AN ABSTRACT OF THE CAPSTONE REPORT OF

Jieliang Wang for the degree of Master of Chemical Sciences

Title: *Peptide Synthesis and Modification as a Versatile Strategy for Probes Construction*

Project conducted at: *University of Pennsylvania,* 231 S. 34 Street, Philadelphia, PA 19104-6323
Supervisor: Professor *E. James Petersson*
Dates of Project: *May 15 2016 to August 7 2017*

Abstract approved:
E. James Petersson, Academic Advisor

Peptide synthesis and modification is a versatile chemical biology strategy to construct probes and sensors of a variety of types of biological activity, including protein/protein interactions, protein localization, and proteolysis. In my thesis work, I have made probes for three distinct biological applications. To do so, I have used a combination of solid phase peptide synthesis (SPPS), native chemical ligation (NCL), protein expression, and *S*-alkylation to construct probes with desired functional groups, while minimizing the perturbation to the native structure. In the first project, I constructed photo-crosslinking probes to study the difference in protein-protein interactions of N-terminal acetylated (N-ac) histone H4 peptide versus non-acetylated histone H4 peptide. One protein was identified by Western blot with binding preference to N-Ace histone H4 peptide. In the second separate project, I constructed probes to study the toxicity mechanism of proline/arginine dipeptide PR_x from amyotrophic lateral sclerosis (ALS) associated gene *C9ORF72*. The preliminary result suggested the PR_x peptide toxicity on proteasome depends on the length of the $(PR)_x$ peptide. In the third project, I synthesized fluorescence sensors to study the positional effects of thioamide on the proteolysis process of chymotrypsin. From hydrolysis studies, my coworkers and I determined that thioamide incorporation at the P1 or P2 positions can greatly inhibit chymotrypsin proteolysis.

Peptide Synthesis and Modification as a Versatile Chemical Tool for Probes Construction

by

Jieliang Wang

A CAPSTONE REPORT

submitted to the

University of Pennsylvania

in partial fulfillment of
the requirements for
the degree of

Masters of Chemical Sciences

Presented *August 9, 2017*

Commencement *August 2017*

Master of Chemical Sciences Capstone Report of *Jieliang Wang*
presented on August 9 2017.

APPROVED:

E. James Petersson, representing Biological Chemistry

I understand that my Capstone Report will become part of the permanent collection of the University of Pennsylvania Master of Chemical Sciences Program. My signature below authorizes release of my final report to any reader upon request.

Jieliang Wang, Author

Acknowledgements

I would like to express my special appreciation and thanks to my advisor Professor E. James Petersson, who have been a tremendous advisor and life coach for me. Despite my training in the Petersson lab is only one year, Dr. Petersson sculptured my thinking of science and guided me to put ideas into practice. I am sure the precious experience will benefit me throughout my life.

I would like to thank my mentors Dr. Xing “Stella” Chen and Daniel Miklos Szantai-Kis, for training me with all the techniques involved in this work. Also for Chunxiao Liu and Taylor Barrett, for synthesizing peptides that were used in this project. I would also like to thank our collaborators, Dr. Ronen Marmorstein at the Perelman School of Medicine at the University of Pennsylvania and his graduate student Gleb Bazilevsky; Dr. Robert G. Kalb at The Children’s Hospital of Philadelphia and his student Matthews Lan. My projects will be useless without their outstanding work in biological experiments.

Finally, I would like to say thank you to all current and previous members of the Petersson laboratory for answering my questions and forgiveness when every time I made a mistake.

Table of Contents

List of Figures

List of Tables

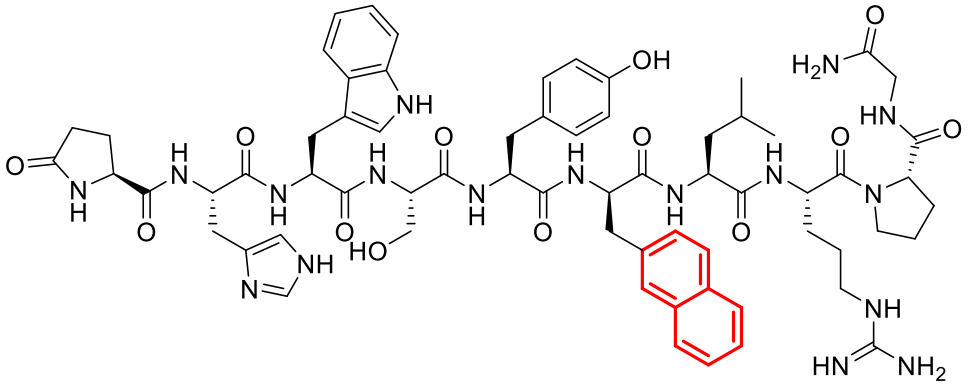
List of Appendices

Chapter 1
Development of Protein Binding Probes

Introduction

Peptide synthesis and protein modification are powerful synthetic strategies that combine the flexibility of organic synthesis and the convenience of recombinant protein expression. By using a combination of synthetic approaches, functional groups can be added to the peptides or proteins of interest, such as fluorophores, unnatural amino acids, photo-crosslinkers, and purification tags. Similarly, synthetic approaches can be used to modify the backbone of the peptide. Modified peptides have wide applications in drug design, live imaging, and other biological studies. For example, Nafarelin® is an analog of the naturally occurring gonadotropin-releasing hormone (GnRH) with modification at position 6 (shown in red) in **Figure 1A**.[1] This modification significantly extends the half-life from 2-4 minutes[2] to 2.6-4 hours.[1] Another example from Nguyen and coworkers[3] showed a caged TEV (*Tobacco Etch* Virus) protease can be activated by release of the caging group under light irradiation, allowing manipulation of protease activity by light, as shown in **Figure 1B.**

A



B

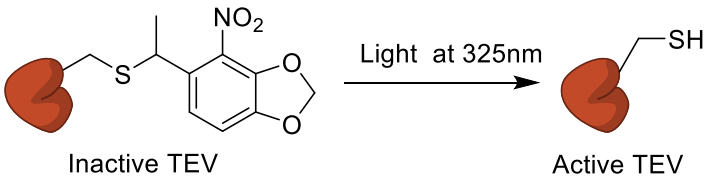


Figure 1. Examples of Modified Peptide and Proteins. (A): a GnRH analog Nafarelin, marketed as Synarel® by Pfizer.[1] (B): The TEV protease is inactivated with a caged cysteine.[3]

I hope to use peptide synthesis and modification methods to construct probes to solve biological questions. Based on the probes applications, this study is divided into projects 1 and 2. Project 1 involves synthesis of photo-crosslinking probes and then studies the functions of histone H4 N-terminal acetylation (N-ac). Project 2 involves synthesis of a photo-crosslinking probe and a fluorescence probe, followed by study on the toxicity mechanism of proline-arginine tandem repeat peptides PR_x on amyotrophic lateral sclerosis (ALS). Since project 1 and 2 uses similar methods including solid phase peptide synthesis (SPPS), native chemical ligation (NCL), *S*-alkylation and maleimide labeling, the methods will be discussed in one section.

Project 1: Interactions of N-Terminally Acetylated Histones

Importance and Significance Although histone N-terminal acetylation (N-ac) was discovered in 1964, the functions of this modification have not been fully characterized.[4] Recently, Marmorstein and co-workers[5] characterized the molecular basis of an N-terminal acetyltransferase NatD, which only acetylates histone H4 and H2A9. Initial evidence suggests that histone N-ac can alter histone in protein bindings, but the potential binding partner(s) was unidentified.[5] Identifying the binding partner(s) can help reveals the functions and mechanism of gene regulation.

Purpose and Hypothesis The purpose of project 1 is to identify the binding partner(s) of N-terminally acetylated histones. The Marmorstein group hypothesized N-ac of histone H4 could change the binding affinity to its binding partner(s) by N-ac, to trigger downstream activities. Therefore, there should be a difference in protein binding between non-acetylated histone H4 versus N-ac histone H4. It was anticipated that at least one protein target could be identified from the photo-crosslinking experiment, possibly a transcription factor or an enzyme that deacetylates the N-ac histone.

Experimental Design To find the putative unknown binding partner(s) regulated by histone H4 N-ac, a photo-crosslinking experiment to crosslink histone H4 with proteins in the cell lysates was proposed. By comparing the binding partner(s) between non-acetylated with N-ac histone H4, protein(s) that shows binding preferences towards one type of histone H4 could be identified. The photo-crosslinking experiment requires a model that mimics the native structure of histone H4. The Petersson laboratory devised a peptide probe consists of the first 22 N-terminal residues of the human histone H4 protein.[6] Only the first 22 residues were chosen to represent the histone H4 protein because the full length of histone H4 protein is over 100 amino acids long, and synthesis of full-length protein is far beyond the capacity of SPPS without the aid of ligation methods. To functionalize the peptide probe, a photo-lysine residue and a biotin tag were incorporated into the sequence. The photo-lysine is a diazirine photo-crosslinker, developed by Yang and co-workers.[7] In photo-lysine, the diazirine group is embedded in the lysine sidechain, having close structural proximity to the native lysine. The Petersson laboratory previously synthesized histone H4 peptide with benzophenone photo-crosslinker, then a photo-crosslinking experiment was

performed by the Marmorstein group. Due to non-specific binding, no useful data could be extrapolated (Marmorstein laboratory, University of Pennsylvania, PA, USA, unpublished work, 2016). Compared to the benzophenone, the photo-lysine can replace the natural lysine residue with minimal perturbation to the native peptide, and hence generate more biologically relevant data. The diazirine is also more reactive than benzophenone and the binding study requires a shorter UV irradiation time. The reduction of time can also reduce non-specific binding between the probe and the cell lysate. Therefore, the photo-lysine was chosen as the photo crosslinker in this study. The biotin tag was introduced at the C-terminal end of the peptide probe to pair with the photo-lysine for pull down experiments. My task in this study was to synthesize the probe, and the photo-crosslinking experiment was performed by the Marmorstein laboratory.

Background in Histones N-Terminal Acetylation In eukaryotic cells, histones provide the scaffolds for wrapping and compacting DNA and to form the structure of nucleosomes. Histones play critical roles in gene expression and are highly regulated by amino acid sidechain post-translational modifications (PTMs), which have been studied extensively in the past few decades. Different PTMs patterns can trigger distinct downstream effects.[8] Therefore, the function of a specific PTMs is only meaningful in the context of a peptide or protein. A particularly elegant approach to study the function of PTMs patterns is to use synthetic peptides with defined amino acid sidechain modifications. For example, Muir and co-workers[9] synthesized histone proteins H2B with ubiquitylated lysine 120 and then evaluated the effect on protein recruitment and gene transcription. In many cases, PTMs in this histone protein resulted in the recruitment of target proteins for gene regulation.

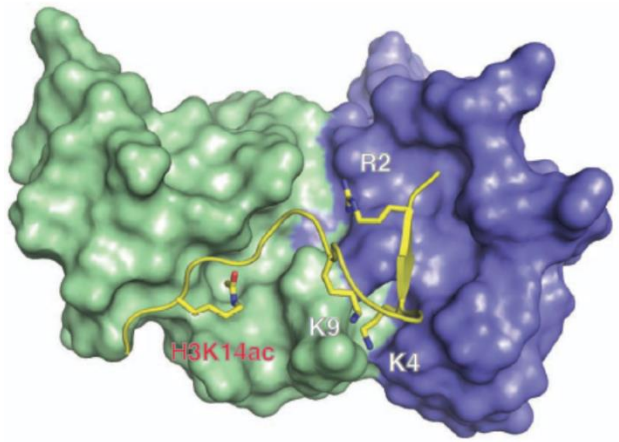


Figure 2. Space-filled PHD12 Structure, Highlighting the H3K14ac Peptide (yellow) Bound Across the Unified Structure of PHD1 (green) and PHD2 (blue).[15]

Source: Adapted with permission from Zeng, L.; Zhang, Q.; Li, S.; Plotnikov, A. N.; Walsh, M. J.; Zhou, M. Mechanism and Regulation of Acetylated Histone Binding by the Tandem PHD Finger of DPF3b. *Nature* **2010**, *466* (7303), 258–262. Copyright 2017 Nature Publishing Group.

The function of histone sidechain acetylation was characterized. For example, the H3K14ac modification can change the histone binding to PHD finger motif

(**Figure 2**). Compared to sidechain acetylation, the effect of protein N-ac has not been fully characterized. In N-ac, an acetyl group from acetyl-coenzyme A (Ac-CoA) is transferred to the α-amino group on the first residue of the protein substrate. N-acetylation occurs on about 80% of human proteins, either post- or co-translationally.[10] The functions of N-ac are related to protein stability, translocation, and protein binding.[11,12] Shemorry and coworkers reported that N-ac is one of the many factors affecting the ubiquitination pathway.[9] Depending on other conditions, N-ac can either stabilize the protein or act as a degradation signal.[13,14]

There is some preliminary evidence showing histone N-ac can change its protein binding partners. Marmorstein and co-workers[5] characterized the molecular basis of an N-terminal acetyltransferase NatD, which only acetylates histone H4 and H2A9. The role of NatD in gene regulation is not fully understood, but initial evidence suggests that histone N-ac may regulate the binding of PHD finger proteins. The PHD finger is a motif often related to chromatin-mediated gene regulation. Another study by Zeng and co-workers[15] showed that the PHD finger can selectively bind to histones H3 and H4, where the interaction of the PHD fingers with H3 is promoted by acetylation at lysine 14, but inhibited by sidechain methylation at lysine 4. The differential binding of the PHD fingers alters the gene transcription of mouse DPF3b downstream genes Pitx2 and Jmjd1c.

Project 2: Binding Study of $(PR)_x$ Peptide to Proteasome

Importance and Significance Amyotrophic lateral sclerosis (ALS, or Lou Gehrig's disease) is a progressive, fatal neurodegenerative disease characterized by the loss of motor neurons in the brain and spinal cord.[16] The prevalence of ALS is about 5 cases per 100,000 people and death occurs typically 3–5 years after diagnosis. *Riluzole*[17] and *Edaravone*[18] are the only ALS medicaitons approved by FDA. The mechanism of ALS is not fully understood, mutations of several genes can lead to the development of ALS.[19] The microsatellite expansion in *C9ORF72* gene was found to be the most important cause of ALS, which accounts for approximately 34.2% of ALS cases.[20,21] Undestanding of the mechnism of the *C9ORF72* gene can provide thoerical background for ALS treatments development.

Purpose and Hypothesis The *C9ORF72* gene is transcribed via a repeat-associated non-AUG (RAN) translation mechanism, where translation can occur in both sense and anti-sense transcripts. The translation products of *C9ORF72* are five different poly-dipeptides: glycine-alanine (GA), glycine-arginine (GR), proline-arginine (PR) proline-alanine (PA), and glycine-proline (GP). The length of the peptides varies among patients, so PR_x is used to represent peptides without defined length.

There are two theories about the mechanism of the disease. The RNA toxicity theory suggests the disease is the result of RNA toxicity from *C9ORF72* transcription.[22] In contrast, the peptide toxicity theory suggests that cell death is caused by the translation products of *C9ORF72*.[23] Shi and co-workers[24] reported that the toxicity of the PR_x peptide comes from plugging the nuclear pore, and prehibiting communication between the nucleus and cytoplasm. In their study, the

stimulated emission depletion (STED) microscopy image shows co-localization of the nucleoporin gp210 and the PR_{20} peptide indicates the PR_{20} peptide is binding to the nuclear pore, as shown in **Figure 3**.

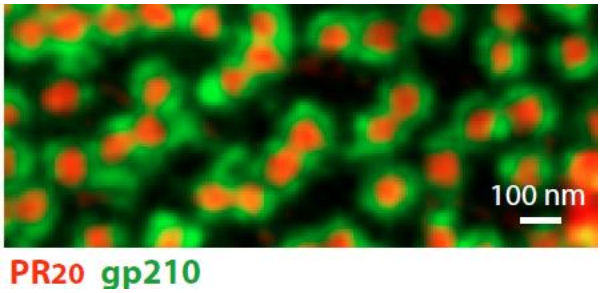


Figure 3. Merged Confocal Microscopic Image of Fluorescein Isothiocynate Labeled PR_{20} Peptide (red) and STED Image of Nucleoporin gp210 (green).[25]

Source: Adapted with permission form Taylor, J. P. A PR plug for the nuclear pore in amyotrophic lateral sclerosis. *Proc. Natl. Acad. Sci.* **2017**, *114* (7), 201621085. Copyright 2017 PNAS.

A separate study by Kalb and co-workers[26] reported that the PR_x peptide binds to the proteasome to inhibit degradation of ubiquitinated substrates, which leads to cytotoxicity. These results do not necessarily conflict with each other; it is possible that the PRx peptide can bind to several targets in cells. This is likely, given its very simple sequence.

The Kalb group hypothesized that the PR_x peptide can bind to certain subunit(s) of the proteasome, possibly by plugging the proteasome entrance site and resulting in cytotoxicity. This hypothesis can be confirmed by identifing the binding pattern of PR_{20} to the proteasome using photo-crosslinking probes.The Kalb group also expect that the PR_x peptide can colocalize with the proteasome in the cell for the binding event to occur. The second hypothesis can be confirmed by observing the colocation of a fluorocence PR_{20} peptide with protesome in cell under a confocal microscope. My taks in this project was to develop the chemical probes needed for the *in vitro* experiments. All *in vitro* experiments was performed by the Kalb group.

Experimental Design To confirm the hypothesis that the PR_x peptide can bind to certain subunit(s) of the proteasome, a photo-crosslinker probe can be used to identify the binding partners(s) of PR_{20} peptide. Similar to project 1, a photo-crosslinking probe with a photo-crosslinker and a biotin tag was incorporated into the PR_{20} peptide. Since it is difficult to perform the PR_{20} peptide sidechain modification, the photo-corsslinker and biotin tag was added to the terminal of the sequence. In the fluorocence PR_{20} peptide design, a fluorophore was attached to a PR_{20} peptide at the C-terminal.

Methods in Peptide Synthesis and Functionalization

One advantage of peptide synthesis and modification is that once the standard protocols are established, peptides with desired sequence, functional groups and

structure can be synthesized by the same approach. The methods described below apply for both Chapter 1 and Chapter 2.

Peptide Backbone Construction and Peptide Modification Solid phase peptide synthesis (SPPS) can generate short peptides fragments. The maximum length of peptide generated by SPPS is around 40 amino acids with useful yield. Longer peptides or proteins can be synthesized by recombinant protein expression or by ligation of shorter fragments. Among different ligation methods, Native Chemical Ligation (NCL) is the most popular method to form a native amide bond between the two fragments, by joining a C-thioester peptide fragment to an N-cysteine modified peptide fragment.

Installing chemical modifications depends on the nature of the modification. Some chemical modifications can be installed during SPPS by using unnatural amino precursors. In protein expression, unnatural amino acid mutagenesis can be used. Since unnatural amino acid mutagenesis is typically restricted to the inclusion of a single non-natural amino acid, we will restrict our discussion to SPPS and ligation. If the modification is unstable in SPPS condition or too large and sterically hindering to the next coupling reaction in SPPS, other methods such as click reactions and cysteine *S*-alkylation can be used to install those modifications.[27,28]

Backbone Construction by Solid Phase Peptide Synthesis (SPPS) First introduced by Merrifield and co-workers[29] in 1963, solid phase peptide synthesis (SPPS) has become the most popular synthetic approach for making short peptides. The yield of stepwise SPPS decrease as the length of peptide chain increases. In general, the maximum length of the peptide with useful yield is around 40 amino acids. The use of solid support has huge advantages over solution phase peptide synthesis, for the latter often requires purification after each standard synthetic cycle and overnight reactions. SPPS eliminates the purification steps in each synthetic cycle. SPPS can be accelerated to 4 minutes per synthetic cycle,[30] and performed by automation at small scale or at industrial scale.[31]

Based on the choice of N-terminal protecting group, SPPS can be divided into two strategies: tert-butyloxycarbonyl (Boc) strategy and 9H-fluoren-9-yl-methoxycarbonyl (Fmoc) strategy. The Boc strategy is the original method developed by Merrifield, which uses the Boc protecting group as the N-terminal protecting group. The Boc group is removed in mild acid condition. The negative aspect of the Boc strategy is that it requires the use of hazardous hydrofluoric acid (HF) to remove sidechain protecting groups and cleavage from the solid support.[29] On the other hand, the Fmoc strategy uses a base liable N-terminal protecting group. The removal of sidechain protecting groups and cleavage from the solid support is achieved by treatment with trifluoroacetic acid (TFA), which is a much milder condition compare to HF. The Fmoc strategy is more popular than the Boc strategy for safety reasons, and the Petersson laboratory uses the Fmoc strategy exclusively. **Figure 4** shows the typical synthetic cycle of Fmoc based SPPS synthesis, where PG stands for a generic side-chain protecting group and AA stands for a generic amino acid. The C-terminal amino acid is first attached to the resin via a cleavable linker. In the next step, the Fmoc N-terminal protecting group is removed by base without removing the acid liable side-chain protecting groups.

The peptide chain is extended by coupling the next amino acid after removal of reaction side products and excess reagents by rinsing and filtration. The same deprotecting and coupling process is cycled until the completion of the entire sequence of the peptide. At the end of the cycle, the cleavage from the solid support and the side-chain protecting groups are typically removed simultaneously. The synthesized crude peptide is often purified by precipitation and High-performance liquid chromatography (HPLC), then lyophilized to a powder. For peptides requiring ligation steps, NCL is often used to link fragments of peptides.

Backbone Extension by Native Chemical Ligation Developed by Kent and co-workers[32] in 1994, NCL can be used to ligate two peptide fragments with a native peptide bond without epimerization.[33] NCL is mostly used to ligate peptides that are too long for SPPS or even an expressed protein to a synthetic peptide fragment. The main advantage of NCL is its selectivity towards a thioester and N-terminal cysteine, allowing both peptide fragments to be unprotected. Using unprotected peptides eliminates steps for protection/deprotection and prevents solubility issues associated with bulky hydrophobic sidechain protecting groups. In addition, NCL is a convenient one-pot reaction with a high conversion rate. NCL has expanded the application of SPPS and enabled the synthesis of long peptides or even full-length proteins.

Figure 5 shows the steps of the NCL method commonly employed in the Petersson laboratory.[34,35] Fragment A is a C-terminal modified peptide hydrazide generated from modified 2-chlorotrityl chloride resin, fragment B is a peptide with an N-terminal cysteine. First, the acyl hydrazide group from the fragment is converted to an azide group, then converted to a peptide thioester by 4-mercaptophenylacetic acid (MPAA) (**Figure 5B**). Next, the thiol group of the N-terminal cysteine of fragment B attacks the peptide thioester and forms a native peptide bond via an N- to S- acyl shift (**Figure 5C**). To prepare the peptide hydrazide, the 2-chlorotrityl chloride resin is treated with hydrazine hydrate before loading the first amino acid, as shown in **Figure 5A**.[35] The SPPS product of modified resin gives the C-terminal peptide hydrazide. If the thiol group from the cysteine is not desired in the final product, the thiol group can be removed by desulfurization.[36] Desulfurization greatly extends the possible ligation sites of NCL. Besides NCL, there are other traceless peptide ligation methods, such as traceless Staudinger ligation[37] and methionine ligation.[32,38] All ligation methods involve a chemo-selective capture step followed by an invariable intramolecular acyl transfer reaction. These steps are followed by a final variable step to convert the auxiliary to a natural amino acid.

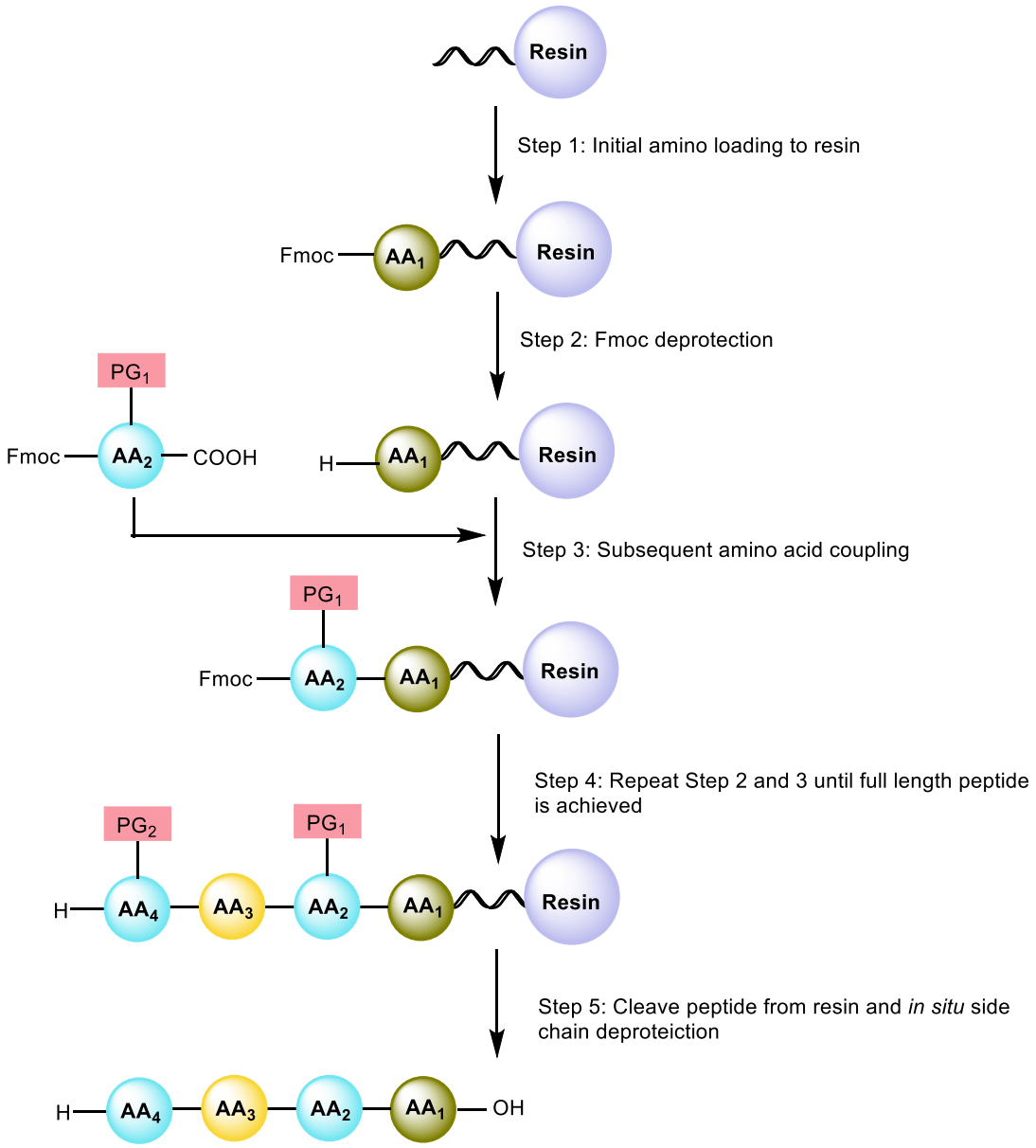


Figure 4. Scheme of Fmoc Based SPPS.[39]

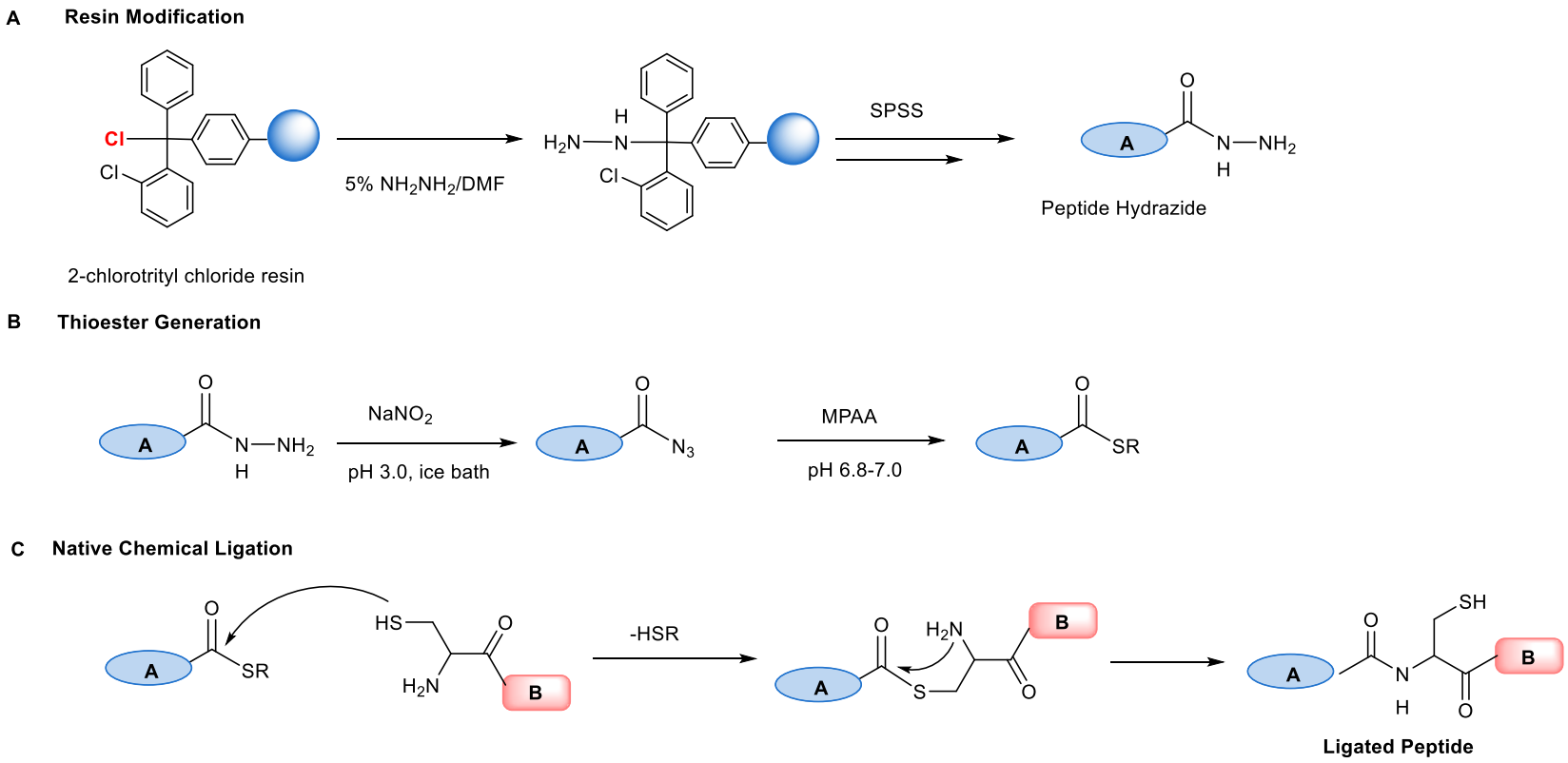


Figure 5. Native Chemical Ligation Scheme.[35] (A) Generation of hydrazine modified 2-chlorotrityl chloride resin. (B) Generation of peptide thioester from peptide hydrazide, and (C) Peptide bond formation.

Common Peptide and Protein Modification Methods. After the construction of the peptide backbone, residues can be modified at various reactive sites. Cysteine is a common handle for modifying peptides. There are several approaches to conjugate building blocks or modify at the cysteine residue. *S*-alkylation and maleimide crosslinking are the two methods that were used in this study. Other possible methods include solid phase cross-coupling[40] and enzyme-catalyzed macrocyclization[41].

S-alkylation conjugates the peptide with the fragment of interest at the cysteine thiol sidechain, establishing a thioether bond at basic conditions.[27] **Figure 6** is an example of using *S*-alkylation to couple an alkyne to the cysteine residue sidechain. The linker is small but the requirement for basic reaction conditions is not suitable for base labile molecules.

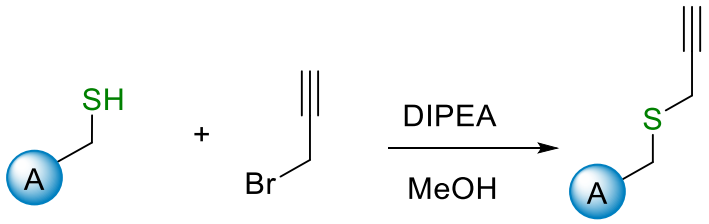


Figure 6. *S*-alkylation of Cysteine Containing Peptide.[27] A represents a protein or peptide with a cysteine.

Maleimide crosslinking is a common method to label a protein with a biotin or a fluorophore.[42] This method is also used to generate antibody−drug conjugates for cancer treatment.[43] The main advantage of maleimide crosslinking reaction is the reaction can occur in neutral aqueous conditions. One negative aspect of maleimide crosslinking is it leaves a relatively large linker compared to *S*-alkylation, as shown in **Figure 7**.

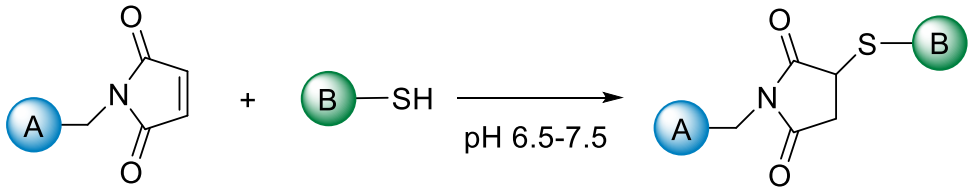


Figure 7. Maleimide Crosslinking Reaction.[43] A represents a labeling reagent with the maleimide reactive group; B represents a protein or other molecule that contains the target functional group.

Azide-Alkyne Click Chemistry Even though cysteine is a useful handle for bioconjugation, selectively labeling peptides or proteins with multiple cysteine residues is non-trivial. In this case, click chemistry is another common peptide and protein modification method. The term "click chemistry" does not refer to one specific reaction, rather, any ligation reaction that is robust, easy to perform, with high yield, minimum byproducts formation, and broad substrate spectrum can be referred to as a “click chemistry reaction”. The classical click chemistry reaction is the copper catalyzed azide-alkyne cycloaddition shown in **Figure 8**, in which a Cu(I) catalyst is generated from Cu(II) *in situ*.[44] The main advantage to this reaction is

that there are no azides or alkynes in most biological systems, so this reaction can be selective in biomolecules like proteins, and can even occur *in vivo* without perturbing the biological system (although Cu(I) can be toxic).[45] More importantly, the copper catalyzed click reaction is orthogonal to S-alkylation and maleimide crosslinking, which enable us to selectively modify peptides and proteins at multiple sites.

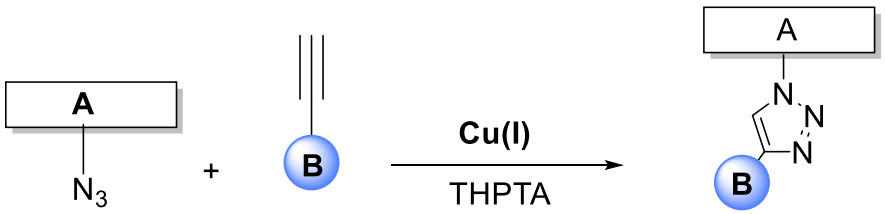


Figure 8. Scheme of Click Reaction.[46] THPTA refers to 3-hydroxypropyltriazolylmethyl)amine.

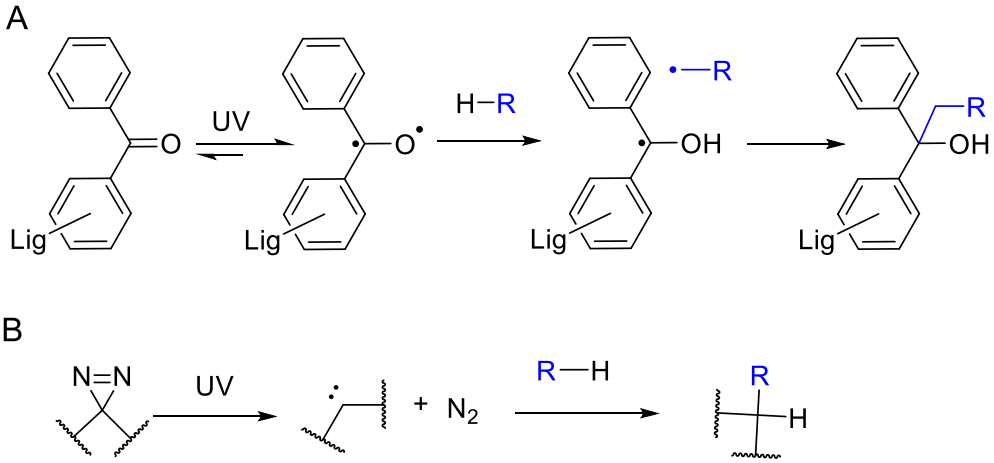


Figure 9. Mechanisms of Photo-crosslinking.[47] (A): benzophenone based crosslinking. (B): diazirine based crosslinking. Lig refers generic ligands.

Photo-Crosslinking Probes for Protein-protein Interaction Studies Unnatural amino acids can be introduced during SPPS or at modification steps. One class of unnatural amino acids is photo-crosslinkers. A photo-crosslinker can generate a covalent linkage between the ligand and the binding target based on spatial proximity under UV irradiation. The captured target(s) can be isolated and identified by mass spectrometry (MS),[48] which can be useful in drug design and protein function studies. Common photo-crosslinkers can be characterized into three families: benzophenone, aryl azide, and diazirine. These photo-crosslinking groups have differing selectivity and reactivities.[47] The mechanism of benzophenone and diazirine crosslinking is shown in **Figure 9**. In both cases, UV irradiation generates a reactive radical, but the homolytic bond cleavage of benzophenone is reversible, allowing it to return to the ground state if no reactive partner is present. On the other hand, the expulsion of N_2 in diazirine photolysis

makes it irreversible, so the resulting carbene is much more reactive. Thus, benzophenones are more selective crosslinkers.

Using the methods of peptide synthesis and modification mentioned above, I synthesized a series of probes to study the functions of histone H4 N-terminal acetylation (N-ac) in project 1. The result of project 1 can help explaining the role of histone co- and post-translation in gene regulation. In project 2, I synthesized a different set of probes to study the toxicity mechanism of proline-arginine tandem repeat peptides PR_x. The result of project 2 can provide mechanism foundation for developing amyotrophic lateral sclerosis (ALS) treatments.

Materials and Methods

SPPS Materials Fmoc-Lys(biotinyl)-OH was purchased from TCI America (Portland, OR, USA). Fmoc-PhotoLys(Boc)-OH was kindly provided by Dr. Xiang David Li from the University of Hong Kong (Hong Kong, China). 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 2-Chlorotrityl chloride resin (100-200 mesh, 1% DVB) were purchased from Novabiochem (currently EMD Millipore; Billerica, MA, USA). Standard amino acids for Fmoc peptide synthesis were also purchased from Novabiochem (currently EMD Millipore; Billerica, MA, USA), including Fmoc-Ala-OH, Fmoc-Arg(Pbf)-OH, Fmoc-Gly-OH, Fmoc-His(Trt)-OH, Fmoc-Leu-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH and Fmoc-Pro-OH. Piperidine was purchased from American Bioanalytical (Natick, MA, USA). Sigmacote®, *N,N*-diisopropylethylamine (DIPEA), and Trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Glass peptide synthesis reaction vessels (RVs) were treated with Sigmacote® prior to use. Triisopropylsilane (TIPS) was purchased from Santa Cruz Biotechnology (Dallas, TX, USA). *N*-Methylmorpholine (NMM) and 4-mercaptophenylacetic acid were purchased from Acros (Currently Fisher Scientific, Hampton, NH, USA). Acetic anhydride, hydrazine hydrate (55% in water), Bond-Breaker™ TCEP Solution, Ellman's reagent, Acetonitrile (ACN), Dimethylformamide (DMF), Methanol (MeOH) and Tetramethylrhodamine-5-Maleimide (TAMRA) were purchased from Fisher Scientific (Hampton, NH, USA). Guanidine hydrochloride (Gn·HCl) was purchased from Invitrogen (Carlsbad, CA, USA) $Cys(ProArg)_{20}$ peptide was purchased from GenScript (Piscataway, NJ, USA).

Instruments Histone H4 peptide residues 1 to 18 was synthesized using a Liberty synthesizer (CEM Corporation, Matthews, NC). Crude peptides were purified with a Varian ProStar High-Performance Liquid Chromatography (HPLC) instrument outfitted with a diode array detector (currently Agilent Technologies) using aqueous A (H_2O + 0.1% TFA) and organic B (CH_3CN + 0.1% TFA) phases. Purification testing was performed in an Agilent 1100 Series HPLC with a 4.6nm Phenomenex Luna C8 (2) column.

UV absorbance spectra were obtained with a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA). Matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) data

were collected with a Bruker Ultraflex III MALDI-TOF-TOF mass spectrometer (Billerica, MA, USA).

Reagent Setup TFA cleavage cocktail A: TFA/DCM/TIPS/Water=90/5/2.5/2.5 (vol/vol/vol/vol) TFA cleavage cocktail B: TFA/DCM/TIPS/Water=85/10/2.5/2.5 (vol/vol/vol/vol) TFA cleavage cocktail C: TFA/DCM/TIPS/Water=30/65/2.5/2.5 (vol/vol/vol/vol) Acylating cocktail: DMF/NMM/acetic anhydride=84/6/10 (vol/vol/vol/vol) 5% Hydrazine in DMF: 0.4 mL of hydrazine hydrate was mixed with 7.6 ml of DMF. TAMRA-maleimide in DMSO: TAMRA-maleimide was dissolved in DMSO to a final concentration of 4.5 mM. Two different phosphate solutions (0.2M) containing 6M Gn•HCl (pH 3.0-3.1 and pH 6.9-7.0, respectively): For a 10-ml solution, 312 mg of NaH_2PO_4•H_2O and 5.74 g of Gn•HCl was mixed into a 15-ml centrifuge tube and pH was adjusted to 3.0-3.1 and 6.9-7.0 with 6M NaOH and 1M HCl. The solution was filtered by 0.22 µm microporous membrane filter. $NaNO_2$, 0.5M: 17mg of $NaNO_2$ was dissolved in 0.5 ml of water before use. 50% ACN in water with 0.1%TFA: 5 ml of water, 5ml of CAN and 10 µl of TFA were mixed. -15 °C ice bath was prepared by mixing 12.5 g of NaCl with 25 g of ice. Milli-Q filtered (18 MΩ) water was used for all solution preparations (EMD Millipore, Billerica, MA, USA).

Standard Synthesis Cycle Standard synthesis cycle was used for all deprecating and coupling reactions except the first amino acid and unnatural amino acids. In each synthesis cycle, the resin was washed three times with (dimethylformamide) DMF under gentle agitation. Followed by three washes of dichloromethane (DCM) and another three washes of DMF. To remove the Fmoc deprotecting group, 2 ml of 20% piperidine in DMF (vol/vol) was added to the reaction vessel (RV) and agitated for 20 minutes. Then the resin was drained and washed with the same DMF-DCM-DMF sequence as mentioned above. For each 25 µmol scale synthesis, 125 µmol (5 equiv) of the appropriate amino acid and 125µmol of HBTU (5 equiv) was dissolved in 2 ml of DMF, with 44 µl of DIPEA (10 equiv) added. This step pre-activates the amino acid and usually gives the solution a light-yellow color. The fully dissolved amino acid solution was then added to the RV. The coupling reaction time is 30 minutes. This cycle was repeated until the full length of peptide is synthesized. After the last amino acid coupling, step 1-6 was performed to remove the Fmoc group.

Synthesis of PR_3 and PR_5 Peptides 7, 8 PR_3 and PR_5 peptides were synthesized in 25 µmol scales each, using 2-Chlorotrityl chloride resin under standard SPPS condition. To couple on the first amino acid, 42mg of 2-Chlorotrityl chloride resin (estimated resin loading capacity of 0.6mmol/g) was added to a dry reaction vessel (RV) with prior Sigmacote® treatment. Then the resin was washed with DMF and then swelled in DMF for two successive of 15 minutes with gentle agitation. Then 125 µmol (5 equiv) of Arg(Pbf) was dissolved in 2 ml of DMF, then the amino acid solution and 44µL of DIPEA (10 equiv) were added to RV. The 2-Chlorotrityl chloride resin does not require HBTU in the first amino acid coupling step. The coupling time was 30 minutes. The rest of the sequence was synthesized using the Standard Synthesis Cycle. After the final Fmoc deprotection, the resin was drained and dried by vacuum filtration. To remove the Pbf protecting group and cleaved from the resin, 2ml of TFA cleavage cocktail A was added to the RV

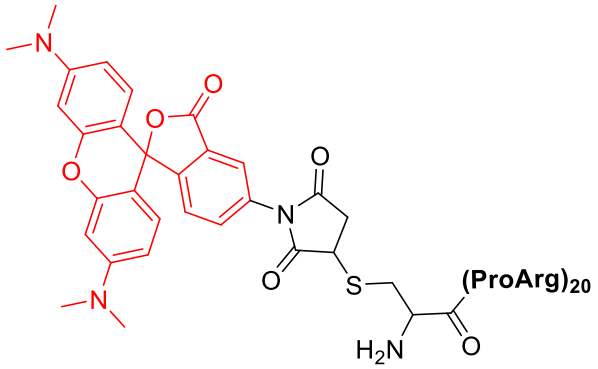
and agitated for 90 minutes. At the end of the cleavage, the cleavage solution was collected from the bottom of the RV, and then two small amounts of DCM were added to rinse the RV. The combined cleavage solution was concentrated by a rotavapor. The concentrated peptide was diluted to 5 ml with 50% ACN in water with 0.1%TFA. The crude peptide was then filtrated with a 0.20µM syringe filter to remove solid particles before HPLC purification.

PR_3 and PR_5 Peptide Purification and Characterization The crude peptides were purified by reverse-phase HPLC using a Vydac 218TP C18 prep column (Grace/Vydac; Deerfield, IL, USA). The flow rate was 15 ml/min, the gradient was 2%B in 0-5 min, 2-10%B in 5-10 min, 10-40%B in 10-40 min, 40-100%B in 40-45 min, 100%B in 45-50 min, and 100-2%B in 50-55 min, where A is water with 0.1% TFA and B is ACN with 0.1% TFA. Fractions were tested by MALDI-MS, fractions with pure peptide were combined and lyophilized. The lyophilized products were dissolved with 5.2 ml of 50% ACN in water (0.1%TFA), then 5 ml of the solution was transferred to five microcentrifuge tubes with 1 ml solution in each tube. The microcentrifuge tubes were dried down with a table top vacuum centrifuge (Savant/Thermo Scientific; Rockford, IL, USA). The rest of the solution was used for MALDI characterization and analytical HPLC analysis. For HPLC characterization, the sample was diluted to 800 µL final volume with water with 0.1% TFA. MALDI characterization was performed using standard instrument set-up under positive mode with reflector enabled, using α-Cyano-4-hydroxycinnamic acid (CHCA) matrix. HPLC analysis was performed using a 4.6mm VYDAC HPLC C18 column from Grace Davison Discovery Science at 1 ml/min flow rate, measuring absorbance at 215 nm, using the same gradient in protein preparation.

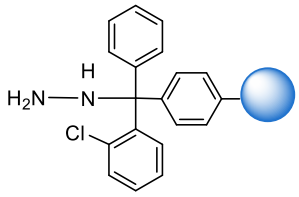
Systhesis of Histone H4 Peptides 1, 2 Synthesis of Histone H4 peptide uses the synthesis cycle as above, with minor differences in certain amino acids coupling steps mentioned below. For Lys(biotinyl), 3 equiv of the amino acids for coupling reaction was used. For Photo-Lysine, we used 2 equiv of the amino acids in coupling reactions. 5 equiv of the Fmoc amino acids for all other amino acids. Residues 2-18 were synthesized by the Liberty peptide synthesizer following manufactory's procedures. For the first amino coupling, 84mg of 2-Chlorotrityl chloride resin was added to a dry reaction vessel (RV) with prior Sigmacote® treatment. 5 ml of 50% DMF in DMSO was added with magnetic stirring for two successive of 15 minutes to swell the resin, then drained by vacuum filtration. Next, 120 mg of Fmoc-Lys(Biotin)-OH was added to 6 ml of 50% DMSO in DMF. To help solubilization of the amino acid, the solution was warmed in 37 °C water bath for about 15 minutes. 3 ml of the amino acid solution and 175 µl of DIPEA were added to the RV. After 45 minutes of coupling, the resin was drained by vacuum filtration. The rest of Fmoc-Lys(Biotin)-OH solution and 175µl of DIPEA were added to the RV. After 45 minutes of coupling, the resin was drained and washed by the DMF-DCM-DMF sequence as mentioned before. After loading the first amino acid to the resin, the resin was swelled in 6 ml of DMF for 30 minutes then drained. To deactivate the unreacted site on the resin, a capping procedure was performed by adding 2 ml of 10% MeOH/DMF and 175 µl of DIPEA to the RV. The capping reaction was 20 minutes and then the resin was washed with the DMF-DCM-DMF sequence. Next, the resin was transferred from the RV to the Liberty peptide

synthesizer to synthesize residues 2-18 using manufacturer's standard method. After the 18th residue, the resin was removed from the peptide synthesizer and manual synthesis was performed using the standard synthesis cycle, since the photo-lysine is a light sensitive amino acid, the handling of photo-lysine and reaction steps after photo-lysine coupling reaction were performed under red LED with the RV covered with aluminum foil. After the final Fmoc deprotection reaction, half of the resin was acetylated at N-terminal, giving the N-acetylated Histone H4 peptide. The other half of the resin undergoes cleavage directly. To acetylate the peptide at the N-terminal. Two portions of 5 ml of the acylating cocktail were added to the RV and agitated for 20 minutes with washing in between. To cleave the peptide from the resin and remove the protecting groups, 2 ml of TFA cleavage cocktail B was added to the RV and agitated for 90 minutes. At the end of the cleavage, the cleavage solution was collected from the bottom of the RV, and then two small amounts of DCM were added to rinse the RV. The cleavage solutions were then concentrated using a rotavapor. The concentrated peptide solution was transferred to a 15-ml centrifugal tube, and then about 10 ml of cold ether was added to precipitate down the crude peptide. The volume of cold ether should be at least 10 times larger than the concentrated peptide solution. The peptide was incubated at 4 °C for 20 minutes and then centrifuged for 5 minutes. The supernatant was discarded and the pellet was dissolved in minimal amount of 50% ACN in water with 0.1%TFA. The crude peptide was then filtrated with a 0.20µM syringe filter to remove solid particles before HPLC purification.

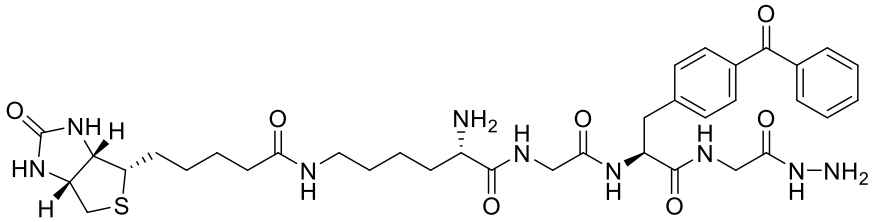
Peptide Purification and Characterization The crude peptides were purified by reverse-phase HPLC using a Vydac 218TP C18 prep column (Grace/Vydac; Deerfield, IL, USA). The flow rate was 15 ml/min, the gradient was 2%B in 0-5 min, 2-10%B in 5-10 min, 10-40%B in 10-40 min, 40-100%B in 40-45 min, 100%B in 45-50 min, and 100-2%B in 50-55 min, where A is water with 0.1% TFA and B is ACN with 0.1% TFA. Fractions with pure peptide were combined and lyophilized. The lyophilized products were dissolved with 5.2 ml of 50% ACN in water (0.1%TFA), then 5 ml of the solution was transferred to five Eppendorf tubes with 1 ml solution in each tube. The Eppendorf tubes were dried down by a table top vacuum centrifuge (Savant/Thermo Scientific; Rockford, IL, USA). The rest of solution was used for MALDI characterization and analytical HPLC analysis. For HPLC characterization, the sample was diluted to 800µL final volume with water (0.1% TFA). MALDI and HPLC characterization was performed using the same condition as for PR_3 and PR_5 peptides.



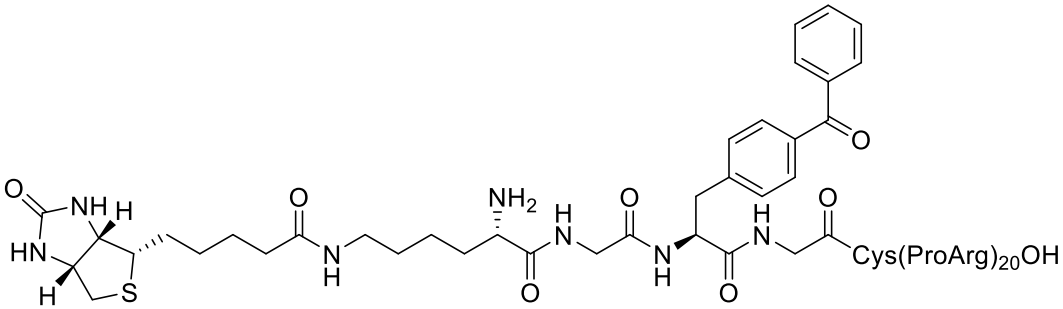
Synthesis of TAMRA-PR$_{20}$ Probe 3 Crude Cys(PR)$_{20}$ peptide solid was purchased from GeneScript (Piscataway, NJ, USA) and purified by reverse-phase HPLC using a Vydac 218TP C18 prep column (Grace/Vydac; Deerfield, IL, USA). The flow rate was 12 ml/min, the gradient was 2%B in 0-5 min, 2-10%B in 5-10 min, 10-40%B in 10-40 min, 40-100%B in 40-45 min, 100%B in 45-50 min, and 100-2%B in 50-55 min. The purified peptide was lyophilized. Quantification was performed by using Ellman's reagent following Sigma-Aldrich manual #22582, with the exception that pretreatment with immobilized TCEP was not performed. 0.2 mmol of Cys(PR)$_{20}$ peptide was dissolved in 900 µl of 20 mM Tris buffer, pH 8. 0.5 µl of TCEP bond breaker and 36 µl of TAMRA-maleimide in DMSO was added. The reaction was performed at room temperature with magnetic stirring, using MALDI-MS to monitor the reaction process. The labeled peptide **3** was purified using the same condition for Cys(PR)$_{20}$ peptide, followed by characterized with MALDI-MS with expected m/z of 5652.



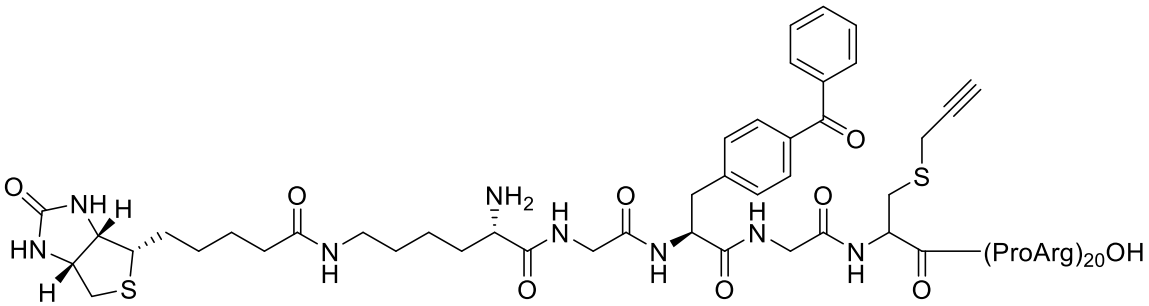
Preparation of Hydrazine Modified 2-Chlorotrityl Chloride Resin The warhead **4** was synthesized by SPPS using hydrazine modified 2-Chlorotrityl chloride resin. First, 100 mg of 2-Chlorotrityl chloride resin was added to a dry reaction vessel (RV) with prior Sigmacote® treatment. The resin was swelled by DMF for two successive of 15 minutes. Then, the resin is treated with two successive of 4 ml of 5% (vol/vol) hydrazine in DMF for 20 minutes under gentle agitation. The resin was washed with DMF and then DCM carefully.



Synthesis of Warhead 4 To couple the first amino acid, 125 µmol (5 equiv) of the appropriate amino acid and 125 µmol of HBTU (5 equiv) was dissolved in 2 ml of DMF, then 44 µl of DIPEA (10 equiv) was added to the amino acid solution. The amino acid solution was added to the RV and agitated for 30 minutes. The rest of the sequence was synthesized by standard synthesis cycle. The cleavage and purification condition was identical to the PR_3 peptide. The expected m/z is 752, monitored by MALDI-MS.



Synthesis of Molecule 5 0.6 µmol of peptide hydrazide **4** was dissolved with 0.4 ml of 0.2M phosphate buffer (pH 3) in a 2 ml microcentrifuge tube. pH was adjusted to 3.0-3.1 with 6M NaOH and 1M HCl. The microcentrifuge tube was incubated on -15 °C ice bath for 15 minutes. Around 0.5 µmol of $Cys(PR)_{20}$ peptide and 13.6 mg of MPAA (4-Mercaptophenylacetic acid were dissolved in 0.4 ml of pH 7 phosphate buffer and adjusted pH to 6.5. To oxidize the peptide hydrazide to azide, 40 µl of 0.5M $NaNO_2$ solution was added to the solution of compound **4**. The pH of the mixture was adjusted to 3-3.1 and stirred for 15 minutes in -15 °C ice bath. To convert the peptide azide to thioester, $Cys(PR)_{20}$ and MPAA mixture solution was added to peptide azide **4**. The microcentrifuge tube was removed from ice bath and warmed to room temperature. pH was adjusted to 6.8-7.0 with 6M NaOH. After overnight stirring at room temperature, 80 µl of TCEP bond breaker was added. The mixture was diluted to 5 ml final concentration before HPLC purification. For MALDI-MS characterization, 1 µl of the reaction mixture was diluted 30 times with water. The product **5** was purified using the same HPLC gradient to compound **3** and then lyophilized. The expected m/z is 5908.



S-Alkylation 4 Lyophilized peptide **5** was dissolved in a round bottom flask with 1 ml of MeOH. With stirring, 10 µl propargyl bromide, 5 µl of DIPEA and 1 mg of TCEP was added to the flask. The reaction progress was monitored by MALDI-MS until finished. When the reaction is finished, MeOH was removed by a rotary evaporator and product **4** was dissolved in 50% ACN in water and then purified by HPLC using the same gradient to compound **3**. The expected m/z is 5976.

Peptide Characterizations All peptides were characterized with MALDI-MS as shown in **Table 1**. For all peptides, expected m/z was observed, which suggests I have successfully synthesized the peptides of interest.

Table 1. Characterizations of Synthetic Peptides

Molecule	*Calculated (+m/z)*	*Measured (+m/z)*
3	5652	5652
4	752	752
5	5908	5908
6	5976	5976
7	778	778
8	1284	1284

Results and Discussion

Project 1: Interactions of N-Terminally Acetylated Histones

For comparison between the non-acetylated histone H4 peptide and N-ac histone H4 peptide, two versions of the probe was designed. The non-acetylated version is referred to as compound **1** and N-ac version as compound **2**. The sequence is R-Ser-Gly-Arg-Gly-PhotoLys-Gly-Gly-Lys-Gly-Leu-Gly-Lys-Gly-Gly-Ala-Lys-Arg-His-Arg-Lys-Gly-Gly-Lys(Biotin)-COOH. R refers to a hydrogen atom for non-acetylated histone peptide H4 probe **1** and an acetyl group for N-ac histone peptide H4 probe **2.** The structures are shown in **Figure 10**.

Peptide **1** and **2** were synthesized by SPPS and characterized by MALDI-MS. We observed m/z of 2457 and 2499 for compound **1** and **2** on MALDI-MS, respectively. This two masses correspond to $[M+H]^+$ minus 28 Da, the minus 28 Da is the loss of two nitrogen atoms from diazirine functional group, which is expected for the unstable crosslinking group.[49] The MS data confirmed that I have successfully synthesized the peptides of interest, and the peptides were used in *in vitro* crosslinking experiments.

Cell Lysate Western Blot The cell lysate crosslinking test was performed by Gleb Bazilevsky in the Marmorstein group and the result is shown in **Figure 11**. In brief, the probes were incubated with cell lysates under UV irradiation, then a streptavidin Western blot was performed to image the captured protein(s). The photo-crosslinking experiments were performed with *E. coli* cell lysate and Hela cells lysate, respectively. In the *E. coli* cell lysate pull-down experiment, no difference was observed between the treatment with peptide **1** and **2**, indicating no protein shows binding preferences between the two versions of histone H4 peptide in *E. coli.* The result that proteins from prokaryotic *E. coli* do not bind to the histone H4 peptide meets the expectation since histone proteins only exist in eukaryotic organisms. Thus, this serves as an effective negative control for non-specific binding of the peptides. For the HeLa cell lysate pull-down experiment, a few bands were found only in lane 12 (the N-ac H4 peptide) but not in lane 10 (the non-acetylated H4 peptide). This suggests that proteins bind exclusively to the N-ac H4 peptide. The identities of the protein bands have not yet been determined by MS. It is concerning that many of the bands in lane 12 are also present in lane 11, non-irradiated N-ac H4 peptide. Further experiments are necessary to fully characterize these interactions. I hypothesize that the identified protein might be an enzyme that can remove the N-terminal acetyl group or a transcription factor.

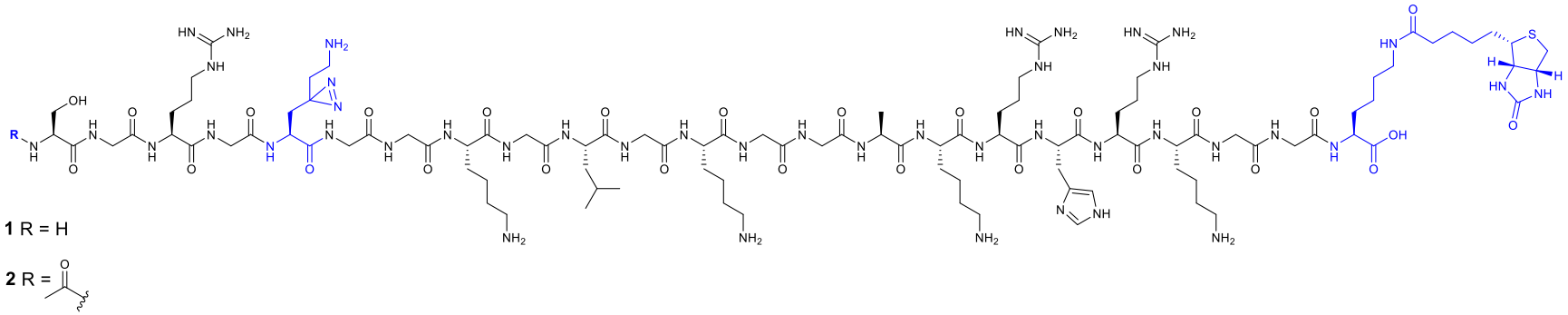


Figure 10. Design of Histone H4 Peptide Probe

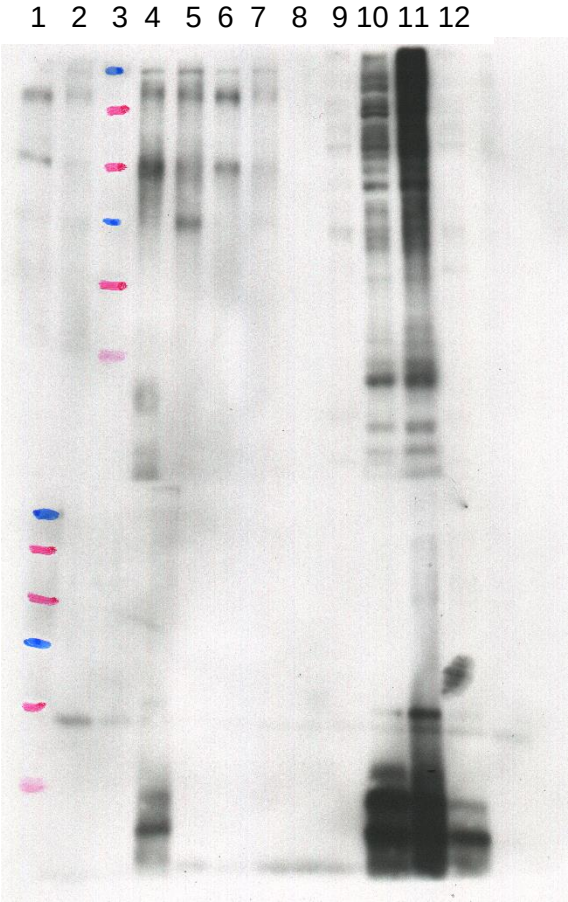


Figure 11. Cell lysate Western Blot experiment. Lane 1 to 6 is *E. coli* lysate with either 0 or 30 min UV irradiation. Lane 7 to 12 are HeLa cell lysate with either 0 or 30 min UV irradiation. Lane 1: cell lysate only, no probe, 0 min; Lane 2 cell lysate only, no probe, 30 min; Lane 3: none-acetylated H4 peptide, 0 min; Lane 4: non-acetylated H4 peptide, 30 min; Lane 5: N-ac H4 peptide, 0 min; Lane 6: N-ac H4 peptide, 30 min; Lane 7: no peptide, 0 min; Lane 8: no peptide, 30 min; Lane 9: non-acetylated H4 peptide, 0 min; Lane 10: non-acetylated H4 peptide, 30 min; Lane 11: N-ac H4 peptide, 0 min; Lane 12: N-ac H4 peptide, 30 min. (Bazilevsky, G. University of Pennsylvania, PA, USA. Unpublished work, 2016)

Project 2: Binding Study of $(PR)_x$ Peptide to Proteasome

PR_{20} Probe for Cell Imaging Several peptides were synthesized to investigate the toxic effects of $(PR)_X$ peptides, including their localization in cells and their sites of interaction with the proteasome. To perform a fluorescence microscopy imaging study, probe **3** was designed and synthesized to monitor the location of the peptide in cells. This probe can also be used to determine the affinity of the probe for purified proteasomes using flurescemce polarization. This probe consists of the PR_{20} peptide with a fluorophore attached at one end. In probe **3** design, a tetramethylrhodamine-5-Maleimide (TAMRA) fluorophore (highlighted in red) is linked to the N-terminal cysteine of the PR_{20} peptide via maleimide crosslinking (**Figure 12**). I chose TAMRA as the fluorophore because of its red fluorescence can be distinguished from the GFP labeled proteasomes expressed in cells. The proposed experiment is to add TAMRA-labeled probe **3** to the cell growth media, and the cellular uptake and localization of probe **3** can be monitored by a confocal microscope. Colocalization of probe **3** with the proteasome in cells can support the proteasome toxicity theory.

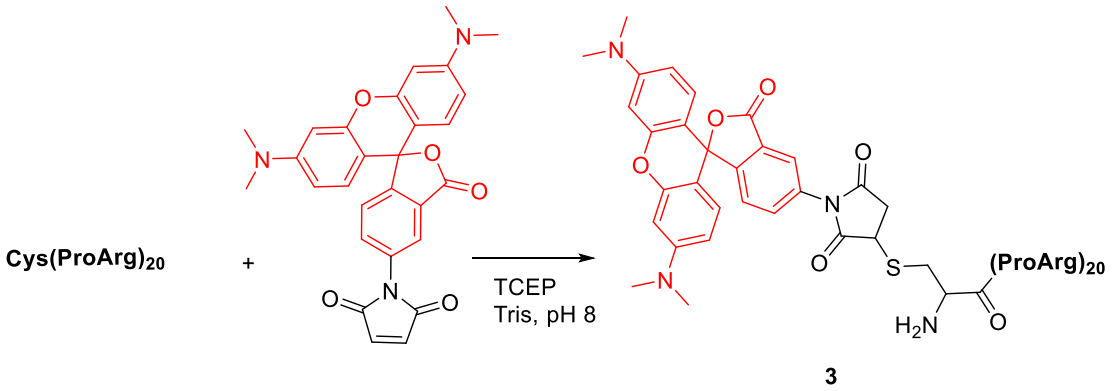


Figure 12. Synthesis Route of TAMRA-PR_{20} Probe **3**

Development of Photo-Crosslinking PR_{20}-X Probe To verify the hypothesis that the PR_x peptide can bind to certain subunit(s) of the proteasome, a second probe was designed and synthesized, which consists of a PR_{20} peptide, a photo-crosslinker, and a biotin affinity purification tag. The identification of the binding site(s) can provide critical information on the molecular mechanism of PR_{20} toxcity. It could also be used in design of PR_{20} peptide binding antagonists, a potential treatment for ALS.

Previously, Xing Chen from the Petersson group synthesized a photo-crosslinking probe by using NCL to ligate a PR_{20} peptide to a “warhead”. The warhead is a short peptide containing a biotin (highlighted in red) and a benzophenone group (highlighted in blue), with a glycine residue in between each

functional motif (**Figure 13**). This probe is referred to as the PR_{20}-X probe. I synthesized a batch of the PR_{20}-X probe as a replicate to Xing Chen's work.

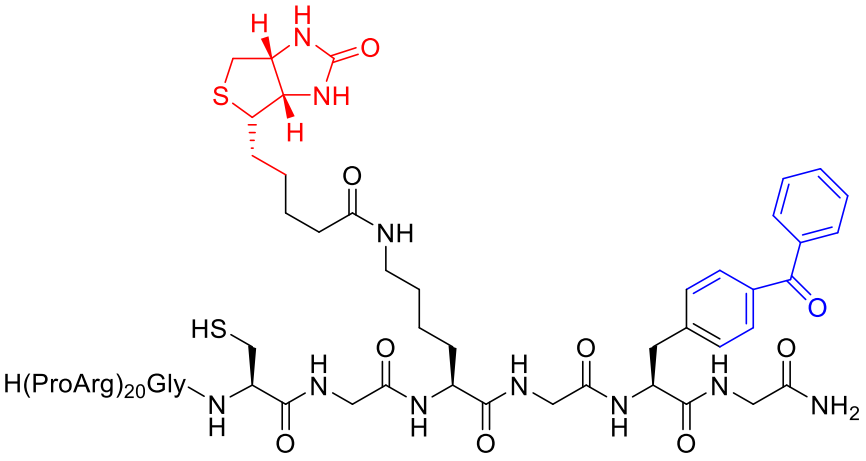


Figure 13. PR_{20}-X Probe Design. (Petersson laboratory, University of Pennsylvania, PA, USA. Unpublished work, 2016).

Development of Photo-Crosslinking Probe with a Bio-orthogonal Handle
In the PR_{20}-X probe design, the warhead is linked to the C-terminal end of the PR_{20} peptide. One negative aspect of this probe design is there is a cysteine residue in the sequence, and the cysteine thiol group can potentially form disulfides or bind to the proteasome non-specifically. To overcome this problem, probe **4** was designed to verify the result obtained from using the PR_{20}-X probe (**Figure 14**). Probe **4** has similar functional groups to PR_{20}-X, but the warhead is linked to the N-terminal end instead of the C-terminal end of the PR_{20} peptide. If the proteasome binding patterns of PR_{20}-X and **4** are consistent, we can conclude that the binding of the PR_{20} peptide to the proteasome subunits is not an artifact due to the placement of warhead. Otherwise, the probe design must be modified so that the probe does not perturb the function of the PR_{20} peptide. Probe **4** also contains a click reaction handle at the cysteine residue, so that the cysteine thiol group can be neutralized. Neutralizing this reactive thiol group not only prevents non-specific binding but also provides a click handle for fluorophore attachment. The fluorophore attached probe **4** can be a substitute to probe **3**. For the photo-crosslinker, benzophenone was chosen because benzophenone is more stable than the diazirine photo-crosslinker used in project 1, and therefore the peptide probe is likely to have selective crosslinking. The peptide is also more stable under ambient light.

The synthesis route of **4** is summarized in **Figure 14**. First, warhead **5** was synthesized by SPPS, then **5** was ligated with the $Cys(PR)_{20}$ peptide to form compound **6**, followed by *S*-alkylation with propargyl bromide to form the final product **4**. I expected that probe **4** could bind to certain proteasome subunit(s) in the same manner as PR_{20}-X. The band from the blotting experiment can be further analyzed by MS to identify the proteasome subunit.

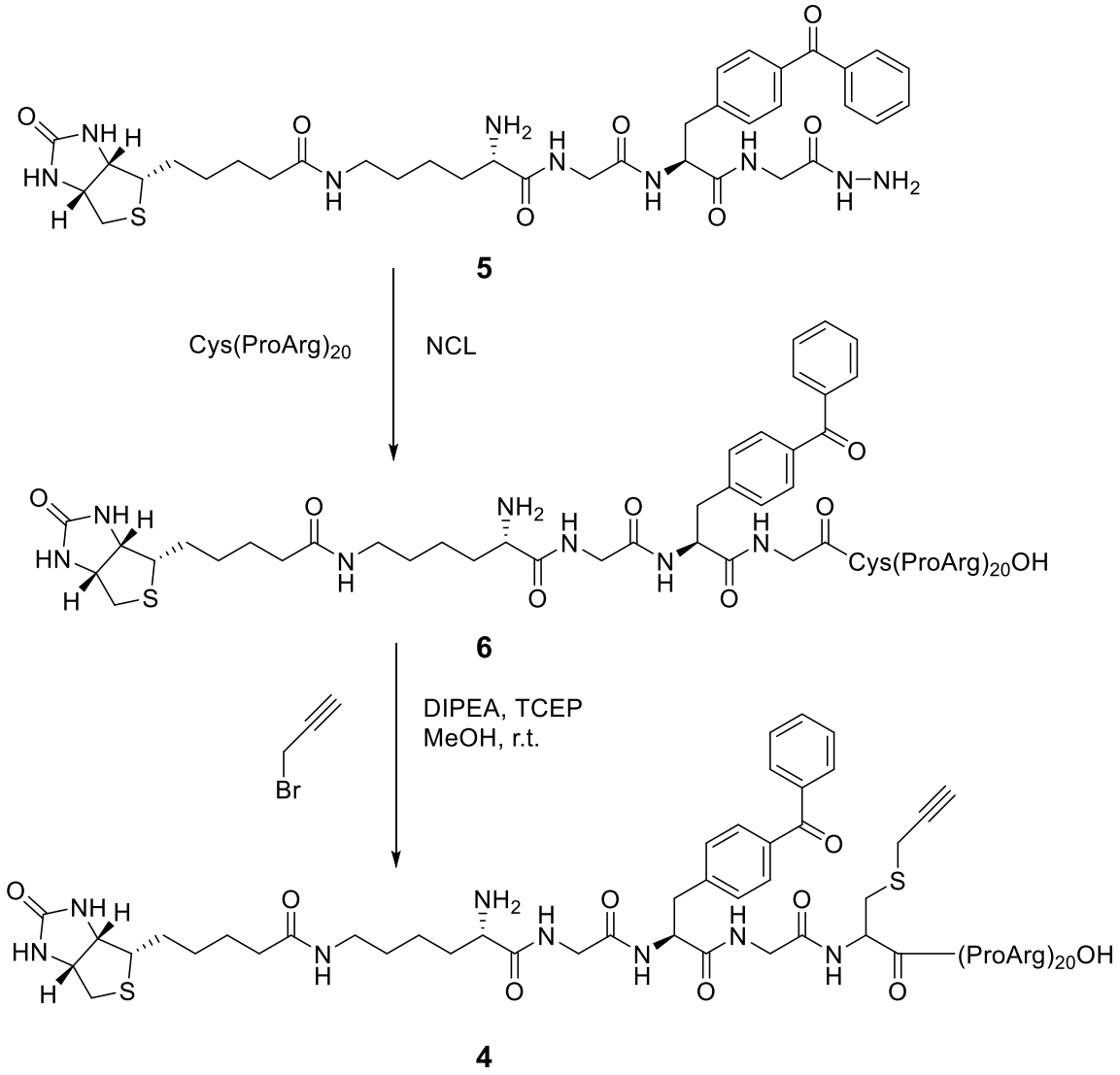


Figure 14. Synthesis Route of Warhead-PR_{20} Probe **4**

The preparation of the PR_{20} peptide suffers from low yield and difficulties in quantification. I initially synthesized the PR_{20} peptide by using a CEM Liberty peptide synthesizer but less than 1% yield product was recovered. One reason for the low yield is the length of the PR_{20} peptide is at the maximum limit of SPPS. As the peptide chain grows, the polymer may aggregate or become sterically inaccessible, and prevents further chain extension. Another reason is that the arginine 2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl (Pbf) protecting group is not very acid lable, the removal of Pbf groups requiring prolonged treatment of high concentration TFA. However, the same condition can also hydrolyze the peptide backbone, so the deprotection time was limited to a few hours. As a result, we observed incomplete deprotection by-products from the crude peptide, having an m/z of 252 Da or 504 more than the expected mass. 1,2-Dimethylindole-3-sulfonyl (MIS) was reported to be a more acid lable substitution

for the Pbf group, but there is no commercially available Fmoc-Arg(MIS)-OH amino acid in the market as of 2017.[50]

In probe **4** synthesis, the Cys(PR)$_{20}$ peptide was quantified by Ellman's reagent without pretreatment. According to Ellman's reagent standard protocols (Sigma-Aldrich manual #22582), the peptide should be treated by immobilized TCEP to remove possible disulfide bonds. However, I could not elute the peptide from the immobilized TCEP resin. The most likely reason is that the guanidinium group from the arginine sidechain binds to the resin tightly, preventing elution from the resin. Without TCEP treatment, the quantification will underestimate the concentration of the Cys(PR)$_{20}$ peptide. To compensate for this problem, I used an excess amount of the compound **5** during the NCL process.

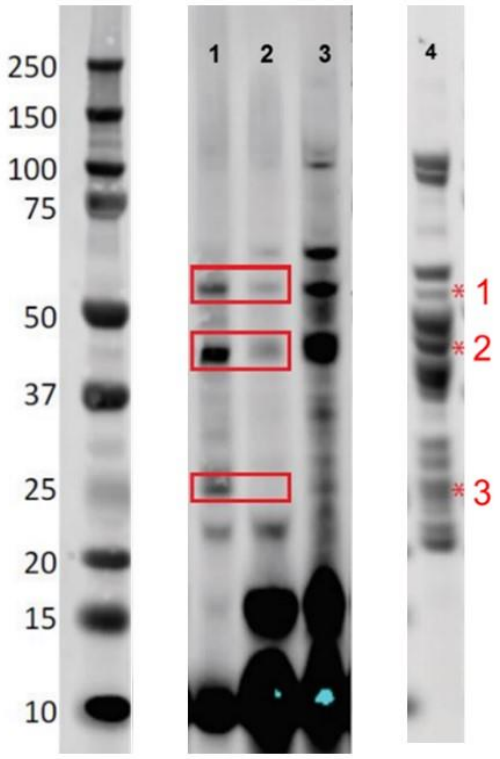


Figure 15. PR$_{20}$ Peptide Crosslinks to Proteasome Subunits. Lane 1: proteasome crosslinked with the PR$_{20}$-X probe. Lane 2: same to lane 1 plus excess unlabeled PR$_{20}$ peptide to compete off PR$_{20}$-X probe binding. Lane 3: proteasome crosslinked with the PR$_{20}$-X probe plus an excess amount of unlabeled GR$_{20}$ peptide. Lane 4: Coomassie stained proteasome subunits. The red boxes indicate the corresponding bands from lanes 1 and 2 to lane 4. (Kalb, R. Children's Hospital of Philadelphia, PA, USA. Unpublished work, 2016).

Proteasome Binding Assay The proteasome binding assay using the PR$_{20}$-X probe was performed by the Kalb group from the Children's Hospital of Philadelphia. **Figure 15** is a streptavidin Western blot of proteasome subunits, where the probes and the cell lysate were crosslinked by UV irradiation, then separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS

Page) (Kalb, R. Children's Hospital of Philadelphia, PA, USA. Unpublished work, 2016). The three clear bands in lane 1 suggested that three proteasome subunits can bind to the PR_{20}-X probe, labeled as 1, 2, and 3. The bands are less intense in lane 2, suggesting excess unlabeled PR_{20} can compete with PR_{20}-X for proteasome binding sites, confirming that the PR_{20}-X probe behaves similarly to the unlabeled PR_{20} peptide. The patten of bands in lane 3 is similar to lane 1, suggesting that the GR_{20} peptide cannot compete with PR_{20}-X for proteasome binding sites. This result agrees with the fact that the GR20 peptide is non-toxic while the PR_{20} peptide is highly toxic.[26] This result indicates the PR_{20} peptide can bind to the proteasome subunits while the non-toxic GR_{20} peptide can not.

The three bands indentifed by the blotting experiment was anlyzed by the Proteomics Core Facility at the Children's Hospital of Philadelphia (CHOP) Research Institute. Serveal protein candidates were identified based on the molecular weight of the protein target (**Table 2**).[51] The binding pattern is not obvious according to the MS result. I hypothesize that all subunits that bind to the PR_{20} peptide have a common motif, but further experiments with probe **4** is needed to confirm the crosslinked protein identities and to explain the binding mechanism.

Table 2. Crosslinked Proteasome Proteins Candidates

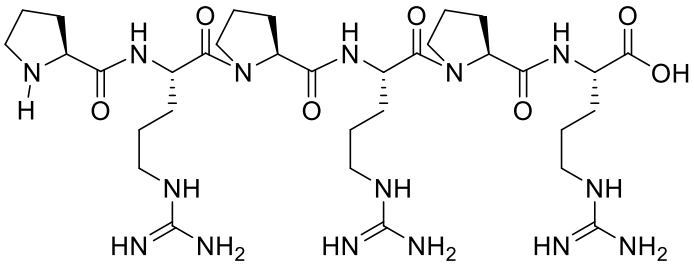
Band Number	*candidate proteins*	*molecular weights (Kda)*	*UniProt accession numbers*
Band 1	26S protease regulatory subunit 4	49	P62191
	26S proteasome non-ATPase regulatory subunit 2	100	Q13200
Band 2	26S proteasome non-ATPase regulatory subunit 11	47	O00231
	Isoform 2 of 26S protease regulatory subunit 8	45	P62195-2
Band 3	Proteasome subunit alpha type-6	27	P60900
	Proteasome subunit alpha type-3	28	P25788
	Proteasome subunit alpha type-5	26	P28066
	Proteasome subunit beta type-7	30	Q99436
	Proteasome subunit alpha type-7	28	O14818

Source: Adapted from Lan, M. The Role of the Proline-Arginine Dipeptide Repeat in Amyotrophic Lateral Sclerosis. M.S. Dissertation, University of Pennsylvania, Philadelphia, PA, USA, 2017.

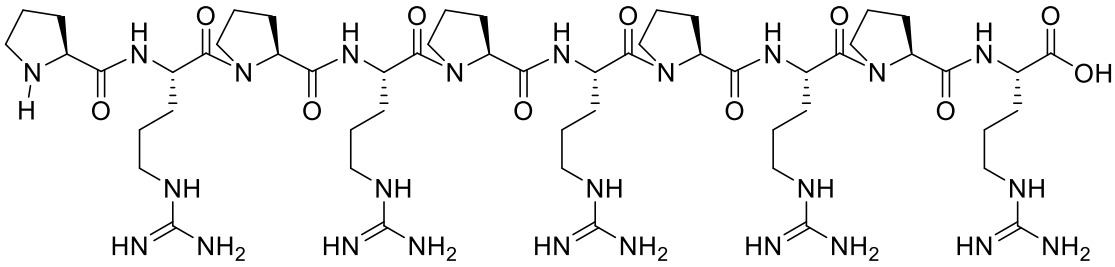
Binding Competition of PRx Peptides with Different Lengths The final goal of this project was to verify if the length of the PR_x peptides determines the binding activity and toxicity. In most previous studies, the PR_{20} peptide was used as the model for toxic PR_x.[24,26] This gave rise to the question if the length of PR_x peptides can have effects in proteasome binding. To answer this question, I synthesized two unlabeled versions of PR_x peptide: $(Pro\text{-}Arg)_3$ (**7**) and $(Pro\text{-}Arg)_5$ (**8**) in this project (**Figure 16**). A binding competition assay using **7** and **8** against the PR-X

probe was performed to determine if peptide **7** and **8** can bind to the proteasome and compete out the PR_{20}-X probe (Kalb, R. Children's Hospital of Philadelphia, PA, USA. Unpublished work, 2017).

Binding Competition Assay The binding competition assay was performed by incubating the PR_{20}-X probe, PR_3 (**7**) and PR_5 (**8**) peptides with proteasome subunits. Then the mixture was crosslinked by UV irradiation followed by avidin Blot (**Figure 17**). Lane 2 and lane 3 indicated the natural PR_{20} peptide could compete out the PR_{20}-X probe in proteasome binding. This serves as a control to prove that the PR_{20}-X probe and PR_{20} peptide have binding affinity. It can be observed that the PR_3 and the PR_5 peptides do not compete with the PR_{20}-X probe to bind to the proteasome, as indicated by comparing lane 2 to lanes 6 and 7. This suggested that PR_x peptide binding to the proteasome is length dependent and sufficiently cooperative; multiple PR_3 or PR_5 peptides cannot simply occupy the binding site for PR_{20}.



7



8

Figure 16. The Structure of PR_3 Peptide **7** (top) and PR_5 Peptide **8** (bottom)

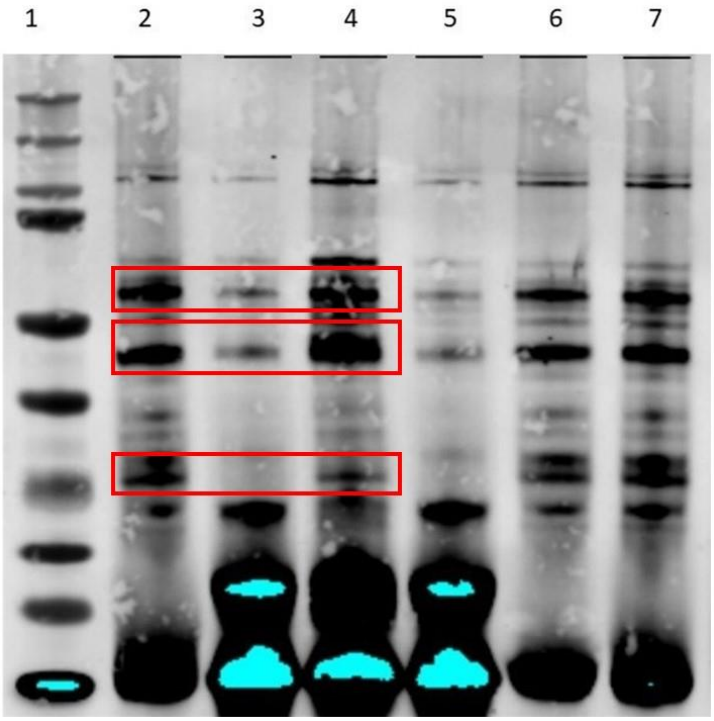


Figure 17. Binding Competition Assay by Streptavidin Blot. Lane 1 to 7 (left to right) represents: ladder; PR_{20}-X probe with proteasome; PR_{20}-X peptide with proteasome with extra PR_{20} peptide; GR_{20}-X peptide with proteasome; identical to lane 3 but with extra incubation time before UV irradiation; PR_3 peptide and PR_{20}-X probe with proteasome; PR_5 peptide and PR_{20}-X probe with proteasome. All samples were irradiated by UV at 365 nm for 30 min for crosslinking before the streptavidin blot. The red boxes indicate the same bands as in **Figure 15**. (Kalb, R. Children's Hospital of Philadelphia, PA, USA. Unpublished work, 2017)

Conclusion and Future Work

To summarize this study, our synthetic probes provided tools to study the mechanism of ALS disease and histone function *in vitro* or in cells. In project 1, I used SPPS to synthesize two 23-mer probes **1** and **2** to perform a crossslinking experiment to identify the protein(s) that bind to an N-ac histone H4 peptide. In the blotting experiment, some protein targets were found with binding preference to the N-ac histone peptide. The next step would be to replicate these results and use MS to identify the proteins. In project 2, I synthesized another set of probes, including a 41-mer TAMRA fluorophore labeled PR_{20} peptide **3**, and a 46-mer PR_{20} peptide with a photo-crosslinking warhead probe **4**, a 6-mer PR3 peptide **7,** and a 10-mer PR_5 peptide **8**. This set of probes were synthesized for the study of the toxicity mechanism of PR_x peptide. The preliminary *in vitro* result demonstrated that the length of the PR_x peptide plays in role in proteasome binding. The PR_x peptide binds to the proteasome with a defined pattern. The probes **3** and **6** will be tested by the Kalb group.

PR_x Peptide Synthesis by Recombinant Protein Expression Peptides in the 30 to 40 amino acid range are usually too short for protein expression and are generated by SPPS. In my syntheses of the PR_{20} peptide by SPPS, I found the yields to be low due to the difficult removal of arginine Pbf sidechain protection groups. To synthesize unconventional peptides like the PR_x peptides, we would like to use recombinant protein expression to synthesize the peptide. Purification tags can facilitate protein expression and stabilize the peptide. Previously, the Petersson laboratory successfully used an intein as a traceless purification tag to express α-synuclein (140 amino acids) containing the unnatural amino acid *p*-cyanophenylalanine at position 136 using orthogonal aminoacyl-tRNA synthetase/tRNA pairs.[52,53] I could use inteins and the Small Ubiquitin-like Modifier (SUMO) proteins as a generalizable purification tag for medium size peptides that are unsuitable for NCL, but too long for efficient SPPS.

We performed the preliminary experiment of PR_{20} peptide biosynthesis but the result was unsuccessful. We transformed pET His6 Sumo TEV CPR_{20} plasmid (**Appendix A**) into *E. coli.* BL21(DE3). **Figure 18** shows the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS Page) gel image of the SUMO-CPR_{20} fusion protein expression and Ni column purification elution, stained with Coomassie blue. Lane 1 and 2 is the cell lysate before and after transformation but before Isopropyl β-D-1-thiogalactopyranoside (IPTG) induction, respectively. The intense induction band near to the top of the gel at lane 2 suggests the fusion protein can be readily induced by IPTG. However, the induction band could not be observed in lane 3, which is from to harvested cell lysate and Ni column purification elution, respectively. The loss of the induction band suggests that the target protein was lost during the process of cell harvesting. We believe this is because the PR_{20} fusion protein has a high arginine content. It was known that arginine has a high binding affinity towards nucleic acids, so a protein with high arginine content is likely to precipitate with other cellular components during cell harvesting process.[54] To overcome this problem, we need to find the cell lysing condition that prevents Arginine from binding to nucleic acid. The addition of glycerol and high concentration of potassium chloride into the cell lysing buffer can prevent the interactions between arginine-rich peptides and nucleic acid (Skordalakes, E. Wistar Institute, Philadelphia, PA. Personal communication, 2017.). This strategy will be tested in the future.

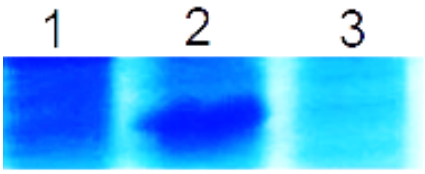


Figure 18. SDS Page of His6-SUMO-CysPR_{20} Fusion Protein Expression and Ni Column Purification.

Chapter 2
Thioamide as a Proteolysis-resistant Modification

Introduction

Importance and Significance Peptide-based medicine has a wide range of applications and huge value in marketing. The global annual sale of peptide-based medicine was $14.1 billion in 2011 and was predicted to be $25.4 billion in 2018.[55] One of the challenges to peptide-based medicine is that most naturally occurring peptides are susceptible to proteolysis in plasma and have a very short half-life. For example, the incretin hormone glucagon-like peptide-1 (GLP-1) has a half-life of 1.5 to 5 minutes in plasma.[56] The use of peptide cyclization, D- and β-amino acids substitution can stabilize peptides in plasma.[57,58] These three approaches have had success in the past but they all change the structure or conformation of the target molecules, which may lead to unpredictable effect on bioactivity. The Petersson laboratory has recently showed that the single-atom substitution of the thioamide can be introduced into protease substrates to stabilize peptides against proteolysis. The thioamide modification can be used to stabilize peptides against proteolysis in peptide-based medicine while introducing minimal perturbation to the structure.

Thioamide Physical Properties The physical differences between the thioamide bond and amide bond are summarized in **Table 3**. The thioamide bond and amide bond are nearly isosteric, but the bond length of the thioamide bond is longer. The thioamide bond is weaker and more polarizable compared to the amide bond.[59] The most significant differences between thioamide and amide are in the hydrogen bond donor-acceptor properties. The sulfur atom in the thioamide is a weaker hydrogen bond acceptor, but the adjacent N-H group is a stronger hydrogen bond donor.[60]

Table 3. Thioamide Properties

X=O or S	*Oxoamide*	*Thioamide*
van der Waals radius (Å)	1.40	1.85
C=X bond length (Å)	1.24	1.54-1.68
C=X bond dissociation energy (kcal/mol)	170	130
N-H pK_a	17	11-13
C-N rotational barrier (kcal/mol)	≈15-19	≈18-22
π to π* absorption (nm)	190	270
Oxidation potential (eV)	3.29	1.21

Source: Adapted from Petersson, E. J.; Goldberg, J. M.; Wissner, R. F. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6827–6837.

Thioamides as Fluorescence Quenchers Sulfur substitution shifts the amide π to π* transition from 190 nm (oxo) to 270 nm (thio), giving it spectral overlap with UV wavelength fluorophores that permits Förster resonance energy transfer (FRET). In addition, there is a 2.1 eV change in the oxidation potential that makes

thioamides able to act as electron donors in photo-induced electron transfer (PET) reactions. These two properties allow the thioamide to be used as a quencher of fluorophores including cyanophenylalanine (Cnf), acridonylalanine (Acd), fluorescein (Fam), 7-methoxycoumarin (Mcm), tryptophan, and tyrosine.[60,61] These quenching interactions can be used to monitor conformational changes in proteins or cleavage by proteases, using doubly labeled peptides.[62] The Petersson laboratory previously developed a real-time proteolysis assay using fluorescence labeled thioamide peptide.[61] The probe was designed so that the Mcm fluorescence was quenched by the thioamide. After the cleavage of the scissile bond by the protease, the fluorescence was turned on and gave an increased fluorescence signal that could be analyzed by a plate-reader, as shown in **Figure 19**. I used this approach to monitor the protease activity in this study.

General Proteases Properties Proteases, also referred as peptidases, or proteinases are enzymes that catalyze the hydrolysis of peptide bonds. Protease signaling pathways are vital to many biological processes and the activity of proteases is highly regulated in the body.[63] For this reason, many protease inhibitors have been developed as therapeutics. *Captopril* is an angiotensin-converting enzyme (ACE) protease inhibitor for hypertension treatment;[63] *Ritonavir* is a (human immunodeficiency virus) HIV protease inhibitor for (acquired immunodeficiency syndrome) AIDS treatment;[63] *Sivelestat* is a neutrophil elastase inhibitor for acute respiratory failure treatment.[63]

The substrates of protease follow a special nomenclature: residues in the N-terminal direction from the cleaved bond are designated P1, P2, P3, P4 etc.; the residues in C-terminal direction are designated P1', P2', P3', P4'. etc.[64] Based on the site of cleavage, proteases can be divided in to endopeptidase and exopeptidase. Endopeptidase cleave peptide bonds of at non-terminal amino acids while exopeptidase cleave peptide bonds of terminal amino acids. Based on the mechanism of catalysis, proteases in mammals can be classified into five classes: aspartic, cysteine, serine, threonine, and metalloproteases proteases.[65] Among these classes, serine protease class is the best studied class, which uses the classical Ser/His/Asp catalytic triad mechanism, where the serine residue is used as the nucleophile.[66] Since proteases degrade peptide-based medicine *in vivo*, it is critical to understand the effect of thioamide incorporation with different classes of proteases.

Previous Thiopeptide Protease-Resistant Studies Since hydrogen bonds play critical roles in biological activities, the altered hydrogen bond properties can change the potency and stability of the peptide. The changes of properties depend on the site of incorporation. A thioamide at position 2 of the Leu-enkephalin (Leu-Enk) peptide increased the potency, but a thioamide at position 1 or 3 decreased the potency, as shown in **Figure 20**.[67] Certain thioamide peptides has been shown to be resistant against carboxypeptidase A,[68–71] HIV-1 proteases,[72] membrane dipeptidase,[73] papain,[74] trypsin,[75] leucine aminopeptidase,[76] and prolyl oligopeptidase.[77] Some studies have also previously evaluated the effectiveness of thio-peptide therapeutics *in vivo*. Zhang and coworkers[78] showed that polybia-MPI thio-peptides has increased the half-life and anticancer activity. Zacharie and coworkers[79] reported thioamide Imreg peptides have slightly improved *in vivo*

immune-stimulant effect and serum half-life. In all studies, the activity and proteolysis stability of thioamide largely depends on the position of thioamide incorporation.

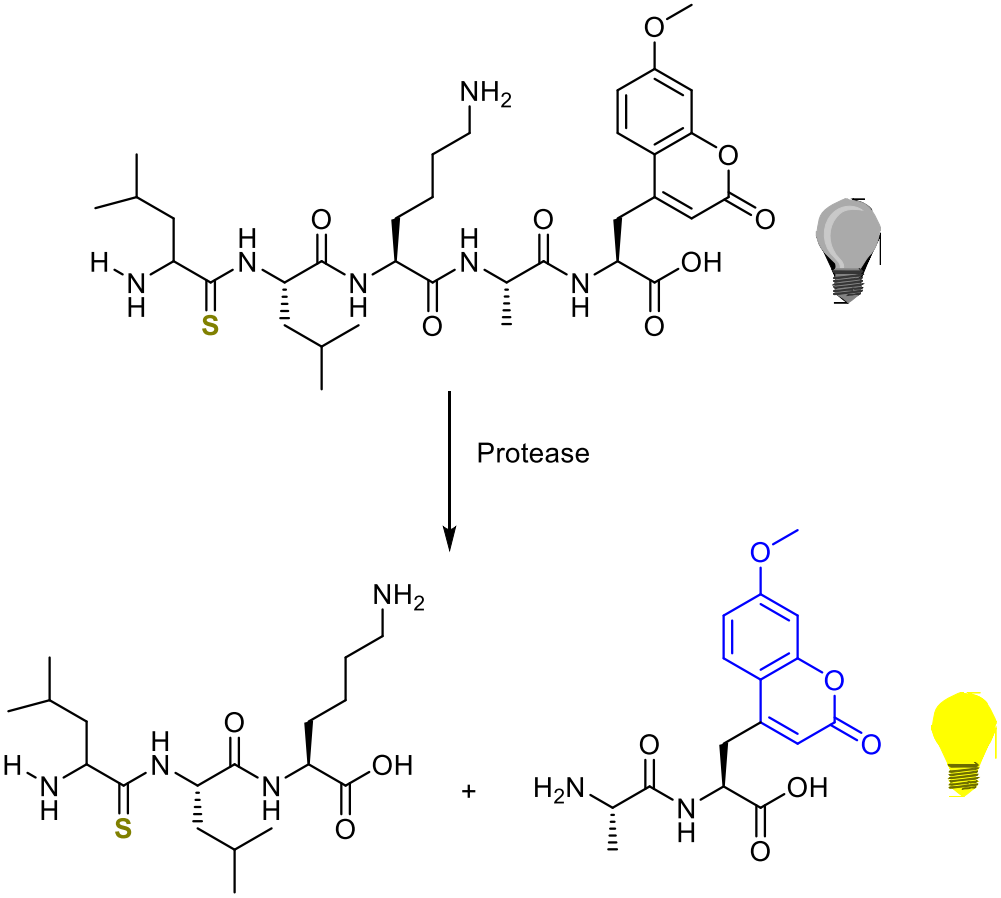


Figure 19. Mechanism of Singly Labeled Fluorescence Probe.[61]

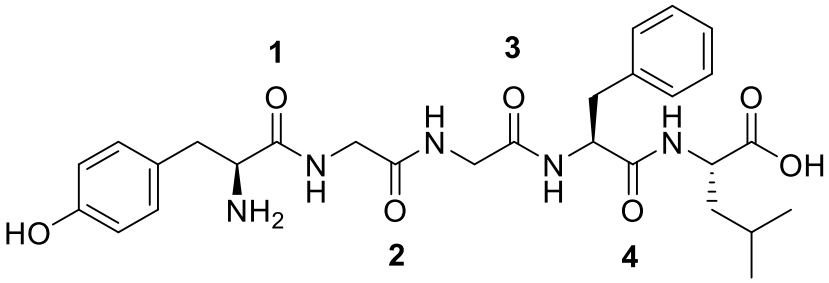


Figure 20. Chemical Structure of Leu-enkephalin Peptide.[67]

Purpose For applications in peptide therapeutics, the positioning of the thioamide must be disruptive to interactions with the protease, but not with the peptide receptor. The Petersson laboratory has demonstrated thioamide peptide can stabilize GLP-1 without disrupting activity in cell based assays.[80] In separate work, the Petersson laboratory has demonstrated that thioamide quenchers can be used in fluorescent peptide sensors to probe to protease activity.[61] In the case of protease sensors, one wishes the thioamide to be as close as possible to the scissile bond without altering k_{cat} or K_M compared to the parent peptide. To fulfill both requirements for thioamide, it is important to gain a predictive understanding of which thioamide positions are disruptive for a given protease. To gain a better understanding of the positional effects of thioamide, I synthesized a set of probes with thioamide at each peptide bond positions and measure the rates of proteolysis.

Experimental Design The Petersson laboratory previously demonstrated that thioamides can be used as fluorescence quenchers in protease activity sensors using the probe shown in **Figure 19**. This sensor can monitor the activity of protease, but cannot be used to study the positional effect of thioamide due to limitations to the possible positions of thioamide incorporation. The singly labeled probe cannot provide kinetic information for the oxo-peptide control. To overcome these two problems, a doubly labeled peptide sensor was designed with identical fluorophores at each end, as shown in **Figure 21** (fluorophores highlighted in blue). 7-methoxycoumarin (Mcm) was chosen as the fluorophore for the sensor since Mcm can be quenched by both thioamide and by intramolecular self-quenching. This probe design can monitor the rate of proteolysis with thioamide incorporated at any peptide bond position. The synthesis and proteolysis assays followed the previous established methods.[61]

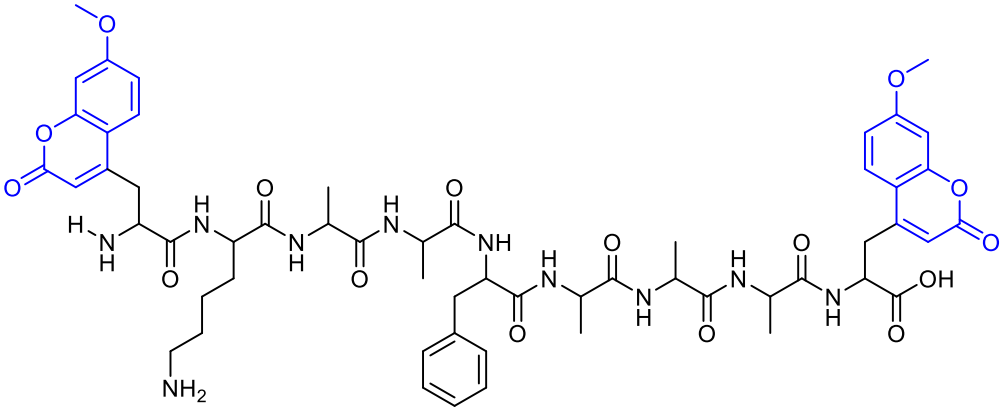


Figure 21. Proposed Template of Doubly Labeled Thioamide Proteolysis Sensor

Materials and Methods

General Information Fmoc-β-(7-methoxycoumarin-4-yl)-Ala-OH (μ) was purchased from Bachem (Torrance, CA, USA). Fmoc-Lys(biotinyl)-OH was purchased from TCI America (Portland, OR, USA). 2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU), 2-Chlorotrityl chloride resin (100-200 mesh, 1% DVB), Fmoc-Ala-OH, Fmoc-Lys(Boc)-OH, Fmoc-Phe-OH were purchased from Novabiochem (currently EMD Millipore; Billerica, MA, USA). Piperidine was purchased from American Bioanalytical (Natick, MA, USA). Sigmacote®, *N,N*-diisopropylethylamine (DIPEA), 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU) and Trifluoroacetic acid (TFA) were purchased from Sigma-Aldrich (St. Louis, MO, USA). Glass peptide synthesis reaction vessels (RVs) were treated with Sigmacote® prior to use. Triisopropylsilane (TIPS) was purchased from Santa Cruz Biotechnology (Dallas, TX, USA). Chymotrypsin (lyophilized powder; from Bovine Pancreases) was purchased from MP Biomedicals (Santa Ana, CA, USA). Nα-Fmoc-L-Alanine-2-amino-5-nitroanilide and Nα-Fmoc-L- Phenylalanine-2-amino-5-nitroanilide were synthesized by former lab member Yanxin Wang. All other reagents including solvents were purchased from Fisher Scientific (Hampton, NH, USA) unless specified otherwise.

Instruments Crude peptides were purified with a Varian ProStar High-Performance Liquid Chromatography (HPLC) instrument outfitted with a diode array detector (currently Agilent Technologies) using aqueous A (H_2O + 0.1% TFA) and organic B (CH_3CN + 0.1% TFA) phases. Purification testing was performed in an Agilent 1100 Series HPLC with a 4.6nm Phenomenex Luna C8 (2) column. UV absorbance spectra were obtained with a Hewlett-Packard 8452A diode array spectrophotometer (currently Agilent Technologies; Santa Clara, CA, USA). Matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) data were collected with a Bruker Ultraflex III MALDI-TOF-TOF mass spectrometer (Billerica, MA, USA). Fluorescence spectra were collected with a Tecan M1000 plate reader (San Jose, CA, USA)

Peptide Synthesis The synthesis of thioamide peptide was performed following a published method.[81] For amino acids other than thioamide-containing amino acids, a standard synthetic method identical to Chapter 1 was performed. To couple the thioamide, 2 equiv. of pre-activated precursors was dissolved in 2 ml of dry dichloromethane (DCM) and then added to the RV. 2 equiv. of DIPEA was added to the RV and stirred for 30 min. The RV was drained, then another 2 equiv. of pre-activated precursors in dry DCM was added to the RV with 2 equiv. of DIPEA, stirred for 30min and then drained. The Fmoc deprotection for thioamide amino acids was performed by treatment with 2% 1,8-Diazabicyclo[5.4.0]undec-7-ene (DBU) in DMF for 2 min, for three times.

After the last synthesis step, the peptide was cleaved from the resin by treatment of 2 ml of cleavage cocktail for 30 minutes with stirring (30% TFA, 2.5% TIPS, 2.5% water, 65% DCM). The TFA from crude peptide was removed by adding 2 ml of ACN and evaporated in a rotary evaporator twice. The crude peptide was then purified by HPLC and freeze dried. MALDI-TOF was used to confirm the identity of the peptide and purity was verified by an analytical HPLC.

Steady State Protease Assays To obtain kinetic data of peptide hydrolysis, the peptide was dissolved in 100 mM Tris-HCl buffer (pH 7.8) and standardized the concentration by measuring the UV absorbance at 325 nm to 5.0 µM. The peptides were added to a 96-well plate (Greiner Bio-One, 675096) and incubated in the Tecan plate reader with shaking between each measurement at 25 °C. The fluorescence was monitored as a function of time using excitation wavelength of 325 nm and emission wavelength of 390 nm. Chymotrypsin enzyme was dissolved in 1 mM HCl, 10mM $CaCl_2$, to a final concentration of 2.5 mg/ ml. The hydrolysis was initiated by adding 4 µl of the enzyme to each 46 µl peptide, making a final enzyme concentration of 0.2 mg/ml. For control groups, 4 µl of 1 mM HCl, 10mM $CaCl_2$ was added to the peptide. Each hydrolysis trail was performed in triplicates and control in duplicates. The result of the three trials was averaged and fitted to the one phase decay exponential function using GraphPad Prism to calculate the half-life.

Results and Discussion

Peptide Characterization To monitor the positional effects of thioamides on chymotrypsin proteolysis, doubly Mcm-labeled peptides with a thioamide at each peptide position from P4 to P3' were synthesized (**10**-**16**). The corresponding oxo-peptide was also synthesized as the control (**9**). Peptide **9**, **13**-**15** were synthesized by Chunxiao Liu, Peptide **15** was synthesized by Xing Chen. I synthesized peptide **10**-**12**. All peptides synthesized were characterized by MALDI-MS, expected $[M+H]^+$ ions were observed as shown in **Table 4**, which µ refers to Mcm derivative β-(7-methoxycoumarin-4-yl)-Ala and superscript S refers to position of thioamide.

Table 4. Characterization of Thioamide Peptides

Molecule	*Sequence*	*Calculated (m/z)*	*Observed (m/z)*
9	µKAAFAAAµ	1139	1139
10	µK^SAAFAAAµ	1155	1155
11	µKA^SAFAAAµ	1155	1155
12	µKAA^SFAAAµ	1155	1155
13	µKAAF^SAAAµ	1155	1155
14	µKAAFA^SAAµ	1155	1155
15	µKAAFAA^SAµ	1155	1155
16	µKAAFAAA^Sµ	1155	1155

Enzyme Activity Assays Positional screening of thio-peptides hydrolysis from P4 to P3' positions was performed using chymotrypsin. Chymotrypsin hydrolysis was carried out with 45 µl of the 5µM peptide concentration in 100 mM HCl, pH 7.8, at room temperature in the presence or absence of 0.2 mg/ml chymotrypsin. **Figure 22** shows the fluorescence traces of peptide hydrolysis. The fluorescence intensity of peptides without protease (orange traces) was approximately 60%

lower than the intensity after hydrolysis (red traces), indicating the doubly labeled thioamide peptides was quenched by approximately 60%. The fluorescence intensities of peptides without protease (orange traces) remained constant over the course of the experiment, demonstrating that the probes were stable during the experiment. Upon addition of the protease, the fluorescence signals increased exponentially over time and reached plateaus (red traces). This gain of fluorescence corresponded to the cleavage of the peptides. For peptides showing proteolysis inhibition, the growth of fluorescence intensities was slower and the traces were almost linear. The half-lives were calculated by fitting curves to a single exponential function using GraphPad Prism, and are shown in **Table 5**. The oxo-peptide **5** could not be fully dissolved in the buffer, hence the hydrolysis experiment was not performed. Data for P4 to P1 positions were obtained from former lab member Xing Chen.[80] The calculated half-lives showed that for chymotrypsin, a thioamide at the P1 or P2 position can greatly inhibit proteolysis.

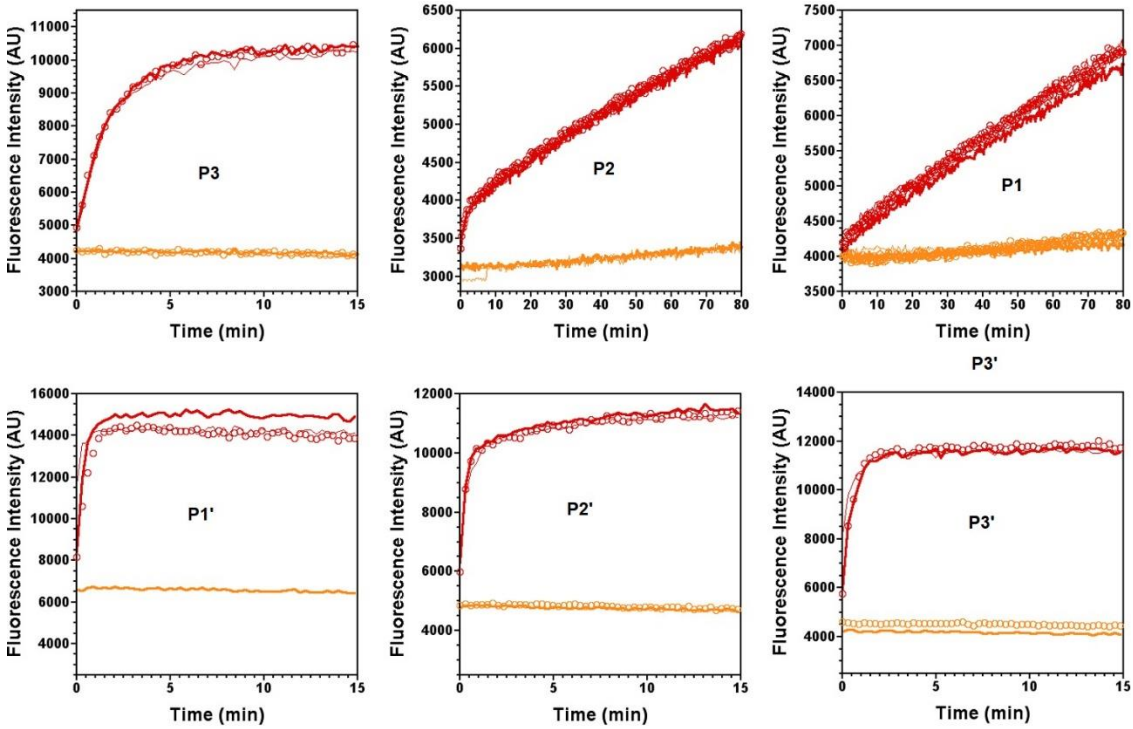


Figure 22. Representative Results for Chymotrypsin Proteolysis.

Top (left to right): thioamide incorporated to P3, P2, P1 positions, respectively. Bottom (left to right): thioamide incorporated to P1', P2', P3' positions, respectively. Red traces: Hydrolysis with chymotrypsin; Orange traces: Peptides without chymotrypsin. Three independent trials for each protease are shown as open circles, thin lines, and thick lines.

Table 5. Thioamide Positional Effects on Chymotrypsin Activity

Molecule	*Thioamide Position*	*Sequence*	*Chymotrypsin t 1/2 (min)*
9	Control	μKAAFAAAμ	not determined
10	P4	μK^{S}AAFAAAμ	0.37±0.01
11	P3	μKA^{S}AFAAAμ	1.19±0.02
12	P2	μKAA^{S}FAAAμ	>40
13	P1	μKAAF^{S}AAAμ	>40
14	P1'	μKAAFA^{S}AAμ	0.19±0.06
15	P2'	μKAAFAA^{S}Aμ	0.25±0.05
16	P3'	μKAAFAAA^{S}μ	0.32±0.06

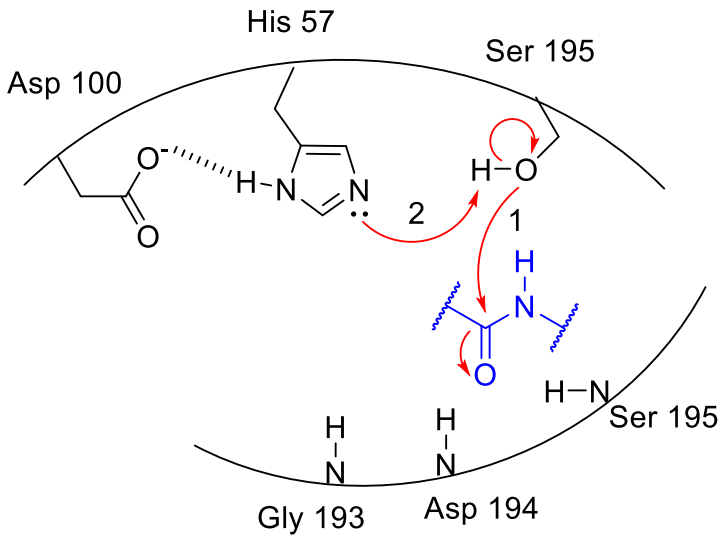


Figure 23. Nucleophilic Attack Step of Serine Protease.[83]

Sources: Adapted with permission from 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.

The protease resistance at the P1 position can be explained by the catalytic mechanism. **Figure 23** shows the nucleophilic attack step of the serine protease, were the serine sidechain serve as a nucleophile and attacks the peptide bond. His 57 is pulling a proton from Ser 195 (labeled as 2 in **Figure 23** and **24**), the protonated Ser 195 can act as a nucleophile to attack the amide bond (labeled as 1). A close examination of the crystal structure of α-chymotrypsin bound with its inhibitor shows the P1 position peptide bond is closely interacting with protein backbone at Gly 193, Asp 194 and Ser 195 positions, labeled as 3, 4, and 5, respectively in **Figure 24**.[82] Since the bond length of thioamide is longer than oxoamide and sulfur atom is larger than oxygen, the substrate binding can be sterically prohibited and inhibit proteolysis. In addition, the sulfur atom is a weaker

hydrogen bond acceptor, the reaction intermediates are destabilized in thioamide bonds, making the hydrolysis unfavorable. The inhibition mechanism on the P2 position is not obvious by inspecting the crystal structure. It is possible that the thioamide bond has a higher rotation energy barrier than the amide bond, and the substrate with thioamide at the P2 position cannot adapt the correct confirmation.

For FRET experiment applications, thioamide has the advantage of having a smaller size over other FRET quenchers. The typical fluorophores for FRET experiments are larger than the size of an amino acid while the thioamide is only a single atom substitution. The introduction of a large fluorophore molecule to the protein may perturb the structure of the protein and makes the measurements unreliable. In contrast, the isosteric thioamide is minimally-perturbing. For peptide-based medicine applications, thioamides can prevent the drug from protease degradation, if incorporated at the right position. However, since the thioamide bond is weaker than the oxoamide bond, thioamide peptide drugs may potentially have a shorter shelf life.

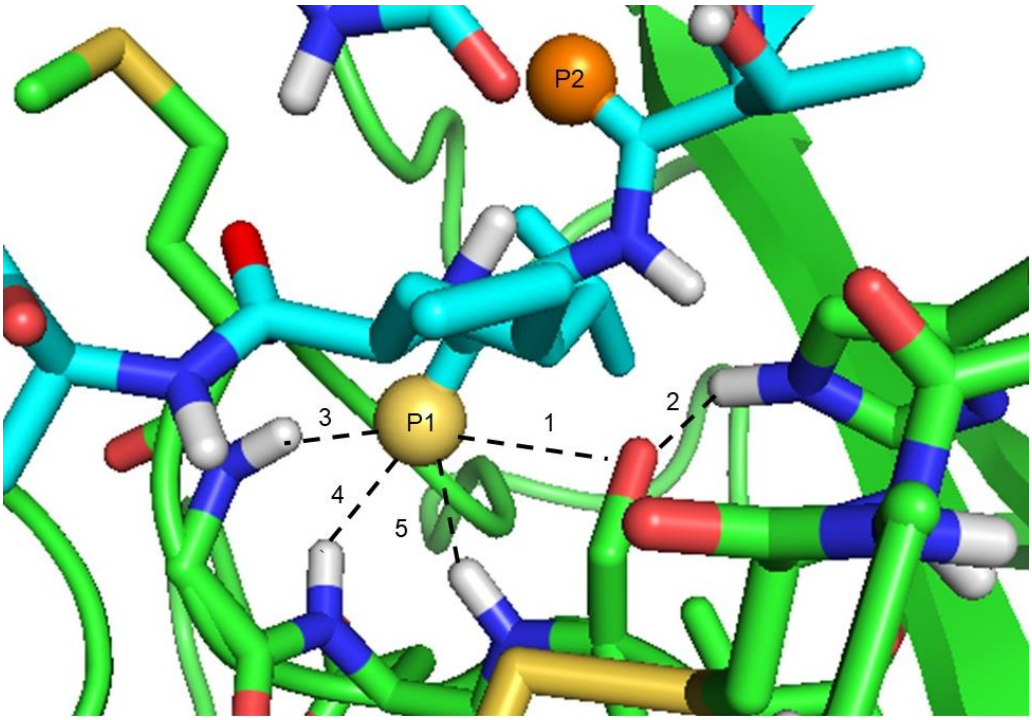


Figure 24. Crystal Structure of Chymotrypsin with Inhibitor from Insect *Locusta migratoria* (PMP-C).[82]

Sources: Adapted from Roussel, A.; Mathieu, M.; Dobbs, A.; Luu, B.; Cambillau, C.; Kellenberger, C. Complexation of Two Proteic Insect Inhibitors to the Active Site of Chymotrypsin Suggests Decoupled Roles for Binding and Selectivity. *J. Biol. Chem.* **2001**, *276* (42), 38893–38898. (PDB ID: 1GL1)

Conclusion and Future Work

To evaluate the effect of thioamides to the proteolysis process, collaborators and I synthesized one set of hydrolysis sensors with thioamide substitution from P4 to P3' positions, and performed hydrolysis studies using chymotrypsin. The result revealed that thioamide at the P1 or P2 positions can greatly inhibit chymotrypsin hydrolysis. This initial result of protease inhibition is useful for peptide based drug design and other hydrolysis probe design. If peptides of therapeutic interest are known to be cleaved by chymotrypsin-like enzymes, they could be modified with thioamides at the P2 or P1 positions for stabilization.

Several future experiments are suggested by our results. Obtaining the Michaelis–Menten kinetics parameters of thioamide probes can provide information about the mechanism of thioamide proteolysis resistance. Michaelis-Menten kinetics can be obtained by performing the photolysis experiment at different concentrations of the substrates and fit into a Michaelis–Menten model.[84]

Enzymatic Thioamide Synthetic Route Although we successfully coupled thioamides into peptides, the yields were only approximately 10-25% of the corresponding oxo-peptide yield. To make thioamide peptides a more generally-accessible tool, it is important to optimize the yield of the synthesis. Using an enzymatic reaction to introduce a backbone thioamide is a possible way to increase yield. Dunbar and Mitchell[85] reported an ATP-dependent cyclodehydratase that can cyclize a peptide on Cys, Ser, and Thr residues, forming a heterocyclic azoline structure (**Figure 25**). This azoline intermediate can be released with KHS, forming a backbone thioamide. This reaction can introduce backbone thioamide modifications to an expressed protein at specific locations. Currently, such protein can only be made by synthesizing peptide fragments and then ligating the fragments. This process is labor intensive and involves multiple purification steps. Incorporating thioamide to an expressed protein with enzyme is far more efficient.

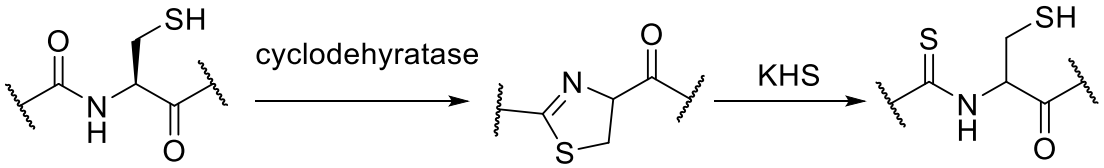


Figure 25. Scheme of Thioamide Formation by Cyclodehyratase.[85]

Sources: Adapted with permission from Dunbar, K. L.; Mitchell, D. A. Insights into the Mechanism of Peptide Cyclodehydrations Achieved through the Chemoenzymatic Generation of Amide Derivatives. *J. Am. Chem. Soc.* **2013**, *135* (23), 8692–8701. Copyright 2013 American Chemical Society.

References

1. Brown, J.; Farquhar, C. An Overview of Treatments for Endometriosis. *JAMA* **2015**, *313* (3), 296–297.
2. Kumar, P.; Sharma, A. Gonadotropin-Releasing Hormone Analogs: Understanding Advantages and Limitations. *J. Hum. Reprod. Sci.* **2014**, *7* (3), 170–174.
3. Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W. Genetic Encoding of Photocaged Cysteine Allows Photoactivation of TEV Protease in Live Mammalian Cells. *J. Am. Chem. Soc.* **2014**, *136* (6), 2240–2243.
4. Verdin, E.; Ott, M. 50 Years of Protein Acetylation: From Gene Regulation to Epigenetics, Metabolism and Beyond. *Nat. Rev. Mol. Cell Biol.* **2014**, *16* (4), 258–264.
5. Liszczak, G.; Goldberg, J. M.; Foyn, H.; Petersson, E. J.; Arnesen, T.; Marmorstein, R. Molecular Basis for N-Terminal Acetylation by the Heterodimeric NatA Complex. *Nat. Struct. Mol. Biol.* **2013**, *20* (9), 1098–1105.
6. Natsume, R.; Eitoku, M.; Akai, Y.; Sano, N.; Horikoshi, M.; Senda, T. Structure and Function of the Histone Chaperone CIA/ASF1 Complexed with Histones H3 and H4. *Nature* **2007**, *446* (7133), 338–341.
7. Yang, T.; Li, X.; Bao, X.; Fung, Y. M. E.; Li, X. D. Photo-Lysine Captures Proteins That Bind Lysine Post-Translational Modifications. *Nat. Chem. Biol.* **2015**, *12* (12), 70–72.
8. Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J. Post-Translational Modification: Nature's Escape from Genetic Imprisonment and the Basis for Dynamic Information Encoding. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **2012**, *4* (6), 565–583.
9. Brown, Z. Z.; Muller, M. M.; Kong, H. E.; Lewis, P. W.; Muir, T. W. Targeted Histone Peptides: Insights into the Spatial Regulation of the Methyltransferase PRC2 by Using a Surrogate of Heterotypic Chromatin. *Angew. Chem., Int. Ed.* **2015**, *54* (22), 6457–6461.
10. Arnesen, T. Towards a Functional Understanding of Protein N-Terminal Acetylation. *PLoS Biol.* **2011**, *9* (5).
11. Yu, M.; Gong, J.; Ma, M.; Yang, H.; Lai, J.; Wu, H.; Li, L.; Li, L.; Tan, D. Y. Immunohistochemical Analysis of Human Arrest-Defective-1 Expressed in Cancers in Vivo. *Oncol. Rep.* **2009**, *21* (4), 909–915.
12. Yi, C. H.; Pan, H.; Seebacher, J.; Jang, I. H.; Hyberts, S. G.; Heffron, G. J.; Vander Heiden, M. G.; Yang, R.; Li, F.; Locasale, J. W.; Sharfi, H.; Zhai, B.; Rodriguez-Mias, R.; Luithardt, H.; Cantley, L. C.; Daley, G. Q.; Asara, J. M.; Gygi, S. P.; Wagner, G.; Liu, C. F.; Yuan, J. Metabolic Regulation of Protein N-Alpha-Acetylation by Bcl-xL Promotes Cell Survival. *Cell* **2011**, *146* (4), 607–620.
13. Shemorry, A.; Hwang, C. S.; Varshavsky, A. Control of Protein Quality and Stoichiometries by N-Terminal Acetylation and the N-End Rule Pathway. *Mol. Cell* **2013**, *50* (4), 540–551.

14. Hwang, C.-S.; Shemorry, A.; Varshavsky, A. N-Terminal Acetylation of Cellular Proteins Creates Specific Degradation Signals. *Science* **2010**, *327* (5968), 973–977.
15. Zeng, L.; Zhang, Q.; Li, S.; Plotnikov, A. N.; Walsh, M. J.; Zhou, M. Mechanism and Regulation of Acetylated Histone Binding by the Tandem PHD Finger of DPF3b. *Nature* **2010**, *466* (7303), 258–262.
16. Taylor, J. P.; Brown, R. H.; Cleveland, D. W. Decoding ALS: From Genes to Mechanism. *Nature* **2016**, *539* (7628), 197–206.
17. Mathis, S.; Couratier, P.; Julian, A.; Corcia, P.; Le Masson, G. Current View and Perspectives in Amyotrophic Lateral Sclerosis. *Neural Regen. Res.* **2017**, *12* (2), 181.
18. Abe, K.; Itoyama, Y.; Sobue, G.; Tsuji, S.; Aoki, M.; Doyu, M.; Hamada, C.; Kondo, K.; Yoneoka, T.; Akimoto, M.; Yoshino, H. Confirmatory Double-Blind, Parallel-Group, Placebo-Controlled Study of Efficacy and Safety of Edaravone (MCI-186) in Amyotrophic Lateral Sclerosis Patients. *Amyotroph. Lateral Scler. Front. Degener.* **2014**, *15* (7–8), 610–617.
19. Yacila, G.; Sari, Y. Potential Therapeutic Drugs and Methods for the Treatment of Amyotrophic Lateral Sclerosis. *Curr. Med. Chem.* **2014**, *21* (31), 3583–3593.
20. DeJesus-Hernandez, M.; Mackenzie, I. R.; Boeve, B. F.; Boxer, A. L.; Baker, M.; Rutherford, N. J.; Nicholson, A. M.; Finch, N. A.; Flynn, H.; Adamson, J.; Kouri, N.; Wojtas, A.; Sengdy, P.; Hsiung, G. Y. R.; Karydas, A.; Seeley, W. W.; Josephs, K. A.; Coppola, G.; Geschwind, D. H.; Wszolek, Z. K.; Feldman, H.; Knopman, D. S.; Petersen, R. C.; Miller, B. L.; Dickson, D. W.; Boylan, K. B.; Graff-Radford, N. R.; Rademakers, R. Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron* **2011**, *72* (2), 245–256.
21. Mori, K.; Weng, S.-M.; Arzberger, T.; May, S.; Rentzsch, K.; Kremmer, E.; Schmid, B.; Kretzschmar, H. A.; Cruts, M.; Van Broeckhoven, C.; Haass, C.; Edbauer, D. The C9orf72 GGGGCC Repeat Is Translated into Aggregating Dipeptide-Repeat Proteins in FTLD/ALS. *Science* **2013**, *339* (6125), 1335–1338.
22. Donnelly, C. J.; Zhang, P. W.; Pham, J. T.; Heusler, A. R.; Mistry, N. A.; Vidensky, S.; Daley, E. L.; Poth, E. M.; Hoover, B.; Fines, D. M.; Maragakis, N.; Tienari, P. J.; Petrucelli, L.; Traynor, B. J.; Wang, J.; Rigo, F.; Bennett, C. F.; Blackshaw, S.; Sattler, R.; Rothstein, J. D. RNA Toxicity from the ALS/FTD C9ORF72 Expansion Is Mitigated by Antisense Intervention. *Neuron* **2013**, *80* (2), 415–428.
23. Jovičić, A.; Mertens, J.; Boeynaems, S.; Bogaert, E.; Chai, N.; Yamada, S. B.; Paul, J. W.; Sun, S.; Herdy, J. R.; Bieri, G.; Kramer, N. J.; Gage, F. H.; Van Den Bosch, L.; Robberecht, W.; Gitler, A. D. Modifiers of C9orf72 Dipeptide Repeat Toxicity Connect Nucleocytoplasmic Transport Defects to FTD/ALS. *Nat. Neurosci.* **2015**, *18* (9), 1226–1229.
24. Shi, K. Y.; Mori, E.; Nizami, Z. F.; Lin, Y.; Kato, M.; Xiang, S.; Wu, L. C.; Ding, M.; Yu, Y.; Gall, J. G.; McKnight, S. L. Toxic PRn Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Block Nuclear Import and Export. *Proc.*

Natl. Acad. Sci. U. S. A. **2017**, *114* (7), 1111–1117.
25. Taylor, J. P. A PR Plug for the Nuclear Pore in Amyotrophic Lateral Sclerosis. *Proc. Natl. Acad. Sci.* **2017**, *114* (7), 1445–1447.
26. Gupta, R.; Lan, M.; Mojsilovic-Petrovic, J.; Choi, W. H.; Safren, N.; Barmada, S.; Lee, M. J.; Kalb, R. The Proline/Arginine Dipeptide from Hexanucleotide Repeat Expanded C9ORF72 Inhibits the Proteasome. *eneuro* **2017**, *4* (1), ENEURO.0249-16.2017.
27. Calce, E.; De Luca, S. The Cysteine S-Alkylation Reaction as a Synthetic Method to Covalently Modify Peptide Sequences. *Chem. Eur. J.* **2017**, *23* (2), 224–233.
28. Castro, V.; Rodríguez, H.; Albericio, F. CuAAC: An Efficient Click Chemistry Reaction on Solid Phase. *ACS Comb. Sci.* **2016**, *18* (1), 1–14.
29. Merrifield, R. B. B. Solid Phase Peptide Synthesis. I. The Synthesis of. *J. AM. CHEM. SOC.* **1963**, *86* (14), 2149–2154.
30. Pedersen, S. L.; Tofteng, A. P.; Malik, L.; Jensen, K. J. Microwave Heating in Solid-Phase Peptide Synthesis. *Chem. Soc. Rev.* **2012**, *41* (5), 1826–1844.
31. Bray, B. L. Large-Scale Manufacture of Peptide Therapeutics by Chemical Synthesis. *Nat. Rev. Drug Discov.* **2003**, *2* (7), 587–593.
32. Hackenberger, C. P. R.; Schwarzer, D. Chemoselective Ligation and Modification Strategies for Peptides and Proteins. *Angew. Chem., Int. Ed.* **2008**, *47* (52), 10030–10074.
33. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of Proteins by Native Chemical Ligation. *Science* **1994**, *266* (5186), 776–779.
34. Thapa, P.; Zhang, R. Y.; Menon, V.; Bingham, J. P. Native Chemical Ligation: A Boon to Peptide Chemistry. *Molecules* **2014**, *19* (9), 14461–14483.
35. Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L. Chemical Synthesis of Proteins Using Peptide Hydrazides as Thioester Surrogates. *Nat. Protoc.* **2013**, *8* (12), 2483–2495.
36. Wan, Q.; Danishefsky, S. J. Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem., Int. Ed.* **2007**, *46* (48), 9248–9252.
37. Saxon, E.; Armstrong, J. I.; Bertozzi, C. R. A “traceless” Staudinger Ligation for the Chemoselective Synthesis of Amide Bonds. *Org. Lett.* **2000**, *2* (14), 2141–2143.
38. Tam, J. P.; Xu, J.; Eom, K. D. Methods and Strategies of Peptide Ligation. *Biopolymers - Peptide Science Section*. 2001, pp 194–205.
39. Coin, I.; Beyermann, M.; Bienert, M. Solid-Phase Peptide Synthesis: From Standard Procedures to the Synthesis of Difficult Sequences. *Nat. Protoc.* **2007**, *2* (12), 3247–3256.
40. Mapelli, C.; Natarajan, S. I.; Meyer, J.-P.; Bastos, M. M.; Bernatowicz, M. S.; Lee, V. G.; Pluscec, J.; Riexinger, D. J.; Sieber-McMaster, E. S.; Constantine, K. L.; Smith-Monroy, C. a; Golla, R.; Ma, Z.; Longhi, D. a; Shi, D.; Xin, L.; Taylor, J. R.; Koplowitz, B.; Chi, C. L.; Khanna, A.; Robinson, G. W.; Seethala, R.; Antal-Zimanyi, I. a; Stoffel, R. H.; Han, S.; Whaley, J. M.; Huang, C. S.; Krupinski, J.; Ewing, W. R. Eleven Amino Acid Glucagon-like

Peptide-1 Receptor Agonists with Antidiabetic Activity. *J. Med. Chem.* **2009**, *52* (23), 7788–7799.
41. Zhang, C.; Dai, P.; Spokoyny, A. M.; Pentelute, B. L. Enzyme-Catalyzed Macrocyclization of Long Unprotected Peptides. *Org. Lett.* **2014**, *16* (14), 3652–3655.
42. Atanasijevic, T.; Shusteff, M.; Fam, P.; Jasanoff, A. Calcium-Sensitive MRI Contrast Agents Based on Superparamagnetic Iron Oxide Nanoparticles and Calmodulin. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (40), 14707–14712.
43. Behrens, C. R.; Ha, E. H.; Chinn, L. L.; Bowers, S.; Probst, G.; Fitch-Bruhns, M.; Monteon, J.; Valdiosera, A.; Bermudez, A.; Liao-Chan, S.; Wong, T.; Melnick, J.; Theunissen, J. W.; Flory, M. R.; Houser, D.; Venstrom, K.; Levashova, Z.; Sauer, P.; Migone, T. S.; Van Der Horst, E. H.; Halcomb, R. L.; Jackson, D. Y. Antibody-Drug Conjugates (ADCs) Derived from Interchain Cysteine Cross-Linking Demonstrate Improved Homogeneity and Other Pharmacological Properties over Conventional Heterogeneous ADCs. *Mol. Pharm.* **2015**, *12* (11), 3986–3998.
44. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective “Ligation” of Azides and Terminal Alkynes. *Angew. Chem., Int. Ed.* **2002**, *41* (14), 2596–2599.
45. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. Copper-Free Click Chemistry for Dynamic in Vivo Imaging. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (43), 16793–16797.
46. Presolski, S. I.; Hong, V. P.; Finn, M. G. Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. In *Current Protocols in Chemical Biology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2011; Vol. 3, pp 153–162.
47. Dorman, G.; Prestwich, G. D. Using Photolabile Ligands in Drug Discovery and Development. *Trends Biotechnol.* **2000**, *18* (2), 64–77.
48. Kramer, K.; Sachsenberg, T.; Beckmann, B. M.; Qamar, S.; Boon, K.-L.; Hentze, M. W.; Kohlbacher, O.; Urlaub, H. Photo-Cross-Linking and High-Resolution Mass Spectrometry for Assignment of RNA-Binding Sites in RNA-Binding Proteins. *Nat. Methods* **2014**, *11* (10), 1064–1070.
49. Gomes, A. F.; Gozzo, F. C. Chemical Cross-Linking with a Diazirine Photoactivatable Cross-Linker Investigated by MALDI- and ESI-MS/MS. *J. Mass Spectrom.* **2010**, *45* (8), 892–899.
50. Isidro, A.; Latassa, D.; Giraud, M.; Alvarez, M.; Albericio, F. 1,2-Dimethylindole-3-Sulfonyl (MIS) as Protecting Group for the Side Chain of Arginine. *Org. Biomol. Chem.* **2009**, *7* (12), 2565–2569.
51. Lan, M. The Role of the Proline-Arginine Dipeptide Repeat in Amyotrophic Lateral Sclerosis. M.S. Dissertation, University of Pennsylvania, Philadelphia, PA, USA, 2017.
52. Batjargal, S.; Walters, C. R.; Petersson, E. J. Inteins as Traceless Purification Tags for Unnatural Amino Acid Proteins. *J. Am. Chem. Soc.* **2015**, *137* (5), 1734–1737.
53. Liu, C. C.; Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu.*

Rev. Biochem. **2010**, *79* (1), 413–444.
54. Luscombe, N. M.; Laskowski, R. a; Thornton, J. M. Amino Acid-Base Interactions: A Three-Dimensional Analysis of Protein-DNA Interactions at an Atomic Level. *Nucleic Acids Res.* **2001**, *29* (13), 2860–2874.
55. Fosgerau, K.; Hoffmann, T. Peptide Therapeutics: Current Status and Future Directions. *Drug Discov. Today* **2015**, *20* (1), 122–128.
56. Hui, H.; Farilla, L.; Merkel, P.; Perfetti, R. The Short Half-Life of Glucagon-like Peptide-1 in Plasma Does Not Reflect Its Long-Lasting Beneficial Effects. *Eur. J. Endocrinol.* **2002**, *146* (6), 863–869.
57. Weinstock, M. T.; Francis, J. N.; Redman, J. S.; Kay, M. S. Protease-Resistant Peptide Design-Empowering Nature's Fragile Warriors against HIV. *Biopolymers* **2012**, *98* (5), 431–442.
58. Xu, P.; Xu, M.; Jiang, L.; Yang, Q.; Luo, Z.; Dauter, Z.; Huang, M.; Andreasen, P. A. Design of Specific Serine Protease Inhibitors Based on a Versatile Peptide Scaffold: Conversion of a Urokinase Inhibitor to a Plasma Kallikrein Inhibitor. *J. Med. Chem.* **2015**, *58* (22), 8868–8876.
59. Sifferlen, T.; Rueping, M.; Gademann, K.; Jaun, B.; Seebach, D. β-Thiopeptides: Synthesis, NMR Solution Structure, CD Spectra, and Photochemistry. *Helv. Chim. Acta* **1999**, *82* (12), 2067–2093.
60. Petersson, E. J.; Goldberg, J. M.; Wissner, R. F. On the Use of Thioamides as Fluorescence Quenching Probes for Tracking Protein Folding and Stability. *Phys. Chem. Chem. Phys.* **2014**, *16*, 6827–6837.
61. Goldberg, J. M.; Chen, X.; Meinhardt, N.; Greenbaum, D. C.; Petersson, E. J. Thioamide-Based Fluorescent Protease Sensors. *J. Am. Chem. Soc.* **2014**, *136* (5), 2086–2093.
62. Wissner, R. F.; Batjargal, S.; Fadzen, C. M.; Petersson, E. J. Labeling Proteins with Fluorophore/Thioamide Förster Resonant Energy Transfer Pairs by Combining Unnatural Amino Acid Mutagenesis and Native Chemical Ligation. *J. Am. Chem. Soc.* **2013**, *135* (17), 6529–6540.
63. Turk, B. Targeting Proteases: Successes, Failures and Future Prospects. *Nat. Rev. Drug Discov.* **2006**, *5* (9), 785–799.
64. Harper, E.; Berger, A. On the Size of the Active Site in Proteases: Pronase. *Biochem. Biophys. Res. Commun.* **1972**, *46* (5), 1956–1960.
65. López-Otín, C.; Bond, J. S. Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **2008**, *283* (45), 30433–30437.
66. 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.
67. Maziak, L.; Lajoie, G.; Belleau, B. Productive Conformation in the Bound State and Hydrolytic Behavior of Thiopeptide Analogs of Angiotensin-Converting Enzyme Substrates. *J. Am. Chem. Soc.* **1986**, *108* (1), 182–183.
68. Mock, W. L.; Chen, J. T.; Tsang, J. W. Hydrolysis of a Thiopeptide by Cadmium Carboxypeptidase A. *Biochem. Biophys. Res. Commun.* **1981**, *102* (1), 389–396.
69. Bond, M. D.; Holmquist, B.; Vallee, B. L. Thioamide Substrate Probes of Metal-Substrate Interactions in Carboxypeptidase a Catalysis. *J. Inorg.*

Biochem. **1986**, *28* (2–3), 97–105.
70. Bartlett, P. a; Spear, K. L.; Jacobsen, N. E. A Thioamide Substrate of Carboxypeptidase A. *Biochemistry* **1982**, *21* (7), 1608–1611.
71. Nashedt, N. T. Carboxypeptidase A Catalyzed Hydrolysis of Thiopeptide and Thionester Analogues of Specific Substrates. An Effect on. *Society* **1982**, No. 22, 5221–5226.
72. Yao, S.; Zutshi, R.; Chmielewski, J. Endothiopeptide Inhibitors of HIV-1 Protease. *Bioorg. Med. Chem. Lett.* **1998**, *8* (6), 699–704.
73. McElroy, J.; Guthrie, D. J.; Hooper, N. M.; Williams, C. H. Gly-(CSNH)-Phe Resists Hydrolysis by Membrane Dipeptidase. *Biochem. Soc. Trans.* **1998**, *26* (1), S31.
74. Foje, K. L.; Hanzlik, R. P. Peptidyl Thioamides as Substrates and Inhibitors of Papain, and as Probes of the Kinetic Significance of the Oxyanion Hole. *Biochim. Biophys. Acta - Gen. Subj.* **1994**, *1201* (3), 447–453.
75. Cho, K. Synthesis of Fluorescent Peptidyl Thioneamides and the Assay of Papain in the Presence of Trypsin. *Anal. Biochem.* **1987**, *164* (1), 248–253.
76. Beattie, R. E.; Elmore, D. T.; Williams, C. H.; Guthrie, D. J. The Behaviour of Leucine Aminopeptidase towards Thionopeptides. *Biochem. J.* **1987**, *245* (1), 285–288.
77. Schutkowski, M.; Jakob, M.; Landgraf, G.; Born, I.; Neubert, K.; Fischer, G. Probing Substrate Backbone Function in Prolyl Oligopeptidase Catalysis. Large Positional Effects of Peptide Bond Monothioxylation. *Eur. J. Biochem.* **1997**, *245*, 381–385.
78. Zhang, W.; Li, J.; Liu, L. W.; Wang, K. R.; Song, J. J.; Yan, J. X.; Li, Z. Y.; Zhang, B. Z.; Wang, R. A Novel Analog of Antimicrobial Peptide Polybia-MPI, with Thioamide Bond Substitution, Exhibits Increased Therapeutic Efficacy against Cancer and Diminished Toxicity in Mice. *Peptides* **2010**, *31* (10), 1832–1838.
79. Zacharie, B.; Lagraoui, M.; Dimarco, M.; Penney, C. L.; Gagnon, L. Thioamides: Synthesis, Stability, and Immunological Activities of Thioanalogues of Imreg. Preparation of New Thioacylating Agents Using Fluorobenzimidazolone Derivatives. *J. Med. Chem.* **1999**, *42* (11), 2046–2052.
80. Chen, X. Thioamides in Protease Studies: Many Applications for a Single-Atom Substitution. Ph.D. Dissertation, University of Pennsylvania, Philadelphia, PA, USA, 2017.
81. Mukherjee, S.; Verma, H.; Chatterjee, J. Efficient Site-Specific Incorporation of Thioamides into Peptides on a Solid Support. *Org. Lett.* **2015**, *17* (12), 3150–3153.
82. Roussel, A.; Mathieu, M.; Dobbs, A.; Luu, B.; Cambillau, C.; Kellenberger, C. Complexation of Two Proteic Insect Inhibitors to the Active Site of Chymotrypsin Suggests Decoupled Roles for Binding and Selectivity. *J. Biol. Chem.* **2001**, *276* (42), 38893–38898.
83. Ekici, Ö. 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.

84. Roskoski, R. Michaelis-Menten Kinetics. In *xPharm: The Comprehensive Pharmacology Reference*; 2011; pp 1–10.
85. Dunbar, K. L.; Mitchell, D. A. Insights into the Mechanism of Peptide Cyclodehydrations Achieved through the Chemoenzymatic Generation of Amide Derivatives. *J. Am. Chem. Soc.* **2013**, *135* (23), 8692–8701.

Appendices

Appendix A. pET His6 Sumo TEVCysPR20 Plasmid Map