Spring 5-2-2018

Elucidating Ubiquitin Chain Formation Catalyzed by E2 Enzymes through use of a Di-Ubiquitin Probe

Tomaya Carpenter
tcarpenter4@student.gsu.edu

ELUCIDATING UBIQUITIN CHAIN FORMATION CATALYZED BY E2 ENZYMES THROUGH USE OF A DI-UBIQUITIN PROBE

by

TOMAYA CARPENTER

Under the Direction of Jun Yin, PhD

ABSTRACT

Ubiquitin plays a major role in many cellular processes including protein degradation. Ubiquitin attaches to a lysine group of a substrate molecule through its carboxyl terminal end, resulting in monoubiquitination and the subsequent post-translational modification of the protein. Substrate modifications are far more complex due to seven ubiquitin lysine residues which serve as acceptors for the conjugation with other ubiquitin molecules. The focus of this thesis is centered on the elucidation of ubiquitin chain formation through catalysis of E2 cascade enzymes. Here, we incorporate (N^{ε}-L-Thiaprolyl-L-lysine) site specifically at the K11 position of ubiquitin. The incorporated protein is deprotected forming an acceptor ubiquitin molecule, which we utilize for ligation with a donor molecule to construct a di-ubiquitin probe and conjugate with a specific E2 enzyme. Future studies are aimed at retrieving a three-dimensional structure for this DiUb-E2 conjugate to understand the mechanisms of multiple proteins involved in ubiquitin chain synthesis.

INDEX WORDS: Ubiquitin pathway, Aminoacyl tRNA synthetase, Conjugating enzymes, Unnatural amino acid, Post-translational modifications, Aminoacylating

ELUCIDATING UBIQUITIN CHAIN FORMATION CATALYZED BY E2 ENZYMES THROUGH USE OF A DI-UBIQUITIN PROBE

by

TOMAYA CARPENTER

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2018

ELUCIDATING UBIQUITIN CHAIN FORMATION CATALYZED BY E2 ENZYMES THROUGH USE OF A DI-UBIQUITIN PROBE

by

TOMAYA CARPENTER

Committee Chair: Jun Yin

Committee: Kathryn Grant

Hao Xu

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2018

DEDICATION

To my mom, who showered me with unconditional love even through my toughest times

-Your Butterfly

ACKNOWLEDGEMENTS

I would like to personally thank Dr. Yin who granted me the opportunity to perform research in his collaborative group.

Among many professors, I thank Dr. Xao Hu and Kathryn Grant for their expertise in teaching organic and inorganic concepts, which were both useful in my work.

I would also like to thank my lab partners Han Zhou and Stacey Jones who provided me with much needed support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATP	adenosine triphosphate
Boc	tert-butyloxycarbonyl
Cbz	carboxybenzyl
Cdc34	ubiquitin conjugating enzyme E2 R1
di-Ub	diubiquitin
DMAP	4-dimethylaminopyridine
DTT	dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA	ethylenediaminetetraacetic acid
E1	ubiquitin activating enzyme
E2	ubiquitin conjugating enzyme
E3	ubiquitin protein ligase
Gly	glycine
HECT	homologous to E6-AP carboxyl terminus
HPLC	high performance liquid chromatography
IMAC	immobilized-metal affinity chromatography
IPTG.	isopropyl-Beta-d-thiogalactopyranoside
Lys	lysine
Met-tRNA	methionyl-transfer ribonucleic acid
mRNA	messenger ribonucleic acid
NTA	nitrilotriacetic acid
NCL	native chemical ligation

PAGE	polyacrylamide gel electrophoresis
PMSF	phenylmethylsulfonyl fluoride
Pyl	pyrrolysine
RINGs	really interesting new gene E_3 ligase
RBR	ring between ring E_3 ligase
SUMO	small ubiquitin like modifier
TCEP	tris(2-carboxyethyl) phosphine hydrochloride
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLC	thin layer chromatography
tRNA	transfer ribonucleic acid
UAP	ubiquitin associated protein
UBD	ubiquitin binding domains
UPS	ubiquitin proteasome system

1 INTRODUCTION

1.1 Ub Cascade Pathway

Ubiquitination is highly regulated by a series of cascade proteins: activating (E1), conjugating (E2), and ligating enzymes (E3), which all assist in the transfer of ubiquitin to a target substrate. There are approximately 2, 40, and 600 known E1, E2, E3 enzymes, respectively. Initially, a Ub molecule becomes activated by an E1 enzyme in the presence of ATP to form a thioester linkage between the carboxyl-terminal glycine of ubiquitin and a specific cysteine residue in the active site of the E1 protein. The activated molecule is then transferred to an E2 enzyme where it becomes covalently attached via transthiolation or aminolysis. Lastly, an E3 ligase enzyme facilitates the transfer of Ub either directly or indirectly to a target substrate, forming an isopeptide bond. Ub most commonly associates with a specific amino group from a lysine on the target molecule, although attachment to other groups.[6]

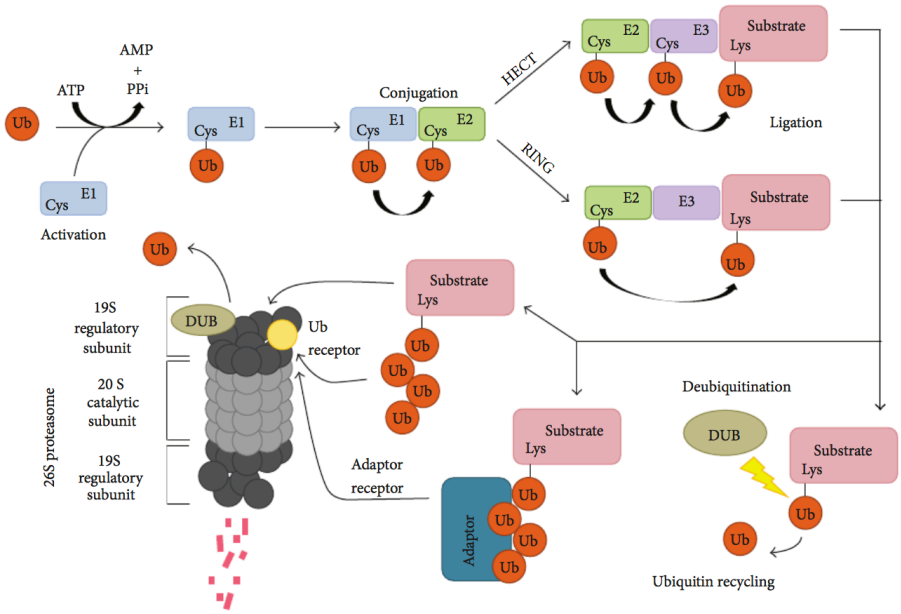


Figure 1.1 Ubiquitin Cascade

Ubiquitin is transferred from multiple proteins (E1, E2, and E3) to a target substrate. A thioester bond is formed between the carboxyl-terminal end of a glycine residue on ubiquitin and a cysteine in the active site of an E1 protein. Ub is then transferred to an E2 via transthiolation. Lastly, ubiquitin is the transferred to a target molecule, in this case which creates an isopeptide bond between a lysine residue on the substrate and the carboxyl end of ubiquitin.[4]

1.2 Ub Structure and the Correlation to Cellular Functions

Ub is a highly conserved protein made up of 76-amino acids and found in most eukaryotic cells. The stability of Ub makes up for its rather small sequence. The protein is not easily denatured at high temperatures and can sustain a wide range of pH levels. This is a result of stable hydrogen bonding of protons in alpha and beta helices as well as at the helix/sheet interface. Amide protons located in the back-bone region also contribute to the stability of ubiquitin due to hydrogen bonds formation upon refolding. The structure of Ub is quite interesting, which is composed of 1 β-sheet, 3 helices, and 6 β-turns, all surrounding a hydrophobic core.[4]

Protein modifications are diversified considering variations of different ubiquitination sites. Ub contains seven different lysine ubiquitination sites (K6, K11, K27, K29, K33, K48, K63) in addition to a methionine (M1) site, which differ in cellular functions and contributes to the overall complexity of the ubiquitination process. Thus, Ub can form eight different linkage types using any of the residues. Each linkage type results in specific interactions and facilitates the regulation of various biological functions. The methionine residue is associated with cell signaling and DNA repair. Lysine 6, 11, 27, 29, and 48 have been shown to contribute to cell cycle proteasome degradation. Lysine 33 and 63 facilitates in protein interactions and cell

signaling/trafficking, respectively.[16] Ubiquitination at the K63 position has also been shown to correlate with cell autophagy. Degradation of substrate proteins has been shown to involve multiple ubiquitin chains which are joined specifically by Lys^{48} on the N terminus of one Ub molecule and the carboxyl group of Gly^{76} of another Ub molecule. The post translational modification made on a substrate, whether degradation or cellular transportation to a certain site, is dependent on the polymerization state via monomeric or polyubiquitinated linkages to a target protein. The attachment of Ubto the lysine residue of a substrate molecule acts a signal usually to target the substrate for degradation by the 26S proteasome, which relies heavily on multiple protein interactions.[20]

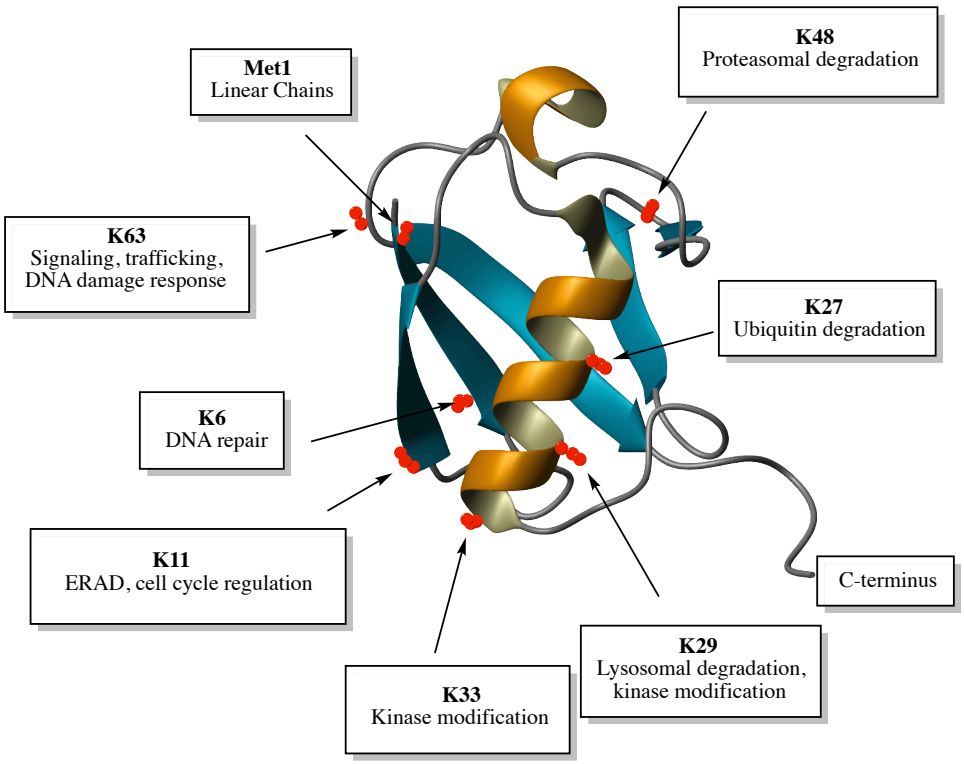


Figure 1.2 Structure of Ubiquitin

Ubiquitin and its seven lysine residues and one methionine binding sites shown with depicted cellular function. Cellular modifications are dependent on the linkage type of Ub to a particular substrate. Substrate molecules can be ubiquitinated at any one of the lysine or methionine binding residues. Substrates can even be polyubiquitinated, which is associated with the multiple attachment of ubiquitin molecules. This contributes to the overall complexity of the ubiquitination process considering different mechanisms and cellular modifications.[4,5]

1.3 Ub Chain Formation

Ub associates with a substrate molecule through an isopeptide bond between the carboxyl end of ubiquitin and a primary amine of the substrate, resulting in monoubiquitination of the substrate. The seven different lysine positions all serve as acceptors for conjugation with other ubiquitin molecules. Any of the exposed amino ends can function as an acceptor to the carboxyl group of a Gly residue of other Ub molecules, which adds multiple chains onto a substrate with different linkage types. As a result, substrates can be labelled for any particular function, predominantly for degradation. Essentially, the post-translational modification of a substrate is dependent on the chain linkage type.[6]

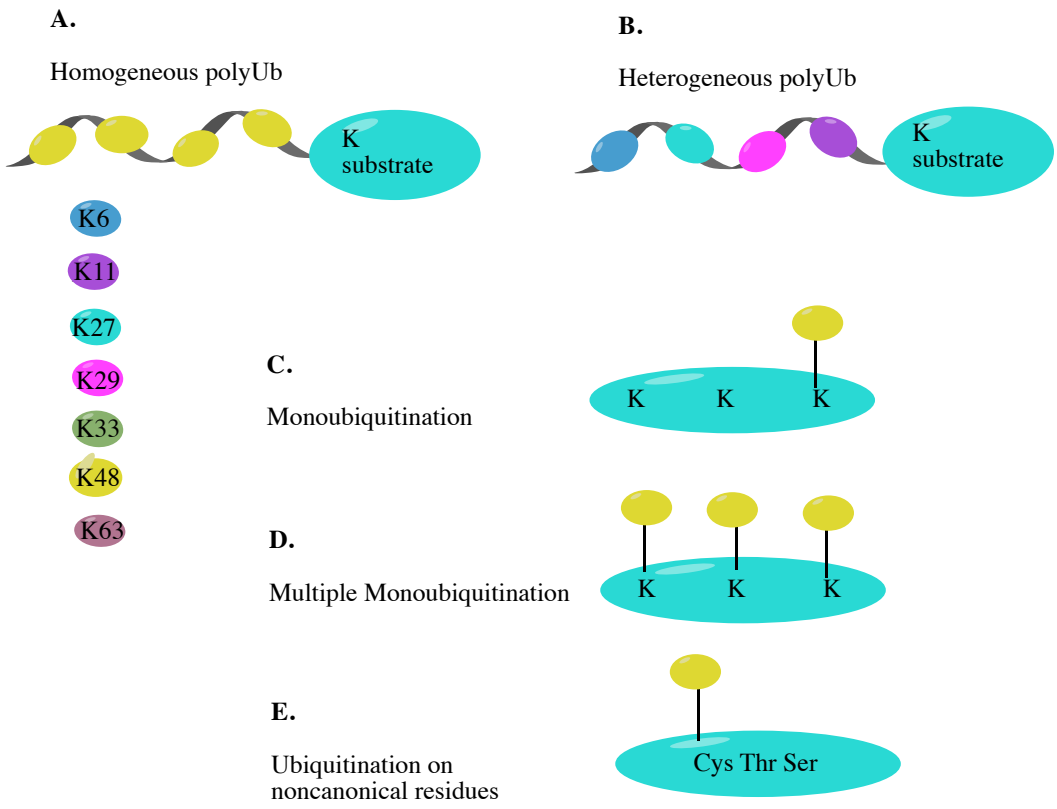


Figure 1.3 Ubiquitin Linkages

Different types of ubiquitination. **A.** Homogeneous polyubiquitination, which involves the multiple attachment of ubiquitin molecules at the same linkage or lysine residue. The exposed amino end of the same lysine residue (K48) acts as an acceptor to the carboxyl end of glycine residue on the previous ubiquitin molecule. **B.** Heterogeneous polubiquitination, which involves the multiple attachment of ubiquitin molecules via different linkage residues. **C.** Monoubiquitination of a substrate at the K48 position of ubiquitin. Monoubiquitination of this sort involves the covalent linkage of the carboxyl end if ubiquitin to the lysine end of a substrate. **D.** Mulitple monoubiquitination of a substrate. Instead of ubiquitin molecules associating with each other to create poly uB chains, multiple ubiquitin molecules associate with a specific lysine residue on the substrate. **E.** Noncanonical ubiquitination, which involves attachment of ubiquitin to a residue other than lysine on the target substrate.[6]

1.4 Currently Known Mechanisms for Ub Chain Synthesis

The conjugation of Ub to a target substrate is facilitated by a series of E1, E2, and E3 enzymes that function to activate, transfer, and ligate Ub to a target substrate. The mechanisms by which ubiquitin is transferred to a substrate have been publicized by recent studies. The area of numerous studies has been centered heavily on the transfer of Ub from an E2 to E3 and from an E3 to a substrate molecule.

Recent studies have shown two different mechanisms in the transfer of Ub to substrate molecules. The really interesting new gene (RING) and related U-box E3 ligase proteins catalyze the transfer of Ub from an E2 to a substrate directly. This is carried out by binding of the E3 ligase to the E2-Ub intermediate and to the substrate simultaneously. UBE2L3, an E2 conjugating enzyme, interacts with E3 RING-containing ligases. Instead of forming a thioester complex with ubiquitin, RING-like E3 proteins act as mere scaffolds to allow the direct transfer of ubiquitin from an E2 protein to a substrate.[24] The interaction between UBE2L3 and RING-like E3 proteins is facilitated by non-covalent interactions in which two E3 zinc (Zn^{2+}) ions coordinate to form a platform for binding.[15,25] The interaction of E2s with substrates either directly with E3 ligases or indirectly, suggests a variation in the mechanism for these proteins.[11,13,19]

Unlike RING-E3 ligases, HECT and RING-between-RING ligase proteins utilize a two-step mechanism for the ubiquitination of substrates. Members of the HECT E3s contain a HECT domain, which is characterized by a C-terminal region comprised of 350 amino acids. The N-terminal region of the HECT domain determines substrate specificity and is based on a specific amino acid sequence. As a result, HECT E3s are categorized into three sub families: Nedd4 (contains WW domain), HERC (contains RLDs), and “other” HECT E3s (contains neither WW

or RLD domain).[14] Ub is initially transferred from the E2 to the active site cysteine within the E3, which results in the formation of an E3-Ub intermediate.[3] Ub is then transferred from the E3 to the substrate. UbcH7, a well-known ubiquitin-conjugating E2 enzyme, forms a thioester bond with ubiquitin and associates with a specific domain of HECT E3 ligase proteins. The association of the UbcH7-ubiquitin conjugate with an HECT E3 ligase protein catalyzes the final attachment of ubiquitin to substrates, considering the formation of an intermediate thioester complex. It has been shown that the HECT domain of E3 ligase proteins contain a catalytic cysteine residue in the C-terminal lobe and an E2 binding domain in the N-terminal lobe. The lobes for each domain are flexible enough to allow suitable positioning of the catalytic cysteine towards the E2-Ub bond to allow transthiolation of ubiquitin from the E2 to E3 protein. It is apparent that E2 proteins must also contain two important domains: one containing the catalytic cysteine residue for the transfer of Ub from an E1 protein and the other bearing the binding site for an appropriate E3 protein considering the structural similarity of E2 and E3 enzymes. In all cases, interaction and binding of proteins to facilitate Ub transfer is specific and must be enzymatically sufficient.[19,23]

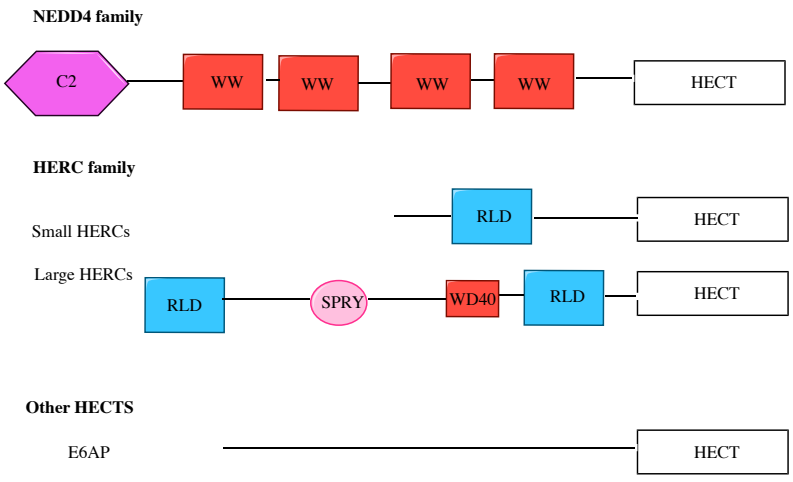


Figure 1.4 Members of Human E3 Ligase Proteins

The 3 subfamilies: NEDD4, HERC, and "other" HECTs. which are grouped according to the N-terminal region of the HECT domain. Substrate binding is based on the specific amino acid sequence located in the N-terminal to the HECT domain. Nedd4 family is characterized by an N-terminal C2 and WW domains. HERC contains an n-terminal RLD domain, while "Other" HECT E3s contain neither.[23]

1.5 Restrictions and Regulations for Ub Catalytic Chain Synthesis

Binding of substrates and the interaction with proteins for any given E1, E2, or E3 is distinct and specific and may even require the aid of other small molecules. For example, all members of HECT E3 ligases are characterized by the HECT domain which contains the catalytic cysteine residue in the C-terminal lobe and E2 binding domain in the N-terminal, but not all interact effectively with E2 proteins.[24] The HECT domain of SMURF2, a member of the Nedd4 sub-family of E3s, interacts ineffectively with the E2 protein UbcH7 compared to other E3 ligases. SMAD7 binds to both the HECT domain of SMURF2 and UbcH7 and acts as an auxiliary protein to allow necessary interaction between SMURF2 and UbcH7. In the absence of SMAD7, SMURF2 exists in an inactive form in which the interaction between the C2 and HECT domains serve as an auto-inhibitory function. In the presence of SMAD7, the C2 and HECT domain of SMURF2 become displaced. The interaction of SMAD7 with SMURF2 is seen in the regulation of TGF-β signaling pathway. SMAD7 consists of an amino-terminal domain (NTD), a linker region containing a PY motif that binds to the WW domains of SMURF2, and an MH2 domain that interacts with TGF-β receptor complex.[26] Crystal structures reveal NTD binding of SMURF2 to both HECT and UbcH7, suggesting catalyzed efficiency of the domain for E2

binding and ubiquitination. This shows the regulation for some HECT domains by interaction with certain molecules, which allows ubiquitination of a substrate only when necessary.

Although a general concept for the ubiquitination process is defined, the specifics for how a substrate is recognized and the role different proteins play during interaction with others is still quite ambiguous. Members of the Nedd4 family predominantly form K63-linked ubiquitin chains, while other HECT E3 ligases, such as E6AP, form K48-linked chains. The mechanisms by which a substrate is ubiquitinated is still not clear and remains somewhat unresolved.[14]

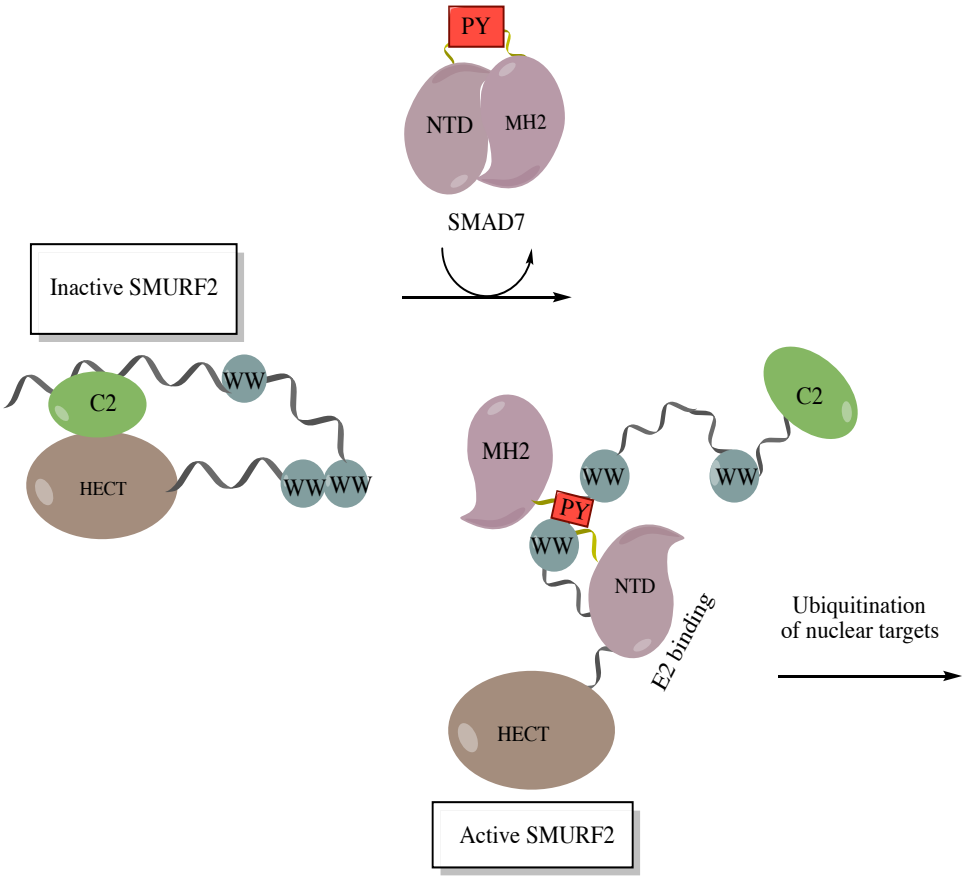


Figure 1.5 SMAD7 Interactions with SMURF2

Interaction of SMAD7 with SMURF2, a member of the Nedd4 E3s, and the E2 UbcH7. The interaction of SMAD7 with these certain proteins is still not fully understood, but current studies suggest a capable mechanism. The binding of SMURF2 HECT and C2 domains results in the auto inhibition of the E3 and its inactive form, which occurs in the absence of SMAD7. In the presence of SMAD7, the two domains become dissociated and the E3 becomes active and responsive to E2 binding. SMAD7 consists of an NTD and MH2 region connected by a PY motif. The PY motif interacts with two WW domains of the E3, while the MH2 domain has been shown to interact with TGF-β receptor complex in the TGF-β signaling pathway. NTD plays a major role in ubiquitination as it specifically binds to both the HECT region of SMURF2 and the E2 UbcH7, allowing efficient interaction between the E2 and E3 proteins.[26]

1.6 An Approach towards Ub Catalytic Chain Synthesis

It is understood that ubiquitin chain formation is catalyzed by a series of cascade proteins, however, the mechanisms behind these interactions are poorly understood. Here, we focus primarily on E2 proteins because they are key mediators in the ubiquitination process and govern the switch between chain initiation and elongation. During ubiquitin elongation, it is the E2 protein which catalyzes reaction of an acceptor Ub molecule bearing an exposed lysine residue to a donor Ub molecule. The acceptor Ub molecule nucleophilically attacks the C-terminal end of the donor Ub, resulting in a tetrahedral intermediate. Formation of the tetrahedral intermediate is transient and provides inability to capture a clear image on the structure during Ub chain elongation. Therefore, designing a more stable di-Ub probe **3** which can be conjugated to a specific E2 protein is important for understanding and determining cellular modifications and interactions with other

molecules. This di-Ub-E2 conjugate **4**,**6** functions as a mimic of the transition state **2** during Ub elongation, which grants dichotomizing the roles of Ub chain formation. With this, we can aid in resolving undistinguishable issues, including mechanisms of ubiquitin chain synthesis, Ub assembly on a substrate, and other regulatory events that occur. Our approach is centered on generating a di-Ub-E2 conjugate in hopes of retrieving X-ray crystallography to provide better understanding the mechanisms behind Ub chain linkages.[2]

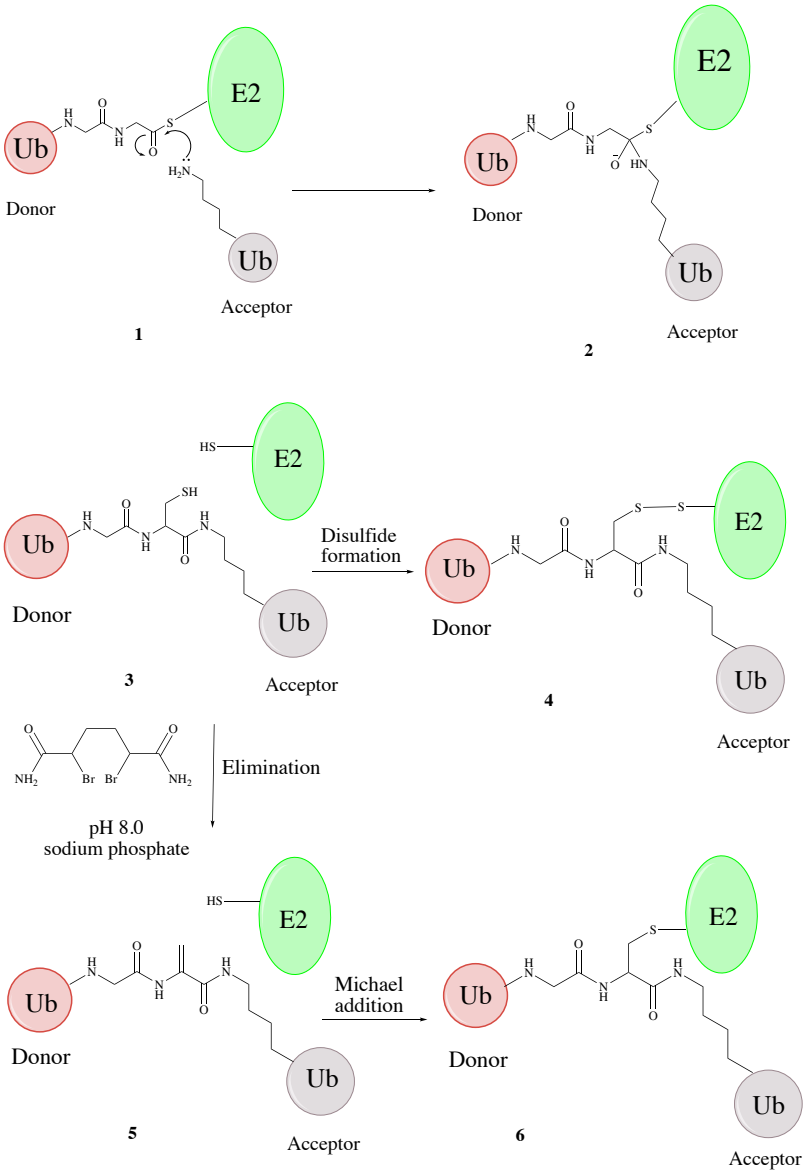


Figure 1.6 Elucidating Formation of E2 Catalyzed Ubiquitin

1.7 Overview and Strategy on Di-Ub Probe

Construction of our Di-Ub probe **3** is centered on four basic steps involving synthesis, incorporation, deprotection, and ligation of an unnatural amino acid, which will be discussed in greater detail in chapters 2-4. A specified amino acid can be incorporated into ubiquitin through an isopeptide bond or another stable, complementary group in response to an amber stop codon. After generating the protein of interest containing the incorporated UAA, the acceptor Ub molecule can be ligated to a donor Ub utilizing native chemical ligation to generate the di-Ub probe. This is useful as it allows the various roles of ubiquitination to be accessed. The framework of the di-Ub probe is outlined.

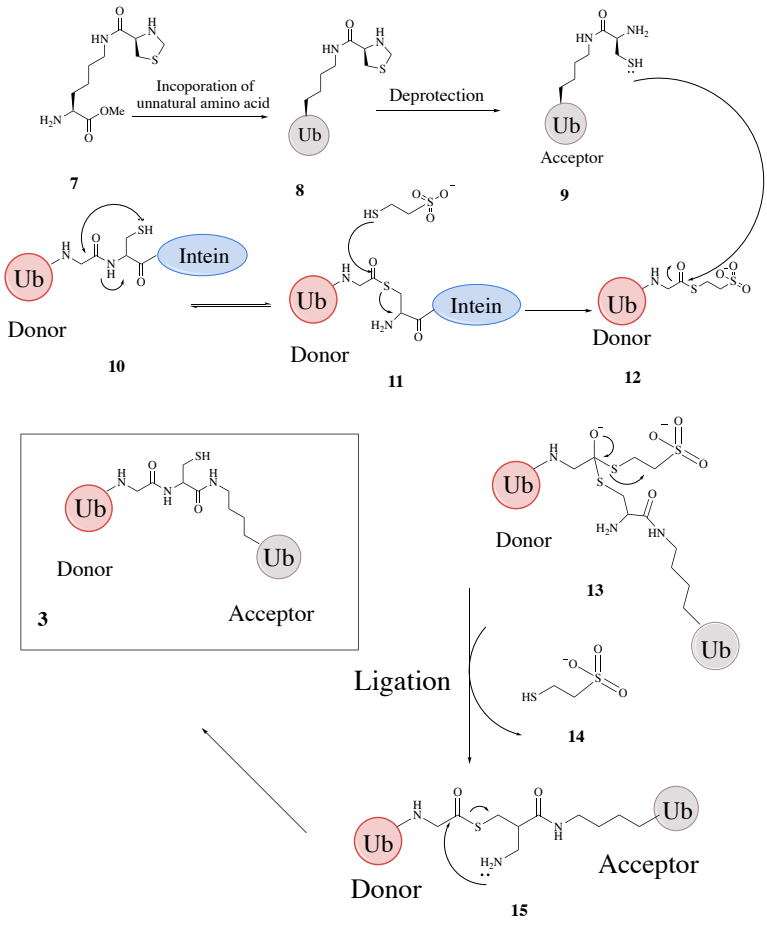


Figure 1.7 Synthesis of Di-Ub Probe for Conjugation with an E2 Enzyme

Steps for construction of the di-Ub probe. The first step includes synthesis of the UAA, followed by incorporation of the UAA into ubiquitin at a specific site. Incorporation of the UAA is achieved by integration into a host genome and generated in large amounts. The UAA is deprotected to create a reactive molecule which acts as an acceptor ubiquitin to another ubiquitin molecule.

1.7.1 Design of Di-Ub Probe by UAA Synthesis

Construction of the di-uUb conjugate is initiated with synthesis of an unnatural amino acid. There are 21 amino acids commonly found in human cells which are necessary for the synthesis of proteins. The physical and chemical properties of each amino acid are heavily dependent on the R group side chain, which allows differentiation in hydrophobicity, size, charge, and shape. Naturally occurring amino acids form polypeptide bonds with one another and are incorporated into proteins by RNA transfer.

DNA mutagenesis, which involves the site-specific replacement of any amino acid in a protein sequence, has played a significant role in determining structures and biological activity for proteins. New discoveries and techniques have also allowed the expansion of amino acids, which are not commonly found in human cells. Unnatural amino acids are produced by chemically modifying a natural amino acid to construct a closely similar molecule. Analogue modifications include alanine/glycine derivatives, β-amino acids, and linear core amino acids. Synthesis of our unnatural amino acid, N^{ε}-L-Thiaprolyl-L-lysine involves a modified four-step reaction, which yields a lysine derivative containing a thiazolidine ring.[1]

1.7.2 Design of Di-Ub Probe by UAA Incoporation

Incorporation of the synthesized unnatural amino acid into a specific protein site is the succeeding step for the development of the di-ubiquitin conjugate, which contributes to various strategies and applications in unnatural amino acid mutagenesis.

The central dogma of cellular biology outlines the process for the assembly of proteins. During protein synthesis for natural amino acids, DNA is converted into protein by a two-step reaction with utilization of appropriate enzymes. First, DNA is transcribed into messenger RNA in the nucleus which is then moved to the cytoplasm to be translated into protein. The mRNA strand serves as the template for protein synthesis because it carries the necessary codons, which are sequences of three amino bases. Protein synthesis occurs when the small ribosomal unit attaches to the mRNA strand and recognizes the AUG start codon. The aminoacyl tRNA synthetase contains two binding sites and activates a tRNA/amino acid pair. Protein synthesis begins when the first binding site recognizes a methionine amino acid, resulting in binding of ATP and the subsequent release of a pyrophosphate molecule. Activated tRNA containing the anticodon sequence binds to the second site and the tRNA/amino acid pair (Met-tRNA) is released from the enzyme to initiate the synthetic process. The tRNA bearing the anticodon attaches to the start codon sequence on mRNA with assistance of the large ribosomal unit. Elongation of the peptide sequence occurs when another tRNA carrying molecule bearing an amino acid binds to the ribosomal complex and associates with the mRNA codon sequence. Consequently, each natural amino acid has a specific aminoacyl tRNA synthetase molecule that activates the tRNA/amino acid pair. Protein elongation continues until the ribosomal unit reaches the stop codon UAG on mRNA.[21,22]

Successful incorporation of unnatural amino acids at particular sites in proteins involves construction of a tRNA synthetase that doesn't recognize host tRNA and only aminoacylates the tRNA containing the appropriate UAA. The tRNA/synthetase pair has to be engineered from a host, typically strains of *E. coli* cells, which occurs by mutating host tRNA that isn't recognizable by its synthetase. The production of evolved orthogonal tRNAs from an original source is attained by a series of negative and positive selection. The process of negative and positive selection involves constructing a library of *E. coli* cells containing mutant tRNAs in the presence and absence of cognate synthetases. In negative selection, a tRNA library is introduced into *E. coli* cells with a mutant gene in which amber nonsense codons are added at specific sites. Cells are destroyed when the suppressor tRNA is aminoacylated by an endogenous synthetase. Consequently, cells that undergo aminoacylation of orthogonal tRNAs survive, which are utilized in the positive selection. In this process, tRNA aminoacylation with a cognate synthetase leads to cell survival as a result of positive selectivity. Combination of positive and negative selection also generates an evolution of synthetases which assists the aminoacylation of cognate tRNA with the UAA of interest. Synthetase variants, which are constructed by randomizing residues in the binding pocket, are able to specifically recognize an UAA.[22]

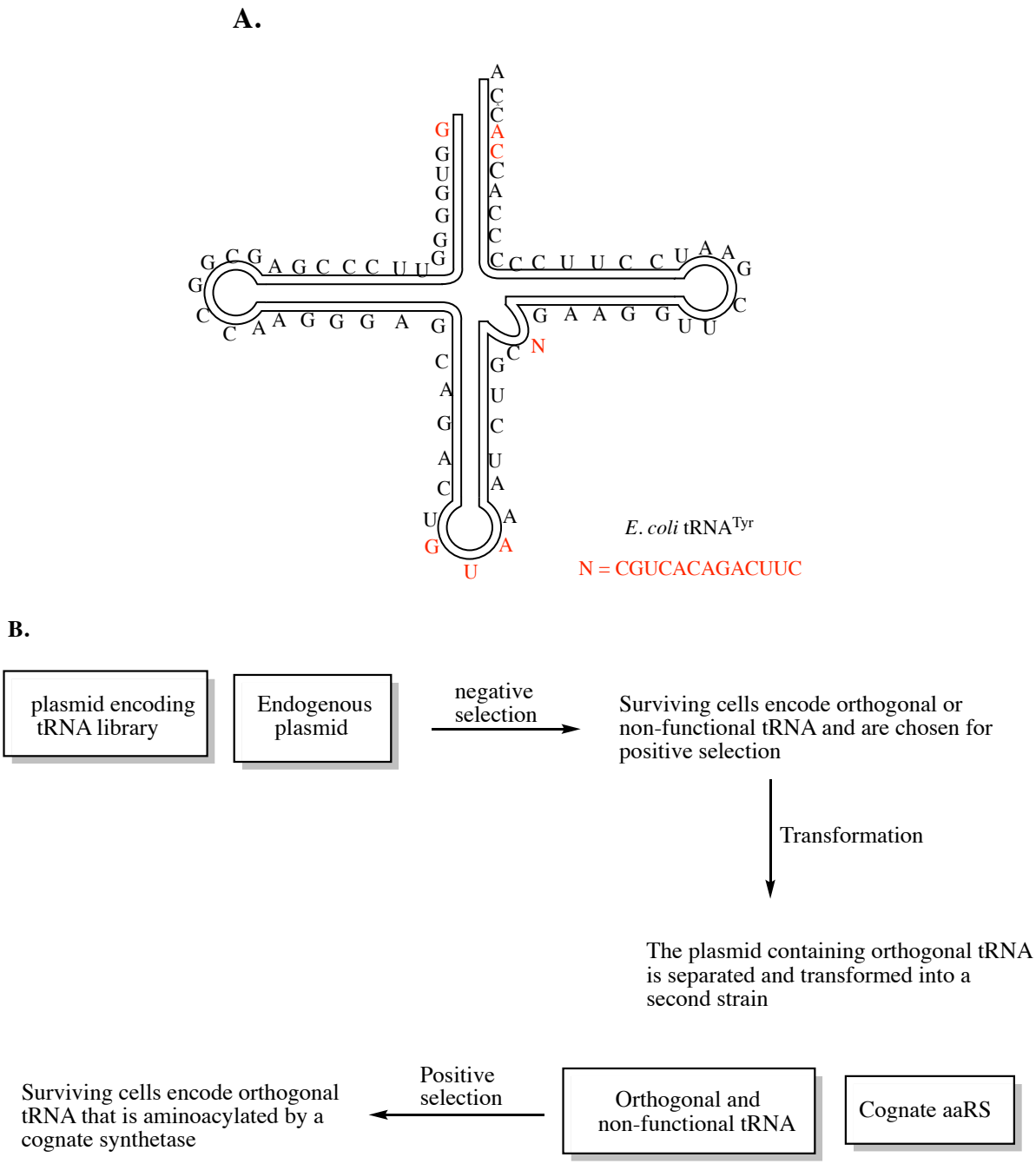


Figure 1.8 Orthogonal tRNA/tRNA Synthetase Pair for UAA Incorporation

A. *E. coli* $tRNA^{Tyr}$, which is one of approximately 20 canonical tRNAs. The evolution of orthogonal tRNA involves generating a suppressor tRNA library which contains a mutant gene in which amber nonsense codons are introduced at sites to substitution by other amino acids. **B.**

The process of negative and positive selection for a suppressor tRNA library. In the negative selection, tRNAs aminoacylated by orthogonal synthetases result in survival for *E. coli* cells. These tRNAS are then subjected to a positive selection in the presence of a cognate synthetase. Again, tRNAS aminoacylated by orthogonal synthetases grant cell survival.[21]

1.7.3 Design of Di-Ub Probe by Native Chemical Ligation

After successful incorporation of the unnatural amino acid, the acceptor Ub is used for ligation with another expressed protein that serves as the donor ubiquitin. This method known as native chemical ligation joins smaller peptide fragments together and involves two fully deprotected molecules. Our acceptor Ub molecule contains an N-terminal cysteine residue, which enables nucleophilic attack on the C-terminal thioester present on the donor ubiquitin. The reaction involves thioesterification to generate an intermediate, which undergoes an S to N acyl transfer mechanism to form the respective peptide bond. This reaction serves to generate our di-Ub conjugate.[8]

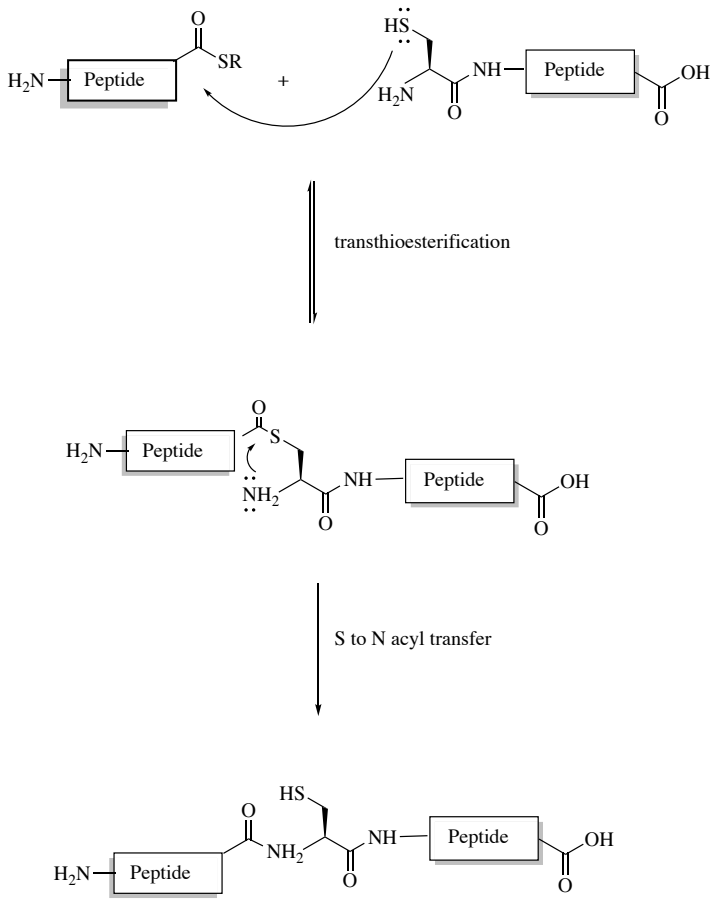


Figure 1.9 Native Chemical Ligation

The mechanism involves transthioesterification in which a nucleophilic N-terminal cysteine reacts with a C-terminal thioester to form a reactive thioester intermediate. The reaction undergoes an irreversible S to N acyl transfer due to the instability of the intermediate, creating a peptide bond.

1.8 Current Models of E2 Catalyzed Ub Chain Synthesis

Mutations and dysfunctions in E2 proteins are becoming a huge concern as they have led to many human diseases, such as cancer and other neurological illnesses. Therefore,

comprehension on the role of E2 proteins in ubiquitination is vital for an approach towards future treatment options for such diseases. Understanding the underlying mechanisms for these proteins is necessary for insight on interactions with other proteins. As a result, scientists can control and regulate certain protein-protein interactions, which aids in future studies and determining unresolved protein structures and functions.

E2 enzymes have an important role in ubiquitination by serving as intermediaries in the three-step cascade, which results in the transfer of ubiquitin from an E1 to an E3 protein. Ubiquitin conjugating proteins are characterized by a conserved 150 amino acid catalytic core known as the ubiquitin conjugation domain (UBC), which contains the catalytic cysteine. This region is what facilitates formation of a thioester bond with ubiquitin. Consequently, ubiquitin transfer relies heavily on the reactivity of E2 proteins and its nucleophilic regions, including cysteine groups.[2,10]

Ub conjugating enzymes possess a wide range of functions. Some E2s have been reported to catalyze Ub chain formation through a non-covalent backside binding site that is distant from the active site cysteine. The backside binding site of an E2 enzyme, Ubc13, positions the ubiquitin E2 Mms2 to facilitate K63 chain synthesis. Other E2s, such as Ube2D3 and Ubc9, behave in the same manner as Ubc13 by using the backside binding site to interact with Ub or SUMO for ubiquitin chain elongation. Current advances in understanding the E3 ligase proteins continue to show new E2 activities during Ub chain synthesis.[2]

Studies and images have provided evidence regarding the structure of ubiquitin conjugating proteins, which hint that other regions of E2 proteins contribute significantly to enzymatic activity. For instance, high resolution three-dimensional images of the E2 protein Ube21 indicates three conserved residues that play a substantial role in the transfer of ubiquitin

like SUMO protein to a substrate by activating the substrate acceptor lysine. The important E2 residues (Asn85, Tyr87, and Asp127) contribute to the activation of the substrate molecule by lowering the substrate lysine pKa, which allows nucleophilic attack of the conjugated Ube21-SUMO thioester bond. These regions of E2 proteins are important because mutations yield changes in their catalytic function. A mutation of Cdc34 (Asp127 homologue) in yeast E2 proteins resulted in a change from poly-ubiquitination to mono-ubiquitination of substrates, although ability of ubiquitin to attach to the substrate was not hindered. When Ser120 (Asp127 homologue for human E2 protein) is mutated to alanine or threonine, the activity of the E2 is either downregulated or upregulated, depending on the position of the amino acid. Both examples illustrate the importance of specific conserved regions of E2 proteins and how disruptions in these sites could affect the enzymatic function of E2 proteins in response to the substrate lysine towards the catalytic cysteine. With this in mind, various E2 proteins and their structures can be analyzed in order to determine how they work and interact with other proteins.[10]

1.9 Structure of Di-Ub Probe and Model for E2 Catalyzed Ub Chain Ligation

Our di-Ub probe can be conjugated to a specific E2 protein through disulfide formation and Michael addition. Construction of the di-Ub-E2 conjugate is necessary for mimicking the transition state during Ub chain elongation, in which an acceptor Ub molecule is added onto the Ub chain. During Ub chain formation, the C-terminal end of the donor Ub molecule is associated with a cysteine in the active site of the E2. The E2 then facilitates appropriate configuration allowing a specific lysine residue of an acceptor Ub to nucleophilically attack the C-terminal end

of the donor Ub, linking the two Ub molecules together. By constructing a di-Ub-E2 conjugate we can access Ub chain formation and better understand the roles of Ub formation and transfer.

1.10 Elucidating the Formation of E2 Catalyzed Ub Synthesis through X-ray Crystallography

Generation of a di-ubiquitin conjugate provides great opportunity on future studies, in terms of interactions with other proteins and enzymatic mechanisms. Here, we hope to reveal mechanisms of E2 catalyzed Ub chain ligation by obtaining a well-defined three-dimensional crystal structure of our di-Ub probe conjugated with a specific E2 enzyme, Ube2s.

2 SYNTHESIS OF THZK

2.1 Synthetic Scheme of ThzK

This chapter will focus on synthesis of the unnatural amino acid, Nε-L-Thiaprolyl-L-lysine (ThzK). Initially, our UAA is synthesized using a five-step mechanism, but later modified via four-step mechanism for improved yields and enhanced incorporation into protein. Synthesis of the UAA is followed by purification via HPLC, therefore the reported values are representative of a purified product. The mechanisms are discussed.

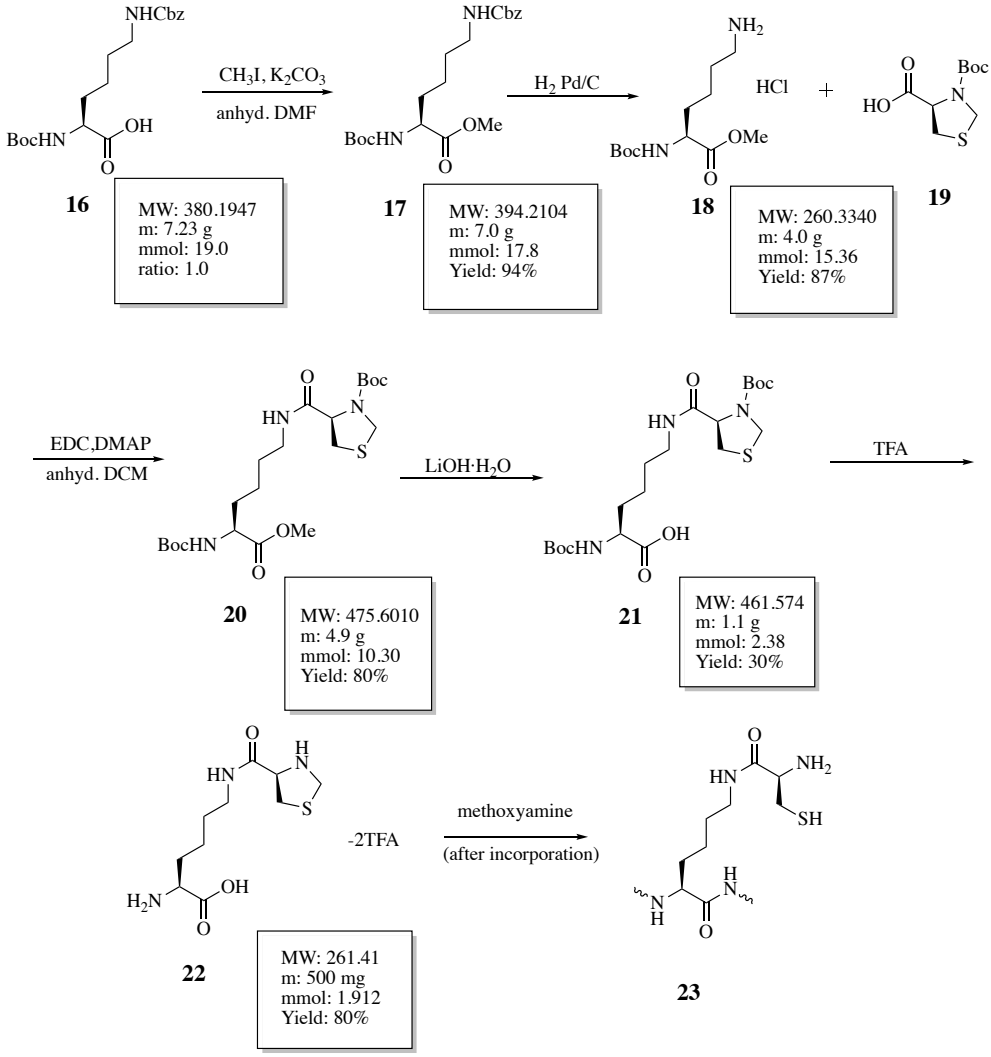


Figure 2.1 Synthetic Scheme for Unnatural Amino Acid

The UAA is generated with a five-step mechanism. Steps include conversion of an OH group to a methylated form followed by reduction to a free amine. The free amine is coupled with a carboxylic acid to yield an amide product. The methylated form is converted back to non-methylated form to a hydroxyl group. The final step undergoes reduction of two Boc groups to yield the target product.

Non-methylated ThzK was synthesized with a five-step mechanism. The reaction involves methylation of a carboxyl group via SN_2 mechanism and the subsequent reduction of a carboxybenzyl (Cbz) protecting group utilizing hydrogen gas. The next step involves coupling between the generated free amine and carboxylic acid, resulting in the formation of an amide product containing a thiazolidine ring. In this step, the carboxylic acid is deprotonated and reacts with a water-soluble carbodiimide to form a reactive o-Acylisourea intermediate. The unstable intermediate is reactive towards a free amine, which allows conjugation for efficient amide formation. This step is critical, as it was modified from the original protocol appropriate to the mechanism. The original protocol involved coupling between the primary amine and carboxylic acid followed by addition of DMAP and EDC. However, the mechanism often resulted in low yield and was modified accordingly. The fourth step involves conversion of the methylated (OMe) group to a hydroxyl group, followed by reduction of tert-butyloxycarbonyl (Boc) under acidic conditions.[7]

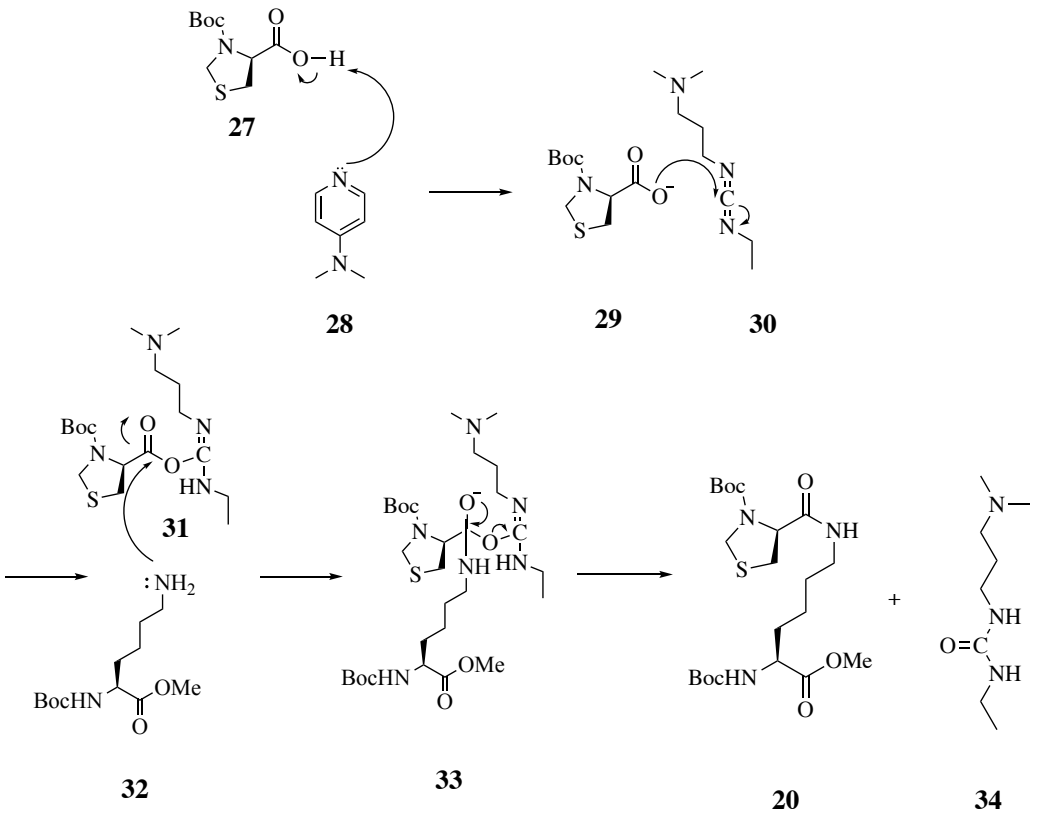


Figure 2.2 EDC Coupling Mechanism

A modification to the third step of the synthetic scheme was obtained according to the coupling mechanism, as stated above. Initially, DMAP **28** deprotonates the carboxylic acid **27**, which enables a nucleophilic attack on EDC **30**. Addition of the primary amine **32** creates an unstable o-Acylisourea intermediate **33**, which is removed to generate an isourea by-product **34**. This also results in formation of the coupled amide product **20**. Following this protocol resulted in higher yields for the third coupling step.

2.1.1 Deprotection using TFA as a Reducing Agent

The most common removal for Boc groups require the use of strong acids, such as TFA. Tert-Butyl carbamates are cleaved under these conditions to yield free amines. However, use of TFA as a reducing agent can produce undesirable side products, such as TFA salts. Purification of the final step via ether precipitation (precipitating the product from the reaction mixture with cold ether) was rather challenging, often resulting in low yields (less than one gram). This is likely due to the production of TFA salts and the inability of product to precipitate out of solution. Solubility of a molecule plays an important role in precipitation and it highly possible that the high solubility of the product makes ether precipitation somewhat difficult. It is also important to consider the potential harmful effects of trifluoroacetic acid salts, which are detrimental to growing cells. Such contaminants could lead to a substantial decrease in cell proliferation and a subsequent decrease in expressed protein. As a result, the procedure was improved with intentions of increasing the overall yield of the UAA and limiting the amount of harmful side products.[7,21]

2.2 An Improvised Mechanism on the Synthesis of ThzK

2.2.1 Deprotection using Hydrochloride as a Reducing Agent

The modified protocol involved using a 4M solution of hydrogen chloride as a reducing agent, instead of TFA. Hydrogen chloride is a more suitable reducing agent as undesirable side product are limited. When HCl is used for the deprotection, the product is obtained as a HCl salt, which is less harmful, more soluble and easier to handle. Therefore, a modified procedure was generated in anticipation for increasing the amount of UAA produced and suitable incorporation into cell for protein expression.

2.3 Yield Analysis on Improvised Synthetic Scheme

The yield for the modified protocol produced comparable results to the original synthetic scheme instead of substantially increasing the yield. Similarly, the amount of UAA obtained was less than a gram. According to the data in in Figures 2.2 and 2.4, the yield of the unnatural amino acid shows a significant decrease during the fourth step, involving conversion of the methoxy group to hydroxyl group. The step involves generation of carboxylate ions which are collected in an aqueous layer during extraction and purification techniques. The aqueous layer is acidified with a 5% citric acid solution to give the target product., which then becomes soluble in organic layer. The decrease in yield could be a result of many steric and electronic factors in the reaction. Also, the low polarity of THF solvent utilized in the reaction could hinder product formation. Many studies have also revealed that methylated forms of UAA display higher yields and better incorporation into cells. Methylated UAAs are less charged and are believed to be better transferred into cells. As a result, the protocol was modified further using a four-step reaction that yielded a methylated unnatural amino acid in attempts for substantially increasing the overall yield.[21]

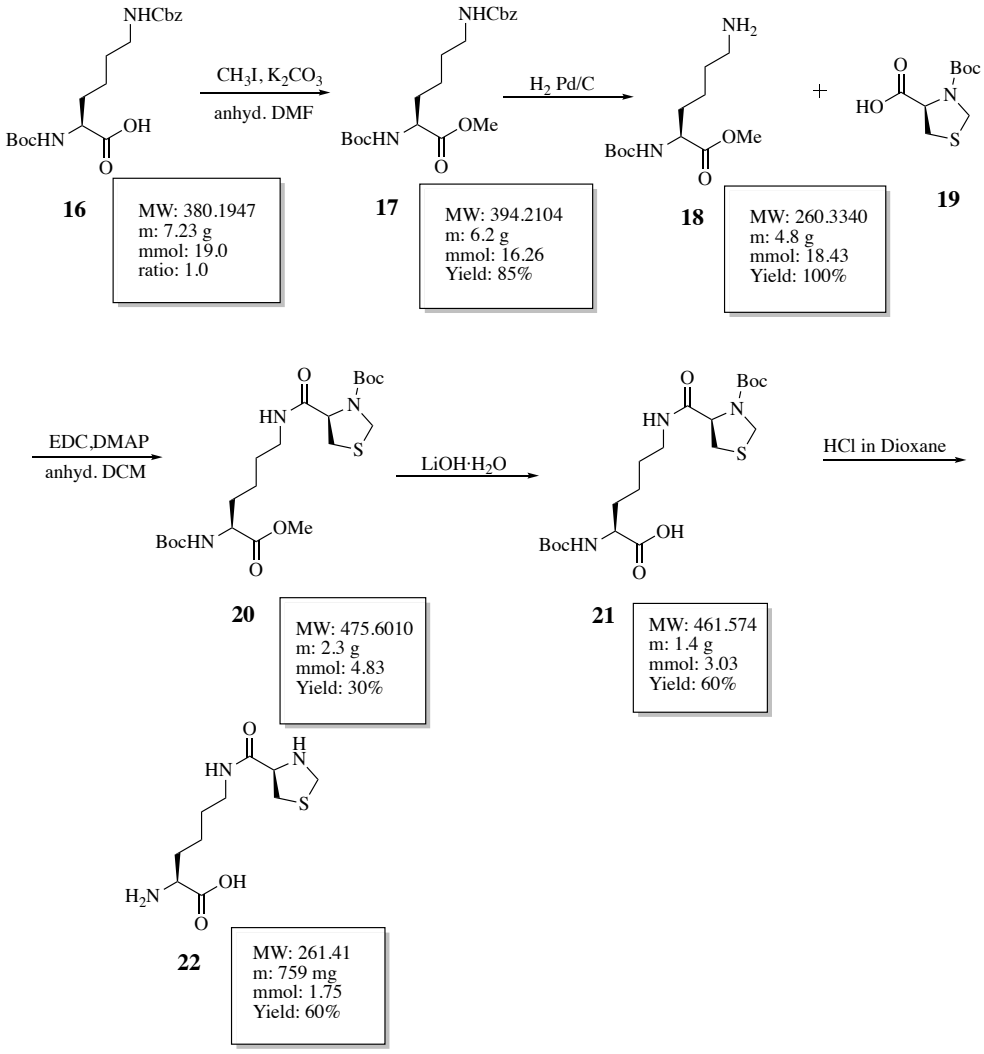


Figure 2.3 Yield Analysis for Modified Approach using HCl as a Reducing Agent

Data represents the first attempt using modified protocol. Each step shows the formula weight, mass, and mmol in which the yield is calculated based on the preceding step. Results using HCl as a reducing agent produces similar results as in original scheme, although detrimental side products are reduced.

Synthesis of the second modified protocol involved production of a methylated form of the unnatural amino acid, which was shown to not only have a higher yield during synthesis but

also during incorporation (will be discussed further in chapter 3). The first three steps of the initial protocol involving methylation, reduction, and coupling were unchanged. However, the fourth step involving conversion from a methoxy to OH group was depleted. To help in removal of impurities, the third step was subjected to purification via column chromatography. The final step involved reduction of the Boc group using a 4M solution of hydrogen chloride, as mentioned before.

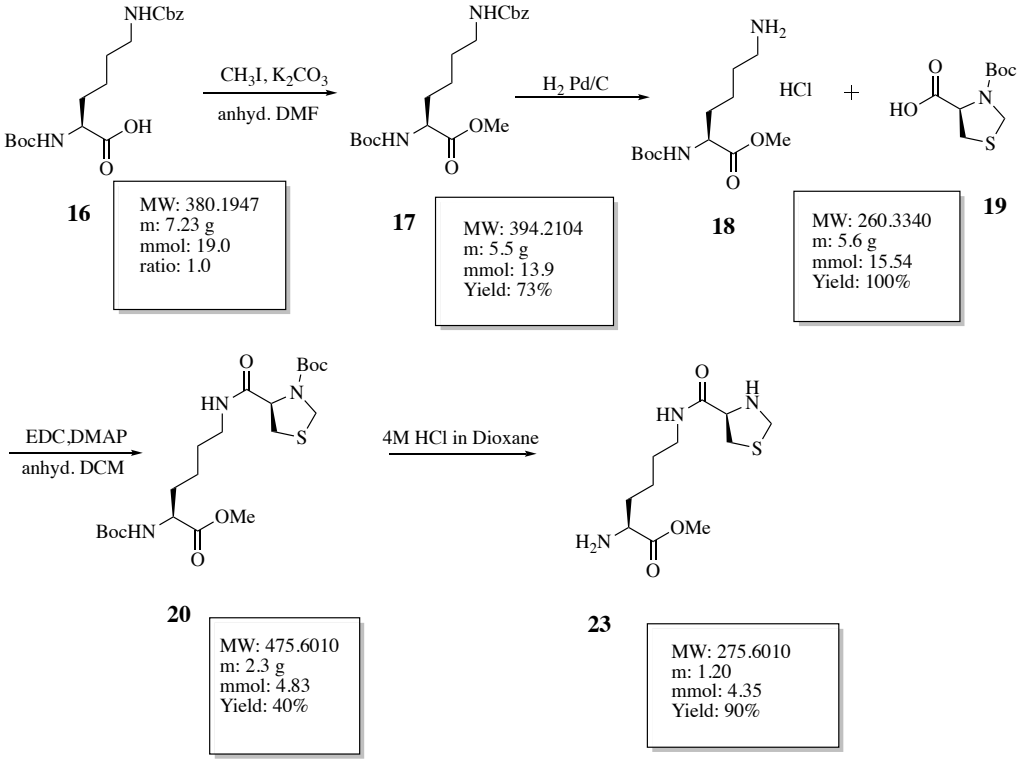


Figure 2.4 Yield Analysis for Methylated UAA

The data includes the yield for each compound is shown, which is calculated from the preceding step. The mmol for each compound is determined by the molecular weight and amount produced.

2.4 Conclusions

An efficient method for synthesizing and purifying the UAA was established after numerous trials, which generated a higher yield in product and decrease in unfavorable side products. A convenient and suitable synthesis method is essential for a large-scale production of UAA.

2.5 Methods and Materials

^{1}HNMR was analyzed using Bruker 400 MHz spectrometer in Chloroform-*d* and deuterium oxide.

Compound 17 - Synthesis of Boc-L-Lysine(Cbz)Methylester

BocLys(Cbz)OH (7.23g, 19.0 mmol) was dissolved in DMF (40 ml), followed by addition of potassium carbonate (5.25g, 38.0 mmol). The solution was allowed to cool to 0^{o} C before dropwise addition of iodomethane (1.05 ml, 24.7 mmol). Completion of the reaction was monitored after overnight stirring at room temperature using thin layer chromatography (Hex:EA = 1:1). The reaction mixture was dissolved in ethyl acetate (50 ml) and extracted twice using water (100 ml) to remove excess DMF. The organic layer was then extracted using brine (100 ml) and the combined organic layers were concentrated to dryness and purified using column chromatography (hexane: ethyl acetate = 1:1).

Compound 18 - Synthesis of Boc-L-Lysine Methylester

Boc-L-Lysine(Cbz)Methylester (1.0 eq) was dissolved in a mixture of ethyl acetate (40 ml) and ethanol (10 ml). Palladium (10%, 0.39 mmol) was added to the stirred solution and hydrogen was bubbled through the solution three consecutive times to promote efficient reduction. Completion of the reaction was monitored using TLC analysis (Hex:EA = 1:1). A

final ninhydrin stain (C:M = 9:1) confirmed reduction of the carboxybenzyl group to a free amine. The palladium was removed by filtration through a pad of celite and the product was concentrated to dryness.

Compound 20 - Synthesis of N^{ε}-NBoc-L-thiazolidine-Methylester

N-boc-L- thiazolidine-4-carboxylicacid (1.0 eq.) was dissolved in dichloromethane (40ml). Next, DMAP (0.5 eq.) was added followed by subsequent addition of EDC (1.2 eq.). After an hour of stirring at 0^{o} C, Boc-L-Lysine Methylester (previously dissolved in 5 mL DCM) was added dropwise. Completion of the reaction was monitored using TLC analysis (Hex:EA = 1:1, iodine stain). The mixture was diluted with dichloromethane (50 ml) and extracted with 1M HCl (50 ml) and brine (50 ml). The organic fractions were dried with magnesium sulfate and concentrated to dryness to give the crude product. The product was purified using column chromatography hexane: ethyl acetate = 3:1 in which mass spec and NMR analysis helped in identifying of the compound.

Compound 21 - Synthesis of Nε-(N-boc-thiazolidine)

Nε-(N-boc-L-thiazolidine)]-O-methylester (1 eq.) was dissolved in THF:H_2O (3:1, 36 mL), followed by addition of $LiOH \cdot H_2O$ (2 eq.). The reaction was stirred at room temperature overnight and the reaction was monitored using TLC analysis (hexane: ethyl acetate = 1:1, iodine stain). The reaction mixture was diluted with ethyl acetate (30 ml) and water (30 ml) and the aqueous layer with acidified using a 5% citric acid solution, (pH ~ 4.0). The aqueous layer was then extracted twice with ethyl acetate (30 ml) and then with brine (30 ml). The combined layers were dried, filtered, and concentrated to dryness.

Compound 22 - Synthesis of N^{ε}-L-Thiaprolyl-L-lysine

[Nε-(N-boc-thiazolidine)] (1 eq.) was dissolved in dichloromethane (10 mL) and cooled to 0^{O} C. Afterwards, trifluoroacetic acid (5 mL) was added dropwise and the reaction was stirred overnight. The oil product was dissolved in H_2O (20 mL) and extracted twice with diethyl ether (20 mL). The recovered aqueous layer was concentrated to dryness under reduced pressure, to yield L-lysine(Nε-thiazolidine) (N^{ε}-L-Thiaprolyl-L-lysine). NMR and mass spec analysis confirmed synthesis of the product.

Compound 23 - Synthesis of N^{ε}-L-Thiaprolyl-L-lysineOMe

[Nε-(N-boc-thiazolidine)] (1 eq.) was dissolved in 1,2-dioxane and stirred at room temperature. 4M hydrogen chloride in dioxane was added dropwise and the reaction was allowed to stir overnight, which resulted in formation of a white precipitate. Excess solvent was removed through evaporation and the crude was further purified using ether precipitation (3700 rpm, 10 min, 4^{o} C). The product was allowed to dry and NMR analysis granted successful synthesis of the product.

3 INCORPORATION OF THZK INTO UBIQUITIN

3.1 Expression of UbK11ThzK via Incorporation of UAA

Both synthesized unnatural amino acids (methylated and non-methylated forms) were incorporated site-specifically into ubiquitin at the K11 position. Formation of the peptides were confirmed by MALDI.

3.1.1 Selection Markers

Selectable markers are utilized to identify cells containing a specific trait. Introducing a selectable marker allows certain colonies to uptake a plasmid containing the gene of choice. Most of these markers promote resistance to specific antibiotics and colonies that successfully take up the plasmid will most likely develop resistance against a particular antibiotic. In the expression of UbK11ThzK, plasmids encoding the vectors pMyo-Ubk11TAG-Tet and pBk-ThzPylr-Amp contained tetracycline and ampicillin resistant genes, respectively.[21]

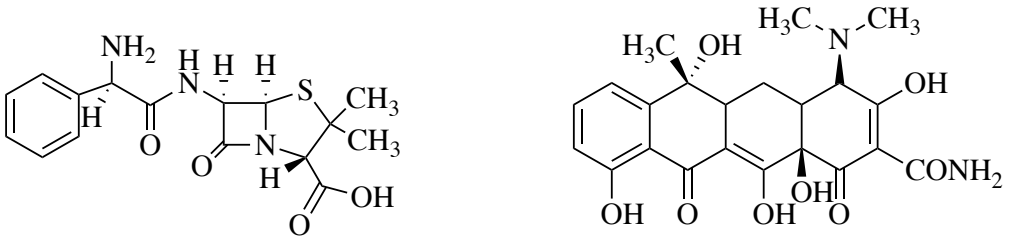


Figure 3.1 Structures of Ampicillin and Tetracycline

The antibiotics are used as selectable markers for gene resistance in protein expressions. Introduction of the antibiotic gene into the vector is achieved by recombination techniques, which is then transformed into cells. Replicating cells containing recombinant DNA molecules develop resistance against these antibiotics.

3.1.2 Translation

Incorporation of UAAs can be utilized to create proteins with unique functions. We have shown successful selective incorporation for both methylated and non-methylated unnatural amino acids into ubiquitin at the K11 position, although the yield reported for the methylated form shows higher selectivity. Site-specifically introducing an UAA requires construction of an orthogonal aaRS-tRNA pair that uses an amber suppressor codon to incorporate an UAA into the protein of interest. The amber codon triplet in DNA is TAG and in mRNA UAG. Instead of terminating protein translation, use of amber codon suppression incorporates the UAA as a coding codon in translation. Therefore, evolution of an aaRS-tRNA pair is necessary so that it no longer recognizes natural amino acids, only the UAA of interest. Here, our engineered aaRS, Pyrrolysyl tRNA synthetase, is developed through cloning techniques and encoded in the vector pBk-ThzPylr-Amp. The cognate tRNA can be aminoacylated by the aaRS and charged with our synthesized UAA to be incorporated into ubiquitin at the K11 position through amber suppression. The amber codon triplet is encoded in the vector pMyo-Ubk11TAG-Tet.[22]

Incorporation of the UAA involves aminoacylation of $tRNA^{Pyl}$ with the cognate PylRS. The tRNA is charged with the amino acid, which is incorporated into the protein sequence site-specifically through recognition of an amber stop codon.[21,22]

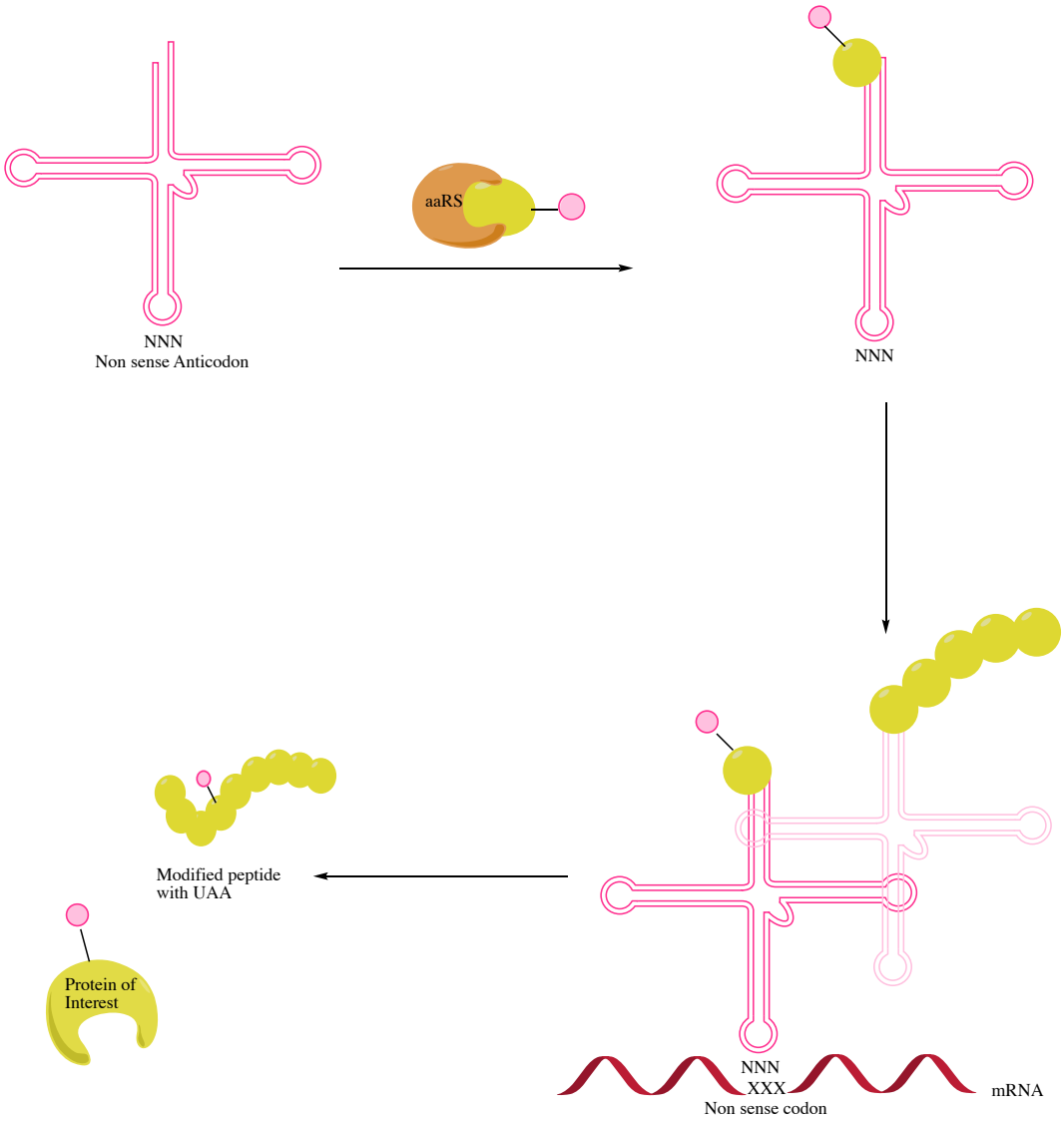


Figure 3.2 Nonsense Amber Suppression

Occurs when a stop codon, such as UAG, is decoded as an amino acid instead of terminating protein synthesis. This allows incorporation of an UAA into the peptide sequence for a given protein. The tRNA bearing the anticodon sequence is aminoacylated by its cognate aaRS and charged with the appropriate amino acid. The mRNA codon is matched with the anticodon of the tRNA carrying the UAA, which facilitates the incorporation of a UAA at a specific site in

a peptide sequence. Incorporation of the UAA site-specifically yields the protein of interest, which can be isolated through purification techniques.[21]

3.1.3 Incorporation of Methylated ThzK

Our results show successful incorporation of the methylated UAA into ubiquitin at the K11 position. Here, we focus on utilizing cultures inoculated with DH10B *E. coli* stock solutions containing the recombinant plasmids as indicated.

The following expression is presented to show how the yield for each expression is calculated. Following multiple successive lysis and wash buffers during purification of the expressed protein and is eluted using 5 mL intervals. The absorbance for each elution is measured at OD_{595}. Appropriate calculations are determined at an absorbance above 0.100, as an absorbance under this value reciprocate insignificant yield values. The total yield is calculated based on each elution measured. Incorporation of the synthesized UAA is shown in Figure

Elution (5 ml)	UbK11ThzKOMe-4	
1	OD_{595} = 0.596	= mg
2	OD_{595} = 0.361	= mg
3	OD_{595} = 0.138	= mg
4	OD_{595} = 0.122	= mg
5	OD_{595} = 0.104	= mg
6	OD_{595} = 0.096	insignificant
		Total = 66 mg

Figure 3.3 Yield for Incorporation of Methylated UAA

$$\frac{OD_{595} + 0.0245}{0.03945} = X \text{ mg/ml}$$

2.4 ul = mg/ml ÷ 3

4.8 ul = mg/ml ÷ 6

8.0 ul = mg/ml ÷ 10

Figure 3.4 Calculation from the Standard Curve for Yield Calculation

The concentration of the sample is monitored at OD_{595} and utilized in the equation above. Note that the equation produces values in mg/ml and the overall yield is determined depending on the total volume of elution buffer utilized.

Our yield calculations are defined from a standard graph plotting absorbance versus concentration, where the linear equation of increasing absorbance in relation to concentration is generated. The equation can be arranged to allow calculation of the yield utilizing the absorbance of the elution. According to Beer's law, absorbance of a sample is proportional to the concentration. Therefore, the absorbance is reflective of the concentration at any point in the curve, where decreasing concentrations correspond to a lower absorbance and increasing concentrations have a higher absorbance. The optical density of any region should be proportional to the surface protein concentration as the concentration approaches 1. Therefore, appropriate yields are calculated at a concentration at or above 0.100, but no larger than 1.

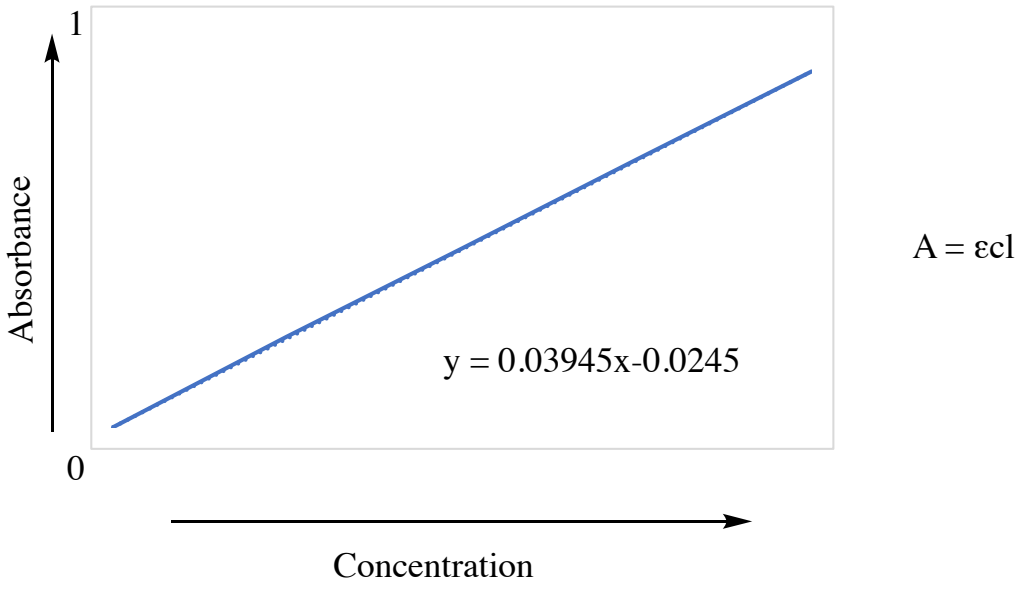


Figure 3.5 Linear Equation Developed from Molar Concentration in Relation to Absorbance

The linear equation can be arranged and applied to calculate the yield for an entire protein expression using the absorbance of the sample at a given OD. The linear equation was obtained from measuring the absorbance for a series of protein expressions and calculating the concentration utilizing Beer's law.

Incorporation of the methylated UAA showed higher yields compared to the non-methylated form, which indicates improved target affinity and membrane permeability for the methylated UAA. Yields for the methylated UAA were favorable. Proteins bound to the Ni^{2+} resin were purified with lysis and two successive portions of wash buffers. The proteins were then eluted with six 5mL portions of elution buffer, notated as E_1 through E_6. The first elution band (E_1) shows the overexpressed protein of interest at 11kDa. Impurities are few, compared to the incorporation of non-methylated UAA. The results are obtained from a 2L culture.

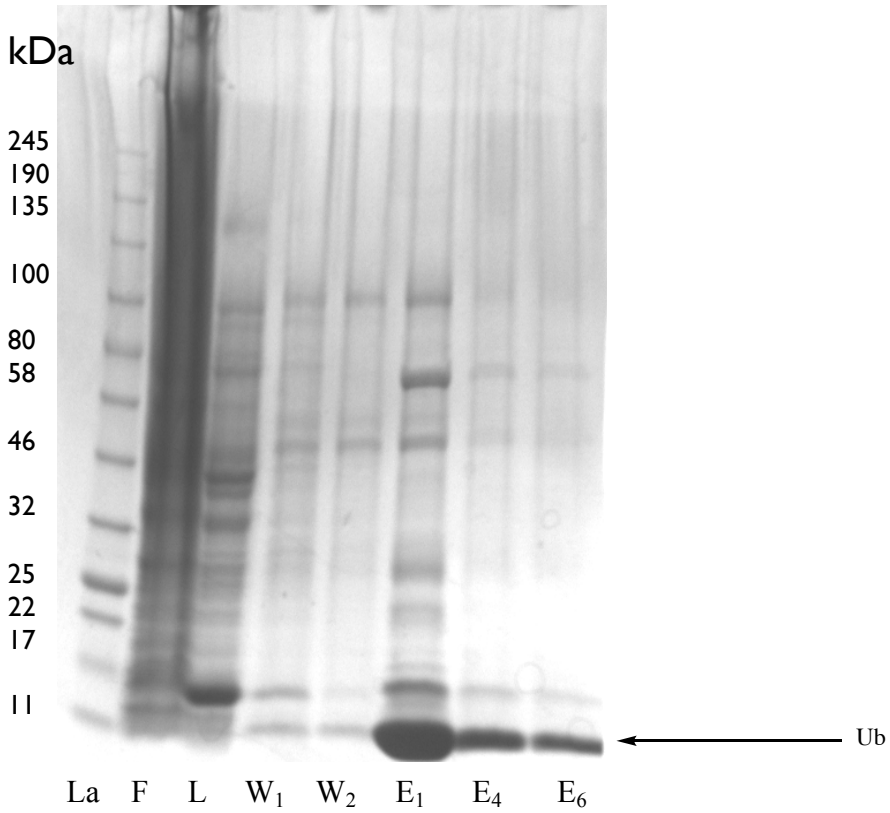


Figure 3.6 Expressed Protein from Incorporation of Methylated UAA

Proteins tagged with an HA hemagglutinin tag can be easily identified in Western blot techniques. This antibody is used to detect HA tag of target protein expressed in DH10B cells. Cell lysates were loaded onto a 4-20% Tris-HCl polyacrylamide gel. Proteins were separated using SDS-PAGE, transferred to a PVDF membrane, and blocked with 5% nonfat milk diluted in 0.1% TBS. The membrane was probed with the HA tag at a dilution of 1:500 overnight at 4^{o} C, and then probed with a secondary antibody at a dilution of 1:10,000 for an hour at room temperature. The membrane was washed with a 0.1% TBS buffer before and after addition of the secondary antibody. Lysates containing a multi-epitope tagged protein are visualized in Figure 3.7, although the band of interest is strongly exhibited at 11kDa.

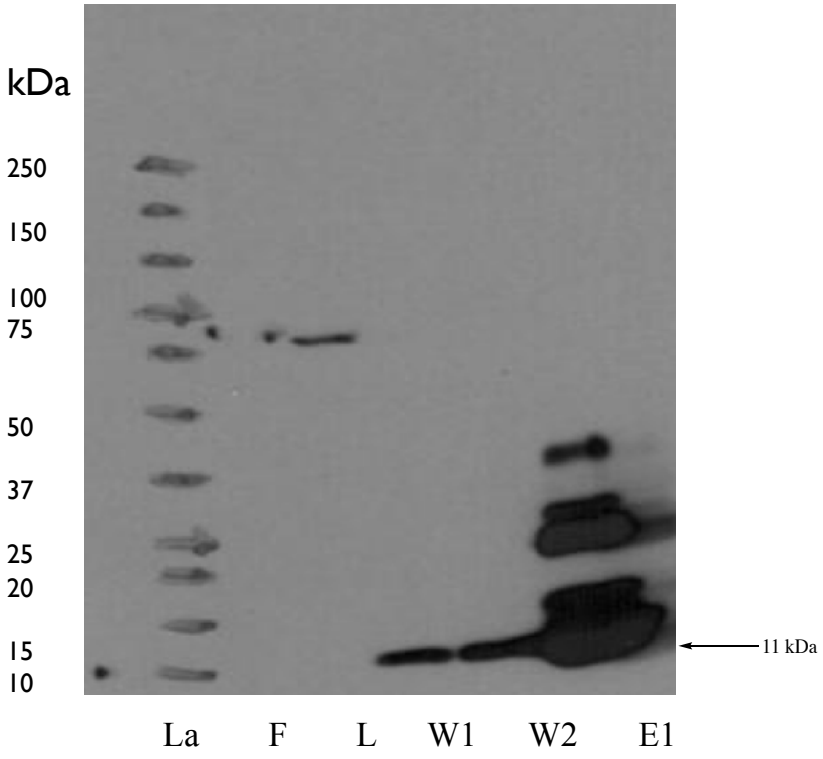
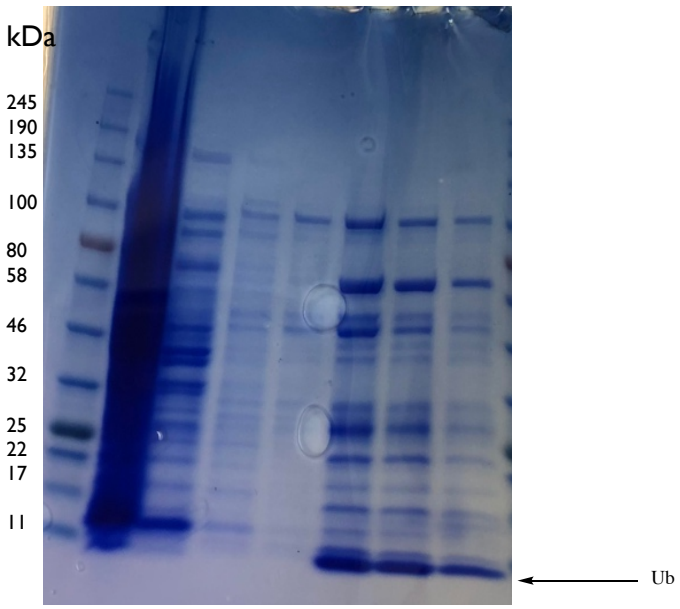


Figure 3.7 Western Blot Analysis on Incorporated UAA

3.1.4 Incorporation of Non-Methylated ThzK

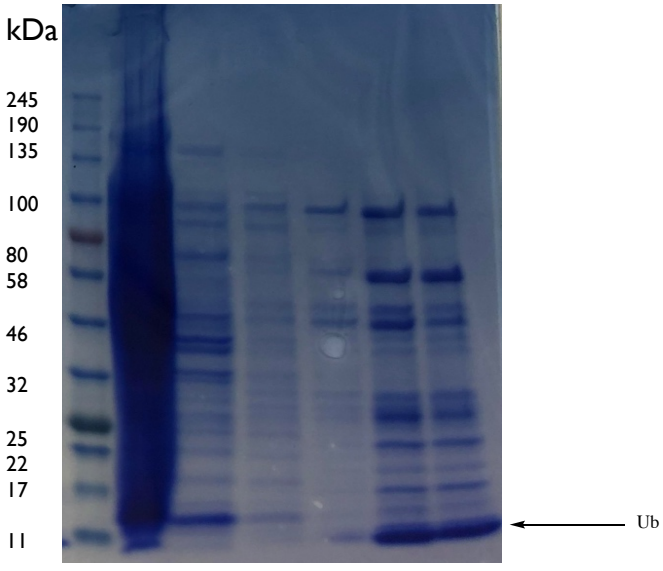
Non-methylated UAAs were incorporated into ubiquitin at the K11 position, as indicated by coomassie staining. Proteins bound to a Ni^{2+} resin were purified with lysis and two successive portions of wash buffers. The proteins were then eluted with 1mL portions of elution buffer, notated as E_1 through E_3. Results indicate three attempts for incorporation of the UAA for a 500 ml culture. The yield for each trial averages to about 7 mg from a 500 ml culture. Elution bands show small amount of expressed protein at 11kD, and multiple impurities, suggesting low incorporation and selection for the synthesized UAA.

A



Yield = 7.065 mg

B



Yield = 5.795 mg

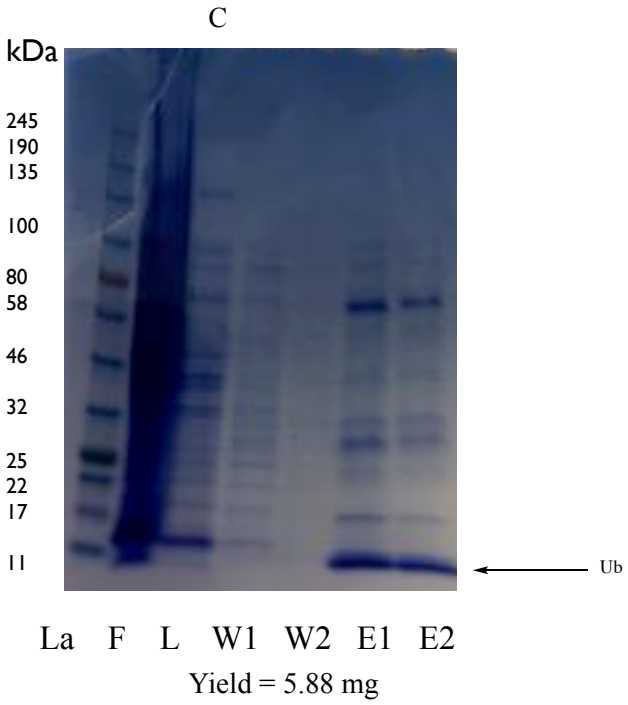
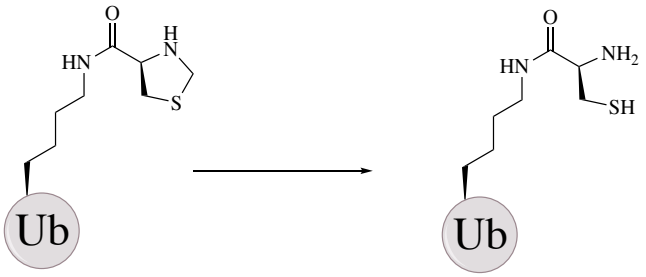


Figure 3.8 Incorporation of Non-Methylated UAA

Three separate trials for incorporation of non-methylated UAA were performed. Coomassie staining indicates low affinity and permeability for non-methylated UAA. The calculated yield for each expression averages around 6mg for a 500 mL culture. Multiple impurity bands are also visualized in the elution lanes, suggesting decreased incorporation of the UAA. Incorporation of methylated UAA is believed to show better incorporation in cells as it is less charged compared to the non-methylated UAA, which could possibly facilitate better delivery into cells. Cells also contain an esterase enzyme to easily cleave the methoxy group.

3.2 Deprotection of UbK11ThzK

The incorporated UAA undergoes de-protection to generate a nucleophilic N-terminal cysteine residue utilizing a *O*-methylhydroxyamine in a 6M solution of guanidium chloride. Generation of the cysteine residue permits ligation of the acceptor Ub to a donor Ub via native chemical ligation, creating our di-Ub probe **3**, essential for conjugation to an E2 enzyme. Deprotection of ThzK to CysK is confirmed by MALDI analysis.



8 **9**

Figure 3.9 Deprotection of ThzK to CysK

Generation of a cysteine residue that is capable of nucleophilic attack on a donor ubiquitin molecule. The acceptor molecule becomes ligated to the donor ubiquitin, facilitating production of a di-Ub conjugate.

3.3 Conclusions

Incorporation of the methylated UAA into ubiquitin at the K11 position is more sufficient compared to the non-methylated UAA, as exhibited in coomassie and MALDI analysis. Higher yields and fewer impurities are obtained.

3.4 Methods and Materials

Expression of UbK11ThzK

A starting culture was prepared with 5 ml of 2XYT media containing tetracyclin and ampicillin (1:1000 dilution). The media was inoculated with DH10B *E. coli* cells to yield tetracyclin and ampicillin resistant cells, encoding vectors pMyo-UbK11TAG-Tet and pBk-ThzPyls-Amp. After overnight growth (37^{o} C, 280 rpm), 1 ml of the starting culture was inoculated into four 500 mL (2L) flasks containing sterile media. To this, the antibiotics (1:1000 fold) were added and the cultures were incubated (37^{o} C, 280 rpm) until OD_{600} = 0.6 – 0.8. The previously synthesized amino acid Nε-L-Thiaprolyl-L-lysine (2 mM, 1.42g) was dissolved in H_2O and NaOH (3 ml, 1 ml, pH 8.0) and added to the cultures. After 30 minutes, the cells were induced with Isopropyl D-1-thiogalactopyranoside and 20% arabinose (1:1000 dilution) and incubated overnight (30^{o} C, 280 rpm). The cells were harvested by centrifugation at rcf 8000 x gravity (4 ^{o}C, 30 min) and then re-suspended in lysis buffer. Lysosome (10 ml/L), magnesium chloride (10 uL/L), calcium chloride (10 uL/L), phenylmethylsulfonyl fluoride (200 uL/L) and dithiothreitol (20 uL/L) were added to promote cell lysis. The cells were sonicated (19%, 30 min., 10s impulses) and centrifuged (12000 rpm, 30 min, 4 ^{o}C), in which the supernatant was collected and incubated overnight in a prepared Ni^{2+} resin (4 ^{o}C). The protein was purified with portions of lysis, wash, and elution buffers, respectively and confirmed by MALDI analysis. The protein was dialyzed (1L, 50 mM TRIS, 50 mM NaCl, 1 mM DTT, pH 8.0). After dialysis, the protein was lyophilized, deprotected (5h, 37^{o}, C), dialyzed again, and purified via HPLC.

Deprotection of ThzK

UbK11ThzK was dissolved in 1 mL 6M solution of guanidium chloride in 100 mM phosphate buffer, pH 8.0. *O*-methylhydroxyamine hydrochloride was added to the solution at a

final concentration of 200 mM. The pH was adjusted to 4.0 and the solution was incubated at RT for 4h. The protein was dialyzed into PBS buffer containing 1 mM DTT and then into 20 mM ammonium bicarbonate containing 1 mM DTT.

4 SYNTHESIS OF UBK11CYSUB

This chapter will focus on synthesis of the di-Ub probe, which involves expression of the donor Ub, UBSR, and conjugation to the acceptor Ub, CysK. The two ubiquitin molecules are ligated and undergo S-N acyl transfer. Ligation of the two deprotected peptides was confirmed with MALDI/MS.

4.1 Linkage Specific Di-Ub Synthesis through Native Chemical Ligation

Synthesis of **3** involves expression of a donor Ub bound to intein tag for affinity purification of the fusion protein on chitin resin. Addition of a thiol compound results in cleavage of the donor Ub from the intein tag, which is purified via affinity chromatography. The donor Ub can then be ligated to our acceptor Ub via native chemical ligation to yield **15**, which undergoes S to N acyl transfer giving the resulting product **3.**

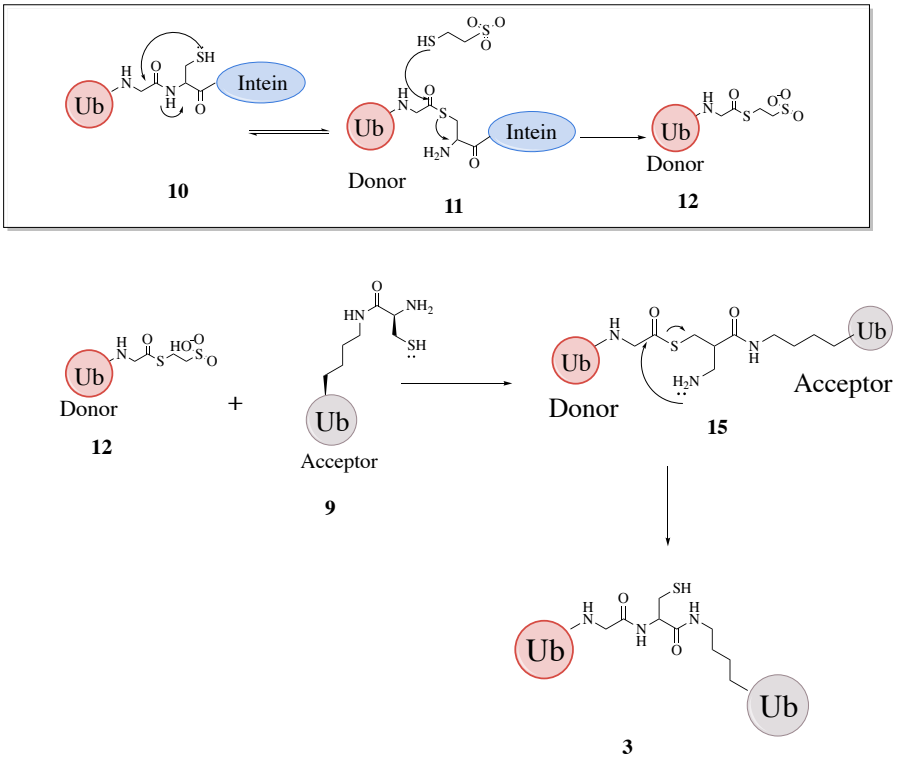


Figure 4.1 Intein Mediated Thiol Cleavage

4.2 Expression of UBSR

Early engineered inteins have allowed the development of self-cleaving affinity tags and methods for combining protein segments through protein ligation. UBSR expression involves a modified intein that is expressed in fusion to an affinity tag. Cleavage of the intein and tag from the target protein occurs during induction of the intein and after the fusion if affinity-purified. The cleavage activity of inteins are triggered by the addition of thiol compounds. The addition of MESNA via thiol mediated cleavage separates our target protein, UBSR, from the intein-tag complex and generates a good sulfonate leaving group upon addition of an acceptor peptide during chemical ligation.[8]

During the purification of UBSR via affinity column, the intein and affinity tag binds to a chitin resin. The resin is washed with a wash buffer containing 1 mM EDTA and TCEP to elute undesirable proteins. Elution buffer containing 200mM MESNA is added to the resin and capped overnight at room temperature to allow sufficient thiol mediated cleavage of the target protein. The protein is eluted, leaving behind the intein and afftinity tag. The donor Ub can then undergo chemical ligation with an acceptor Ub to generate a di-Ub conjugate.

The donor Ub can be ligated to an acceptor Ub only after successful cleavage from the intein. Before cleavage, the protein-intein complex undergoes N-S acyl transfer to generate a thioester product. Addition of MESNA separates the target protein from the intein, facilitating a thiol mediated cleavage. An acceptor Ub molecule can be ligated to the donor Ub by nucleophilic attack on the C-terminal of the donor Ub. The two ligated peptides undergo another S-N acyl transfer to produce the target di-Ub probe.

The target protein was purified via column purification after cleavage with MESNA. The results are visualized in Figure 4.2. In the fourth and fifth wash lanes, a band near 30 kDa is

shown, representative of fused intein bound UBSR protein obviously not bound to the chitin resin. UBSR, 8.5 kDa, is visualized in the last elution lane. There are small traces of protein at a higher molecular weight, corresponding to fused intein-UBSR and intein only at kDa, respectively. The sixth lane shows a sample taken directly from the chitin resin with the same bands indicated in the elution band. However, the intein band is much more intense compared to the elution band.

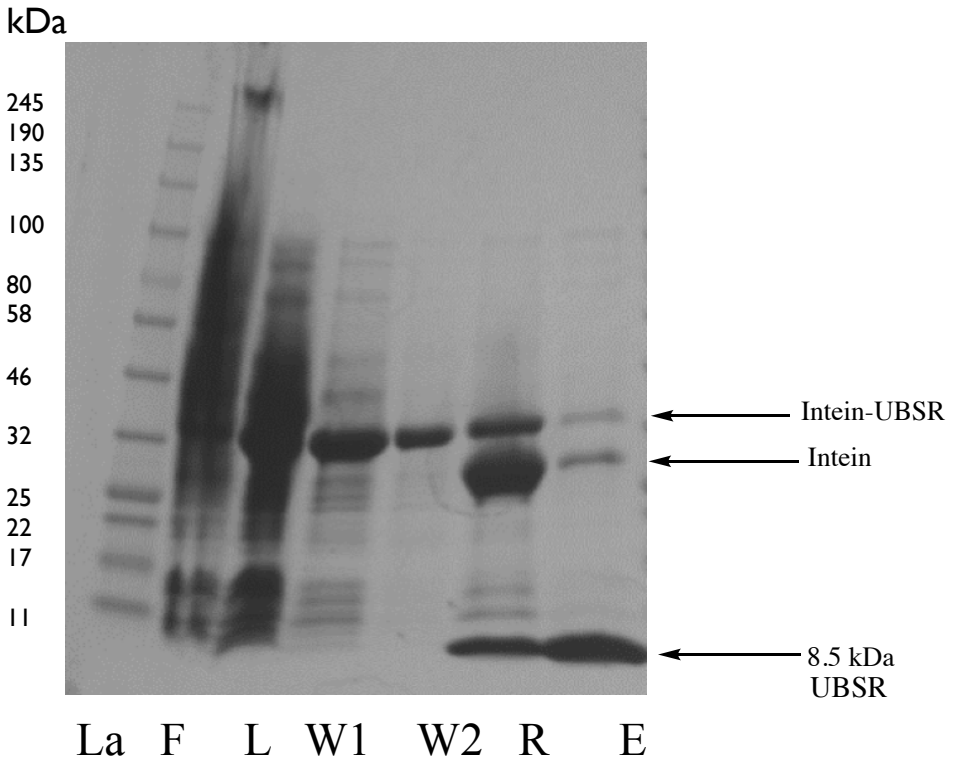


Figure 4.2 Expression of Donor Ubiquitin

Coomassie analysis for expressed donor molecule, visualized at 8.5 kDa. Protein is treated with MESNA (1 mM, overnight, RT) and subjected to purification via Ni^{2+} resin.

4.3 Native Chemical Ligation of UbK11CysUb

Here, we prepare **3**, which we can further conjugate with a specific E2 enzyme to construct our target di-Ub-E2 probe. Formation of our di-Ub probe is discussed.

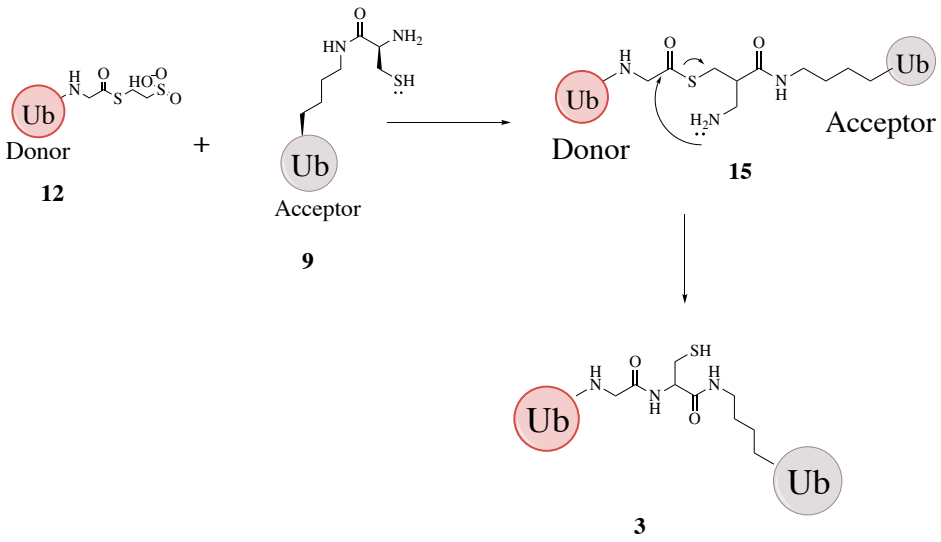


Figure 4.3 Formation of DiUb Probe through Native Chemical Ligation

Both expressed donor and acceptor Ub molecules were successfully ligated using native chemical ligation. Formation of the di-Ub product is monitored at 0, 2, and 4 hours and overnight. The donor molecule is used in excess (1.5 eq.) to promote formation of the di-Ub product, which is visualized by overexpression at 8.5kDa during all time intervals. The acceptor **9** (11 kDa) is diminished over the time period a, while formation of **3** (19 kDa) increases. acceptor and di-Ub product itself are similar in structure and each carry a bound His tag. The donor Ub is easily purified using affinity purification considering the lack of a bound His tag. Purification of **12** is shown in the flowthrough and lysis lanes in Figure. It is rather challenging to separate compounds **9** and **3** via affinity purification alone considering bound His tag. The elution lanes contain overexpression of **3**, although some unconverted acceptor Ub is visiualized.

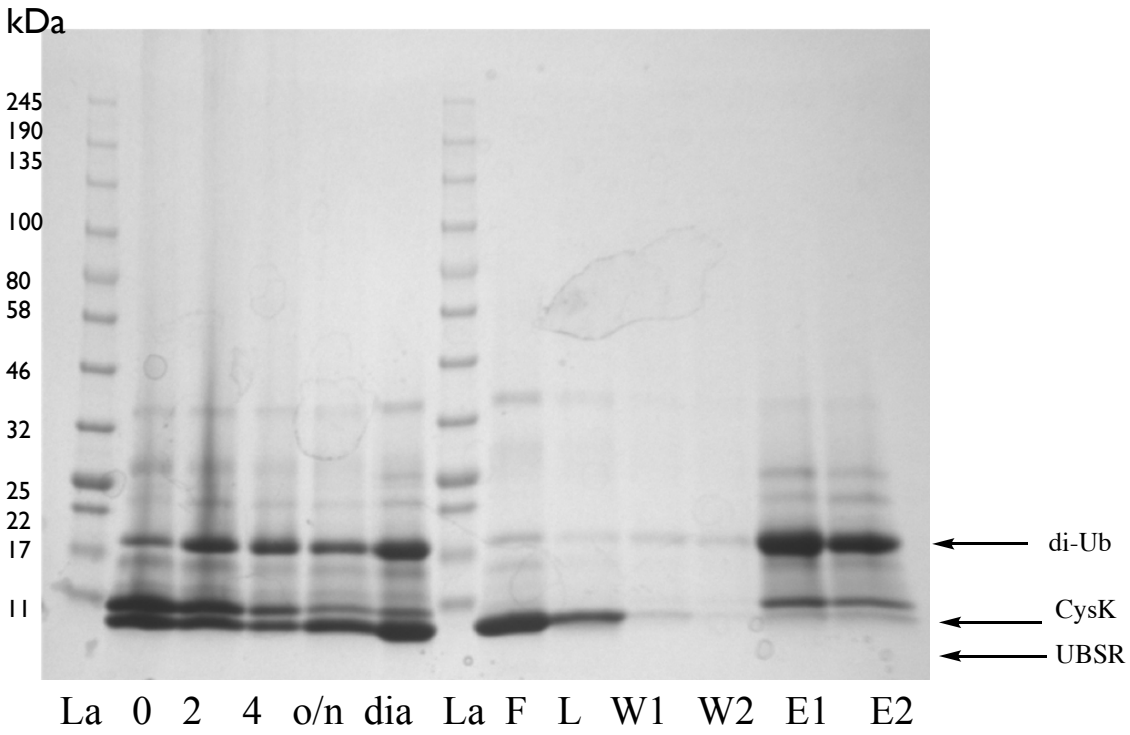


Figure 4.4 Expression of UbK11CysUb

The donor Ub molecule (8.5 kDa) is added in excess to ensure most of reactant **9** is converted into product **3**. Formation of the di-Ub conjugate is monitored over a period of 0, 2, and 4 hours as well as an overnight reaction. Depletion of reactant **9** is visualized over the time intervals as formation of **3** increases. The acceptor and di-Ub conjugate both contain a His tag, while the donor lacks this tag. Therefore, unutilized donor **12** can be purified via column chromatography rather easily. Overexpression of **3** is visualized in the two elution bands, although a small trace amount of **9** is present to difficulty in separation of the two compounds via affinity purification.

4.4 Conclusions

The Di-Ub conjugate **3** was successfully synthesized after expression and purification of the donor Ub molecule **12**, following native chemical ligation to the acceptor Ub **9**.

4.5 Methods and Materials

Expression of UBSR

A 5 ml starting culture was prepared with 2XYT broth and inoculated with ampicillin (1:1000 fold) and BL21 cells containing the vector pBk-ThzPyls-Amp. 1 ml of the starting culture was inoculated into four 1L flasks (4L total), which was incubated (37^{o} C, 280 rpm) until OD_{600} = 0.8. The cells were induced with IPTG (1:1000) and after 30 minutes transferred to the refrigerated shaker overnight. Harvesting of the cells required centrifugation at rcf 8000 x gravity (4 ^{o} C, 30 min) and resuspension in column buffer (20 mM Na_2HPO_4, 200 mM NaCl, 1 mM EDTA, 1 mM TCEP, 20 uM PMSF) for an hour. After resuspension, the cells were sonicated for 30 minutes (19%, 10s impulses) and spun down via centrifugation (12000 rpm, 30 min, 4 ^{o} C). The protein was purified using Int^{c} resin with two successive portions of wash buffer: wash I (1mM EDTA, 1 mM TCEP, pH 7.2) wash II (1mM EDTA, 10 mM TCEP, pH 7.2), respectively. Elution buffer (1mM EDTA, 10 mM TCEP, 200 mM MesNa) was added to the column and capped for overnight incubation at room temperature, which was eluted the following day.

Ligation – UbK11CysUb

CysK (11.5kDa, 340 nmol, 3.92 mg) was dissolved in 200 ul of ligation buffer (recipe) and the absorbance was monitored at OD_{595}. Purified UBSR (8.5kDa, 1367 nmol,11.62 mg) measured at OD_{595} was allowed to react (in 1.5 equivalence) with the acceptor ubiquitin overnight under inert conditions. Fractions taken at 0, 2, 4, and overnight intervals (5 ul dissolved in 80 ul H_2O) were collected to be analyzed for coomassie and western blot analysis. The protein was dialyzed (100 ml 10x phosphate, 900 ml H_2O, 1mM DTT, pH 8.0) and allowed to incubate on Ni^{2+} overnight at 4^{o} C. The protein was purified with portions of lysis, wash

(5ml), and elution buffer (1.5ml) and the elution fractions containing an absorbance > 0.1 at OD_{595} were dialyzed even further.

5 PREPARATION OF DIUB-E2 CONJUGATE

This section covers the expression of an E2 enzyme, Ube2s, which is further conjugated with our di-Ub probe to better understand how ubiquitin chains are formed through E2 catalyzed reactions.

5.1 Expression of Ube2s

The E2 enzyme Ube2s is bound to a SUMO-His tag which is cleaved with addition of Ulp1 to generate the target E2 protein. Catalytic activity of Ulp1 involves hydrolysis of the alpha-linked peptide bond in the sequence Gly-Gly-Ala-Thr-Tyr at the C-terminal end of small ubiquitin like modifier protein (SUMO) and cleaves between the second Gly and Ala residues. The protease is highly specific for SUMO protein fusion and recognizes the tertiary structure of SUMO rather than amino acid sequence.[17,27]

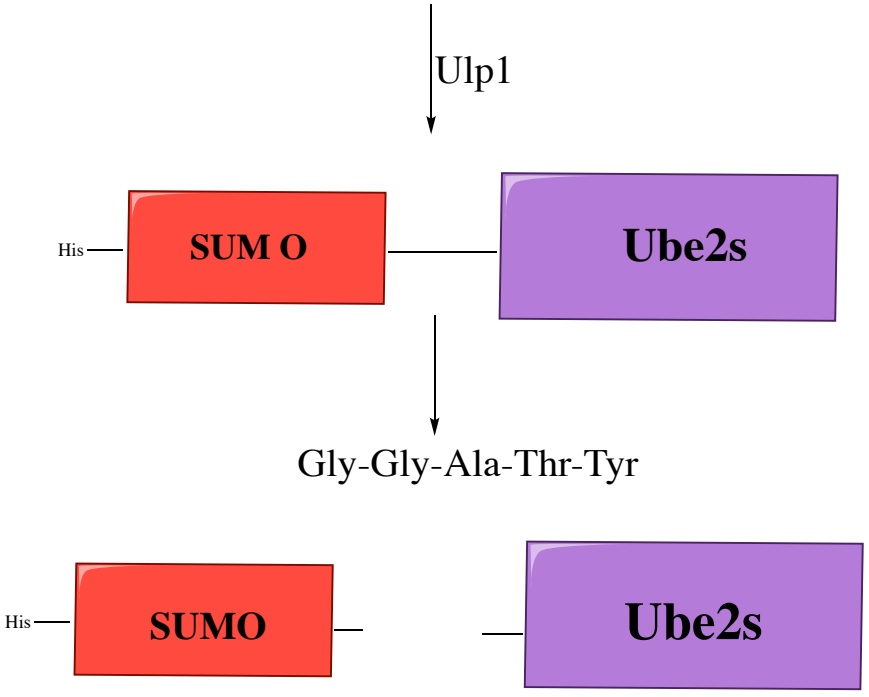


Figure 5.1 Cleavage of Ube2s by Addition of Ulp1

The expressed protein was purified via affinity column using Ni^{2+} resin. Eluted samples containing Ube2s with bound SUMO peptide (E_1 to E_6) were subjected to dialysis (1x TBS, 1 mM DTT) twice and treated with Ulp1 (2 mL). Purification of the target protein (17 kDa) was achieved through additional affinity chromatography leaving the SUMO-His bound to Ni^{2+} resin. The expressed cleaved protein is visualized in elutions 7 through 10 (E_1 – E_{10}), although high yield of the protein was obtained in elution 9.

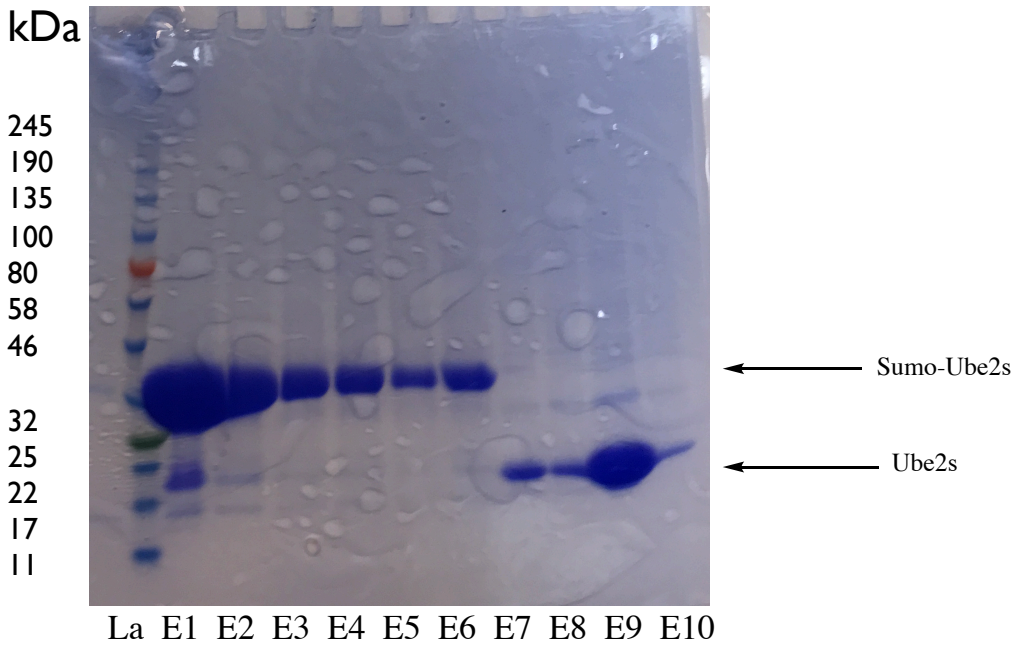


Figure 5.2 Expression of Ube2s

Coomassie Stain showing expressed protein before and after addition of Ulp1. Before addition of the protease, the protein is purified via affinity purification using Ni^{2+} resin and eluted as shown from E1 to E6. The eluted proteins are treated with a 2 mL solution of Ulp1 and purified further via Ni^{2+} resin in which the SUMO tag has high affinity for the column. The target protein is visualized at kDa 17.

The addition of di-Ub probes to covalently trap enzymes could address structural and biochemical studies of Ub linkage specificity. Current work is centered on conjugating our expressed E2 protein **24** with the synthesized di-Ub probe **3** through disulfide bond formation in hopes of retrieving a clear crystal structure through X-ray crystallography of the intact di-Ub-E2 conjugate. The ability of E2-Ub conjugates to transfer Ub can be explored by assessing their reactivity towards nucleophiles such as free lysine. Ubiquitin associates with a cysteine in the active site of an E2 through its C-terminal end. During Ub chain synthesis catalytic activity of the E2 enzyme requires proper conformation and orientation to allow nucleophilic attack and subsequent linkage of the acceptor Ub to an existing chain. In situ, the point at which an acceptor Ub molecule becomes covalently linked to a donor Ub within the active site of the E2 is transient and rather difficult to map. Therefore, construction and conjugation of a stable di-ub probe to an E2 enzyme provides leeway for addressing areas of ubiquitin elongation involving multiple protein interactions. The di-Ub-E2 probe can be exposed to a multitude of substrates, deubiquitinating enzymes (DUBs), and other cascade proteins to better understand mechanisms of ubiquitin transfer.

Construction of the di-Ub-E2 conjugate **25** through disulfide bridge linkage is derived by coupling through oxidation of two thiol groups. Formation of a disulfide bond provides stability of a protein in its respective conformation and plays an important role in folding and stability. Another approach for synthesis of of the di-Ub-E2 conjugate is through elimination of cysteine to dehydroalanine (Dha), which is a useful precursor to a range of post translational modifications of proteins by conjugation of a thiol. Our E2 can then be conjugated with the reduced product via Michael addition reaction.[28]

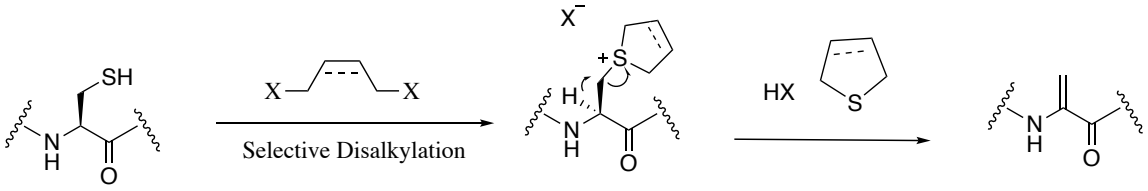


Figure 5.3 Bis-Alkylation of Cysteine to Dha

The cysteine of the di-Ub probe **3** undergoes an elimination reaction, similar to Figure 5.3, with α, α'– dibromoadipic-bis-amide (DBAA), which activates the cysteine residue for elimination under basic conditions. The product can then be coupled with an E2 enzyme forming product **26** in Figure 5.4.[1]

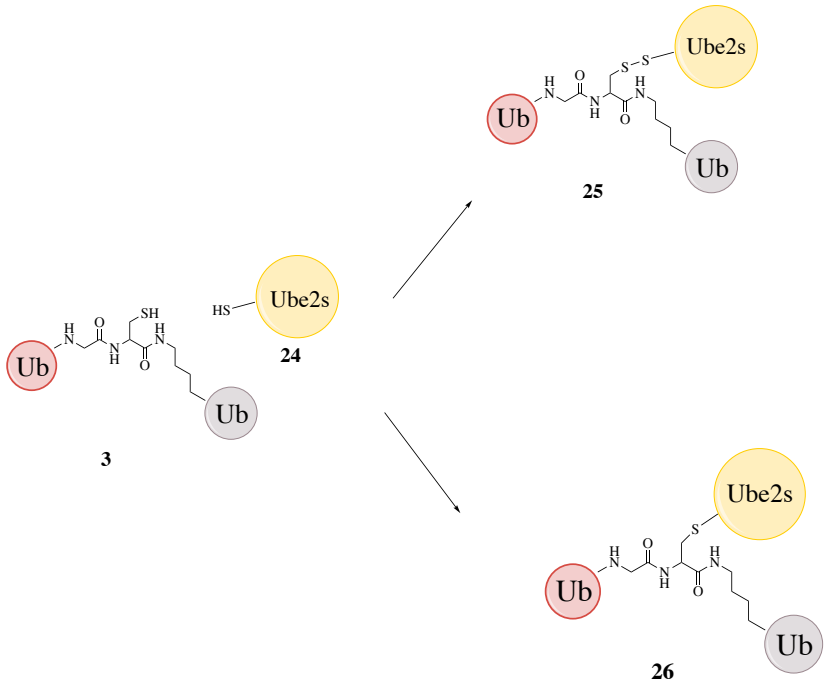


Figure 5.4 Formation of DiUb-E2 Conjugate via Disulfide and Michael Addition Reactions

5.2 Conclusions

Our Di-Ub-E2 probe is derived from conjugating a specific E2 enzyme with our synthesized Di-Ub. Construction of the Di-Ub involves chemically modifying a natural amino acid and site specifically incorporating it into the K11 position of ubiquitin with the use of an orthogonal tRNA/tRNA synthetase pair. Once incorporated, our acceptor Ub is de-protected to generate a reactive cysteine species which undergoes native chemical ligation with an expressed donor Ub, facilitating production of the Di-Ub molecule.

The stable di-Ub-E2 conjugate can be utilized to mimic the transition state during linkage of an acceptor Ub molecule to a donor Ub catalyzed by an E2 enzyme. Here we hope to uncover a three-dimensional structure of the conjugate to expand understanding on the mechanisms behind ubiquitin chain synthesis. Current work involves expressing proteins for the conjugated diUb-E2 probe.

5.3 Methods and Materials

Expression of Ube2s

A 20 ml starting culture was prepared with 2XYT broth and inoculated with kanamycin (1:1000 fold) into a BL21 cell line. The starting culture was equally inoculated into four 1L flasks (4L), which was incubated (37^{o} C, 280 rpm) until OD_{600} = 0.8. The cells were induced with IPTG (1:1000) and after 30 minutes transferred to the refrigerated shaker (18^{o} C) overnight. Harvesting of the cells required centrifugation at rcf 8000 x gravity (4^{o} C, 30 min) and resuspension in lysis buffer for an hour. After resuspension, the cells were sonicated for 30 minutes (19%, 10s impulses) and spun down via centrifugation (12000 rpm, 30 min, 4 ^{o}C). The protein was purified using Ni^{2+} resin with lysis and two successive portions of wash buffer,

respectively. The eluted proteins were dialyzed (50 mM TRIS, 50 mM NaCl, 1 mM DTT, pH = 8.0) twice and treated with a solution (2 mL/mg) of Ulp1. Proteins were further purified using affinity purification with Ni^{2+} resin to facilitate removal of SUMO tag.

REFERENCES

1. Jongkees, S.; Umemoto, S.; Suga, H. Linker-Free Incorporation of Carbohydrates into *in vitro* displayed macrocyclic peptides. *Chem Sci.* **2017**,8 (2), 1474-1481.
2. Stewart, M.; Ritterhoff, T.; Klevit, R.; Brzovic, P. E2 Enzymes: More than Just Middle Men. *Cell Res.* **2016**, 26 (4), 423-440.
3. Lin, A.;Man H. Ubiquitination of Neurotransmetter Receptors and Postsynaptic Scaffolding Proteins. *Neural Plasticity.* **2013**, (17):432057.
4. Ciechanover, A.; Stanhill, A. The Complexity of Recognition of Ubiquitinated Substrates by the 26 Proteasome. *Elsevier*. **2014**, 86-96.
5. Komander, D. The Emerging Complexity of Protein Ubiquitination. *Biochemical Society Transactions.* **2009**, 37(5), 937-953.
6. Kleiger, G.; Mayor, T. Perilous Journey: A Tour of the Ubiquitin-Proteasome System. *PlumX Metrics*. **2014**, 24(6), 352-359.
7. Mitra, N. Incorporating Unnatural Amino Acids into Recombinant Proteins in Living Cells. *Labome.* **2017**, 3:204.
8. Dawson, PE.; Muirir, C. Synthesis of Proteins by Native Chemical Ligation. *Science*. **1994**, 266(5186) 776-779.
9. Cook, W.; Jeffrey, L.; Carson, M.; Chen, Z.; Pickhart, C. *The Journal of Biological Chemistry.* **1992**, 267(23), 16467-16471.
10. Valimberti, I.; Tiberti, M.; Lambrughi, M. Sarcevic, B.; Papaleo, E. E2 Superfamily of Ubiquitin-Conjugating Enzymes: Constitutively Active or Activated through Phosphorylation in the Catalytic Cleft. *Sci Rep.* **2015**, 5:14848.

11. Metzger, M.; Pruneda, J.; Levit, R.; Weissman, A. RING-Type E3 Ligases: Master Manipulators of E2 Ubiquitin-Conjugating Enzymes and Ubiquitination. *Elsevier.* 1843(1) 47-60.

12. Soss, S.; Yue, Y.; Paganon, S. E2 Conjugating Enzyme Selesctivity and Requirements for Function of the E3 Ubiquitin Ligase CHIP. *Journal of Biological Chemistry.* **2011**, 286, 21277-21286.

13. Kumar, S.; Kao, W.; Howley, P. Physical Interaction between Specific E2 and HECT E3 Enzymes Determines Functional Cooperativity. *Journal of Biological Chemistry.* **1997**, 272, 13548-13554.

14. Schwarz, N.; Nuber, U.; Kaiser, P.; Schneider, R.; Scheddner, M. Cloning of Human Ubiquitin-Conjugating Enzyme UbcH6 and UbcH7 (E2-F1) and Characterization of their Interaction with E6-AP and RSP5. *J. Biol. Chem.* **1996**, 271: 2795-2800.

15. Cole, M.; O'Neil, D.; Markham, A.F.; Robinson, P.A. Fine-Mapping, Genomic Organization, and Transcription Analysis of the Human Ubiquitin-Conjugating Enzyme Gene UBE2L3. *Genomics.* **1998**, 51: 124-127.

16. Park, C.W. Cellular Ubiquitin Pool Dynamics and Homeostasis. *BMB Reports*. **2014**, 47(9) 128.

17. Wijk, S.; Muller, S.; Dikic, I. Shared and Unique Properties of Ubiquitin and SUMO Interaction Networks in DNA Repair. *Genes and Development.* **2011**. 25:1763-1769.

18. Ashida, H.; Kim, M.; Sasakawa, C. Exploitation of the Host Ubiquitin System by Human Bacterial Pathogens. *Nature Reviews Microbiology.* **2014**, 12, 399-413

19. Scheffner, M.; Kumar, S.Mammalian HECT Ubiquitin-Protein Ligases: Biological and Pathophysiological Aspects. *Elsevier.* **2014**, 1843(1) 61-74.

20. Suryadinata, R.; Holien, J.; Yang, G.; Parker, M.; Papaleo, E.; Sarcevic, B. Molecular and Structural Insight into Lysine Selection on Substrate and Ubiquitin Lysine 48 by the Ubiquitin-Conjugating Enzyme Cdc34. *Cell Cycle.* **2013**, 12(11) 1732-1744

21. Chin, J.W.; Deiters, A. Cropp, A.; Mukherji, M.; Anderson, C.; Schultz, P. Adding Amino Acids with Novel Reactivity to the Genetic Code of Saccharomyces Cerevisiae. *J. Am. Chem. Soc.* **2003**, 125(39) 11782-11783.

22. Nguyen, D.; Lusic, H.; Neumann, H.; Kapadnis, P.; Deiters, A.; Chin, J.W. 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.* **2009**, 131(25) 8720-8721.

23. Banka P.A.; Behera AP; Sarkar S.; Datta A.B. RING E3-Catalyzed E2 Self-Ubiquitination Attenuates the Activity of Ube2E Ubiquitin-Conjugating Enzymes. *J Mol Biol.* **2015**; 427:2290–2304.

24. Schumacher F.R.; Wilson G.; Day C.L. The N-Terminal Extension of UBE2E Ubiquitin-Conjugating Enzymes Limits Chain Assembly. *J Mol Biol.* **2013**; 425:4099–4111.

25. Ulrich H.D; Jentsch S. Two RING Finger Proteins Mediate Cooperation Between Ubiquitin-Conjugating Enzymes in DNA Repair. *EMBO J.* **2000**; 19:3388–3397

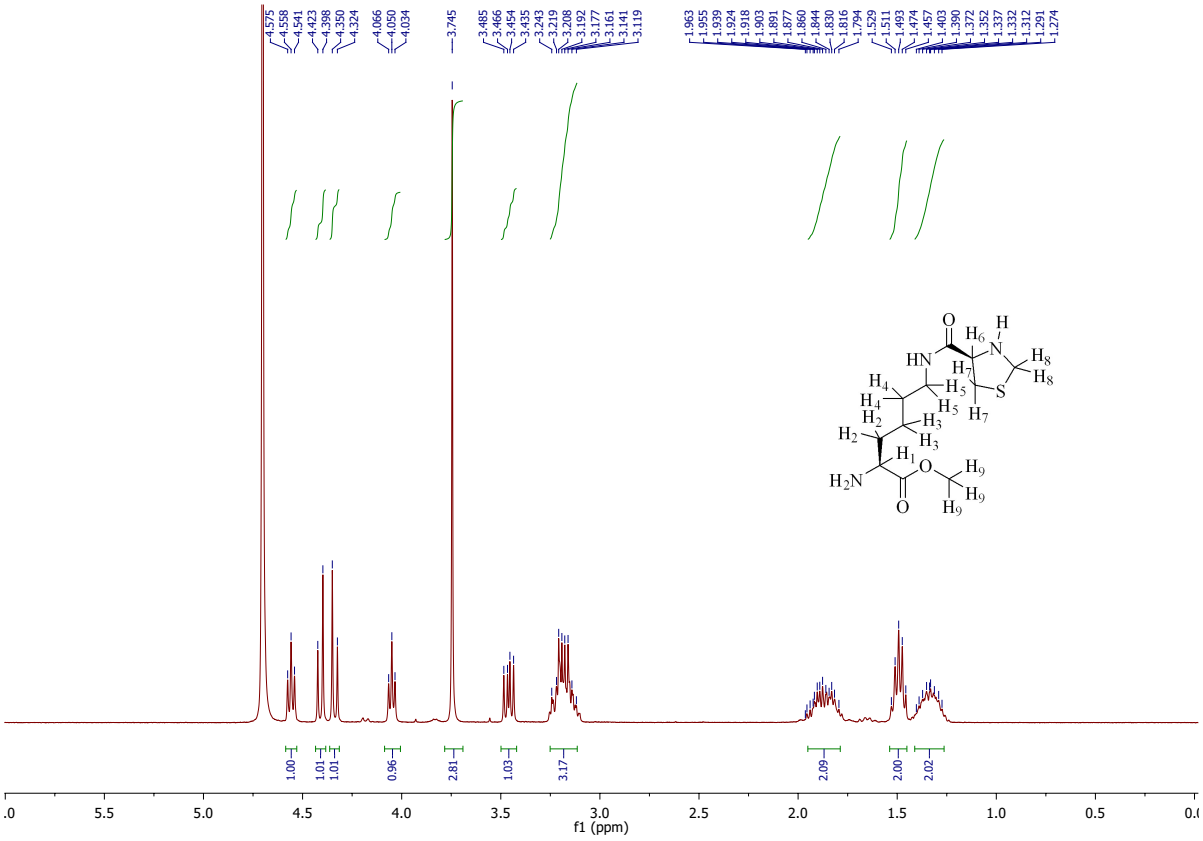
26. Kavask, P.; Rasmussen, R.; Causing, C.; Bonni, S.; Zhu, H.; Thomsen, G.; Wrana, J. Smad7 Binds to Smurf2 to Form and E3 Ubiquitin Ligase that Targets the TGFβ Receptor for Degradation. *Molecular Cell.* **2000**. 6(6) 1365-1375

27. Hall, J. SUMO Gene Fusion: New Methods for Enhancing Protein Expression and Purification in Prokaryotes. *Life Sensors.* **2004**

28. Wood, D.; Camarero, J. Intein Applications: From Protein Purification and labeling to Metabolic Control Methods. *J. Biol. Chem.* **2014.** 10 1074

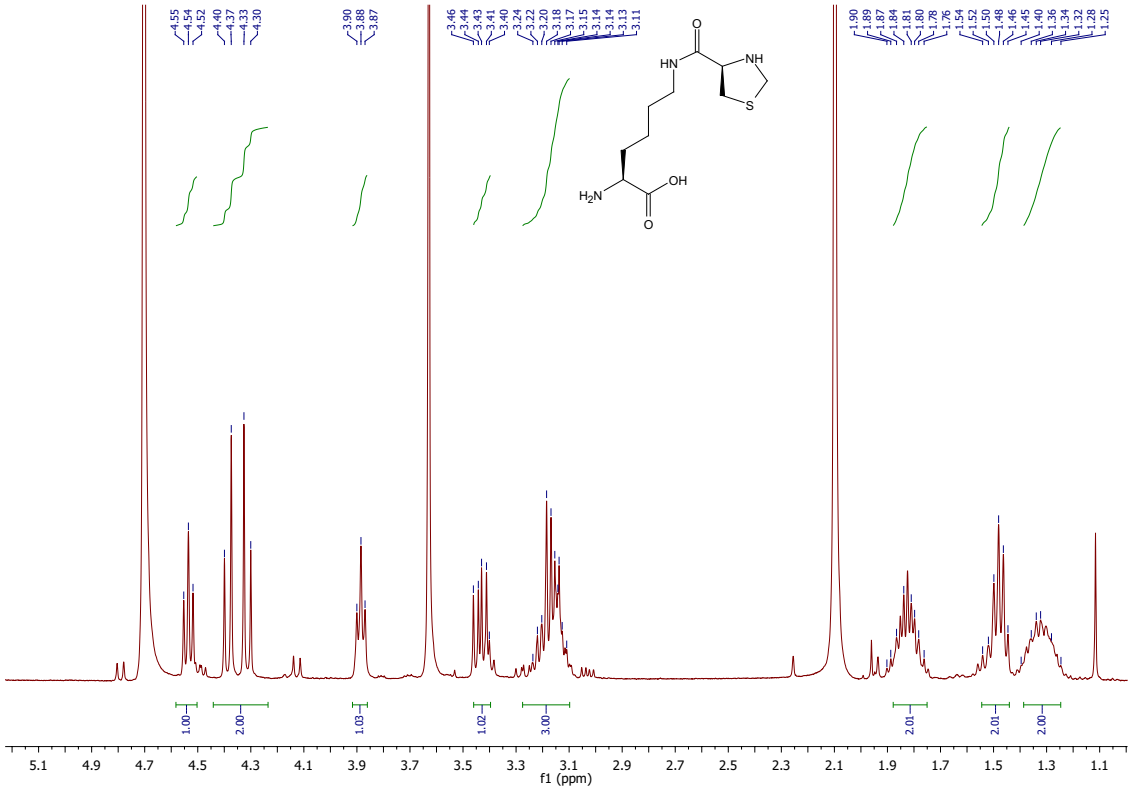
APPENDICES

^{1}H-NMR Analysis of Compound 23 ThzKOMe



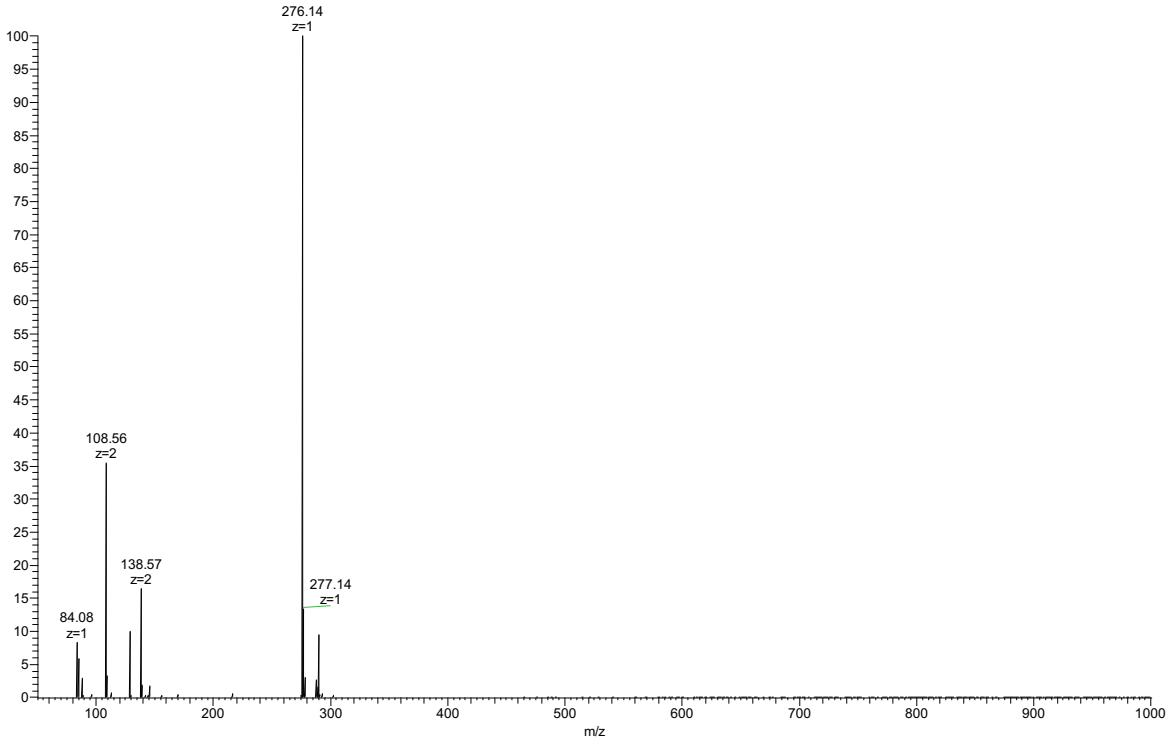
^{1}H NMR (400 MHz, D_2O) δ 4.55 (t, J = 6.8 Hz, 1H), δ *4.41* (d, J = 10 Hz, 1H), δ 4.33 (d, J = 10 Hz, 1H), δ 4.05 (t, J = 6.0, 1H), δ 3.74 (s, 3H), δ 3.48- 3.43 (m, 1H), δ 3.24 – 3.11 (m, 3H), δ 1.96 – 1.79 (m, 2H), δ 1.52 – 1.40 (m, 2H), δ 1.39 – 1.27 (m, 2H)

[1]H-NMR Analysis of Compound 22 ThzKOH

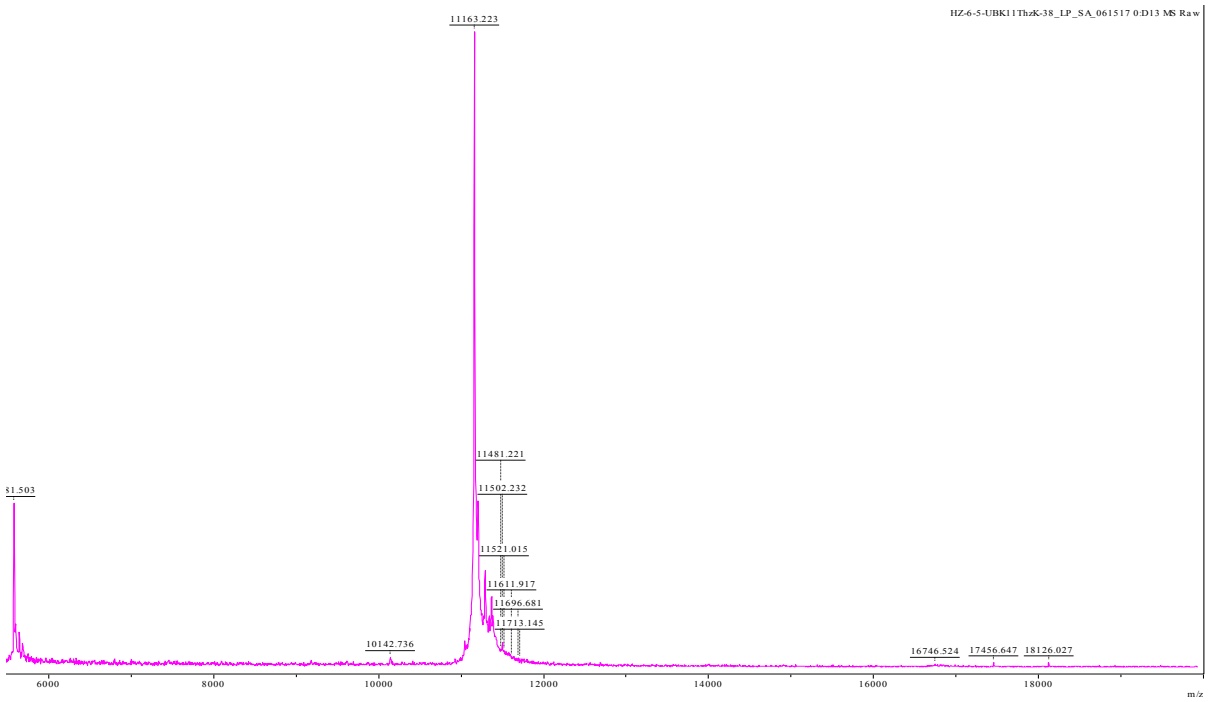


[1]H NMR (400 MHz, D_2O) *δ* 4.55 (t, J = 6.8 Hz, 1H), *δ 4.38* (d, J = 10 Hz, 1H), *δ* 4.32 (d, J = 10 Hz, 1H), *δ* 3.88 (t, J = 6.0, 1H), *δ* 3.46- 3.40 (m, 1H), *δ* 3.24 – 3.11 (m, 3H), *δ* 1.90 – 1.76 (m, 2H), *δ* 1.54 – 1.45 (m, 2H), *δ* 1.40 – 1.25 (m, 2H)

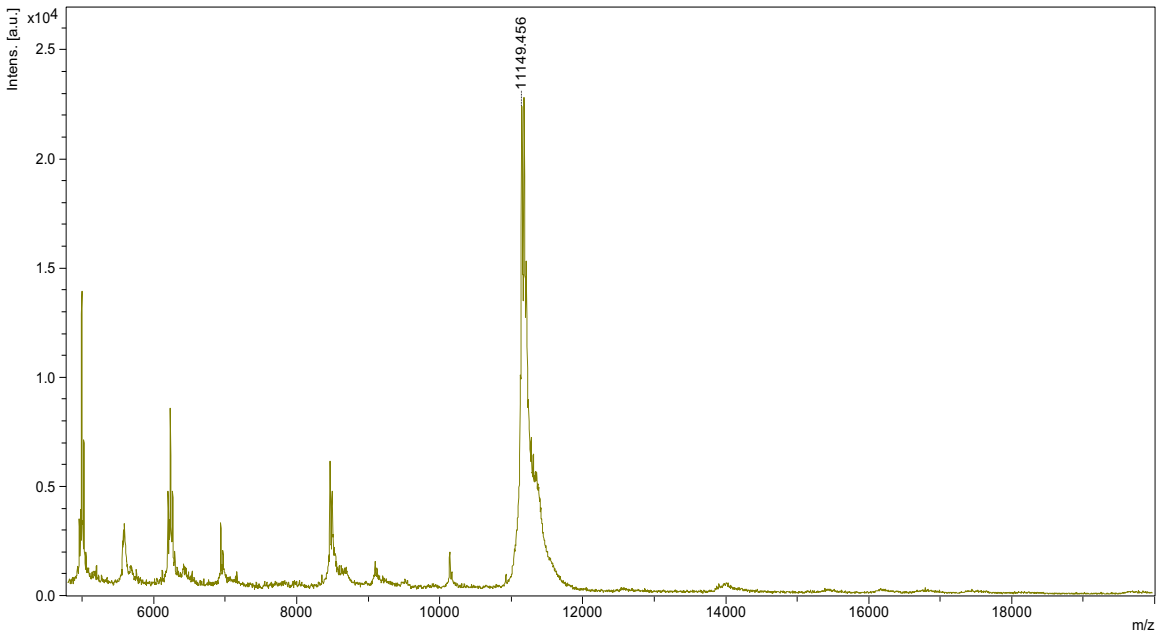
Electrospray Ionization for ThzKOMe



MALDI Characterization for Incorporated UAA – UbK11ThzK



MALDI Characterization of Deprotected CysK



MALDI Characterization of UbK11CysUb