DISTRIBUTION AGREEMENT

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Ogonna Nwajiobi 07/12/22

Development of Chemical Methods for Labeling Monomethyl Lysine Post-Translational Modification and Serine and their Applications.

By

Ogonna Nwajiobi

Doctor of Philosophy

Chemistry

Dr. Monika Raj
Advisor

Dr. Jen Heemstra
Committee Member

Dr. Vincent Conticello
Committee Member

Accepted:

Kimberly Jacob Arriola, Ph.D, MPH

Dean of the James T. Laney School of Graduate Studies

Date

THE ABSTRACT COVER PAGE

Development of Chemical Methods for Labeling Monomethyl Lysine Post-Translational Modification and Serine and their Applications.

By

Ogonna Nwajiobi

B.Sc., Nnamdi Azikiwe University, 2010

M.Sc., Nnamdi Azikiwe University, 2014

Advisor: Monika Raj, PhD

An abstract of

A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry 2022

ABSTRACT

Development of Chemical Methods for Labeling Monomethyl Lysine Post-Translational Modification and Serine and their Applications.

By

Ogonna Nwajiobi

Chemoselective modification of peptides and proteins offers a powerful way of studying protein functions and generating hybrid molecules with diverse functions. In this dissertation, I present chemoselective techniques for labeling monomethyl lysine, the novel synthesis of smart macrocyclic peptides, and methods for site-specific modification of serine for the generation of homogeneous biologics.

Lysine monomethylation (Kme) is an impactful post-translational modification (PTM) responsible for regulating biological processes and implicated in diseases, thus there is great interest in identifying these methylation marks globally. Current methods for detecting methyl lysine include Mass Spectrometry (MS) and antibodies, both of which are limited in efficiently characterizing the low abundant monomethyl lysine in a complex mixture. To tackle this, we developed a chemical technology (STaR chemistry) in which an electron-rich diazonium ion covalently labels the monomethyl lysine on peptides/proteins generating a triazene-chromophore in solution-phase or solid support, and the bioconjugate can be released to afford highly pure unmodified peptides from complex mixtures.

The STaR chemistry was further adapted to the synthesis of smart macrocyclic peptides. Myriads of methods exist for the synthesis of macrocyclic peptides, but most of these cannot generate a macrocyclic peptide that is responsive to external stimuli and contains an in-built chromophore. I developed a highly chemoselective, rapid, intramolecular reaction of secondary amines and p-amino phenylalanine that generates an inbuilt chromophoric triazene cyclic peptide. The triazene cyclic peptides can be opened in response to acidic conditions and UV radiations and can be preferentially reclosed when restored to neutral pH.

Lastly, I developed a site-specific modification of serine residues on peptides by enhancing the nucleophilicity of the hydroxyl group on serine residues within the catalytic triad. Through rational finetuning of the microenvironment through spatial positioning of amino acids around serine (H*X*S*X*H), serine was made very reactive over other serine residues. 2-aminoethyl) benzene sulfonyl fluoride (AEBSF) was found to react with the peptides containing (H*X*S*X*H) motif over other nucleophilic residues on the same peptide. Thus, this strategy can be utilized to preferentially incorporate desired moieties at specific sites on proteins such as in ADCs to generate homogeneous products.

Development of Chemical Methods for Labeling Monomethyl Lysine Post-Translational Modification and Serine and their Applications.

By

Ogonna Nwajiobi

B.Sc., Nnamdi Azikiwe University, 2010

M.Sc., Nnamdi Azikiwe University, 2014

Advisor: Monika Raj, PhD

A dissertation submitted to the Faculty of the James T. Laney School of Graduate Studies of Emory University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Chemistry 2022

ACKNOWLEDGEMENT

I believe that the words of Isaac Newton: “If I have seen further, it is by standing on the shoulders of Giants” applies to me more than any other living human. I have indeed been a beneficiary of untold kindness from uncountable persons; thus, I have the obligation of gratitude for "No duty is more urgent than giving thanks.”, James Allen. I have a litany of people I owe my success to.

Starting from the beginning of this journey, no one deserves better recognition than Sr. AmaraKristi Onyido, PhD who insisted that I applied to Auburn University when my attention was turned to schools in the UK. She read my application materials, provided feedback, and persisted with me in prayers. When I was successfully accepted into Auburn she made sure I had everything I needed before I landed at Auburn. I owe you eternal gratitude for your labor of love.

Upon joining the Chemistry Department, I needed a PI to groom me into an enviable scientist and I am so glad I chose to work for Dr. Monika Raj. Being the first graduate student to join the lab, I still vividly remember the two of us setting up the lab. You taught me everything with untold devotion and clarity. Thank you for being present and available, an excellent teacher, an inspiring mentor, and a model of hard work. Thank you for pushing me hard to be the best version of myself. For all the members of Raj lab, past, and present: Postdocs (Mahesh Sriram, Ashish Kumar, Samrat Sahu and Pratap Rajendra), Grad students (Victor Adebomi, Kuei Tang, Rachel Wills, Benjamin Emenike, Angele Bruce, Lyndsey Prosser, Patricia Rodriguez and Jonathan Farhi), Undergrad (Xavier Streety, Catherine Spancake, Patrick Czabala, Jake Moreh, Sophia Shahin, Matthew Buxton, Julia Donovan, Daniel Wang), you are all part of my success story. I am glad for the culture of conviviality and collaborative spirit we consciously created in the labor.

I have been blessed with very supportive committee members that pride themselves on excelling in mentorship: Dr. Jen Heemstra, Dr. Vincent Conticello, Dr. Christopher Easley, Dr. Holly Ellis and Dr. Evert Duin. Your constructive feedback on my research, encouragement and recommendations have pushed me on the advantageous path toward actualizing my dreams. I am grateful to all of you. In the same light, I want to thank the numerous industry mentors that have blessed me with their time, resources, and experience. Of special mention is Dr. Duncan Coffey who has been more than a mentor, I am indeed grateful for preparing me for life after graduate school. Also, I wish to acknowledge our collaborators, Drs Jae Won Chang and Sikh Cheung Hahm of Winship Cancer Institute for a successful collaboration.

To my most beloved wife, Sommy, you have been with me throughout my graduate school providing your unalloyed support and being my number one fan. I would not have done this without your partnership. I am very grateful to you for making our home a peaceful place where impactful undertakings can be pursued and to my boys, Chidum and Chibuifem, the desire to give you a good life makes me wake up in the morning without the need for alarm: thanks for inspiring me in this manner. To my most beloved parents: Late John Nwajiobi thank you for inspiring me with your unparalleled intelligence and love for education. My dear mother, Ezinne Calista Nwajiobi, thank you for teaching me resilience and dedication, virtues that have helped me through graduate school. Also, I have been hugely supported by my siblings, thanks for believing in me.

In a special way, I owe a debt of gratitude to the dear sister of mine, Sr Jovita Nwajiobi who has been a strong pillar of support since this journey started.

To the Almighty God, the source and summit of science, thank you for giving me the excellent spirit to discover you in created things through the lens of chemistry. I owe all the good things I have received to you. You have established my feet on solid ground.

Table of Contents

List of Tables

CHAPTER ONE: CHEMICAL METHODS FOR TAGGINGMONOMETHYL LYSINE POST-TRANSLATIONAL MODIFICATIONS

1.0. Introduction

Dynamic PTMs regulate protein-protein interactions, protein localization, protein stability, and enzymatic activities[1]. Post-translational modification (PTM) represents a biochemical event that converts a ribosomal encoded amino acid residue into non-standard amino acid residue by enzymatic activity. PTMs play a vital role in the transmission of biological signals[2] and more than twenty (20) different types of modifications on proteins have been documented[3]. It is now known that one of the major types of PTMs occurring on histone and non-histone proteins is the methylation of lysine[4]. Lysine can undergo multiple methylations such as monomethylation, dimethylation or trimethylation and each methylation state correlates with distinct functions (**Figure 1.1**). Protein Lysine Methyl Transferases (PKMTs) generally catalyze the transfer of methyl groups to acceptor lysine residues on histone and non-histone protein substrates using S-adenosyl-L-methionine as cofactor. They are referred to as *writers* while the Protein Lysine Demethylases (PKDMs) are referred to as *erasers* since they remove the methylation marks. The methylation events are linked to biological outcomes by methyl binding domains (MBDs) that recognize methylated lysine residues in a state-and sequence-dependent manner and they are called *reader* proteins. These events have huge relevance in biological processes and disease states[5] as PKMTs and PKDMs play critical roles in epigenetic regulation.

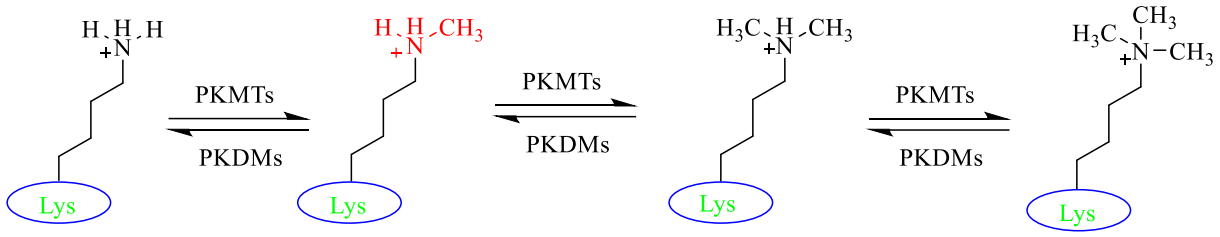


Figure 1.1: The different methylation states of lysine. Protein Lysine Methyltransferases (PKMTs) catalyze the transfer of methyl groups to lysine while Protein Lysine Methyl Demethylases (PKDMs) remove the methyl groups from lysine.

There has been a rigorous interest in mapping out the epigenetic role of PTMs in disease and development[6]. Lots of proteins have been discovered to be methylated; ranging from proteins involved in DNA transcription and RNA processing to cytosolic or membrane proteins involved in cell-cell interaction, cellular development and signal transduction[7]. With recent advances in innovative strategies and mass spectrometry, more than 5000 lysine methylation sites have been identified in proteins. Compared to that, Huang and Berger asserted that more than 10 lysine methylation sites have been found on proteins which extrapolates to 40,000 methylation sites within the 20,000 proteins in the human proteome[8]. So, it goes without saying that many lysine methylated sites on proteins remain uncharacterized. Amino acid sequencing[9], radio-labelled techniques[10], or immunoblotting[11] were the earliest methods used to identify lysine methylation. Edman sequencing takes a long time and requires a large number of target proteins, whereas immunoblotting and radioactive methylation assays do not reveal the methylation location.

Efficient capturing of global lysine methylation will deepen our understanding of their biological importance, but tools to detect methylated lysines are limited. Detection of lysine methylation is compounded by two major factors: first, the methylation of lysine does not add a substantial steric bulk (i.e the molecular weight of methyl group is small in comparison to other PTMs). It can be a challenge to isolate post-translationally modified peptides in which the PTM induces a small change12. Between the possible methylation states, this challenge is more pronounced on

monomethylation. Second, methylation on lysine does not neutralize the charge on the amino acid. So, there is no charge difference between methylated and unmethylated residues. Because of the aforementioned factors, it is difficult to detect lysine monomethylation PTMs in a selective manner. Current methods for detecting monomethyl lysines include Mass Spectrometry (MS) and antibodies but are limited in efficiently characterizing the low abundant monomethyl lysines in a complex mixture. Consequently, the identification of mono-methylated proteins in the proteome is highly challenging. Therefore, there is a great need to develop new methods for selective detection of lysine methylome in proteome.

1.1. Background

Site-selective covalent labeling of PTMs represents an innovative tool which has the potential to expand the chemical tool kit available for epigenetics research[13]. Numerous reports have confirmed that lysine methylation PTM is important in a variety of biological processes including but not limited to DNA repair, RNA processing, translation, cellular localization and other signaling pathways[14,15,16,17,18]. Methylation of lysine residues is thought not to change the overall charge nor the conformation of the substrate proteins[15] but significantly change the hydration energy and hydrogen potential of the lysine residues[19]. However, it is believed to modulate the surface architecture, thereby, leaving a mark that can either inhibit or promote binding of a substrate protein. Another implication of protein methylation is that it inhibits the attachment of the other possible PTMs on lysines[8]. Lysine methylation is very well documented on histone proteins but a variety of non-histone proteins are equally monomethylated such as transcription factors (TAF10, TAF7 and eEF1a), p65, estrogen receptor α, Retinoblastoma tumor suppressor protein (Rb), HIV Tat and Heat shock 90 (HS90) proteins. The major concern is that dysregulation of lysine methylation PTMs has significant roles in tumorigenesis[20,21,22]. A typical protein that is

extensively monomethylated and involved in tumorigenesis is p53. The over-expression of SMYD2 which monomethylates p53 at K370 has been found in hepatocellular carcinoma, primary tumor samples of esophageal squamous cell carcinoma and in pediatric lymphoblastic leukemia[23]. Moreover, SMYD2 methylation events have been found to promote cyst growth in autosomal dominant polycystic kidney disease[24]. Also, the over-expression of G9a (EHMT2) which is responsible for methylation of p53 at K382 has been implicated in different cancers[25]. Documented records have widely established the biological roles of lysine protein methylation and its involvement in human diseases, it is therefore pertinent that robust methods are developed to identify methylated lysines on proteins and their site of methylation.

1.1.1. Reported Methods and Limitations

Mass spectrometry has been the mainstay for proteomic scale studies of PTMs[26]. Despite the overwhelming advantages of mass spectrometry-based proteomics, it is time consuming and expensive. Coupled to the fact that it is limited for detecting low abundant monomethyl lysines in a complex mixture the need for a new technique is abundantly. This challenge can be mitigated by protein enrichment. Protein enrichment is a technique by which specific proteins often of low abundance are concentrated to improve the ease of their identification.

The second method involves the enrichment of post-translationally modified proteins using antibodies. While antibodies have been quite successful for acetylation, ubiquitination, phosphorylation and even arginine methylation[15,27], there have been huge challenges in developing high affinity pan-antibodies for methyl detection due to the small size[9]. Also, antibody affinity enrichment suffers from other serious setbacks such as they are unable to distinguish between different methylated states of lysine, adjacent PTMs can negatively affect antibody recognition of the modification of interest[28,29], cross-reactivity with off target antigens, and batch

to batch irreproducibility[30]. Moreover, antibodies cannot be used to identify unknown PTMs due to their sequence specificity[31].

The third method involves the use of natural methyl-lysine binding domains (MBDs). These domains are known to recognize methylated lysine through a hydrophobic binding pocket that forms limited contact with proximal residues[32]. However, the major limitation associated with MBDs is that it cannot identify the site of modification from the enriched protein[33]. Also, aside from being a non-covalent interaction it cannot distinguish between mono- or di-methylated peptides nor does it determine the stoichiometry of methylation at any site[34]. The group led by Marcey Waters has used small molecules (synthetic receptors) that mimic chromodomain to target and identify trimethyl lysine on proteins in a non-covalent manner. They reported the use of dynamic combinatorial chemistry (DCC) to identify receptors for trimethyl lysine that have similar binding affinity as native HP1 chromodomain[35]. Recently, they engineered chromodomains from *Drosophila melanogaster* and humans with high binding affinity to H3K9me3[36]. As elegant as those reports are, they only efficiently target and selectively label trimethyl lysines.

Wu *et al.* developed an elegant indirect chemical protocol for detecting monomethyl lysines on proteins by derivatizing the gamma-amine group of the lysine with propionyl moiety and subsequently enriching them using a pan anti-propionyl monomethyl lysine antibody[37]. However, the method is not specific since the propionylation reaction occurs at monomethyl lysine as well as on unmodified lysine. Also, there is a requirement for antibody for the enrichment. Additionally, other groups have explored functionalization of the methyl donor, S-adensoyl-methionine with biorthogonal groups like alkynes or azides which are transferred to the substrate proteins in place of the methyl group[38,39,40]. However, it is reported that the unnatural SAM analogues have cross-reactivity with native SAM and not all PKMTs accept the cofactors[41]. Together, none of these

methods are suitable for selective tagging of monomethyl lysine in a Pan specific manner, for determining the site of modification and their individual roles in controlling biological functions. Carlson and Gozani equally opined that research into non-histone protein lysine methylation has stalled because of limited strategies of identifying monomethylated lysines across the entire proteome[42].

To effectively study and understand the role of protein lysine monomethylation in various biosignaling and pathologies, there is a need for a chemoselective method that can covalently tag these methyl marks and enrich them even in their low abundant states. Currently, there are no methods available in the literature to achieve this goal. I present the efforts we have made towards tackling this challenge by developing new chemical tools with unique chemoselectivity toward monomethyl lysines. This chemical technology forms a strong covalent bond with monomethyl lysine in a selective manner and has been employed in labeling monomethyl lysine on a global scale. We started with the development of Secondary Amine Selective Petasis (SASP) Bioconjugation.

1.2. **Secondary Amine Selective Petasis (SASP) Bioconjugation**

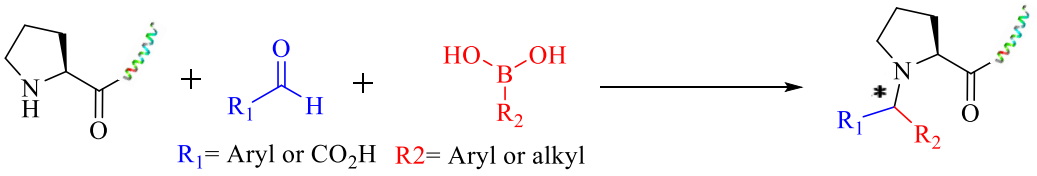


Figure 1.2: Secondary Amine Selective Petasis Bioconjugation (SASP) involving a secondary amine, an aldehyde and a boronic acid to form a stable bioconjugate with a chiral center at the site of modification

The SASP bioconjugation was developed in our lab to covalently label secondary amine on peptides and proteins.[43] The SASP bioconjugation is a multicomponent reaction involving a secondary amine, aldehyde, and organoboron to afford a stable product under physiological conditions (**Figure 1.2**). The secondary amine reacts with an aldehyde to form a reversible

iminium ion which further undergoes a nucleophilic attack by an organoboronate to form a stable C-C bond. The SASP bioconjugation strategy was the first utilization of the Petasis reaction under physiological conditions for modifying unprotected proteins and peptides. The bioconjugation strategy has the following attractive features: First, it is very selective for the N-terminal proline and does not cause any other amino acids to be modified. Second, it is a multicomponent reaction that dually labels peptides and proteins under ambient biological conditions without the use of any catalysts. Also, with strong stereoselectivity (de >99 %), the SASP reaction generates a new chiral center at the conjugation site. Because of the secondary amine selectivity of this reaction, we reasoned that it could be particularly beneficial for selectively tagging/enriching proteins with monomethyl lysine PTMs to discover complete sites of monomethyl lysine Kme in the entire proteome.

1.2.1. Development of Secondary Amine Trapping Reagents (STRap) for Tagging Monomethyl Lysine

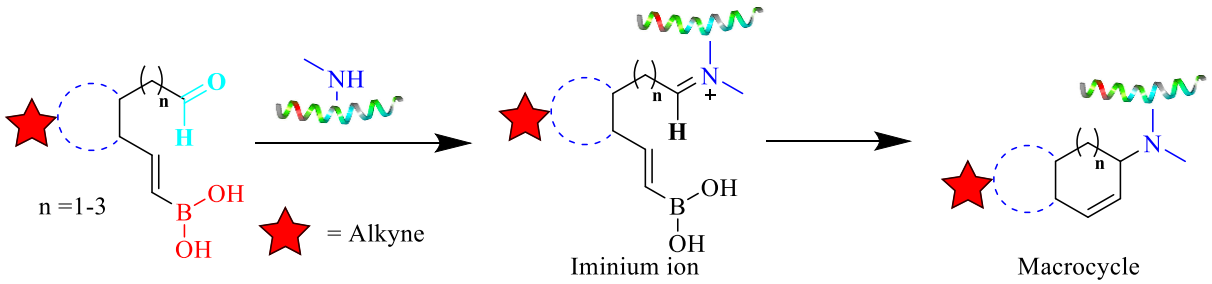


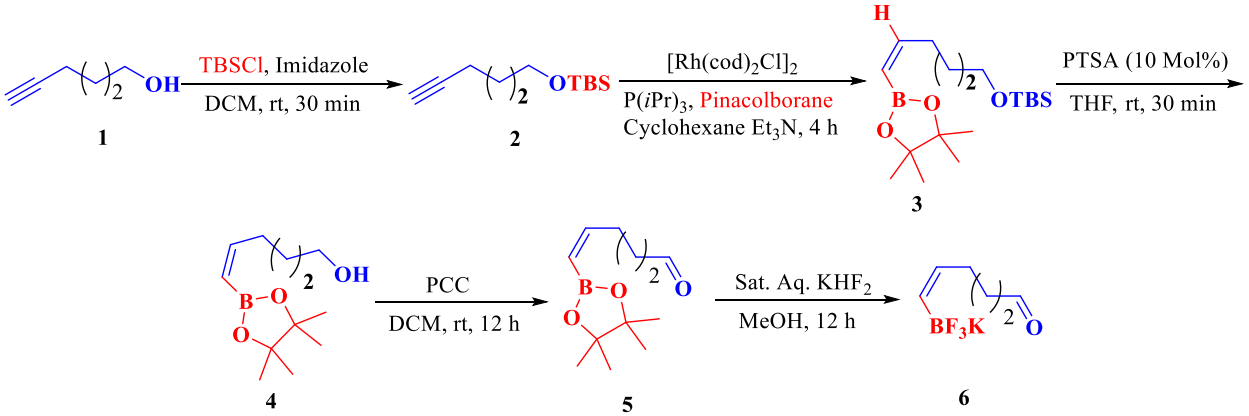
Figure 1.3: Bifunctional probe for rapid trapping of secondary amines. The probe reacts with monomethyl lysine to form an iminium ion and intramolecular nucleophilic attack from the organoboronate will form a stable macrocycle.

The SASP bioconjugation proved to be an efficient strategy for selectively labeling secondary amines such as N-terminal prolines, even achieving selectivity for proline in the presence of competing cysteine and lysine residues. Inspired by this observation, we sought to further develop this bioconjugation strategy for selective detection of monomethyl lysine post-translational modification (PTM) to determine their involvement in regulating various physiological processes.

Since SASP Bioconjugation is a multicomponent reaction, we anticipated that the relatively low rate of reaction could present a challenge in the trapping of low abundant monomethyl lysine peptides. The low natural abundance of monomethyl lysine PTM in the cell is one of the stumbling blocks in their global identification. Thus, we rationalized that converting the three-component reaction to a two-component reaction by introducing two moieties (an aldehyde and boronic acid) in one probe could enhance the rate of the reaction (**Figure 1.3**). We rationalized the proximity effect to be conferred by tethering these functional groups on the same molecule could accelerate the rate of the reaction. These bifunctional STRap reagents would significantly accelerate the rate of the reaction and be able to enrich low abundant monomethyl lysine proteolytic fragments from the complex mixture.

The general design of SA-traps consists of the incorporation of two functional groups; (1) aldehyde for the formation of reversible iminium ion with Kme and (2) vinyl or aryl boronic acid for iminium ion trapping via a proximity-induced intramolecular nucleophilic reaction to generate a stable C-C bond. We made aliphatic, aryl and heteroaryl STRap reagents and reacted them with peptides to determine their chemoselectivity towards secondary amines.

1.2.1.1. Aliphatic STRap



Scheme 1.1: Synthesis of vinyl boronic acid aliphatic STRaP

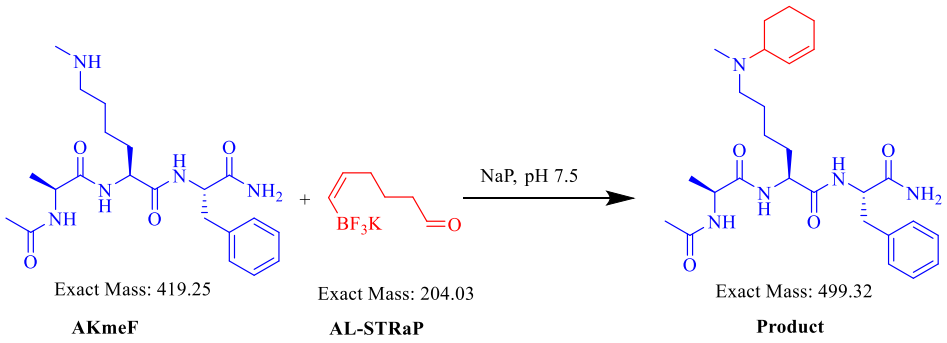
The aliphatic STRap (Al-STRap) was synthesized by following the steps outlined in **Scheme 1.1** in which an aldehyde and a cis-vinyl boronic were fused together with 4-Pentyn-1-ol as the starting reagent for this reaction. The reaction of this probe with the secondary amine substrate is envisaged to lead to the formation of a stable 6-membered cycle. For the synthesis of the AL-STRap, two considerations were taken: first, careful design ensured that the reagent is water-soluble since the ideal condition for tagging Kme should be in physiological condition. Secondly, that the reagent possesses cis-configuration of a vinyl boronic acid to facilitate the nucleophilic reaction with iminium ions.

Chemoselectivity Studies of AL-STRap with Different Peptides

Next, we carried out reactions with peptides containing secondary amine and primary amine under different conditions to determine chemoselectivity towards monomethyl lysine (Kme). We started with the peptide AKmeF acetylated at the N-terminal to avoid confounding results with the primary amine at the N-terminus. The reactivity of the peptide with the AL-STRap was determined at physiological temperature and different solvent combinations. Though the desired product was observed (**Table 1.1**), the percentage conversion to the product as determined by HPLC and MS was poor with the highest (25 %) being in the combination of MECN: Tris buffer. We went ahead to evaluate if Al-STRap reacts with primary amines on peptides. We employed the peptide, AAF with free amine at the N-terminus and exposed it to our reaction conditions as outlined in **Table 1.2**. From HPLC and MS analysis, single modification and double modification of the peptide were observed. In buffer, the conversion of the first Petasis product to the second product was not observed within the reaction time (6 and 12 hours) . The reaction worked better when buffer was used in combination with organic solvents, dimethylformamide (DMF), and acetonitrile (MeCN) and the highest overall conversion was observed by DMF. Our studies establish that the AL-

STRap reacted moderately with primary amines as well. From these findings, the rate of reaction of AL-STRap with Kme was not fast to efficiently label monomethyl lysine containing peptides neither was it selective for Kme over primary amines.

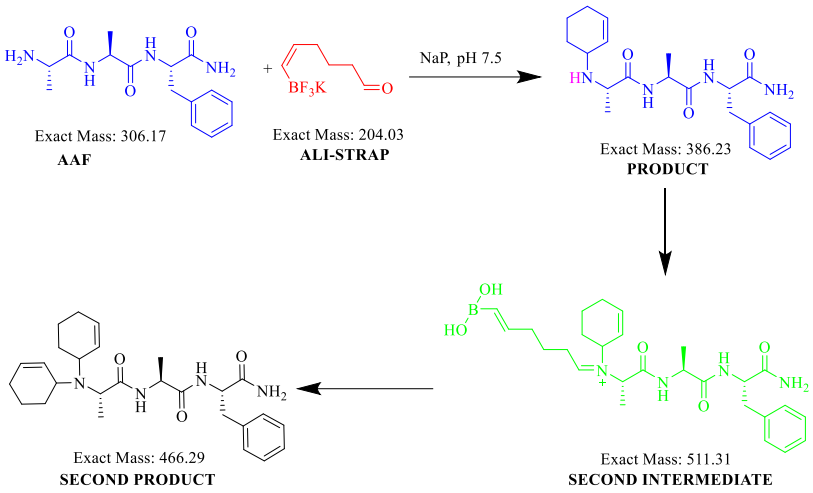
Table 1.1: Reaction of Aliphatic STRap (AL-STRap) with AKmeF peptide under different reaction conditions.



Entry	Peptide	Solvent	Temperature	Time	Conversion (%)
a	AKmeF	Tris Buffer	37 ºC	12	18
c	AKmeF	MeCN: Buffer (1:1)	37 ºC	12	25
d	AKmeF	DMF: Buffer (1:1)	37 ºC	12	18

Conditions: 1 eq of peptide, 4 mM and 2 eq of AL-STRap, 8 mM Tris Buffer, NaP, 25 mM, pH 7.5

Table 1.2: Reaction of Aliphatic STRap (AL-STRap) with AAF peptide under different reaction conditions.

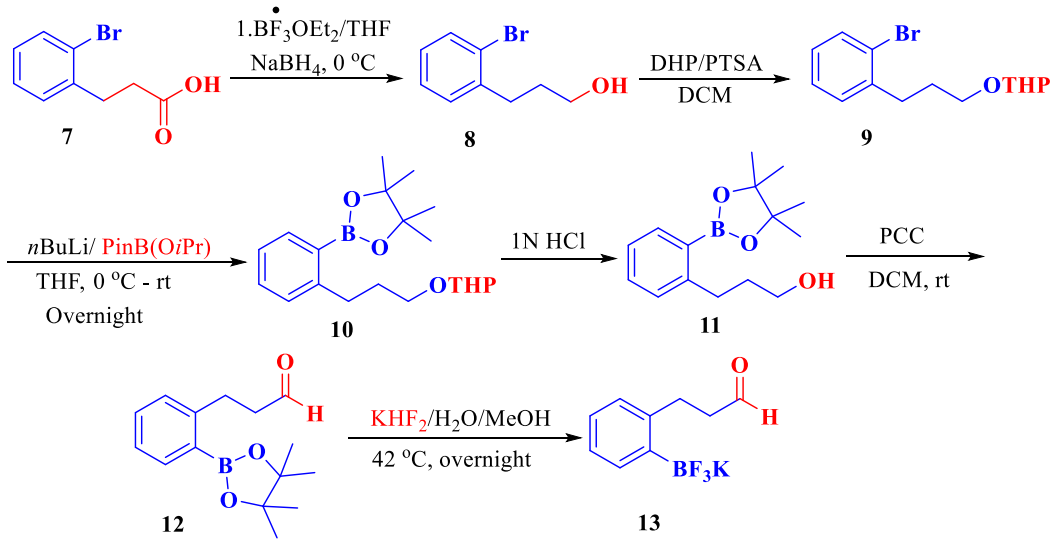


Entry	Peptide	Solvent	Temperature	Time (h)	Conversion (%)
a	AAF	Tris Buffer	37 °C	6	Product (11) 2nd Intermediate (7)
b	AAF	Tris Buffer	37 °C	12	Product (25) 2nd Intermediate (15)
c	AAF	MeCN: Buffer (1:1)	37 °C	6	Product (18) 2nd Intermediate (27) 2nd Product (7)
d	AAF	MeCN: Buffer (1:1)	37 °C	12	Product (28) 2nd Intermediate (0.5) 2nd Product (15)
e	AAF	DMF: Buffer (1:1)	37 °C	6	Product (16) 2nd Intermediate (21)

					2nd Product (16)
f	AAF	DMF: Buffer (1:1)	37 °C	12	Product (16) 2nd Intermediate (33) 2nd Product (25)

Conditions: 1 eq of peptide, 4 mM and 2 eq of AL-STRap, 8 mM Tris Buffer, NaP, 25 mM, pH 7.5

1.2.1.2. ARYL STRap

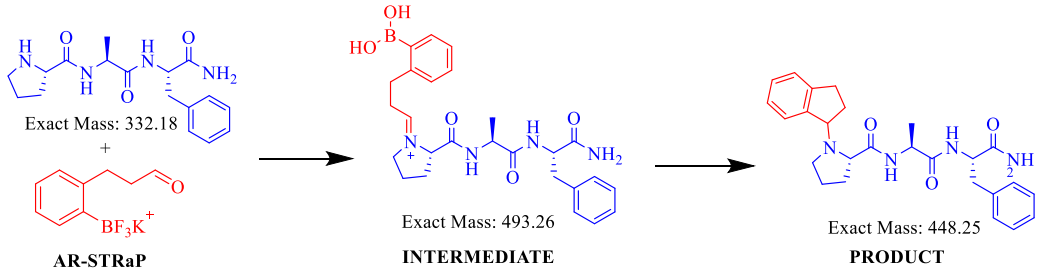


Scheme 1.2: Synthesis of Aromatic STRaP (AR-STRAP) (13)

We synthesized the AR-STRap according to the procedure outlined in **Scheme 1.2** starting with the reduction of compound **7** to afford the alcohol which is further oxidized to give the aldehyde. In the same manner, boronation of compound 10 gave rise to boronate ester enabling the synthesis of the final product as a potassium trifluoride **13**. With the AR-STRap in hand, we reacted it with a peptide containing PAF, containing a secondary amine under several conditions to ascertain its selectivity towards secondary amine (**Table 1.3**). We started with exposing the peptide and AR-STRap to sodium phosphate buffer in both 37 °C and 60 °C for 12 hours but observed no

modification of the peptide. In a bid to facilitate the reactivity of the probe, we employed HFIP as a co-solvent for the reaction. HFIP has been reported to accelerate the rate of Petasis reaction shortening the reaction time and increasing yields.[44] However, we recorded only 30 % conversion to the product with HFIP as a co-solvent with sodium phosphate buffer in 1:1 ratio at 60 ^{o}C with no reaction occurring at 37 ^{o}C. Interestingly, with HFIP as the only solvent, the reaction goes to completion in 6 hours at 60 ^{o}C for both PAF and AAF, a peptide containing a primary amine. Since HFIP can only afford high reactivity at an elevated pH and marred by lack of selectivity for secondary, we sought other conditions that could enhance the rate of the reaction while favoring reactivity with secondary amine. We first started with combining MEOH and buffer at 60 ^{o}C for the reaction between PAF and AR-STRAP and observed poor modification (15 %) of the peptide after 12 hours. However, by using ethanol as the solvent we observed very high conversion (94 %) to the product in 6 h at 60 ^{o}C and (44 %) at 37 ^{o}C. Since ideal bioconjugation reactions should be operational at ambient temperature and water, we examined the possible solvent combinations that can get us closer to the ideal conditions. From the study, we observed that the most tolerable condition closest to the physiological condition is a combination of ethanol : buffer (1:1) with moderate conversion (50 %) of the peptide to the product in 12 hours at 37 ^{o}C. However, when we subjected AAF to the same conditions, we observed moderate modification of primary amines. Taken together, though AR-STRap has high reactivity towards secondary amine but can as well react with primary amines. Thus, this ruled it out as a reagent for the chemoselective tagging of monomethyl lysine (Kme).

Table 1.3: Reaction of Aryl STRap (AR-STRap) with different substrates under different reaction conditions.



Entry	Peptide	Solvent	Temperature	Time	Conversion (%)
a	PAF	Buffer	37 ºC	12 h	ND
b	PAF	Buffer	60 ºC	12 h	ND
c	PAF	HFIP: Buffer (1:1)	37 ºC	6h	ND
d	PAF	HFIP: Buffer (1:1)	60 ºC	12 h	30
e	PAF	HFIP	60 ºC	6 h	100
f	PAF	HFIP	37 ºC	12 h	50
g	AAF	HFIP	60 ºC	6 h	100
h	PAF	MeOH: Buffer(1:1)	60 ºC	12 h	15
i	PAF	Ethanol	37 ºC	6 h	44
j	PAF	Ethanol	60 ºC	6 h	94
k	PAF	Ethanol: DMF (1:1)	37 ºC	12 h	58
l	PAF	Ethanol: Buffer (1:1)	37 ºC	12 h	50
l	PAF	MeCN: Buffer(1:1)	37 ºC	12 h	45
m	AAF	Ethanol: DMF (1:1)	37 ºC	12 h	30

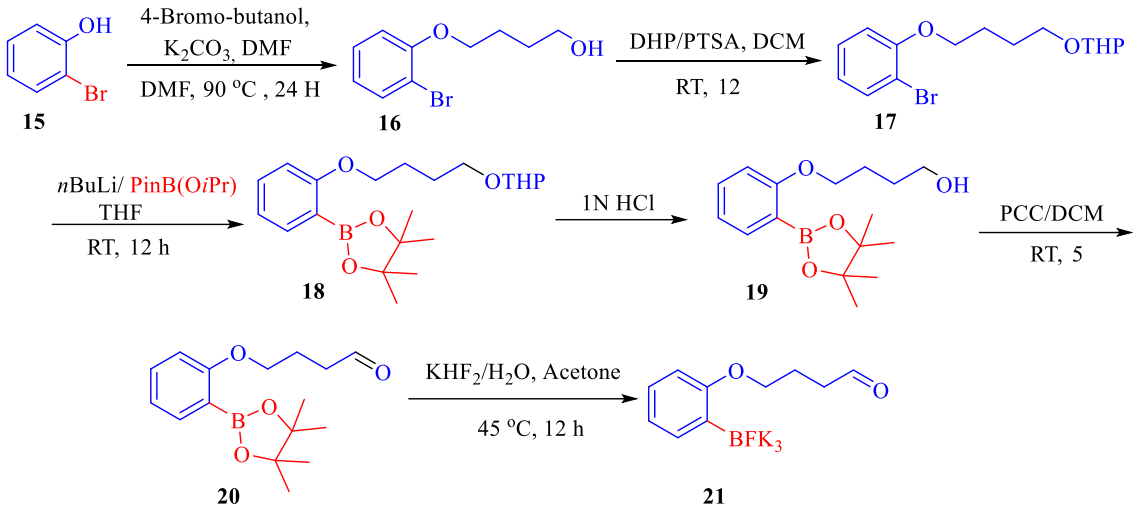
Conditions: 1 eq of peptide, 4 mM and 2 eq of AR-STRap, 8 mM

Buffer: Sodium Phosphate Buffer, NaP, 25 mM, pH 7.5

HFIP: Hexafluoroisopropanol

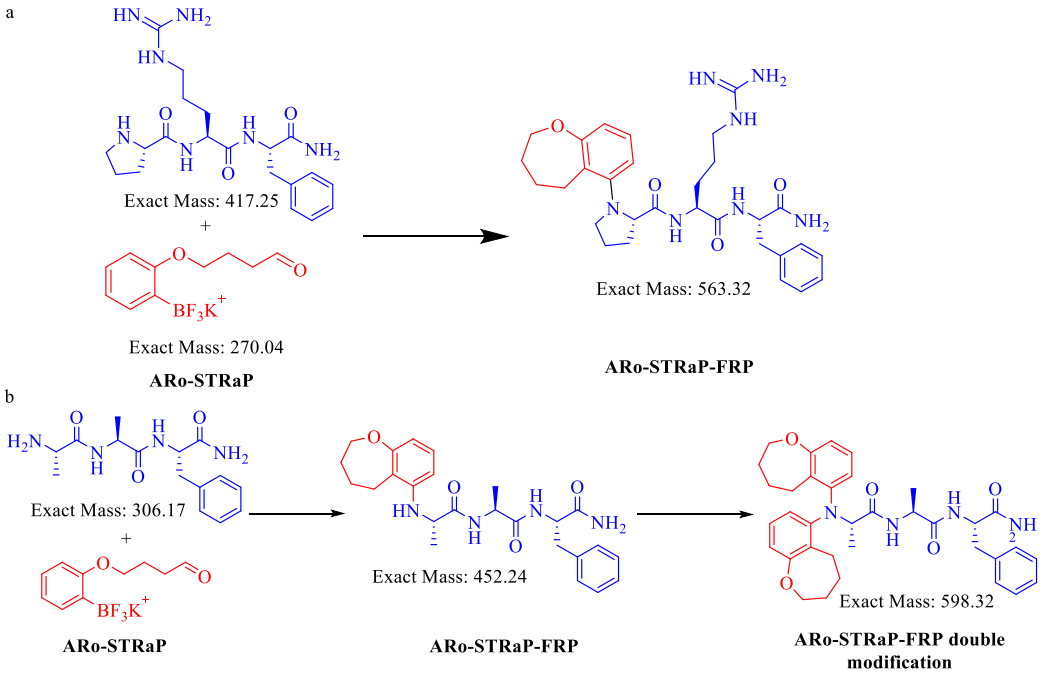
1.2.1.3. **HETEROARYL STRap**

Bothered by the lack of chemoselectivity of the AL-STRap and AR-STRap, we sought to make a probe that has a Lewis base close to the reactivity center. It has been noted that an important feature of the Petasis reaction is its requirement for a proximal Lewis basic group in the carbonyl component which has been proposed to activate and orient the organoboron species prior to addition. In agreement with this, report from literature confirms that hydroxy-substituted benzaldehydes particularly salicylaldehyde worked well for Petasis reaction[45, 46] and we observed this from SASP bioconjugation in which salicylaldehyde achieve high selectivity towards secondary amine.[3] So, we designed an heteroaryl probe having both aldehyde and boronic acid in the same molecule and the schematic pathway as shown in **Scheme 1.3** starting with 4-bromobutanol nucleophilic substitution reaction with bromophenol (**15**) to afford hydroxyl which was protected with OTHP to give compound **16**. Compound **16** underwent boronation reaction with pinacolborane and butyllithium to obtain a boronate ester **17**. With deprotection of OTHP followed by oxidation and deprotection of boronate ester, the desired AR-STRap (**20)** was obtained.



Scheme 1.3: Synthesis of heteroaryl STRaP

Table 1.4: Reaction of peptides with Aromatic STRap with heteroatom (ArO-STRap) under different reaction conditions.



Entry	Peptide	Solvent	Temperature	Time	Conversion (%)
a	PRF	HFIP	70 ^{o}C	12 h	91
b	PRF	HFIP: Buffer (1:1)	RT	12 h	79
c	PRF	HFIP: MECN (1:1)	RT	12 h	78
d	AAF	HFIP: Buffer (1:1)	70 ^{o}C	12 h	34 (**Single Modification**) 35 (**Double Modification**)

Conditions: 1 eq of peptide, 4 mM and 2 eq of ARo-STRap, 8 mM

Buffer: Sodium Phosphate Buffer, NaP, 25 mM, pH 7.5

HFIP: Hexafluoroisopropanol

With the probe in hand, we determined the conditions that engender rapid reaction rate as well as chemoselectivity towards secondary amine by applying HFIP as the solvent and as a co-solvent in different reaction conditions (**Table 1.4**). We used PRF as the secondary amine containing peptide and observed high conversion to the peptide in both high temperature and room temperature for 12 hours with HFIP (91 %) as the solvent; and with one-to-one combination with sodium phosphate buffer (79 %) and acetonitrile (78 %). Encouraged by this, we sought to rule out reactivity with primary amines by subjecting a primary amine containing peptide, AAF to our reaction conditions with a combination of sodium phosphate buffer and HFIP and observed single and double modification of the amine. We reasoned that the addition of an electron-donating substituent such as a para-alkoxy increased the reactivity of both secondary and primary amines. It is well established that Petasis reaction with aryl and heteroaryl boronic acids tends to provide higher yields if the boronic acid is electron-rich because there is greater electron density about the π-system and migrating substituent.[47] So, we could not achieve the chemoselective modification

of secondary amine using the ARo-STRap as we did in SASP bioconjugation by using salicylaldehyde which has hydroxyl group at the ortho-position.

Thus, sadly, none of the STRap reagents was selective towards secondary making us reconsider our strategy and search for alternative chemistries for selective tagging of mono methyl lysine.

1.3.0. Development of Selective Triazenation Reaction (Star) of Secondary Amines for Tagging Monomethyl Lysine Post-Translational Modifications (PTMs)

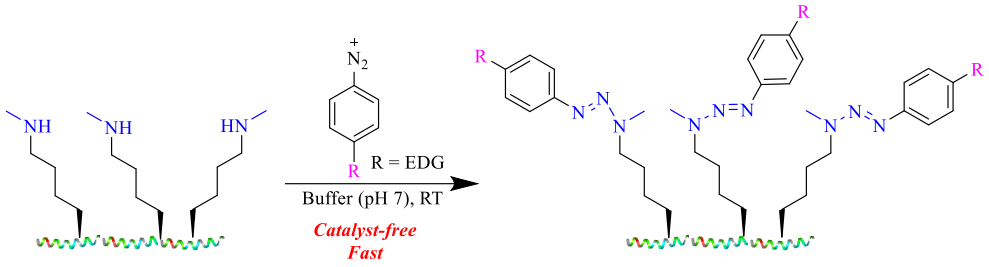


Figure 1.4: Triazenation reaction between electron rich diazonium ion and secondary amine functional group on monomethyl lysine

We reasoned that by exploiting the nature and concentration of arene diazonium ion and pH, we could selectively modify secondary amines such as monomethyl lysine over unmodified lysine and other reactive groups on proteins. From literature, arene diazonium ion form triazene with N-terminal amines and side chain of lysines but has never been pursued as a bioconjugation method due to the reversible nature of the reaction with primary amines[48]. Also, azo coupling of tyrosine with arene diazonium ion has been a hotpot for bioconjugation[49]. However, the latter is known to have a requirement for elevated pH (pH 9.0) and electron-withdrawing groups on the diazonium ion. The azo coupling is thought to proceed in the presence of a strongly electron-deficient diazonium partner owing to the moderate nucleophilicity of the phenol groups in tyrosine[48]. Drawing inspiration from these observations (**Figure 1.4**), we rationalized that the identification of arene diazonium ions with electron-donating substituents could form the basis of a novel chemoselective reaction that can detect monomethyl lysine on proteins. The reaction will form a

strong irreversible bond with secondary amines under physiological conditions in an efficient manner.

To test our hypothesis, we started by evaluating the reactivity of different secondary amines with electron-rich arene diazonium ions. Our model reaction was between 4-methoxybenzenediazonium tetrafluoroborate (4MDz) and monomethyl lysine which gave 99 % conversion to the product in 30 minutes under our optimized conditions.[11] The optimization studies showed that the reaction between the two partners in 0.6 mM requires the presence of Na_2CO_3 (0.6 mM) in phosphate buffer (100 mM, pH 7) at room temperature for 30 min. Having identified the ideal conditions, we proceeded to the peptide level by carrying out a reaction between 4MDz (0.6 mM) with monomethyl lysine peptide AKmeF (0.6 mM). We monitored the reaction for 24 h by taking aliquots at regular intervals of time. We compared the rate of the formation of the triazene on the peptide with the rate of formation of azo-coupled product on tyrosine. We used spectrophotometry to confirm the formation of the chromophoric triazene-coupling product with increased absorption at 330-390 nm and we recorded 93% conversion to triazene-coupling adduct in just 5 min as analyzed by HPLC (Figure 1.5, Supplementary Figure 1). Interestingly, even after 24 hours, the azo-coupled product from the tyrosine was not observed.

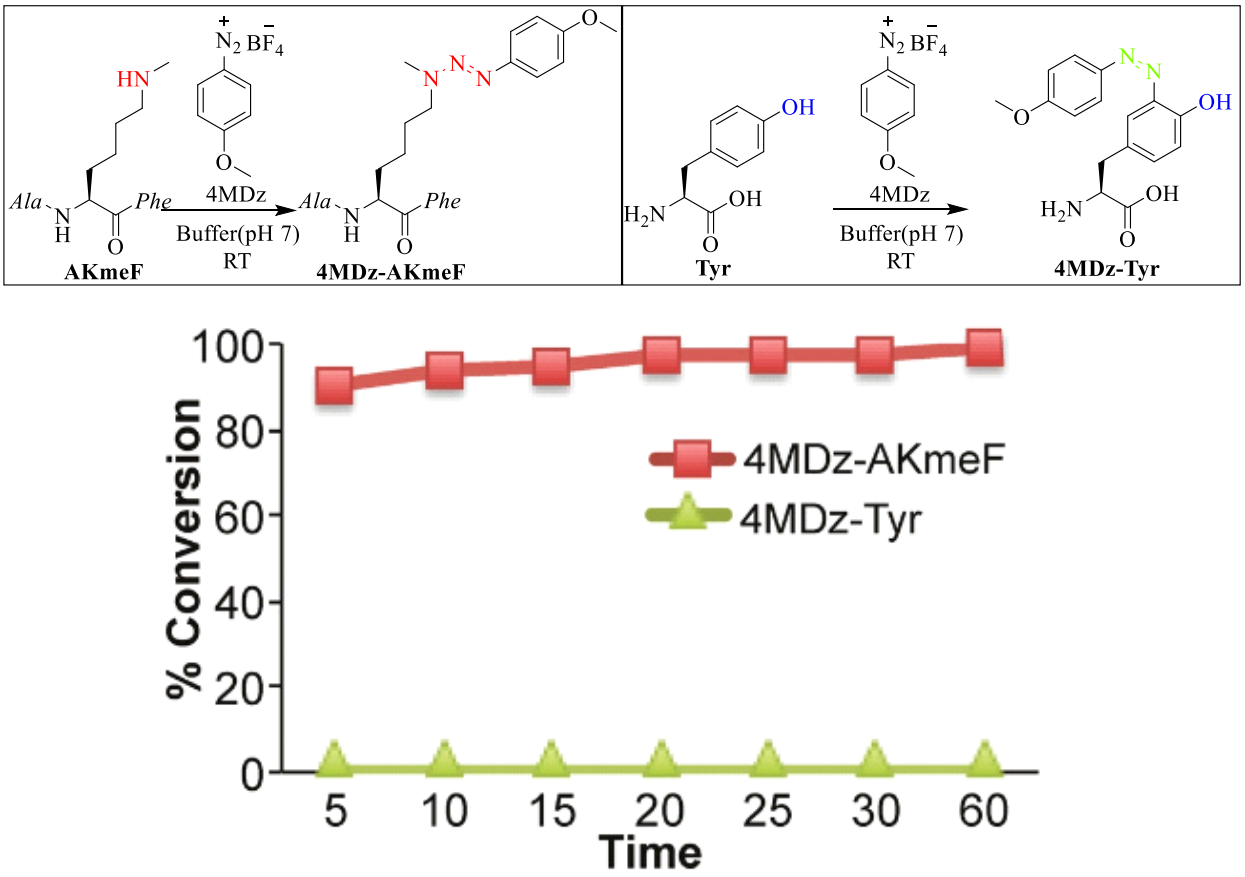


Figure 1.5: Rate of the coupling reaction of 4MDz with AKmeF and Tyr

For full structural characterization of the triazene products formed from our reaction conditions, we carried out large scale reactions between 4-methoxybenzene diazonium salt (4MDz) with different small molecules and amino acids having secondary amines including 2-(methylamino) ethanol, proline methyl ester, and a dipeptide, (NHMe)Ala-Phe-OMe. The reactions were left for 30 minutes and the formation of triazene products was confirmed by NMR (^{1}H and ^{13}C) (*Supplementary Figure 3*). We subjected the amino acid, alanine with a primary amine to the same reaction conditions but did not observe the formation of a triazene product(*Supplementary Figure 3*). This served as further proof of the literature report that triazene products formed between diazonium ions and primary amines are reversible in nature.[49]

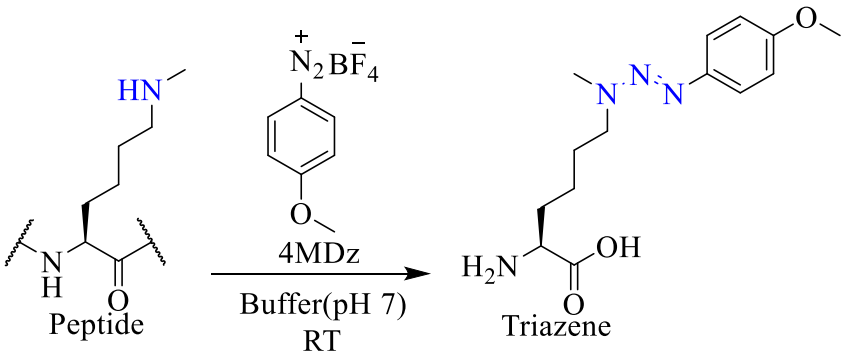
1.3.1 Chemoselectivity Studies

Next, to determine the chemoselectivity nature of the triazenation reaction, we carried out a reaction of 4MDz with varying reactive amino acids such as Ala, Arg, Asp, Asn, Cys, His, Pro, Ser, Trp, and Tyr. We also carried out reactions with the different methylated states of lysine (Kme, Kme_1, Kme_2, and Kme_3). The reactions were monitored for 30 mins, 2 hours, and 24 hours by LCMS. The data showed that the triazene product is formed only with secondary amines such as proline and monomethylated lysine (Supplementary Figure 5). These studies confirmed that the reaction between 4MDz and secondary amines is highly chemoselective.

1.3.2. Pan-Specific Nature of Triazene Reaction

With this knowledge in hand, we screened various peptides of different sizes containing secondary amine and monomethylated lysine at different positions with 4MDz and 4CDz under optimized triazenation reaction conditions to evaluate how the amino acids close to the monomethylated lysine are tolerated by our chemistry. From the data generated (Table 1.**5**), we observed that both aliphatic and aromatic amino acids when close to Kme are well tolerated by our chemistry. The peptides gave impressive conversion to the products between 85-98% (*Table 1*, HRMS-*Supplementary Figure 8*). It is worthy of note that the peptide GKmeAKmeF, which has two monomethyl lysine with one amino acid apart, produced a dual modification resulting from the generation of a triazene-coupled product with both monomethyl lysine (89%, **Table 1.4**, HRMS-*Supplementary Figure 8).* Also, the reactions with a peptide GEPGIAGFKmeGEQGPK (collagen fragment) containing reactive amino acids such as Glu, Gln, and Lys gave a desired triazene-coupling result with the excellent conversion of Kme to the triazene product (89%, **Table 1.4**,

Table 1.5: Chemoselective and pan-specific nature of STAR chemistry



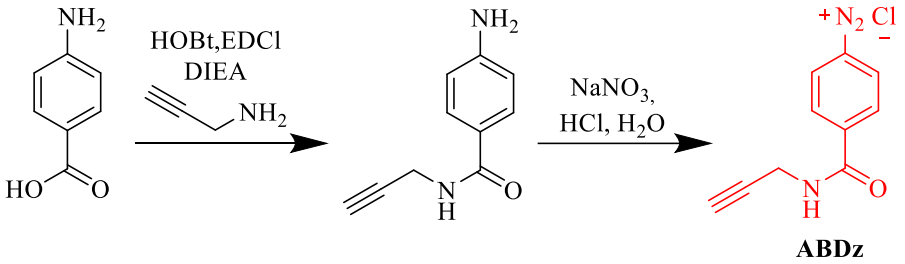
ENTRY	PEPTIDE SEQUENCE	CONVERSION (%)
1	PAF	98
2	A**Kme**F	95
3	GA**Kme**F	93
4	GDAKF	0
5	GYARF	0
6	GCAMF	0
7	GWAHF	0
8	GSANF	0
9	GG**Kme**GKF	91
10	G**Kme**A**Kme**F	89
11	**Kme$_2$**G**Kme**GKF	90
12	DRVYIHPF	0
13	GEPGIAGF**Kme**GEQPK	85

Reaction conditions: peptide (0.6 mM), 4MDz (0.6 mM), sodium carbonate (0.6 mM) in sodium phosphate buffer (0.1 mM ,pH 7.),stirred at room temperature for 1h

HRMS-*Supplementary Figure 8*). Next, we carried out the triazenation reaction on peptides with varying amino acid compositions without any secondary amine (GHAKF, GYARF, GMACF, GSAWF, GNADF) under optimized reaction conditions. We did not observe the modification of any of these peptides **(Table 1.4**, *Supplementary Figure 7).* These findings validated our chemistry's excellent chemoselectivity and pan specificity for secondary amines since we observed the modification of only monomethyl lysine in all the cases irrespective of the amino acid residues flanking the Kme. This is a game-changer in the detection of lysine methylation since antibodies as recognition molecules often require specific flanking sequences for the identification of their targets.

Following this observation, we sought to determine the selectivity and compatibility of our system for labeling Kme in a complex environment by incubating a mixture of peptides PAF, AKmeF, GAKmeF and GKmeAKmeF with 4MDz. Excitingly, the data showed complete labeling of all the peptides when the reaction mixture was analyzed by LC-MS (*Supplementary Figure 9*). Thus, these studies confirmed the robustness of our approach to label Kme in a pan-specific manner from a single peptide to a complex mixture under mild physiological conditions.

1.3.3. Synthesis of arene diazonium affinity tags for monomethyl lysine



Scheme 1.4: Synthesis of ABDz

Affinity tags have evolved into being useful biological research tools for purifying and detecting proteins of interest, as well as separating protein complexes. So, we synthesized diazonium ions with an alkyne functionalized affinity tag, 4-aminobenzoic acid-derived alkyne diazonium (ABDz). The first step involved the amidation of 4-aminobenzoic acid with propargyl amine to form alkyne functionalized aniline, **ABDz** (**Scheme 1.4**, *Supplementary Figure 10*). Treatment of the resultant product, ABDz *in situ* with sodium nitrite affords the diazotization product under acidic conditions (*Supplementary Figure 11*). Upon the formation of the diazonium ion within 10 mins, peptides were added to the reaction mixture and allowed to stir at room temperature. We went ahead and exposed five (5) peptides: GAKmeF, AKmeF, GGKmeGKF, GKmeAKmeF and Kme2GGKmeGKF to the alkyne functionalized diazonium ion generated *in situ* and we observed excellent conversion to the products in a very selective manner (*Supplementary Figure 11*).

1.3.4. Selective Tagging of Monomethyl Lysine in the Complex Mixture Using Alkyne-Derived Diazonium Ion

Next, we attempted tagging multiple peptides in the same solution to test the potential of our method for enrichment in a complex mixture. The mixture of peptides AK***me***F, GAK***me***F and PAF were incubated with ABDz for 1 hour under the optimized reaction conditions. Employing 5-azidopentanoic acid as the azide handle, the reaction mixture was subsequently subjected to the copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) for 2 hours. The reaction products were analyzed by MS and the data showed the tagging with ABDz and enrichment with 5-azidopentanoic acid for all the peptides containing secondary amines in the reaction mixture *(***Figure 1.6**, *Supplementary Figure 12).*

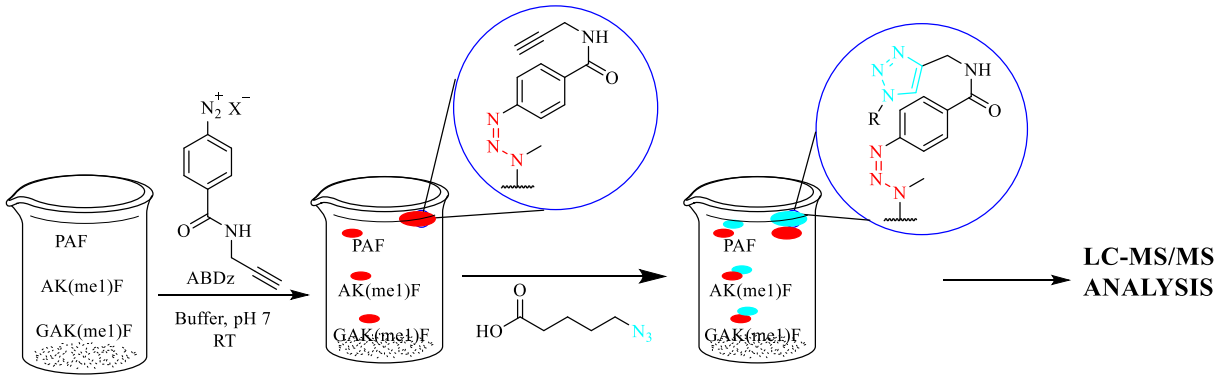


Figure 1.6: Selective tagging and enrichment of secondary amine containing peptides from a mixture of peptides by the use of alkyne-derived diazonium ion

1.3.5. Traceless Enrichment of Monomethyl Lysine Using Star

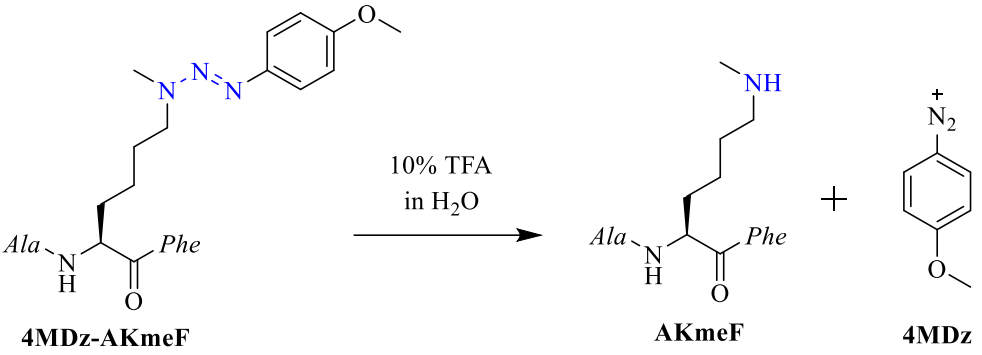


Figure 1.7: Traceless cleavage of triazene-coupled product 4MDz-AKmeF under acidic conditions

To employ the STAR chemistry for the enrichment of Kme from the proteome, we sought to understand the chemical stability of the triazene-coupling product under different conditions. It is well known in the literature that triazene product is used as a protecting group for secondary amines and tolerates various coupling reagents and bases used in solid-phase peptide synthesis (SPSS) such as N, N′-Diisopropylcarbodiimide (DIC), N, N-Diisopropylethylamine (DIEA) and piperidine (PPR). We subjected these reagents to a solution of 4MDz-Proline triazene adduct and found them stable in all the cases including strong protein disulfide reducing agent tris(2-carboxyethyl)phosphine (TCEP) and click chemistry (CuAAC) conditions (*Supplementary Figure 13*). To sodium dithionate, the triazene-adduct showed reasonable stability and was stable for 1 hour (*Supplementary Figure 14*). This contrasts with the azo-coupling product with Tyr, which degrades rapidly in the presence of sodium dithionate to produce modified tyrosine.[13,14]

Furthermore, the triazene-coupling product 4MDz-Pro-OMe was incubated in an ACN: Water mixture for 24 hours and showed no signs of degradation. Even though the triazene-secondary amine linkage is stable under physiologically relevant conditions, we wanted to see if the conjugate could be dissociated to regenerate back the partners in their unchanged nature. Much to our delight, upon the addition of acid (10% TFA in H_2O) to the triazene adduct peptide 4MDz-AKmeF, the

conjugate got hydrolyzed to the unmodified peptide in 5 minutes as determined by HPLC and MS analysis (**Figure 1.7**, *Supplementary Figure 15*). Thus, our approach is a traceless type of bioconjugation. It's worth noting that the acidic cleavage produced original coupling partners in an unaltered manner. Few bioconjugation techniques, on the other hand, are reversible, and the decoupling products are somewhat altered copies of the original reactants.[50] Our approach's impressive rate of conversion of secondary amine-containing peptides to the triazene adducts and exquisite chemoselective towards secondary amine, as well as the ability to regenerate unmodified starting materials by cleaving the resulting triazene-linkage in a traceless way, highlight the strategy's unique advantages.

1.3.6. Enrichment of Monomethyl Lysine Kme from Complex Mixture

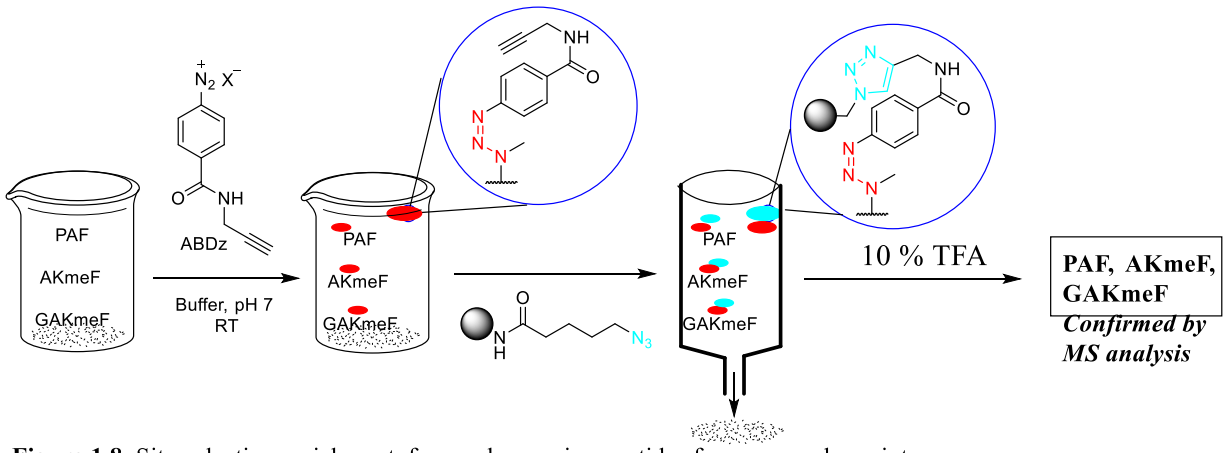


Figure 1.8: Site-selective enrichmentof secondary amine peptides from a complex mixture

Encouraged by these results, we utilized our chemistry to establish rapid, near quantitative, and site-specific enrichment of monomethyl lysine peptides from a heterogeneous mixture of proteolytic fragments. Three secondary containing peptides: PAF, AKmeF and GAKmeF were added to trypsin digested proteolytic mixture and incubated with ABDz for 1h. Thereafter, azide-functionalized resins were used as the azide handle in click chemistry (CUAAC) reaction with the reaction mixture for 2 hours. After the click chemistry reaction, different solvents were used to

wash the resin to remove nonspecific binders. The trapped peptides were eluted and recovered by the addition of acid (10% TFA in DCM) to the resin for 10 mins followed by LCMS analysis(**Figure 1.8**, *Supplementary Figure 16*). All the three peptides containing secondary amine were efficiently enriched and captured from the complex mixture.

1.3.7. Enrichment on Solid Support

We started out our studies with the overarching goal of developing a robust chemical technology for efficiently capturing and enriching Kme from the proteome. We reasoned that a potent means of complimenting enrichment on solution-phase would be to design a system in which the diazonium ion is immobilized onto a solid support. In this system, monomethyl lysine peptides directly bind to the solid support followed by their detachment from solid support in a traceless manner under acidic condition.

1.3.7.1. Synthesis of arene diazonium ion-functionalized resin

Chlorobenzylalcohol was used to make arene diazonium ion-functionalized resin in two steps (*Supplementary Figure 17*). With this in hand, a mixture of four (4) peptides AKmeF, SVF, NAF, and RAF were incubated with the arene diazonium ion-functionalized resin for 16 hours followed by washing of the resin with various solvents including ACN, DMF, MeOH and DCM to remove non-specific binders from the resin. By exploiting the sensitivity of triazene under acidic conditions, we employed 10% trifluoroacetic acid (TFA) in DCM for 10 minutes to cleave off the tagged peptide from the resin. LCMS analysis showed that only peptide AKmeF containing secondary amine attached to the resin (**Figure 1.9**, *Supplementary Figure 18*). The peptides with primary amines SVF, NAF, RAF were not trapped by the diazonium ion tethered to the resin and thus were not enriched.

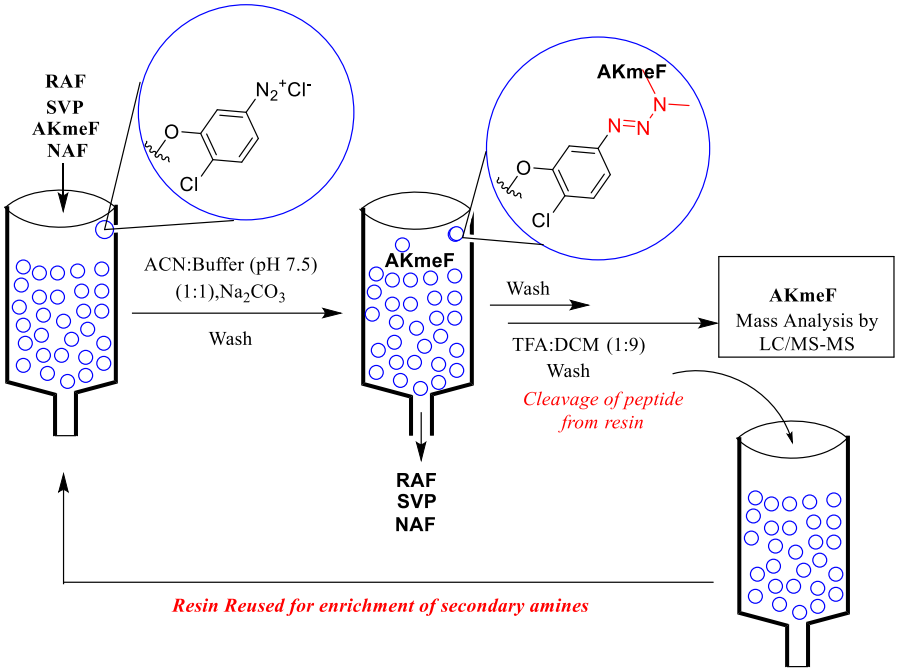


Figure 1.9: Selective enrichment of AKmeF from a mixture of peptides using an arene-diazonium-functionalized solid support

1.3.7.2. Enrichment of Peptide from Complex Mixture

Having established this protocol, we examined the compatibility of this method in a complex mixture. A protein digest was spiked with AK(***me1***)F and incubated with the diazonium ion tethered onto solid support for 16h followed by washing of the resin to remove unattached proteolytic fragments from the resin (*Figure 10*). Interestingly, the trapped peptides were released under mild circumstances (10% TFA/DCM) and characterized clearly by MS and only AKmeF was enriched from the complex mixture validating our method as being robust enough for capturing and enriching of monomethyl lysine (**Figure 1.10**, *Supplementary Figure 19*).

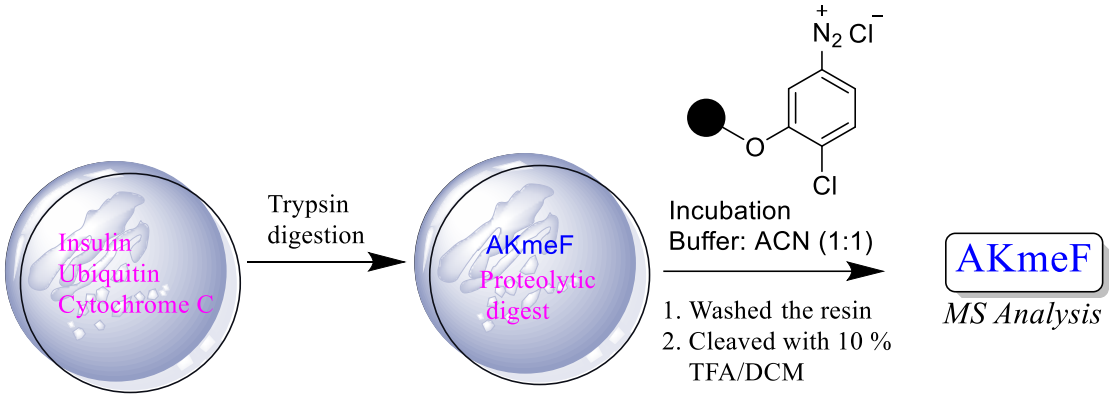


Figure 1.10. Enrichment of AKmeF peptide from a mixture of proteolytic fragments

1.3.7.3. Reusability of the diazonium functionalized resin

Next, we evaluated the reusability of the arene diazonium ion-functionalized resin for capturing monomethyl lysine. We first incubated the diazonium ion-functionalized resin for 1 hour with peptides AKmeF and GAKmeF and Na_2CO_3, then washed and cleaved the peptides from the resin under mildly acidic conditions (10 % TFA in DCM in 10 mins). AKmeF and GAKmeF peptides were efficiently enriched from the resin as confirmed by MS analysis. In the second re-use of the resin, we incubated the resin with PAF to avoid any confounding results due to the first peptide capture event. From MS analysis, PAF was efficiently enriched from the resin. In the third cycle, we incubated AKmeF and equally observed the release of the peptide in an unchanged manner from the resin (*Supplementary Figure 20*). Thus, we were able to reuse the resin three times to enrich monomethyl lysine/secondary amine peptides with the resin retaining its functional integrity.

1.3.7.4. Quantification of Peptides Enriched from the Diazonium Functionalized Resin

Then, using LCMS, the peptide trapping was quantified by incubating peptides (NHMe)AVF and GAKmeF with arene-diazonium ion resin for 2 hours at room temperature and quantifying the

reduction in peptide concentration in solution using anisole as a standard. The results showed that 56 % of peptides were trapped in 1 hour and 87 % of peptides were trapped in 2 hours (*Supplementary Figure 22*).

1.3.8. Enrichment with Biotin-functionalized Resin

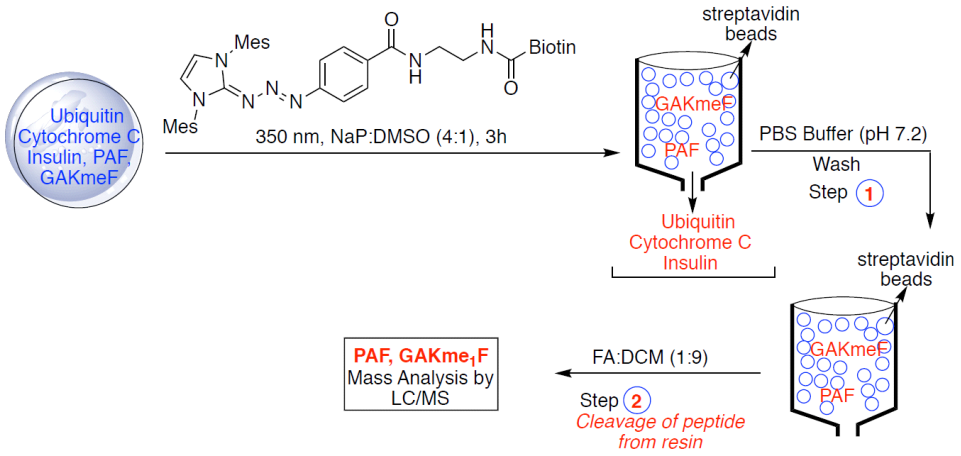


Figure 1.11. Enrichment of GAKmeF and PAF peptides with biotin functionalized resin from a mixture of proteins

Finally, we explored the compatibility of our approach with streptavidin-biotin purification method. Many proteome studies rely on streptavidin-based purifications since it is a powerful approach for identifying biotinylated biomolecules.[51] A solution of two peptides containing secondary amine, PAF and GAKmeF; cytochrome C, ubiquitin and insulin was incubated with a biotin-functionalized diazonium ion and allowed to react. Under acidic condition, the trapped peptides were released in a traceless manner from streptavidin beads and the eluent were analyzed by LCMS (**Figure 1.11**, *Supplementary Figure 21*). Excitingly, only PAF and GAKmeF were identified from the eluted solution confirming that our method is selective and powerful for enriching secondary amine containing peptides.

1.4. Proteome-wide Profiling of Monomethyl Lysine with ABDz

We applied our probe, ABDz for profiling monomethyl lysine across different cell lines. In our previous studies, the diazonium ion (ABDz) wass generated in solution, and peptides were added immediately to the reaction without isolating the probe. However, for cell studies, to ensure that the pH of the cells was not perturbed and that a definite amount of ABDz was used for the different experiments, ABDz was isolated in its salt with tetrafluoroborate as the counterion. With the diazonium salt in hand, we started our initial experiment by assessing the reactivity of Human Embryonic Kidney 293T (HEK293T) and SKOV3 cell lysates with ABDz in a concentration-dependent manner (**Figure 1.12**, Panel A). The result of the experiment showed that the proteome had the most optimal reactivity with ABDz at 100 uM. In the same manner, a time-dependent reactivity of ABDz was carried out with the whole cell lysate from the cell lines (HEK293T and SKOV3). From this experiment, it was observed that the proteomes from the two cell lines had reasonable reactivity with ABDz from 10 minutes with progressively increased reactivity to the maximal reactivity at 30 minutes (**Figure 1.12**, Panel B). This was very encouraging and sets ABDz apart as an efficient probe with fast reaction kinetics for probing the human proteome. With these results from cell lysates, we sought to determine the ideal concentration and time frame for profiling in live cells. Live SKOV3 cells were treated with ABDz in a concentration and time-dependent manner. Interestingly, we observed similar trends in live cells as was recorded in cell lysate with ideal concentration observed at 100 uM of ABDz and near-optimal time at 30 minutes of the reaction (**Figure 1.13**).

To confidently establish the reactivity of ABDz with different cell lines, whole-cell lysates from four different cell lines (HeLa, SKOV3, PC3 and A549) were treated with ABDz and DMSO in separate experiments. Excitingly, no protein bands were seen in the experiments where DMSO was substituted for ABDz (**Figure 1.14**, Panel A) ruling out nonspecific binding of ABDz with the human proteome. Next, whole-cell lysates from nine different cell lines (HEK293t, HeLa, SKOV3, PC3, MDA-MB-231, MCF7, A549, H82, K562) were treated with ABDz, and the proteome bands visualized by employing in-gel fluorescence. The result showed efficient labeling of the proteome from the nine cell lines (**Figure 1.14**, Panel B). This result establishes that ABDz has good reactivity with human proteomes from diverse cell lines. To the best of our knowledge, this is the first time, a chemical probe has been employed for the covalent tagging of monomethyl lysine from the proteome. Chemical proteomics work is ongoing toward establishing the identities of the proteins labeled by our probe and further understanding the implications of these lysine monomethylation marks in the human physiological system.

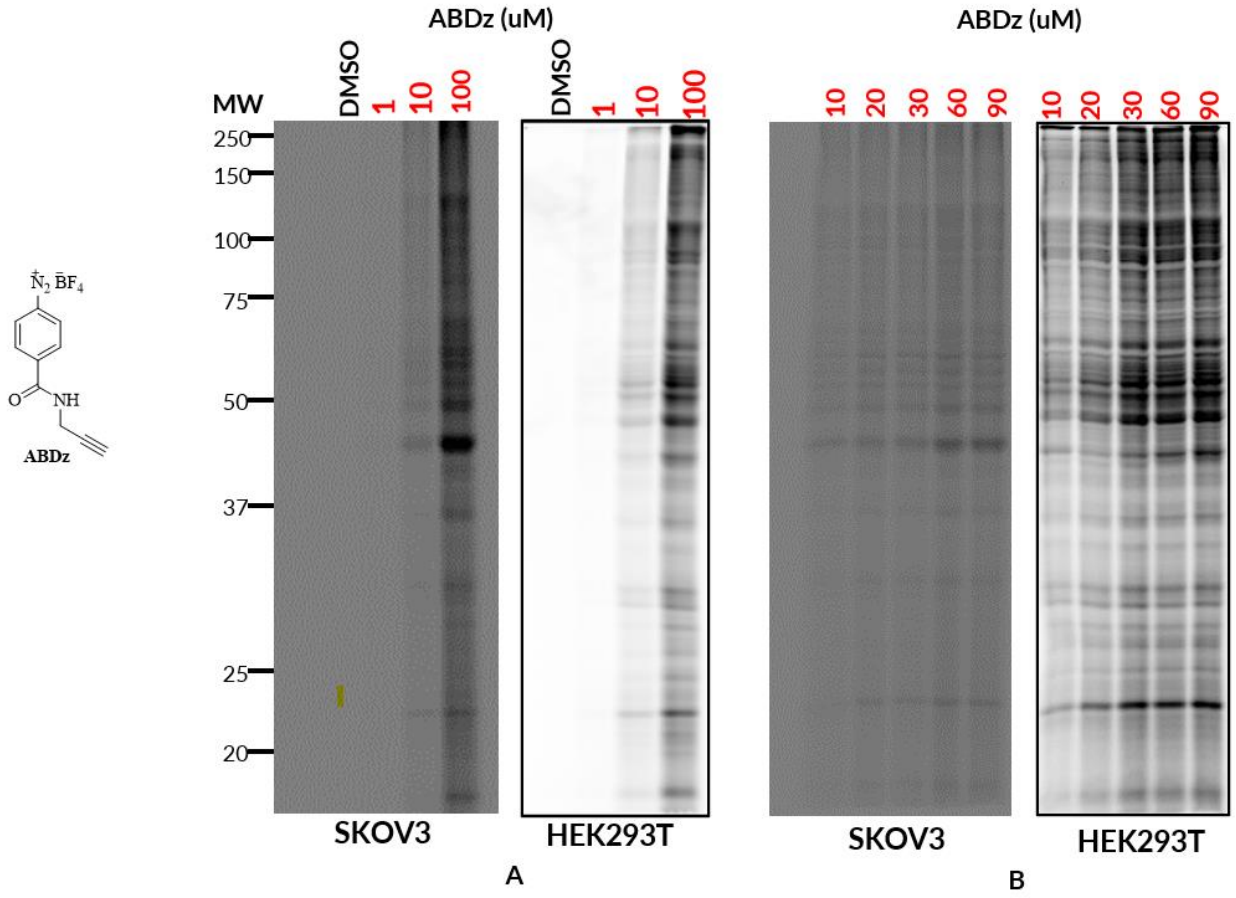


Figure 1.12: Proteome-wide profiling of monomethyl lysine. A. Concentration-dependent labeling of SKOV3 and HEK293T whole lysates treated with ABDz. B. Time-dependent labeling of SKOV3 and HEK293T whole lysates treated with ABDz.

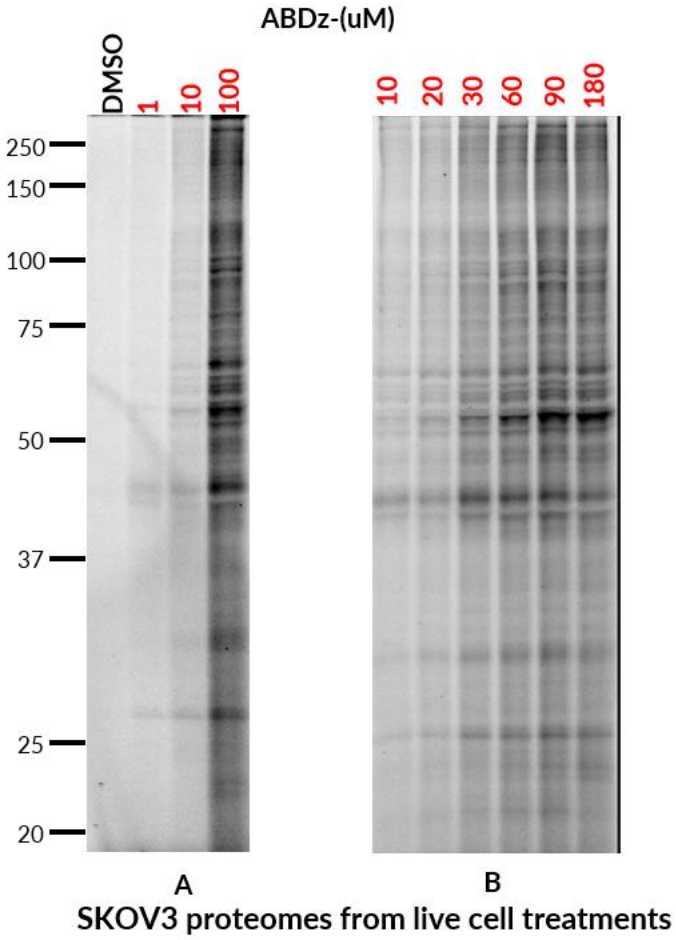


Figure 1.13: Concentration and time-dependent labeling of live SKOV3 cells treated with ABDz. A. Concentration-dependent labeling of live SKOV3 cells treated with ABDz. B. Time-dependent labeling of live SKOV3 cells treated with ABDz

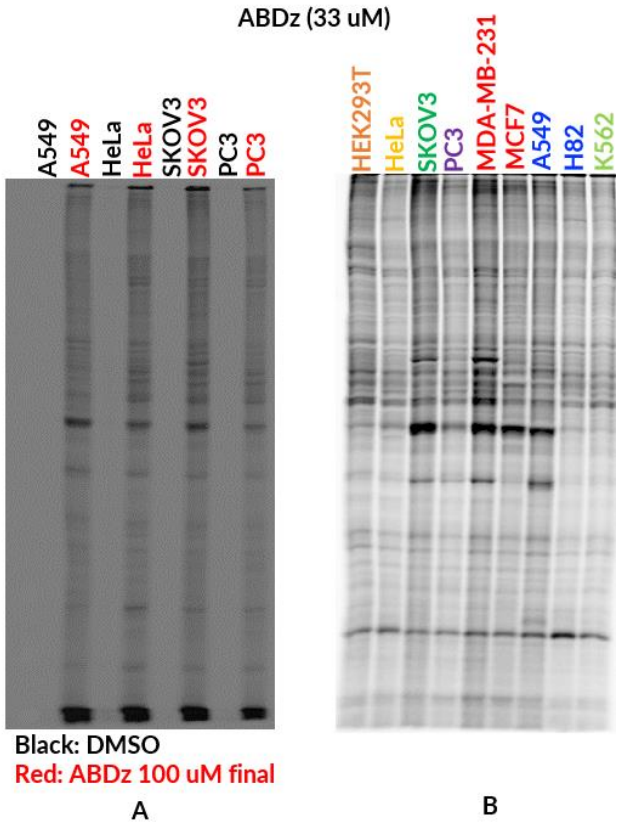


Figure 1.14: Comparison of the *in vitro* proteome-wide labeling of monomethyl lysine with ABDz. A. Monomethyl lysine profiling in A549, Hela, SKOV3 and PC3 cell lines treated with ABDz with DMSO as control. B. Profiling of monomethyl lysine with ABDz in nine cell lines treated with ABDz.

CONCLUSION

To effectively study and understand the role of protein lysine monomethylation in cell signaling and pathologies, there is a need for a chemoselective method that can covalently tag these methyl marks and enrich them even in their low abundant states. We started our studies by further developing the Secondary Amine Selective Petasis (SASP) bioconjugation strategy for efficient labeling of monomethyl lysine by incorporating the aldehyde and organoboron into one molecule called STRaP reagent. The SASP bioconjugation is a multicomponent reaction between the secondary amine, aldehyde, and organoboron; thus, it suffered from a low rate of reaction and

consequently an inability to detect the low abundance monomethyl lysine. We had anticipated that the STRaP reagents would increase the rate of trapping monomethyl lysine to enable the detection of low abundant monomethyl lysine from a complex mixture in an efficient manner. However, the STRaP reagents we synthesized suffered from chemoselectivity towards secondary amine against primary amine.

With this setback, we diverted our attention to the development of Selective Triazenation Reaction (STaR) of Secondary Amines for Tagging Monomethyl Lysine Post-Translational Modifications. STAR involves the formation of a stable triazene between monomethyl lysine and electron-rich arene diazonium ion. By carefully exploiting the nature and concentration of arene diazonium ion and pH, a robust approach was developed that selectively modifies secondary amines such as monomethyl lysine over unmodified lysine and other reactive groups on proteins. The reaction is biocompatible and occurs under mild conditions characterized by rapid kinetics, generation of stable products, pan-specific and traceless in the sense that the reacting partners can be decoupled following successful bioconjugation unchanged.

These excellent attributes make our method an elegant approach for proteome-wide profiling of ‘lysine monomethylome’. Our method directly gives residue-specific information (shows the location of the methylation event) instead of total protein methylation. This is important because knowing that a protein is methylated is very necessary, however, addressing specific biological questions often requires the detection of the site of methylation. Our method provides the tool for understanding the diverse roles the monomethylation of lysine can play in the biological system. This is possible because our chemical probes have the potential to add an affinity tag that can be functionalized for interrogating the biological system. Success with this will lead to the possible identification of monomethylated proteins involved in different pathogenesis and those can be used

as biomarkers to monitor disease onset, progression, and efficacy of drug treatment. With these chemical tools, it would be possible to profile monomethylation between disease and non-disease states.

Excitingly, our probe demonstrated reactivity with human proteomes from different cell lines and chemical proteomics work is currently ongoing toward establishing the identities of the proteins labeled by our probe and further understanding the implications of these lysine monomethylation marks in the human physiological system.

CHAPTER ONE REFERENCES

1. Zhang, X.; Wen, H.; Shi, X. Lysine Methylation : Beyond Histones Modifying Enzymes for Lysine Methylation. *Acta Biochim. Biophys. Sin.* **2012**, 14-27.

2. Allis, C. D.; Berger, S. L.; Cote, J.; Dent, S.; Jenuwien, T.; Kouzarides, T. New Nomenclature for Chromatin-Modifying Enzymes. *Cell*, **2007**, *131*, 633-636.

3. Qian, C.; Zhou, M. M. SET Domain Protein Lysine Methyltransferases: Structure, Specificity and Catalysis. *Cell and Mol Life Sci*. **2006**, *63(23)*, 2755-2763.

4. Wu, Z.; Connolly, J.; Biggar, K. K. Beyond Histones – The Expanding Roles of Protein Lysine Methylation. *FEBS J*. **2017**, *284*, 2732–2744.

5. Greer, E. L.; Shi, Y. Histone methylation: A dynamic mark in health, disease and inheritance, *Nat Rev Genet.*, **2012**, *13(5)*, 343–357.

6. Sidoli, S.; Cheng, L.; Jensen, O. N. Proteomics in Chromatin Biology and Epigenetics: Elucidation of Post-translational Modifications of Histone Proteins by Mass Spectrometry. *J Proteomics*. **2012**, *75(12)*, 3419-3433.

7. Ko, S.; Ahn, J.; Song, C. S.; Kim, S.; Knapczyk-Stwora, K.; Chatterjee, B. Lysine Methylation and Functional Modulation of Androgen Receptor by Set9 Methyltransferase. *Mol Endocrinol*. **2011**, *25(3)*, 433–444.

8. Huang, J.; Berger, S, L. The Emerging Field of Dynamic Lysine Methylation of Non-histone Proteins. *Curr Opin Gen Dev*. **2008**, *18(2)*, 152-158.

9. Schaefer, W. H.; Lukas, T. J.; Blair, I. A.; Schultz, J. E.; Watterson, D. M. Amino acid sequence of a novel calmodulin from Paramecium tetraurelia that contains dimethyllysine

in the first domain. *J Biol Chem* **1987**, *262* (3), 1025-1029.

10. S Wang, C.; Lazarides, E.; O'Connor, C.; Clarke, S. Methylation of chicken fibroblast heat shock proteins at lysyl and arginyl residues. *Journal of Biological Chemistry* **1982**, *257* (14), 8356-8362.

11. Iwabata, H.; Yoshida, M.; Komatsu, Y. Proteomic analysis of organ-specific post-translational lysine-acetylation and-methylation in mice by use of anti-acetyllysine and-methyllysine mouse monoclonal antibodies. *Proteomics* **2005**, *5* (18), 4653-4664.

12. Lanouette, S.; Mongeon, V.; Figeys, D.; Couture, J.-F. The functional diversity of protein lysine methylation. *Molecular systems biology* **2014**, *10* (4), 724-724. DOI: 10.1002/msb.134974 PubMed.

13. Gober, I. N.; Waters, M. L. Supramolecular Affinity Labeling of Histone Peptides Containing Trimethyllysine and Its Application to Histone Deacetylase Assays. *J Am Chem Soc*. **2016**; *138(30),* 9452-9459.

14. Liu, Q.; Dreyfuss, G. In Vivo and in Vitro Arginine Methylation of RNA-binding Proteins. *Mol Cell Biol*. **1995**; *15(5),* 2800-2808

15. Ong, S. E.; Mittler, G.; Mann, M. Identifying and Quantifying in Vivo Methylation Sites by Heavy Methyl SILAC. *Nat Methods*. **2004**; *1(2),* 119-126

16. Uhlmann, T.; Geoghegan, V. L.; Thomas, B.; Ridlova, G.; Trudgian, D. C.; Acuto, O. A. Method for Large-scale Identification of Protein Arginine Methylation. *Mol Cell Proteomics*. **2012**; *11(11),* 1489-1499.

17. Côté, J.; Boisvert, F. M.; Boulanger, M. C.; Bedford, M. T.; Richard, S. Sam68 RNA

Binding Protein Is an in Vivo Substrate for Protein Arginine N-methyltransferase 1. *Mol Biol Cell*. **2003**; *14(1),* 274-287.

18. McBride, A. E.; Silver, P. A. State of the Arg: Protein Methylation at Arginine Comes of Age. *Cell*. **2001**. *106(1),* 5-8.

19. Hamamoto, R.; Saloura, V.; Nakamura, Y. Critical roles of non-histone protein lysine methylation in human tumorigenesis. *Nat Rev Cancer*, **2015,** *15(2),* 110-124.

20. Hamamoto, R.; Furukawa, Y.; Morita, M.; Iimura, Y.; Silva, F. P.; Li, M. SMYD3 Encodes a Histone Methyltransferase Involved in the Proliferation of Cancer Cells. *Nat Cell Biol*. **2004**, *6(8),* 731-740.

21. Kotake, Y.; Cao, R.; Viatour, P.; Sage, J.; Zhang, Y.; Xiong, Y. pRB Family Proteins Are Required for H3K27 Trimethylation and Polycomb Repression Complexes Binding to and silencing p16INK4a Tumor Suppressor Gene. *Genes Dev*. **2007,** *21(1),* 49-54.

22. Hamamoto, R.; Silva, F. P.; Tsuge, M.; Nishidate, T.; Katagiri, T.; Nakamura, Y. Enhanced SMYD3 Expression Is Essential for the Growth of Breast Cancer Cells. *Cancer Sci*. **2006,** *97(2),* 113-118.

23. Ferguson, A. D.; Larsen, N. A.; Howard, T.; Pollard, H.; Green, I.; Grande, C. Structural basis of Substrate Methylation and Inhibition of SMYD2. *Structure*. **2011,** *19(9),* 1262-1273.

24. Li, L. X.; Fan, L. X.; Zhou, J. X.; Grantham, J. J.; Calvet, J. P.; Sage, J. Lysine Methyltransferase SMYD2 Promotes Cyst Growth in Autosomal Dominant Polycystic Kidney Disease. *J. Clin. Invest.* **2017**, *127(7),* 2751–2764.

25. Casciello, F.; Windloch, K.; Gannon, F.; Lee, J. S. Functional role of G9a histone methyltransferase in cancer. *Front Immunol*. **2015**, *25(6)*, 487.

26. Moore, K. E.; Gozani, O. An Unexpected Journey-Lysine Methylation Across the Proteome. *Biochim. Biophys. Acta*. **2014**, *1839(12)*, 1395-1403.

27. Guo, A.; Gu, H.; Zhou, J.; Mulhern, D.; Wang, Y.; Lee, K. A. Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Mol Cell Proteomics*. **2014**, *13(1)*, 372-387.

28. Bock, I.; Dhayalan, A.; Kudithipudi, S.; Brandt, O.; Rathert, P.; Jeltsch, A. Detailed Specificity Analysis of Antibodies Binding to Modified Histone Tails with Peptide Arrays. *Epigenetics*. **2011**, *6(2)*, 256-263.

29. Fuchs, S. M.; Krajewski, K.; Baker, R. W.; Miller, V. L.; Strahl, B. D. Influence of Combinatorial Histone Modifications on Antibody and Effector Protein Recognition. *Curr Biol*. **2011**, *21(1)*, 53-58.

30. Rothbart, S. B.; Dickson, B. M.; Raab, J. R.; Grzybowski, A. T.; Krajewski, K.; Guo, A. H. An Interactive Database for the Assessment of Histone Antibody Specificity. *Mol Cell*. **2015**, *59(3)*, 502-511.

31. Blow, N. The generation game. *Nature*. **2007**. *447(7145)*, 741-744.

32. Nady, N.; Krichevsky, L.; Zhong, N.; Duan, S.; Tempel, W.; Amaya, M. F. Histone Recognition by Human Malignant Brain Tumor Domains. *J. Mol. Biol*. **2012**; *423(5)*, 702-718.

33. Carlson, S. M.; Moore, K. E.; Green, E. M.; Martín, G. M.; Gozani, O. Proteome-wide

Enrichment of Proteins Modified by Lysine Methylation. *Nat. Protoc*. **2014**; *9(1)*, 37-50.

34. Carlson, S, M.; Gozani, O. Emerging Technologies to Map the Protein Methylome, *J Mol Biol.* **2014**; *426(20)*, 3350-3362.

35. Ingerman, L. A.; Cuellar, M. E.; Waters, M. L. A Small Molecule Receptor that Selectively NRecognizes Trimethyl Lysine in a Histone Peptide with Native Protein-like affinity. *Chem Commun.* **2010**; ***46,*** 1839-1841.

36. Katherine, I. A.; Mackenzie, W. K.; Christopher, J. P.; Madison, M. P.; Brian D. S.; Eric, M. B.; Marcey, L. W. 2020. Engineered Reader Proteins for Enhanced Detection of Methylated Lysine on Histones. *ACS Chem. Biol.* **2020,** *15 (1)*, 103-111.

37. Wu, Z.; Cheng, Z.; Sun, M.; Wan, X.; Liu, P.; He, T. A Chemical Proteomics Approach for Global Analysis of Lysine Monomethylome Profiling. *Mol Cell Proteomics*. **2015**;14(2):329-339.

38. Islam, K.; Bothwell, I.; Chen, Y.; Sengelaub, C.; Wang, R.; Deng, H. Bioorthogonal profiling of protein methylation using azido derivative of S -adenosyl- l -methionine. *J Am Chem Soc*. **2012**; *134*, 5909-5915.

39. Islam, K.; Chen, Y.; Wu, H.; Bothwell, I. R.; Blum, G. J.; Zeng, H. Defining Efficient Enzyme-cofactor Pairs for Bioorthogonal Profiling of Protein Methylation. *Proc Natl Acad Sci U S* A. **2013**, *110(42)*, 16778-16783.

40. Peters, W.; Willnow, S.; Duisken, M.; Kleine, H.; Macherey, T.; Duncan, K. E. Enzymatic Site-specific Functionalization of Protein MethyltransferaseSsubstrates with Alkynes for Click labeling. *Angew Chem Int Ed Engl*. **2010**, *49(30)*, 5170-5173.

41. Binda, O.; Boyce, M.; Rush, J. S.; Palaniappan, K. K. A Chemical Method for Labeling Lysine Methyltransferase Substrates. *Chembiochem*. **2011**, *12(2),* 330-334.

42. Carlson, S. M.; Gozani, O. Nonhistone Lysine Methylation in the Regulation of Cancer Pathways. *Cold Spring Harb Perspect Med*. **2016**; *6(11),* 1–16.

43. Sim, Y. E.; Nwajiobi, O.; Mahesh, S.; Cohen, R. D.; Reibarkh, M. Y.; Raj, M. Secondary amine selective Petasis (SASP) bioconjugation. Chemical Science 2020, 11 (1), 53-61, 10.1039/C9SC04697F. DOI: 10.1039/C9SC04697F.

44. Kabalka, G. W.; Venkataiah, B.; Dong, G. The use of potassium alkynyltrifluoroborates in Mannich reactions. Tetrahedron Letters 2004, 45 (4), 729-731. DOI: https://doi.org/10.1016/j.tetlet.2003.11.049.

45. Petasis, N. A.; Boral, S. One-step three-component reaction among organoboronic acids, amines and salicylaldehydes. Tetrahedron Letters 2001, 42 (4), 539-542. DOI: https://doi.org/10.1016/S0040-4039(00)02014-1.

46. Yadav, J. S.; Reddy, B. V. S.; Lakshmi, P. N. Ionic liquid accelerated Petasis reaction: A green protocol for the synthesis of alkylaminophenols. Journal of Molecular Catalysis A: Chemical 2007, 274 (1), 101-104. DOI: https://doi.org/10.1016/j.molcata.2007.04.026.

47. Petasis, N. A.; Goodman, A.; Zavialov, I. A. A new synthesis of α-arylglycines from aryl boronic acids. Tetrahedron 1997, 53 (48), 16463-16470. DOI: https://doi.org/10.1016/S0040-4020(97)01028-4.

48. Nwajiobi, O.; Mahesh, S.; Streety, X.; Raj, M. Selective Triazenation Reaction (STaR) of Secondary Amines for Tagging Monomethyl Lysine Post-Translational Modifications. Angew Chem Int Ed Engl 2021, 60 (13), 7344-7352. DOI: 10.1002/anie.202013997.

49. Sonousi, A.; Crich, D. Selective Protection of Secondary Amines as the N-Phenyltriazenes. Application to Aminoglycoside Antibiotics. Organic Letters 2015, 17 (16), 4006-4009. DOI: 10.1021/acs.orglett.5b01902.

50. Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. Journal of the American Chemical Society 2017, 139 (34), 11670-11673. DOI: 10.1021/jacs.7b05125.

51. Chen, S.; Tsao, M.-L. Genetic Incorporation of a 2-Naphthol Group into Proteins for Site-Specific Azo Coupling. Bioconjugate Chemistry 2013, 24 (10), 1645-1649. DOI: 10.1021/bc400168u.

CHAPTER TWO: RAPID ARENE TRIAZENE CHEMISTRY FOR MACROCYCLIZATION

2.0. Introduction

Polypeptide chains with a cyclic ring structure are known as cyclic peptides[1]. Cyclic peptides generally have a molecular weight of 500 to 2000 Da and are made up of 5 to 14 amino acids[2]. Cyclic peptides have conformational rigidity, receptor selectivity, and biochemical stability, and some cyclic peptides are membrane-permeable, imparting drug-like properties to them. When compared to a corresponding linear sequence, cyclization often results in a 10-1000-fold increase in relative affinity.[3] Macrocyclic peptides are an intriguing molecular format for drug discovery, combining the benefits of small molecules and biological therapeutics: low immunogenicity, synthetic accessibility and toxicity, high binding affinity and selectivity, and the ability to target protein surfaces previously thought to be "undruggable." Because of their propensity to disrupt protein-protein interactions, cyclic peptides have gotten a lot of interest in recent years.

2.1.Peptide-based Drugs and Advantages of Cyclic Peptides

Peptides and peptidomimetics are gaining popularity in the chemical and pharmaceutical industries as promising sources of therapeutic candidates and/or biological instruments.[4] When compared to small molecule drugs, peptides are less hazardous and do not accumulate in organs. Peptide medicines may be less toxic since they are swiftly removed after acting on target molecules due to proteolytic breakdown. Remarkably, the breakdown products are just amino acids with no toxicity[5]. Also, peptides can work very selectively on their targets since their interactions with them are quite specific[6]. Due to these great attributes, peptides were widely seen as the medications of the future from the 1960s through the end of the Millennium.[5] In 2020, the peptide therapeutics

market was estimated to be worth USD 28,510.60 million, with a CAGR of 9.66 percent projected during the forecast period.

However, linear peptides are generally considered weak drug candidates because of the following reasons: first, peptides are poorly absorbed through the digestive tract, making them poor candidates for oral absorption. As a result, injection remains the preferred method of delivery. Second, peptides that are successfully ingested are metabolized by proteolytic enzymes. Third, peptides seldom pass the cell membrane, rendering them unsuitable for cytoplasmic targets[1].

Cyclic peptides have even better pharmacological attributes compared to their linear counterparts: Because of their conformational rigidity, cyclic peptides have more biological activity than their linear counterparts. This is so because the stiffness of cyclic peptides reduces the Gibbs free energy's entropy term, allowing for improved binding to target molecules or receptor selectivity[7]. This is because, in aqueous solvents, linear unconstrained oligopeptides are disordered: they can rapidly sample a range of conformations and consequently exist in high entropy states. However, macrocyclization induces conformational constraint and functional group preorganization, which facilitates target engagement by reducing entropic loss due to binding[8]. Thus, the interaction between the target protein and macrocyclic peptide is stronger and more selective. Also, cyclic peptides are more stable to proteolytic degradation than linear peptides. Part of the reason for this is the fact that most cyclic peptides lack both amino and carboxyl termini and the structure is less flexible than linear peptides[1].

Another feature that makes cyclic peptides an attractive drug molecule is their ability to act as inhibitors of protein-protein interactions. Cyclic peptides in the 700-2000 molecular weight range have sizes that are typically 3-5 times larger than conventional small-molecule drugs and a

balanced conformational flexibility/rigidity, allowing them to bind to the flat PPI surfaces with affinity and specificity.[9] PPI interfaces typically contain α-helices, hence stapled α-helical peptide ligands that disrupt such PPIs can be rationally synthesized.

In addition to this, some cyclic peptides such as cyclosporin A can cross the cell membrane.[10] Cyclosporin A was one of the foremost cyclic peptides that increased interest in the medicinal property of cyclic peptides. Cyclosporin A is a cyclic peptide that was isolated from a fungus *Tolypocladium inflatum* and used globally used as an immunosuppressant to prevent graft rejection. In the USA alone, cyclosporin A generates an annual sale of more than US$ 1.0 billion[11]. In the last two decades, 28 new peptide-based drugs have gotten clinical approval worldwide while many others are still in development—more than 200 peptides are still in the preclinical stage, and about 170 are at various stages of clinical trials[12].

2.2. Synthetic Methods for Generating Macrocyclic Peptide

The large range of uses for cyclic peptides has sparked interest in the search for efficient and feasible macrocyclization methodologies from both academia and industry.[13] Peptides can be cyclized head-to-tail, head-to-sidechain, sidechain-to-tail or sidechain-to-sidechain (stapled peptide). Lactamization is the commonest means of macrocyclization reaction for one-component head-to-tail cyclic peptides in which the C-terminal carboxylate and N-terminal amino group are linked together to form a constrained ring. Many physiologically active cyclic peptides are generated through cyclization N-to-C (or head-to-tail) through Lactamization. The latter, sidechain-to-sidechain cyclization has been extensively used to stabilize secondary structures, such as α-helices and β-sheets, yielding so-called ‘stapled peptides[14] and they have been successfully applied as protein-protein interaction inhibitors.

Because the repertoire of amino acids and ways of creating cyclic peptides is broad, the synthetic technique can generate more versatile cyclic peptide compounds. Solid-phase peptide synthesis has been employed to make cyclic peptides. Typically, cyclic peptides are made through Fmoc-based solid phase peptide synthesis and the macrocyclization step carried out either in solution or on-resin. Oligomerization due to competition between intermolecular and intramolecular reactions poses a challenge to the cyclization step leading to side products. To circumvent this, most macrocyclization reactions in solution phase are carried out under diluted reaction conditions. Equally, on-resin macrocyclization overcomes this limitation by affording pseudo diluted conditions as well as facilitates the purification of the cyclopeptide.[13]

A number of stapling (two-component) strategies have equally been developed which rely on external linkers such as lysine and cysteine residues as nucleophiles.[15,16] Also, Yudin and co-workers successfully carried out three-component macrocyclization affording aziridine- and oxadiazole-containing peptide macrocycles. In one approach, they employed a linear peptide, an aldehyde and (N-isocyanimino) triphenyl phosphorane to generate a macrocyclic compound with a 1,3,4-oxadiazole grafted within the peptide backbone.

A lot of effort has been made towards the development of chemoselective reactions for peptide macrocyclization. The orthogonality conferred by these chemoselective reactions obviate the need for protecting group strategies. Some of these chemoselective reactions include internal disulfide bond formation[17], thioether formation[18], ring-closing metathesis[19] as well as catalyzed cycloaddition of azides to alkynes[20], click chemistry[21], dichloroacetone for cyclization[22], thiol–ene reaction[23], Petasis-boron-Mannich reactions[24], photochemical methods[25], etc. Recently, our group developed an innovative intramolecular chemoselective cyclization technique “CyClick” for the

synthesis of different cyclic peptides at high concentrations without generating any dimers or oligomers as a side product.[26]

2.3. Need for Cyclic Peptides for Drug Discovery

Although numerous approaches for the chemoselective synthesis of stable cyclic peptides are available, further work is needed to create synthetic methodologies that give the peptides a more drug-like character. For example, there are very few cyclic peptides that respond to changes in the external stimulus by opening and reopening the macrocycle ring. Bandyopadhyay and Gao refer to such cyclic peptides as "smart" and among other things can enable targeted drug delivery or serve as reporters of local environment.[27] Apart from the method they demonstrated for responsive peptide cyclization via spontaneous and reversible formation of intramolecular iminoboronates[27] and disulfide chemistry[28], other strategies for peptide macrocyclization form permanent linkages and thus do not respond to biological stimuli.

Also, there has been an earnest interest in the introduction of reporter groups into cyclic peptides, most notably in the context of chromophores to provide an emissive or visible spectral signature that not only confirms successful peptide cyclization but as well aids in screening and target identification.[29] Even more valuable are cyclic peptides that generate inbuilt chromophores. This characteristic is very desirable since it eliminates the need for cyclic peptides to contain a large bulky chromophoric group, which can affect their structure, binding affinity, and solubility. Unfortunately, only one method in literature has been reported to achieve this which involved the condensation of ortho-phthalaldehyde (OPA) with a primary amine and a thiol to produce a cyclic peptide with a fluorescent 1-thio-2H-isoindole.[29]

Equally, many macrocyclization strategies suffer from oligomerization due to low rates of the reaction. Thus, macrocyclization strategies that have fast kinetics could lead to the synthesis of cyclic peptide libraries in one pot without any cross-linking byproducts. Thus, there is a great need to develop a single methodology for peptide macrocyclization that fulfill all the above criteria for drug discovery research and development. This need led us to develop an efficient peptide macrocyclization strategy dubbed 'Rapid Arene Triazene Chemistry for Macrocyclization'.

2.5.0. Development of a Rapid Arene Triazene Chemistry for Macrocyclization

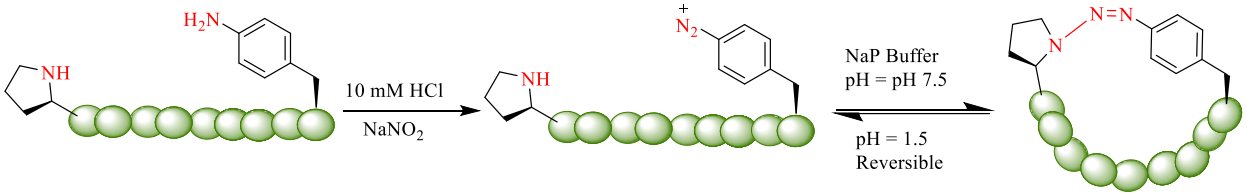


Figure 2.1: Rapid Arene Triazene Chemistry for Macrocyclization Strategy

Arene diazonium ion have been reported to react with a secondary amine to generate a stable triazene under physiological conditions.[30,31,32.] In our hands, we developed an empowering chemical technology for selective triazenation reaction of secondary amines using arene diazonium salts to achieve highly selective, rapid, and robust tagging of mono methyl lysine peptides from a complex mixture under biocompatible conditions.[33] Thus, we got inspired by the great attributes of this reaction and we considered using this chemistry to synthesize macrocyclic peptides. We rationalized that the incorporation of commercially available p-amino phenylalanine (*pAF*) in the peptide chain could serve as a precursor giving easy access to the arene diazonium ion under physiological conditions. Arene diazonium ion would then chemoselectively react with secondary amines such as N-terminal proline at pH-7.5 to generate a stable triazene macrocycle with an inbuilt chromophore at the site of macrocyclization. Since triazene peptide macrocycle is labile to acidic conditions, the macrocycle can be opened when needed. This approach represents a rare example of external stimuli controlled reversible cyclic peptides with inbuilt chromophore that

might offer a general strategy for the efficient synthesis of triazene-fused cyclic peptides. The triazene moiety formed by secondary amines is a recognized structural motif found in a wide variety of medicines and bioactive compounds as prodrugs with anti-tumor and mutagenesis[34] activities such as mitozolomide[34], dacarbazine[35], and temozolomide.[36]

2.5.1. Initial Studies and Structural Characterization of Triazene Cyclic Peptides

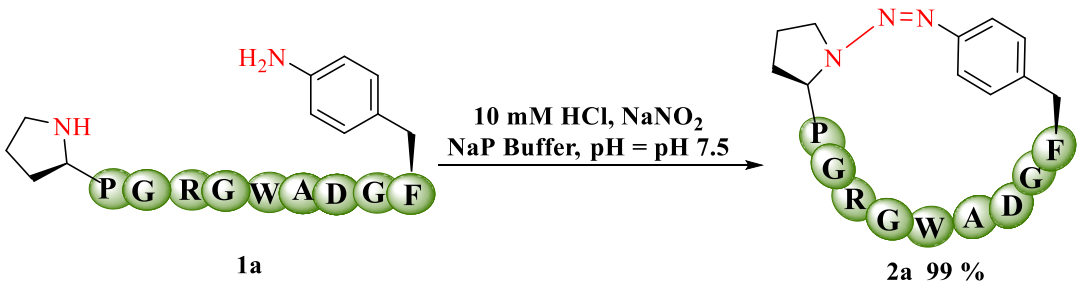


Figure 2.2: Rapid Arene Triazene Chemistry for Macrocyclization One-pot Scheme

As a proof of concept, using the standard Fmoc SPSS, we synthesized a 10-mer peptide with the linear sequence of PGRGWADGA(pAF) **1a**. The *pAF* residue present on the sequence generated arene diazonium ions immediately upon the addition of 10 mM HCl and sodium nitrite ($NaNO_2$). Upon adjusting the pH of the solution to 7.5, we observed a fast formation of the head-to-side chain triazene cyclic product **2a** with N-terminal proline with >99% conversion in 5 min without the formation of any byproducts (**Figure 2.2**, *Supporting Information Figure S1*). High-Performance Liquid Chromatography (HPLC) and High-Resolution Mass Spectrometry (HRMS) were used to confirm the formation of the triazene cyclic peptide **2a** (*Supporting Information Figure S1 and S2*). It is noteworthy that the peptide was completely unprotected and had reactive residues such as Asp (D), Arg (R), and Trp (W) and we did not observe modification on any of those. Also, the conditions employed were very mild without any requirements for coupling reagents, organic solvents, metal catalysts, and high temperatures.

For elaborate structural characterization, we cyclized 8-mer linear peptide PAMGLAE(pAF) **1b** on a large scale and subjected it to our cyclization conditions resulting in the formation of triazene cyclic peptide **2b** with >99% conversion as analyzed by HPLC (*Supporting Information Figure S1 and S2*) and isolated the pure triazene cyclic peptide **2b** in 58% yield. We carried out Nuclear Magnetic Resonance (NMR) spectroscopy analysis as a more re-assuring way of confirming the formation of the triazene cyclic peptide since the *m/z* of the intermediate diazonium ion is equal to *m/z* of triazene cyclic peptide (*Supporting Information Table S2*). We elected to compare the NMR spectra of the linear PAMGLAE(pAF) peptide **1b** with those of cyclic peptide **2b** (*Supporting Information Figure S2*). We reasoned that the NH proton of proline in the linear peptide **1b** could be diagnostic in confirming the formation of triazene cyclic peptide and we observed it at δ 2.32–2.24 in the linear peptide which disappeared in the ^{1}H NMR spectrum of the triazene cyclic **2b** product. Also, the diagnostic alpha protons (2.44–2.33 ppm) and alpha carbon (30.06 ppm) on the proline in the triazene cyclic peptide **2b** moved upfield as compared to the linear peptide **1b** alpha protons (3.24–3.17, ppm) and alpha carbon (42.31 ppm). Likewise, significant changes were observed in the ^{1}H and ^{13}C NMR of *p*-NH_2-phenylalanine in cyclized **2b** and linear peptide **1b** (*Supporting Information Figure S2).*

For further confirmation of the formation of the triazene cyclic product, we synthesized a small triazene cyclic peptide PGG(pAF), **2c** and got 53% yield. We went ahead to characterize it using HPLC, MS, and NMR (*Supporting Information Figure S3).* We observed similar trends when the NMR of linear peptide PGG(pAF) **1c** was compared with that of cyclized peptide **2c** like we obtained in the ^{1}H and ^{13}C NMR spectra of **2b**. Since 2D-NMR give more detailed structural information, we employed Heteronuclear Single-Quantum Correlation (HSQC), Heteronuclear Multiple-Bond Correlation (HMBC), and Rotating-frame Overhauser Enhancement Spectroscopy

(ROESY) NMR experiments to confirm the formation of the triazene cyclic peptides (*Supporting Information Figures S2 and S3*).

2.5.2. Investigation into the Reactivity of Primary Amines with Diazonium Ion

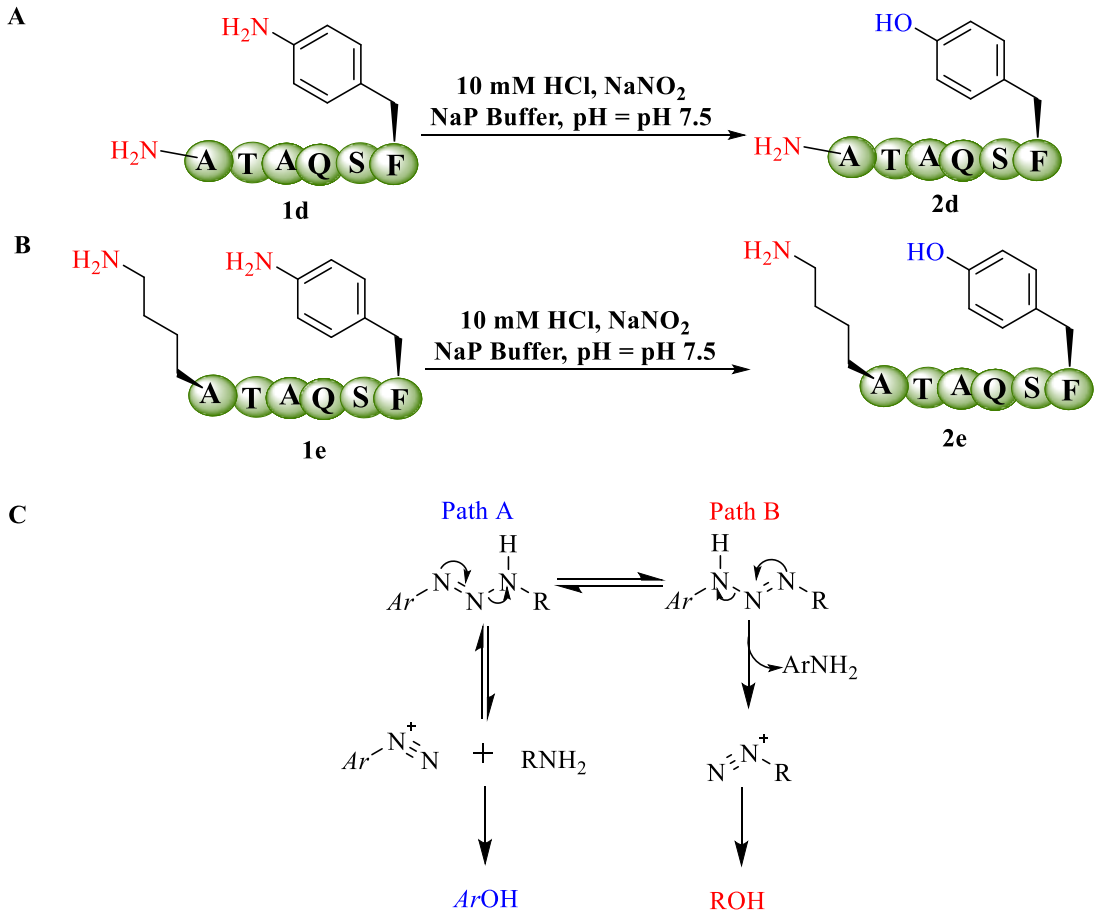


Figure 2.3: Reaction with control peptides: a. Reaction scheme for the formation of alcohol when diazonium ion reacts with N-terminary primary amine. b. Reaction scheme for the formation of alcohol when diazonium ion reacts epsilon-amino group of lysine. c. Mechanistic pathways for the hydrolysis of triazene formed between primary amine and diazonium ion.

Primary amines have been reported to react with diazonium ion to give unstable adducts that dissociate in an aqueous solution. Thus, we attempted the macrocyclization of linear peptides containing primary amines such as N-terminal alanine and lysine ATAQS(pAF) **1d** and Ac-KTAQS(pAF) **1e** under various reaction conditions to confirm to the nature of products formed with them (**Figure 2.3**). As expected, we did not observe the formation of the macrocyclic products

of **1d** and **1e** between the diazonium ion and N-terminal alanine and the side chain of lysine (**Figure 2.3**, *Supporting Information Figure S4*). However, we observed the formation of the corresponding peptide alcohols **2d** and **2e** as confirmed by the HPLC and MS (**Figure 2.3**, *Supporting Information Figure S4 and S2*). To confirm the nature of the alcohol we were forming, we synthesized the peptide alcohol product, **2d**, in 35% yield from the linear peptide, ATAQS(pAF) **1d**, on a large scale and characterized it by NMR (Supporting Information Figure S5). From NMR analysis, we observed that the NH_2 protons of the N-terminal alanine remained unchanged in both **1d** and **2d** and a significant change in the ^{1}H NMR spectrum for *p*-NH_2-phenylalanine in **2d** with the peak at δ 9.14 for a single proton suggested that NH_2 of the phenyl ring converted into the −OH group. This was further supported by the observation that there was a significant change in the ^{13}C NMR spectrum of phenyl carbons from δ 137.61 in **1d** to δ 155.73 in **2d** (*Supporting Information Figure S5*). We postulated that this might be because the electronegative oxygen de-shields the carbon on the phenyl ring, and consequently, moves the signals downfield. To exhaustively confirm the structure of alcohol **2d**, we made the **1d** analogue, by replacing *p*-NH_2-phenylalanine with tyrosine ATAQSY. Interestingly, the NMR of the **1d** analogue ATAQSY was identical to that of **2d** thus proving that NH_2 of *p*-NH_2-phenylalanine was converted to tyrosine (−OH) due to the degradation of the triazene cyclic product obtained between primary amine and diazonium ions in aqueous medium via **path A** in the mechanistic pathway (**Figure 2.3c**, *Supporting Information Figure S4*).

Also, we examined the reactivity of lysine towards our chemistry by carrying the macrocyclization reaction with a peptide PGVSAKGA(pAF)G **1f** containing both lysine and N-terminal proline under the reaction conditions. Much to our satisfaction, the triazene cyclic peptide **2f** was formed with proline with >99% conversion (44% yield) as analyzed by HPLC and HRMS; and confirmed

with NMR (Supporting Information Figure S7, Table S1 and S2). Following the trend, the diagnostic NH proton of proline of **1f** was observed at δ 2.32–2.27 but absent in the ^{1}H NMR spectrum of triazene cyclic product **2f**. Also, in ^{13}C NMR, we observed that the diagnostic alpha carbon (29.2 ppm) on the proline in the cyclized product **2f** moved upfield as compared to the linear peptide **1f** alpha carbon (46.15 ppm) (Supporting Information Figure S7). However, the chemical shifts observed for values for lysine side chain in the NMR of cyclic peptide **2f** and **1f** were not significantly different, confirming that proline is the site of cyclization. Thus, we exhaustively demonstrated that the arene triazene cyclization reaction is highly selective for secondary amines over the primary amines.

2.5.3. Scope of Arene Triazene Chemistry

Next, we examined the breadth of substrates that can react with our chemistry.

2.5.3.1. Triazene Cyclic Peptides with Different Ring Sizes and Amino Acid Composition

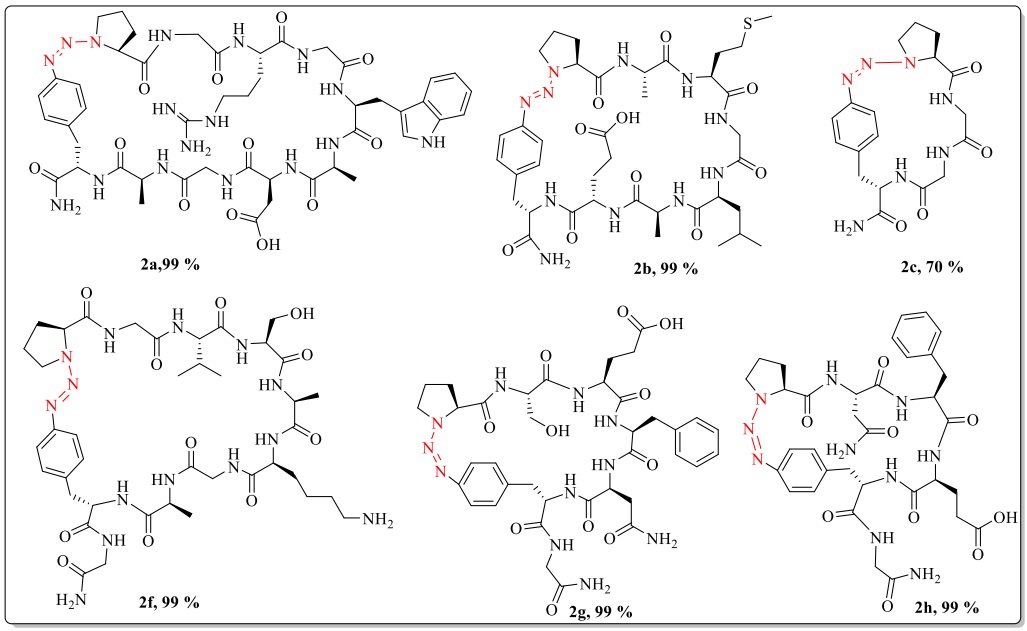


Figure 2.4: Triazene cyclic peptide of various ring sizes and amino acid composition

We made various unprotected linear peptides of different lengths and amino acid composition: PGRGWADGA(pAF) **1a**, PAMGLAE(pAF) **1b**, PGG(pAF) **1c**, PGVSAKGA(pAF)G **1f**, PSEFN(pAF)G **1g**, and PNFE(pAF)G **1h** being 10-mer, 8-mer, 9-mer, 6-mer and 5-mer respectively. In general, they had reactive amino acids such as Trp, Arg, Lys, Asp, Asn, Glu, Ser, and Met. The linear peptides were subjected to our reaction conditions, and they were efficiently cyclized to the corresponding triazene macrocycles (**2a-2c** and **2f-2h**) with complete conversions as determined by LC-MS analyses of the crude reaction mixtures (**Figure 2.4**, *Supporting Information Figure S8, Tables S1* and *S2*). It is important to note that none of the peptides had turn inducers and we still recorded complete conversion to the cyclic peptides in minutes without the formation of any side products. This is remarkable because most of the successful methods for macrocyclization embrace the incorporation of turn-inducing elements capable of promoting macrocyclic ring closure by bringing both termini.[37] Also, from our analyses, there were no side reactions with any of the reactive amino acids.

2.5.3.2. Triazene Cyclic Peptide Containing Different Secondary Amines

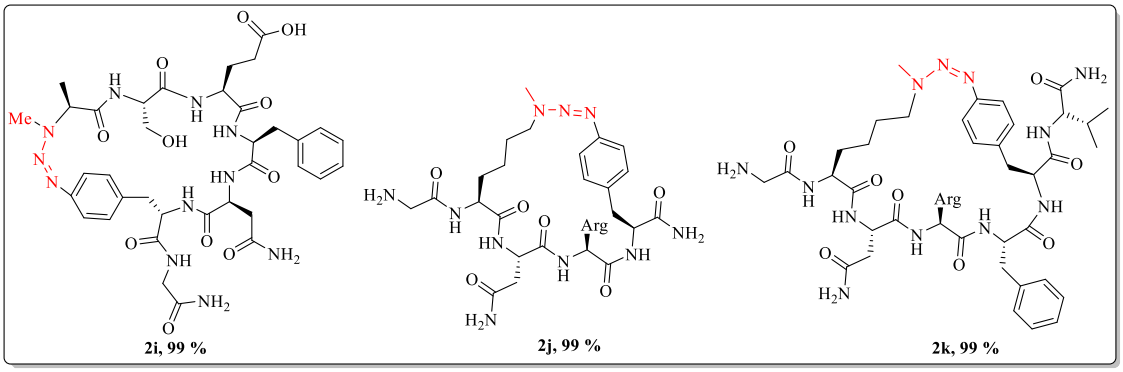


Figure 2.5: Triazene cyclic peptide with different secondary amine

Next, the reactivity of the arene triazene reaction toward different secondary amines was investigated. We started by introducing N-methylated alanine at the N-terminus A(me)SEFN(pAF)G **1i** and under the reaction conditions, high conversion to the head-to-side

chain triazene-cyclic product **2i** was recorded within 10 minutes (**Figure 2.5**, *Supporting Information S8, Tables S1* and *S2*). In the same manner, we subjected two unprotected peptides: GK(me)NR(pAF) **1j** and GK(me)NRF(pAF)V **1k** with a free primary N-terminus and monomethyl lysine to our reaction condition and observed complete conversion to the products as analyzed by HPLC and MS (**Figure 2.5**, *Supporting Information S8, Tables S1* and *S2*). Thus, our strategy can produce both head-to-side chain and side chain-to-side chain triazene cyclic peptides because it does not distinguish between cyclic secondary amines (proline) and non-cyclic secondary amines (N-methyl amine), as well as the side chain amino group and the N-terminal amine. N-methylation has proven to be an effective method for controlling the permeability of cyclic peptides/peptidomimetics.[38] Also, in contrast to other inbuilt-chromophore macrocyclization techniques, orthogonal protecting groups are not required with any reactive amino acids for diazonium ions, including the free N-terminus or side chains of lysine and tyrosine.

2.5.3.3. Triazene Cyclic Peptides Made from Peptide Sequences Reported to Be Difficult to cyclize

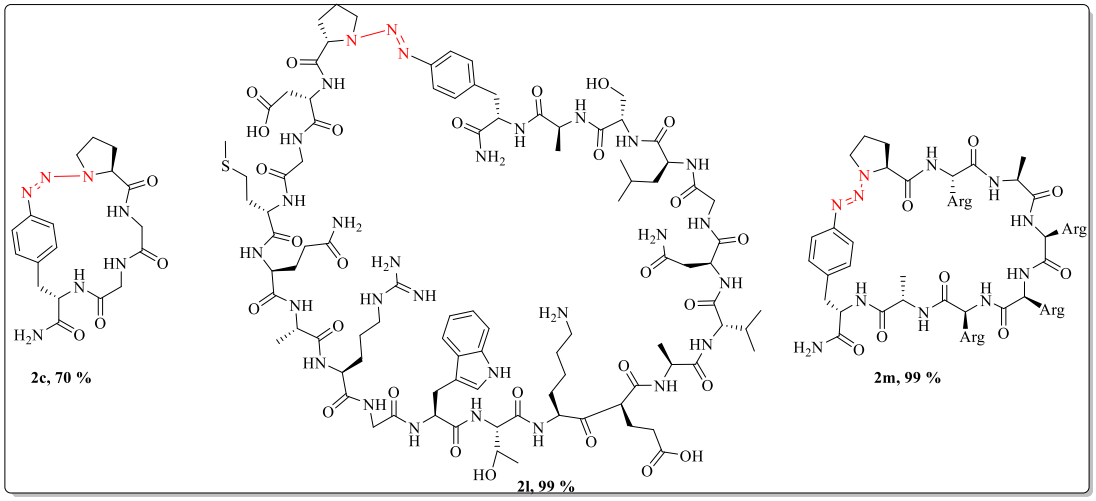


Figure 2.6: Cyclization of difficult sequences by RARE strategy

Small ring peptide cyclization is difficult and often results in considerable cyclic dimer formation.[39] Thus, we carried out the cyclization of a short peptide with only 4 amino acids,

PGG(pAF) **1c,** and observed 70% conversion to the cyclic product **2c** and 30% was cyclic dimer **3c** (**Figure 2.6**, Supporting Information S8, Tables S1 and S2)**.** Also, peptides with long linear amino acid sequences are reputed to be harder to cyclize but, in our hands, we were able to a cyclize a 20 amino acid peptide PDGMQARGWTKEAVNGLSA(pAF) **1l** to full conversion (>99%) to the triazene cyclic products **2l** without the formation of any dimers or oligomers (**Figure 2.6**, Supporting Information S8, Tables S1 and S2). Amit *et al*. reported that peptides that are generally rich in Lys and Arg residues are difficult to cyclize[39] and thus, we explored whether our chemistry could mitigate against this by subjecting a multiarginine-containing peptide PRARRRA(pAF) **1m** to our cyclization condition and we observed full conversion (>99%) to the triazene cyclic products **2m** without the formation of any dimers or oligomers (**Figure 2.6**, Supporting Information S8, Tables S1 and S2). As a result, the chemoselective arene triazene cyclization technique makes it easy to quickly cyclize unprotected native peptides, which are otherwise difficult to achieve.

2.5.4. Rate of Arene Triazene Cyclization

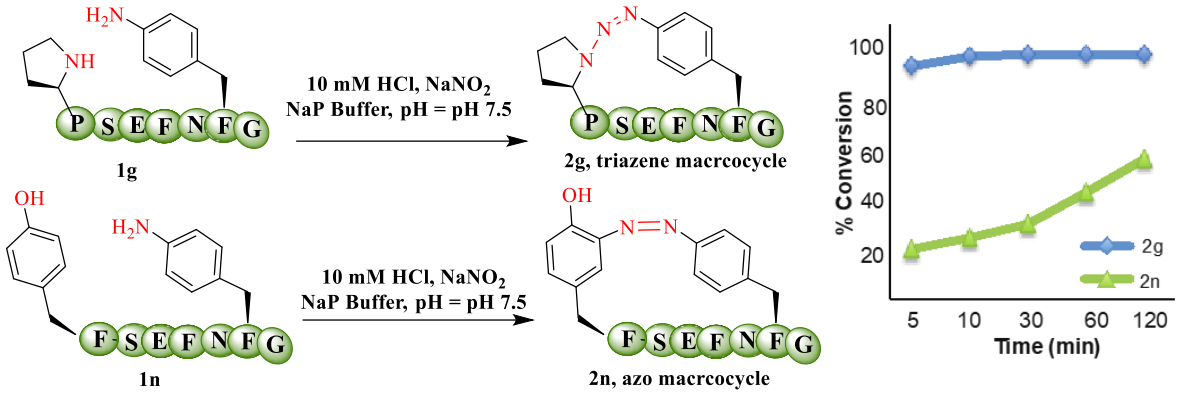


Figure 2.7: Rate study comparison between triazene versus azo cyclization

Macrocyclization reactions with a rapid rate of conversion of the linear peptide to the cyclized product are desirable as they mitigate the formation of oligomers by favoring intramolecular reactions over intermolecular reactions. So, we evaluated the rate of the arene triazene reaction by

carrying out a macrocyclization of a peptide PSEFN(pAF)G **1g** (3 mM in 10 mM HCl) in the presence of $NaNO_2$ (1.5 equivalent of 4.5 mM) and with the formation of the diazonium ion, we adjusted the overall pH of the solution to 7.5. The reaction was allowed to stir at room temperature and aliquots were taken at regular intervals of time and analyzed with HPLC and MS to monitor the reaction progress. To our greatest surprise, we observed that the cyclization of the peptide **1g** to **2g** was near completion with >99% conversion in less than 5 min (**Figure 2.7**, *Supporting Information S9*). To compare this with possible cyclization resulting from the azo coupling between diazonium ion and tyrosine residue on a peptide, we made a linear peptide analog of **1g** with proline replaced by tyrosine, YSEFN(pAF)G **1n.** We subjected **1n** to our cyclization reaction and observed only 20% conversion to azo cyclic product 2n after 10 min and 50% conversion in 2 h (**Figure 2.7**, *Supporting Information S9*). It is important to mention that both **1g** and **1n** were exposed to similar reaction conditions and we observed the remarkable formation of **2g** over 2**n**. We proceeded to evaluate the extent of conversion of **1n** to **2n** under elevated pH since literature reports have hinted that azo coupling between arene diazonium ion and tyrosine is more favorable under elevated pH[40, 41]. Aligning with previous reports, we recorded the formation of the azo cyclic product with Tyr in high conversion at elevated pH only (pH 9). We were unable to isolate the diazonium ion in the reaction with proline since it was a highly rapid reaction that produced **2g** triazene cyclic product in minutes, so, we synthesized a peptide AcATAQS(pAF) **1o** with arene amine but no free secondary or primary amine. The peptide AcATAQS(pAF) **1o** was subjected to our reaction conditions by first converting it to the diazonium adduct **2o** followed by adjusting the pH to 7.5. We were able to trap the diazonium ion since there was no reactive group it could react with, and we confirmed the formation of AcATAQS(pAF) diazonium ion **2o** and characterized it by HPLC and MS (*Supporting Information S10*). The m/z of the diazonium ion (M) and M-N2

peaks were detected in MS spectra, confirming the formation of diazonium adduct **2o** in HPLC at 8 minutes. For more elaborate analysis, we added piperidine and KI separately to diazonium ion **2o** solutions and observed the formation of triazene **3o** and iodinated product **4o** (*Supporting Information Figure S10*), respectively further confirming that the intermediate was diazonium ion. As a definite proof of the formation of diazonium ions, we made diazonium-ion tetrafluoroborate salt synthesized from small molecule Fmoc-Phe (4-NH2)-OH on a large scale and carried out NMR on it (*Supporting Information Figure S11*). The NMR studies compared favorably with commercially available p-methoxyphenyl diazonium salt and conclusively confirmed the formation of the diazonium adduct under our reaction conditions (*Supporting Information Figure S11*).

2.5.5. Chromophoric Property of Triazene Cyclic Peptides

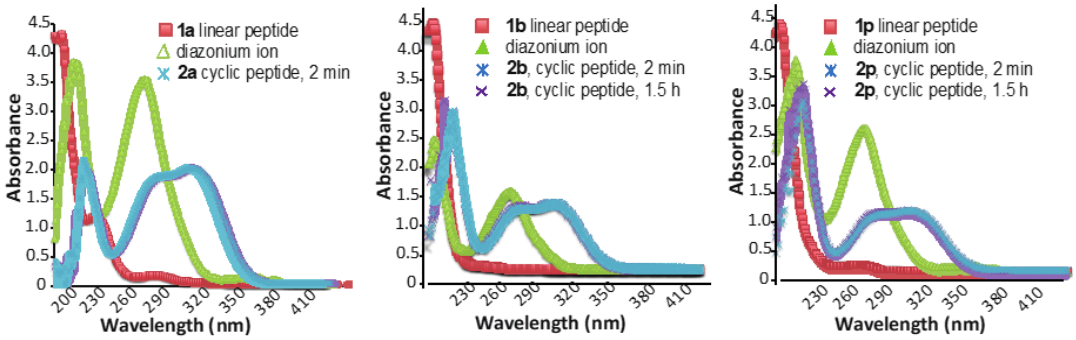


Figure 2.8: Chromophoric studies of linear peptides to triazene cyclic peptides via diazonium ion

Cyclic peptides that generate inbuilt chromophores eliminate the need for cyclic peptides to contain a large bulky chromophoric group, which can affect their structure, binding affinity, and solubility. Also, the chromophoric property aids in confirming successful peptide cyclization and as well aids in screen and target identification.[29] The triazene moiety generated in our cyclic peptide is a chromophore, so we evaluated the photophysical properties of triazene cyclic peptides using UV analysis. The UV analysis on the triazene cyclic peptide PGG(pAF) **2c** revealed that

triazene had a distinctive absorbance at 289 and 313 nm wavelengths. The linear equivalent **1c**, on the other hand, did not demonstrate any absorption in this UV region (**Figure 2.8**, *Supporting Information S12*). Having confirmed a distinct absorbance in UV corresponding to triazene cyclic peptide **2c**, we monitored the course of reaction on three linear peptides PGRGWADGA(pAF) **1a**, PAMGLAE(pAF) **1b**, and PTAQS(pAF) **1p** by directly carrying out the cyclization reaction in UV cuvettes (**Figure 2.8**, *Supporting Information S13*). Following the addition of $NaNO_2$ under acidic circumstances, all three peptides produced a peak at 273 nm, which is indicative of diazonium ions. After changing the pH of the solution to 7.5, additional peaks at 289 and 313 nm, which are indicative of the triazene cyclic products for all three cyclic peptides **2a–2b** and **2p**, were seen within 30 seconds (**Figure 2.8**, *Supporting Information S13*). We discovered that all three peptides converted completely to triazene cyclic products within a minute, regardless of the size or sequence of the amino acids. HPLC and MS analyses of the resultant solutions confirmed the full conversion to the triazene cyclic compounds **2a–2b** and **2p**. We then used linear regression to determine the extinction coefficient of triazene cyclic peptides **2a** and **2c** (58,500–59,900) by monitoring UV at various concentrations of the triazene cyclic products **2a** and **2c** (*Supporting Information Figure S15*)

Earlier before this discovery, reaction mixtures were allowed to stir for up to 10 minutes, but encouraged by this, we employed HPLC and MS analysis to investigate the fate of the peptide **1b** cyclization reaction at room temperature within 1 minute. We found that the full conversion to triazene cyclic product **2b** (>99 percent) occurred within a minute as determined by HPLC and MS analysis (*Supporting Information Figure S14*). Taken together, all these studies confirmed that the chemistry of arene triazene cyclization is fast and generates chromophores in triazene cyclic peptides.

2.5.6. Response of Triazene Cyclic Peptide to External Stimuli

2.5.6.1. Response Triazene Cyclic Peptide to Change in pH

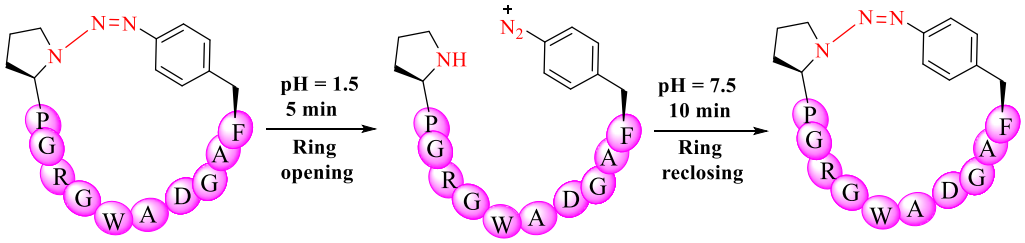


Figure 2.9: Dynamic response of triazene macrocyclic peptides to pH change

There is an interest in methodologies for the synthesis of smart cyclic peptides that respond to changes in the external stimuli by opening and re-closing the macrocycle ring. Inspired by the observation from our previous work on diazonium ions for labeling monomethyl lysine, we rationalized that triazene cyclic peptides will respond to a change in the pH by the protonation of the triazene. It was our belief that the triazene peptide will undergo protonation at acidic conditions (low pH), leading to the opening of the triazene ring to an unchanged starting linear peptide with diazonium ions and N-terminal proline; and by re-adjusting the pH to neutral pH, we can re-close the ring to afford the triazene cyclic peptide. This was tested by subjecting PGRGWADG(pAF)**1a** to our cyclization conditions (**Figure 2.9**). At pH 7, the formation of the triazene cyclic peptide **2a** was observed and had a retention time of 26.379 on HPLC (*Supporting Information S16*). The ring-opening step was carried out by lowering the pH to 1.5 and we observed a significant shift in the retention time in the HPLC (R_t = 22.829) within 5 min, and the *m/z* of the new peak corresponded to the linear peptide M–N_2 adduct. In the mass spectrometer, it is common to observe −N_2 adducts of the diazonium salts. To have confidence in this finding and confirm that it is the diazonium ion that is generated from this step and not the proto-diazotization product that could lead to the generation of phenylalanine at the site of pAF in the peptide, we first adjusted the pH to 7.5 and excitingly, observed the formation of the cyclic triazene peptide **2a** within 5 min as

analyzed by both HPLC retention time (R_t = 25.81) and mass spectrometry, which clearly showed the M + H adduct of the cyclic peptide **2a** (*Supporting Information S16*). Secondly, to completely rule out the formation of the proto-diazotization product, we made the linear peptide PGRGWADGAF with Phe (F) at the site of pAF. From the HPLC analysis of this peptide, we observed a peak with a retention time of 21.36 which was different from the retention time of the ring-opening product (*Supporting Information S16*). Thus, we concluded that our method gives rise to a cyclic peptide that can be opened and re-closed in response to a change in pH and represents a significant technical advance in the known methodologies for the generation of cyclic peptides. This is particularly important in drug discovery, where this unique feature of triazene cyclic peptides could be used to sequence hit cyclic peptides discovered from library screening.

2.5.6.2. Response of Triazene Cyclic Peptide to UV Radiation

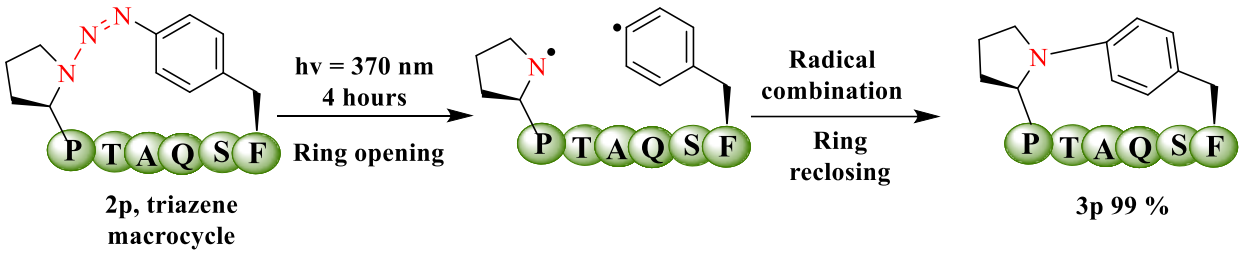


Figure 2.10: Conversion of triazene macrocycle to another ring structure by diradical combination

Next, we evaluated whether triazene cyclic peptide responds to UV irradiation. We started our study with triazene cyclic peptide **2p** and observed that the ring opens completely in 4 hours to form a diradical intermediate, which then rejoins to form a new cyclic peptide **3p** (35 % yield) with a C–N bond between the proline and phenyl ring of p-NH2-Phe at the site of cyclization (**Figure 2.10,** *Supporting Information S17*). To fully characterize the full structure of **3p**, we carried out 1D and 2D NMR on **3p** and compared it with the linear peptide **1p** (*Supporting Information Figure S17).* The NH proton of proline in **1p** was found as a multiplet at 2.36–2.31,

however it vanished in the ^{1}H NMR spectra of product *3p*. When compared to linear peptide **1p** alpha protons (3.26–3.22 and 3.19 ppm) and alpha carbon (46.32 ppm), the diagnostic alpha protons (4.02–3.96, ppm) and alpha carbon (54.12 ppm) on the proline in the cyclized product **3p** shifted more downfield, indicating more deshielding (*Supporting Information Figure S17*).This was in contrast to triazene cyclic product **2p**, where the alpha proton and carbon of proline are more protected and an upfield shift was detected in all the NMR spectra of the cyclic triazene peptides. The interaction of the alpha proton of proline with protons of the phenyl ring of pAF is indicated by another distinctive shift of **3p** in ROESY spectra. Because these protons were not near the aromatic ring in triazene cyclic products, no such contact was seen, implying the creation of a C–N bond between the proline and phenyl ring of p–NH2–Phe (*Supporting Information Figure S17*).

2.5.7. Cyclization in One Pot and Sequencing of Triazene Cyclic Peptides

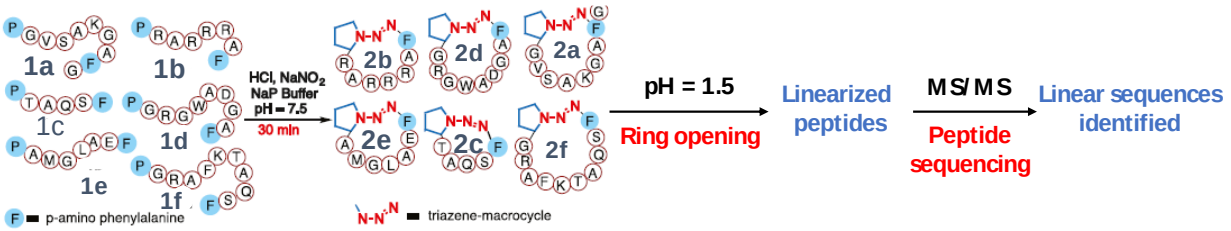


Figure 2.11: Multiple peptide cyclization in one-pot via RARE

Our lab has been interested in the site-specific linearization of cyclic peptides as a way of facilitating the analysis and sequencing of hits from peptide library screening. The sequence determination of hit peptides frequently slows down the screening of these cyclic peptide libraries against biological targets. A cyclic peptide undergoes ring-opening at many places in a mass spectrometer, yielding a family of mass degenerate ions, each of which fragments further into a complicated mixture of shorter peptides, making spectral interpretation extremely challenging. Also, because cyclic peptides (N-to-C cyclization or N-to-Side) lack a free N-terminus, they are

unable to be sequenced using the Edman degradation. Thus, a method of linearization of cyclic peptides could overcome these challenges. Reckoning by the chemoselective nature of our chemistry towards secondary at pH 7.5 and the reaction rate, we proceeded to evaluate if under our reaction conditions we could synthesize multiple cyclic peptides from their corresponding linear peptides in one pot and under acidic conditions linearize them. Six linear peptides **1a–1b, 1f, 1m, 1p,** and **1q** were dissolved in one pot and subjected to our reaction conditions (**Figure 2.11**, *Supporting Information Figure S18*; conc. 3 mM of each peptide). We observed successful conversion of the linear peptides to the triazene cyclic products **2a–2b, 2f, 2m, 2p,** and **2q** without observing any intermolecular cross-linking products as seen from both HPLC and MS analysis (Figure 4a, Supporting Information Figure S18). We used 10 % TFA in buffer to open the library of six cyclic peptides to linear counterparts and observed full ring-opening as analyzed by LC-MS to demonstrate the utility of our method for the sequencing of HIT cyclic peptides after screening (**Figure 2.11**, *Supporting Information Figure S18*). The sequences of three of the linearized peptides **1b, 1p,** and **1q** were then determined by LC-MS/MS. (*Supporting Information Figure S18).* Thus, we proved that our method can substantially help in elucidating the structure of cyclic peptide hits after library screening.

2.5.8. Stability of Triazene Cyclic Peptides

The stability of cyclic peptides is a major concern for pharmaceutical applications. We incubated cyclic peptide **2b** in aqueous solution and a water: ACN mixture, to determine the stability of the triazene cyclic peptide. The triazene cyclic peptide **2b** remained stable for 24 hours and showed no signs of degradation (*Supporting Information Figure S19*). We also subjected triazene cyclic peptide **2b** to sodium dithionite (5 equiv), and it remained stable for 2 hours without degradation, whereas azo-cyclic peptide **2n** generated between the Tyr side chain and diazonium ions degraded

completely in 5 minutes. We further demonstrated the orthogonality of our strategy by exposing triazene cyclic peptide **2a** to the piperidine utilized in Fmoc-SPPS for 12 hours and found no degradation of **2a** as determined by HPLC and MS (*Supporting Information Figure S20*).

Next, we compared the proteolytic stability of a cyclic peptide to that of its linear counterpart to assess the potential of triazene cyclic peptides for biological applications. Trypsin is traditionally employed to hydrolyze peptide bonds at the C-terminal side of lysine and arginine in peptides and proteins. So, we subjected linear peptide PGRAFKTAQS(pAF) **1q** and the corresponding triazene cyclic adduct **2q** to incubation with a solution of trypsin. According to HPLC and MS analyses, triazene cyclic peptide **2q** remained totally intact for up to 2 hours in the presence of trypsin, whereas its linear counterpart **1q** was completely proteolytically degraded (*Supporting Information Figure S21*). From these results, the triazene moiety formed during cyclization considerably increased the stability of cyclic peptides against proteolysis and degradation under a variety of harsh conditions. Thus, our chemistry can be used to synthesize potentially bioactive cyclic peptidomimetics that can be used to probe biological systems.

2.5.9. Synthesis of Bicyclic Peptides

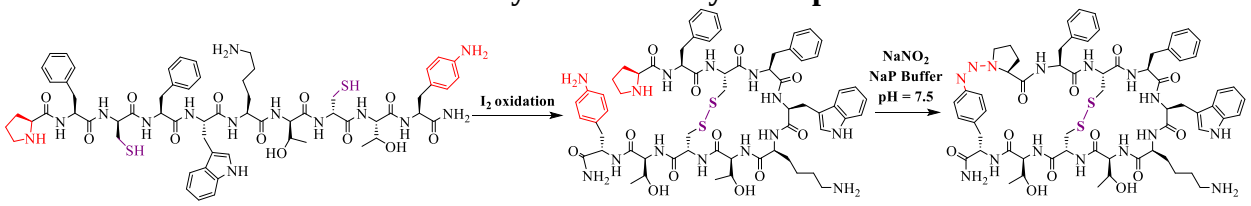


Figure 2.12: Peptide bicyclization using Arene Triazene chemistry

Bicyclic peptides have two macrocyclic rings and have advantages over linear and monocyclic peptides[42] as each ring of this bicyclic structure can act separately, making these peptides bifunctional.[43] We employed our RARE strategy for the synthesis of bicyclic peptides. PFCFWKTCT(pAF) **1r** was synthesized as an analogue of octreotide with two cysteine residues, then cyclized under oxidative conditions to generate a monocyclic peptide containing a disulfide

bond **2r.** Further arene triazene cyclization of the disulfide cyclic peptide **2r** produced bicyclic product **3r** with >90% conversion according to LCMS analysis (**Figure 2.12**, *Supporting Information Figure S22*).

2.5.10. Late-Stage Derivatization of Triazene Cyclic Peptides

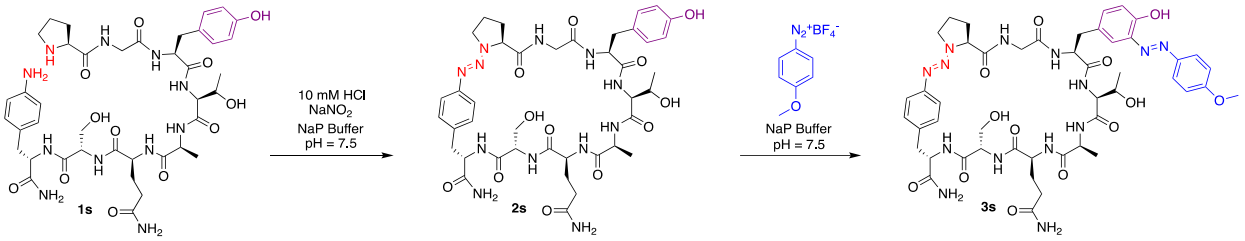


Figure 2.13: Late-stage diversification of triazene macrocyclic peptide through azo coupling of tyrosine residue

Late-Stage Diversification (LSD) of structurally complex peptides has emerged as a promising platform for molecular engineering and drug development.[44] Late-stage functionalization increases the structural complexity of cyclic peptides, which can be used to study structure-activity relationships and lead to the discovery of high-affinity binders. The late-stage derivatization of triazene cyclic peptides was carried out by utilizing known chemistries. To make triazene cyclic peptide **2s**, a linear peptide containing tyrosine PGYTAQS(pAF) **1s** was synthesized, then cyclized by employing arene triazene chemistry. By the addition of the electron-rich diazonium ion 4-MDz, the tyrosine side chain of the triazene cyclic peptide **2s** was modified to afford azo-cyclic peptide **3s** analyzed by HPLC and MS (53 % conversion (**Figure 2.13**, *Supporting Information Figure S23*). The lower conversion recorded might be due to the inefficiency of tyrosine labeling with an electron-donating arene diazonium ion 4-methoxy phenyl diazonium ion (4MDz). To evaluate the relative reactivity between proline and tyrosine, we cyclized a linear peptide PGYTAQS(pAF) **1s** containing both Tyr and Pro and observed the formation of 81 % of the triazene cyclized product **2s** (49 % yield) and 19 % azo cyclic product

2s′ analyzed using HPLC and MS. We confirmed the structures of these adducts using NMR (*Supporting Information S24*).

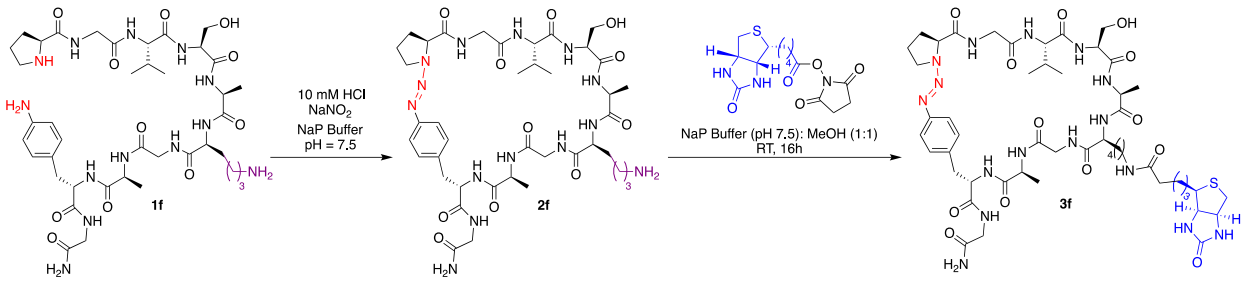


Figure 2.14: Late-stage diversification of triazene macrocyclic peptide via coupling of biotin-NHS ester to the lysine residue

Next, we added biotin NHS ester to triazene cyclic peptide PGVSAKGA(pAF)G **2f** and observed the modification of the lysine side chain, forming a stable amide bond 3f with 66% conversion as analyzed by LCMS (**Figure 2.14**, *Supporting Information S25*). It is worthy of note that by successively adding 4MDz and NHS-ester to the reaction mixtures, we were able to execute the arene triazene cyclization and post cyclization modification in one pot.

2.6. Conclusion

In summary, the Rapid Arene Triazene Chemistry for Macrocyclization strategy is a highly chemoselective rapid intramolecular reaction of secondary amines such as proline on a peptide and p-amino phenylalanine that generates an inbuilt chromophoric triazene cyclic peptide. The triazene cyclic peptides can be opened in response to acidic conditions and UV radiations and can be preferentially reclosed when restored to neutral pH. This unique quality makes our approach stand out as a method of choice synthesis of cyclic peptides for targeted drug delivery and could act as sensors for the local environment in various disease states. Also, the method has high reaction kinetics enabling the efficient cyclization of peptides of short and long sequences within minutes.

With this method synthesis of 18-66-membered monocycles and bicycles was carried and some of the triazene cyclic peptides were further derivatized with various functional groups.

CHAPTER TWO REFERENCES

(1) Joo, S. H. Cyclic peptides as therapeutic agents and biochemical tools. *Biomol Ther (Seoul)* **2012**, *20* (1), 19-26. DOI: 10.4062/biomolther.2012.20.1.019.

(2) Gang, D.; Kim, D. W.; Park, H.-S. Cyclic Peptides: Promising Scaffolds for Biopharmaceuticals. *Genes (Basel)* **2018**, *9* (11), 557. DOI: 10.3390/genes9110557.

(3) Khan, A. R.; Parrish, J. C.; Fraser, M. E.; Smith, W. W.; Bartlett, P. A.; James, M. N. Lowering the entropic barrier for binding conformationally flexible inhibitors to enzymes. *Biochemistry* **1998**, *37* (48), 16839-16845. DOI: 10.1021/bi9821364.

(4) Gracia, S. R.; Gaus, K.; Sewald, N. Synthesis of chemically modified bioactive peptides: recent advances, challenges and developments for medicinal chemistry. *Future Med Chem* **2009**, *1* (7), 1289-1310. DOI: 10.4155/fmc.09.97.

(5) Loffet, A. Peptides as drugs: is there a market? *J Pept Sci* **2002**, *8* (1), 1-7. DOI: 10.1002/psc.366.

(6) Hummel, G.; Reineke, U.; Reimer, U. Translating peptides into small molecules. *Mol Biosyst* **2006**, *2* (10), 499-508. DOI: 10.1039/b611791k.

(7) Horton, D. A.; Bourne, G. T.; Smythe, M. L. Exploring privileged structures: the combinatorial synthesis of cyclic peptides. *J Comput Aided Mol Des* **2002**, *16* (5-6), 415-430. DOI: 10.1023/a:1020863921840.

(8) Vinogradov, A. A.; Yin, Y.; Suga, H. Macrocyclic Peptides as Drug Candidates: Recent Progress and Remaining Challenges. *Journal of the American Chemical Society* **2019**, *141* (10), 4167-4181. DOI: 10.1021/jacs.8b13178.

(9) Qian, Z.; Dougherty, P. G.; Pei, D. Targeting intracellular protein-protein interactions with cell-permeable cyclic peptides. *Curr Opin Chem Biol* **2017**, *38*, 80-86. DOI: 10.1016/j.cbpa.2017.03.011.

(10) Sazykin Iu, O.; Telsnina, G. N.; Zaslavskaia, P. L.; Lasova, T. V.; Ivanov, V. P.; Elkina, E.; Navashin, S. M. [Cyclosporin A and the permeability of the cytoplasmic membrane in Aspergillus niger]. *Antibiot Khimioter* **1995**, *40* (10), 19-24.

(11) GIROUX, R. Cyclosporine. *Chemical & engineering news* **2005**, *83* (25), 56-56.

(12) Moiola, M.; Memeo, M. G.; Quadrelli, P. Stapled Peptides—A Useful Improvement for Peptide-Based Drugs. *Molecules* **2019**, *24* (20), 3654.

(13) Posada, L.; Serra, G. Three Methods for Peptide Cyclization Via Lactamization. *Methods Mol Biol* **2022**, *2371*, 3-17. DOI: 10.1007/978-1-0716-1689-5_1.

(14) Schafmeister, C. E.; Po, J.; Verdine, G. L. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *Journal of the American Chemical Society* **2000**, *122* (24), 5891-5892. DOI: 10.1021/ja000563a.

(15) Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y. S.; Pentelute, B. L. A perfluoroaryl-cysteine S(N)Ar chemistry approach to unprotected peptide stapling. *J Am Chem Soc* **2013**, *135* (16), 5946-5949. DOI: 10.1021/ja400119t .

(16) Lautrette, G.; Touti, F.; Lee, H. G.; Dai, P.; Pentelute, B. L. Nitrogen Arylation for Macrocyclization of Unprotected Peptides. *J Am Chem Soc* **2016**, *138* (27), 8340-8343. DOI: 10.1021/jacs.6b03757 .

(17) Góngora-Benítez, M.; Tulla-Puche, J.; Albericio, F. Multifaceted roles of disulfide bonds. Peptides as therapeutics. *Chem Rev* **2014**, *114* (2), 901-926. DOI: 10.1021/cr400031z.

(18) Jo, H.; Meinhardt, N.; Wu, Y.; Kulkarni, S.; Hu, X.; Low, K. E.; Davies, P. L.; DeGrado, W. F.; Greenbaum, D. C. Development of α-helical calpain probes by mimicking a natural protein-protein interaction. *J Am Chem Soc* **2012**, *134* (42), 17704-17713. DOI: 10.1021/ja307599z.

(19) Blackwell, H. E.; Sadowsky, J. D.; Howard, R. J.; Sampson, J. N.; Chao, J. A.; Steinmetz, W. E.; O'Leary, D. J.; Grubbs, R. H. Ring-closing metathesis of olefinic peptides: design, synthesis, and structural characterization of macrocyclic helical peptides. *J Org Chem* **2001**, *66* (16), 5291-5302. DOI: 10.1021/jo015533k.

(20) Dharanipragada, R. New modalities in conformationally constrained peptides for potency, selectivity and cell permeation. *Future Med Chem* **2013**, *5* (7), 831-849. DOI: 10.4155/fmc.13.25.

(21) Cistrone, P. A.; Silvestri, A. P.; Hintzen, J. C.; Dawson, P. E. Rigid peptide macrocycles from on-resin Glaser stapling. *ChemBioChem* **2018**, *19* (10), 1031-1035.

(22) Assem, N.; Ferreira, D. J.; Wolan, D. W.; Dawson, P. E. Acetone-Linked Peptides: A Convergent Approach for Peptide Macrocyclization and Labeling. *Angewandte Chemie International Edition* **2015**, *54* (30), 8665-8668. DOI: https://doi.org/10.1002/anie.201502607.

(23) Wang, Y.; Chou, D. H. C. A thiol–ene coupling approach to native peptide stapling and macrocyclization. *Angewandte Chemie International Edition* **2015**, *54* (37), 10931-10934.

(24) Ricardo, M. G.; Llanes, D.; Wessjohann, L. A.; Rivera, D. G. Introducing the petasis reaction for late-stage multicomponent diversification, labeling, and stapling of peptides. *Angewandte Chemie International Edition* **2019**, *58* (9), 2700-2704.

(25) Raynal, L.; Rose, N. C.; Donald, J. R.; Spicer, C. D. Photochemical Methods for Peptide Macrocyclisation. *Chemistry* **2021**, *27* (1), 69-88. DOI: 10.1002/chem.202003779.

(26) Adebomi, V.; Cohen, R. D.; Wills, R.; Chavers, H. A. H.; Martin, G. E.; Raj, M. CyClick Chemistry for the Synthesis of Cyclic Peptides. *Angewandte Chemie International Edition* **2019**, *58* (52), 19073-19080. DOI: https://doi.org/10.1002/anie.201911900.

(27) Bandyopadhyay, A.; Gao, J. Iminoboronate-Based Peptide Cyclization That Responds to pH, Oxidation, and Small Molecule Modulators. *Journal of the American Chemical Society* **2016**, *138* (7), 2098-2101. DOI: 10.1021/jacs.5b12301.

(28) Ikeda, M.; Tanida, T.; Yoshii, T.; Hamachi, I. Rational molecular design of stimulus-responsive supramolecular hydrogels based on dipeptides. *Advanced Materials* **2011**, *23* (25), 2819-2822.

(29) Todorovic, M.; Perrin, D. M. FlICk (fluorescent isoindole crosslinking) for peptide stapling. *Methods Enzymol* **2020**, *639*, 313-332. DOI: 10.1016/bs.mie.2020.04.017.

(30) Bräse, S.; Dahmen, S.; Lauterwasser, F.; Leadbeater, N. E.; Sharp, E. L. Polymer-bound 1-aryl-3-alkyltriazenes as modular ligands for catalysis. Part 1: Synthesis and metal coordination. *Bioorganic & medicinal chemistry letters* **2002**, *12* (14), 1845-1848.

(31) Lazny, R.; Nodzewska, A.; Klosowski, P. A new strategy for synthesis of polymeric supports with triazene linkers. *Tetrahedron* **2004**, *60* (1), 121-130.

(32) Nwajiobi, O.; Mahesh, S.; Streety, X.; Raj, M. Selective Triazenation Reaction (STaR) of Secondary Amines for Tagging Monomethyl Lysine Post-Translational Modifications. *Angewandte Chemie* **2021**, *133* (13), 7420-7428.

(33) Nwajiobi, O.; Mahesh, S.; Streety, X.; Raj, M. Selective Triazenation Reaction (STaR) of Secondary Amines for Tagging Monomethyl Lysine Post-Translational Modifications. *Angew Chem Int Ed Engl* **2021**, *60* (13), 7344-7352. DOI: 10.1002/anie.202013997 From NLM.

(34) Marchesi, F.; Turriziani, M.; Tortorelli, G.; Avvisati, G.; Torino, F.; De Vecchis, L. Triazene compounds: mechanism of action and related DNA repair systems. *Pharmacological research* **2007**, *56* (4), 275-287.

(35) Domingues, V. O.; Hörner, R.; Reetz, L. G.; Kuhn, F.; Coser, V. M.; Rodrigues, J. N.; Bauchspiess, R.; Pereira, W. V.; Paraginski, G. L.; Locatelli, A. In vitro evaluation of triazenes: DNA cleavage, antibacterial activity and cytotoxicity against acute myeloid leukemia cells. *Journal of the Brazilian Chemical Society* **2010**, *21* (12), 2226-2237.

(36) Seiter, K.; Liu, D.; Loughran, T.; Siddiqui, A.; Baskind, P.; Ahmed, T. Phase I study of temozolomide in relapsed/refractory acute leukemia. *Journal of clinical oncology* **2002**, *20* (15), 3249-3253.

(37) White, C.; Yudin, A. Contemporary strategies for peptide macrocyclization. Nat Chem 3 (7): 509–524. 2011.

(38) Li, Y.; Li, W.; Xu, Z. Improvement on Permeability of Cyclic Peptide/Peptidomimetic: Backbone N-Methylation as A Useful Tool. *Mar Drugs* **2021**, *19* (6), 311. DOI: 10.3390/md19060311 PubMed.

(39) Thakkar, A.; Trinh, T. B.; Pei, D. Global Analysis of Peptide Cyclization Efficiency. *ACS Combinatorial Science* **2013**, *15* (2), 120-129. DOI: 10.1021/co300136j.

(40) Schlick, T. L.; Ding, Z.; Kovacs, E. W.; Francis, M. B. Dual-Surface Modification of the Tobacco Mosaic Virus. *Journal of the American Chemical Society* **2005**, *127* (11), 3718-3723. DOI: 10.1021/ja046239n.

(41) Fridkin, G.; Gilon, C. Azo cyclization: peptide cyclization via azo bridge formation. *J Pept Res* **2002**, *60* (2), 104-111. DOI: 10.1034/j.1399-3011.2002.02993_1.x From NLM.

(42) Ahangarzadeh, S.; Kanafi, M. M.; Hosseinzadeh, S.; Mokhtarzadeh, A.; Barati, M.; Ranjbari, J.; Tayebi, L. Bicyclic peptides: types, synthesis and applications. *Drug Discovery Today* **2019**, *24* (6), 1311-1319. DOI: https://doi.org/10.1016/j.drudis.2019.05.008.

(43) Quartararo, J. S.; Eshelman, M. R.; Peraro, L.; Yu, H.; Baleja, J. D.; Lin, Y.-S.; Kritzer, J. A. A bicyclic peptide scaffold promotes phosphotyrosine mimicry and cellular uptake. *Bioorganic & Medicinal Chemistry* **2014**, *22* (22), 6387-6391. DOI: https://doi.org/10.1016/j.bmc.2014.09.050.

(44) Wang, W.; Wu, J.; Kuniyil, R.; Kopp, A.; Lima, R. N.; Ackermann, L. Peptide Late-Stage Diversifications by Rhodium-Catalyzed Tryptophan C7 Amidation. *Chem* **2020**, *6* (12), 3428-3439. DOI: https://doi.org/10.1016/j.chempr.2020.10.026.

CHAPTER THREE: ENZYME-INSPIRED SITE-SPECIFIC MODIFICATION OF SERINE

3.0. Introduction

Nature has devised a complex system for covalently attaching various functional groups to proteins after they have been synthesized by the ribosome in a process called post-traditional modification. Proteins can have their structure and functional roles expanded by chemically altering them after they have been translated. Chemists have leveraged this attribute to create techniques to covalently endow proteins with various functionalities in the laboratory, in an attempt to follow nature's lead[1]. Modifying such chemically and physically complex biopolymers without disrupting their function is a difficult task. The challenges of achieving this are two-pronged: first, to be relevant, a site-specific protein modification reaction must take place under physiological conditions (i.e., 20–37 °C, pH 6–8, aqueous solvent) in order to avoid disrupting protein structure and function.[2] Secondly, reactions must be site-specific for one amino acid side chain on the protein surface over all others. This makes amino acids with low abundance on protein surfaces well suited for site-specific modification chemistries. Cysteine is the amino acid that is most readily targeted for site-specific modification because of its low abundance (< 2%) in proteins as well as because of its status as the most nucleophilic amino acid[3]. One of the drug modalities that require exquisite control over the site that is modified to achieve homogeneity in the product is Antibody-Drug Conjugates.

3.1. Methods for Achieving Site-Specific Modification on Proteins

The most common examples of protein modification in the past have relied on inherent distinctions in reactivity among the amino acid side chains to determine conjugation. The inherent reactivity

of canonical amino acids like lysine and cysteine has traditionally been used to make protein conjugates.[10] Because of their natural abundance, surface accessibility, and nucleophilicity of the e-amino side chain, lysine residues provide a simple technique for bioconjugation. About 85 lysine residues are found in IgG1 antibodies, with more than 40 of them being changeable and thus it is almost impossible to control the site of conjugation site and millions of different species can be generated in every synthetic batch.[11] Furthermore, because lysine residues are abundantly present on an antibody, their alteration can prevent antigen recognition, decreasing the antibody's potency. Nevertheless, three of the FDA approved ADCs: Mylotargs, Kadcylas and Besponsas all employ lysine conjugation and are therefore administered as a heterogeneous mixture of products[12]. Non-selective techniques are now regarded unsatisfactory for developing next-generation ADCs, therefore, site-specific conjugation strategies to link the cytotoxic drug to the antibody are required in order for ADCs to achieve their full potential[13].

Also because of their low natural abundance and the extraordinarily high nucleophilicity of the deprotonated thiolate side chain, cysteine residues are particularly appealing target for site-specific modification of proteins. For 1g1, the four interchain disulfides on antibodies can be reduced to eight free thiols which are reacted with soft electrophiles to generate ADCs with a homogenous DAR of 8. Studies show that ADCs with higher DAR have increased clearance and less tolerated than ADCs with lower ones[14] and is theorized to cause steric hindrance which can prevent proper binding of the target. Also, high drug load can lead to aggregation if the drug is very hydrophobic[15]. Adcetriss, Polivys, Padcevs, Enhertus, Trodelvys, and Blenreps, as well as the majority of ADCs now in clinical trials, are all made through maleimide chemistry of cysteines[12].

While these approaches have generally had good chemoselectivity, they have failed to sustain regioselectivity at high conversion rates[4]. Although there are numerous techniques for stochastically labeling proteins, these random alterations can block enzymatic active sites and binding pockets or modify the protein's 3D structure, resulting in a decrease in or full obliteration of activity[4].

Alternatively, using site-directed mutagenesis, cysteine residues can be engineered at specific sites on a protein. Junutula *et al*. published the first demonstration of this approach in 2008 coined THIOMABs[16]. The engineered cysteine was inserted on an anti-MUC16 antibody by mutation of heavychain alanine 114 (HC-A114). Using engineered cysteine, ADCs were generated comprising monomethyl auristatin E (MMAE), a potent tubulin inhibitor, and an anti-MUC16 mAb. The ADCs were reported to be more than 90% homogeneous with an average drug-to-antibody ratio of 1.9 and equally active in both vitro and in vivo[14]. However, this method necessitates a reduction-oxidation phase, which may leave native cysteines with residual reactivity. ADC heterogeneity may also be influenced by the presence of unpaired cysteines in expressed IgGs[17].

Genetic Encoding of Unnatural Amino Acids is another strategy that has been used in introducing site-specific modification on proteins. This method enables the ribosome integration of non-canonical amino acids into proteins with functions not ordinarily found in living cells[9]. Proteins generated through this approach possess chemical functionalities that are not usually encountered within a cell, such as alkynes, boronic acids or aryl iodides[7]. Reactions that are highly chemoselective or even mechanistically specific for the endowed chemical functionality are carried out in the second step to site-selectively install modifications on individual proteins in complex biological mixtures, including in living organisms[8]. The insertion of non-natural amino acids into proteins has opened possibilities for site-selective antibody modification[21]. Two

unnatural amino acids, p-azidophenylalanine and p-acetylphenylalanine, have been particularly useful in generating bioconjugates through azide-alkyne cycloaddition and oxime ligation, respectively[19]. ADCs produced through this approach with a drug-to-antibody ratio of ~2.0[22]. Despite the fact that this method allows for the creation of homogeneous conjugates by attaching a drug to almost any accessible site on a mAb, ADCs might cause immunogenic responses[13].

Another approach that has been used to site-specifically attach drugs to antibodies is utilizing enzymes that react with particular amino acid sequences[13]. Transglutaminases (TGs) catalyze the formation of amide bonds between the primary amine of lysine and the amide group of glutamine. Interestingly, the TG from *Streptoverticillium mobaraense* only recognizes the glutamine residue within the constant regions for an antibody referred to as 'glutamine (Q) tag' (that is, LLQG) as opposed to other naturally occurring glutamine residues in the constant regions of the antibody[18]. In the same manner, formylglycine-generating enzymes (FGE) catalyze the selective conversion of a cysteine CXPXR sequence (where X is usually serine, threonine, alanine, or glycine) to an aldehyde which is further functionalized with aminooxy or hydrazine-functionalized molecules[19]. The limitation of this approach is that the hydration of the aldehyde from formylglycine in water to form an unreactive gem diol lowers the yield of the product formed from this approach[20].

3.2. Our Approach: Mimicking the Catalytic Triad

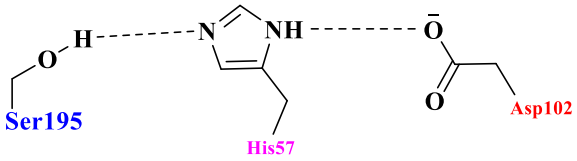


Figure 3.1: The Catalytic Triad

Modifying the local microenvironment around desired reaction sites can improve specificity. Miller *et al.* in the 1990s discovered that some peptide hydroxyl groups have increased intrinsic chemical reactivity and can be O-acylated by biotin N-hydroxysuccinimide esters[24]. They made

an unexpected discovery when they reacted Gonadotropin-releasing hormone (GnRH) and two of its sulfonated N-hydroxysuccinimide esters analogs and did observe the occurrence of multiple biotinylated derivatives for each of the peptides[25]. Using mass spectrometric analysis, they confirmed that there was substantial O-biotinylation of residue Ser4 and to a less minor extent Tyr5 of the GnRH and its analogues. They went ahead to show that the increased reactivity of the seryl hydroxyl groups occurred in the linear triads His-Xaa-Ser or Ser-Xaa-His (Xaa = any amino acid)[26]. They subsequently reported that threonine and tyrosine residues when within the triad motif in a peptide also show enhanced reactivity[24]. Also, they observed the modification of Arg in DMF by biotin N-hydroxysuccinimide esters[27]. They hypothesized that hydrogen bonding could form between hydrogen of the serine hydroxyl group and imidazole nitrogen of the histidine. The hydrogen bonding increases the nucleophilicity of the serine hydroxyl thereby conferring enhanced reactivity on it. They opined that this could be reminiscent of the Asp-His-Ser catalytic triad seen in proteases and esterases[26].

According to data from human genome sequencing, there are about 500 proteases in human cells[29]. Proteases are classified into four categories based on their mechanisms: cysteine, serine, metallo, and aspartic acid proteases[30]. The most well-known class of proteases is the serine protease (EC3.4.21), which uses the classic Ser/His/Asp catalytic triad mechanism, in which serine is the nucleophile, histidine is the general base and acid, and aspartate aids in orienting the histidine residue and neutralizing the charge that develops on it during transition states[30]. There are serine proteases that use catalytic residue arrangements other than the standard triad (Ser/His/Asp). Novel triads, such as Ser/His/Glu, Ser/His/His, or Ser/Glu/Asp; dyads, such as Ser/Lys or Ser/His, or a single Ser catalytic residue are used by these atypical serine proteases. The nucleophilic hydroxyl can also be produced from threonine rather than a serine residue in some proteases[30]. There are

two families of serine proteases: the trypsin and subtilisin families unrelated in sequence and three-dimensional structure, but they use the identical Asp-His-Ser triad for catalysis[31]. Chymotrypsin and trypsin are two of the most well-known serine proteases that use the Ser/His/Asp triad. Ser195 is the catalytic nucleophile in chymotrypsin, His57 is the general base, and Asp102 is the aspartate residue. The oxyanion hole formed by the backbone amide nitrogen of Gly193 and Ser195 and hydrogen bond interactions have been proposed to play important roles.

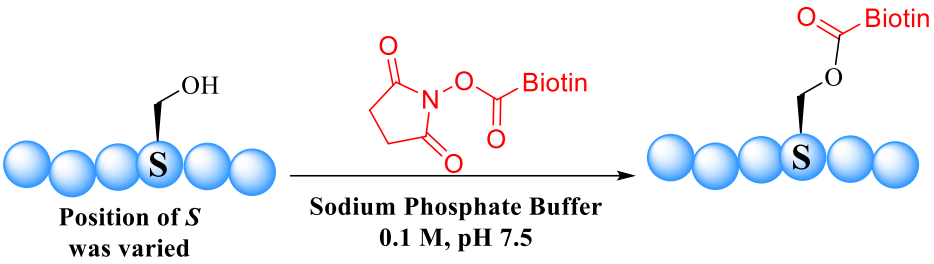
The catalytic triad has also been observed in other enzyme classes: an Asp–His–Ser triad in active site of lipases[32]; a closely related Glu–His–Ser triad at the active site of acetylcholinesterase[33]; serine protease activity in alcohol dehydrogenase of Drosophila[34] and peptide bond-cleaving activity using Ser and His residues in a serine protease-type mechanism has been observed with catalytic autoantibodies in patients with autoimmune disease[35]. Because the catalytic triad can be viewed as an independent catalytic motif linked to different binding sites to perform different hydrolytic functions, Iengar and Ramakrishnan reasoned that rational design and protein engineering to introduce the protease triad into non-proteases can lead to the design of novel catalysts.[31] Inspired by this, we sought to rationally design a peptide sequence which can mimic the catalytic triad to make serine more nucleophile so that it can be modified in a chemoselective and regioselective manner. The knowledge that the ability of enzymes to catalyze reactions is based on the spatial organization of the active atoms, which orchestrates the chemical and structural steps of the reaction[28] encouraged that the right microenvironment can make one serine more reactive over other serine residues in a peptide or protein.

3.3. Design of the HB-Hub by Finetuning the Microenvironment of Serine for Site-Specific Modification

Modifying the local microenvironment around desired reaction sites can improve specificity, but it comes at the cost of protein engineering. Penteluete group has utilized such approach to selectively modify one cysteine site in proteins containing multiple endogenous cysteine residues.[37] The tetrapeptide peptide sequence, *Phe-Cys-Pro-Phe*, which they dubbed the "π-clamp" fine-tunes the reactivity of its cysteine thiol for site-selective conjugation with perfluoroaromatic reagents. With this approach, they were able to site-specifically modify antibodies and cysteine-based enzymes. Encouraged by this, we identified a pentapeptide, H*X*S*X*H (HB-Hub), which mediates specific serine modification in physiological conditions without the need for enzymes or other catalysts. Our strategy relies on a network of hydrogen bonding between serine and histidine to significantly increase the nucleophilicity of the hydroxyl group of serine thereby making it amenable for covalent conjugation with known serine protease inhibitors. The *X* represents any of the amino acids making our approach quite tolerable without compromising the site-selectivity it achieves. We started our investigation by rationally designing peptide sequences that could mimic the catalytic triad by bringing together residues known to mediate the serine proteases' reactivity. We considered the spatial arrangements of these amino acids in the known catalytic triads. In a typical catalytic triad, the aspartate interacts with the histidine residue through hydrogen bonding and puts it in the proper orientation. In this orientation, the imidazole nitrogen of the histidine and serine hydroxyl group engage in relay of hydrogen bonding with the concomitant effect of increased nucleophilicity of the serine hydroxyl such that it is far more reactive than other inert hydroxyl groups on proteins. Serine can then nucleophilically attack and react with known serine protease covalent inhibitors. Thus, our design effort was to place the

relevant amino acids found in the catalytic at proper distances and orientation to maximize the relay of hydrogen bonding. We considered the following factors:

Table 3.1: Identification of the HB-Hub sequence



ENTRY	PEPTIDE SEQUENCE	CONVERSION (%)
a	Ac-SPGHADA-$CONH_2$	55 (1 h)
b	Ac-SAGHADA-$CONH_2$	38 (1 h), 61 (6 h)
c	Ac-SPGHALA-$CONH_2$	67 (1 h), 78 (6 h)
d	Ac-SPGFADA-$CONH_2$	NR
e	NH_2-SPGFADA-$CONH_2$	82 (1 hr) >99 (6 hr) (NH_2 modified)
f	Ac-HASGHADA-$CONH_2$	75 (1 hr) >99 (6 hr)
g	Ac-HAGSGAHA-$CONH_2$	32 (1 hr) 82 (6 hr)
h	Ac-SHALA-$CONH_2$	NR
i	Ac-HSHA-$CONH_2$	NR
j	Ac-HDSDHA-$CONH_2$	84 (1 h)
k	Ac-HNSNHA-$CONH_2$	75
l	Ac-HASDHA-$CONH_2$	75
m	Ac-HWSWHA-$CONH_2$	87
n	Ac-HDSWHA-$CONH_2$	88
o	Ac-HASGHMFADSG-$CONH_2$	90
p	Ac-GTNFMHDSDHA-$CONH_2$	82
q	Ac-GYNFMHDSDHA-$CONH_2$	85
r	Ac-ASFMSLAFMHDSDHA-$CONH_2$	85
s	Ac-ASAWMHDSDHMFNSA-$CONH_2$	64
t	Ac-AHDSDHWVQSMFNSA-$CONH_2$	80

Peptide (2 mM), Biotin-NHS ester (10 mM), pH 7.3, 0.1M

Conversion determined by HPLC and MS analysis

3.3.1. Effect of Turn-Inducer

We first investigated the role of turn inducers in bringing these hydrogen-bonding partners in proximity. We reasoned that a peptide with a turn-inducer could bring serine closer to histidine and we synthesized two peptides: one with a turn-inducer, Ac-SPGHADA-$CONH_2$ **1a,** and one without a turn inducer, Ac-SAGHADA-$CONH_2$, **1b**. For the proof of concept, we started our preliminary investigation with Biotin NHS ester (BIO-NHS) as the electrophile source. So, these two peptides were exposed to Biotin-NHS ester, and much to our delight we observed modest modification of the peptide with the turn-inducer achieving 55 % conversion (as determined by HPLC and MS analysis) to the product **2a** in 1 hour while it took the peptide without a turn-inducer 6 hours to achieve a similar result (61 %) (*Table 3.1*). Thus, the result suggested that the turn-inducer element influenced the reactivity of the serine.

3.3.2. Role of Histidine and Aspartic acid

Having established this, we went on to examine the individual contribution of each of the catalytic triad amino acids. So, we mutated the peptide **1a** by replacing aspartic acid with leucine to get Ac-SPGHALA-$CONH_2$ **1c** and observed an increased reactivity (67 % in one hour) in the presence of Biotin-NHS ester (*Table 5.1*). However, when the same peptide, **1a** was mutated by substituting histidine with phenylalanine to afford Ac-SPGFADA-$CONH_2$ **1d** no reaction was observed when subjected to our reaction conditions. Thus, these results show that histidine is responsible for mediating the increased reactivity of serine with Biotin-NHS with no contribution from aspartic acid. To compare the reactivity of serine versus that of a primary amine with Biotin-NHS ester, we allowed an analog of **1d**, NH_2-SPGFADA-$CONH_2$ **1f** with a free N-terminal amine to react with our probe and found slightly higher conversion to the product (82 % in 1 hour) (*Table 3.1*).

3.3.3. Effect of the Distance between Serine and Histidine in the Peptide Motif

Being convinced that we can further increase the intrinsic reactivity of serine, we sought to optimize the relay of hydrogen bonding in our peptide motif by flanking the serine residue with histidine on both sides with one amino acid apart, Ac-HASGHADA-$CONH_2$ **1f.** Much to our delight, when the Biotin NHS ester probe was allowed to react with **1f** we observed complete conversion to the desired product in 6 hours and 75 % in 1 hour. With this knowledge, we studied the relevant distance for optimal reaction our peptides with Biotin NHS ester. With two amino acids between serine and histidine on both sides (Ac-HAGSGAHA-$CONH_2$), there was poor conversion to the product in 1 hour (32 %) which significantly increased after 6 hours (82 %) (*Table 3.1*). However, no reaction was observed when histidine was next to serine (Ac-SHALA-$CONH_2$ **1h**) nor when two histidine residues were next to serine on both sides of the peptide (Ac-HSHA-$CONH_2$ **1i**). This we hypothesized could be due to the side chains of histidine and serine being in opposite orientations because of the trans nature of the peptide bond. Taken together, we found an ideal motif, HXSXH which significantly enhances the reactivity of the hydroxyl group of serine with electrophiles such as Biotin NHS ester. This unique motif we call HB-Hub. It is worthy of note that this motif lacks a turn-inducer element and has no requirement for aspartic acid.

3.3.4. Impact of the Amino Acid in-between Serine and Histidine

Though we have established that the amino acids between serine and histidine in the HB-Hub can be very flexible, we sought to further ascertain if the reactivity of serine can be improved by unique amino acids. We started by placing aspartic acid in-between serine and histidine on both sides of serine and observed improved modification (84 %) of the linear peptide (Ac-HDSDHA-$CONH_2$ **1j**), 75 % conversion with two asparagine in-between (Ac-HNSNHA-$CONH_2$ **1k**), and 87 % conversion with two tryptophan in-between (Ac-HWSWHA-$CONH_2$ **1m**) and we observed

similar conversion when one of the tryptophan residues in **1m** was replaced with aspartic acid to afford (Ac-HDSWHA-$CONH_2$ **1n**). Thus, it is evident from the result that the HB-Hub motif tolerates different types of the canonical amino acids with slightly increased reactivity when aspartic acid and tryptophan are between serine and tryptophan in the HB-Hub. This is surprising since the amino acids adjacent to the serine in the HB-Hub are out of phase with serine in orientation. More computational study is required to account for this observation.

3.3.5. Selective Modification of Serine in the HB-Hub over other Serine Residues in a Peptide

To be worth its salt, the HB-Hub should confer selective reactivity on serine found within its motif over other serine/threonine/tyrosine residues on proteins and peptides. To evaluate this, we started by ruling out cross-reactivity with threonine and tyrosine by synthesizing two peptides (Ac-GTNFMHDSDHA-$CONH_2$ **1p** and Ac-GYNFMHDSDHA-$CONH_2$ **1q**) containing HB-Hub but differing only in possessing either tyrosine or threonine in a different site on the peptides; and a third peptide, Ac-HASGHMFADSG-CONH2 (**1o**) with two serine residues with one within the HB-Hub motif. These peptides were reacted with our Biotin NHS ester, and we observed a single modification of the peptides with high conversion to the products in one hour (*Table 3.1*). Next, we synthesized three peptides, each with three serine residues on it with one of the serine residues in the HB-Hub: Ac-ASFMSLAFMHDSDHA-$CONH_2$ **1r**, Ac-ASAWMHDSDHMFNSA-$CONH_2$ **1s** and Ac-AHDSDHWVQSMFNSA-$CONH_2$ **1t.** Under our reaction conditions with Biotin NHS ester probe, we observed excellent modification of the peptides in 1 hour with 85 %, 64 %, and 80 % respectively. Taken together, we achieved selective modification of serine residue within the HB-Hub over other serine residues, tyrosine, and threonine. Also, the HB-Hub can be tolerated at different positions within the peptide with the C-terminus and N-terminus being slightly more favorable.

3.3.6. Confirmation of the Site of Modification

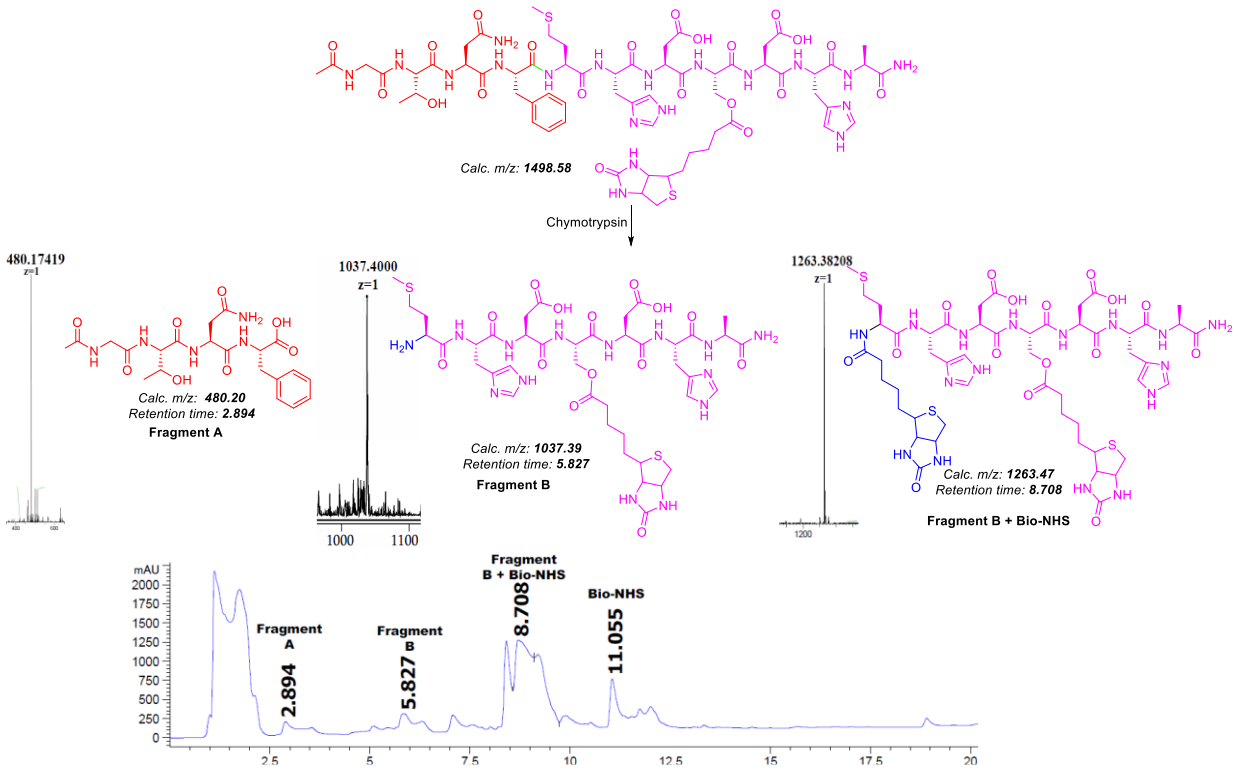


Figure 3.2: Chymotrypsin mediated cleavage of Ac-GTNFMHDSDHA-Biotin revealing the part of the peptide containing the biotinylated residue.

Having confirmed the modification of peptides containing the HB-Hub using HPLC and MS spectrometry analysis, we went ahead to confirm the exact site being modified on the peptide using two approaches: enzymatic degradation of the peptide and tandem mass spectrometry (MS/MS). In the first approach, the Ac-GTNFMHDSDHA-$CONH_2$ **1p** containing serine and threonine; and phenylalanine within the peptide sequence was reacted with Biotin NHS ester and allowed to reach completion. A solution of chymotrypsin was introduced into the reaction mixture and stirred for 2 hours. Chymotrypsin is known to catalyze the hydrolysis of the C-terminus of the aromatic amino acids on peptides and proteins, so, we looked for possible proteolysis of the modified peptide into two fragments. As expected, we identified through HPLC analysis a peak at 2.894 minutes with *m/z* values 480.17419 corresponding to **fragment A** which is the N-terminal part of the peptide

containing threonine. The fragment accordingly did not have any modification on them meaning that the hydroxyl group on threonine was inert to our condition. Also, at 5.827 minutes, we identified a fragment with the *m/z* 1037.4000 corresponding to the expected mass of fragment B with modification on the serine hydroxyl group. Interestingly, we observed further biotinylation of the fragment B (retention time of 8.708 minutes) following the enzymatic cleavage since the Bio-NHS ester was used in excess for the reaction. While this result was encouraging, it could not be conclusive evidence for the site of the reaction on the peptide.

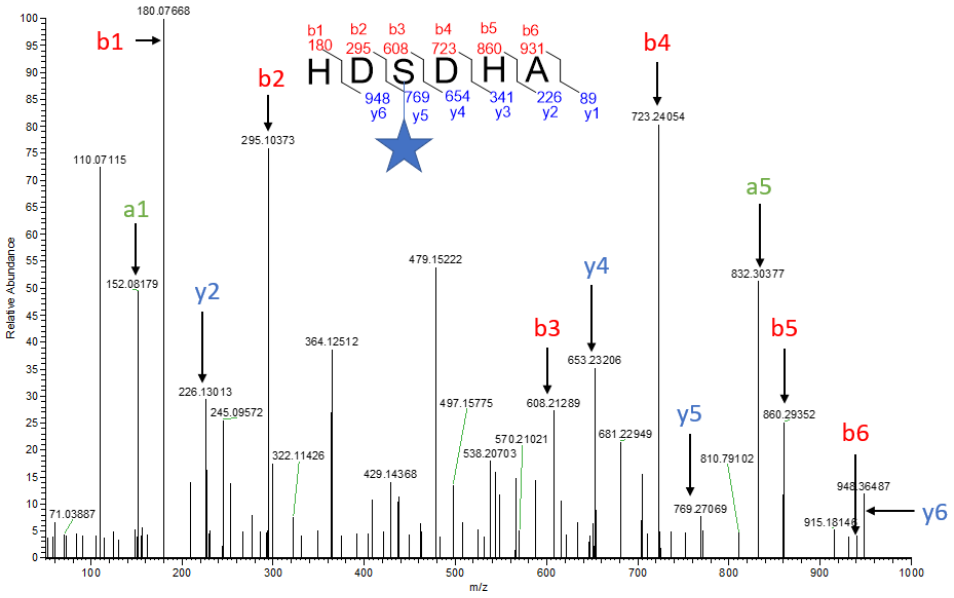


Figure 3.3: Tandem mass (MS/MS) analysis showing serine as the site of modification on the modified peptide, Ac-HDSDHA **2j**

Thus, we further employed MS/MS analysis to confidently identify the exact site of modification on the peptide. To achieve this, we carried out MS/MS on the bioconjugate **2j** being Ac-HDSDHA-$CONH_2$ **1j** modified with Biotin NHS ester. To our delight, we completely identified all the *b ions* and most of the *y ions* that correspond to the modification of serine on the peptide (*Figure 5.2*). The most relevant ions that gave us confidence in this analysis were *b3*, *b4*, *b5*, *b6*,

y4, *y5*, and *y6* as they represent the segment of the peptide containing an additional 226 representing the modification of the serine residue. Taken together, the combination of the enzymatic cleavage of **2p** showing the modification of fragment B and MSMS identifying serine as the site of modification, confirmed that the HB-Hub makes the otherwise inert hydroxyl group on serine a reactive site that can be modified in the presence of other hydroxyl groups on peptides.

3.4. Screening of Known Protease Inhibitors for Selective Reaction with Peptides with HB-Hub

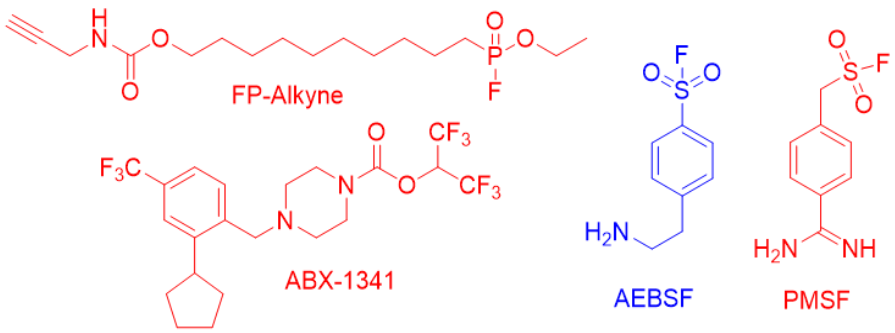
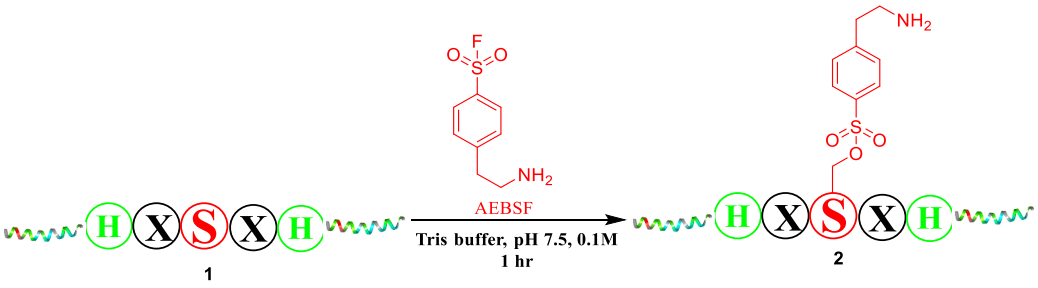


Figure 3.4: Known serine protease inhibitors screened for reactivity with peptides containing the HB-Hub

Our overarching goal was to find the right peptide motif that could mimic the catalytic triad such that they can site-specifically react with serine protease inhibitors. Luckily, a lot of groups have synthesized such probes with most being commercially available. Thus, we obtained four known serine protease inhibitors (**Figure 3.4**) that have been demonstrated in the literature to react covalently with active sites of serine proteases. These probes are believed to irreversibly inhibit serine proteases by acylation of the serine on the active site of the enzymes.[38] We started our investigation by reacting them with our model peptide, **1A** (*Table 3.2*). With a series of conditions, we were unable to observe any modifications on the peptide with FP-alkyne (fluorophosphonate probe), ABX-1341 (carbamate-based probe) and PMSF (sulfonyl fluoride-based probe). However, AEBSF showed moderate reactivity with **1A** with 48 % conversion as determined by MS and HPLC analysis (*Table 3.2*). Though PMSF is reported to be more reactive than AEBSF, it is not

readily soluble in water and rapidly degrades in aqueous solutions.[39] In our hands, we observed the insolubility challenge and that could account for the lack of reactivity as we had to pre-dissolve it in organic solvents before use and in all the cases precipitated out of solution once the buffer is added. The same was obtainable with both FP-alkyne and ABX-1341 as well.

Table 3.2: Reactions between HB-Hub peptides and AEBSF, a serine protease probe



ENTRY	SEQUENCE	CONVERSION (%)
A	Ac-HASGHADA-$CONH_2$	Single modification: 48 Double modification:15 Triple modification:3
B	Ac-HWSWHA-$CONH_2$	Single modification: 33 Double modification:28 Triple modification:2
C	Ac-HASGHMFADSG-$CONH_2$	Single modification: 40 Double modification:15 Triple modification:2
D	Ac-GTNFMHDSDHA-$CONH_2$	Single modification: 47 Double modification:13 Triple modification:2
E	Ac-GYNFMHDSDHA-$CONH_2$	Single modification: 47 Double modification:22 Triple modification:3
F	Ac-ASFMSLAFMHDSDHA-$CONH_2$	Single modification: 42 Double modification:11
G	Ac-ASAWMHDSDHMFNSA-$CONH_2$	Single modification: 45 Double modification:9 Triple modification:2
H	Ac-AHDSDHWVQSMFNSA-$CONH_2$	Single modification: 45 Double modification:4 Triple modification:1

Next, we tried out the reaction of AEBSF with peptides containing two hydroxyl groups in their chain sides with one serine residue located within the HB-Hub, **1C, 1D** and **1E.** The peptides were moderately modified with percentage conversion to the product of 40, 47 and 47 respectively

(*Table 3.2*). Following this, we examined whether the position of the HB-Hub in the linear sequence of the peptide affects the reactivity of the serine by carrying out reactions with **1F**, **1G** and **1H** , which have the HB-Hub positioned at the C-terminus, middle of the peptide and N-terminus respectively. Interestingly, we registered a similar conversion (42-45 %) to the product with these peptides (*Table 3.2*) suggesting that the position of the HB-Hub does not impart the reactivity of the activated serine with AEBSF.

Unsurprisingly, we observed that most of peptides were doubly and triply modified. From literature reports, AEBSF has been reported to modify not only reactive serine, but also context-specific threonine, lysine, tyrosine, cysteine, and histidine residues. Since our HB-Hub has two histidine residues within the HB-Hub motif, we reasoned that the histidine could be undergoing modification to a lesser extent. To ascertain whether the major modification was occurring on the serine in the HB-Hub, we carried out an MS/MS analysis on the singly modified peptide Ac-GTNFMHDSDHA **2D.** The analysis clearly confirmed that the modification was on serine within the HB-Hub as indicated by the the *b9, y4, y5 ions* (*Figure 5.3*). Also, the MS/MS analysis of the doubly modified product confirmed that the second modification was on histidine closer to the C-terminus of the peptide. The *y2, y3* and *b10 ions* gave us confidence in this assignment. Excitingly, the *b* and *y ions* clearly showed that threonine was not modified in both products. Unfortunately, we were unable to generate good MSMS data from the triply modified product probably due to the low abundance of its occurrence.

Thus, with the MSMS analysis, it was confirmed with confidence that the major modification occurs site-specifically at the activated serine within the HB-Hub. Work is still ongoing towards site-specifically modifying a protein with engineered HB-Hub motif on it. We envisage that such

can be replicated on antibodies and thus, our chemistry can be applied as a strategy for attaching moieties to antibodies in ADCs.

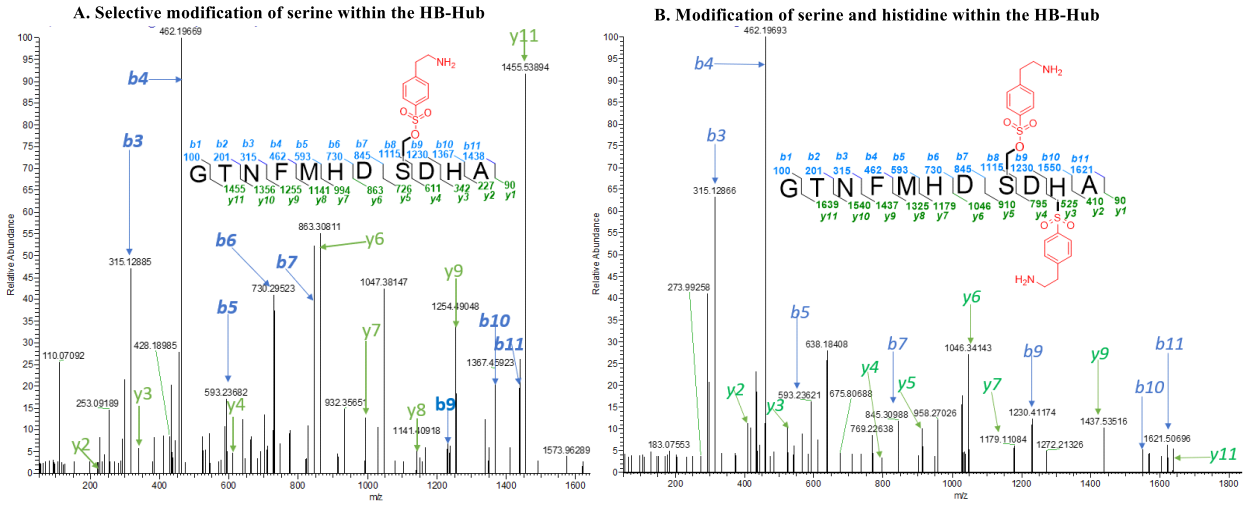


Figure 3.5: Tandem mass (MS/MS) of the products formed from the reaction reaction of AEBSF and Ac-GTNFMHDSDHA **1D.** A. The singly modified product is confidently mapped to the serine within the HB-Hub by the y4 and b9. B. The singly modified product is confidently mapped to the serine (by the *b9 and y4, y5* ions) and histidine (*y2, y3* and *b10*) within the HB-Hub.

CHAPTER THREE REFERENCES

(1) Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nature Chemistry* **2016**, *8* (2), 103-113. DOI: 10.1038/nchem.2393.

(2) Spicer, C. D.; Davis, B. G. Selective chemical protein modification. *Nature Communications* **2014**, *5* (1), 4740. DOI: 10.1038/ncomms5740.

(3) Chalker, J. M.; Bernardes, G. J.; Lin, Y. A.; Davis, B. G. Chemical modification of proteins at cysteine: opportunities in chemistry and biology. *Chem Asian J* **2009**, *4* (5), 630-640. DOI: 10.1002/asia.200800427

(4) Shadish, J. A.; DeForest, C. A. Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter* **2020**, *2* (1), 50-77. DOI: https://doi.org/10.1016/j.matt.2019.11.011.

(5) Strebhardt, K.; Ullrich, A. Paul Ehrlich's magic bullet concept: 100 years of progress. *Nat Rev Cancer* **2008**, *8* (6), 473-480. DOI: 10.1038/nrc2394

(6) Aretin, M.-B. Antibody–drug conjugates—the magic bullet? *memo - Magazine of European Medical Oncology* **2022**. DOI: 10.1007/s12254-021-00780-8.

(7) Baslé, E.; Joubert, N.; Pucheault, M. Protein Chemical Modification on Endogenous Amino Acids. *Chemistry & Biology* **2010**, *17* (3), 213-227. DOI: https://doi.org/10.1016/j.chembiol.2010.02.008.

(8) Elliott, T. S.; Townsley, F. M.; Bianco, A.; Ernst, R. J.; Sachdeva, A.; Elsässer, S. J.; Davis, L.; Lang, K.; Pisa, R.; Greiss, S.; et al. Proteome labeling and protein identification in specific

tissues and at specific developmental stages in an animal. *Nat Biotechnol* **2014**, *32* (5), 465-472. DOI: 10.1038/nbt.2860

(9) Davis, L.; Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nature Reviews Molecular Cell Biology* **2012**, *13* (3), 168-182. DOI: 10.1038/nrm3286.

(10) Walsh, S. J.; Bargh, J. D.; Dannheim, F. M.; Hanby, A. R.; Seki, H.; Counsell, A. J.; Ou, X.; Fowler, E.; Ashman, N.; Takada, Y.; et al. Site-selective modification strategies in antibody-drug conjugates. *Chem Soc Rev* **2021**, *50* (2), 1305-1353. DOI: 10.1039/d0cs00310g

(11) Wang, L.; Amphlett, G.; Blättler, W. A.; Lambert, J. M.; Zhang, W. Structural characterization of the maytansinoid–monoclonal antibody immunoconjugate, huN901–DM1, by mass spectrometry. *Protein Science* **2005**, *14* (9), 2436-2446. DOI: https://doi.org/10.1110/ps.051478705.

(12) Walsh, S. J.; Bargh, J. D.; Dannheim, F. M.; Hanby, A. R.; Seki, H.; Counsell, A. J.; Ou, X.; Fowler, E.; Ashman, N.; Takada, Y.; et al. Site-selective modification strategies in antibody–drug conjugates. *Chemical Society Reviews* **2021**, *50* (2), 1305-1353, 10.1039/D0CS00310G. DOI: 10.1039/D0CS00310G.

(13) Chudasama, V.; Maruani, A.; Caddick, S. Recent advances in the construction of antibody–drug conjugates. *Nature Chemistry* **2016**, *8* (2), 114-119. DOI: 10.1038/nchem.2415.

(14) Kaur, S.; Xu, K.; Saad, O. M.; Dere, R. C.; Carrasco-Triguero, M. Bioanalytical assay strategies for the development of antibody-drug conjugate biotherapeutics. *Bioanalysis* **2013**, *5* (2), 201-226. DOI: 10.4155/bio.12.299

(15) Reardon, D. A.; Zalutsky, M. R.; Akabani, G.; Coleman, R. E.; Friedman, A. H.; Herndon, J. E., 2nd; McLendon, R. E.; Pegram, C. N.; Quinn, J. A.; Rich, J. N.; et al. A pilot study: 131I-

antitenascin monoclonal antibody 81c6 to deliver a 44-Gy resection cavity boost. *Neuro Oncol* **2008**, *10* (2), 182-189. DOI: 10.1215/15228517-2007-053

(16) Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; et al. Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nature Biotechnology* **2008**, *26* (8), 925-932. DOI: 10.1038/nbt.1480.

(17) Chumsae, C.; Gaza-Bulseco, G.; Liu, H. Identification and localization of unpaired cysteine residues in monoclonal antibodies by fluorescence labeling and mass spectrometry. *Anal Chem* **2009**, *81* (15), 6449-6457. DOI: 10.1021/ac900815z

(18) Jeger, S.; Zimmermann, K.; Blanc, A.; Grünberg, J.; Honer, M.; Hunziker, P.; Struthers, H.; Schibli, R. Site-specific and stoichiometric modification of antibodies by bacterial transglutaminase. *Angew Chem Int Ed Engl* **2010**, *49* (51), 9995-9997. DOI: 10.1002/anie.201004243

(19) Wu, P.; Shui, W.; Carlson, B. L.; Hu, N.; Rabuka, D.; Lee, J.; Bertozzi, C. R. Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. *Proceedings of the National Academy of Sciences* **2009**, *106* (9), 3000-3005. DOI: doi:10.1073/pnas.0807820106.

(20) Rabuka, D.; Rush, J. S.; deHart, G. W.; Wu, P.; Bertozzi, C. R. Site-specific chemical protein conjugation using genetically encoded aldehyde tags. *Nature Protocols* **2012**, *7* (6), 1052-1067. DOI: 10.1038/nprot.2012.045.

(21) Axup, J. Y.; Bajjuri, K. M.; Ritland, M.; Hutchins, B. M.; Kim, C. H.; Kazane, S. A.; Halder, R.; Forsyth, J. S.; Santidrian, A. F.; Stafin, K.; et al. Synthesis of site-specific antibody-drug

conjugates using unnatural amino acids. *Proceedings of the National Academy of Sciences* **2012**, *109* (40), 16101-16106. DOI: doi:10.1073/pnas.1211023109.

(22) Zimmerman, E. S.; Heibeck, T. H.; Gill, A.; Li, X.; Murray, C. J.; Madlansacay, M. R.; Tran, C.; Uter, N. T.; Yin, G.; Rivers, P. J.; et al. Production of Site-Specific Antibody–Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System. *Bioconjugate Chemistry* **2014**, *25* (2), 351-361. DOI: 10.1021/bc400490z.

(23) Jain, N.; Smith, S. W.; Ghone, S.; Tomczuk, B. Current ADC Linker Chemistry. *Pharm Res* **2015**, *32* (11), 3526-3540. DOI: 10.1007/s11095-015-1657-7

(24) Miller, B. T.; Rogers, M. E.; Smith, J. S.; Kurosky, A. Identification and Characterization of O-Biotinylated Hydroxy Amino Acid Residues in Peptides. *Analytical Biochemistry* **1994**, *219* (2), 240-248. DOI: https://doi.org/10.1006/abio.1994.1263.

(25) Miller, B. T.; Collins, T. J.; Nagle, G. T.; Kurosky, A. The occurrence of O-acylation during biotinylation of gonadotropin-releasing hormone and analogs. Evidence for a reactive serine. *Journal of Biological Chemistry* **1992**, *267* (8), 5060-5069. DOI: https://doi.org/10.1016/S0021-9258(18)42730-5.

(26) Miller, B. T.; Kurosky, A. Elevated Intrinsic Reactivity of Seryl Hydroxyl Groups within the Linear Peptide Triads His-Xaa-Ser or Ser-Xaa-His. *Biochemical and Biophysical Research Communications* **1993**, *196* (1), 461-467. DOI: https://doi.org/10.1006/bbrc.1993.2272.

(27) Miller, B. T.; Collins, T. J.; Rogers, M. E.; Kurosky, A. Peptide Biotinylation with Amine-Reactive Esters: Differential Side Chain Reactivity. *Peptides* **1997**, *18* (10), 1585-1595. DOI: https://doi.org/10.1016/S0196-9781(97)00225-8.

(28) Dodson, G.; Wlodawer, A. Catalytic triads and their relatives. *Trends Biochem Sci* **1998**, *23* (9), 347-352. DOI: 10.1016/s0968-0004(98)01254-7

(29) Puente, X. S.; Sánchez, L. M.; Gutiérrez-Fernández, A.; Velasco, G.; López-Otín, C. A genomic view of the complexity of mammalian proteolytic systems. *Biochem Soc Trans* **2005**, *33* (Pt 2), 331-334. DOI: 10.1042/bst0330331

(30) Ekici, O. D.; Paetzel, M.; Dalbey, R. E. Unconventional serine proteases: variations on the catalytic Ser/His/Asp triad configuration. *Protein Sci* **2008**, *17* (12), 2023-2037. DOI: 10.1110/ps.035436.108

(31) Iengar, P.; Ramakrishnan, C. Knowledge-based modeling of the serine protease triad into non-proteases. *Protein Eng* **1999**, *12* (8), 649-656. DOI: 10.1093/protein/12.8.649

(32) Blow, D. More of the catalytic triad. *Nature* **1990**, *343* (6260), 694-695. DOI: 10.1038/343694a0.

(33) Sussman, J. L.; Harel, M.; Frolow, F.; Oefner, C.; Goldman, A.; Toker, L.; Silman, I. Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. *Science* **1991**, *253* (5022), 872-879. DOI: 10.1126/science.1678899

(34) Fibla, J.; Atrian, S.; Gonzàlez-Duarte, R. Evidence of serine-protease activity closely associated with Drosophila alcohol dehydrogenase. *Eur J Biochem* **1993**, *211* (1-2), 357-365. DOI: 10.1111/j.1432-1033.1993.tb19905.x

(35) Paul, S.; Volle, D. J.; Beach, C. M.; Johnson, D. R.; Powell, M. J.; Massey, R. J. Catalytic hydrolysis of vasoactive intestinal peptide by human autoantibody. *Science* **1989**, *244* (4909), 1158-1162. DOI: 10.1126/science.2727702

(36) Dorywalska, M.; Strop, P.; Melton-Witt, J. A.; Hasa-Moreno, A.; Farias, S. E.; Galindo Casas, M.; Delaria, K.; Lui, V.; Poulsen, K.; Sutton, J.; et al. Site-Dependent Degradation of a Non-Cleavable Auristatin-Based Linker-Payload in Rodent Plasma and Its Effect on ADC Efficacy. *PLoS One* **2015**, *10* (7), e0132282-e0132282. DOI: 10.1371/journal.pone.0132282

(37) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. Π-Clamp-mediated cysteine conjugation. *Nature chemistry* **2016**, *8* (2), 120-128. DOI: 10.1038/nchem.2413

(38) James, G. T. Inactivation of the protease inhibitor phenylmethylsulfonyl fluoride in buffers. *Analytical Biochemistry* **1978**, *86* (2), 574-579. DOI: https://doi.org/10.1016/0003-2697(78)90784-4.

(39) Narayanan, A.; Jones, L. H. Sulfonyl fluorides as privileged warheads in chemical biology. *Chemical Science* **2015,** 6 (5), 2650-2659, 10.1039/C5SC00408J. DOI: 10.1039/C5SC00408J.