gel. While many proteins in the lysate were present at higher expression levels, in-gel fluorescence imaging showed that only the expressed OmpX-AcrK could be specifically labeled, giving nearly no detectable background (**Figure 2-16A**). With the success of the cell lysate labeling, experiments were conducted to fluorescently label living cells with **1** by incubating the PBS-washed *E. coli* that overexpressed OmpX-AcrK with **1** (0.5 mM) at RT for 1 h. The cells were then washed with an isotonic saline solution and subjected to confocal microscopy imaging. *E. coli* cells expressing OmpX-AcrK showed detectable fluorescence (**Figure 2-16B**), indicating the cycloaddition is biocompatible and highly selective. The low variation in florescence lifetime profiles for all cells indicates a uniformity of dye states consistent with single-site covalent binding of the dye to the cell membrane and not of multiple dye environments typical of electrostatically bound or internalized dyes (**Figure 2-17**). We also noticed that the labeled OmpX-AcrK is not uniformly distributed on the *E. coli* cell membrane. This preferential accumulation of OmpX at the two poles of *E. coli* needs to be further investigated. The same reaction to label *E. coli* cells without OmpX-AcrK did

not yield any fluorescent cells that could be detected by confocal fluorescent microscopy (**Figure 2-18**).

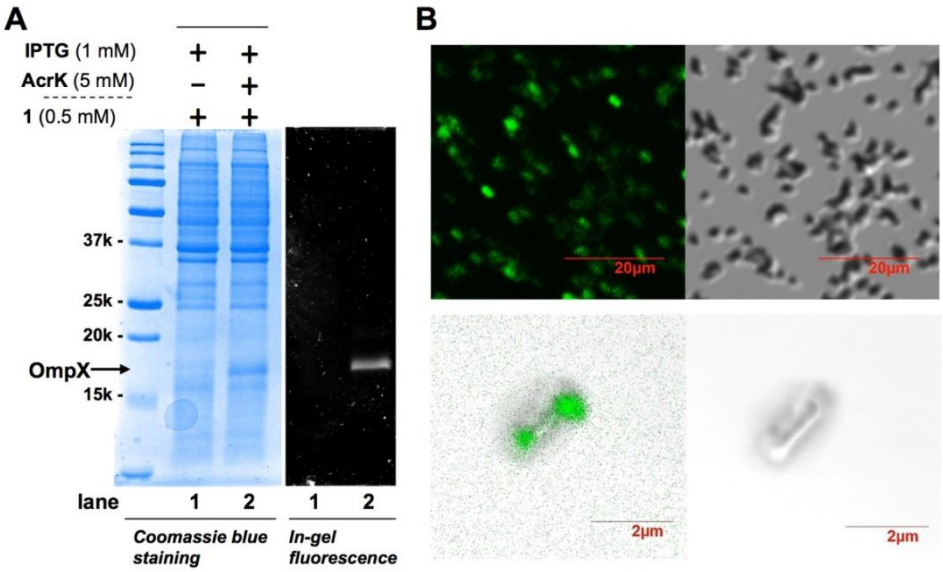


Figure 2-16: (**A**) Labeling OmpX-AcrK in a cell lysate with **1**. The first lane shows a control without expressed OmpX-AcrK. (**B**) Confocal fluorescent imaging of cells overexpressing OmpX-AcrK and labeled with **1**. The top left image shows the fluorescent imaging of multiple cells; the top right is the corresponding DIC image; the bottom right is the DIC image of a single cell; the bottom left is the overlay of the DIC and the confocal fluorescent image of a single *E. coli*.

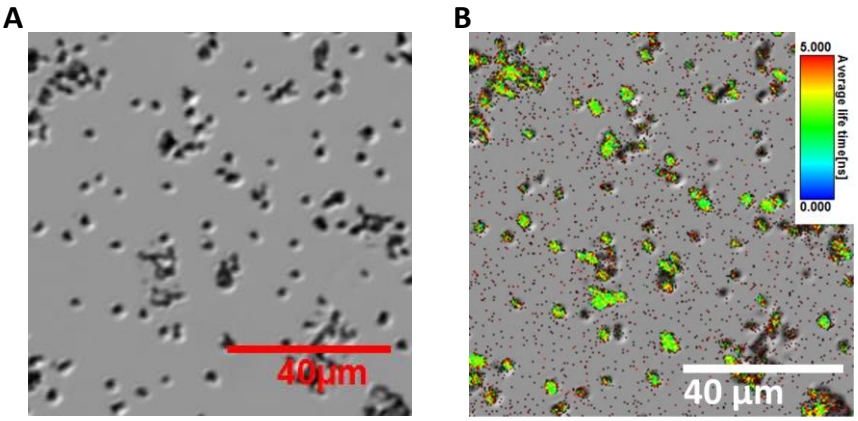


Figure 2-17. (A) DIC images of a *ca.* 100 individual labeled *E. coli.* **(B)** FLIM imaging overlaid on DIC. Note the low variation in lifetime profiles for all cells present, indicating a uniformity of dye states consistent with single site binding and not indicative of multiple dye environments typical of electrostatically bound or internalized dyes.

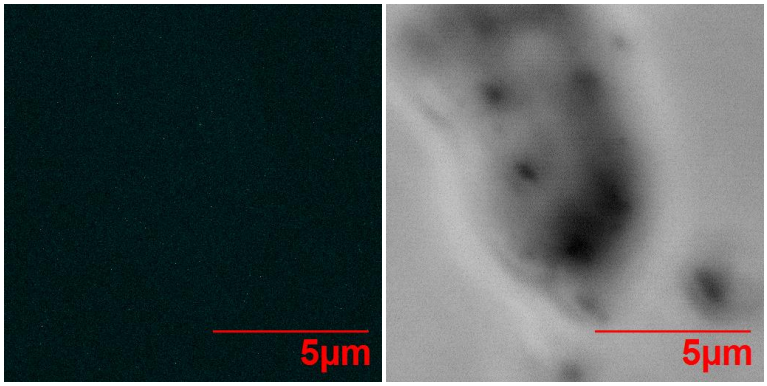


Figure 2-18. Fluorescence confocal and DIC image of single *E. coli* from the control group, which is not expressing OmpX-AcrK. Fluorescence collection parameters were similar to the OmpX-AcrK provided in figure 2-16B with identical scan rates, excitation power and monochrometer settings. Fluorescence intensity was not above background nor was it co-localized with the *E. coli* control group.

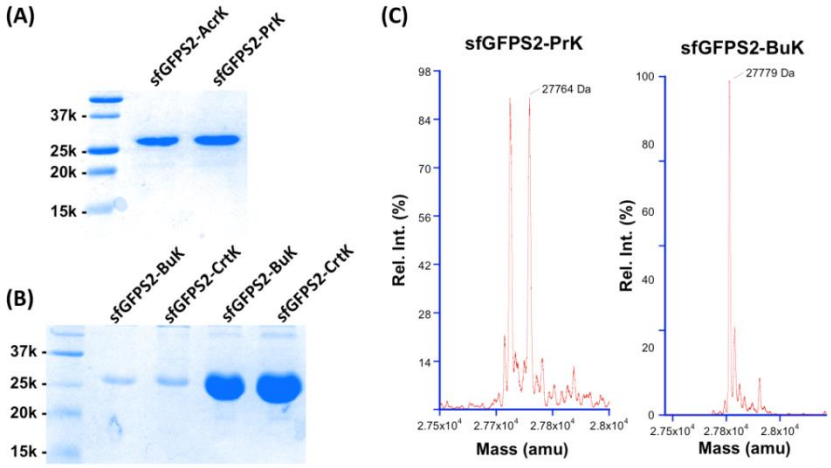


Figure 2-19. (A) The genetic incorporation of AcrK and PrK into the S2 position of sfGFP. Proteins were expressed in cells that contained pEVOL-PrKRS and pET-sfGFPS2TAG and were grown in LB media supplemented with 1 mM AcrK or PrK. **(B)** The genetic incorporation of BuK and CrtK into the sfGFPS2 position. Proteins in third and fourth lanes were expressed in cells that contained pEVOL-BuKRS and pET-sfGFPS2TAG and were grown in LB media supplemented with 1 mM BuK or CrtK. The first two lanes indicated the sfGFP expression levels using wild type PylRS. **(C)** ESI-MS spectra of sfGFP2 with PrK incorporated at S2 (sfGFPS2-PrK) and sfGFP with BuK incorporated at S2 (sfGFPS2-BuK). Calculated molecular weights are 27,766 Da for sfGFPS2-PrK and 27,780 Da for sfGFPS2-BuK.

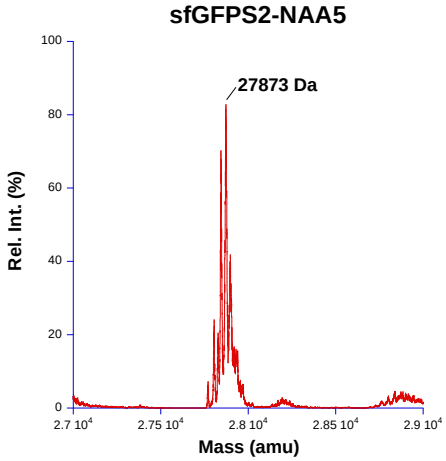


Figure 2-20. ESI-MS spectrum of sfGFPS2 incorporated with NAA **5**. The theoretical molecular weight of sfGFPS2-NAA**5** is 27,874 Da.

Conclusion

In summary, we report a method for the genetic installation of a small electron deficient olefin, acrylamide, into recombinant proteins in *E. coli*. Being a multifaceted chemical group for functionalizing proteins, the acrylamide moiety is relatively stable in physiological conditions and undergoes a number of biocompatible and unique reactions. We demonstrate that acrylamide-containing AcrK can be genetically encoded using an evolved PrKRS-$tRNA_{CUA}^{Pyl}$ pair and the genetically incorporated AcrK mediated efficient 1,4-addition, radical copolymerization, and 1,3-diplor cycloaddition reactions on proteins. These unique reactions provide a direct route to applications that include synthesizing PEGylated therapeutic proteins for disease intervention and generating protein-immobilizing hydrogels for ELISA, microarray/microfluidic detections. One reaction is proven viable for site-selective protein labeling in living cells for protein functional investigations. The 1,4- addition reaction to PEGylate an acrylamide in a protein produces a selective and simple thioether linkage that is likely well tolerated in the biological system, potentially providing an advantage over the existing PEGylation techniques. The nitrile imine-acrylamide cycloaddition reaction presented in this study displays fast kinetics and high selectivity and is performed without catalyst or light activation. This reaction can be potentially used as an excellent alternative for *in vivo* protein labeling to the most popular click reaction type, the azide-alkyne cycloaddition that requires a Cu(I) catalyst and therefore may disrupt certain cellular functions.[14, 91] The current study also identified mutant PylRS variants that allow the synthesis of proteins incorporated with PrK, BuK, and CrtK. Propionylation, butylation, and

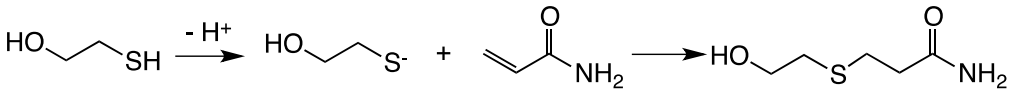
crotonylation are naturally existing posttranslational lysine modifications of proteins.[92, 93] In line with others,[86, 87, 94] we have provided approaches that can be applied to investigate these novel posttranslational modifications.

Experimental Section

Kinetic Analysis of Acrylamide-Thiol 1,4-Addition Reaction

Reaction conditions: In a NMR tube, 5 mM acrylamide in 0.1 M sodium phosphate buffer (pH 7.4 or pH 8.0) was added 30 mM β-ME solution. A trace amount of acetonitrile was added as the internal 1H standard. The reaction was maintained at RT and subjected to 1H NMR analysis at different times. NMR spectra were acquired every 5 minutes during the first 2 h and acquired every 10-30 minutes after the first 2 h. The appearance of a chemical shift at 3.88 ppm represents the sp^3 protons on the β carbon of the product. Signals from this chemical shift were integrated and drawn against time shown in SI Figures 1 & 2. These time-dependent chemical shift signal increase data were fitted to a single exponential equation to give pseudo first order reaction rate constants as 0.000120 ± 0.000009 s^{-1} at pH 7.4 and 0.00038 ± 0.00002 s^{-1} at pH 8.0. Since k' = k × [β-ME], where k' is the pseudo first order rate constant, k the second order rate constant, [β-ME] the β-ME concentration (30 mM), the first calculated second order rate constants are 0.0040 ± 0.0003 s^{-1} at pH 7.4 and 0.013 ± 0.001 s^{-1} at pH 8.0. If the acrylamide-thiol 1,4-addition reaction followed the reaction mechanism shown in Scheme 1, the logarithm of the determined second order rate constants between acrylamide and β-ME at different pH values would be linearly dependent on pH. Based

on this assumable linear dependence, theoretically the second order rate constant at pH 8.0 would be 4 times of that at pH 7.4, which is close to the 3.1 times difference determined experimentally.



Scheme 2-2. The pH dependent acrylamide-thiol 1,4-addition reaction.

Identification of PylRS Mutants

Library Construction

The plasmid pBK-MmAcKRS1 was derived from pBK-MmPylRS plasmid reported previously.[85] The gene *MmAcKRS1* was amplified from pAraCB2[82] using a 5'-end primer containing a NdeI site: 5'-gaatcccatatggataaaaaaccactaaacactctg-3' and a 3'-end primer containing a NsiI site: 5'-gtttgaaaatgcatttacaggttggtagaaatccc-3'. The amplified gene was cloned into pBK plasmid to replace the original synthetase gene. The resulting plasmid, pBK-MmAcKRS1, has synthetase expression under the control of constitutive *E. coli glnS* promoter and terminator.

To identify mutant synthetases specific for recognizing AcrK and CrtK, a MmAcKRS1 library was constructed with randomization at Y384. Two primers mAcKRS-Y384NNK-F: 5'-gattcctgcatggctNNKggggataccttgatg-3' and mAcKRS-Y384NNK-R: 5'-catcaagggtatccccMN Ngaccatgcaggaatc-3' (N=A/C/T/G; T, K=G/T; M=C/A) were used to amplify pBK-MmAcKRS1. The PCR product was phosphorylated

with T4 PNK, ligated with T4 DNA ligase and heat killed before transformation. The same primers were used to create a pBK-MmPylRS plasmid library with randomization at Y384 of its encoded *MmPylRS* gene.

Library Screening

The pBK-MmAcKRS1 or pBK-MmPylRS plasmid library and the positive selection plasmid pY+ were used to cotransform *E. coli* Top10 cells. The plasmid pY+ encodes the $tRNA_{CUA}^{Pyl}$ gene (PylT), a chloramphenicol acetyltransferase gene with an amber mutation at position D112, a T7 RNA polymerase gene with two amber mutations at its N-terminus, and a green fluorescent protein (GFP) gene under control of a T7 promoter.[85] The transformed cells were plated on LB agar plates supplemented with kanamycin (Kan, 50 μg/ mL) and tetracycline (Tet, 25 μg/ mL). The afforded single colonies were transferred to 96-well plates in which each well contains LB broth with 50 ug/mL Kan and 25 ug/mL Tet and incubated at 37^{o}C to allow cell propagation. After 6 hours, the cells on the 96-well plate were dotted on a screening plate containing GMML agar, Cam (68 μg/mL), Kan (50 μg/mL), Tet (25 μg/mL) and a NAA (2 mM). A plate without NAA served as the control. Plates were incubated at 37^{o}C for overnight. The cell phenotypic differences of viability and fluorescence intensity were analyzed and photographed. The clones with strong fluorescence on the NAA-containing plate and low viability and fluorescence on the control plate were selected as the active clones and were amplified from the corresponding wells in the 96-well plate. The pBK plasmid of

the active clones was isolated and sequenced. Most of the active clones were converged to contain a Y384W mutation (**Figure 2-2**).

sfGFP Protein Expression and Purification

DNA Sequence of *sfGFP*

atgtagaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtcc
gtggagagggtgaaggtgatgctacaaacggaaaactcacccttaaatttatttgcactactggaaaactacctgttccgtggcc
aacacttgtcactactctgacctatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgcc
atgcccgaaggttatgtacaggaacgcactatatctttcaaagatgacgggacctacaagacgcgtgctgaagtcaagtttgaa
ggtgatacccttgttaatcgtatcgagttaaagggtattgattttaaagaagatggaaacattcttggacacaaactcgagtacaac
tttaactcacacaatgtatacatcacggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacgttgaag
atggttccgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgt
cgacacaatctgtcctttcgaaagatcccaacgaaaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattaca
catggcatggatgagctctacaaaggatcccatcaccatcaccatcactaa

Protein sequence of sfGFP:

MASKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFI**C**TTGKL
PVPWPTLVTTLTYGVQ**C**FSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY
KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGI
KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKR
DHMVLLEFVTAAGITHGMDELYKGSHHHHHH

*The ncAA incorporation site is highlighted in red.

Plasmid pEVOL-PylT-PrKRS Construction

The pEVOL-PylT-PrKRS plasmids were derived from the pEVOL-PylT plasmid reported previously. The PrKRS gene was PCR amplified from the identified pBK-PrKRS plasmids using primers pEVOL-PylRS-SpeI-F: 5'-gaggaaactagtatggataaaaaaccactaaacactctg-3' and pEVOL-PylRS-SalI-R: 5'-tgatggtcgactcacaggttggtagaaatcccgtt-3'. The PCR product was digested by SpeI and SalI restriction enzymes, and ligated to a similarly digested pEVOL-PylT plasmid.

Plasmid pEVOL-PylT-BuKRS Construction

The pEVOL-PylT-BuKRS plasmid was derived from the pEVOL-PylT plasmid reported previously.[74] The *BuKRS* gene was PCR amplified from the identified pBK-BuKRS plasmids using primers pEVOL-PylRS-SpeI-F: 5'-gaggaaactagtatggataaaaaaccactaaacactctg-3' and pEVOL-PylRS-SalI-R: 5'-tgatggtcgactcacaggttggtagaaatcccgtt-3'. The PCR product was digested by SpeI and SalI restriction enzymes, and was ligated to a similarly digested pEVOL-PylT plasmid.

Protein Expression in *E. coli* and Protein Purification

The plasmid pET-sfGFPS2TAG reported previously was used to express sfGFP proteins for NAA incorporation and characterization.[74] *E. coli* BL21(DE3) cells were transformed with pEVOL-pylT-AcrKRS and pET-sfGFPS2TAG and plated on LB agar plate containing chloramphenicol (Cam) (34 μg/mL) and ampicillin (Amp) (100 μg/mL).

A 5 mL o/n culture was prepared from a single colony. This overnight culture was used to inoculate 500 mL of LB medium supplemented with Cam and Amp. Cells were grown at 37 ^{o}C in an incubator (250 r.p.m.) and the protein expression was induced with 1 mM IPTG, 0.2% arabinose and 2 mM AcrK when OD_{600} reached 0.6. After 8h induction, cells were harvested, resuspended in a lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and sonicated in an ice/water bath four times (4 min each, 10 min interval to cool the suspension below 10 ^{o}C between each pulse). The cell lysate was centrifuged (60 min, 11,000 g, 4 ^{o}C). The supernatant was incubated with 3 mL Ni-NTA resin (Qiagen) (2 h, 4 ^{o}C). The slurry was then loaded to a column and the protein-bound resin was washed with 30 mL of the wash buffer. Protein was finally eluted by the lysis buffer with 250 mM imidazole. Eluted fractions were collected, concentrated and buffer exchanged to 10 mM ammonium bicarbonate using an Amicon Ultra-15 Centrifugal Filter Devices (10k MWCO cut, Millipore). The protein purity was analyzed by 15% SDS-PAGE. sfGFP proteins incorporated with CrtK (sfGFP-CrtK) was similarly expressed and purified by supplementing 5 mM CrtK in the growth medium.

sfGFPS2AcrK Protein Labeling with β-Mercaptoethanol via 1,4-Addition

β-Mercaptoethanol (40 mM) was added to the solution of sfGFP-AcrK (40 μM) in 0.1 M Tris-HCl buffer (pH 8.8). The solution was incubated at 37 ^{o}C for 8h. DTT (5mM) was then added and incubated at RT for an additional 1h. The sample was further treated by Zip-tip (Millipore) before subjected to the ESI-MS analysis (**Figure 2-4**).

sfGFPS2AcrK Protein PEGylation

Methoxypolyethylene glycol thiol 5k (Nanocos Inc.) (50 mM) was added into the solution of sfGFP-AcrK (40 μM) in 0.1 M Tris-HCl buffer (pH 8.8). The solution was incubated at 37 ^{o}C for 8h. The solution was then added DTT (5 mM) and was incubated at RT for an additional 1h before subjected to the SDS-PAGE analysis. An equal amount of wild-type sfGFP was treated in parallel as the control experiment.

sfGFPS2AcrK Protein Immobilization on Hydrogel via Radical Copolymerization

To the solution of sfGFP (5 μg) in PBS buffer (15 μL) was added equal volume of 30% acrylamide/bis-acrylamide solution (29:1) and β-ME (0.5% of final conc.). *N,N,N',N'*-tetramethylenylenediamine (TEMED) (0.1 μL) and ammonium persulfate (1 μL) was added into the mixture before loaded into the sample well of a 15% native polyacrylamide gel. The gel was left on bench top for 30 min to allow completion of the copolymerization. The gel was then applied under electrophoresis conditions with native running buffer (5 mM Tris-HCl, 30 mM glycine, pH 8.6). An equal amount of wild-type sfGFP was loaded next to immobilized wild-type sfGFP hydrogel as the control. After electrophoresis, the whole gel was visualized under blue light and photographed.

Kinetics Analysis of Nitrile Imine-Olefin Cycloaddition

The reactions between 5 μM 1 and varying concentrations of olefins in a 2 mL buffer that contained PBS and acetonitrile (1:1) were fluorescently monitored using a PTI QM-40 fluorescent spectrophotometer. The fluorescence increment from the

cyclized pyrazoline product was monitored with excitation at 315 nm and emission at 480 nm. Measurements were stopped once the signal reaches maximum mostly 60 to 90 min. The fluorescent increment data were fitted to a single exponential increase equation $F = F_1 - F_2 \times e^{(-k' \times t)}$, where F is the detected fluorescent signal at a give time, F_1 the final fluorescence, F_1-F_2 the background fluorescent signal, and *k'* the apparent pseudo first order reaction rate constant. The resulted values for *k'* were plotted against olefin concentrations and fitted to equation $k' = k_1 \times [olefin] + k_2$, where k_1 is the second order rate constant for the reaction between an olefin and nitrile imine 2 and k_2 is the hydrolysis rate constant of 2. All the data are shown in **Figures 2-8 - 2-12**.

Hydrolysis of Nitrile Imine

The UV absorbance decay of 50 μM 1 in a 2 mL buffer that contained PBS and acetonitrile (1:1) was monitored at 315 nm for about 4 h (**Figure 13**). The curve shows a single exponential decay with a *k* value as 0.0034 ± 0.00001 s^{-1}.

Simulation of Nitrile Imine Labeling of Proteins with Olefin Incorporated

The simulation was carried out using the program COPASI. The starting reagents were set as 500 μM for the hydrazonoyl chloride **1** and 1 μM for a corresponding olefin. The reaction time was set as 2 h. The simulation data are shown in **Figure 2-15**.

In vitro Fluorescent Protein Labeling with Hydrazonoyl Chloride 1

Compound 1 was added to a solution of sfGFP-AcrK in 0.1 M phosphate buffer (pH 7.4) and incubated at RT for 1 h. The protein was quickly purified using Ni-NTA resin and was subjected to SDS-PAGE analysis. In-gel fluorescence was detected by BioRad ChemiDoc XRS+ system under UV irradiation. All other sfGFPs incorporated with CrtK, **3**, **4**, or **5** were treated in parallel.

OmpX-AcrK Expression and Fluorescent Labeling

pETDuet-OmpX plasmid Construction

OmpX gene with C-terminal 6xHis Tag was amplified from BL21 (DE3) cells using an 5'-end primer containing an NdeI site: 5'-aggtggttcatatgaaaaaaattgcatgtctttcagcactgg-3' and a 3'-end primer with a KpnI site: 5'-caaagtgaggtaccttagtgatggtgatggtgatggaagcggtaa ccaacaccggcaat-3'. The PCR product was cloned into the NdeI and KpnI sites of pETDuet-1 to afford pETDuet-OmpX. The pETDuet-OmpXTAG was generated by Phusion high fidelity DNA polymerase-based site-directed mutagenesis with primer pairs 5'-aagccgtactgcaagcg ccgcagccgcataggccgcagcttctggtgactacaac-3' and 5'-ttctcggtgtaagtgaaagaaccgatc-3'. The pETDuet-OmpXTAG has an AAAAXAA (A denotes alanine; X denotes an amber mutation) insertion between its coded OmpX residues S53 and S54.

Fluorescent Protein Labeling on Cell Lysates

BL21(DE3) cell was transformed to carry two plasmids pEVOL-PylT-PrKRS and pETDuet-OmpXTAG. The OmpX-AcrK protein was overexpressed in LB medium supplemented with IPTG (1 mM) and AcrK (5 mM) at 30 ^{o}C for 10 h. The cell pellet was washed with PBS 3 times and lysed. The cell lysate was incubated with **1** (0.5 mM) at RT for 1 h and subject to SDS-PAGE analysis. In-gel fluorescence was detected by BioRad ChemiDox XRS+ system under UV irradiation. The cell culture without AcrK supplement in the medium served as the control.

Fluorescent Protein Labeling on Living Cell

BL21(DE3) cell was transformed to carry plasmids pEVOL-PylT-PrKRS and pETDuet-OmpXTAG. The OmpX-AcrK protein was overexpressed in LB medium supplied with IPTG (1 mM) and AcrK (5 mM) at 30 ^{o}C for 10 h. The cell pellet was washed with PBS 3 times and incubated with **1** (0.5 mM) at RT for 1 h. The cell pellet was then washed with isotonic saline solution 3 times and subjected to the confocal microscopy imaging. The cell culture without AcrK supplement in the medium served as the control and treated following the same protocol.

Confocal Imaging

An IX-81 base (Olympus) housing a full spectral FV-1000 (Olympus) was used to image the fluorescently labeled *E. coli* and confirm a lack of autofluorescence in the unlabeled *E. coli*. A 405 nm ps-pulsed diode laser was used for the excitation of the

sample, with collection being obtained through the use of a monochrometer as a band pass filter (420-520 nm) followed by a photomultiplier tube (PMT). A 20x UPlanSApo (na 0.75, infinity corrected) objective was used for all measurements. Scan rates varied from 1 to 20 μm per second based on field of view and maximization of S/N (**Figure 2-16**). The unlabeled cells were unable to be observed via confocal fluorescence. For additional confirmation of targeted binding, the *E. coli* cells were also imaged by fluorescence lifetime imaging microscopy (FLIM) (**Figure 2-17**). This was accomplished using the same excitation regime and a time correlated single photon counting system (Picoquant) for detection, consisting of a gated PMT and a band-pass filter (470 ±35 nm) and with an laser repetition rate of 40 MHz and analyzed *via* the PicoHarp software package (Picoquant).

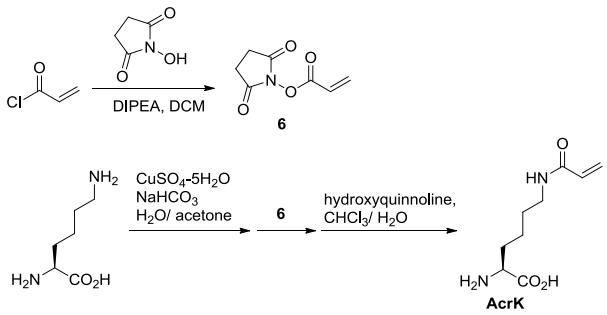
Chemical Synthesis

Materials and General Experimental

Unless otherwise noted, synthetic reagents were purchased from general chemical manufactures including Sigma-Aldrich, TCI, Alfa Aesar, and Chem-Impex and were used as received without further purification. Anhydrous DMF and methanol were purchased from Sigma-Aldrich in sealed bottles and transferred into reaction systems by using air-free technique. All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under argon atmosphere. Thin layer chromatography (TLC) was performed on silica 60 F-254 plates. Visualization of TLC was accomplished by UV irradiation at 254 nm or stained with *p*-anisaldehyde in acidic ethanol. Flash

column chromatography was performed with silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc. (Atlanta, GA).

All ^{1}H and ^{13}C NMR spectra were obtained on Varian Mercury-300, INOVA-300 and INOVA-500 MHz spectrometers. NMR spectra chemical shifts were reported in parts per million (ppm) and referenced to solvent peaks: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.8 ppm for ^{1}H). A minimal amount of 1,4-dioxane was added as the reference standard (67.19 ppm for ^{13}C) for carbon NMR spectra in deuterium oxide. ^{1}H NMR spectra are tabulated as follows: chemical shift, multiplicity (s = singlet, bs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). Mass spectra were obtained at the Laboratory for Biological Mass Spectrometry at Department of Chemistry, Texas A&M University.



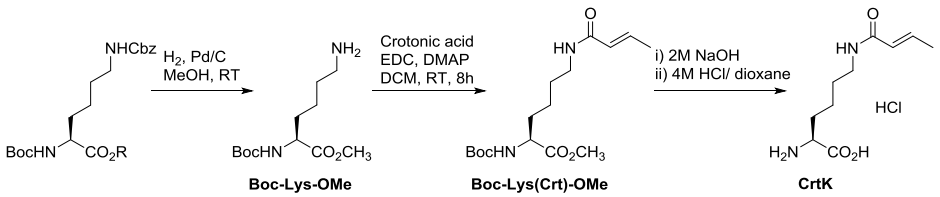
Scheme 2-3

Synthesis of N^{ε}-acryloyl-L-lysine (AcrK) – **Scheme 2-3**

***N*-Succinimidyl acrylate (6):** To a solution of N-hydroxysuccinimide (NHS) (2.6 g, 22.6 mmol) and *N*,*N*-diisopropylethylamine (DIPEA) (3.1 mL, 17.8 mmol) in

anhydrous dichloromethane (50 mL) was added acryloyl chloride (1.5 mL, 18.5 mmol) dropwise in ice bath. The mixture was warmed to RT and stirred for 8h. The reaction was then diluted with ethyl acetate, washed with saturated aqueous ammonium chloride, brine, and dried over anhydrous $MgSO_4$. After filtration, the filtrate was concentrated under reduced pressure to give a clear oil (3.4 g) that was subjected to next step without further purification.

***N*$^{\varepsilon}$-acryloyl-L-lysine (AcrK):** Copper(II) sulfate pentahydrate (2.5 g, 10 mmol) in water was added into a aqueous solution of lysine hydrochloride (3.7 g, 20 mmol) and sodium bicarbonate (4.7 g, 56 mmol). The blue mixture was stirred at RT for 15 min, and then **6** from previous step (40 mL) was added as a solution in acetone. The reaction was stirred at RT for 8h. The blue suspension was then filtered and washed with water and acetone. To the afforded blue solid was added water and chloroform in equal amount (200 mL). The solution was stirred vigorously at RT. Solid 8-hydroxyquinoline (4.1 g, 28 mmol) was added in one portion. The mixture was gradually turned green and the solution was stirred at RT for additional 30 min. The resulted in heterogeneous green mixture was filtered. The filtrate was washed with chloroform, concentrated *in vacuo* and further purified by ion-exchange chromatography to give a white powder as the desired product (3 g, 74% for two steps). ^{1}H NMR (D_2O, 300 MHz) δ 6.2-5.43 (m, 2H), 5.59 (dd, 1H, J = 11.4, 1.8 Hz), 3.58 (t, 1H, J = 6 Hz,), 3.13 (t, 2H, J = 6.9 Hz), 1.71 (m, 2H), 1.46 (m, 2H), 1.24 (m, 2H). ^{13}C NMR (75 MHz, D_2O) δ 175.4, 169.1, 130.6, 127.7, 55.3, 39.6, 30.7, 28.6, 22.4. HRMS (ESI) calcd. for $C_9H_{17}N_2O_3$ $[M+H]^+$ 201.1234, found 201.1241.



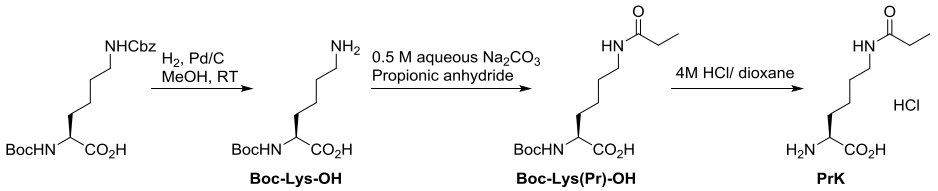
Scheme 2-4

Synthesis of N^{ε}-crotonyl-L-lysine (CrtK) – **Scheme 2-4**

Boc-Lys-OMe: To a solution of Boc-Lys(Cbz)-OMe (3.9 g, 10 mmol) in methanol was added palladium on activated carbon (Pd 10%, 0.6 g, 0.6 mmol) and the suspension was stirred with hydrogen bubbled through for 3 h. The product was filtered with celite and concentrated to be an almost colorless oil. The product was directly used in the next step without further purification.

Boc-Lys(Crt)-OMe: To a solution of Boc-Lys-OMe (2.6 g, 10 mmol) in dichloromethane (40 mL, dried with calcium hydride) was added crotonic acid (1.76 g, 20 mmol), followed by 4-dimethylaminopyridine (50 mg, 0.4 mmol) and N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC, 2.5 g, 13 mmol) in dichloromethane (10 mL) dropwise. The reaction mixture was stirred at RT under a balloon of argon overnight. The product was extracted by ethyl acetate (100 mL) and washed with water (60 mL twice), brine (60 mL), dried with sodium sulfate, concentrated and chromatographed (EtOAc/hexanes, 1:3) to give a colorless oil (2.8 g, 85%).^{1}H NMR ($CDCl_3$, 300 MHz) δ 6.83 (m, 1H), 5.79 (d, 1H, J = 15.0 Hz), 5.66 (s, 1H), 5.13 (d, 1H, J = 8.1 Hz), 4.26 (m, 1H), 3.72 (s, 3H), 3.29 (quart, 2H, J = 6.3 Hz), 1.83 (dd, J = 6.9, 1.8 Hz, 3H), 1.69-1.43 (m, 4H), 1.41 (s, 9H), 1.39 (m, 2H).

***N*$^{\varepsilon}$-Crotonyl-L-lysine (CrtK):** To a solution of Boc-Lys(Crt)-OMe (2.8 g, 8.5 mmol) in THF (4 mL) was added sodium hydroxide solution (2 M, 20 mL, 40 mmol) dropwise in an ice bath. The reaction mixture was stirred at 0 ^{o}C for 1h then at RT for another 2h. The organic impurities were washed by ethyl acetate (40 mL twice) and then the pH of the aqueous layer was adjusted to 3 with 3 M aqueous HCl. Then the product was extracted by ethyl acetate (60 mL then another 20 mL), washed with water (40 mL twice), brine (40 mL), dried with anhydrous sodium sulfate, and concentrated to give a sticky colorless oil, which was then dissolved in dioxane (5 mL). Hydrochloric acid in dioxane (4 M, 5 mL, 20 mmol) was then added and the reaction mixture was stirred at RT for 1h. The solvent was removed to give an almost colorless oil (1.7g, 94%). The product was applied to biological studies without further purification. ^{1}H NMR (D_2O, 300 MHz) δ 6.78 (m, 1H), 5.93 (d, 1H, J = 15.0 Hz), 4.01 (t, 1H, J = 6.3 Hz), 3.26 (t, 2H, J = 6.6Hz), 1.95 (m, 2H,), 1.85 (dd, 3H, J = 6.9, 1.5 Hz), 1.59 (m, 2H), 1.46 (m, 2H). ^{13}C NMR (75 MHz, D_2O) δ 175.8, 169.6, 142.2, 124.5, 55.3, 39.5, 30.9, 28.7, 22.4, 17.7. HRMS (ESI) calcd. for $C_{10}H_{19}N_2O_3$ $[M+H]^+$ 215.1390, found 215.1387.



Scheme 2-5

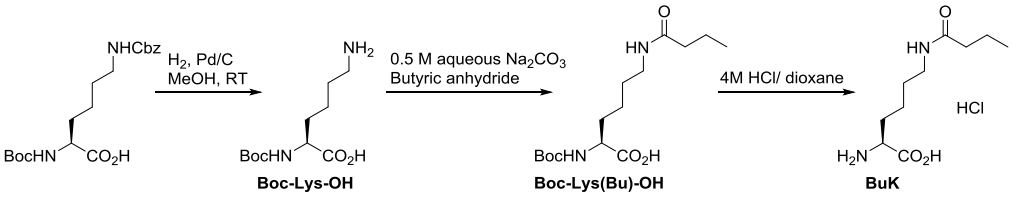
Synthesis of N^{ε}- Propionyl-L-lysine (PrK) – **Scheme 2-5**

Boc-Lys-OH: To a solution of Boc-Lys(Cbz)-OH (3.8 g, 10 mmol) in methanol (50 mL) was added palladium on activated carbon (Pd 10%, 0.6 g, 0.6 mmol) and the suspension was stirred with hydrogen bubbled through for 3 h. The product was filtered with celite and concentrated to be almost colorless oil. The product was directly used in the next step without further purification.

Boc-Lys(Pr)-OH: Boc-Lys-OH (2.4 g, 9.8 mmol) was dissolved in the solution of sodium carbonate (0.5 M, 40 mL) and then propionic anhydride (2.0 mL, 15.7 mL) was added dropwise into the solution. The reaction mixture was stirred at RT for 2h. Sodium hydroxide (2 M, 20 mL) solution was added and the excessive propanoic anhydride was extracted by ethyl acetate. Then the pH of the aqueous phase was adjusted by hydrochloric acid to pH 3. The product was extracted by ethyl acetate (80 mL) and concentrated to give a colorless oil. The product was directly used in the next step without further purification.

***N^{ε}*- Propionyl-L-lysine (PrK)**: To a solution of Boc-Lys(pr)-OH (2.8 g, 9.3 mmol) in dioxane (5 mL) was added hydrochloric acid in dioxane (4 M, 5 mL, 20 mmol). The reaction mixture was stirred at RT for 2h. A white sticky solid precipitated out as the reaction went. The liquid was filtered off and the solid was washed with excess ethyl acetate (200 mL). The product was then dried under oil pump overnight to afford a slightly yellow solid (1.7 g, 90%). ^{1}H NMR (D_2O, 300 MHz) δ 4.05 (t, 1H, J = 6.3 Hz), 3.16 (t, 2H, J = 6.9Hz), 2.20 (quart, 2 H, J=7.8Hz,), 1.93 (m, 2H), 1.52 (m, 2H),

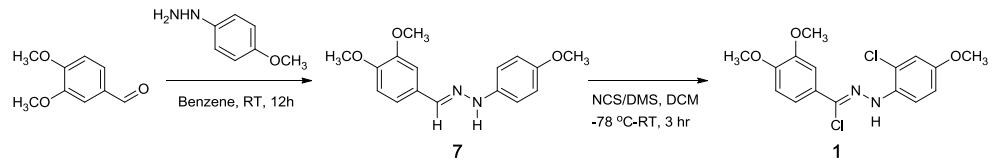
1.44 (m, 2H). ^{13}C NMR (75 MHz, D_2O) δ 175.6, 173.0, 53.6, 39.3, 30.0, 29.8, 28.5, 22.1, 10.3. HRMS (ESI) calcd. for $C_9H_{19}N_2O_3$ $[M+H]^+$ 203.1390, found 203.1403.



Scheme 2-6

Synthesis of N^{ε}- butyryl-L-lysine (BuK) – **Scheme 2-6**

***N*$^{\varepsilon}$- Butyryl-L-lysine (BuK)** was prepared in the same manner as PrK except that a cation exchange column was applied to purify the final product. ^{1}H NMR (D_2O, 300 MHz) δ 3.72 (t, 1H, J = 6.3 Hz), 3.20 (t, 2H, J = 6.9 Hz), 2.21 (t, 2H, J = 7.5 Hz,), 1.87 (m, 2H), 1.59 (m, 4H), 1.42 (m, 2H), 0.90 (t, 2H, J = 7.5 Hz). ^{13}C NMR (75 MHz, D_2O) δ 177.6, 175.3, 55.3, 39.5, 38.3, 30.7, 28.7, 22.4, 19.7, 13.3. HRMS (ESI) calcd. for $C_{10}H_{21}N_2O_3$ $[M+H]^+$ 217.1547, found 217.1550.



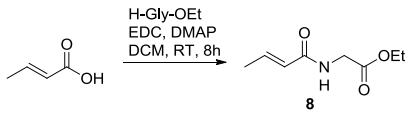
Scheme 2-7

Synthesis of Hydrazonoyl Chloride **1** – **Scheme 2-7**

Hydrazonoyl chloride **1** was synthesized according to a literature procedure.[27]

Hydrazone 7: To commercially available 3,4-dimethoxybenzaldehyde (0.59 g, 3.6 mmol) in 20 mL of benzene was added 4-methoxyphenylhydrazine hydrochloride (0.62 g, 3.6 mmol) at RT. The solution was stirred for 12 hours. The resulting heterogeneous mixture was concentrated to give a brownish solid **7** which was used in the next step without further purification.

Hydrazonoyl chloride 1: To a solution of N-chlorosuccinimide (NCS) (0.71 g, 5.3 mmol) in anhydrous dichloromethane (15 ml) was added dimethyl sulfide (DMS) (0.78 mL, 10.6 mmol) slowly in an ice bath. The mixture was stirred at 0°C for 20 min and cooled to -78°C. A solution of **7** in anhydrous dichloromethane was then added slowly. After complete addition, the solution was stirred for an additional 3h at the same temperature. The solvent-free crude product was purified by flash column chromatography (10% EtOAc/Hexane) to give a white powder as the desired product. ^{1}H NMR (300 MHz, $CDCl_3$) δ 8.27 (s, 1H), 7.45-7.52 (m, 3H), 6.92 (d, 1H, J = 2.76), 6.89 (d, 1H, J = 8.27), 6.87 (dd, 1H, J = 8.97, J = 2.77), 3.98 (s, 3H), 3.94 (s, 3H), 3.80 (s, 3H). ^{13}C NMR (75 MHz, $CDCl_3$) δ 154.1, 150.6, 149.0, 134.1, 127.4, 126.5, 120.1, 118.5, 115.6, 114.8, 114.6, 110.8, 109.3, 56.2, 56.2, 56.1.



Scheme 2-8

Synthesis of Crotonamide 8 – **Scheme 2-8**

Crotonamide 8: To a solution of crotonic acid (0.5 g, 5.8 mmol), glycine ethyl ester hydrochloride (0.87 g, 6.2 mmol), and 4-dimethylaminopyridine (DMAP) (30 mg, 0.25 mmol) in anhydrous dichloromethane was added 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC) (1.8 g, 9.4 mmol) in an ice bath. The reaction mixture was warmed to RT and stirred for 8 h. The solution was quenched with a dichloromethane/water mixture. The organic layer was collected, washed with water, brine, dried over anhydrous Na_2SO_4 and concentrated. The crude product was subjected to flash column chromatography (eluted at 25% EtOAc/hexane) to afford a yellowish solid as the desired product. ^{1}H NMR (300 MHz, $CDCl_3$) δ 6.90-6.80 (m, 1 H), 6.11 (bs, 1H), 5.88 (dq, 1H, J = 15, 3 Hz), 4.22 (q, 2H, J = 6 Hz), 4.09 (d, 2H, 3 Hz), 1.87 (dt, 2H, J = 9, 3 Hz), 1.29 (t, 3H, 6 Hz). ^{13}C NMR (75 MHz, $CDCl_3$) δ 170.4, 166.1, 141.0, 124.6, 61.7, 41.6, 18.0, 14.3.

CHAPTER III

GENETICALLY ENCODED UNSTRAINED OLEFINES FOR LIVE CELL LABELING WITH TETRAZINE DYES*

Introduction

Owing to its high selectivity, fast reaction kinetics, and non-catalytic nature, the inverse electron-demand Diels-Alder cycloaddition reaction between a tetrazine and an olefin has emerged as a state-of-the-art approach for selective bioconjugation in live cells.[31, 95-99] Previous efforts focused on applications of strained cyclic olefins or alkynes that react rapidly with tetrazines.[21, 22, 30, 56, 76-79, 100-104] However, similar applications with unstrained olefins have been largely overlooked except a recent report of metabolic glycan labeling,[105] due to the common notion that the reaction between an unstrained olefin and a tetrazine is not kinetically favored. Here, we systematically analysed unstrained olefin-tetrazine reactions and showed that several unstrained olefins readily reacted with a tetrazine dye and a number of unstrained olefin-bearing ncAAs can be genetically incorporated into proteins in *Escherichia coli* for selective labeling with tetrazine dyes.

*Reprinted with permission from Lee, Y. J., Kurra, Y., Yang, Y., Torres-Kolbus, J., Deiters, A., and Liu, W. R. (2014) Genetically encoded unstrained olefins for live cell labeling with tetrazine dyes, *Chem. Commun. 50,* 13085-13088. Copyright 2014 Royal Society of Chemistry.

Results and Discussion

We first characterized the reaction kinetics between a tetrazine dye **10** (**Figure 3-1**) and a number of unstrained olefins in phosphate buffered saline (PBS) at ambient temperature. Three alkenyl alcohols (allyl alcohol, 3-buten-1-ol, and 4-penten-1-ol) and ncAAs **5-8** (**Figure 3-1**) were chosen to undergo kinetic analysis. The alkenyl alcohols were used instead of their corresponding ncAAs **1-3** due to their high solubility in PBS buffer. Both **2** and **3** could hardly dissolve to more than 5 mM in PBS buffer. With this low solubility, their reactions with **10** were too long for reliable data collection. To test whether terminal and non-terminal olefins react differently with **10**, 2-buten-1-ol was also included in the analysis. Since the inverse electron-demand Diels-Alder reaction is highly influenced by the dienophile electron density, vinyl ethers with electron-rich olefins, 2-hydroxyethyl vinyl ether that corresponds to ncAA **4**, and **9** were also selected to undergo assays with **10**. Conjugation of fluorescein to a tetrazine moiety such as in **10** inherently quenches the fluorescence from the fluorophore.[31] Cycloaddition of the tetrazine moiety and an olefin abolishes this intrinsic quenching effect, leading to the rescue of fluorescence that can be readily detected with a fluorospectrometer. The reactions of all olefins with **10** were carried out in pseudo-first-order conditions in which the olefin concentration was at least 20-fold higher than the concentration of **10** (**Figures 3-2 - 3-11**). The determined second-order reaction rate constants are presented in **Table 3-1**.

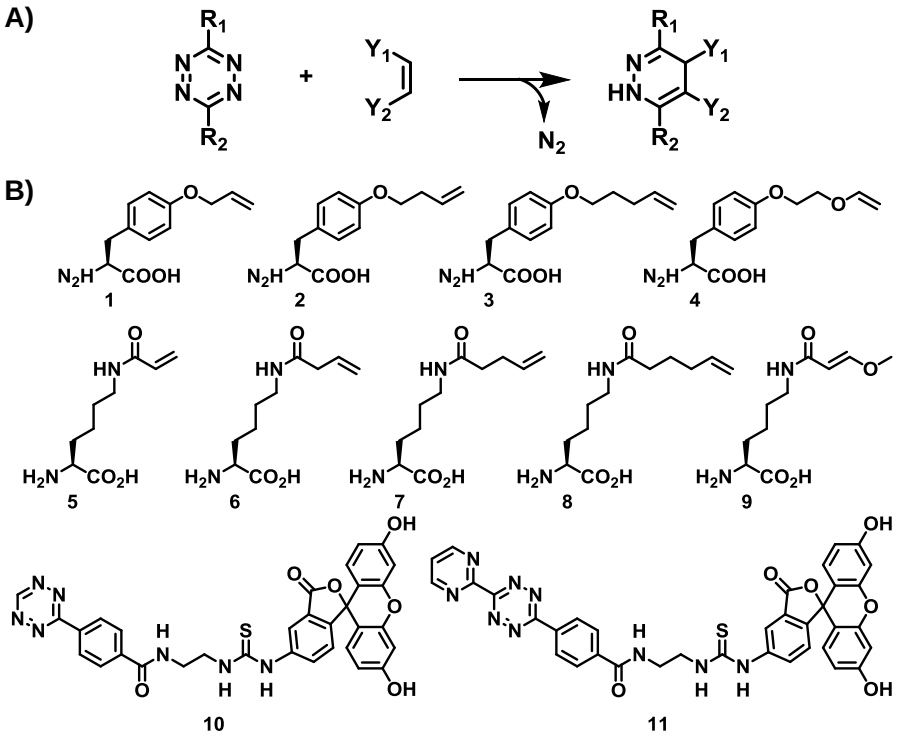
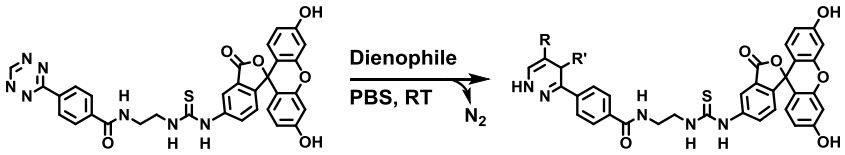


Figure 3-1. (A) The tetrazine-olefin reaction. (B) Unstrained olefin-bearing ncAAs. (C) Two fluorescein-tetrazine dyes used in this study.



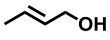
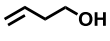
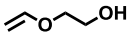
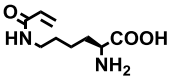
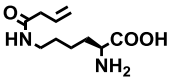
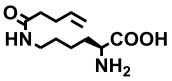
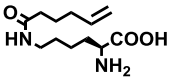
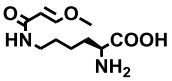
Entry	Dienophile	$k\ x10^3$/ $M^{-1}\ s^{-1}$
1		**9.4** ±1.2
2		**1.9** ±0.1
3		**26** ±5
4		**36** ±4
5		**81** ±1
6		**1.2** ±0.1
7		**1.7** ±0.2
8		**11** ±2
9		**16** ±1
10		**19** ±1

Table 3-1. Determined second-order reaction rate constants between various olefin dienophiles and **10**.

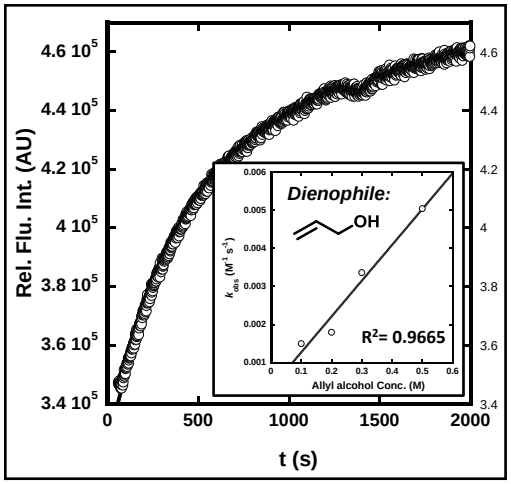


Figure 3-2. Kinetics analysis of the inverse electron demand Diels-Alder reaction (DAinv) between tetrazine 10 and allyl alcohol dienophile in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

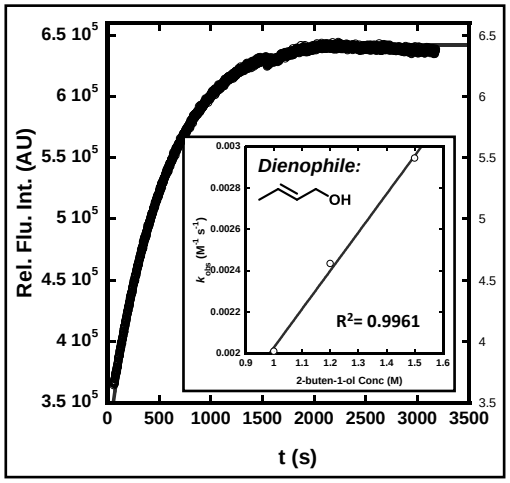


Figure 3-3. Kinetics analysis of the DAinv between tetrazine 10 and 2-buten-1-ol dienophile in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

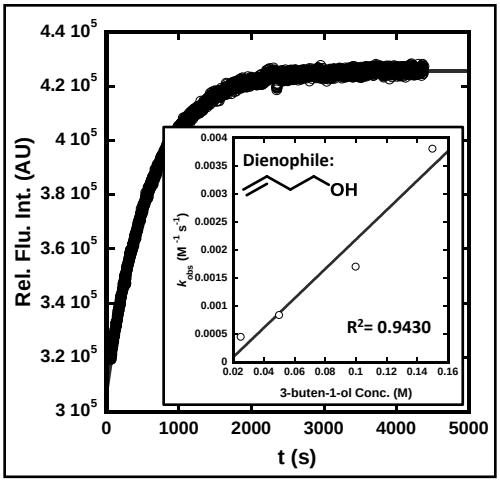


Figure 3-4 Kinetics analysis of the DAinv between tetrazine 10 and 3-buten-1-ol dienophile in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

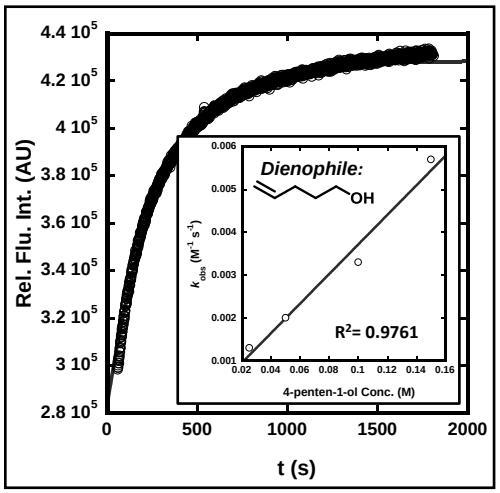


Figure 3-5 Kinetics analysis of the DAinv between tetrazine 10 and 4-penten-1-ol dienophile in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

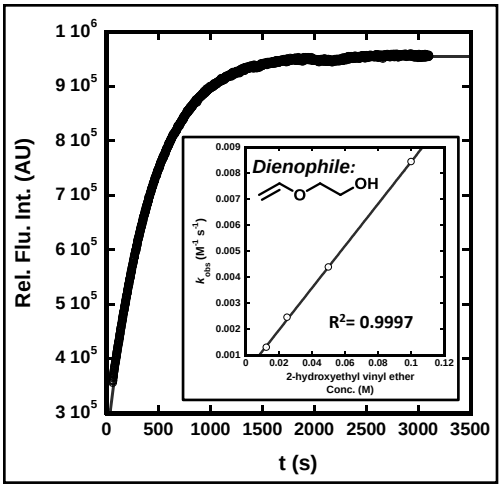


Figure 3-6 Kinetics analysis of the DAinv between tetrazine 10 and 2-hydroxyethyl vinyl ether dienophile in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

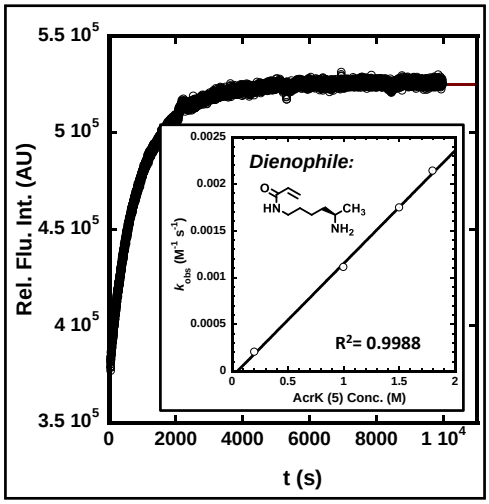


Figure 3-7 Kinetics analysis of the DAinv between tetrazine 10 and ncAA 5 in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

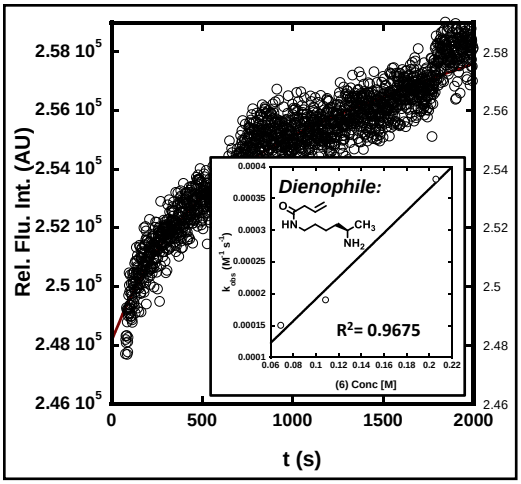


Figure 3-8 Kinetics analysis of the DAinv between tetrazine 10 and ncAA 6 in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

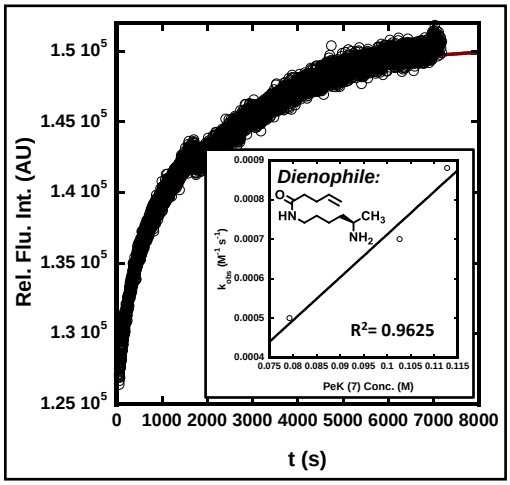


Figure 3-9 Kinetics analysis of the DAinv between tetrazine 10 and ncAA 7 in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

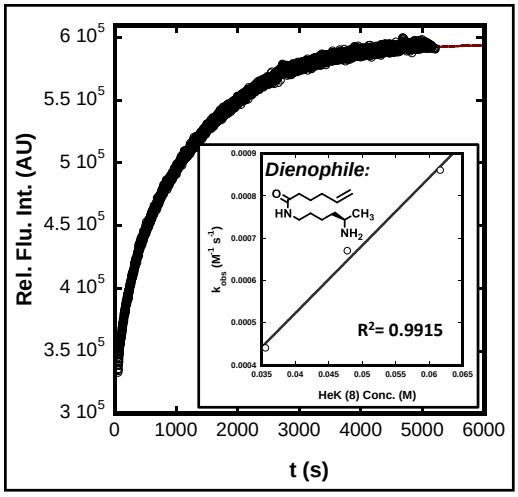


Figure 3-10 Kinetics analysis of the DAinv between tetrazine 10 and ncAA 8 in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

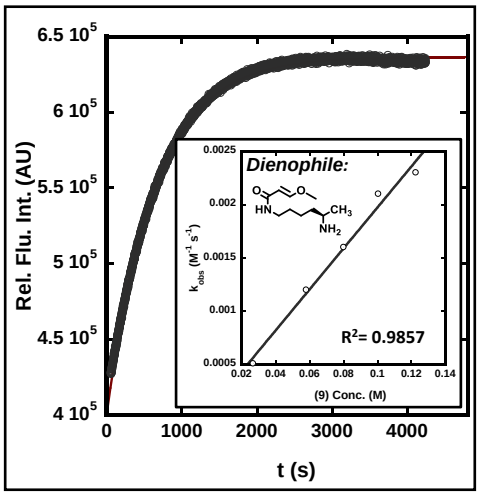


Figure 3-11 Kinetics analysis of the DAinv between tetrazine 10 and ncAA 9 in PBS buffer at RT. The fluorescence increase at 521 nm was monitored as excited at 493 nm. The inset plot presents the linear dependence of the reaction apparent rate (k_{obs}) with the concentration of dienophile.

As shown in **Table 3-1**, all tested unstrained olefins reacted readily with **10** with second-order rate constants more than 0.001 M^{-1} s^{-1}. Both inductive and resonance effects of an olefin substituent significantly influence its reactivity with respect to **10**.[106, 107] Electron withdrawing effects of the hydroxyl group and the amide group gradually diminish, correlating well to gradually improved reactivities from allyl alcohol to 4-penten-1-ol and **6** to **8**, respectively. Two vinyl ethers, 2-hydroxyethyl vinyl ether and **9**, display relatively higher reactivities toward **10** in comparison to other tested olefins due to the strong electron donating resonance effect of the vinyl ether oxygen atom. The sluggish reaction between **5** and **10** is attributed to the electron withdrawing resonance effect of the amide group. In comparison to the allyl alcohol, 2-buten-1-ol reacts significantly slower with **10**, indicating that the methyl substituent adds steric hindrance. Although our kinetic analysis indicates much slower reactions with a tetrazine for unstrained olefins compared to strained ones,[98] the determined second-order rate constants are comparable to those of the Staudinger ligation (k = 0.002 $M^{-1}s^{-1}$)[16] and the copper-free dibenzocyclooctyne-azide cycloaddition (k = 0.0565 $M^{-1}s^{-1}$),[108-112] two reactions that have been well adopted for bioconjugation in live cells. In comparison to strained olefins, unstrained olefins can be more easily prepared and are generally more stable toward cellular nucleophiles (except **5**).[113] Their genetic installation in proteins followed by selective bioconjugation with tetrazine dyes can potentially provide a simple and readily adoptable protein labeling approach in live cells.

Encouraged by our kinetic analysis, we proceeded to recombinantly synthesize proteins carrying site-specifically incorporated unstrained olefins in *E. coli*. We

previously reported that a rationally designed pyrrolysyl-tRNA synthetase (PylRS) mutant with two mutations, N346A and C348A, (PylRS(N346A/C348A)) together with $tRNA_{CUA}^{Pyl}$ is able to mediate the genetic incorporation of **1**-**4** into proteins at an amber codon in *E. coli*.[70, 74, 114, 115] This system was applied to synthesize four superfolder green fluorescent protein (sfGFP) variants with **1**, **2**, **3**, or **4** incorporated at their S2 positions (sfGFP-**1**, sfGFP-**2**, sfGFP-**3**, and sfGFP-**4**). To synthesize the sfGFP variant with **5** incorporated at its S2 site (sfGFP-**5**), a previously evolved PylRS mutant PrKRS together with $tRNA_{CUA}^{Pyl}$ was used.[26] Although systems for the genetic incorporation of **6**-**9** has not been described in the literature, our studies indicated that another previously evolved PylRS mutant BuKRS is able to recognize them as substrates.[26] BuKRS together with $tRNA_{CUA}^{Pyl}$ was therefore applied for the synthesis of sfGFP variants with **6**, **7**, **8**, and **9** at their S2 positions (sfGFP-**6**, sfGFP-**7**, sfGFP-**8**, and sfGFP-**9**) and the site-specific incorporation of **6**-**9** was verified by electrospray ionization mass spectrometry (ESI-MS) analysis (**Figures 3-17 - 3-20**). The sfGFP variants were purified and then subjected to the fluorescent labeling with **10** at physiological pH in PBS buffer before they were denatured and analyzed by SDS-PAGE. Direct in-gel fluorescence detection showed strong fluorescent labeling of sfGFP-**2**, sfGFP-**3**, sfGFP-**7**, and sfGFP-**8** (**Figure 3-12**). Labeling of sfGFP-**1**, sfGFP-**5**, and sfGFP-**6** could be barely detected and, importantly, no labeling was observed for wild-type sfGFP. This labeling observation is consistent with the kinetic results in which 3-buten-1-ol, 4-penten-1-ol, **7**, and **8** react more favorably with **10** compared to allyl alcohol, **5**, and **6**. Although 2-hydroxyethyl vinyl ether displayed fast kinetics when reacted with **10**, sfGFP-**4** did not show detectable

labeling with **10** on the gel. This is likely due to the cleavage of the cycloaddition product from sfGFP-**4** to generate an aromatized 1,2-diazine-fluorescein final product and a sfGFP-**4** derivative with its vinyl group removed.[106, 107] Although a similar diazine formation reaction that followed the reaction between sfGFP-**9** and **10** would retain the 1,2-diazine-fluorescein moiety to sfGFP-**9**, we did not observed fluorescein-labeled sfGFP-**9** on the gel. Since a fluorescent small molecule band was detected in the gel (**Figure 3-13**), we suspect that the final fluorescein-labeled sfGFP-**9** was not stable and hydrolysed during the SDS-PAGE analysis that required harsh treatments such as heating at 100 degree for 5 min. In gels, sfGFP-**2** and sfGFP-**7** displayed integrated fluorescence intensities that were about 40% and 80% in comparison to sfGFP-**3** and sfGFP-**8** after their labeling with **10**. The labeled sfGFP-**3** was further subjected to the ESI-MS analysis that indicated close to 50% labeling efficiency (**Figure 3-14**).

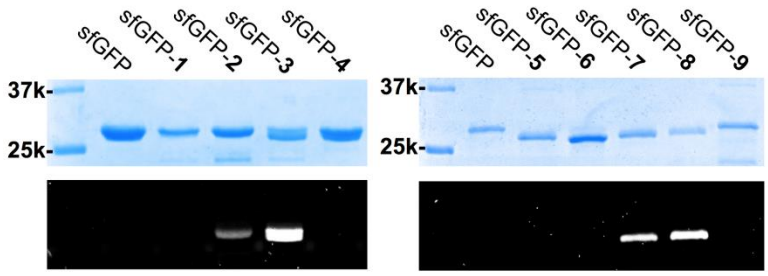


Figure 3- 12 Labeling of sfGFP variants with **10**. The top panel shows Coomassie blue stained gels and the bottom panel presents the fluorescent imaging of the same gels before they were stained with Coomassie blue. Each protein (35 μM) was incubated with 0.5 mM of **10** in a 50 mM Tris-Cl buffer at pH 7.4 for 8 h before they were precipitated in 10% TFA, fully denatured with heating at 100 °C for 5 min, and then analyzed by SDS-PAGE. sfGFP with no ncAA incorporated was used as a control.

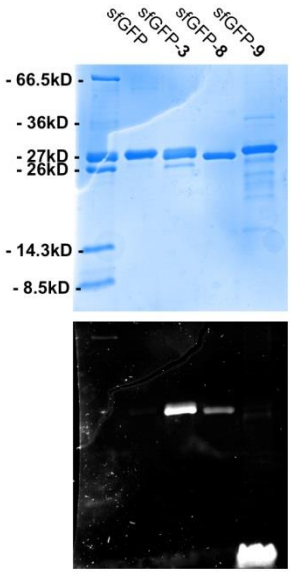


Figure 3-13 Labeling sfGFP variants with **10**. The top: Coomassie blue stained gel and the bottom: the fluorescent imaging of the same gel before they were stained with Coomassie blue. sfGFP with no ncAA incorporated was used as a control. The fluorescent signal of the gel indicated that sfGFP-**3** and sfGFP-**8** were fluorescently labeled with **10**. sfGFP-**9** did not show fluorescent signal at its corresponding sfGFP protein band. However, a fluorescent signal was observed at low molecule weight region suggesting the decomposition of the product after the reaction between sfGFP-**9** and **10**.

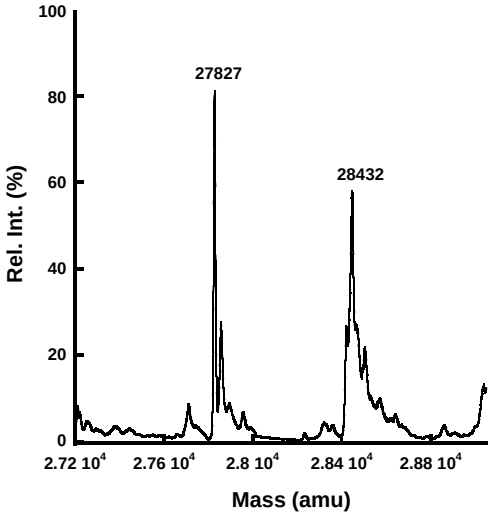


Figure 3-14. ESI-MS analysis of the sfGFP-3 modified with tetrazine dye 10. Calculated mass for unmodified sfGFP-3: 27826 Da; found: 27827 Da. Calculated mass for sfGFP-3 and tetrazine 10 conjugate: 28431 Da; found: 28432 Da.

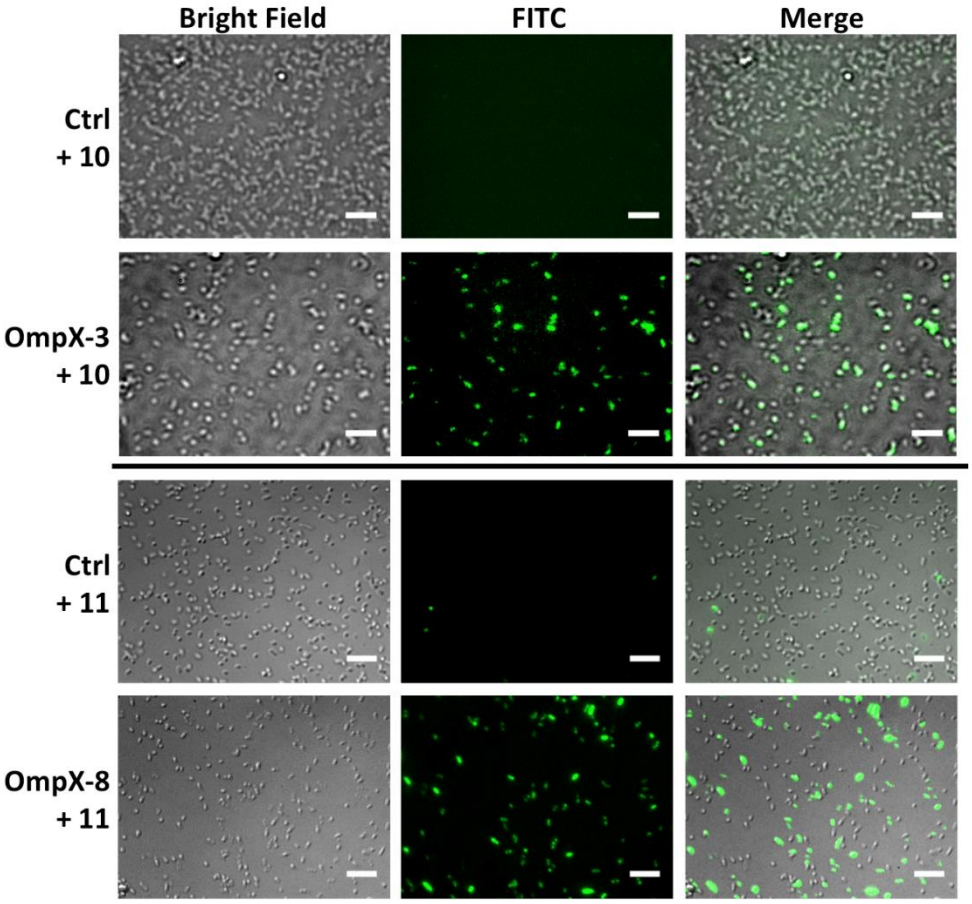


Figure 3-15. Selective labeling of *E. coli* cells expressing OmpX-**3** and OmpX-**8**. Labeling reactions were carried out in the presence of 0.5 mM **10** or **11** in the PBS buffer at pH 7.4 for 8 h. *E. coli* cells grown in corresponding media without NCAA supplemented served as controls (Ctrls). The left panels show bright-field images of *E. coli* cells, the middle panels show green fluorescent images of the same cells, and the right panels shows composite images of bright-field and fluorescent images. Scale bars represent 10 um.

With the demonstration of the tetrazine reaction of unstrained alkenes for protein labeling *in vitro*, we then proceeded to test this reaction to label proteins bearing unstrained alkenes in live cells. Two previously constructed plasmids pEVOL-pylT-PylRS(N346A/C348A)[70] and pETDuet-OmpXTAG[26] were used to transform *E. coli* BL21 cells to express an *E. coli* outer membrane protein OmpX with **3** incorporated at its extracellular domain (OmpX-**3**). Plasmid pEVOL-pylT-PylRS(N346A/C348A) carried genes for PylRS(N346A/C348A) and $tRNA_{CUA}^{Pyl}$ and plasmid pETDuet-OmpXTAG contained a gene coding OmpX with an AAAXAA (A denotes alanine and X denotes an amber mutation) insertion between two extracellular residues 53 and 54. The transformed cells were grown in M9 minimal medium supplemented with 1% glycerol, 1 mM IPTG, 0.2% arabinose, and 5 mM **3** to express OmpX with **3** incorporated. After the expression of OmpX-**3**, cells were washed with PBS buffer and incubated with **10** (0.5 mM) in PBS buffer at room temperature for 8 h. The residual dye was removed by washing cells with isotonic saline. Finally the labeled cells were subjected to fluorescent microscope imaging. *E. coli* cells expressing OmpX-**3** showed detectable fluorescence (**Figure 3-15**). In contrast, control cells that were grown in the absence of **3** did not display any fluorescence. A similar labeling reaction was carried out to label *E. coli* cells that expressed OmpX with **8** incorporated (OmpX-**8**). To express OmpX-**8**, plasmids pEVOL-pylT-BuKRS and pETDuet-OmpXTAG were used to transform *E. coli* BL21 cells. Plasmid pEVOL-pylT-BuKRS carried genes coding BuKRS and $tRNA_{CUA}^{Pyl}$. OmpX-**8** was expressed similarly as OmpX-**3** except that the expression was done in LB medium. Since we noticed some background reaction

between **10** and bovine serum albumin, we employed another tetrazine-fluorescein dye **11** that displayed a much lower reactivity toward cysteine than **10** for the following cell labeling analysis (**Figure 3-16**). Cells were incubated with **11** (0.5 mM) for 8h, washed with isotonic saline, and then subjected to fluorescent microscope imaging. Cells expressing OmpX-**8** displayed strong green fluorescence. However, cells grown in the absence of **8** were not visibly labeled with **11**. These results indicate that the unstrained alkene labeling reaction with a tetrazine is biocompatible and highly selective.

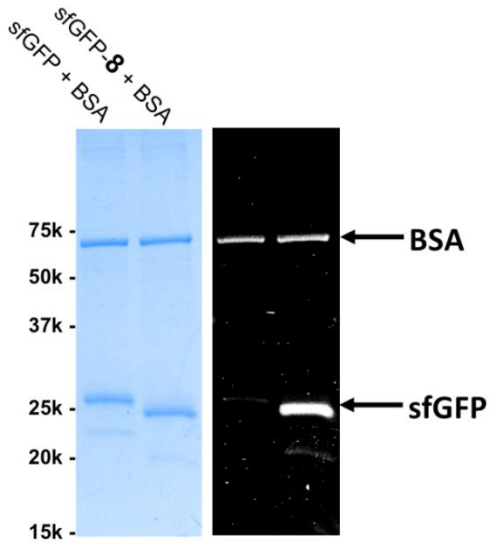


Figure 3-16. Labeling of sfGFP variants and bovine serum albumin (BSA) mixture with **10**. The left panel shows Coomassie blue stained gel and the right panel presents the fluorescent imaging of the same gel before it was stained with Coomassie blue. Each mixture of 35 uM sfGFP and 30 ug BSA (Thermo-Fisher Scientific Inc., Rockford, IL, USA) was incubated with 0.5 mM 10 in 50 mM Tris-HCl buffer at pH 7.4 for 8 h before they were precipitated in 10% TFA, fully denatured with heating at 100 oC for 5 min, and then analyzed by SDS-PAGE.

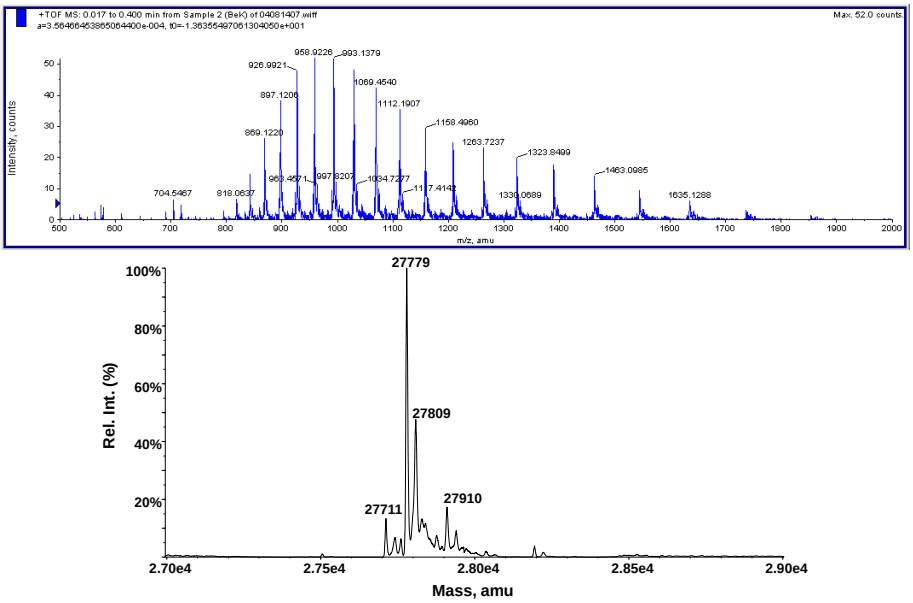


Figure 3-17 ESI-MS analysis of the purified sfGFP incorporated with NCAA 6 (sfGFP-6). Top panel is the protein charge ladder and the bottom panel is the corresponding deconvoluted mass spectrum. Calcd. mass 27776 Da, found 27779 Da.

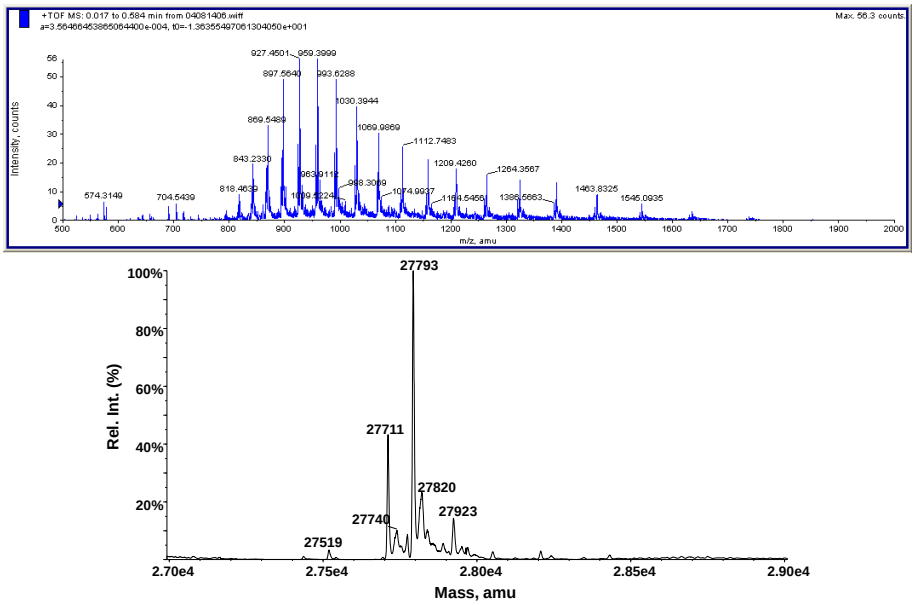


Figure 3-18 ESI-MS analysis of the purified sfGFP incorporated with NCAA 7 (sfGFP-7). Top panel is the protein charge ladder and the bottom panel is the corresponding deconvoluted mass spectrum. Calcd. mass 27790 Da, found 27793 Da.

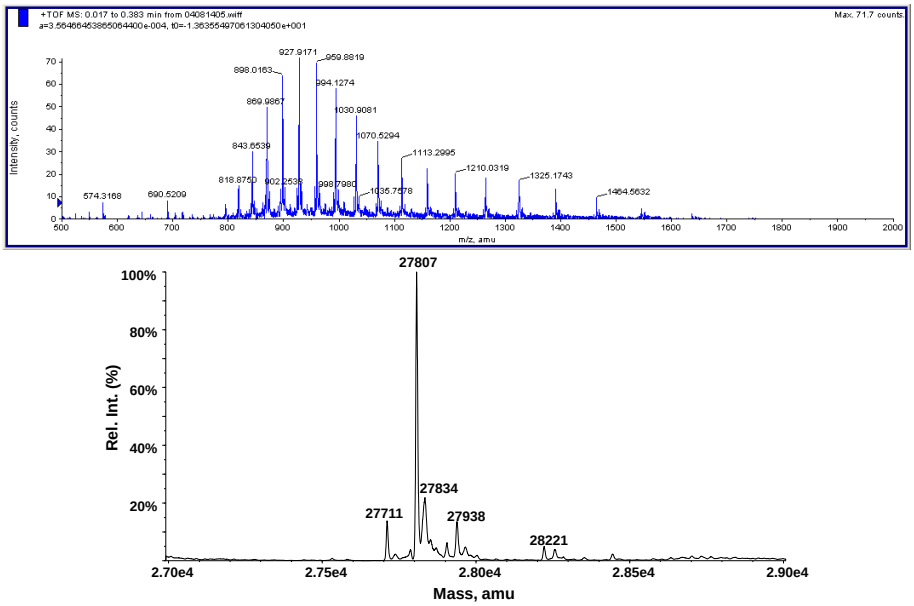


Figure 3-19 ESI-MS analysis of the purified sfGFP incorporated with NCAA 8 (sfGFP-8). Top panel is the protein charge ladder and the bottom panel is the corresponding deconvoluted mass spectrum. Calcd. mass 27804 Da, found 27807 Da.

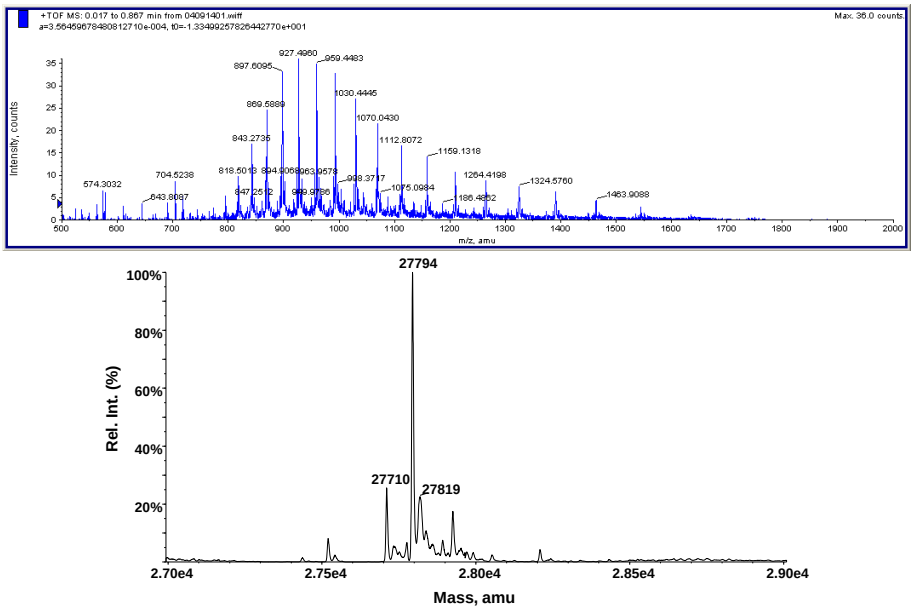


Figure 3-20 ESI-MS analysis of the purified sfGFP incorporated with NCAA 9 (sfGFP-9). Top panel is the protein charge ladder and the bottom panel is the corresponding deconvoluted mass spectrum. Calcd. mass 27792 Da, found 27794 Da.

Conclusion

In summary, we demonstrate that an unstrained olefin can react with a tetrazine dye with an appreciable reaction rate that allows selective labeling of proteins on the surface of live bacterial cells. Given its small size, easy preparation, and high stability, an unstrained olefin provides a significant advantage over a strained olefin in promoting the general adoption of the tetrazine labeling approach for cell biology studies. With a size comparable to alkyne and azide, an unstrained olefin can be similarly integrated into metabolic surrogates for cell labeling.[16, 105, 116-119] Its biocompatible reaction with a tetrazine probe potentially provides an alternative tool for metabolic labeling research. That can be applied to investigate the temporal and spatial synthesis of DNA, RNA, proteins, and carbohydrates in live cells.

Experimental Section

Kinetic analysis of DAinv between tetrazine and acyclic alkenes

Kinetic analysis of the inverse electron demand Diels-Alder reaction (DAinv) was performed on a PTI QM-40 spectrofluorometer (PTI, Edison, NJ, USA). The fluorescence increase at 521 nm was monitored as exciting the mixed reaction solution at 493 nm in PBS buffer at room temperature. The reaction was set to be performed in pseudo first order kinetics manner, namely the dienophile concentration is at least 20 times higher than that of the diene. Specifically, 0.1 mM of tetrazine **10** was mixed with each acyclic alkene at designated concentration (ranging from 20 mM to up to 0.5 M). For each diene and dienophile pair, at least three experiments with different concentration of subjected alkene were carried out. The measurement was stopped when the fluorescence reached saturation, generally in 2-3h. The fluorescent increment data were fitted to a single exponential increase equation $F = F_1 - F_2 \times e^{(-k' \times t)}$, where F is the detected fluorescent signal at a given time, F_1 the final fluorescence, F_2 the background fluorescent signal, and k' the apparent pseudo first order reaction rate constant. The resulted values for k' were plotted against alkene concentrations and fitted to equation $k' = k \times [dienophile] + C$**,** where k is the second order rate constant for the reaction between an olefin and tetrazine **10.** All the data are shown in **Figures 3-2 - 3-11**.

Genetic incorporation of ncAAs into sfGFP and protein expression

Proteins Sequences

sfGFP:

MASKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFI**C**TTGKL

PVPWPTLVTTLTYGVQ**C**FSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY

KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGI

KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKR

DHMVLLEFVTAAGITHGMDELYKGSHHHHHH

*The ncAA incorporation site is highlighted in red.

Incorporation of ncAA-**1**, **-2**, **-3**, **-4,** and **-5** into sfGFP

A previously reported protocol was followed to genetically incorporate ncAA-**1**, **-2**, **-3**, **-4**, **-5** into sfGFP.[74] Briefly, *E. coli* BL21(DE3) cells were transformed with plasmids pET-PylT-sfGFPS2TAG and pEVOL-PylT-PylRS(N346A/C348A). The transformed cell was used to inoculate the LB media and was grown at 37°C with agitation until the OD_{600} reached 1.0-1.2. Then the cells were pelleted and washed with PBS buffer (x3). After medium shift to a minimal medium supplemented with 5 mM ncAA, 1 mM IPTG, and 0.2% arabinose, the cells were incubated at 37^{o}C with agitation for 8h. All ncAAs were purchased from Chem-Impex International Inc. (Wood Dale, IL, USA) and used without further purification. After centrifugation, the cell pellet was resuspended in ice cold lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM

imidazole, pH 8.0) and lysed by sonication. The lysate was centrifuged (10000 xg, 4^{o}C, 1h) and the supernatant was subjected to Ni-NTA affinity chromatography according to manufacturer's protocol. The supernatant was incubated with Ni-NTA resin (Qiagen) (2h, 4^{o}C). The slurry was then loaded to a column and the protein-bound resin was washed with washing buffer (50 mM NaH2PO4, 300 mM NaCl, 45 mM imidazole, pH 8.0) followed by eluting the bound protein with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein was conducted buffer exchange to Tris-HCl buffer (50 mM, pH 7.4) and concentrated using an Amicon Ultra-15 Centrifugal Filter (10 kD MWCO, Millipore). The quality of the purified protein was analyzed by SDS-PAGE and ESI-MS. The sfGFP concentration was quantified by UV absorbance at 485 nm using the literature reported molar extinction coefficient $8.33x10^4$ M^{-1} cm^{-1}.[120]

Incorporation of ncAA-**6**, -**7**, -**8**, -**9** into sfGFP

A previously reported protocol was followed to genetically incorporate ncAA **6**, **7**, **8, 9** into sfGFP.[26] Briefly, *E. coli* BL21(DE3) cells were transformed with plasmids pET-PylT-sfGFPS2TAG and pEVOL-PylT-BuKRS. The transformed cell was used to inoculate LB media and was grown at 37 °C with agitation until the OD_{600} reached 0.6-0.8. The medium was then supplemented with 5 mM ncAA, 1 mM IPTG, and 0.2% arabinose, followed by incubating the cells at 37 ^{o}C with agitation for 8 h for protein expression. The purification procedures are the same as described above.

Protein labeling with tetrazine dye on sfGFP-ncAA mutants

sfGFP (WT or incorporated with ncAA, 20 ug) and tetrazine **10** (0.5 mM) in Tris-HCl buffer (50 mM, pH 7.4) were incubated at room temperature for 8 h. After the reaction time, protein was precipitated down using aqueous trichloroacetic acid (TCA) solution, followed by centrifugation to remove excessive dyes in the supernatant. The obtained protein pellet was further washed with ice-cold acetone, dried and resuspended in 1x SDS-PAGE sample loading buffer (250 mM TrisHCl pH 6.8, 250 mM DTT, 2.5% SDS, 10% glycerol). After fully heat-denatured, the protein was then subjected to the SDS-PAGE analysis. In-gel fluorescence detection was performed before coomassie blue staining on a Bio-Rad ChemiDoc XRS+ system (Bio-Rad, Hercules, CA, USA) under UV irradiation.

E. Coli OmpX-ncAA mutants protein expression and protein labeling with tetrazine dye

The construction of OmpX expression plasmid pETDuet-OmpXTAG was described previously.[26] The expression of the OmpX with ncAA incorporation followed the same reported procedure with minor modification for expression optimization.

<u>OmpX-ncAA-5 Protein Expression</u>

BL21(DE3) cells transformed to carry plasmid pEVOL-PylT-PylRS(N346A/C348A) and pETDuet-OmpXTAG were grown in LB medium until

OD_{600} reached 1.0. Then the cells were pelleted and washed with PBS buffer (x3). After medium shift to minimal medium supplemented with 5 mM ncAA, 1 mM IPTG, and 0.2% arabinose, the cells was incubated at room temperature with agitation for 16h. After centrifugation, the pelleted cells were washed with PBS (x3), resuspended in PBS and subjected to the subsequent cell labeling procedure. The cell culture without ncAA **5** supplement in the medium served as the control.

OmpX-ncAA-8 Protein Expression

BL21(DE3) cells transformed to carry plasmid pEVOL-PylT-BuKRS and pETDuet-OmpXTAG were grown in LB medium until OD_{600} reached 0.6-0.8. The medium was then supplemented with 5 mM ncAA **8**, 1 mM IPTG, and 0.2% arabinose, followed by incubating the cells at room temperature with agitation for 16 h. After centrifugation, the pelleted cells were washed with PBS (x3), resuspended in PBS and subjected to the subsequent cell labeling procedures. The cell culture without ncAA **8** supplement in the medium served as the control.

Fluorescent *E. coli* OmpX-ncAA Mutants Labeling with Tetrazine Dyes

Aforementioned *E. coli* cells expression OmpX-**3** or OmpX-**8** and tetrazine dye (**10** for OmpX-**3** and **11** for OmpX-**8**) (0.5 mM) in Tris-HCl buffer (50 mM, pH 7.4) were incubated at room temperature for 8 h with homogeneous mixing. After centrifugation to remove the excessive dye in the solution, the resulting cell pellet was

resuspended in PBS and incubated at room temperature for another 1-2 h. The cell was once again pelleted and washed with PBS or isotonic saline (x5), resuspended in PBS or isotonic saline then subjected to epifluorescence microscopy imaging. The cell culture without ncAA supplemented in the medium was treated with the same protocol and served as control experiment.

Microscopy

Microscopy was performed using an Olympus IXA81 inverted microscope (Olympus America, Center Valley PA, USA) equipped with UPLSAPO 100x/1.4 oil immersion objective, a Rolera XR CCD camera (Qimaging, Surrey BC, Canada) and a Proscan H117 motorized XY stage (Prior Scientific, Rockland, MA, USA) controlled by the μManager freeware (http://www.microAmanager.org). A field lens of 1.6x magnification was used to achieve Nyquist sampling for imaging. The following fluorescence filter sets (Chroma technology Corp., Bellows Falls, VT, USA) were used, with the central wavelength and bandwidth of the excitation and emission filters as indicated: FITC (Ex. 470/40; Em. 525/50).

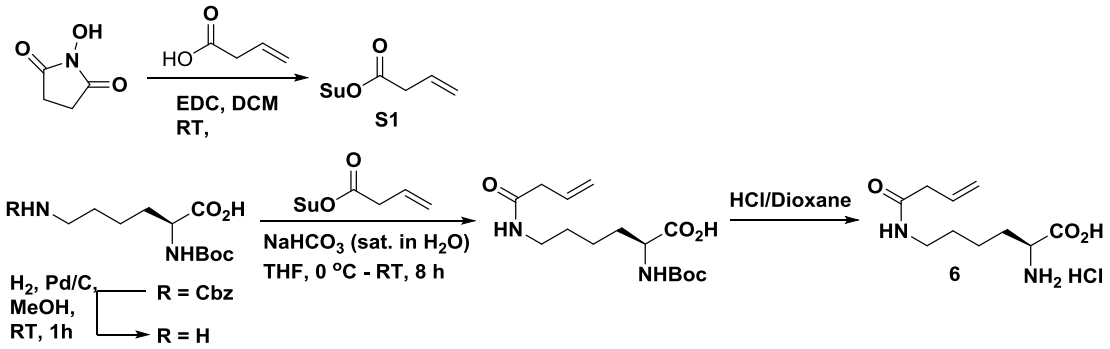
Chemical synthesis

General Experimental

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through

standard laboratory protocols. Analytical thin-layer chromatography (TLC) was performed on EMD Millipore silica gel 60 F_{254} plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.80 ppm for ^{1}H). A minimal amount of 1,4-dioxane was added as the reference standard (67.19 ppm for ^{13}C) for carbon NMR spectra in deuterium oxide, and a minimal amount of sodium hydroxide pellet or concentrated hydrochloric acid was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral conditions. ^{1}H NMR spectra are tabulated as follows: chemical shift, multiplicity (s = singlet, bs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). Mass spectra were obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.



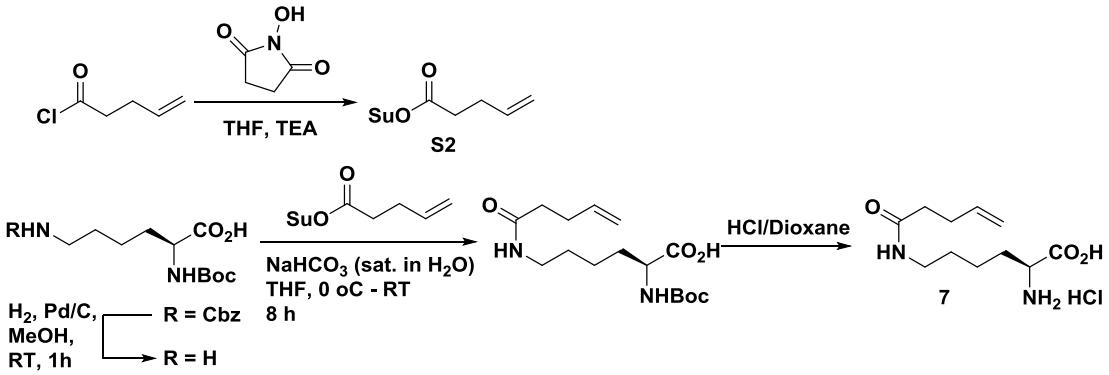
Scheme 3-1

Synthetic of N^{ε}-3-Buteenoyl-L-lysine (**6**) – **Scheme 3-1**

3-Butenoic acid succinimidyl ester (S1): To the solution of 3-butenic acid (0.5 mL, 5.9 mmol) in anhydrous dichloromethane (50 mL) was cooled to 0°C and was added EDC·HCl (1.51 g, 7.9 mmol). The solution was stirred at 0°C for 10 min followed by addition of *N*-hydroxysuccinimide in one portion in an ice bath. The mixture was stirred at RT for 8h before the reaction was diluted with addition of water. After washing with water, brine, the combined organic layer was dried over anhydrous sodium sulfate, concentrated to afford a crude product which was further purified by flash column chromatography (eluted by 50% EA/Hexane) to give a white solid as the desired product (980 mg, 90.7%). **^{1}H NMR** (300 MHz, $CDCl_3$): δ 6.71 (ddt, *J* = 10, 7, 4 Hz, 1H), 6.08-5.99 (m, 2H), 2.84 (s, 4H), 2.20 (d, *J* = 7 Hz, 2H).

***N*$^{\varepsilon}$-3-Butenoyl-L-lysine (6):** The mixture of Boc-Lys(Cbz)-OH (3.8 g, 10 mmol) and palladium on activated carbon (10% Pd basis, 0.6 g, 5 mol%) in methanol (50 mL) was bubbled through hydrogen with stirring at RT for 3h. The solution was filtered through a celite cake before the flow through solution was concentrated. The

afforded Boc-Lys-OH is clear oil and was directly used in the next step without further purification. To the solution of Boc-Lys-OH (1.36 g, 5.52 mmol) dissolved in THF (25 mL) was added $NaHCO_3$ aqueous solution (sat. in water, 25 mL) and cooled in an ice bath. 4-Pentenoic acid succinimidyl ester from previous step (851 mg, 4.6 mmol) was then added into the mixture in on portion and allowed to stir at RT. After 10h, the solution was acidified using 2 M HCl and the THF was removed *in vacuo*. The residue was dissolved in ethyl acetate and the solution was washed with water and brine, dried over anhydrous sodium sulfate and concentrated under reduced pressure. The crude product was further purified by silica gel flash chromatography (eluted at 5% MeOH/DCM) to give a colorless oil as the desired product (1.1 g, 76 %). The afforded product was then dissolved in dioxane (10 mL) and the solution was added hydrogen chloride solution (4 M in dioxane, 10 mL, 40.0 mmol). The mixture was stirred at RT for 8h, and then the resulting white suspension was filtered and washed with diethyl ether to remove organic impurities. The afforded pale white solid was lyophilized overnight to yield the title compound as an HCl salt in quant. yield (875 mg). **^{1}H NMR** (300 MHz, D_2O): δ 6.59 (m, 1H), 5.78 (dd, *J* = 15.4, 1.7 Hz, 1H), 5.68 (dd, *J* = 11.7, 1.7 Hz, 1H), 3.55 (t, *J* = 4.9 Hz, 1H), 3.08 (t, *J* = 6 Hz, 2H), 1.81-1.63 (comp m, 4H), 1.48-1.33 (m, 2H), 1.33-1.13 (m, 2H)



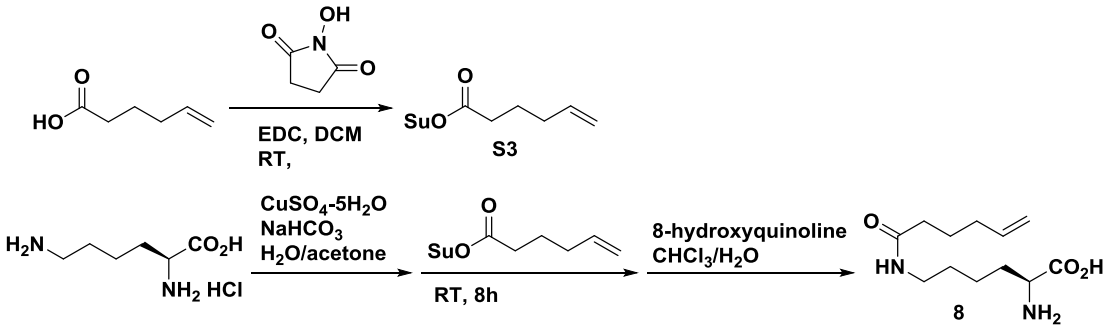
Scheme 3-2

Synthetic of N^{ε}-4-Pentenoyl-L-lysine (**7**) – **Scheme 3-2**

4-Pentenoic acid succinimidyl ester (S2): To the solution of 4-pentenoyl chloride (0.2 mL, 1.8 mmol) and triethylamine (0.28 mL, 2 mmol) in anhydrous THF (10 mL) was added *N*-hydroxysuccinimide (229 mg, 2 mmol) in one portion in an ice bath. The mixture was warmed to RT and stirred for 10h. After removing the THF, the residue was dissolved in ethyl acetate then washed with water and brine. The organic layer was dried over anhydrous sodium sulfate and concentrated to give clear brownish oil that was triturated to give a white solid as the desired product (319 mg, 90%). The afforded product was subjected to the next step without further purification. **^{1}H NMR** (300 MHz, $CDCl_3$): δ 5.84 (m, 1H), 5.19-4.96 (m, 2H), 2.81 (s, 4H), 2.63 (td, *J* = 4, 7 Hz, 2H), 2.47 (m, 2H).

***N*$^{\varepsilon}$-4-Pentenoyl-L-lysine (7):** The mixture of Boc-Lys(Cbz)-OH (3.8 g, 10 mmol) and palladium on activated carbon (10% Pd basis, 0.6 g, 5 mol%) in methanol (50 mL) was bubbled through hydrogen with stirring at RT for 3h. The solution was filtered through a celite cake before the flow through solution was concentrated. The

afforded Boc-Lys-OH is clear oil and was directly used in the next step without further purification. To the solution of Boc-Lys-OH (1.3 g, 5.4 mmol) dissolved in THF (25 mL) was added $NaHCO_3$ aqueous solution (sat. in water, 25 mL) and cooled in an ice bath. 4-Pentenoic acid succinimidyl ester from previous step (880 mg, 4.5 mmol) was then added into the mixture in on portion and allowed to stir at RT. After 10h, the solution was acidified using 2 M HCl and the THF was removed *in vacuo*. The residue was dissolved in ethyl acetate and the solution was washed with water and brine, dried over anhydrous sodium sulfate and concentrated under reduced pressure. The crude product was further purified by silica gel flash chromatography (eluted at 5% MeOH/DCM) to give a colorless oil as the desired product (1.6 g, 96 %). The afforded product was then dissolved in dioxane (12 mL) and the solution was added hydrogen chloride solution (4 M in dioxane, 12 mL, 48.0 mmol). The mixture was stirred at RT for 8h, and then the resulting white suspension was filtered and washed with diethyl ether to remove organic impurities. The afforded pale white solid was lyophilized overnight to yield the title compound in quant. yield (1.2 g). **^{1}H NMR** (300 MHz, D_2O): δ 5.92-5.74 (m, 1H), 5.11-4.97 (m, 2H), 3.70 (t, *J* = 6 Hz, 1H), 3.16 (t, *J* = 7 Hz, 2H), 2.34-2.29 (t, *J* = 7 Hz, 4H), 1.94-1.74 (m, 2H), 1.59-1.44 (m, 2H), 1.44-1.27 (m, 2H); **^{13}C NMR** (300 MHz, D_2O): δ 176.4, 174.7, 136.9, 115.6, 54.6, 38.8, 35.0, 30.0, 29.4, 28.0, 21.7.

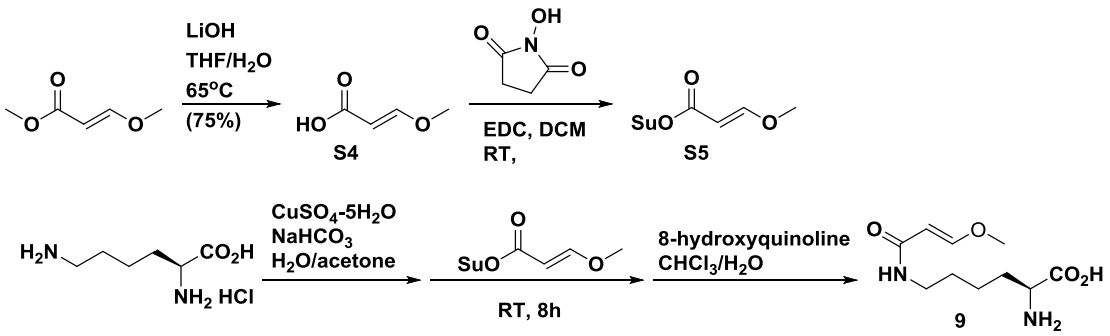


Scheme 3-3

Synthetic of N^{ε}-4-Hexenoyl-L-lysine (**8**) – **Scheme 3-3**

5-Hexenoic acid succinimidyl ester (S3): To the solution of 5-hexenoic acid (2 mL, 16.8 mmol) in anhydrous dichloromethane (80 mL) was added EDC·HCl (4.2 g, 21.9 mmol) and DMAP (107 mg, 0.9 mmol) in an ice bath. The solution was stirred for 10 min and the resulting clear solution was added *N*-hydroxysuccinimide (2.5 g, 21.7 mmol) in one portion in an ice bath. The reaction was warmed to RT and stirred under inert atmosphere for 8h before the reaction was diluted by the addition of water. The solution was then washed with water, brine, dried over anhydrous sodium sulfate and concentrated under reduced pressure. The resulting crude product was subjected to silica gel flash chromatography (eluted with 40% EA/Hexane) to afford an off-white crystal as the desired product (3.3 g, 94%). **^{1}H NMR** (300 MHz, $CDCl_3$): δ 5.87-5.67 (m, 1H), 5.15-5.08 (m, 1H), 5.08-4.96 (m, 1H), 2.84 (s, 4H), 2.63 (t, *J* = 6 Hz, 2H), 2.19 (app dd, *J* = 7 Hz, 2H), 1.86 (m, 2H).

***N*$^{\epsilon}$-4-hexenoyl-L-lysine (8)**: To the solution of lysine monohydrochloride (3.2 g, 17.4 mmol) and sodium bicarbonate (5.4 g, 64.2 mmol) in H_2O (40 mL) was added $CuSO_4 \cdot 5H_2O$ aqueous solution (2.2 g, 8.8 mmol in 20 mL H_2O) at room temperature. The resulting blue solution was stirred at room temperature for 15 minutes. The succinimidyl ester (S3) from previous step (3.5 g, 16.6 mmol) was dissolved in acetone (20 mL) and was added into the aforementioned blue solution dropwise. The mixture was stirred vigorously at room temperature for 10h to give a cloudy solution. After filtration, the pale blue solid was collected, washed with water (30 mL x3), acetone (30 mL x3) and dried to give blue solid as the amino acid copper complex. To remove the copper, the blue solid was added to a 1:1 mixture of H_2O and $CHCl_3$ (160 mL), followed by addition of 8-hydroxyquinoline (3.6 g, 24.8 mmol) in one protein and vigorous stirring at room temperature for 1h. The resulting green suspension was filtered off under reduced pressure. The filtrate was washed with $CHCl_3$ (100 mL x5) and concentrated under high vacuum pump to afford a white solid as the desired product (2.5 g, 59% yield for three steps). **^{1}H NMR** (300 MHz, D_2O): δ 5.96-5.74 (m, 1H), 5.13-4.94 (m, 2H), 3.69 (t, *J* = 5.6 Hz, 1H), 3.17 (t, *J* = 6.7 Hz, 2H), 2.22 (t, *J* = 7 Hz, 2H), 2.05 (app. q, *J* = 7 Hz, 2H), 1.92-1.76 (m, 2H), 1.76-1.59 (m, 2H), 1.59-1.45 (m, 2H), 1.45-1.27 (m, 2H); **^{13}C NMR** (300 MHz, D_2O): δ 176.7, 174.7, 138.4, 115.0, 54.6, 38.8, 35.0, 32.2, 30.0, 27.9, 24.4, 21.7.



Scheme 3-4

Synthetic of N^{ε}-(3-Methoxyacryloyl)-L-lysine (**9**) – **Scheme 3-4**

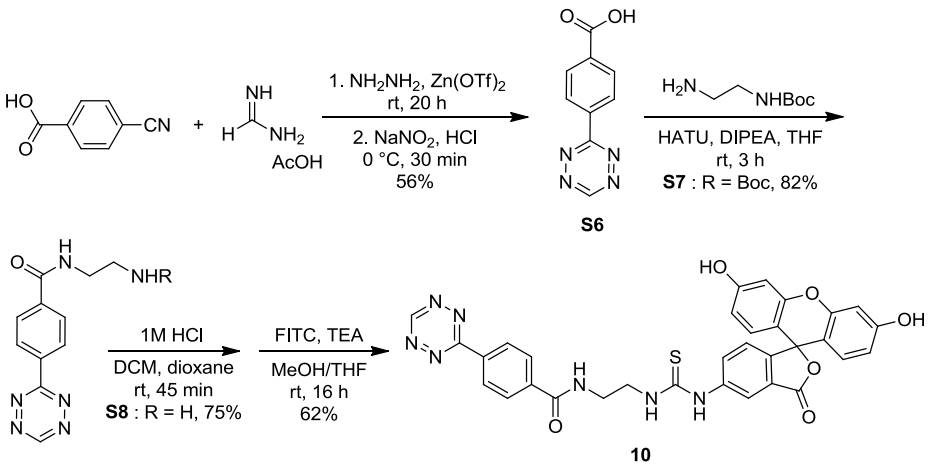
3-Methoxyacrylic acid (S4): The mixed solution of 3-methoxyacrylic acid methyl ester (4 mL, 42.3 mmol) in THF (125 mL) and $LiOH{\cdot}H_2O$ aqueous solution (10.2 g, 243 mmol in 50 mL H_2O) was stirred at 65°C for 24h. After cooling the mixture to 0°C, concentrated hydrochloric acid was added until pH = 3. The mixture was extracted with diethyl ether (120 mL x3), dried over anhydrous magnesium sulfate and concentrated under reduced pressure to give a white crystal as the desired product (3.1 g, 72%). The afforded acid was subjected to the next step without further purification.). **^{1}H NMR** (300 MHz, $CDCl_3$): δ 7.65 (d, *J* = 13 Hz, 1H), 5.14 (d, *J* = 13 Hz, 1H), 3.68 (s, 3H).

3-Methoxyacrylic acid succinimidyl ester (S5) : To the solution of 3-methoxyacrylic acid (3.1 g, 30.4 mmol) in anhydrous dichloromethane (100 mL) was added EDC·HCl (7.7 g, 40.2 mmol) and DMAP (184 mg, 1.5 mmol) in an ice bath. The solution was stirred for 10 min and the resulting clear solution was added *N*-hydroxysuccinimide (4.6 g, 40.0 mmol) in one portion in an ice bath. The reaction was

warmed to RT and stirred under inert atmosphere for 8h before the reaction was diluted by the addition of water. The solution was then washed with water, brine, dried over anhydrous sodium sulfate and concentrated under reduced pressure. The resulting crude product was subjected to silica gel flash chromatography (eluted with 40% EA/Hexane) to afford an off-white crystal as the desired product (3.4 g, 53%). **^{1}H NMR** (300 MHz, $CDCl_3$): δ 7.81 (d, *J* = 12 Hz, 1H), 5.35 (d, *J* = 13 Hz, 1H), 3.78 (s, 3H), 2.81 (s, 4H).

***N*$^{\varepsilon}$-(3-Methoxyacryloyl)-L-lysine (9) :** To the solution of lysine monohydrochloride (3.2 g, 17.4 mmol) and sodium bicarbonate (5.4 g, 64.2 mmol) in H_2O (40 mL) was added $CuSO_4{\cdot}5H_2O$ aqueous solution (2.2 g, 8.8 mmol in 20 mL H_2O) at room temperature. The resulting blue solution was stirred at room temperature for 15 minutes. The succinimidyl ester (S5) from previous step (3.3 g, 16.6 mmol) was dissolved in acetone (20 mL) and was added into the aforementioned blue solution dropwise. The mixture was stirred vigorously at room temperature for 10h to give a cloudy solution. After filtration, the pale blue solid was collected, washed with water (30 mL x3), acetone (30 mL x3) and dried to give blue solid as the amino acid copper complex. To remove the copper, the blue solid was added to a 1:1 mixture of H_2O and $CHCl_3$ (160 mL), followed by addition of 8-hydroxyquinoline (3.6 g, 24.8 mmol) in one protein and vigorous stirring at room temperature for 1h. The resulting green suspension was filtered off under reduced pressure. The filtrate was washed with $CHCl_3$ (100 mL x5) and concentrated under high vacuum pump to afford a white solid as the desired product (2.5 g, 62% yield for three steps). **^{1}H NMR** (300 MHz, D_2O): δ 7.44 (d, *J* = 12 Hz, 1H), 5.38 (d, *J* = 12 Hz, 1H), 3.68 (app s, 4H), 3.22 (t, 6.7 Hz, 2H), 1.84 (m, 2H),

1.55 (m, 2H), 1.39 (m, 2H); **^{13}C NMR** (300 MHz, D_2O): δ 175.0, 169.6, 160.2, 97.8, 57.3, 54.6, 38.7, 30.1, 28.1, 21,7.



Scheme 3-5

Synthesis of Tetrazine Dye (**10**) – **Scheme 3-5**

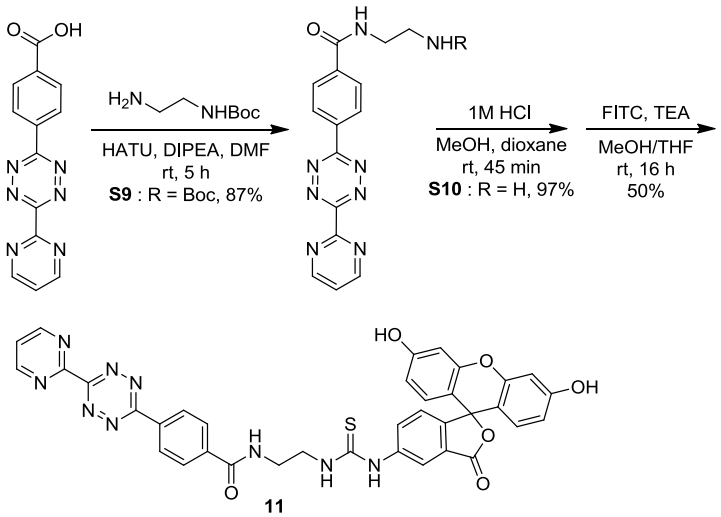
Tetrazine S6: 4-cyanobenzoic acid (0.5 g, 3.4 mmol), formamidine acetate salt (1.76 g, 17.0 mmol) and zinc (II) triflate (0.6 g, 1.7 mmol) were mixed together under nitrogen and cooled to 0 °C. Anhydrous hydrazine (4.2 mL, 136 mmol) was added slowly and the stirring reaction mixture was allowed to slowly warm to r.t. After 20 h, the dark brown mixture was cooled to 0 °C and an ice-cold, aqueous solution of sodium nitrite (2.3 g, 34.0 mmol) was added. This was followed by slow addition of ice-cold 2 M HCl aq, during which toxic nitrogen oxide gases evolved and the mixture turned bright pink. The addition of aqueous HCl was continued until pH 2-3 and stirring was continued for another 30 min at this temperature. Ethyl acetate (20 mL) was added and

the phases were separated. The aqueous layer was extracted with four portions of EtOAc or until the last organic extract was faint pink in color. The organic layers were combined, washed with 1 M HCl, water and brine, dried over Na_2SO_4, and concentrated under reduced pressure. The obtained residue was purified by column chromatography on silica gel using DCM:acetone (3:1) to give tetrazine **S6** (391.4 mg, 56%) as a bright pink solid. Analytical data matched previous literature values (*Bioconjugate Chem.* **2011**, *22*, 2263).

Tetrazine S7: Tetrazine **S6** (100 mg, 0.49 mmol) was dissolved in a solution of DIPEA (0.21 mL, 1.22 mmol) in dry THF (7 mL). HATU (0.28g, 0.73 mmol) was added and the reaction mixture was allowed to stir at r.t. for 2 h under nitrogen before the addition of N-Boc-ethylenediamine (0.115 mL, 0.73 mmol). After 1 h, the reaction mixture was concentrated and the residue was dissolved in EtOAc (15 mL), washed with 5% citric acid, water and brine, dried over Na2SO4, and concentrated under reduced pressure. The obtained residue was purified by column chromatography on silica gel using 20% acetone in DCM to afford tetrazine **S7** (138 mg, 82%) as a bright pink solid. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 10.25 (s, 1H), 8.68 (d, *J* = 8.4 Hz, 2H), 8.06 (d, *J* = 8.4 Hz, 2H), 7.66 (m, 1H), 5.15 (t, *J* = 6.0 Hz, 1H), 3.59 (q, *J* = 4.8, 5.4 Hz, 2H), 3.45 (q, *J* = 4.8, 5.4 Hz, 2H), 1.43 (s, 9H). ^{13}C NMR ($CDCl_3$, 75 MHz) δ 166.8, 166.1, 158.1, 138.3, 134.2, 128.5, 128.1, 80.4, 42.9, 39.9, 28.5. ESI-MS (*m/z*): $[M+K]^+$ calcd for $C_{16}H_{20}N_6O_3$ 383.4662, found 383.1213.

Tetrazine S8: Tetrazine **S7** (90 mg, 0.26 mmol) was suspended in dry MeOH (4 mL) and cooled to 0°C before the slow addition of 4N HCl in 1,4-dioxane (1.3 mL). The reaction mixture was stirred at r.t. for 45 min. The volatiles were removed under reduced pressure and the remaining residue was suspended in diethyl ether. The solvent was decanted and the procedure was repeated two more times. The pink solid was collected and dried under vacuum to yield tetrazine **S8** (55.3 mg, 75%). The crude material was considered pure enough for subsequent reactions and the identity was confirmed by ESI-MS (*m/z*): $[M+H]^+$ calcd for $C_{11}H_{12}N_6O$ 245.1145, found 245.1142.

Tetrazine dye (10): Tetrazine **S8** (12 mg, 0.042 mmol) was dissolved in a solution of TEA (40 µL, 0.28 mmol) in dry MeOH and THF (2:1, 1.2 mL). The reaction mixture was cooled to 0 °C before adding fluorescein 5-isothiocyanate (FITC) (11 mg, 0.028 mmol). The reaction mixture was allowed to warm to rt and stirred overnight before concentrating and purifying by column chromatography on silica gel using a gradient of DCM:acetone:MeOH (80:15:5 to 45:50:5) to afford **10** (10 mg, 62%) as an orange solid. ^{1}H NMR (Acetone-d_6, 300 MHz) δ 10.45 (s, 1H), 9.57 (bs, 1H), 9.02 (bs, 2H), 8.60 (d, *J* = 8.7 Hz, 2H), 8.43 (bs, 0.5H), 8.30 (bs, 0.5H), 8.16 (d, *J* = 8.7 Hz, 2H), 8.05 (bs, 0.5H), 7.88 (bs, 0.5H), 7.19 (d, *J* = 8.1 Hz, 1H), 6.69-6.54 (m, 8H), 3.91 (m, 2H), 3.76 (q, *J* = 6.0, 6.4 Hz, 2H). ^{13}C NMR (Acetone-d_6, 100 MHz) δ 182.2, 168.5, 168.1, 159.6, 158.5, 152.7, 148.8, 141.3, 135.0, 130.6, 129.5, 128.5, 128.1, 124.5, 118.3, 112.6, 111.1, 102.6, 54.8, 39.7, 31.3. ESI-MS (*m/z*): $[M+H]^+$ calcd for $C_{32}H_{23}N_7O_6S$ 634.1503, found 634.1462.



Scheme 3-6

Synthesis of tetrazine dye (**11**) – **Scheme 3-6**

Tetrazine S9: 4-(6-(pyrimidin-2-yl)-1,2,4,5-tetrazin-3-yl)benzoic acid (synthesized according to *Chem. Commun*. **2012**, *48*, 1736) (29 mg, 0.10 mmol) was dissolved in a solution of DIPEA (36 µL, 0.21 mmol) in dry DMF (0.6 mL). HATU (47 mg, 0.12 mmol) was added and the reaction mixture was allowed to stir at r.t. for 1 h under nitrogen before the addition of *N*-Boc-ethylenediamine (20 µL, 0.12 mmol). After 4 h, the reaction mixture was partitioned in EtOAc (10 mL) and water. The organic layer was washed with water (x3) and brine, dried over Na_2SO_4, and concentrated under reduced pressure. The obtained residue was purified by column chromatography on silica gel using 2-10% MeOH in DCM to afford tetrazine **S9** (38 mg, 87%) as a purple solid. ^{1}H NMR ($CDCl_3$, 300 MHz) δ 9.12 (d, *J* = 4.5 Hz, 2H), 8.77 (d, *J* = 8.1 Hz, 2H), 8.08 (d, *J* = 8.7 Hz, 2H), 7.64 (sb, 1H), 7.59 (t, *J* = 5.1 Hz, 1H), 5.12 (t, *J* = 4.2 Hz, 1H), 3.60 (q, *J* = 4.5, 5.7 Hz, 2H), 3.44 (m, 2H), 1.42 (s, 9H). ^{13}C NMR ($CDCl_3$, 100 MHz) δ

166.8, 164.2, 163.3, 159.5, 158.6, 158.0, 138.5, 133.9, 129.0, 128.2, 122.7, 80.3, 42.8, 39.9, 28.5. ESI-MS (*m/z*): $[M+H]^+$ calcd for $C_{20}H_{23}N_8O_3$ 423.18876, found 423.19028.

Tetrazine S10: Tetrazine **S9** (28 mg, 0.066 mmol) was dissolved in dry MeOH (1.5 mL) and cooled to 0 °C before the slow addition of 4N HCl in 1,4-dioxane (0.5 mL). The reaction mixture was stirred at rt for 45 min. The volatiles were removed under reduced pressure and the remaining residue was suspended in diethyl ether. The solvent was decanted and the procedure was repeated two more times. The purple solid was collected and dried under vacuum to yield tetrazine **S10** (23 mg, 97%). The crude material was considered pure enough for subsequent reactions and the identity was confirmed by ESI-MS (*m/z*): $[M+H]^+$ calcd for $C_{15}H_{15}N_8O$ 323.13633, found 323.13746.

Tetrazine 11: Tetrazine **S10** (13 mg, 0.036 mmol) was dissolved in a solution of TEA (39 µL, 0.28 mmol) in dry MeOH and THF (2:1, 1.2 mL). The reaction mixture was cooled to 0 °C before adding fluorescein 5-isothiocyanate (FITC) (11 mg, 0.028 mmol). The reaction mixture was allowed to warm to rt and stirred overnight before concentrating and purifying by column chromatography on silica gel using a gradient of DCM:acetone:MeOH (80:10:10 to 0:90:10) to afford tetrazine **11** (10 mg, 50%) as an orange solid. ^{1}H NMR (CD_3OD, 400 MHz) δ 9.16 (d, 2H, *J* = 4.8 Hz), 8.75 (d, 2H, *J* = 8.4 Hz), 8.12 (d, 2H, *J* = 8.4 Hz), 8.06 (d, br, 1H, *J* =2.4 Hz), 7.81 (t, 1H, *J* = 5.2 Hz), 7.75 (d, br, 1H, *J* = 7.2 Hz), 7.14 (d, 2H, *J* = 8.0 Hz), 6.69-6.63 (m, 5H), 6.51-6.48 (m, 3H), 3.95 (m, 2H), 3.72 (m, 2H). ESI-MS (*m/z*): $[M+H]^+$ calcd for $C_{36}H_{26}N_9O_6S$ 712.17213, found 712.17463.

CHAPTER IV

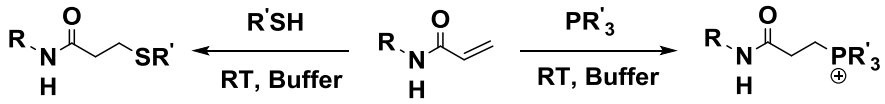
PHOSPHINE-ACRYLAMIDE 1,4-ADDITION AS NEW CLICK REACTION FOR SITE-SPECIFIC PROTEIN LABELING ON A GENETICALLY ENCODED ACRYLAMIDE FUNCTIONALITY

Introduction

Specifically target biomolecules in cellular context is crucial to the investigation of biological events. In particular, precisely probe and track the proteins of interest is essential to visualize, isolate and quantify the proteins from the complex biological mixture, in turn to understand protein dynamic, function and interactive network. [10] Development of toolbox to specifically tag desired proteins accordingly has arisen to an emerging challenge to the research at chemistry-biology interface. Classic residue-specific protein modification using cysteine or lysine chemistry appeared inadequate due to the non-specificity from identical residues of undesired proteins. Whereas, the approach of genetic introduction with a fluorescent protein tag, producing a traceable protein fusion, successfully circumvents the limitation. Though convenient, the large size of fluorescent protein may risk in perturbing native performance of the protein of interest. The development of small molecule based, highly selective, versatile and biocompatible methodology is high desirable. To this end, click chemistry directed bioorthogonal labeling has emerged as an all-around strategy for programmable protein engineering in living system.

Typical bioorthogonal reactions include Cu(I)-catalyzed or strain-promoted azide-alkyne cycloaddition, Staudinger ligation and hydrazide/oxime conjugates, that derive from functional groups of azide-alkyne, aryl phosphine-azide, and aldehyde/ketone-hydrazide/alkoxylamine. Despite there is strong demand on promoting bioorthogonal chemistry, however, there are merely handful functionalities in the inventory. The pressing need to develop new bioorthogonal reaction pairs is prevailing in chemical biology research. Recently alkene has drawn enormous emphasis in this regard because of alkene's inter nature in biological system as well as high versatility that involves many types of organic transformations. Cyclic alkenes such as norbornene, trans-cyclooctene or cyclopropene react rapidly with tetrazine or nitrile imine.[121] Nonetheless, the stability, size and chemical accessibility of these strained alkenes are concerns and may not be optimal options when minimal structural perturbation of proteins is necessary. With regard to this, development of unstrained alkenes as clickable chemical reporter will be beneficial and expand the repertoire of available tools for bioconjugation.[26, 122] Among the alkenyl chemical reporters, acrylamide has advantages over others in its small size and demonstrated multi-faceted reactivity in biological system.[26] Acrylamide is essentially electron-deficient. Given that there is no electrophilic residue presented in canonical amino acids, an acrylamide-containing protein will be unique, serving as a specific acceptor for nucleophilic addition. On the other words, designing of a nucleophilic chemical probe to target acrylamide will potentially be an inimitable click reaction for bioconjugate applications. To explore this, we previously investigated the thiol nucleophilic 1,4-addition to acrylamide and

demonstrated the reaction kinetics (**Scheme 4-1**).[26] While acrylamide is stable against thiols at pH 7.4, the thiol addition reactivity is promoted when pH is elevated to 8.8, which allowed achieving protein PEGylation site-specifically on the acrylamide-incorporated protein. Although thiol-acrylamide 1,4-addition is capable of generating valuable bioconjugates *in vitro*, however, the slow kinetics is insufficient, impeding its application to complicated biological system or living cells due to the stringent reaction parameters. Therefore, the search and development of an alternative nucleophilic probe that gives accelerated kinetics onto the acrylamide at physiological pH is invaluable. Phosphine is a strong Lewis base and has superior reactivity toward α,β-unsaturated compound via conjugate addition. Using trialkylphosphine as catalyst, the hydroalkoxylation of conjugated olefin as well as the nucleophilic thiol-ene reaction were accomplished, where phosphine addition to a α,β-unsaturat alkene is kinetically favored.[123, 124] Based on this observation, we proposed that the trialkylphosphine is an unrivaled candidate to execute the nucleophilic 1,4-addition to an acrylamide in aqueous buffer (**Scheme 4-1**). Here, we report the kinetics analysis and development of phosphine-acrylamide 1,4-addtion as unprecedented click reaction for bioconjugation applications.



Scheme 4-1. Nucleophilic 1,4-addition to acrylamide with thiol and alkyl phosphine.

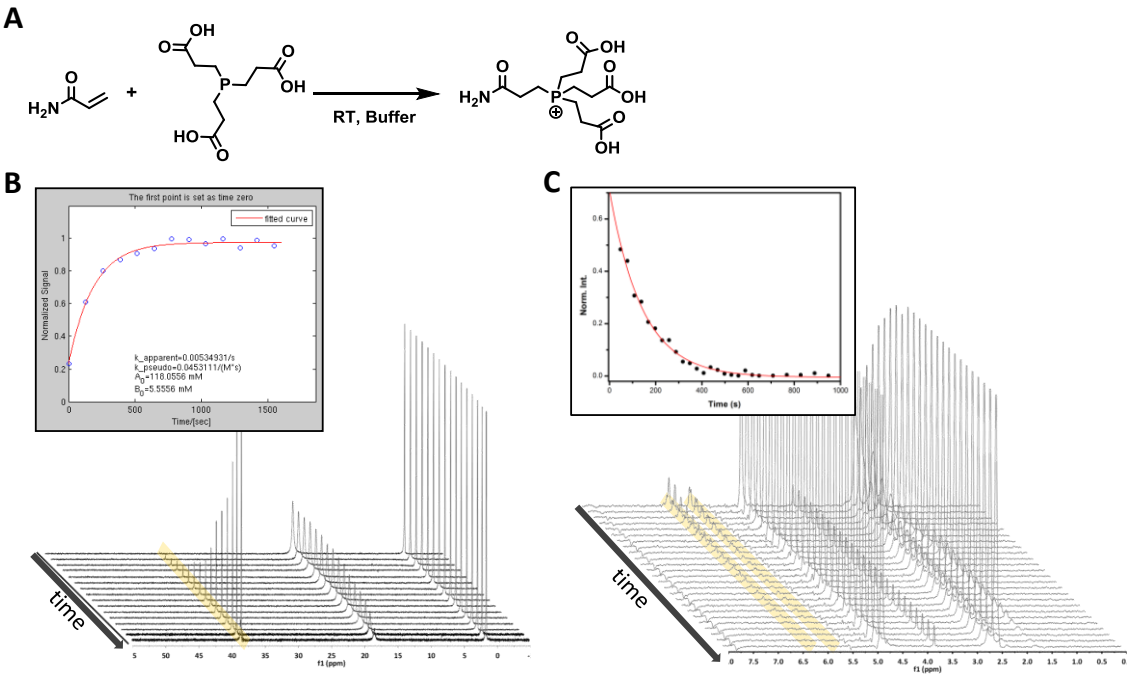


Figure 4-1. Kinetic analysis of nucleophilic 1,4-addition between *tris*(2-carboxyethyl)phosphine (TCEP) and acrylamide by ^{31}P and ^{1}H NMR. (a) Reaction scheme. (b) Time-course ^{31}P NMR experiment of phosphine 1,4-addition on acrylamide. The formation of the product was monitored by ^{31}P NMR signal (c) ^{1}H NMR experiment of acrylamide-phosphine reaction monitored at the proton peak at 5.83-5.74 ppm and 6.35-6.10 ppm. The disappearances of signals were monitored. Integration of the protein signal was plotted.

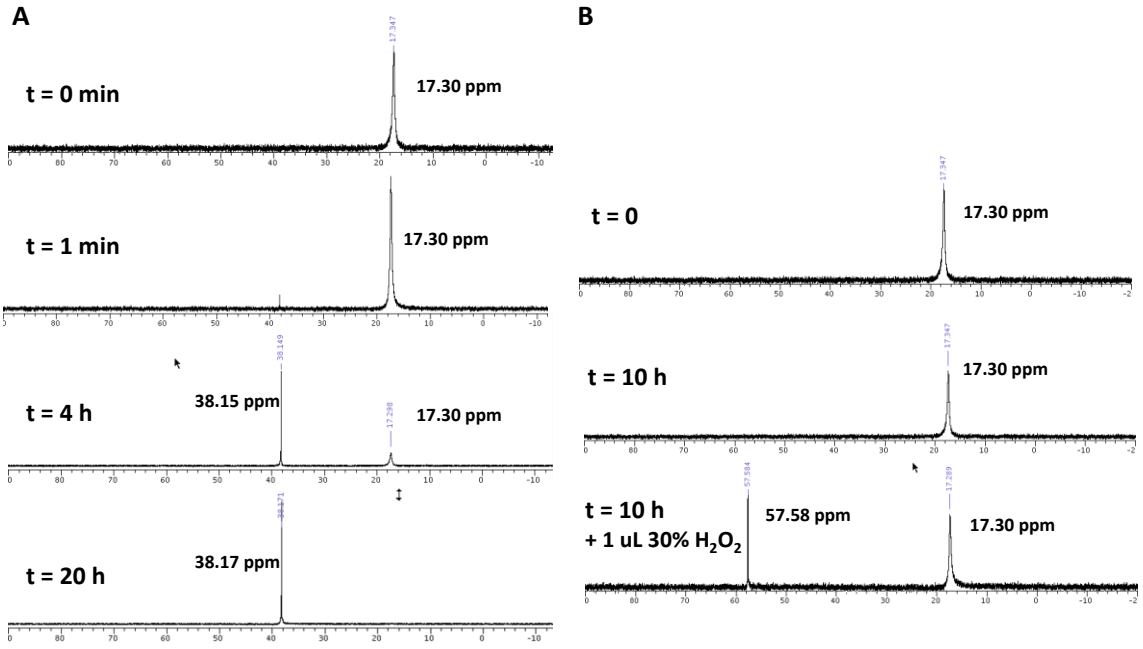


Figure 4-2. ^{31}P NMR analysis of the phosphine nucleophilic 1,4-addition to acrylamide and demonstrated the air stability of TCEP. (A) ^{31}P NMR spectrum of the reaction after 20h reaction. The 1,4-addition product has chemical shift at 38.15 ppm. (B) ^{31}P NMR spectrum of the reaction after 10h incubation, followed by addition of 1 uL H_2O_2 to induce TCEP oxidation where the phosphine oxide has the chemical shift at 57.58 ppm.

Results and Discussion

To test the 1,4-addition reactivity between alkyl phosphine and acrylic alkene in biological setup, we first look into the reaction kinetics of model reactants in aqueous buffer. An air-stable, aqueous-friendly and biocompatible phosphine is highly desirable and we found the *tris*(2-carboxyethyl)phosphine (TCEP) is an exceptionally felicitous and easy accessible candidate for initial study. Given the high solubility and stability in water, TCEP has been a well-adapted mild reductant employed in various biochemical assays both *in vitro* and on live cells. In coupled with acrylamide as the 1,4-addition

acceptor, the reaction between TCEP and acrylamide was investigated in aqueous buffer of physiological pH at room temperature (**Figure 4-1A**). Monitoring chemical reaction with phosphorus-31 nuclear magnetic resonance spectroscopy (^{31}P NMR) is advantageous due to its decent spectrum and facile interpretation. The progression of the reaction is evidenced by the consumption of the starting TCEP ^{31}P NMR signal at 17.6 ppm and the concomitant formation of phosphonium product with ^{31}P NMR signal at 37.4 ppm (**Figure 4-1B** and **Figure 4-2**). Analysis of NMR signal integration indicated the second order rate constant of 1,4-addition determined was 0.060 ± 0.01 M^{-1} s^{-1} (**Figure 4-1B**). It is notable, moreover, that after overnight reaction there is no detectable phosphine oxide byproduct observed, suggesting the stability of TCEP in aqueous condition adopted (**Figure 4-2**). Additionally, a parallel time course experiment using ^{1}H NMR was performed. The depletion of the sp2 proton signal of acrylamide at 6.35-6.10 ppm and 5.83-5.74 ppm were monitored (**Figure 4-1C**) and the kinetic analysis inferred a second order rate constant of 0.067 ± 0.002 M^{-1} s^{-1}, that is in consistent with that of ^{31}P NMR at the same condition. The determined second-order rate constants are comparable to those of the Staudinger ligation (k = 0.002 M^{-1} s^{-1}) and the copper-free dibenzocyclooctyne–azide cycloaddition (k = 0.0565 M^{-1} s^{-1}), two reactions that have been well-endorsed for bioconjugation in live cells. In comparison to the antecedent 1,4-addition with thiol nucleophile, it is more kinetically favored with phosphine, approximately 20 times better, at the same pH. This promoted reaction rate will be a merit whenever live cell labeling is needed. Since there is no electrophilic residue of twenty canonical amino acids inventory, implementation of electrophilic functional

group such as acrylamide onto proteins will be beneficial, granting a unique chemical handle for programmable site-specific bioconjugation.

Acrylamide has recently been introduced into proteins in the form of N^{ε}-acryloyl-L-lysine (AcrK) and *p*-acrylamido-L-phenylanaline.[26, 125] We took a step further to examine if TCEP-acrylamide chemistry can be transplanted and employed in the protein modification. Using an evolved pyrrolysyl-tRNA synthetase and cognate tRNA pair, our group previously reported the genetic incorporation of AcrK into super folder green fluorescence proteins (sfGFPAcrK) and demonstrated the protein bioconjugation with diverse reaction types including protein PEGylation via thiol nucleophilic 1,4-addition, albeit the reactivity is low.[26] Same protocol was followed to recombinantly produce sfGFPAcrK which has an acrylamide displayed at the Ser2 position. TCEP (2 mM) was incubated with sfGFPAcrK (50 uM) in a Tris-HCl buffer (pH 6.8 or pH 8.8) at 37 ^{o}C for 1h. 37 ^{o}C was used to elevate the reaction rate. Following buffer exchange to remove excessive phosphine, the product was subjected to electrospray ionization mass spectroscopy (ESI-MS) analysis. Except a minor peak of unmodified protein (27762 Da), a major detected peak (28013 Da) agrees well with the calculated labeling adduct (28012 Da) (**Figure 4-3**). Moreover, due to the intrinsic acidity of TCEP phosphine in water (pKa = 7.6), the phosphine reactivity/ nucleophilicity is pH dependent. This was in consistent with our observation of reaction conversion at different pH, approximately 80% and 90% modification for pH 6.8 and pH 8.8, respectively (**Figure 4-3**). Additionally, we previously demonstrated thiol 1,4-addition to acrylamide of the same protein using β-mercaptoethanol. In compare to phosphine, β-mercaptoethanol modifies

the protein much slower, giving roughly 50% conversion at 37°C, pH 8.8 and 8h incubation.[26] Therefore, attributed to the high selectivity, phosphine-acrylamide 1,4-addition is not only compatible with protein bioconjugate, but more importantly, the improved kinetics grants a much practical and sensible protocol of biological applications.

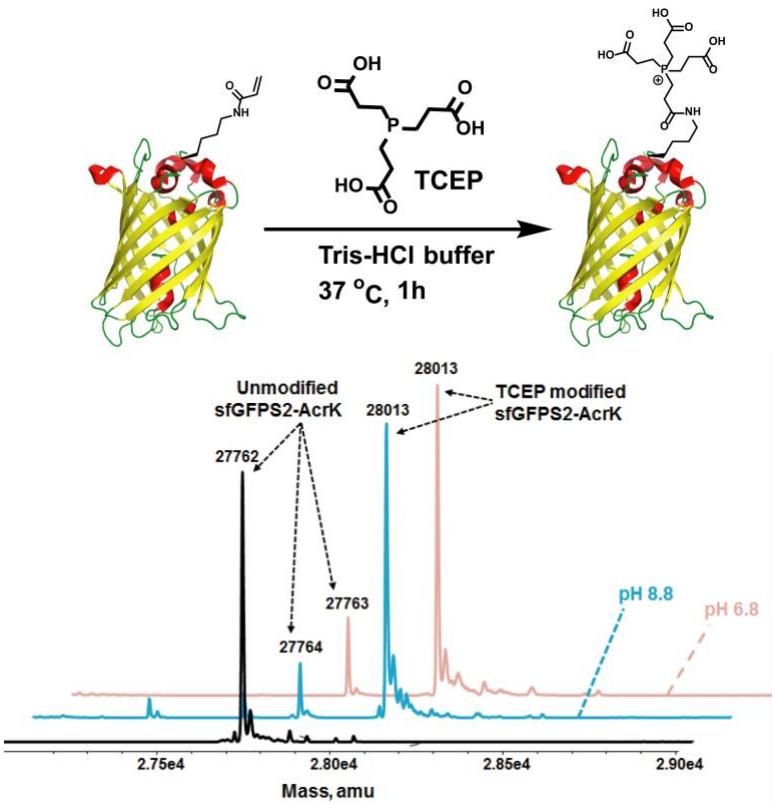


Figure 4-3. Electrospray ionization mass spectroscopy (ESI-MS) analysis of acrylamide-containing super green fluorescence protein (sfGFPS2AcrK) modified with TCEP via nucleophilic 1,4-addition. (A) The reaction scheme. SfGFPAcrK was treated with TCEP in Tris-HCl buffer at 37°C for one hour. After buffer exchange, the product was then subjected to ESI-MS. (B) Stacked MS spectra of original and TCEP-treated sfGFPS2AcrK protein at different pH. Black trace indicates the untreated sfGFPAcrK protein (calculated: 27764 Da; found: 27762 Da). Blue trace and pink trace indicate TCEP treated sfGFPAcrK under buffers of pH 8.8 and pH 6.8, respectively, giving the TCEP conjugated product (calculated: 28012 Da; found: 28013 Da).

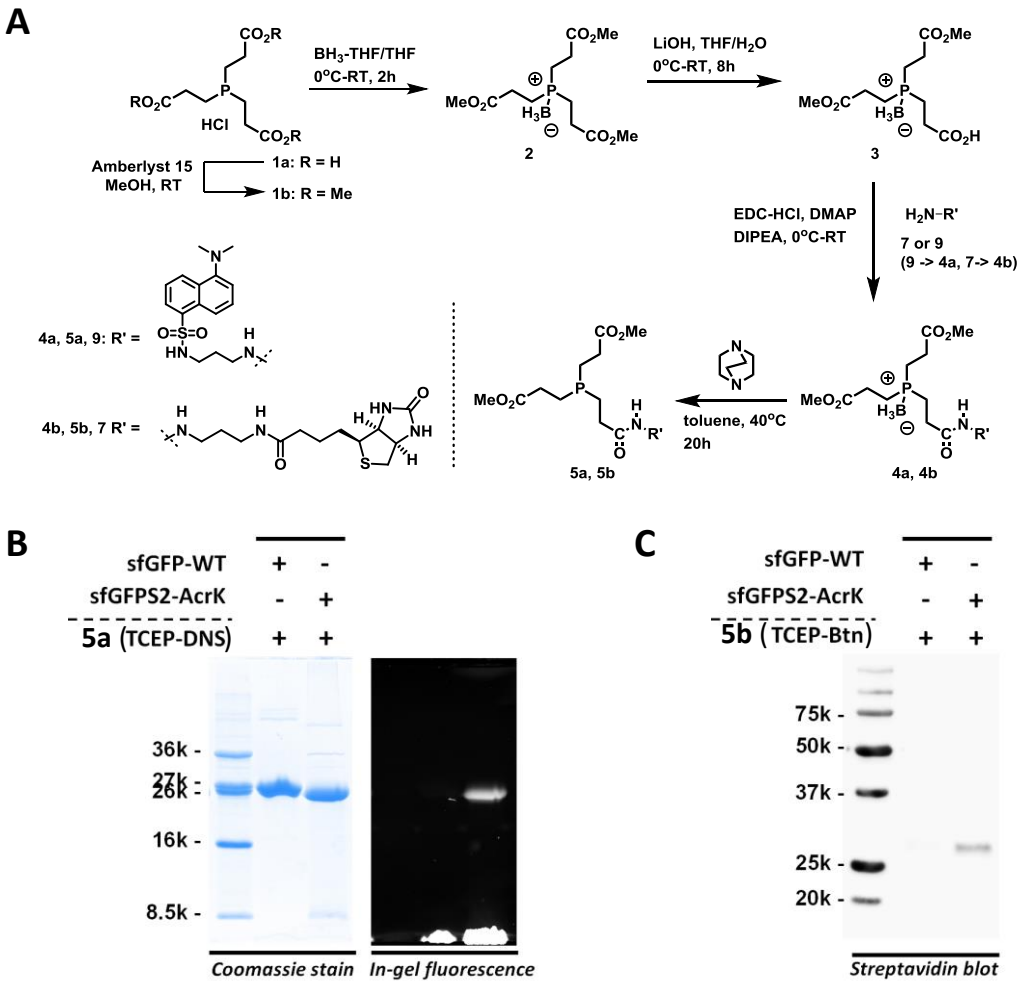


Figure 4-4. (A) Synthetic scheme of phosphine probes. (B) Bioorthogonal fluorescence labeling of the sfGFPS2AcrK protein with phosphine-dansyl probe **5a**. (c) Streptavidin blotting of the sfGFPAcrK protein modified with phosphine-biotin probe **5b**.

Since phosphine and acrylamide form a favorable bioorthogonal reaction pair, to further extend the use of reaction in biochemical interrogation, we proceeded further to design and synthesize functionalized phosphine probes (**Figure 4-4A**). Briefly, utilizing

TCEP as scaffold, TCEP was first methyl-triesterified, followed by treatment of borane tetrahydrofurane solution to form a phosphine-borane complex in order to reinforce phosphine's ambient stability. Stoichiometric hydrolysis of the triester afforded the mono-acid-di-ester phosphine-borane intermediate **3**. The accessible free acid was conjugated with choice of reporters such as fluorophore or biotin using simple amide coupling. Finally, decomplexation of borane by 1,4-diazabicyclo[2.2.2]octane afforded the phosphine **5a** as a dansyl fluorescence probe and **5b** as a biotin probe (**Figure 4-4**).

With **5a** in hand, we proceeded to demonstrate fluorescent protein labeling on the genetically encoded acrylamide. sfGFPAcrK (70 uM) was incubated with probe **5a** (2 mM) in Tris-HCl buffer (pH 7.5) at room temperature (rt) for 5h. After fully denatured, the resulting protein was subjected to SDS-PAGE analysis. In-gel fluorescence detected appreciable signal where the detected area agrees well with the protein band after Coomassie blue staining (**Figure 4-4B**). However, sfGFP protein without AcrK incorporated did not show any positive signal under the same labeling condition (**Figure 4-4B**). Interestingly, the mobility of 5a-modified sfGFPAcrK on SDS-PAGE is slightly higher, migrating farer than that of AcrK-free sfGFP. This may due to the contribution from additional positive charge derived from the formation of phosphonium product after 1,4-addtion. Similarly, biotin probe 5b was incubated with sfGFPAcrK. Followed by buffer exchange and SDS-PAGE, the resulting proteins were transferred onto a nitrocellulose membrane. After sequential treatment of blocking reagent and streptavidin-HRP, the membrane was subjected to chemiluminescent detection. As

expected, the sfGFPAcrK shows detectable signal, whereas sfGFP without AcrK incorporated doesn't (**Figure 4-4C**).

With the demonstration of the phosphine-acrylamide 1,4-addition for protein labeling *in vitro*, we then proceeded to test this reaction to label AcrK of proteins in live cells. Two previously constructed plasmids pEVOL-pylT-PrKRS and pETDuet-OmpXTAG were used to transform *E. coli* BL21 cells to express an *E. coli* outer membrane protein OmpX with AcrK incorporated at its extracellular domain (OmpXAcrK). Plasmid pEVOL-pylT-PrKRS carried genes for PrKRS and $tRNA_{CUA}$ and plasmid pETDuet-OmpXTAG contained a gene coding OmpX with an AAAXAA (A denotes alanine and X denotes an amber mutation) insertion between two extracellular residues 53 and 54. The expression of the protein was performed in LB medium supplemented with 1 mM IPTG and 5 mM AcrK afforded OmpXAcrK overexpression (**Figure 4-4B**). Incubating the lysate of cells expressing OmpXAcrK with **5a** (3 mM) at rt for 7h led to a detectable fluorescent band in a SDS-PAGE gel that corresponded well with the Coomassie blue stained OmpXAcrK in the same gel. While many proteins in the lysate were present at higher expression levels, in-gel fluorescence imaging showed that only the expressed OmpXAcrK could be specifically labeled, giving virtually no discernible background. Following the successful cell lysate labeling, experiments were conducted to fluorescently label living cells with **5a** by incubating the PBS-washed *E. coli* that overexpressed OmpXAcrK with 5a (4 mM) at rt for 1h. The cells were then washed with an isotonic saline solution and subjected to confocal microscopy imaging. *E. coli* cells expressing OmpXAcrK showed strong fluorescence (**Figure 4-4C**).

Nonetheless, cells grown in the absent of AcrK were barely fluorescent which may due to auto-fluorescence from the cell at the detection wavelength.

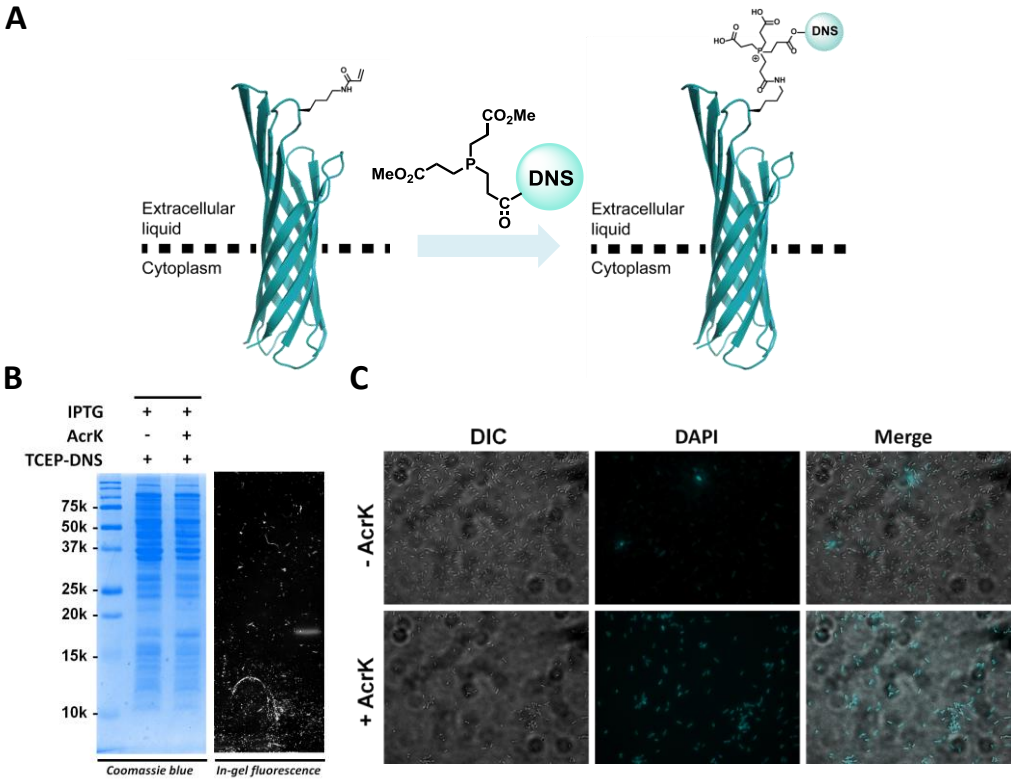


Figure 4-5. (A) Fluorescent labeling of E. coli OmpX-AcrK with phosphine probe **5a**. (B) Labeling OmpX-AcrK in a cell lysate with phosphine probe 5a. The second lane shows control without OmpX-AcrK expressed. (C) Epifluororescent cell imaging of the cells

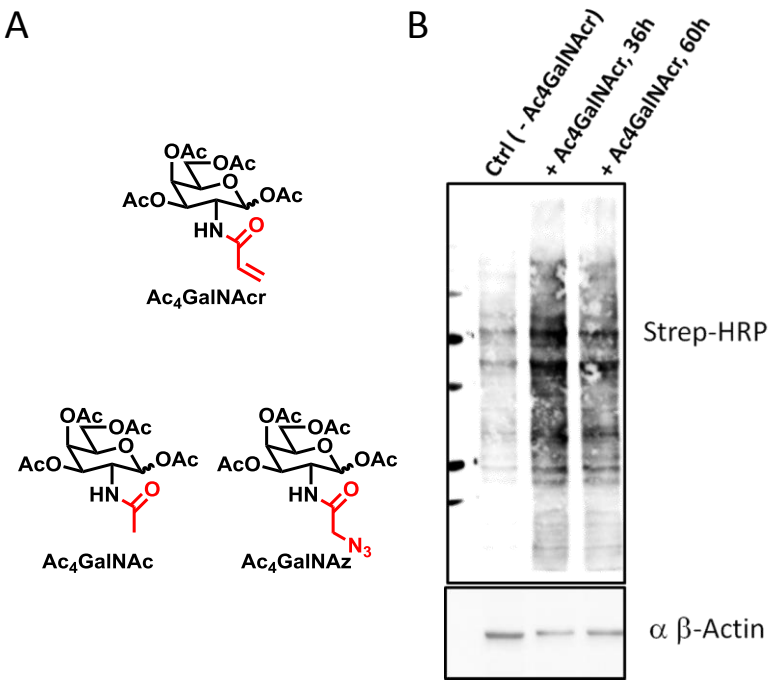


Figure 4-6. Metabolic incorporation of the acrylamide sugar and glycoprotein labeling of whole cell lysate. (A) The structural of native monosaccharide GalNAc (per-*O*-acetylated form) and non-native monosaccharide surrogates with azide, GalNAz (per-*O*-acetylated) or with acrylamide, GalNAcr (per-*O*-acetylated). (B) HEK293T whole cell lysate labeling on the acrylamide glycoprotein using the phosphine probe **5b**. The control (Ctrl) represents the whole cell lysate from the HEK293T without Ac4GalNAcr provided.

Lastly, to evaluate if phosphine-acrylamide can serve as general chemical reporter pair for extended applications, we chose to demonstrate metabolic incorporation of sugar acrylamide and labeling. This state-of-the-art chemical biology methodology involves two steps, metabolic labeling of glycans with a functionalized monosaccharide surrogate followed by labeling of probe via bioorthogonal reaction, allows the detection of specific glycoconjugate on cells or the selective capture of glycoproteins from cell lysates. It was reported previously that the unnatural monosaccharide *N*-

azidoacetylgalactosamine (GalNAz) can be utilized by cells as a substitute for its natural counterpart *N*-acetylgalactosamine (GalNAc). Acrylamide is structurally highly resembled to acetamide, by the same token we speculated the acrylamide version of GalNAc, *N*-acryloylgalactosamine (GalNAcr), will be a granted mimic, seamlessly integrated into the monosaccharide incorporation machinery. Therefore, per-*O*-acetylated *N*-acryloylgalactosamine (Ac_4GalNAcr), acetylated precursor of GalNAcr, was synthesized. HEK293T cell were incubated with Ac_4GalNAcr at a concentration of 50 uM for 36h or 60h. The cells were then washed, lysed and the cell lysates were reacted with **5b** (3 mM) for 1h. Followed by buffer exchange and SDS-PAGE, the proteins were transferred onto a nitrocellulose membrane. After sequential treatment of blocking reagent and streptavidin-HRP, the membrane was subjected to chemiluminescent detection. The labeling of the probe shows stronger signal at the acrylamide sugar supplemented cells in comparison with that of control cells without sugar. Moreover, with longer incubation of the acrylamide sugar, there is slightly reduced signal observed, and this might be due to the medium thiol gradually scavenges the acryloyl functionality.

Conclusion

To recap, we report a new click type reaction which makes use of trialkylphosphine and acrylamide to undergo nucleophilic 1,4-addtion in aqueous buffer. Trialkylphosphine reacts with α,β-unsaturated alkene preferentially to give phosphonium product. Comparing to thiol nucleophilic 1,4-addtion, kinetics analysis of the phosphine nucleophilic 1,4-addition indicates about twenty time faster reaction rate under the same condition. Use a biocompatible tris(2-carboxyethyl)phosphine as nucleophile, the acrylamide containing sfGFP is specifically labeled labeling which fall in live well with the kinetics studies. Given several advantages of acrylamide including small in size and stable in the cellular context, acrylamide and phosphine is ideal chemical reporter pair for bioconjugate applications. We designed the TCEP based fluorescence and biotin probes and demonstrated the site-specific protein labeling in vitro, in the cell lysate and on the bacteria. Moreover, to illustrate the phosphine-acryladmide 1,4-addition can generalized to other biomolecule applications, metabolic acrylamide incorporation and labeling were carried out. This reaction presents a new click labeling protocol which is specific and biocompatible and can be potentially used as an excellent alternative tool for *in vivo* protein labeling and biochemical investigations.

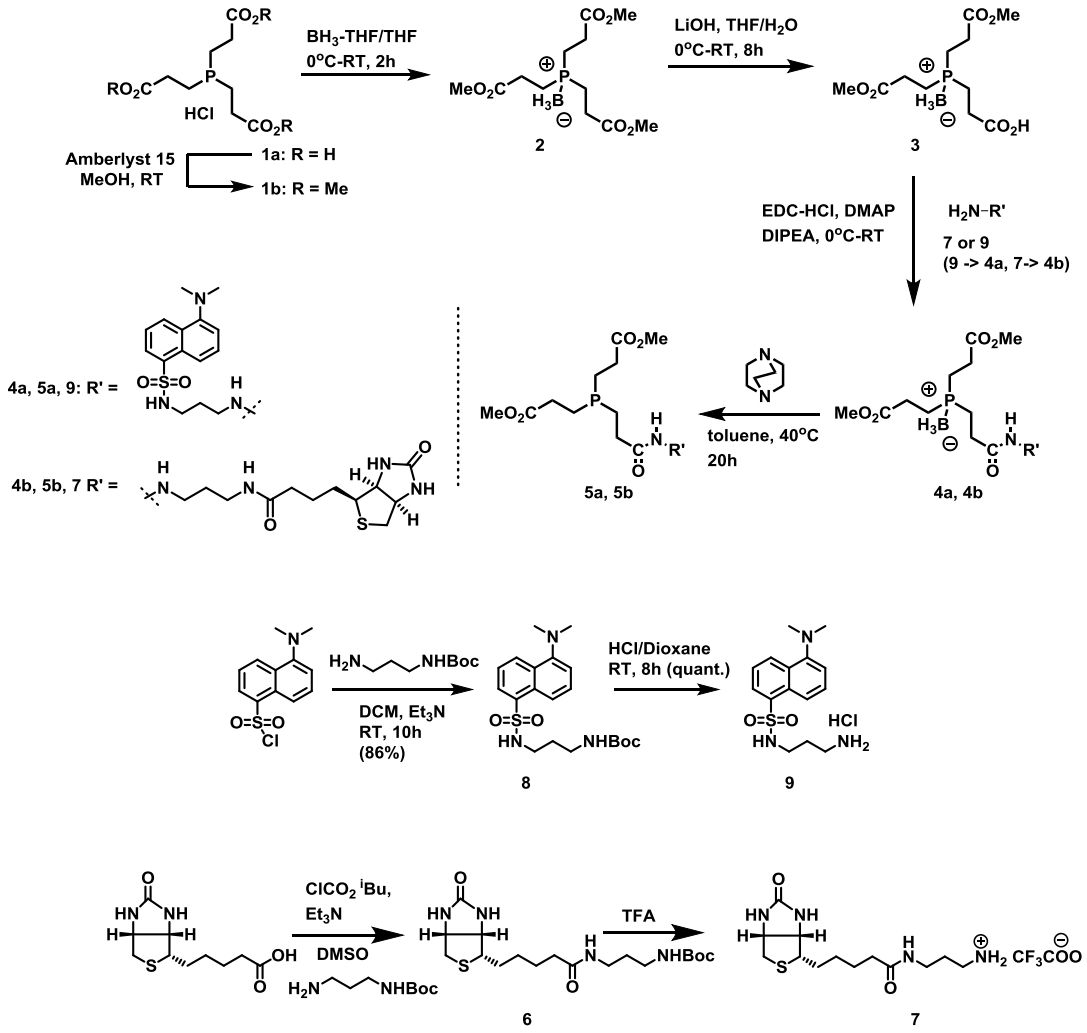
Experimental Section

General experimental

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was performed on EMD Millipore silica gel 60 F_{254} plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.80 ppm for ^{1}H). A minimal amount of 1,4-dioxane was added as the reference standard (67.19 ppm for ^{13}C) for carbon NMR spectra in deuterium oxide, and a minimal amount of sodium hydroxide pellet or concentrated hydrochloric acid was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral conditions. ^{1}H NMR spectra are tabulated as follows: chemical shift, multiplicity (s = singlet, bs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). Mass spectra were obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

Chemical synthesis



Scheme 4-2. Synthesis of Phosphine Probes **5a** and **5b**

Synthesis of Phosphine Probes **5a** and **5b** – **Scheme 4-2**

Synthesis of compound-1b: Tricarboxyethylphosphine hydrochloride (TCEP-HCl, 2.0 g, 2.08 mmol) was stirred with 2.0 mg of sulfonic acid resin (Amberlyst) in 20 mL methanol at room remperature for 2 h minutes. The resin was then removed by filtration and the filtrates, containing a mixture of mono-, di- and tri-methyl ester, were fractionated by flash chromotography (2-3% MeOH in DCM) to give compound **-1** as colorless oil (0.794 mg, 34%). **^{1}HNMR ($CDCl_3$, 300 MHz):** δ 3.70 (s, 9H), 2.56 (t, 6H, J= 9.0 Hz), 2.03-1.92(m, 6H).

Synthesis of compound-2: The compound **1b** (0.7 g, 2.39 mmol) was dissolved in dry THF in a flame-dried round bottom flask under an argon atmosphere. The solution was cooled to 0 ^{o}C and borane-THF (1.0 M in THF, 2.87 mL, 2.87 mmol) was added slowly. The reaction was stirred at 0 ^{o} C for 45 min and then at room temperature for an additional 1.5 h. The solvent was removed under reduced pressure, and the residue was purified by flash chromatography (2-5% MeOH in DCM). The phosphine-borane complex **2** was isolated as a clear oil (0.293 g, 40%). **^{1}HNMR ($CDCl_3$, 300 MHz):** δ 3.69 (s, 9H), 2.55 (t, 6H, J= 9.0 Hz), 2.00-1.91(m, 6H).

Synthesis of compound-3: The compound **2** (0.250 g, 0.81 mmol) was dissolved in dry THF/water (2:1) (18 ml) in a flame-dried round bottom flask under an argon atmosphere. The Lithium hydroxide (0.017 g, 0.408 mmol) was added slowly. The reaction was stirred at room temperature for 15 min. The reaction mixture was quenched with 1 M HCl then extraction with ethyl acetate (2x20 mL). The organic layer was dried over Na_2SO_4 and filtered. The solvent was removed under reduced pressure, and the

residue was purified by flash chromatography (2-6% MeOH in DCM). The mono-acid phosphine-borane complex **3** was isolated as a clear oil along with tri-ester and tri-acid compounds (0.071 g, 30%). **^{1}HNMR ($CDCl_3$, 300 MHz):** δ 3.64 (s, 6H), 2.56-2.47 (m, 6H) 1.96-1.87(m, 6H).

Synthesis of compound-6: To cold (5 ^{o}C) solution of D (+) Biotin (2.5.0 g, 10.25 mmol) in DMSO (10 ml) triethylamine (1.65 mL, 12.5 mmol) was added followed after 5 min stirring by isobutylchloroformate (1.3 mL, 20 mmol) in 2.5 mL dioxane. The viscous mixture was then stirred at 5 ^{o}C for 30 min. A solution of 3-(tert-butyloxycarbonylamino)propane (1.75 g, 10 mmol) and triethylamine (1,35 mL, 10 mmol) in 10 mL dioxane was then added dropwise to the mixture. The cooling bath was removed and stirring continued for 4h. During this process, the solidification of the reaction medium was gradually observed. Cold water (30 mL) was added and the compound **4** was isolated by filtration, washed twice with cold water (5 mL) and dried (3.2 g, 80%). **^{1}HNMR ($CDCl_3$, 300 MHz):** δ 6.58-6.68 (brs, 1H), 6.18-6.22 (brs, 1H), 5.33-3.40 (brs, 1H), 4.98-5.01(brs, 1H), 4.53 (q, 1H, J = 3.3 Hz), 4.34 (t, 1H, J = 4.5 Hz), 3.28-3.18 (m, 2H), 3.17-3.16(m, 2H), 2.92(dd, 1H, J = 3, 7.5 Hz), 2.73(dd, 2H, J = 3.3, 11.1 Hz), 2.25-2.23 (m, 2H), 1.68-1.48 (m, 8H), 1.44 (s, 9H). **^{13}NMR ($CDCl_3$, 75 MHz):** δ 14.9, 172.4, 163.1, 77.9, 61.5, 59.6, 55.8, 38.0, 36.5, 35.6, 33.9, 30.1, 28.6, 28.5, 28.4, 25.7, 24.9.

Synthesis of compound-7: To cold (5 ^{o}C) solution of **4** (3 g, 7.5 mmol), in 10 mL methylene chloride trifluroacetic acid (7 mL) was added and the solution was vigorously stirred for 2h at rt (monitoring by TLC). After completion, the solution was

evaporated in *vacuo* and washes with Et_2O until crystallization occurred (2.3 g, 75%). **^{1}HNMR (CD_3OD, 300 MHz):** δ 4.44-4.52(m, 1H), 4.28-4.32 (m, 1H), 3.23-3.33 (m, 1H), 3.18-3.21 (m, 1H), 2.94-2.90 (m, 2H), 2.70 (d, 1H, J = 7.8 Hz), 2.32 (t, 1H, J = 4.2 Hz), 2.25 (t, 1H, J = 4.5 Hz), 1.82-1.62 (m, 6H), 1.51-1.42 (m, 2H).

Synthesis of compound-4b: To the compound **3** (0.060 g, 0.205 mmol) in DMF solution was added EDC.HCl (0.062 g, 0.328 mmol) and DMAP (0.0025 g, 0.025 mmol) in ice both and the mixture was stirred for 10-15 minutes. The amine salt compound **7** (0.084 g, 0.203 mmol) in anhydrous DMF was added to the above mixture dropwise and followed by addition of DIPEA (0.045 mL, 0.246 mmol) dropwise in ice bath. The mixture was warmed to room temperature and stirred for 16 hours. The solvent was evaporated under reduced pressure and extracted with ethyl acetate. The organic layer was washed with brine, dried over sodium sulfate, concentrated and purified by chromatography (2-3% MeOH/DCM) to give desired product **4b** (0.053 g, 45%). **^{1}HNMR ($CDCl_3$, 300 MHz):** δ 7.09 (t, 1H, J = 3.3 Hz), 6.81 (t, 1H, J = 6.6 Hz), 6.57 (s, 1H), 5.71 (s, 1H), 4.54-4.58 (m,1H), 4.35-4.48 (m, 1H), 3.69 (s, 6H), 3.35-3.201 (m, 4H), 3.15-3.16 (m, 1H), 2.89 (dd. 1H, J = 3.3, 10.2 Hz), 2.71 (d, 1H, J = 7.5 Hz), 2.59-2.54 (m, 4H), 2.46-2.41 (m, 2H), 2.22-2.20 (m, 4H), 2.01-1.93(m, 6H), 1.72-1.61 (m, 6H), 1.46-1.41 (m, 2H). **^{13}CNMR ($CDCl_3$, 75 MHz):** δ 174.0, 172.8, 172.6, 171.5, 164.0, 61.7, 60.2, 55.7, 52.2, 52.1, 40.6, 36.5, 36.2, 35.7, 29.3, 29.1, 28.0, 27.9, 27.4, 25.5, 18.9, 18.7, 18.6, 18.6, 18.3.

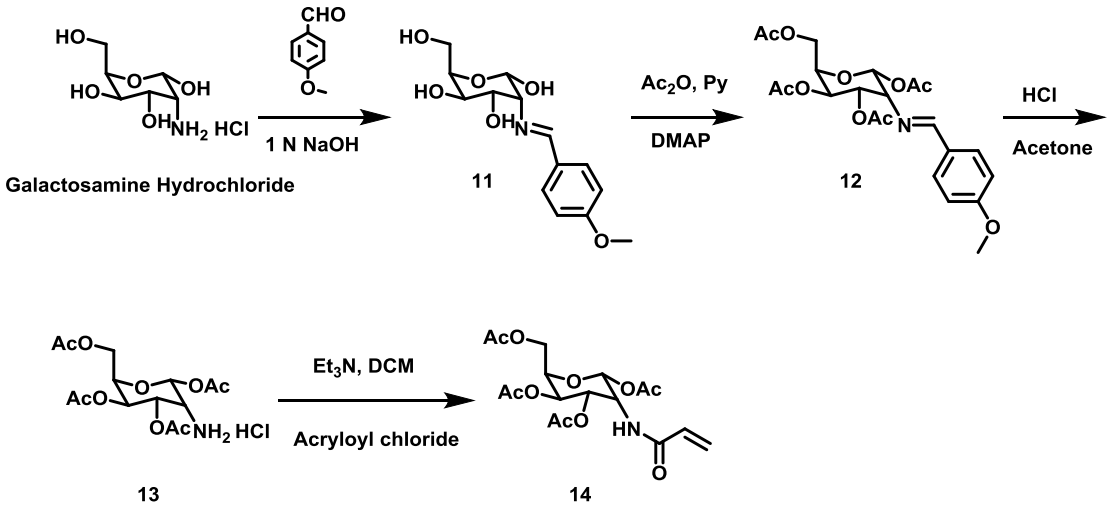
Synthesis of compound-8: To a solution of *tert*-butyl (3-aminopropyl)carbamate (0.64 g 3.71 mmol) and DIPEA (169 mL, 9.19 mmol) in anhydrous CH_2Cl_2 (15 mL) was

added dansyl chloride (1g, 3.71 mmol) in CH_2Cl_2 (10 mL) dropwise over 10 min. The solution was stirred at room temperature for another 2 h. After the removal of solvent, the residue was purified by silica gel column chromatography (hexanes/EtOAc 8:2) to give pure compound **8** (1.2 g, 83%). **^{1}HNMR ($CDCl_3$, 300 MHz):** δ 8.56 (d, 1H, J = 8.4 Hz), 8.36 (d, 1H, J = 8.4 Hz), 8.26 (dd, 1H, J = 1.2, 4.2 Hz), 7.62-7.50 (m, 2H), 7.21 (d, 1H, J= 7.2 Hz), 5.73 (brs, 1H), 4.57 (brs, 1H), 3.11 (q. 2H, J = 3.3, 12.3 Hz), 2.93 (q. 2H, J = 6.6, 13.2 Hz), 2.91 (s, 6H), 1.51-1.49 (m, 2H), 1.37 (s, 9H).

Synthesis of compound-9: **8** (1.0 g, 2.94 mmol) was added to a flask and then added HCl (1.47 mL, 4.0 M in Dioxane). The resulting mixture was stirred at room temperature overnight. The white solid was collected via filtration and dried under high vacuum to give amino acid **9** (0.541 g, 65% yield). **^{1}HNMR (CD_3OD, 300 MHz):** δ 8.91 (d, 1H, J= 8.7 Hz), 8.62 (d, 1H, J = 8.6 Hz), 8.41 (d, 1H,J = 7.2 Hz), 8.12 (d, 1H, J = 6.8 Hz), 7.95-7.87 (m, 2H), 3.50 (s, 6H), 3.03-2.99 (m, 4H),1.89-186 (m, 2H).

Synthesis of compound-4a To the compound **3** (0.050 g, 0.171 mmol) in DMF solution was added EDC.HCl (0.052 g, 0.273 mmol) and DMAP (0.0020 g, 0.0171 mmol) in ice both and the mixture was stirred for 10-15 minutes. The amine salt compound **9** (0.058 g, 0.171 mmol) in anhydrous DMF was added to the above mixture dropwise and followed by addition of DIPEA (0.037 mL, 0.205 mmol) dropwise in ice bath. The mixture was warmed to room temperature and stirred for 16 hours. The TLC assay was showing a new fluorescent spot after 16 hours. Water was added to quench the reaction and extracted with ethyl acetate. The organic layer was washed with brine, dried over sodium sulfate, concentrated and purified by chromatography (50% EA/Hexane) to

give desired product **4a** (0.06 g, 65%). **^{1}HNMR ($CDCl_3$, 300 MHz):** δ 8.55 (d, 1H, J= 5.1 Hz), 5.30 (d, 1H, J = 5.1 Hz), 8.26 (d, 1H, J = 4.2 Hz), 7.60-7.52 (m, 2H), 7.21 (d, 1H, J = 1.5 Hz), 6.01 (brs, 1H), 5.66 (brs, 1H), 3.70 (s, 6H), 3.27 (q, 2H, J = 3.9, 7.5 Hz), 2.92 (q, 2H, J = 3.9, 5.7 Hz), 2.60 (s, 6 H), 2.56-2.58 (m, 4H), 2.37-2.42 (m, 2H), 1.98-1.94 (m, 6H), 1.61-1.56 (m, 2H). **^{13}CNMR ($CDCl_3$, 75 MHz):** 172.7, 172.6, 171.7, 151.9, 134.9, 130.4, 129.8, 129.4, 129.2, 128.4, 123.2, 118.7, 115.2, 52.2, 52.1, 45.4, 40.1, 36.3, 29.3, 29.1, 27.3, 18.8, 18.6, 18.5, 18.3.



Scheme 4-3 Synthesis of Acrylamide Monosaccharide Ac4GalNAcr

Synthesis of Acrylamide Monosaccharide Ac4GalNAcr (**14**) – **Scheme 4-3**

Synthesis of compound 12: To a solution of D-Glactosamine hydrochloride (0.5 g, 2.32 mmol) in a freshly prepared aqueous solution of 1 M NaOH (4.65 mL, 4.65 mmol) under stirring was added *p*-anisaldehyde (0.379 mL, 2.79 mmol). After a short time crystallization began. Then the mixture was refrigerated off and washed with cold water, and followed by a mixture of 1:1 EtOH: Et_2O to give N-protected compound **11** (0.423 g, 65%)**.** This intermediate product **11** (0.4 g, 1.42 mmol) was added successively to a cooled (ice-water) mixture of Pyrindine (1.52 mL) and Ac_2O (2.71 mL). A catalytic amount of DMAP was added and the mixture was stirred in ice-bath for 1 hour and then at room temperature overnight. The solution was poured into 100 mL of ice-water. The precipitated white product was filtered, washed with cold water and dried give compound **2** with (0.398, 60%). %). ^{1}HNMR (CD_3Cl_3, 300MHz): δ 8.18 (s, 1H), 7.66 (d, 2H, J = 7.8 Hz), 6.92 (d, 2H, J = 7.8 Hz), 5.95 (d, 1H, J = 8.4 Hz), 5.45 (t, 1H, J = 9.6 Hz), 5.17 (t, 1H, J= 9.6 Hz), 4.38 (dd, 1H, J = 4.5, 12.3 Hz) 4.14 (dd, 1H, J = 9.0, 12.3 Hz), 4.02-3.97 (m, 1H), 3.86 (s, 3H), 3.47 (dd, 1H, J = 8.4, 9.6 Hz), 2.12 (s, 3H)**,** 2.11 (s, 3H), 2.06 (s, 3H), 1.90 (s, 3H).

Synthesis of compound 13: Concentrated hydrochloric acid was carefully added in an acetone solution of **12** (0.35 g, 0.75 mmol) kept at 0 °C in an ice bath to reach pH 5.The precipitate formed was filtered and washed with acetone to afford **13** (0.23g, 80%). ^{1}HNMR (CD_3OD, 300 MHz): δ 5.89 (d, 1H, J = 8.7 Hz), 5.39 (dd, 1H, J = 9, 10.2 Hz), 5.12 (t, 1H, J = 9.9 Hz), 4.32 (dd, 1H, J = 4.5, 12.6 Hz), 4.12 (dd, 1H, J = 2.1, 12.6

Hz), 4.04-4.08 (m, 1H), 3.64 (dd, 1H, J = 9.0, 10.5 Hz), 2.22 (s, 3H), 2.12 (s, 3H), 2.06 (s, 3H), 2.05 (s, 3H).

Synthesis of compound 14: Et_3N (0.15 mL, 1.04 4 mmol) was added to a stirred solution of **13** (0.2 g, 0.522 mmol). The solution was cooled to 0 °C, and acryloyl chloride (0.046 mL, 0.522 mmol) was added dropwise. The reaction mixture was stirred overnight at room temperature. The solution was concentrated in vacuo and purified on silica gel chromatography (hexanes 100% to hexanes/EtOAc = 1/1) to afford **14** as a light brown color solid (0.136 g, 65%). ^{1}HNMR (CD_3Cl_3, 300MHz): δ 6.31 (d, 1H, *J* = 1.2 Hz), 6.25-5.94 (m, 1H), 5.82-5.92 (brs, 1H), 5.76-5.8 (m, 2H), 5.26-5.10 (m, 2H), 4.43 (dd, *J* = 3.5, 10.5 Hz), 4.20 (ABq, 2H, *J* = 2.1, 10.2 Hz), 3.88-3.84 (m, 1H), 2.12 (s, 3H), 2.12 (s, 3H), 2.07 (s, 3H), 2.04 (s, 3H). ^{13}CNMR (CD_3Cl_3, 75MHz): δ 171.2, 170.6, 169.3, 165.6, 130.2, 127.3, 92.4, 72.7, 68.1, 61.8, 52.5, 20.8, 20.6, 20.6, 20.5.

CHAPTER V

GENETICALLY ENCODED PERFLUORINATED PHENYLALANINES FOR PROBING BIOLOGICAL CATION-Π INTERACTIONS

Introduction

Site-Specific introduction of fluorine labels into proteins is a powerful tool for investigating protein conformational changes and interactions.[70, 126, 127] The remarkable utility of using fluorine substitution as ideal probes mostly due to the unique properties of fluorine element, given its unprecedented combination of small size, low polarity and the strongest electronegativity among all elements.[128] In other words, the fluorine substitution allows the most inductive perturbation to the neighboring moiety but maintain the smallest spatial impact on the host. Therefore, it has been developed into an increasingly popular strategy in studying noncovalent protein interactions, including π-π stacking between aromatic amino acids as well as polar-π interactions between aromatic amino acids and cation/anion.[129, 130] These interactions are highly relevant to biological activities such as protein folding, protein-protein association and protein-ligand recognition. To this end, two methods are commonly used to introduce fluorinated ncAAs which involves semisynthesis or metabolic incorporation via endogenous translational machinery.[126-129] However, semisynthesis suffers from low yield and limitation of protein size. Metabolic incorporation is impeded by background incorporation from canonical amino acids and low yield due to toxicity of fluorinated ncAAs. Development of methodology for programmable fluorinated amino acids

incorporation into protein is eminently desirable. Herein, we report the genetically incorporation of pentafluorophenylalanine into proteins in *Escherichia coli* using a polyspecific pyrrolysyl-tRNA synthetase (PylRS) mutant. Further engineering of the synthetase to optimize the incorporation efficiency, a serious perfluorinated phenylalanines were introduced into proteins site-specifically. To study the protein cation-π interaction between chromodomain and trimethylated lysine on a histone peptide, the M-phase phorsphoprotein 8 was expressed with pentafluorophenylalanine incorporated at the ligand binding cage. The results reveal cation-π interaction is interrupted by the fluorine substitution, leading to abolish of the ligand binding. Incorporation of these fluorinated phenylalanine noncanonical amino acids expands the availability of fluorine probes for studying biological noncovalent biological events.

Results and Discussion

We have previously reported a polyspecific PylRS mutant (PylRS-N346A/C348A), coupled with $tRNA_{CUA}^{Pyl}$, is able to genetically incorporate a wide variety of para-, meta- and ortho- substituted phenylalanine derivatives at amber mutation sites of a protein that brings the size of the substrate inventory of this enzyme close to 60.[70, 74, 114, 115, 122, 131, 132] PylRS(N346A/C348A) displays high activities toward these noncanonical amino acids (ncAAs) while remaining relative orthogonality toward canonical amino acids. Accordingly, to incorporate fluorinated phenylalanines into proteins, particularly pentafluoro-L-phenylalanine (ncAA-**1**), we sought to employ PylRS(N346A/C348A) to mediate the genetic incorporation. To test this assumption, we used a previously generated ncAA incorporation and super green fluorescent protein (sfGFP) expression strain, BL21(DE3) transformed to contain the plasmids of synthetase and tRNA (pEVOL-PylT-PylRS(N346A/C348A)) as well as a plasmid of sfGFP with S2 amber mutation reporter (pET-sfGFPS2TAG), for ncAA-**1** incorporation.[115] Following medium shift protocol, we grown the cell in LB until OD_{600} equals to one. After removed the medium, washed with isotonic saline, the pellet was resuspended in GMML medium supplemented with 0.2% arabinose, 1 mM IPTG, 5 mM ncAA-1 and incubated at 37^{o}C for 20h to induce sfGFP mutant expression. The ncAA-**1** incorporation efficiency is evaluated based on fluorescence intensity of the medium-free cell pellet. In compared to the control expression without providing any ncAA in the medium, indeed, the ncAA1-**1** is able to serve as a substrate of PylRS(N346A/C348A), though the efficiency is not great (**Figure 5-1**). While the hydrogen and fluorine atoms have similar

size, the pentafluorophenylalanine hoards the small inequality from each fluorine substitution, resulting to overall discrimination against phenylalanine for the synthetase. However, in considering of other perfluorophenylalanine ncAAs with fewer fluorine substitutions, PylRS(N346A/C348A) may be inadequate to offer efficient incorporation. Therefore, optimization of the synthase is desirable for extending the perfluorophenylalanine inventory. The crystal structure of a PylRS(N348A/C348A) analogy mutant OMeRS in complex with O-methyltyrosine reveals the synthetase active site configuration coordinating with phenylalanine analogue.[133] **Figure 5-2** presents the active site structure of this enzyme-substrate complex superimposed on the wild-type PylRS with pyrrolysyl-AMP bound. The N346A is a key mutation that allows the synthetase to select phenylalanine analogues.[74] Whereas other residues around the binding pocket don't stringently confine the substrate orientation. On the other words, by engineering the residues around the binding pocket against ncAA-**1**, the synthetase will ideally better accommodate the fluorinated ncAAs, in turn promotes the incorporation efficiency.

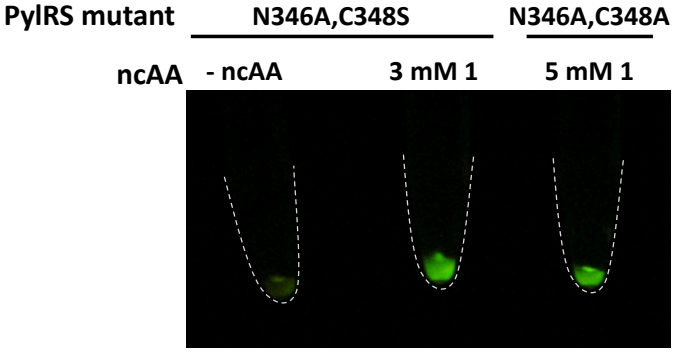


Figure 5-1. Small scale ncAA-**1** incorporation and sfGFP mutant expression test.

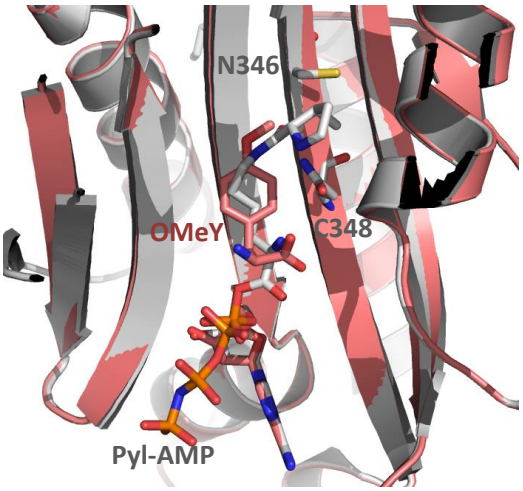


Figure 5-2. Superimposed structures of the OmeRS complex with O-methyltyrosine and the PylRS complex with pyrrolysine-adenosylmonophosphate (Pyl-AMP).

In order to evolve aaRS mutants that present better substrate specificity, directed evolution has proven to be an effective approach, where standard protocols have been established for each newly targeted ncAA of interest.[85] However, the whole evolution protocol is lengthy, normally takes several weeks to finish repeated selections. In the contrary, the rationally designed PylRS(N346A/C348A) mutate has broad substrate scope.[70, 74, 114, 115, 131] The N346 residue of PylRS behaves as key discriminator for selecting lysine derivatives as substrates, whereas the N346 mutation abolishes the selectivity and allows a variety of phenylalanine/tyrosine derivatives to be incorporated. The polyspecificity of the PylRS(N346A/C348A) mutant is convenient to introduce new ncAAs into proteins, bypassing the tedious evolution steps. Nonetheless, the mutant itself encounters bottleneck to further improve its enzymatic efficiency toward those poor ncAA substrates due to the incorporation of canonical amino acid phenylalanine.

This background incorporation outcompetes the PylRS(N346A/C348A) uptake of the ncAA substrate, resulting to inefficient or no convergence after each round of directed evolution. Although it is not practical to evolve a mutant based on PylRS-N346A/C348A, engineering of the mutant catalytic pocket provide a means to fine-tune the specificity to an ncAA substrate. Thereby, to further optimize the incorporation efficiency toward ncAA-**1**, a screening strategy was applied. Briefly, we constructed an active-site mutant library of the mmPylRS gene with randomization at single active site residue, where the N346A mutant was intentionally retained to ensure accommodation of phenylalanine analogues. This single site library in pBK plasmid together with another selection plasmid pY+ was used to transform the Top10 *E. coli* strain. pY+ plasmid contains genes encoding pylT, type I chloramphenicol acetyltransferase with an amber mutation at D112 and green fluorescent protein (GFPuv) with an amber mutation at Q204. The library-containing cell was plated on GMML plate with 1% LB then single colonies were randomly selected and inoculated to LB medium on 96-well plates. After incubating the plates at 37^{o}C for 3h, the cell in each well was dotted onto the screening plates containing chloramphenicol (Cm) and with or without ncAA-**1** (**Figure 5-3**). Both cell viability and fluorescence intensity were evaluated during the screening process. The candidates displayed low background incorporation as well as strong fluorescence were collected, duplicated, plasmid-extracted, and subjected to DNA sequencing. Among various constructed single-site libraries, the randomization at C348 site appeared positive response toward ncAA-**1** supplemented in the medium, where the library converges to a C348S mutant after screening procedure (**Figure 5-3**).

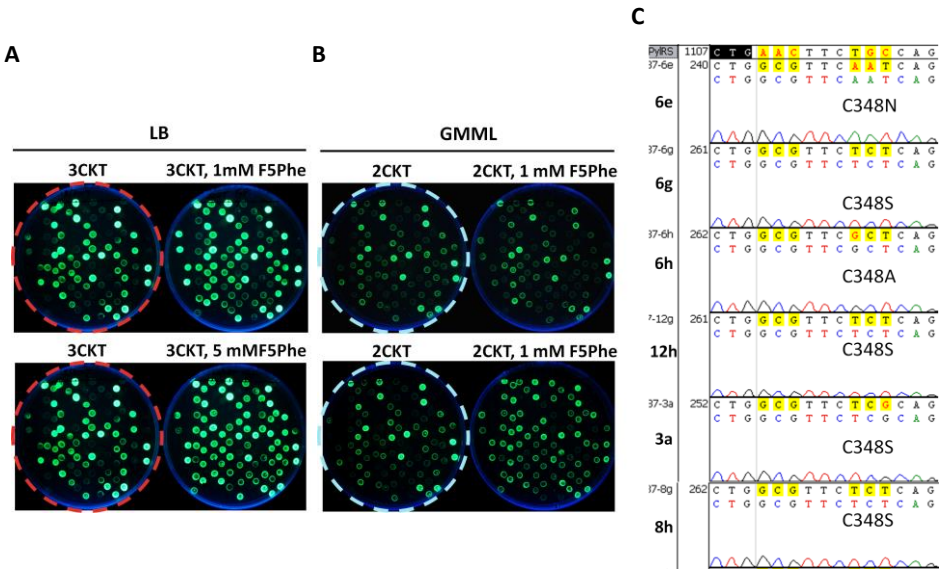


Figure 5-3. Single site randomization library screening experiment for the PylRS mutant optimization.

The afforded PylRS mutant with N346A/C348A double mutations was subsequently cloned into an optimized pEVOL vector under contraol of an araBAD promoter to afford plasmid pEVOL-pylT-PylRS(N346A/C348S) that also contain a $tRNA^{pyl}_{CUA}$ gene under *proK* promoter. Transforming BL21(DE3) cells that contained a reporter plasmid pET-sfGFPS2TAG with pEVOL-PylRS(N346A/C348S) afforded cell strain for further protein expression. PET-sfGFPS2TAG has a His-tagged sfGFP gene that contains an amber mutation at its S2 position and is under control of a IPTG-inducible T7 promoter.[26] Growing cells in LB medium, after medium shift to GMML supplemented with 0.2% arabinose, 1 mM IPTG, and 3 mM of ncAA-**1**resulted in the overexpression of sfGFP, indicated by the strong fluorescence after 20h protein expression. When ncAA-**1** is absent, only a small amount of sfGFP could be observed (**Figure 5-1**). Evidenced by stronger fluorescent signal, the protein expression of ncAA-

1 dependent sfGFP expression using pEVOL-pylT-PylRS(N346A/C348S) is slightly better than that of using previously reported pEVOL-PylRS(N346A/C348A) plasmid under the same expression condition (**Figure 5-1**), indicating PylRS(N346A/C348S) is more specific to ncAA-**1** . This observation is also in line with our assumption that the screening approach is capable of assisting synthetase-ncAA pair optimization. For further characterization of the ncAA-**1** genetic incorporation into sfGFP, the cell was lysed and the lysate was subjected to Ni-NTA affinity purification. After buffer exchange, purified protein was further analyzed by electrospray ionization mass spectrometry (ESI-MS) that showed the expected molecular weights of sfGFP with ncAA-**1** incorporated (calculated 27,819 DA; found 27,817 Da) (**Figure 5-4C** and **Table 5-1**).

Table 5-1. ESI-MS characterization of sfGFP variants incorporated with different fluorinated phenylalanine ncAAs (ncAA **1**-**6**).

ncAA	Calculated Mass (Da)	Observed Mass (Da)
1	**27819**	**27817**
2	**27801**	**27799**
3	**27783**	**27783**
4	**27765**	**27763**
5	**27765**	**27763, (27782, 27800, 27818, 27836, 27845, 27871, 27891, 27911)**
6	**27765**	**27765, (27800, 27835, 27871, 27907, 27932, 27965, 28004)**

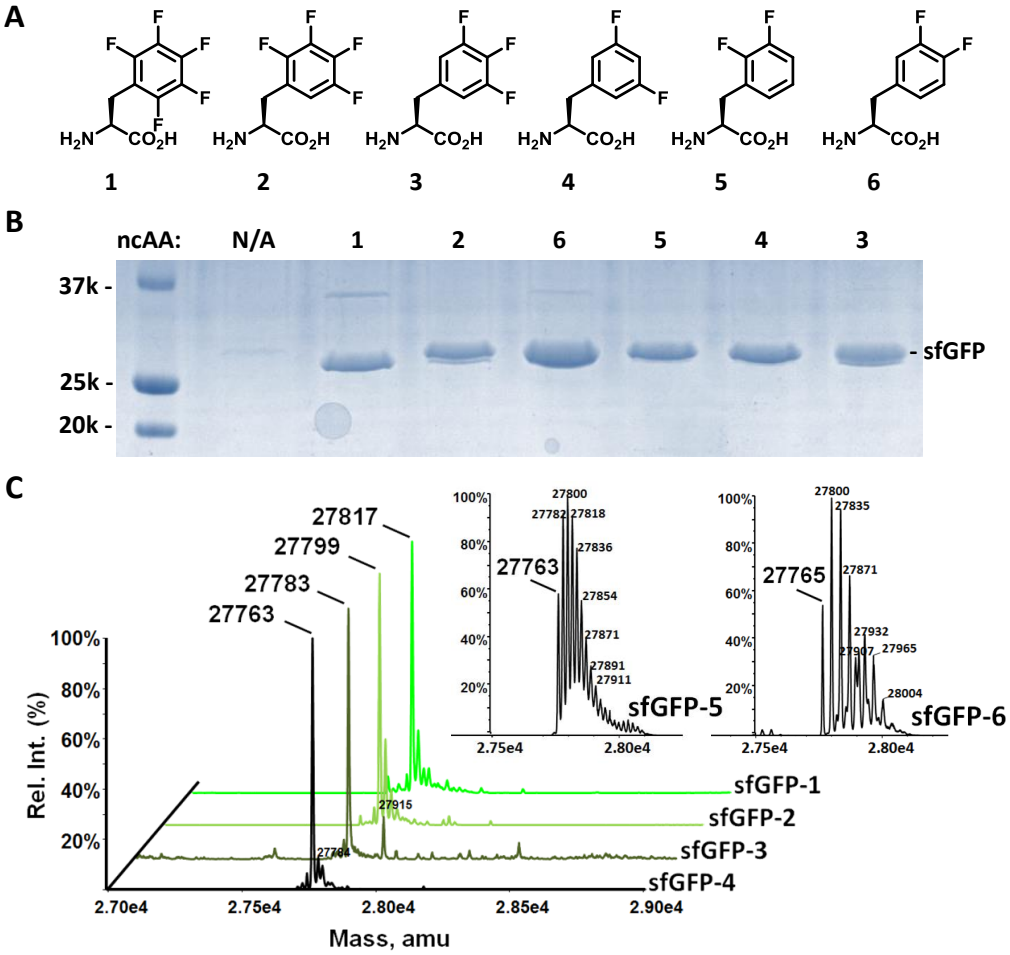


Figure 5-4. (A) Structures of perfluorophenylalanine ncAAs **1**-**6**. (B) Site-specific incorporation of ncAA **1**-**6** into sfGFP at the S2 site. The N/A indicates no ncAA was supplemented in the medium. (C) Deconvoluted ESI-MS of sfGFPS2 mutants incorporated with ncAAs **1**-**6**.

With the success on pentafluorophenylalanine incorporation, we went further to explore the possibility if other fluorine-substituted phenylalanine analogues are able to mediate the ncAA incorporation into proteins. Synthesized or commercially available ncAAs including 2, 3, 4, 5-tetrafluoro-L-phenylalanine (ncAA-**2**, gift from Prof. Jianmin Gao, Boston College), 3,4,5-trifluoro-L-phenylalanine (ncAA-**3**), 2, 3-difluoro-L-phenylalanine (ncAA-**5**), 3, 4-difluoro-DL-phenylalanine (ncAA-**6**), and 3, 5-difluoro-L-phenylalanine (ncAA-**4**) were provided in the expression medium. Following similar expression protocol with the same cell strain, after protein expression, purification and buffer exchange, the afforded proteins were analyzed by ESI-MS (**Figure 5-4** and **Table 5-1**). The MS of the proteins supplemented with ncAA-**2** or ncAA-**3** agree well with theoretical molecular weights, indicating the incorporation of tetra- and tri-fluorine substituted phenylalanines. However, when the medium provided with ncAA-**5** or ncAA-**6**, two difluorophenylalanine isomers, both ESI-MS of the purified proteins show numerous peaks, where each peak differs from roughly 34 Da, indicating multiple difluorophenylalanines are incorporated into sfGFP (**Figure 5-4** and **Table 5-1**). While there are in total of twelves phenylalanines in one sfGFP, the multiple difluorophenylalanines incorporation may result from disincorporation of ncAA-**5** or ncAA-**6** to the original phenylalanine sites. Since *E. coli* endogenous phenylalanyl-tRNA synthetase (PheRS) has known to mischarge the *p*-fluorophenylalanine, little perturbation of difluorine substitutions with high concentration ncAAs provided in the medium may contribute to the mischarge of ncAA-**5** and -**6** on the $tRNA^{Phe}$. Interestingly, when anther isomer 3,5-difluorophenylalanine (ncAA-**4**) was provided in

the medium, after protein expression, purification and ESI-MS analysis, the MS spectrum showed only single protein product, where the observed molecular weight agrees well with that of single difluorophenylalanine incorporated sfGFP (**Figure 5-4** and **Table 5-1**). We assume that the 3,5-difluorine substitution is bulkier than the other isomers, effectively precluding ncAA-**4** being mischarged to the $tRNA^{Phe}$ by endogenous PheRS, leading to site-specific incorporation of ncAA-**4** into sfGFP. This result further expands the perfluorophenylalanine arsenal that can be used for biophysical studies.

The genetic incorporation of ncAA-**1-4** offers an array of perfluorinated phenylalanine probes and enables stepwise change the electron distribution of the aromatic system on phenylalanine. With this tool set in hand, we proceeded to apply it in the investigation of cation-π interactions between chromodomain and trimethyllysine. A chromodomain is a protein structural domain of about 40-50 amino acid residues commonly found in proteins associated with the remodeling and manipulation of chromatin. Chromodomains are highly conserved methylated histone lysine recognition modules that share a common ancestral fold with two to four well-conserved aromatic residues coordinating the methylated lysine. The aromatic residues form a "cage" and binds to the methylated lysine generally believed through the noncovalent cation-π interactions. The M-phase phosphoprotein 8 (MPP8) is a methyl-H3K9-binding protein. MPP8 was found mediating the E-cadherin gene silencing and synergistically promotes tumor cell motility and invasion.[134] E-cadherin gene is a key regulator of tumor cell growth and transition. When MPP8 bind to methylated H3K9, the resulting complexes recruit methyltransferases and methylate E-cadherin gene, in term to repress the gene

function.[134] The MPP8 chromodomain comprises of F59, W80, Y83 aromatic cage and binds tightly with trimethylated H3K9 (H3K9me3) at 0.3 uM of dissociation constant (**Figure 5-5A**).[135] We sought to perturb this binding behavior by substituting the F59 residue to ncAA-**1**. The perfluorination inverts the electronic property of the phenylalanine π system and we predict that the affinity between trimethyllysine and MPP8 will be abolished.

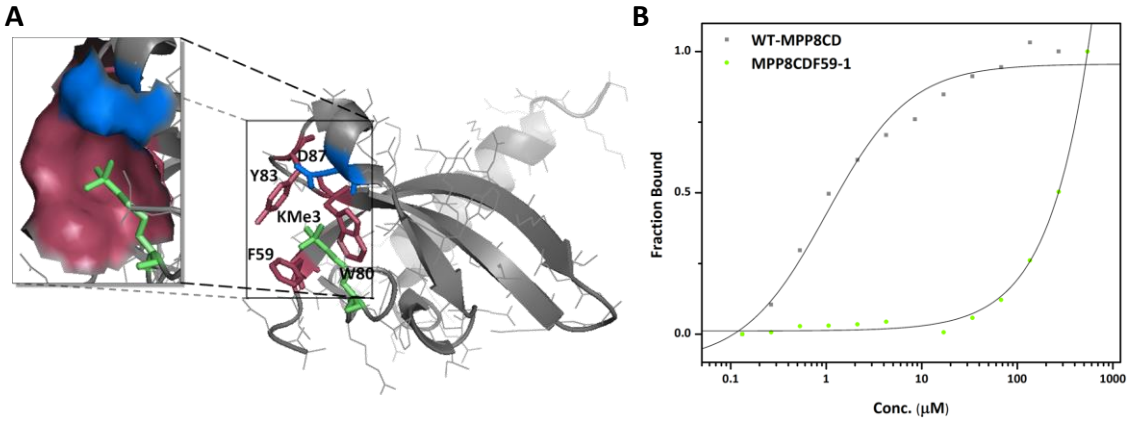


Figure 5-5. (A) The structure of H3K9 trimethyllysine peptide-bound M-phase phorphoprotein 8 chromo domain (MPP8CD). (B) Binding of H3-K9Me3 to wild-type MPP8CD or MPP8CD mutant incorporated with ncAA-**1** as determined by fluorescence polarization assay.

To this end, we constructed an expression plasmid which contains the gene under control of araBAD promoter to express the fusion protein of human MPP8 chromodomain (MPP8CD, residues 55-116) with C-terminal small ubiquitin-like modifier-6xHis (SUMO-6xHis) in which the F59 of MPP8CD gene was mutated to TAG amber codon (pBAD-MPP8CDF59TAG-SUMO). Top10 *E. coli* strain was transformed to contain aforementioned pBAD-MPP8CDF59TAG-SUMO and pEVOL-PylT-PylRS-N346A-C348S for protein expression. The ncAA-**1** incorporation and protein expression follows medium shift protocol. Briefly, the cell was grown in LB until the DO_{600} reach to one. After removing the LB, the cell pellet was washed with isotonic saline for two times and resuspended in GMML medium supplemented with 0.2% arabinose, 2.5 mM ncAA-**1** to induce MPP8SUMO mutant expression. The protein expression was conducted at 37^{o}C for 20h. After collecting the cells, the cell was lysed and subjected to purification procedures including Ni-NTA affinity and Q-sepharose iox-exchange purification to afford the fusion protein MPP8CDF59-**1**-SUMO. Wild-type MPP8CD-SUMO was expressed as a control protein using similar pBAD plasmid without the in-frame amber codon and expressed in LB medium with the same purification protocol. H3K9 trimethyllysine peptide (FAM-ARTKQTAR(KMe3)STGGKAY-NH_2, H3K9Me3) was synthesized with N-terminal end fluorescein modification and C-terminal end amidation. Fluorescence polarization assays were carried out to determine the binding between H3KMe3 and MPP8CD in phosphate buffer (pH 7.5) and 50 mM NaCl. At 25^{o}C, specific binding was observed between wild-type MPP8CD-SUMO and H3K9Me3 peptide at a dissociation constant of 1.01 uM (**Figure 5-5B**) that is consistent

with literature reported binding constant at the degree of order.[135, 136] However, when the ncAA-1 incorporated MPP8CD-SUMO was applied, the binding with H3K9Me3 peptide is significantly dropped, given more than 200 uM dissociation constant (**Figure 5-5B**). Given the similarity in size and shape as well as the greater hydrophobicity of the petntafluorophenylalanine in comparing with phenylalanine, the factors that substantially weaken the binding may be attributed by the inductive effect from the aromatic π system. This observation clearly demonstrates the essential nature of the cation-π component to binding of the lysing-methylated H3 tail to chromodomains.

Conclusion

In summary, we reported the genetic incorporation of an array of perfluorinated phenylalanine ncAAs including penta-, tetra-, tri-, and di-fluorophenylalanine into proteins using an optimized PylRS(N346A/C348S)- $tRNA^{Pyl}_{CUA}$ pair. The fluorine substitution has the minimum structural influence to a molecule but contributes significant electronic perturbation due to fluoride's unrivaled electronegativity. Phenylalanine is involved in a variety of biological noncovalent associations and plays critical role in many protein ligand recognitions in living system. The incorporation of a series of fluorinated phenylalanines has greatly expanded the probe inventory for studying these biological π-interactions. Using this incorporation protocol, we generated pentafluorophenylalanine-incorporated MPP8 chromodomain protein. The perturbation resulted from the fluorine substitutions significantly lower the association between the trimethylated H3K9 peptide, essentially rescinding the cation-π interaction. This observation can provide the insight of quantifying contribution from cation-π interaction in the protein-ligand association event. This method also offers the opportunity to further investigate other biological π-systems of specific biomolecelar recognition.

Experimental Section

Polysepcific PylRS mutant optimization: single-site library construction and library screening

Library Construction

The plasmid pBK-PylRS(N346A) was derived from pBK-MmPylRS plasmid reported previously.[74, 85]

The single site randomization was achieved through Phusion DNA polymerase (NEB, Ipswich, MA, USA) based site-directed mutagenesis protocol with pBK-PylRS(N346A) as template and two synthesized degenerated primers (IDT, Coralville, IA, USA): YJL055 (5'- NNKCAGATGGGATCGGGATGTACAC -3') and YJL056 (5'-GAAGTTCAGCATGGTAAACTCTTCGAGGTGTTC -3') where N=A/C/T/G; T, K=G/T; M=C/A. The afforded PCR product was phosphorylated with T4 PNK, ligated with T4 DNA ligase and heat-inactivation of the enzyme before transformation.

Library Screening

This single site library in pBK plasmid together with another selection plasmid pY+ was used to transform the Top10 *E. coli* strain. pY+ plasmid contains genes encoding pylT, type I chloramphenicol acetyltransferase with an amber mutation at D112 and green fluorescent protein (GFPuv) with an amber mutation at Q204.[85] The library-containing cell was plated on GMML plate with kanamycin (Kan, 50 μg/mL) and tetracycline (Tet, 25 μg/mL) plus 1% LB then single colonies were randomly selected and inoculated to LB medium with kanamycin (Kan, 50 μg/mL) and tetracycline (Tet, 25

μg/mL) on 96-well plates. After incubating the plates at 37^{o}C for 3h, the cell in each well was dotted onto the screening plates containing chloramphenicol (Cm) (68 μg/mL), Kan (50 μg/mL), Tet (25 μg/mL) and with ncAA-**1**. The plates without ncAA-**1** serve as the control. Plates were incubated at 37^{o}C. Both cell viability and fluorescence intensity were evaluated during the screening process. The clones with strong fluorescence on the NAAcontaining plate and low viability and fluorescence on the control plate were selected as the active clones and were amplified from the corresponding wells in the 96-well plate. The pBK plasmid of the active clones was isolated and sequenced. Most of the active clones were converged to contain N346A/C348A mutations.

sfGFP-ncAA mutants protein expression and purification

pEVOL-pylT-PylRS(N346A/C348S) Construction

The pEVOL-PylT-PylRS(N346A/C348S) plasmid was derived from the pEVOL-PylT plasmid reported previously.[74] The PylRS(N346A/C348S) gene were PCR amplified from the identified pBK-PylRS(N346A/C348S) plasmids using primers pEVOL-PylRS-SpeI-F (5'-gaggaaactagtatggataaaaaaccactaaacactctg-3') and pEVOL-PylRS-SalI-R (5'-tgatggtcgactcacaggttggtagaaatcccgtt-3'). The PCR product was digested by SpeI and SalI restriction enzymes, and ligated to a similarly digested pEVOLPylT plasmid.

Protein Sequence of sfGFP:

MA**S**KGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFI**C**TTGKL

PVPWPTLVTTLTYGVQ**C**FSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY

KTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGI

KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKR

DHMVLLEFVTAAGITHGMDELYKGSHHHHHH

*The NCAA incorporation site is highlighted in **red**.

Incorporation of Perfluorinated ncAAs into sfGFP

A previously reported protocol was followed to genetically incorporate ncAA-**1** – ncAA-**6** into sfGFP.[70, 74, 115] Briefly, *E. coli* BL21(DE3) cells were cotransformed with pET-PylT-sfGFPS2TAG and pEVOL-PylT-PylRS(N346A/C348S). The transformed cell was used to inoculate the LB media and was grown at 37°C with agitation until the OD_{600} reached 1.0-1.2. Then the cells were pelleted and washed with isotonic saline or PBS buffer (x3). After medium shift to a minimal medium (GMML) supplemented with 1 mM IPTG, 0.2% arabinose, and one of desired ncAA (3 mM), the cells were incubated at 37^{o}C with agitation for 20h. ncAAs were purchased from commercial sources or synthesized in the lab which were used without further purification. After centrifugation, the cell pellet was resuspended in ice cold lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by sonication. The lysate was centrifuged (10000 g, 4^{o}C, 1h) and the supernatant subjected to Ni-NTA affinity chromatography according to

manufacturer's protocol. The supernatant was incubated with Ni-NTA resin (Qiagen) (2h, 4^{o}C). The slurry was then loaded to a column and the protein-bound resin was washed with washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 45 mM imidazole, pH 8.0) followed by eluting the bound protein with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein was buffer-exchanged to Tris-HCl buffer (50 mM, pH 7.4) and concentrated using an Amicon Ultra-15 Centrifugal Filter (10 kD MWCO, Millipore). The quality of the purified protein was analyzed by SDS-PAGE and ESI-MS. The sfGFP concentration was quantified by UV absorbance at 485 nm using literature reported extinction coefficient $8.33x10^4$ M^{-1} cm^{-1}.

Expression and purification of MPP8CD-SUMO-6xHis proteins

DNA Sequence of MPP8CD

5'-

atgggcgaagacgtgttcgaagtagagaagatcctggacatgaagaccgaaggcggtaaagttctgtacaaagtgcgctgga

aaggctatacttctgacgatgacacctgggaaccagagatccatctggaagattgtaaagaagttctgctggaatttcgcaagaa

gattgccgagaacaaagctaaa- 3'

MPP8CD-SUMO Expression Vector Construction

The codon optimized MPP8 chromo domain (MPP8CD) gene was synthesized by Epoch (Epoch Life Science, Sugar Land, TX, USA). The MPP8CD gene was PCR amplified from the synthetic construct using primers YJL057 (5'-CCACCATG

GGCGAAGACGTGTTCGAAGTAG-3') and YJL058 (5'-AACCATGGC ACCTTGGAAATACAGGTTCTCTTTAGC-3'). A pET28-SUMO-6xHis plasmid (a gift from Prof. C. Hilty, Department of Chemistry, Texas A&M) was digested with NcoI restrict enzyme. The digested site was ligated with the aforementioned NcoI-digested PCR product using T4-DNA ligase to aford pET28b-MPP8CD-SUMO-6xHis.

The MPP8CD-SUMO-6xHis gene was PCR amplified from the template plasmid pET28b-MPP8-SUMO-6xHis using primers YJL057 (5'- CCACCATG GGCGAAGACGTGTTCGAAGTAG-3') and YJL094 (5'- GATCCTGCAGTTAGTGGTGGTGGTGGTGGTG ACCACCAATCTGTTCTCTGTGAGCC-3'). Amplified PCR product was inserted into a pBAD plasmid downstream of an *araBAD* promoter using NcoI and PstI sites to generate pBAD-MPP8CD-SUMO-6xHis.

Protein Sequence of MPP8CD-SUMO-6xHis:

MGEDV**F**EVEKILDMKTEGGKVLYKVR**W**KG**Y**TSDDDTWEPEIHLEDCKEVLLE FRKKIAENKAKELSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTP LRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG HHHHHH

*The underline highlights the MPP8 chromo domain sequence (MPP8CD) and ncAA incorporation site is marked in **red**.

Incorporation of Perfluorinated ncAAs into sfGFP

The purification protocol was modified from the sfGFP-ncAA expression and purification. *E. coli* strain Top10 were transformed with plasmids pBAD-MPP8CDF59TAG-SUMO-6xHis and pEVOL-pylT-PylRS(N346A/C348S) or pEVOL-pylT-PylRS(N346A/C348A). The transformed cell was used to inoculate the LB media and was grown at 37°C with agitation until the OD_{600} reached 1.0-1.2. Then the cells were pelleted and washed with isotonic saline or PBS buffer (x3). After medium shift to a minimal medium supplemented with 1 mM IPTG, 0.2% arabinose, and 3 mM ncAA-**1**, the cells were incubated at 37^{o}C with agitation for 20h.

After centrifugation, the cell pellet was resuspended in ice cold lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by sonication. The lysate was centrifuged (10000 xg, 4^{o}C, 1h) and the supernatant subjected to Ni-NTA affinity chromatography according to manufacturer's protocol. The supernatant was incubated with Ni-NTA resin (Qiagen) (2h, 4^{o}C). The slurry was then loaded to a column and the protein-bound resin was washed with washing buffer (50 mM NaH_2PO_4, 300 mM NaCl, 45 mM imidazole, pH 8.0) followed by eluting the bound protein with elution buffer (50 mM NaH_2PO_4, 300 mM NaCl, 250 mM imidazole, pH 8.0). The eluted protein was buffer-exchanged to Tris-HCl buffer (50 mM, pH 7.4) and concentrated using an Amicon Ultra-15 Centrifugal Filter (10 kD MWCO, Millipore). The quality of purified protein was analyzed by SDS-PAGE.

To further improve the quality of protein, the buffer exchanged MPP8CD was subjected to FPLC mounted with a QSepharose HP column, where the M88P was eluted at roughly 37% NaCl concentration of buffer B (20 mM TrisHCl pH 8, 1 M NaCl; buffer A contains the same composition without NaCl). The collected fractions were dialyzed to a FP assay buffer containing 50 mM potassium phosphate pH 8.0, 25 mM NaCl and concentrated. Protein concentration was determined by absorbance spectroscopy using predicted extinction coefficient (ε_{280} = 15470 M^{-1} cm^{-1}, MW = 19640.12).

Fluorescence polarization binding assays

Fluorescence polarization (FP) was performed on a BioTek Synergy H1 plate reader by setting it on automatic gain and 100 flashes. FP binding assays were performed under conditions of 50 mM potassium phosphate pH 8.0, 25 mM NaCl and in the presence of 100 nM fluorescein-labeled peptide. A series dilution of protein was prepared in the nanomolar-millimolar concentration range on a black solid bottom 96-well plate. Fluorescence polarization (P) values were converted to anisotropy (A) values by the equation A = 2P/(3 - P). Binding curves were analyzed by non-linear least-squares fitting of the raw fluorescence polarization data. Data were fitted using the equation A = Af + (Ab – Af)([protein]/(K_D + [protein])), where Af and Ab represent the anisotropy of the free and bound peptides, respectively.[137]

General experimental and chemical synthesis

All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was performed on EMD Millipore silica gel 60 F_{254} plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.80 ppm for ^{1}H). A minimal amount of 1,4-dioxane was added as the reference standard (67.19 ppm for ^{13}C) for carbon NMR spectra in deuterium oxide, and a minimal amount of sodium hydroxide pellet or concentrated hydrochloric acid was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral conditions. ^{1}H NMR spectra are tabulated as follows: chemical shift, multiplicity (s = singlet, bs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). Mass spectra were obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

Pentafluoro-L-phenylalanine (ncAA-**1**), 3,4,5-trifluoro-L-phenylalanine (ncAA-**3**), 2,3-difluoro-L-phenylalanine (ncAA-**5**) and 3,4-difluoro-L-phenylalanine (ncAA-**6**)

were purchased from ChemImpex Inc. (Wood Dale, IL, USA). All obtained ncAAs were subjected to the protein expression without further treatment.

2,3,4,5-Tetrafluoro-L-phenylalanine (ncAA-**2**) is a gift from Prof. Jianmin Gao (Department of Chemistry, Boston College). The synthesis procedure was published previously. [138] The ncAA-2 was used as obtained without further treatment.

3,5-Difluoro-L-phenylalanine (ncAA-**4**) was purchased from CombiBlocks Inc. (San Diego, CA, USA). It was subjected to the protein expression without further treatment.

CHAPTER VI

CONCLUDING REMARKS AND FUTURE OUTLOOK

In this study, we reported the genetic incorporation of a variety of ncAAs into proteins using genetic code expansion approach with wild-type or evolved PylRS-$tRNA^{Pyl}_{CUA}$ pair. In order to incorporate these novel ncAAs, deviated from previous directed evolution procedure, we demonstrate a small library screen based strategy which is deviated from previous reported directed evolution approach. It allowed quick optimization and fine-tuning the PylRS-ncAA pair, effectively promote the incorporation efficiency. These introduced ncAAs contains novel protein modifications, including native protein posttranslational modification (lysine crotonylation, propionylation and butyrylation), bioorthogonal chemical reports (acryloyllysine, long chain unstrained alkene ncAAs) and biophysical probes (perfluorinated phenylalanines). While the overall emphasis of research is focused on methodology development, the utilization of developed protocol for biologically significant applications, from both fundamental research to translational science and clinical application, have not yet comprehensively explored. Genetic code expansion enables the synthesis of proteins with defined and homogeneous protein with up to now more than 100 functionalities, substantially promoting the ability of controlling biomolecular intracellular activity, though not thoroughly. Backtracking to 10 years ago when alanine-scanning site-directed mutagenesis is the unparalleled protocol for biochemistry studies, nonetheless, there are increasing numbers of research adopting ncAA chemical biology approach.

Because of the uniqueness and selectivity, the ncAA method become a targeted research methodology can be envisioned.

For the development of genetically encoded acrylamide functionality, the incorporation of AcrK has provided diverse application opportunities. Acrylamide has unprecedented combination of small size and reactivity. Our studies show that the acrylamide moiety is stable in cellular condition for recombinant protein expression. The Acrk incorporated protein is able to undergo many-faceted bioconjugation reactions. For example, radical polymerization allows AcrK-containing protein immobilized into a hydrogel. Thiol 1,4-addition on protein acrylamide can provide site specific PEGylation to increase therapeutic protein stability. Moreover, AcrK is considered as lysine acetylation and propionylation PTM surrogate. In combining with the demonstrated bioorthogonal reactions, such as 1,3-dipolar cycloaddition with nitrile imine, the AcrK incorporation into histons can be developed as a chemical biology probe for epigenetics investigation with a high throughput manner.

Besides short-chain acylations, proteins in cells also undergo long-chain acylations. The development of using unstrained alkene as chemical reporter, in coupled with tetrazine DA_{inv} chemistry, provides an easy accessible methodology to address this question. Similar to the AcrK incorporation, this chemical antibody sort of PTM detection can further modified to high throughput fluorescence protocol.

Following the research line of acrylamide incorporation, the developed novel phosphine nucleophilic 1,4-addition is a great improvement of its thiol predecessor, essentially allowing to the practical *in vivo* protein labeling. Phosphine reacts with

acrylamide very selective and in a decent reaction rate which are two key features of ideal click reaction pairs. This development further extends the available tool in the click chemistry inventory. The application of using phosphine and acrylamide pair also can be found on other biomolecule or metabolite research.

Lastly, the site-specific incorporation of fluorine-containing noncanonical amino acid probes into proteins was also developed. Fluorine is referred as “magic bullet” in medicinal organic chemistry and biochemistry and is uniquely in the protein with barely structural perturbation. Phenylalanine is involved in a variety of biological noncovalent associations and plays critical role in many protein ligand recognitions in living system. The incorporation of a series of fluorinated phenylalanines has greatly expanded the probe inventory for studying these biological π-interactions. The site-incorporation of perfluorinated probes is an invaluable tool for the study of noncovalent protein interactions as well as the application on protein designs.

Overall, the collective work presented in this dissertation has improved our knowledge and competency of manipulating site-specific protein modifications with native biological PTMs and handy chemical/ biophysical probes. The developments substantially expand available chemical biology tools for comprehensive applications. The established protocols will help in the future design of probes and strategies on studying significant biological questions.

REFERENCES

[1] Prescher, J. A., and Bertozzi, C. R. (2005) Chemistry in living systems, *Nat. Chem. Biol. 1*, 13-21.

[2] Albery, W. J., and Knowles, J. R. (1976) Free-energy profile of the reaction catalyzed by triosephosphate isomerase, *Biochemistry 15*, 5627-5631.

[3] Nelson, N., and Ben-Shem, A. (2004) The complex architecture of oxygenic photosynthesis, *Nat. Rev. Mol. Cell Biol. 5*, 971-982.

[4] White, M. A., and Anderson, R. G. (2005) Signaling networks in living cells, *Annu. Rev. Pharmacol. Toxicol. 45*, 587-603.

[5] Buehler, M. J., and Yung, Y. C. (2010) How protein materials balance strength, robustness, and adaptability, *HFSP J. 4*, 26-40.

[6] Pauly, H. (1904) On the constitution of histidine i announcement, *H-S. Z. Physiol. Chem. 42*, 508-518.

[7] Stephanopoulos, N., and Francis, M. B. (2011) Choosing an effective protein bioconjugation strategy, *Nat. Chem. Biol. 7*, 876-884.

[8] Moore, D. S. (1985) Amino-acid and peptide net charges - a simple calculational procedure, *Biochem. Educ. 13*, 10-11.

[9] Boutureira, O., and Bernardes, G. J. (2015) Advances in chemical protein modification, *Chem. Rev. 115*, 2174-2195.

[10] Sletten, E. M., and Bertozzi, C. R. (2009) Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality, *Angew. Chem. Int. Ed. Engl. 48*, 6974-6998.

[11] Nienhaus, G. U. (2008) The green fluorescent protein: A key tool to study chemical processes in living cells, *Angew. Chem. Int. Ed. Engl. 47*, 8992-8994.

[12] McKay, C. S., and Finn, M. G. (2014) Click chemistry in complex mixtures: Bioorthogonal bioconjugation, *Chem. Biol. 21*, 1075-1101.

[13] Sharpless, K. B., and Kolb, H. C. (1999) Click chemistry. A concept for merging process and discovery chemistry., *Abstracts of Papers of the American Chemical Society 217*, U95-U95.

[14] Kolb, H. C., Finn, M. G., and Sharpless, K. B. (2001) Click chemistry: Diverse chemical function from a few good reactions, *Angew. Chem. Int. Ed. Engl. 40*, 2004-2021.

[15] Hong, V., Presolski, S. I., Ma, C., and Finn, M. G. (2009) Analysis and optimization of copper-catalyzed azide-alkyne cycloaddition for bioconjugation, *Angew. Chem. Int. Ed. Engl. 48*, 9879-9883.

[16] Saxon, E., and Bertozzi, C. R. (2000) Cell surface engineering by a modified staudinger reaction, *Science 287*, 2007-2010.

[17] Sletten, E. M., and Bertozzi, C. R. (2011) From mechanism to mouse: A tale of two bioorthogonal reactions, *Acc. Chem. Res. 44*, 666-676.

[18] Saxon, E., Armstrong, J. I., and Bertozzi, C. R. (2000) A "traceless" staudinger ligation for the chemoselective synthesis of amide bonds, *Org. Lett. 2*, 2141-2143.

[19] Nilsson, B. L., Kiessling, L. L., and Raines, R. T. (2001) High-yielding staudinger ligation of a phosphinothioester and azide to form a peptide, *Org. Lett. 3*, 9-12.

[20] Lang, K., and Chin, J. W. (2014) Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins, *Chem. Rev.114,* 4764-4806.

[21] Blackman, M. L., Royzen, M., and Fox, J. M. (2008) Tetrazine ligation: Fast bioconjugation based on inverse-electron-demand diels-alder reactivity, *J. Am. Chem. Soc. 130*, 13518-13519.

[22] Taylor, M. T., Blackman, M. L., Dmitrenko, O., and Fox, J. M. (2011) Design and synthesis of highly reactive dienophiles for the tetrazine-trans-cyclooctene ligation, *J. Am. Chem. Soc. 133*, 9646-9649.

[23] Huisgen, R., Grashey, R., Seidel, M., Schmidt, R., Knupfer, H., and Wallbillich, G. (1962) 1.3-dipolare additionen .2. Synthese von 1.2.4-triazolen aus nitriliminen und nitrilen, *Justus Liebigs Ann. Chem. 653*, 105-113.

[24] Song, W., Wang, Y., Qu, J., and Lin, Q. (2008) Selective functionalization of a genetically encoded alkene-containing protein via "photoclick chemistry" in bacterial cells, *J. Am. Chem. Soc. 130*, 9654-9655.

[25] Song, W., Wang, Y., Qu, J., Madden, M. M., and Lin, Q. (2008) A photoinducible 1,3-dipolar cycloaddition reaction for rapid, selective modification of tetrazole-containing proteins, *Angew. Chem. Int. Ed. Engl. 47*, 2832-2835.

[26] Lee, Y. J., Wu, B., Raymond, J. E., Zeng, Y., Fang, X., Wooley, K. L., and Liu, W. R. (2013) A genetically encoded acrylamide functionality, *ACS Chem. Biol. 8*, 1664-1670.

[27] Kaya, E., Vrabel, M., Deiml, C., Prill, S., Fluxa, V. S., and Carell, T. (2012) A genetically encoded norbornene amino acid for the mild and selective modification

of proteins in a copper-free click reaction, *Angew. Chem. Int. Ed. Engl. 51*, 4466-4469.

[28] Yu, Z., Pan, Y., Wang, Z., Wang, J., and Lin, Q. (2012) Genetically encoded cyclopropene directs rapid, photoclick-chemistry-mediated protein labeling in mammalian cells, *Angew. Chem. Int. Ed. Engl. 51*, 10600-10604.

[29] Wang, Y., Song, W., Hu, W. J., and Lin, Q. (2009) Fast alkene functionalization in vivo by photoclick chemistry: Homo lifting of nitrile imine dipoles, *Angew. Chem. Int. Ed. Engl. 48*, 5330-5333.

[30] Liu, D. S., Tangpeerachaikul, A., Selvaraj, R., Taylor, M. T., Fox, J. M., and Ting, A. Y. (2012) Diels-alder cycloaddition for fluorophore targeting to specific proteins inside living cells, *J. Am. Chem. Soc. 134*, 792-795.

[31] Devaraj, N. K., and Weissleder, R. (2011) Biomedical applications of tetrazine cycloadditions, *Acc. Chem. Res. 44*, 816-827.

[32] Liu, S., Hassink, M., Selvaraj, R., Yap, L. P., Park, R., Wang, H., Chen, X., Fox, J. M., Li, Z., and Conti, P. S. (2013) Efficient 18f labeling of cysteine-containing peptides and proteins using tetrazine-trans-cyclooctene ligation, *Mol. Imaging 12*, 121-128.

[33] Merrifield, R. B. (1963) Solid phase peptide synthesis .1. Synthesis of a tetrapeptide, *J. Am. Chem. Soc. 85*, 2149-&.

[34] Dawson, P. E., Muir, T. W., Clark-Lewis, I., and Kent, S. B. (1994) Synthesis of proteins by native chemical ligation, *Science 266*, 776-779.

[35] Durek, T., and Becker, C. F. (2005) Protein semi-synthesis: New proteins for functional and structural studies, *Biomol. Eng 22*, 153-172.

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

[37] Xu, M. Q., and Perler, F. B. (1996) The mechanism of protein splicing and its modulation by mutation, *EMBO J. 15*, 5146-5153.

[38] Chong, S., Mersha, F. B., Comb, D. G., Scott, M. E., Landry, D., Vence, L. M., Perler, F. B., Benner, J., Kucera, R. B., Hirvonen, C. A., Pelletier, J. J., Paulus, H., and Xu, M. Q. (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element, *Gene 192*, 271-281.

[39] Muir, T. W., Sondhi, D., and Cole, P. A. (1998) Expressed protein ligation: A general method for protein engineering, *Proc. Natl. Acad. Sci. U.S.A. 95*, 6705-6710.

[40] Rashidian, M., Dozier, J. K., and Distefano, M. D. (2013) Enzymatic labeling of proteins: Techniques and approaches, *Bioconjugate Chem. 24*, 1277-1294.

[41] Sunbul, M., and Yin, J. (2009) Site specific protein labeling by enzymatic posttranslational modification, *Org. Biomol. Chem. 7*, 3361-3371.

[42] Mao, H., Hart, S. A., Schink, A., and Pollok, B. A. (2004) Sortase-mediated protein ligation: A new method for protein engineering, *J. Am. Chem. Soc. 126*, 2670-2671.

[43] Carlson, B. L., Ballister, E. R., Skordalakes, E., King, D. S., Breidenbach, M. A., Gilmore, S. A., Berger, J. M., and Bertozzi, C. R. (2008) Function and structure of a prokaryotic formylglycine-generating enzyme, *J. Biol. Chem. 283*, 20117-20125.

[44] Wu, P., Shui, W., Carlson, B. L., Hu, N., Rabuka, D., Lee, J., and Bertozzi, C. R. (2009) Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag, *Proc. Natl. Acad. Sci. U.S.A. 106*, 3000-3005.

[45] Chen, I., Howarth, M., Lin, W., and Ting, A. Y. (2005) Site-specific labeling of cell surface proteins with biophysical probes using biotin ligase, *Nat. Methods 2*, 99-104.

[46] Howarth, M., Takao, K., Hayashi, Y., and Ting, A. Y. (2005) Targeting quantum dots to surface proteins in living cells with biotin ligase, *Proc. Natl. Acad. Sci. U.S.A. 102*, 7583-7588.

[47] Fernandez-Suarez, M., Baruah, H., Martinez-Hernandez, L., Xie, K. T., Baskin, J. M., Bertozzi, C. R., and Ting, A. Y. (2007) Redirecting lipoic acid ligase for cell surface protein labeling with small-molecule probes, *Nat. Biotechnol. 25*, 1483-1487.

[48] Noren, C. J., Anthony-Cahill, S. J., Griffith, M. C., and Schultz, P. G. (1989) A general method for site-specific incorporation of unnatural amino acids into proteins, *Science 244*, 182-188.

[49] Mendel, D., Cornish, V. W., and Schultz, P. G. (1995) Site-directed mutagenesis with an expanded genetic code, *Annu. Rev. Biophys. Biomol. Struct. 24*, 435-462.

[50] Liu, C. C., and Schultz, P. G. (2010) Adding new chemistries to the genetic code, *Annu. Rev. Biochem 79*, 413-444.

[51] Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of escherichia coli, *Science 292*, 498-500.

[52] Wang, L., and Schultz, P. G. (2001) A general approach for the generation of orthogonal trnas, *Chem. Biol. 8*, 883-890.

[53] Srinivasan, G., James, C. M., and Krzycki, J. A. (2002) Pyrrolysine encoded by uag in archaea: Charging of a uag-decoding specialized trna, *Science 296*, 1459-1462.

[54] Hao, B., Gong, W., Ferguson, T. K., James, C. M., Krzycki, J. A., and Chan, M. K. (2002) A new uag-encoded residue in the structure of a methanogen methyltransferase, *Science 296*, 1462-1466.

[55] Wan, W., Huang, Y., Wang, Z., Russell, W. K., Pai, P. J., Russell, D. H., and Liu, W. R. (2010) A facile system for genetic incorporation of two different noncanonical amino acids into one protein in escherichia coli, *Angew. Chem. Int. Ed. Engl. 49*, 3211-3214.

[56] Wan, W., Tharp, J. M., and Liu, W. R. (2014) Pyrrolysyl-trna synthetase: An ordinary enzyme but an outstanding genetic code expansion tool, *Biochim. Biophys. Acta 1844*, 1059-1070.

[57] Liu, W. R., Wang, Y. S., and Wan, W. (2011) Synthesis of proteins with defined posttranslational modifications using the genetic noncanonical amino acid incorporation approach, *Mol. Biosyst. 7*, 38-47.

[58] Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P. G. (2007) Genetic incorporation of unnatural amino acids into proteins in mammalian cells, *Nat. Methods 4*, 239-244.

[59] Greiss, S., and Chin, J. W. (2011) Expanding the genetic code of an animal, *J. Am. Chem. Soc. 133*, 14196-14199.

[60] Chin, J. W., Cropp, T. A., Anderson, J. C., Mukherji, M., Zhang, Z., and Schultz, P. G. (2003) An expanded eukaryotic genetic code, *Science 301*, 964-967.

[61] Parrish, A. R., She, X., Xiang, Z., Coin, I., Shen, Z., Briggs, S. P., Dillin, A., and Wang, L. (2012) Expanding the genetic code of caenorhabditis elegans using bacterial aminoacyl-trna synthetase/trna pairs, *ACS Chem. Biol. 7*, 1292-1302.

[62] Wang, L., Zhang, Z., Brock, A., and Schultz, P. G. (2003) Addition of the keto functional group to the genetic code of escherichia coli, *Proc. Natl. Acad. Sci. U.S.A. 100*, 56-61.

[63] Zeng, H., Xie, J., and Schultz, P. G. (2006) Genetic introduction of a diketone-containing amino acid into proteins, *Bioorg. Med. Chem. Lett. 16*, 5356-5359.

[64] Zhang, Z., Smith, B. A., Wang, L., Brock, A., Cho, C., and Schultz, P. G. (2003) A new strategy for the site-specific modification of proteins in vivo, *Biochemistry 42*, 6735-6746.

[65] Huang, Y., Wan, W., Russell, W. K., Pai, P. J., Wang, Z., Russell, D. H., and Liu, W. (2010) Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications, *Bioorg. Med. Chem. Lett. 20*, 878-880.

[66] Chin, J. W., Santoro, S. W., Martin, A. B., King, D. S., Wang, L., and Schultz, P. G. (2002) Addition of p-azido-l-phenylalanine to the genetic code of escherichia coli, *J. Am. Chem. Soc. 124*, 9026-9027.

[67] Deiters, A., and Schultz, P. G. (2005) In vivo incorporation of an alkyne into proteins in escherichia coli, *Bioorg. Med. Chem. Lett. 15*, 1521-1524.

[68] Deiters, A., Cropp, T. A., Mukherji, M., Chin, J. W., Anderson, J. C., and Schultz, P. G. (2003) Adding amino acids with novel reactivity to the genetic code of saccharomyces cerevisiae, *J. Am. Chem. Soc. 125*, 11782-11783.

[69] Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., and Chin, J. W. (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-trna synthetase/trna(cua) pair and click chemistry, *J. Am. Chem. Soc. 131*, 8720-8721.

[70] Wang, Y. S., Fang, X., Chen, H. Y., Wu, B., Wang, Z. U., Hilty, C., and Liu, W. R. (2013) Genetic incorporation of twelve meta-substituted phenylalanine derivatives using a single pyrrolysyl-trna synthetase mutant, *ACS Chem. Biol. 8*, 405-415.

[71] Hao, Z., Song, Y., Lin, S., Yang, M., Liang, Y., Wang, J., and Chen, P. R. (2011) A readily synthesized cyclic pyrrolysine analogue for site-specific protein "click" labeling, *Chem. Commun. 47*, 4502-4504.

[72] Li, Y. M., Yang, M. Y., Huang, Y. C., Song, X. D., Liu, L., and Chen, P. R. (2012) Genetically encoded alkenyl-pyrrolysine analogues for thiol-ene reaction mediated site-specific protein labeling, *Chem. Sci. 3*, 2766-2770.

[73] Zhang, Z., Wang, L., Brock, A., and Schultz, P. G. (2002) The selective incorporation of alkenes into proteins in escherichia coli, *Angew. Chem. Int. Ed. Engl. 41*, 2840-2842.

[74] Wang, Y. S., Fang, X., Wallace, A. L., Wu, B., and Liu, W. R. (2012) A rationally designed pyrrolysyl-trna synthetase mutant with a broad substrate spectrum, *J. Am. Chem. Soc. 134*, 2950-2953.

[75] Ai, H. W., Shen, W., Brustad, E., and Schultz, P. G. (2010) Genetically encoded alkenes in yeast, *Angew. Chem. Int. Ed. Engl. 49*, 935-937.

[76] Lang, K., Davis, L., Wallace, S., Mahesh, M., Cox, D. J., Blackman, M. L., Fox, J. M., and Chin, J. W. (2012) Genetic encoding of bicyclononynes and trans-cyclooctenes for site-specific protein labeling in vitro and in live mammalian cells via rapid fluorogenic diels-alder reactions, *J. Am. Chem. Soc. 134*, 10317-10320.

[77] Lang, K., Davis, L., Torres-Kolbus, J., Chou, C., Deiters, A., and Chin, J. W. (2012) Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction, *Nat. Chem. 4*, 298-304.

[78] Borrmann, A., Milles, S., Plass, T., Dommerholt, J., Verkade, J. M., Wiessler, M., Schultz, C., van Hest, J. C., van Delft, F. L., and Lemke, E. A. (2012) Genetic encoding of a bicyclo[6.1.0]nonyne-charged amino acid enables fast cellular protein imaging by metal-free ligation, *ChemBioChem 13*, 2094-2099.

[79] Plass, T., Milles, S., Koehler, C., Szymanski, J., Mueller, R., Wiessler, M., Schultz, C., and Lemke, E. A. (2012) Amino acids for diels-alder reactions in living cells, *Angew. Chem. Int. Ed. Engl. 51*, 4166-4170.

[80] Giege, R., Sissler, M., and Florentz, C. (1998) Universal rules and idiosyncratic features in trna identity, *Nucleic Acids Res. 26*, 5017-5035.

[81] Chapman, J. S., Diehl, M. A., and Lyman, R. C. (1993) Biocide susceptibility and intracellular glutathione in escherichia-coli, *J. Ind. Microbiol. 12*, 403-407.

[82] Umehara, T., Kim, J., Lee, S., Guo, L. T., Soll, D., and Park, H. S. (2012) N-acetyl lysyl-trna synthetases evolved by a ccdb-based selection possess n-acetyl lysine specificity in vitro and in vivo, *FEBS Lett. 586*, 729-733.

[83] Yanagisawa, T., Ishii, R., Fukunaga, R., Kobayashi, T., Sakamoto, K., and Yokoyama, S. (2008) Multistep engineering of pyrrolysyl-trna synthetase to genetically encode n-epsilon-(o-azidobenzyloxycarbonyl) lysine for site-specific protein modification, *Chem. Biol. 15*, 1187-1197.

[84] Nguyen, D. P., Elliott, T., Holt, M., Muir, T. W., and Chin, J. W. (2011) Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation, *J. Am. Chem. Soc. 133*, 11418-11421.

[85] Wang, Y. S., Wu, B., Wang, Z., Huang, Y., Wan, W., Russell, W. K., Pai, P. J., Moe, Y. N., Russell, D. H., and Liu, W. R. (2010) A genetically encoded photocaged nepsilon-methyl-l-lysine, *Mol. Biosyst. 6*, 1557-1560.

[86] Gattner, M. J., Vrabel, M., and Carell, T. (2013) Synthesis of epsilon-n-propionyl-, epsilon-n-butyryl-, and epsilon-n-crotonyl-lysine containing histone h3 using the pyrrolysine system, *Chem. Commun. 49*, 379-381.

[87] Kim, C. H., Kang, M., Kim, H. J., Chatterjee, A., and Schultz, P. G. (2012) Site-specific incorporation of epsilon-n-crotonyllysine into histones, *Angew. Chem. Int. Ed. Engl. 51*, 7246-7249.

[88] Brueggemeier, S. B., Wu, D., Kron, S. J., and Palecek, S. P. (2005) Protein-acrylamide copolymer hydrogels for array-based detection of tyrosine kinase activity from cell lysates, *Biomacromolecules 6*, 2765-2775.

[89] Brueggemeier, S. B., Kron, S. J., and Palecek, S. P. (2004) Use of protein-acrylamide copolymer hydrogels for measuring protein concentration and activity, *Anal. Biochem. 329*, 180-189.

[90] Bast, K., Christl, M., Huisgen, R., and Mack, W. (1973) 1,3-dipolar cycloadditions .73. Relative dipolarophile activities in cycloadditions of benzonitrile oxide, *Chem. Ber. Recl. 106*, 3312-3344.

[91] Link, A. J., and Tirrell, D. A. (2003) Cell surface labeling of escherichia coli via copper(i)-catalyzed [3+2] cycloaddition, *J. Am. Chem. Soc. 125*, 11164-11165.

[92] Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C., Falck, J. R., Peng, J., Gu, W., and Zhao, Y. (2007) Lysine propionylation and butyrylation are novel post-translational modifications in histones, *Mol. Cell. Proteomics 6*, 812-819.

[93] Xie, Z., Dai, J., Dai, L., Tan, M., Cheng, Z., Wu, Y., Boeke, J. D., and Zhao, Y. (2012) Lysine succinylation and lysine malonylation in histones, *Mol. Cell. Proteomics 11*, 100-107.

[94] Hoffman, M., Rajapakse, A., Shen, X., and Gates, K. S. (2012) Generation of DNA-damaging reactive oxygen species via the autoxidation of hydrogen sulfide under

physiologically relevant conditions: Chemistry relevant to both the genotoxic and cell signaling properties of h(2)s, *Chem. Res. Toxicol. 25*, 1609-1615.

[95] Thalhammer, F., Wallfahrer, U., and Sauer, J. (1990) Reactivity of simple open-chain and cyclic dienophiles in inverse-type diels-alder reactions, *Tetrahedron Lett. 31*, 6851-6854.

[96] Knall, A. C., and Slugovc, C. (2013) Inverse electron demand diels-alder (iedda)-initiated conjugation: A (high) potential click chemistry scheme, *Chem. Soc. Rev. 42*, 5131-5142.

[97] Debets, M. F., van Berkel, S. S., Dommerholt, J., Dirks, A. T., Rutjes, F. P., and van Delft, F. L. (2011) Bioconjugation with strained alkenes and alkynes, *Acc. Chem. Res. 44*, 805-815.

[98] Karver, M. R., Weissleder, R., and Hilderbrand, S. A. (2011) Synthesis and evaluation of a series of 1,2,4,5-tetrazines for bioorthogonal conjugation, *Bioconjugate Chem. 22*, 2263-2270.

[99] Carlson, J. C., Meimetis, L. G., Hilderbrand, S. A., and Weissleder, R. (2013) Bodipy-tetrazine derivatives as superbright bioorthogonal turn-on probes, *Angew. Chem. Int. Ed. Engl. 52*, 6917-6920.

[100] Devaraj, N. K., Weissleder, R., and Hilderbrand, S. A. (2008) Tetrazine-based cycloadditions: Application to pretargeted live cell imaging, *Bioconjugate Chem. 19*, 2297-2299.

[101] Royzen, M., Yap, G. P., and Fox, J. M. (2008) A photochemical synthesis of functionalized trans-cyclooctenes driven by metal complexation, *J. Am. Chem. Soc. 130*, 3760-3761.

[102] Devaraj, N. K., Hilderbrand, S., Upadhyay, R., Mazitschek, R., and Weissleder, R. (2010) Bioorthogonal turn-on probes for imaging small molecules inside living cells, *Angew. Chem. Int. Ed. Engl. 49*, 2869-2872.

[103] Liang, Y., Mackey, J. L., Lopez, S. A., Liu, F., and Houk, K. N. (2012) Control and design of mutual orthogonality in bioorthogonal cycloadditions, *J. Am. Chem. Soc. 134*, 17904-17907.

[104] Seitchik, J. L., Peeler, J. C., Taylor, M. T., Blackman, M. L., Rhoads, T. W., Cooley, R. B., Refakis, C., Fox, J. M., and Mehl, R. A. (2012) Genetically encoded tetrazine amino acid directs rapid site-specific in vivo bioorthogonal ligation with trans-cyclooctenes, *J. Am. Chem. Soc. 134*, 2898-2901.

[105] Niederwieser, A., Spate, A. K., Nguyen, L. D., Jungst, C., Reutter, W., and Wittmann, V. (2013) Two-color glycan labeling of live cells by a combination of diels-alder and click chemistry, *Angew. Chem. Int. Ed. Engl. 52*, 4265-4268.

[106] Meier, A., and Sauer, J. (1990) Donor-acceptor substituted dienophiles in inverse-type diels-alder reactions, *Tetrahedron Lett. 31*, 6855-6858.

[107] Kovalev, E. G., Postovskii, I. Y., Rusinov, G. L., and Shegal, I. L. (1981) Cycloaddition to sym-tetrazines (the carboni-lindsey reaction) (review), *Chem Hetetrocycl Compd 17*, 1063-1076.

[108] Debets, M. F., van Berkel, S. S., Schoffelen, S., Rutjes, F. P., van Hest, J. C., and van Delft, F. L. (2010) Aza-dibenzocyclooctynes for fast and efficient enzyme pegylation via copper-free (3+2) cycloaddition, *Chem. Commun. 46*, 97-99.

[109] Agard, N. J., Prescher, J. A., and Bertozzi, C. R. (2004) A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems, *J. Am. Chem. Soc. 126*, 15046-15047.

[110] Baskin, J. M., Prescher, J. A., Laughlin, S. T., Agard, N. J., Chang, P. V., Miller, I. A., Lo, A., Codelli, J. A., and Bertozzi, C. R. (2007) Copper-free click chemistry for dynamic in vivo imaging, *Proc. Natl. Acad. Sci. U.S.A. 104*, 16793-16797.

[111] Gordon, C. G., Mackey, J. L., Jewett, J. C., Sletten, E. M., Houk, K. N., and Bertozzi, C. R. (2012) Reactivity of biarylazacyclooctynones in copper-free click chemistry, *J. Am. Chem. Soc. 134*, 9199-9208.

[112] Jewett, J. C., Sletten, E. M., and Bertozzi, C. R. (2010) Rapid cu-free click chemistry with readily synthesized biarylazacyclooctynones, *J. Am. Chem. Soc. 132*, 3688-3690.

[113] van Geel, R., Pruijn, G. J., van Delft, F. L., and Boelens, W. C. (2012) Preventing thiol-yne addition improves the specificity of strain-promoted azide-alkyne cycloaddition, *Bioconjugate Chem. 23*, 392-398.

[114] Tuley, A., Wang, Y. S., Fang, X., Kurra, Y., Rezenom, Y. H., and Liu, W. R. (2014) The genetic incorporation of thirteen novel non-canonical amino acids, *Chem. Commun. 50*, 2673-2675.

[115] Tharp, J. M., Wang, Y. S., Lee, Y. J., Yang, Y., and Liu, W. R. (2014) Genetic incorporation of seven ortho-substituted phenylalanine derivatives, *ACS Chem. Biol. 9*, 884-890.

[116] Breinbauer, R., and Kohn, M. (2003) Azide-alkyne coupling: A powerful reaction for bioconjugate chemistry, *ChemBioChem 4*, 1147-1149.

[117] Wang, Q., Chan, T. R., Hilgraf, R., Fokin, V. V., Sharpless, K. B., and Finn, M. G. (2003) Bioconjugation by copper(i)-catalyzed azide-alkyne [3 + 2] cycloaddition, *J. Am. Chem. Soc. 125*, 3192-3193.

[118] Clark, P. M., Dweck, J. F., Mason, D. E., Hart, C. R., Buck, S. B., Peters, E. C., Agnew, B. J., and Hsieh-Wilson, L. C. (2008) Direct in-gel fluorescence detection and cellular imaging of o-glcnac-modified proteins, *J. Am. Chem. Soc. 130*, 11576-11577.

[119] Dieterich, D. C., Link, A. J., Graumann, J., Tirrell, D. A., and Schuman, E. M. (2006) Selective identification of newly synthesized proteins in mammalian cells using bioorthogonal noncanonical amino acid tagging (boncat), *Proc. Natl. Acad. Sci. U.S.A. 103*, 9482-9487.

[120] Khmelinskii, A., Keller, P. J., Bartosik, A., Meurer, M., Barry, J. D., Mardin, B. R., Kaufmann, A., Trautmann, S., Wachsmuth, M., Pereira, G., Huber, W., Schiebel, E., and Knop, M. (2012) Tandem fluorescent protein timers for in vivo analysis of protein dynamics, *Nat. Biotechnol. 30*, 708-714.

[121] Kurra, Y., Odoi, K. A., Lee, Y. J., Yang, Y., Lu, T., Wheeler, S. E., Torres-Kolbus, J., Deiters, A., and Liu, W. R. (2014) Two rapid catalyst-free click

reactions for in vivo protein labeling of genetically encoded strained alkene/alkyne functionalities, *Bioconjugate Chem. 25*, 1730-1738.

[122] Lee, Y. J., Kurra, Y., Yang, Y., Torres-Kolbus, J., Deiters, A., and Liu, W. R. (2014) Genetically encoded unstrained olefins for live cell labeling with tetrazine dyes, *Chem. Commun. 50*, 13085-13088.

[123] Stewart, I. C., Bergman, R. G., and Toste, F. D. (2003) Phosphine-catalyzed hydration and hydroalkoxylation of activated olefins: Use of a strong nucleophile to generate a strong base, *J. Am. Chem. Soc. 125*, 8696-8697.

[124] Chan, J. W., Hoyle, C. E., and Lowe, A. B. (2009) Sequential phosphine-catalyzed, nucleophilic thiol-ene/radical-mediated thiol-yne reactions and the facile orthogonal synthesis of polyfunctional materials, *J. Am. Chem. Soc. 131*, 5751-+.

[125] Furman, J. L., Kang, M., Choi, S., Cao, Y., Wold, E. D., Sun, S. B., Smider, V. V., Schultz, P. G., and Kim, C. H. (2014) A genetically encoded aza-michael acceptor for covalent cross-linking of protein-receptor complexes, *J. Am. Chem. Soc. 136*, 8411-8417.

[126] Danielson, M. A., and Falke, J. J. (1996) Use of 19f nmr to probe protein structure and conformational changes, *Annu. Rev. Biophys. Biomol. Struct. 25*, 163-195.

[127] Frieden, C., Hoeltzli, S. D., and Bann, J. G. (2004) The preparation of 19f-labeled proteins for nmr studies, *Methods Enzymol. 380*, 400-415.

[128] Buer, B. C., and Marsh, E. N. (2012) Fluorine: A new element in protein design, *Protein Sci. 21*, 453-462.

[129] Pace, C. J., and Gao, J. (2013) Exploring and exploiting polar-pi interactions with fluorinated aromatic amino acids, *Acc. Chem. Res. 46*, 907-915.

[130] Mahadevi, A. S., and Sastry, G. N. (2013) Cation-pi interaction: Its role and relevance in chemistry, biology, and material science, *Chem. Rev. 113*, 2100-2138.

[131] Tuley, A., Lee, Y. J., Wu, B., Wang, Z. U., and Liu, W. R. (2014) A genetically encoded aldehyde for rapid protein labelling, *Chem. Commun. 50*, 7424-7426.

[132] Wang, Y. S., Russell, W. K., Wang, Z., Wan, W., Dodd, L. E., Pai, P. J., Russell, D. H., and Liu, W. R. (2011) The de novo engineering of pyrrolysyl-trna synthetase for genetic incorporation of l-phenylalanine and its derivatives, *Mol. Biosyst. 7*, 714-717.

[133] Takimoto, J. K., Dellas, N., Noel, J. P., and Wang, L. (2011) Stereochemical basis for engineered pyrrolysyl-trna synthetase and the efficient in vivo incorporation of structurally divergent non-native amino acids, *ACS Chem. Biol. 6*, 733-743.

[134] Kokura, K., Sun, L., Bedford, M. T., and Fang, J. (2010) Methyl-h3k9-binding protein mpp8 mediates e-cadherin gene silencing and promotes tumour cell motility and invasion, *EMBO J. 29*, 3673-3687.

[135] Li, J., Li, Z., Ruan, J., Xu, C., Tong, Y., Pan, P. W., Tempel, W., Crombet, L., Min, J., and Zang, J. (2011) Structural basis for specific binding of human mpp8 chromodomain to histone h3 methylated at lysine 9, *PLoS One. 6*, e25104.

[136] Chang, Y., Horton, J. R., Bedford, M. T., Zhang, X., and Cheng, X. (2011) Structural insights for mpp8 chromodomain interaction with histone h3 lysine 9:

Potential effect of phosphorylation on methyl-lysine binding, *J. Mol. Biol. 408*, 807-814.

[137] Jacobs, S. A., Fischle, W., and Khorasanizadeh, S. (2004) Assays for the determination of structure and dynamics of the interaction of the chromodomain with histone peptides, *Methods Enzymol. 376*, 131-148.

[138] Wang, F., Qin, L., Wong, P., and Gao, J. (2011) Facile synthesis of tetrafluorotyrosine and its application in ph triggered membrane lysis, *Org. Lett. 13*, 236-239.