Development of chemical and enzymatic peptide ligation strategies

Cao, Yuan

2015

Cao, Y. (2015). Development of chemical and enzymatic peptide ligation strategies. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/65380

https://doi.org/10.32657/10356/65380

Development of Chemical and Enzymatic Peptide Ligation Strategies

Cao Yuan

Supervisor: A/P Liu Chuanfa

School of Biological Sciences

A thesis submitted to the Nanyang Technological University in fulfillment of the requirement for the degree of Doctor of Philosophy

2015

To all the people that have helped me in completion of this thesis

First of all, I appreciate that my supervisor A/P Liu Chuanfa offers me a research chance in the lab and leads me into the research field of protein chemistry. He gives me a chance to experience the Ph.D research which is fulfilling, rewarding, unforgettable and once-in-life. During the past four years and half, I really thank him for the understanding, caring, mentorship and patience to me.

Second, I would like to thank Prof Tam for giving the collaboration chance to work on projects associated with their discovery – butelase 1. I am also sincerely grateful to Dr. Giang with whom I jointly performed the butelase 1-related work. During the collaboration, Dr. Giang selflessly shared with me the scientific knowledge and hands-on skills, patiently helped me analyze the problems I encountered and encouraged me when I got frustrated. To me, he is not just a collaborator in work but also a true friend I cherish in heart.

Next, I express my appreciation to all my previous and current lab members for their help and also for creating an open “knowledge-sharing” working environment. My special thanks would go to two of them, one is Dr Lv Xiaowei and the other is Ms. Rao Chang. Although busy with her own assignment, Dr. Lv Xiaowei voluntarily taught me the fundamental organic chemistry during my early Ph.D years. This was important to me as I did not have much research background as a fresh researcher and eagerly needed skills to survive the Ph.D program. I always feel a pity that I did not thank her in person while she was working in our lab. I own thanks to Ms Rao Chang for the selfless efforts she puts in maintaining the lab in good order. I benefitted a lot from this as I could really focus on my research work while not

getting distracted by other things. Her help also extends to me personally. I sincerely thank her for her help to solve my problems as her own and to adjust my mentality during the hardship.

My parents are the ones I would like to thank the most. I thank them for their true unconditional support all these years especially my toughest time. My understanding of family value has deepened a lot during my overseas study.

I want to thank my SBS friends as well for their friendship, encouragement and help. They include: Miss. Wen Ting, Mr. Zhen Jie and Ms. Wang Han.

At last with great sincerity, I want to thank Nanyang Technological University and Singapore government for providing us foreign students such a good study opportunity.

Table of contents

Abstract

Proteins are central macromolecules of life that carry out virtually all the functions in almost every aspect of living organisms. To elucidate their roles in biological process, easy and convenient access to the proteins of interest is necessary. Chemical peptide synthesis and DNA-based recombinant technologies are the two major direct accesses to peptides or proteins. However, these two methods also have several shortcomings that may limit their application scopes. Chemical synthesis has size limitation (usually < 60 aa) and DNA-based recombinant technology could not afford proteins with site-specific modifications. Peptide ligation strategies which enable the combined use of the chemical synthesis and DNA-based recombinant technology would be a robust tool in biological research and greatly expand the scope of applications.[1,2] The current ligation methods still have their own challenges, like low efficiency, lack of selectivity, operationally challenging etc. Efforts to improve them are worth pursuing. In my Ph.D research work, I mainly devoted to improvement on two popular ligation strategies: native chemical ligation and enzymatic ligation.

In chapter 1, an overview was given on DNA-based recombinant technology, chemical peptide synthesis and different ligation strategies.

In chapter 2, we presented a new method to prepare peptide thioesters by Fmoc chemistry. A new molecule 2,4-bis(mercaptomethyl)-thiazolidine (BMMT) for peptide thioester preparation was designed. It was synthesized and applied to peptide thioester

synthesis taking advantage of the N-S acyl transfer mechanism. Also a *cis*/*trans* issue associated with the nature of thiazolidine ring was investigated.

In chapter 3, we solved the reversibility problem of the butelase 1-mediated ligation by using thiodepsipeptides as substrate. Our method was found to significantly improve the enzymatic ligation efficiency and be superior to the other substrates. Both ubiquitin and GFP were efficiently labeled using this strategy.

In chapter 4, a new technology enabling protein thioester preparation was described. By using amino thioesters and butelase 1, model proteins were converted to their thioester form with good efficiency and little sequence alteration. By coupling our method and the native chemical ligation, we demonstrated that protein C-terminal labeling could be done. The current popular sortase A-mediated ligation has also been proved to be compatible with the above two strategies. Multiple ligations could be done combining the above three ligation strategies which was practically useful in protein total synthesis.

Abbreviations

aa	amino acid
Ac	acetyl
ACN	acetonitrile
BML	butelase 1-mediated ligation
Boc	tert-butoxycarbonyl
Bzl	benzyl
CBD	chitin-binding domain
Da	dalton
DCC	*N,N'*-dicyclohexylcarbodiimide
DCM	dichloromethane
DIC	*N,N'*-diisopropylcarbodimide
DIEA	*N,N*-diisopropylethylamine
DMAP	4-dimethylaminopyridine
DMF	dimethylformamide
DTT	dithiothreitol
E. coli	Escherichia coli
EDT	1,2-ethanedithiol

ESI-MS electrospray ionization mass spectrometry

EPL expressed protein ligation

Fmoc 9-Fluorenylmethoxycarbonyl

HF hydrofluoric acid

HOBt N-hydroxybenzotriazole

IPTG isopropyl β-D-1-thiogalactopyranoside

LC-MS liquid chromatography mass spectrometry

MALDI-TOF Matrix-assisted laser desorption/ionization-time of flight

MBHA 4-Methylbenzhydrylamine

MESNa sodium 2-sulfanylethanesulfonate

MPA 3-mercaptopropionic acid

MPAA 4-mercaptophenyl acetic acid

MW molecular weight

NCL native chemical ligation

Npys 5-nitro-2-pyridinesulfenyl

PCR polymerase chain reaction

Pybop benzotriazol-1-yloxytripyrrolidinophosphonium hexafluorophosphate

RP-HPLC reverse phase high performance liquid chromatography

rpm runs per minute

rt	room temperature
SPPS	solid phase peptide synthesis
Srt A	sortase A
TCEP	*tris*-(carboxyethyl)phosphine
TFA	trifluoroacetic acid
TFMSA	trifluoromethanesulfonic acid
TIS	triisopropyl silane
Tris	tris(hydroxymethyl)aminomethane
Trt	trityl
Ubi	ubiquitin

Chapter 1 Introduction

Proteins and peptides play crucial roles in living organisms. They carry out almost all the biochemical functions of the cell. These effects derive from their unique folded structures which are in turn defined by their unique amino acid sequences. Due to their important roles in the biological world, technologies to prepare, modify and engineer proteins are of great interest.

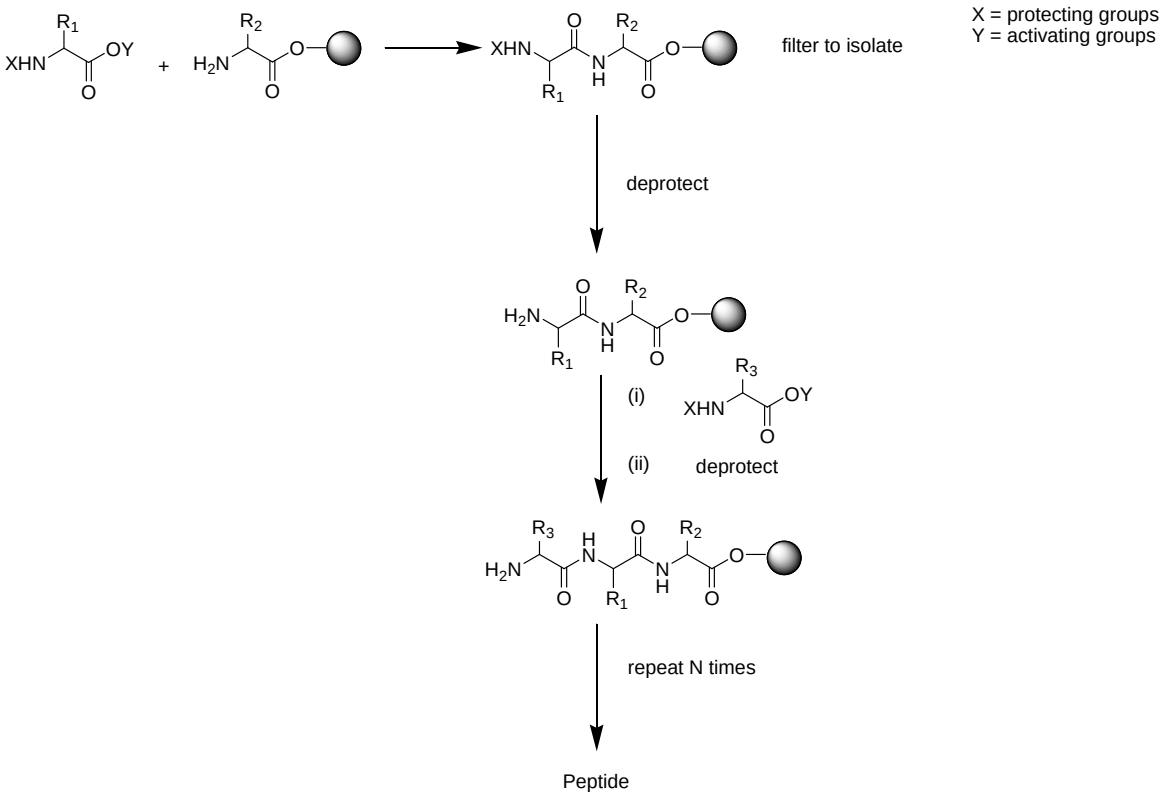
1.1 Direct access to proteins and peptides

To obtain the target proteins or peptides, there are mainly two ways: one is recombinant DNA-based technology and the other is chemical synthesis.

The recombinant DNA-based technology is the most classical tool to study the structure-function relationship of proteins in biological labs. The general idea of this method is to use molecular cloning technique to clone the gene of the target protein into a vector which is then introduced to host cells. The desired protein is expressed by taking advantage of the transcriptional and translational machinery of the host cell. The main drawbacks of this technology are: 1) Not all proteins are accessible in this way, for example the cytotoxic proteins. 2) It is difficult to obtain proteins with site-specific modifications using this method. It is noteworthy that unnatural amino acids can be incorporated into the target proteins by adopting amber stop codon/suppressor strategy. In this way site-specific modification could be made possible.[3-11] However due to the high technical barrier, this technology is still not widely adopted nowadays.

Due to the limitations mentioned above, protein chemical synthesis is now playing a complementary role in biological science.

Peptide synthesis used to be done in solution. However, this method is time and labor consuming thus not suitable for long chain peptide synthesis. Purification is necessary after every step of amino acid coupling. Only well-trained professional chemists can perform this job. In 1963, Dr Merrifield invented solid phase peptide synthesis (SPPS) which revolutionalized the field of peptide chemistry. The principle behind this method resides in the attachment of the first C-terminal amino acid to a solid polymer by a covalent bond, the addition of the succeeding amino acids one by one in a stepwise fashion until the whole sequence is assembled. The desired peptide can be easily cleaved from the solid support. It is much less arduous and more operationally convenient than the traditional solution phase synthesis in the way that the growing peptide chain is firmly attached to the insoluble solid support and the excessive amino acid and soluble by-products can be simply washed and filtered away **(Scheme 1-1)**. This method greatly simplifies the experimental manipulation, shortens the time of the practice and makes peptide synthesis widely accessible.[12]



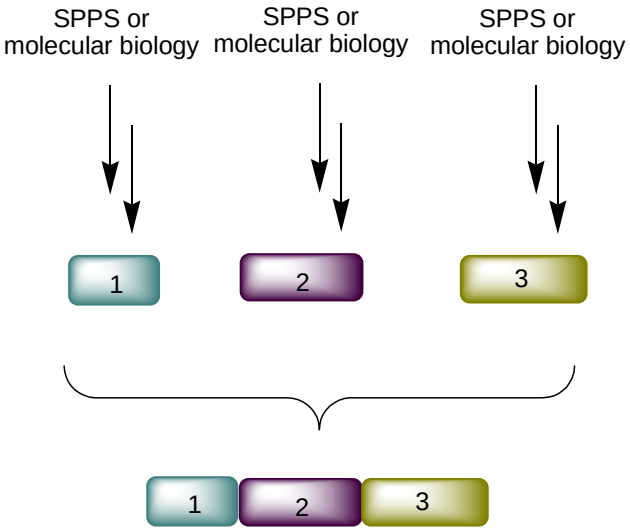
Scheme 1-1. General process of solid phase peptide synthesis.

This SPPS technology enables scientists to prepare various biopolymers with almost all kinds of sequences and site-specific chemical modifications.

Despite the versatility of SPPS, it has a size limitation (usually < 60 aa) on the peptide mainly due to interchain aggregation. One strategy to solve this aggregation problem is by using aggregation breaker during SPPS.[13-15] Usually, only peptides and some small-sized proteins could be prepared using SPPS but most proteins are still beyond its reach.

A combination of SPPS and molecular biology would be highly desirable since it could overcome each one's own limitations to meet the need of high-demanding biological

research nowadays. To achieve this goal, different peptide/protein ligation methods have been sought after by scientists. The benefit of this convergent strategy is that we could prepare different segments by the most suitable methods first and then ligate them together to afford the final full-length product **(Scheme 1-2)**.



Scheme 1-2. Convergent strategy to prepare proteins.

For ligation methods, there are mainly two popular categories: chemical ligation methods and enzyme-mediated ligation methods.

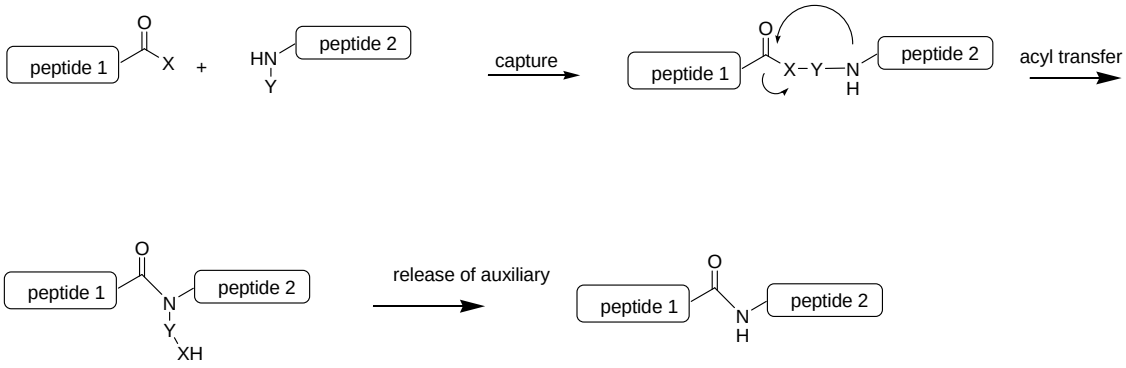
1.2 Peptide/protein ligation methods

1.2.1 Chemical peptide/protein ligation methods

In the beginning, peptide condensations are done in organic solvents with both reactants fully or partially protected. Ligation is done by activation of the C-terminus of one reactant with coupling reagent and subsequently coupling with the other peptide exposing the N-terminus.[16-18] This condensation strategy has mainly four downsides

that impede its wide application: 1 The activation process of the C-terminus of electrophile reactant might give rise to epimerization problem. 2 Fully protected peptides have poor solubility in organic solvents which would set limitations on the reaction concentration, leading to low reaction rate. 3 Usually an excessive amount of one peptide is necessary to drive the reaction to completion since the reactivity between an amino group and an activated carboxyl group is not high enough to compensate the effects brought by the low reactant concentration. 4 Since the condensation reaction is done with peptides protected, a global deprotection step needs to be done after the coupling which certainly compromises the yield of the desired product.

Due to the reasons stated above, orthogonal ligation strategy was developed. The major difference between orthogonal ligation and the classical ligation is the protection scheme. Early orthogonal ligation approaches did not afford native amide bonds at the junction sites, examples are hydrazone, oxime, thioether and thioester.[19-25] Despite their utility that is achieved in some practical applications, the nonnative bond formed after the ligation is always a problem and drives scientists to develop new methods that afford native amide bond. Some such chemical ligation technologies have been successfully developed. They typically comprise three steps: Initially, a pair of reactive nucleophile and electrophile brings the two unprotected peptide segments to proximity. Then an acyl transfer step would occur spontaneously within the molecule. At last, the capture moiety would be released to generate the amide bond **(Scheme 1-3)**.

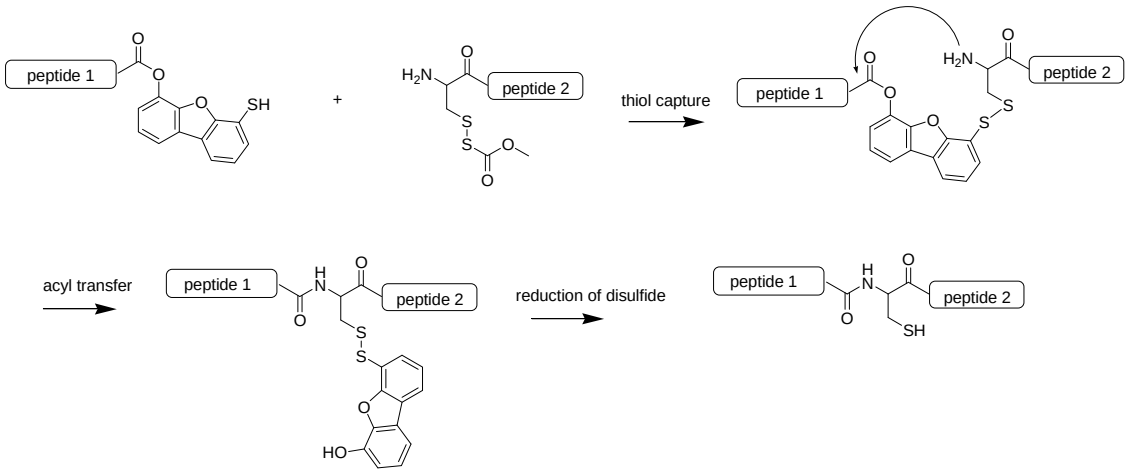


Scheme 1-3. Illustration of the amide bond-forming ligation strategies.

In the following paragraphs, three important chemoselective ligation methods utilizing this concept would be introduced.

1.2.1.1 Prior thiol capture ligation

Kemp's prior thiol capture strategy is the first demonstration of chemoselective ligations between unprotected peptide segments. In this strategy, a highly efficient thiol disulfide exchange reaction is used to bring the two segments into proximity. Then an acyl transfer would occur with assistance from a tricyclic auxiliary - 4-hydroxy-6-mercaptodibenzofuran. For the last step of auxiliary release, disulfide reduction by phosphine is used to afford the desired ligation product.[26,27]



Scheme 1-4. Prior thiol capture ligation developed by Kemp.

1.2.1.2 Thioester-mediated peptide ligation (native chemical ligation)

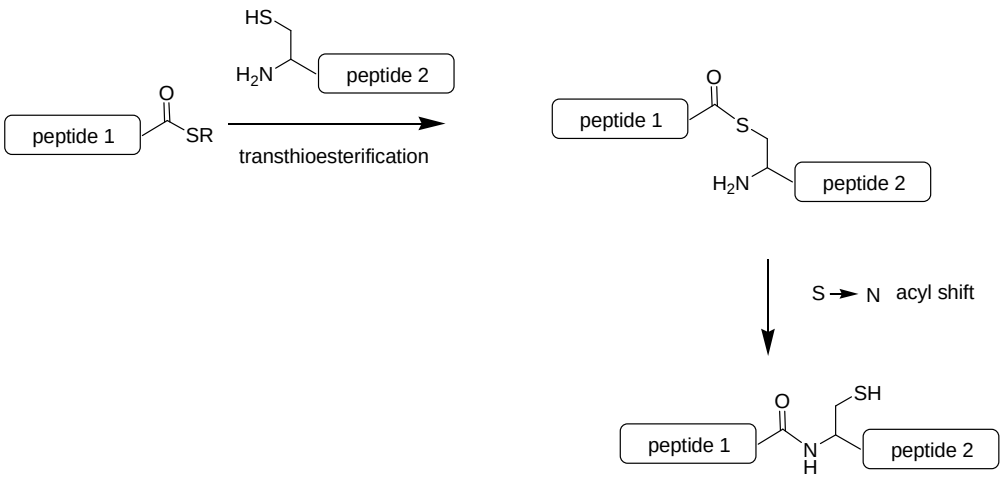
The thioester-cysteine ligation method first described by Wieland in 1953[28] and then by Kent, et al (1994)[29] and Tam, et al (1995)[30] is the most popular and widely adopted ligation method nowadays. This ligation method is better known as "Native chemical ligation" (NCL) as it bears the feature of traceless and generates native chemical bond at the ligation site. In this condensation reaction, an unprotected C-terminal peptide thioester reacts with another peptide containing an N-terminal cysteine residue to give a longer full-length peptide or protein with a native amide bond at the ligation site. The general principle behind this reaction is shown in **Scheme 1-5**, the first step is the chemoselective reaction of an unprotected peptide-alpha-thioester with another unprotected peptide with a cysteine residue at the amino–terminus, affording a thioester-linked intermediate. Then, the acyl group on the thioester moiety would spontaneously migrate from sulfur to the nitrogen via a five-member ring intermediate to form a peptide bond. In the end, a longer natural peptide product is formed.

Due to its robustness and usefulness, extensive investigation has been done into NCL. Standard Boc chemistry offers an enabling tool for preparation of normal peptide thioesters which makes this technology practically accessible.[31] The reaction condition for the NCL has also been well studied. It was found that mild basic condition is suitable for this reaction and Tris(2-carboxylethyl)phosphine (TCEP) could be used to prevent the thiol groups oxidizing into the nonreactive disulfide form. Normally in order to obtain a more reactive thioester, a large amount of thiol additives are added. Different thiol additives have been examined for this reaction. Kent's group reported that 4-mercaptophenyl acetic acid (MPAA) would be an ideal candidate mainly because it is well soluble in basic condition.[32]

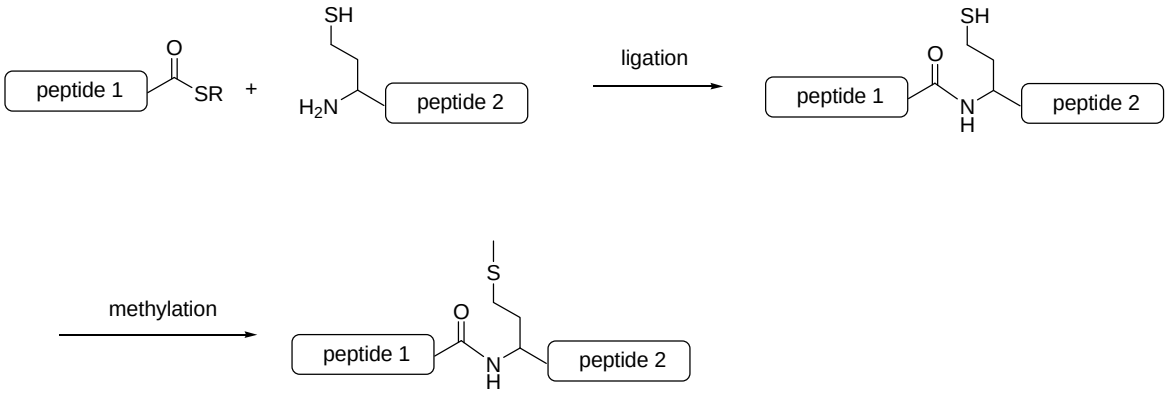
There are two major issues associated with NCL technology which are related to the two reactants of the reaction. The first issue is that NCL requires a cysteine residue at the ligation site which is actually one of the rare occurrences in nature proteins. Different strategies have been developed to expand this technology to other amino acids. Tam's group used homocysteine as a replacement of normal cysteine in NCL.[33] After ligation, methylation of the thiol group on the homocysteine could afford methonine at the ligation site **(Scheme 1-6)**. Nα-auxiliary mediated strategies are also developed.[34-38] In this technology, an auxiliary comprising a thiol group was installed at the N-terminal residue to facilitate the ligation reaction. After the ligaiton is done, removal of the auxiliary gives a natural peptide junction. This strategy has only been applied to residues that are less hindered because the existence of the auxiliary poses great sterical hinderance to the N-terminal amine. Combination of different desulfurization chemical reactions with different unnatural amino acids provides another method to expand the NCL's application scope **(Scheme 1-7)**.[39-47] A thiol

auxiliary group situated at the side chain of an N-terminal amino acid could be used to facilitate the NCL. After ligation, desulfurization could be performed to remove the thiol group, affording the natural amino acids at the ligation site **(Scheme 1-8)**. Unlike the above Nα-auxiliary mediated strategy, this strategy brings little steric hinderance problem as the thiol group is quite small. Ligations at Ala, Phe, Thr, Val, Leu and Lys have been actualized by this strategy.

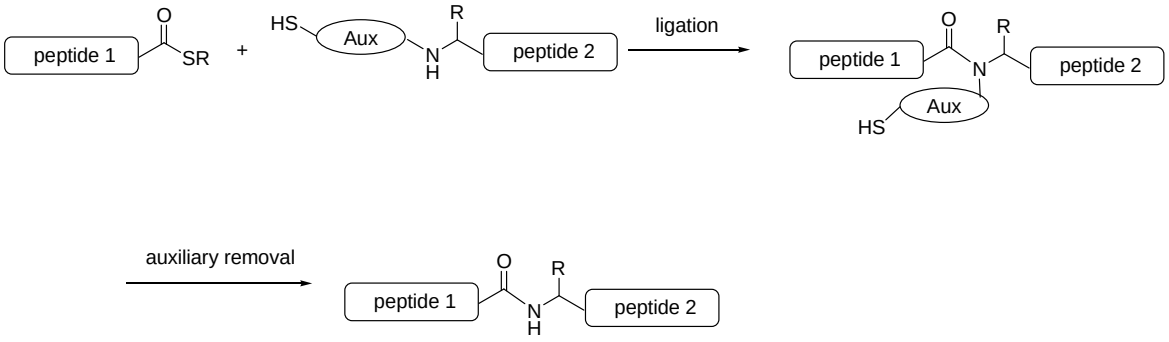
Another issue comes from the peptide thioester part. Although normal peptide thioester could be prepared by standard Boc chemistry, several important posttranslational modifications are not compatible with Boc chemistry. Development to solve this problem would be depicted in chapter 2. This issue is also one of my research aims during Ph.D.



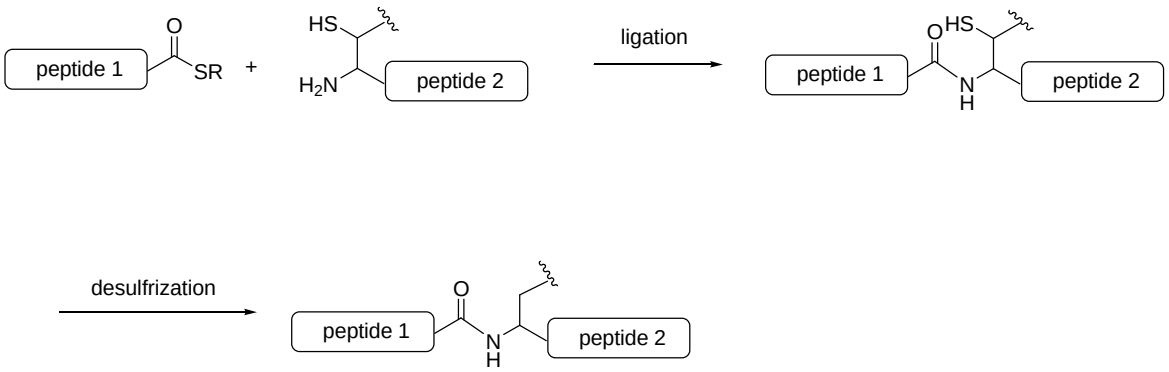
Scheme 1-5. General principle of native chemical ligation.



Scheme 1-6. General scheme for NCL at methonine.



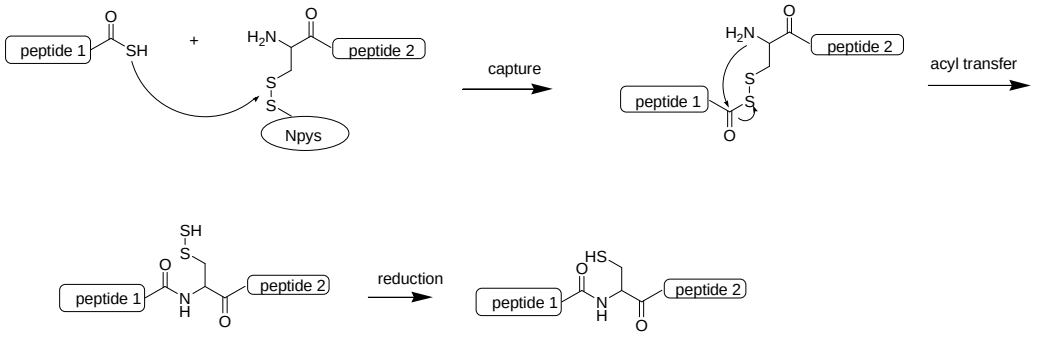
Scheme 1-7. General scheme for Nα-auxiliary mediated ligation.



Scheme 1-8. General scheme for ligation/desulfurization strategy.

1.2.1.3 Thioacid capture ligation

Another chemoselective ligation method worth mentioning is the thioacid capture ligation developed by Tam, et al.[48] In this ligation method, the thiol group at the N-terminus is first activated by 5-nitropyridyl-2-sulfenyl (Npys) group and this highly electrophile mercaptan group would be attacked by a highly nucleophilic thioacid group at the C-terminus of the other reactant. This disulfide intermediate would undergo an intramolecular acyl transfer via a six-member ring to give an amide bond at the junction site. The disulfide bond is reduced at the last step, giving the final desired product.



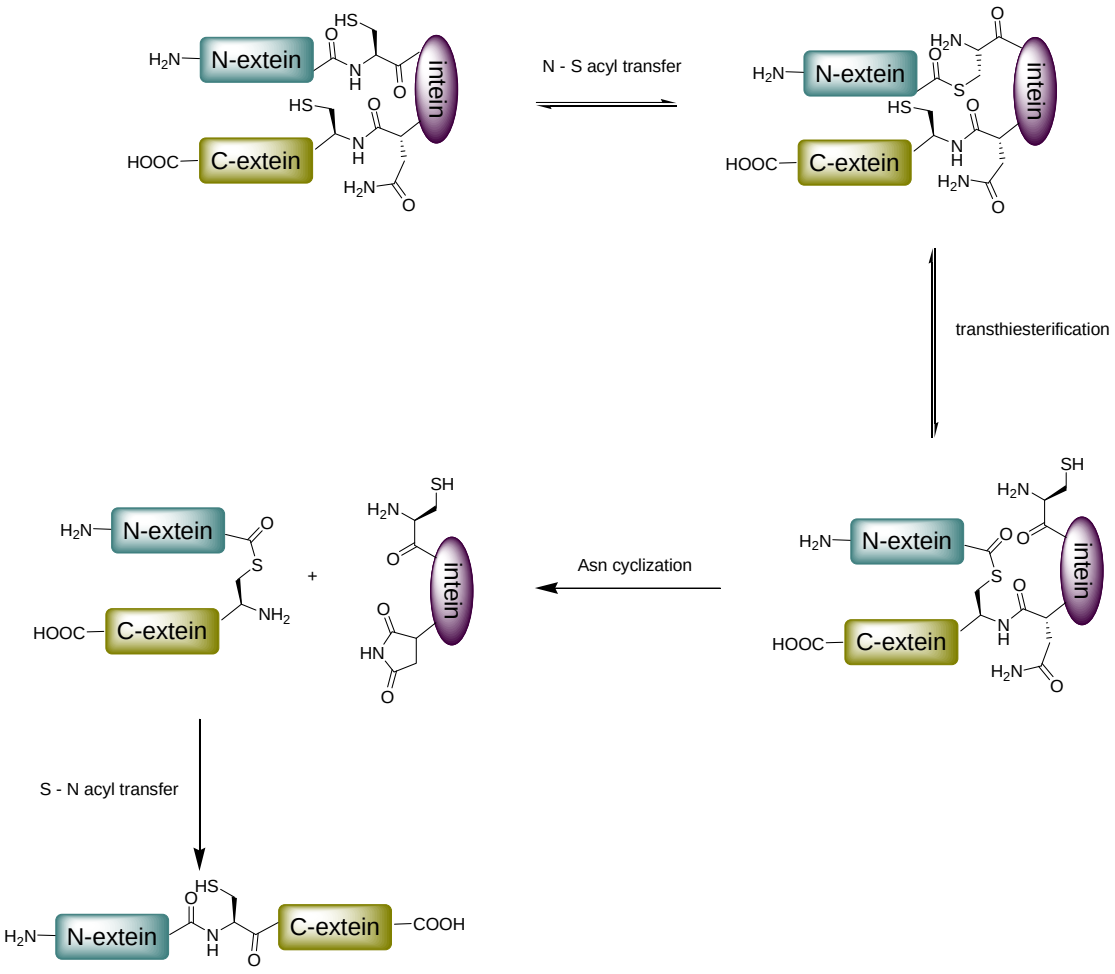
Scheme 1-9. Scheme of thioacid-capture ligation.

1.2.2 Enzymatic method to ligate peptides/proteins

Besides chemical ligation methods, enzymes are also employed to achieve the same goal. Enzymatic methods for peptide/protein ligation require minimal protection of peptide/protein segments and also give great specificity of the coupling reaction. In the following few paragraphs, three important enzymes used for peptide/protein ligation: intein, Sortase A, butelase 1 would be introduced.

1.2.2.1 Intein

Inteins are a group of naturally occurring self-splicing proteins that mediate amide bond formation between the two flanking protein fragements, namely N-extein and C-extein.[49-68] The names of intein and extein are given by analogy to intron and extron in nucleic acid. This process is done through a series of acyl shifts: 1 The N-terminal cysteine residue of intein would undergo a self-catalyzed N to S acyl transfer mainly through amide distortion, affording a thioester junction at the N-intein splitting site. 2 Another cysteine residue at the C-terminus of intein would have a thiol-thioester exchange through nucleophilic attack, leading to a branched intermediate. 3 The intein would then excise itself from the two exteins by cyclization of asparagine (or glutamate) located at its C-terminus. In this way, two exteins are linked together through a thioester bond. 4 Finally, an S-N acyl transfer would happen chemically spontaneously to form a natural amide bond in a NCL-like way **(Scheme 1-10)**.[69-71]

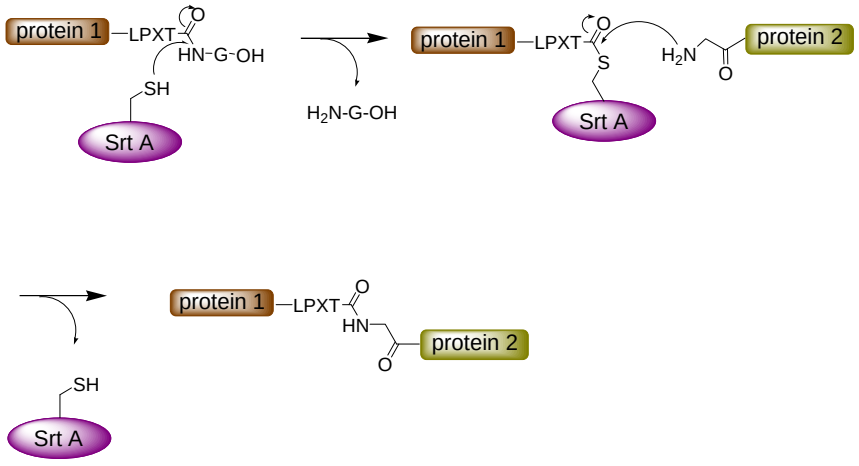


Scheme 1-10. Intein-mediated protein ligaiton.

1.2.2.2 Sortase A

Sortase A (Srt A), produced by gram-positive bacteria *Staphylococcus areus*, is particularly popular in various biochemical applications these years. In nature, this ligase mediates the reversible anchoring of virulence factors to the cell wall. This transpepidation reaction known as “sortagging” begins with the nucleophilic attack on a LPXTG recognition motif of the target protein by the cysteine residue at the Srt A catalytic site. The amide bond between threonine and glycine would be cleaved to form

an acyl-enzyme intermediate. Another peptide/protein comprising an N-terminal glycine would then react with the intermediate to covalently attach itself to the target protein while releasing the enzyme at the same time **(Scheme 1-11)**.[72-79] Many applications have been actualized by Srt A. They include: engineering of bacterial surfaces, C-terminal labeling, N-terminal labeling, labeling at N and C termini, post-translational modification, protein expression and purification etc.[80-89]

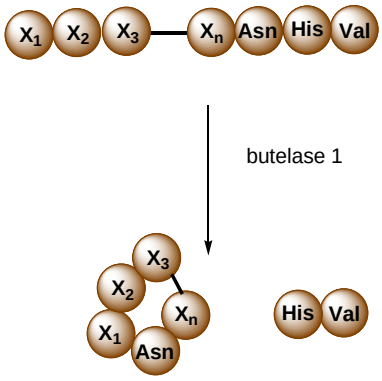


Scheme 1-11. Mechanism of Srt A-mediated ligation.

1.2.2.3 Butelase 1

Recently, Tam's group isolated an Asn/Asp (Asx) specific peptide cyclase/ligase from a cyclotide-producing plant *Clitoria ternatea*.[90] This enzyme, named as butelase 1 according to the local name of this plant, cyclizes the backbone of a group of circular defense peptides in the plant. In its natural bioprocess, butelase 1 recognizes a –N(D)HV motif at the C-terminus of a peptide and cleaves the Asn/Asp(Asx)-His bond to conjugate with an incoming N-terminal amino group. In the end, an Asx-X_1 bond is formed at the ligation junction **(Scheme 1-12)**, leaving a dipeptide –HV as the side

product. Its discovery has drawn special attentions mainly for two reasons: 1 It displays the highest catalytic efficiency (K_{cat} values up to 17 s^{-1} and catalytic efficiencies as high as 542000 $M^{-1}s^{-1}$) among all the ligases discovered so far, including Srt A. 2 Since it shows good tolerance to the N-terminal Xaa residue of the acyl acceptor, butelase 1 can be nearly qualified as a traceless ligase. Due to its merits, we believe this enzyme harbors great potentials in biochemical research.



Scheme 1-12. The natural bioprocess of butelase 1-mediated peptide cyclization.

1.3 Aims of our study

Native chemical ligation strategy and enzymatic ligation strategy are two widely used ones for protein semisynthesis and total synthesis. Despite their versatility, they have their own associated challenges. For native chemical ligation strategy, the building block peptide thioester is not convenient to obtain. For enzymatic ligation technology, the reaction's reversibility is a problem which limits the ligation yield. In my Ph.D work, I aim to provide solutions to the above problems.

My first objective is to develop a method to prepare peptide thioester by Fmoc chemistry. Although traditionally peptide thioeseters could be prepared by Boc chemistry, the harsh cleavage conditions are not compatible with some important posttranslational modifications like glycosylation and phosphorylation. We set to develop a strategy to synthesize peptide thioester using the milder Fmoc chemistry. This part of work is described in chapter 2.

Site specifically introducing functional groups into proteins of interest is crucial to elucidate the structure-function relationship. The second objective of my Ph.D work is to develop a method for site-specific modification of proteins which could afford homogeneous samples for research purpose. Various chemical ligation strategies have been reported but many of them suffer from nonspecificity and operational difficulty. Since enzyme has the feature of site-specificity, we aim to take advantage of it to achieve our goal. This part of work is presented in the third chapter of this thesis.

Chemical method are only limited to peptide thioesters or small-sized protein thioesters. For average-sized protein thioesters, intein-mediated technology is so far the only direct access. My third objective is to develop a method to could prepare protein thioesters by using enzymes as a complement to the intein-mediated technology. This work is shown in the forth chapter of this thesis.

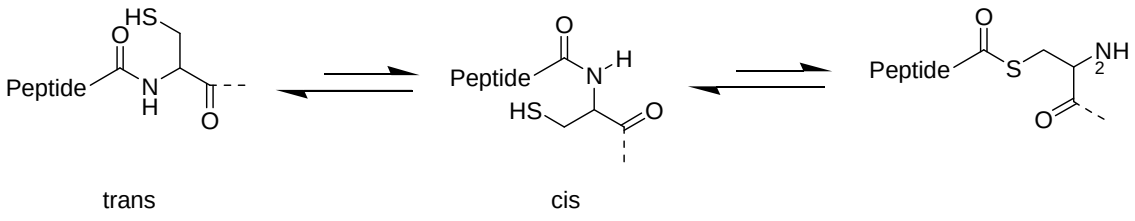
Chapter 2: A novel chemical method to prepare peptide/protein thioesters by N – S acyl shift mechanism

2.1 Introdution

Among all the chemical ligation strategies developed so far, the native chemical ligation (NCL) is the most widely used one. This popular method requires two segments, one is a cysteinyl peptide and the other is a peptidyl thioester. Since the cysteinyl peptide can be prepared by standard SPPS. The success of this method's application lies in the preparation of the other reactant which is the peptidyl thioester.

Typically, peptidyl thioesters can be directly prepared by Boc chemistry.[31] This method requires a strong acid like HF for final cleavage. A special apparatus is needed for this treatment which is a deterring element to many laboratories. Also, the strong acidic condition is not compatible with some post-translational modifications, like glycosylation and phosphorylation.[91] The milder Fmoc chemistry is thus preferable. But the frequent use of a basic reagent piperidine for Fmoc removal leads to the decomposition of thioester moiety. There has been increasing interest in developing new methods to avoid this problem. Some methods focus on the choice of the deprotecting reagent for Fmoc removal.[92-94] The undesired cleavage of the peptide thioester can be reduced when less nucleophilic bases are used. Some of the systems developed in this way do remove the Fmoc protecting group with less ability to cleave the thioester linkage. However, the aminolysis is always a problem. This aminolytic cleavage can be avoided if the thioester moiety is introduced after assembly of the

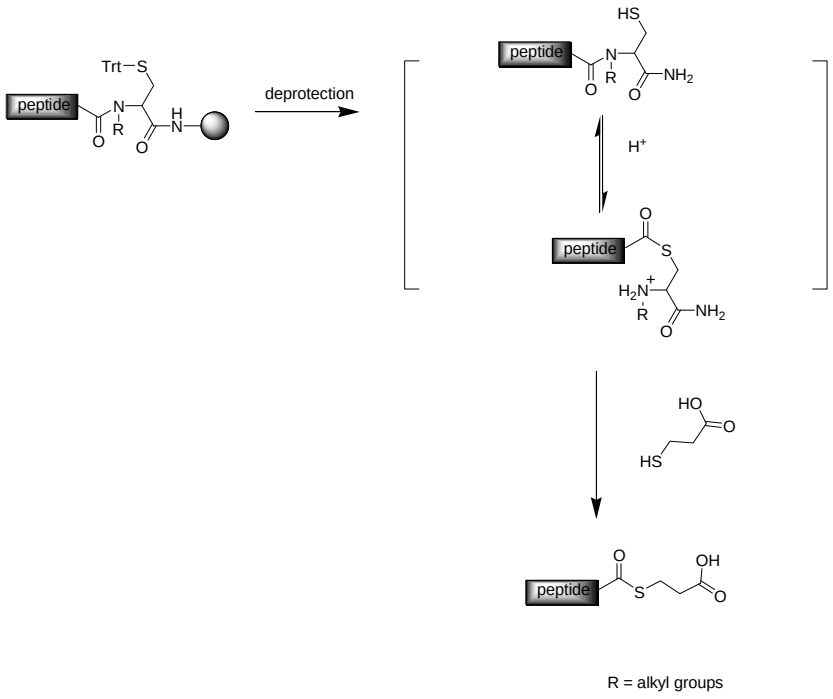
whole peptide. This method relies on converting a fully protected peptide into its thioester form in solution after liberation from resins. Usually it is done by using thiols and activating reagents.[95-98] However this method is difficult to handle in practice because of the poor solubility of fully protected peptides. Recently, the interest of preparing thioesters has much shifted to the N-S acyl transfer mechanism which is the reverse process of the native chemical ligation.[99-113] In this approach, a peptide amide containing a neighboring protected thiol is first prepared by solid phase synthesis. After liberated from the resin, the thiol group serves as an internal nucleophile to attack the amide bond, affording a thioester bond. Normally this step is thermodynamically disfavored. Mechanistically, in order for the N-S acyl transfer to take place, the planar amide bond must be in the conformation where the thiol–bearing N-substituent is anti to the carbonyl oxygen, as shown in **Scheme 2-1**. This needs a *cis* form of the secondary amide which it rarely adopts because of lower stability and high energy barrier for *trans-cis* conformation conversion.[114] In biological organisms, this process is catalyzed by intein mainly through amide distortion. However in chemical practice, this process does not happen spontaneously and requires activation and conformational assistance. Some examples are presented in the following paragraphs to prepare peptidyl thioesters taking advantage of the N-S acyl transfer reaction.



Scheme 2-1. N-S acyl transfer mechanism.

2.1.1 N-Alkyl cysteine-assisted thioesterification of peptides

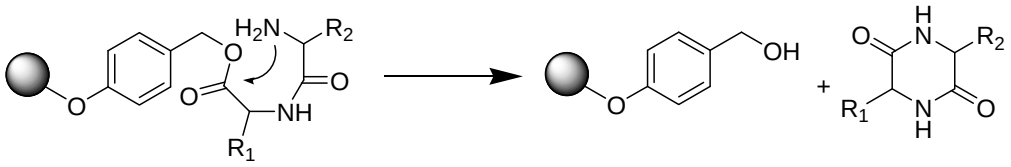
Yoshiaki Nakahara's group used N-alkyl cysteine derivatives as building block to prepare peptidyl thioesters. Presumably, N-alkylation would increase the ratio of the productive *cis* amide form to the non-productive *trans* amide form in the resulting tertiary amide, thereby making thioesterification of the amide bond easier. The peptide is first assembled on the secondary amine of the N-alkyl cysteine derivative using standard SPPS. After trityl protecting group removal in the cleavage step, the liberated thiol could attack the neighboring amide bond to generate a thioester moiety. However this process is reversible. The thioester form and amide form may co-exist in equilibrium. In the presence of an excess amount of an external thiol compound such as 3-mercaptopropionic acid (MPA), intermolecular thiol-thioester exchange could occur, driving the equilibrium towards the direction of the thioester formation. In the end, a stable peptide MPA thioester could be obtained.[101]



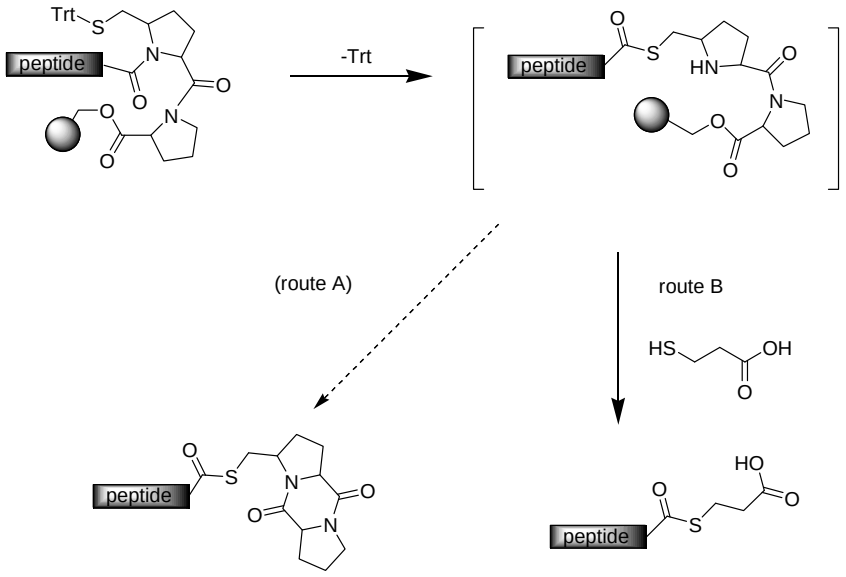
Scheme 2-2. N-Alkyl cysteine-assisted thioesterification of peptides.

2.1.2 Thioester preparation using a proline surrogate

In SPPS, it is a common case that when the last two residues at the C-terminus contain one or two proline(s), diketopiperazine will be formed if the C-terminal linkage is an unhindered ester **(Scheme 2-3)**. Inspired by this reaction, Yoshiaki Nakahara's group used a mercaptomethylated proline derivative to prepare peptide thioesters. The diketopiperazine reaction would "trap" the transiently exposed secondary amine during the N to S acyl transfer facilitated by the ring conformation. The formation of diketopiperazine on resin might be slower than the reverse S to N acyl migration, so excess amount of thiol (3-mercaptopropionic acid for example) was added to drive the thioesterification reaction in its forward direction **(Scheme 2-4)**. In this fashion, a stable peptidyl thioester could be obtained.[106]



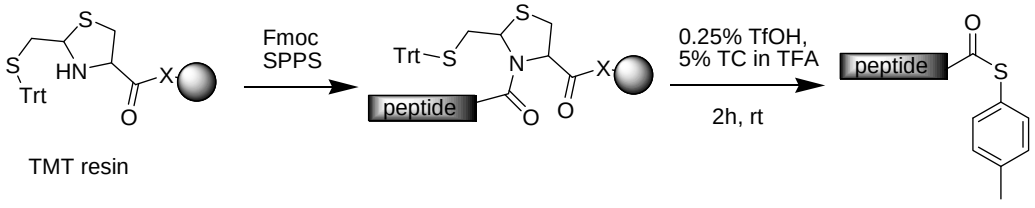
Scheme 2-3. Diketopiperazine formation during SPPS.



Scheme 2-4. The principle of the diketopiperazine-promoted thioester preparation.

2.1.3 Preparation of peptide thioesters by acyl transfer on thiazolidine

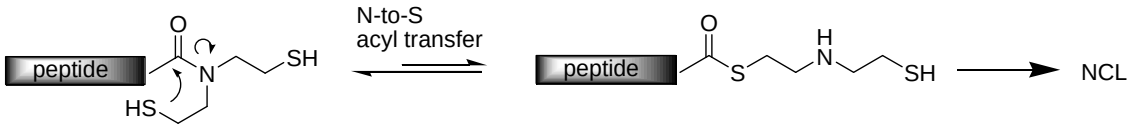
Tam's group reported a method that also employed the proline surrogate to prepare peptide thioesters. In their work, a thiazolidine derivative with a 2-thiomethyl group was used. Proximity-driven N-S acyl shift could occur due to the ring's conformational assistance **(Scheme 2-5)**. Also in the presence of strong acid, the carbonyl at the C-terminus of the peptide would be "trapped" through protonation, driving the amide-thioester equilibrium towards the direction of thioster formation. Thiol-thioester exchange with thiocresol would afford a stable peptide thioester.[111]



Scheme 2-5. Acid –catalyzed N-S acyl transfer and a thioester exchange reaction in tandem to prepare thioester. TC = thiocresol.

2.1.4 Peptidyl N,N-Bis(2-mercaptoethyl)-amide (BMEA) as thioester precursors

Our group and Melnyk's group brought in a "dithiol" concept at almost the same time.[114,115] This method cleverly obviates the "*cis-trans* conversion" issue: With two mercaptoethyl N-substitutions, the BMEA amide always has one thiol group that is anti to the carbonyl oxygen which, as mentioned above, is necessary for the intramolecular thiolysis reaction.



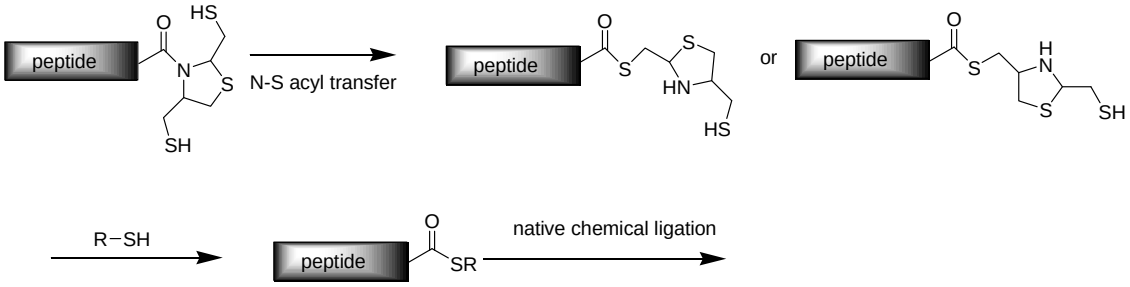
Scheme 2-6. N-to-S acyl transfer in a peptidyl N,N-Bis(2-mercaptoethyl)-amide.

2.2 Results and discussion

2.2.1 Design rationale

Inspired by the two methods developed by Tam's and our group, we designed a molecule which combined two elements favoring N-S acyl transfer, proximity-driven and "dithiol" concept. This molecule 2,4-bis(mercaptomethyl)-thiazolidine (BMMT) at the C-terminus of one peptide has two thiol groups at two sides of the thiazolidine ring. Both thiol groups gain the ring's conformational assistance and have a chance to attack the carbonyl. This would, in principle, double the chance of N-S acyl transfer.

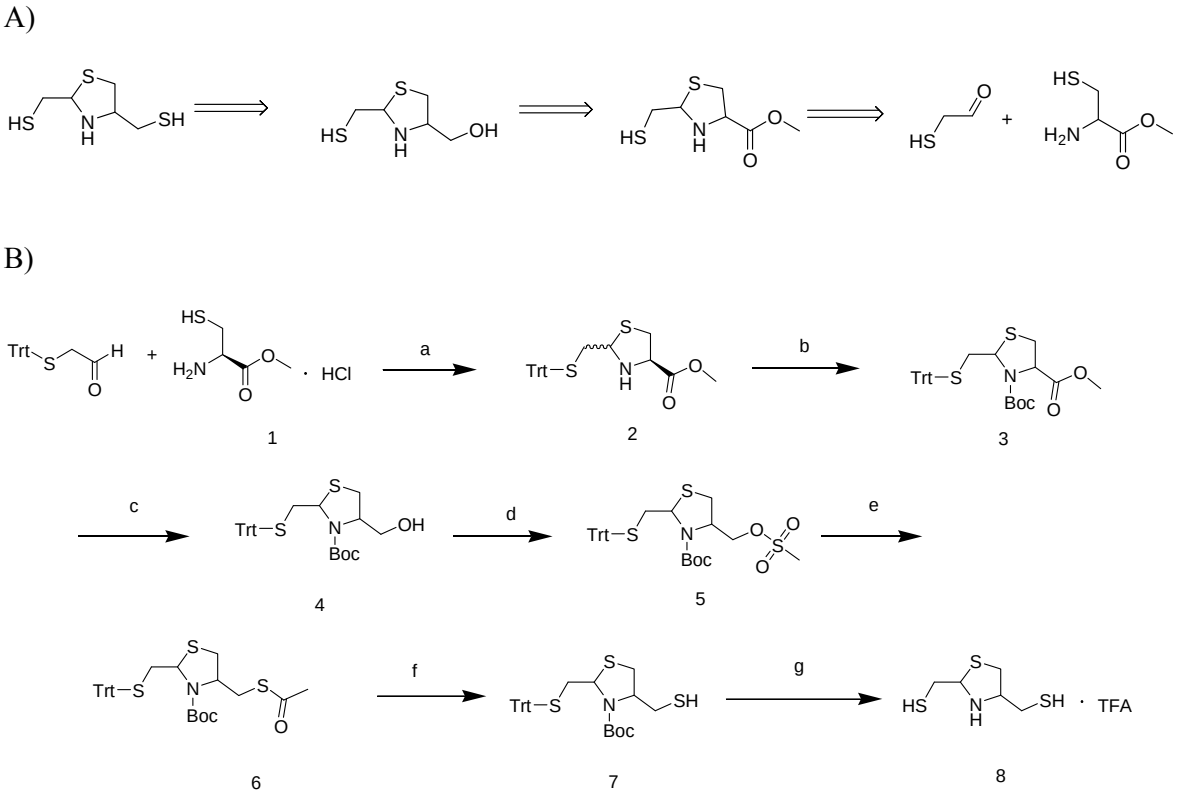
Our expectation is that the peptidyl thioester could be spontaneously formed under certain conditions and in the presence of an external thiol and when mixed with a cysteinyl peptide, the ligation reaction would occur in a one-pot manner **(Scheme 2-7)**.



Scheme 2-7. Proposed scheme for BMMT-based NCL.

2.2.2 Total synthesis of the linker BMMT

Since the molecule BMMT is not commercially available, it has to be synthesized. It is the most time-consuming and also challenging part of this project. According to the retrosynthetic route shown in **Scheme 2-8 A**, the synthesis can be mainly divided into three parts: 1 thiazolidine ring formation; 2 reduction to afford the hydroxyl group; 3 conversion of the hydroxyl group to the thiol group. The synthesis route is shown in **Scheme 2-8 B**.



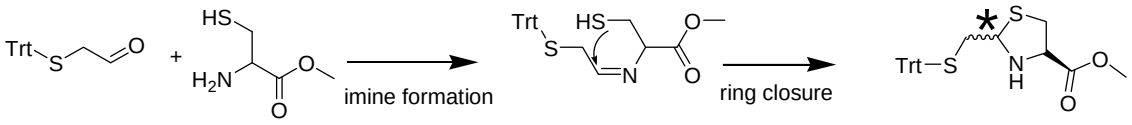
Scheme 2-8. BMMT synthesis.

A)Retrosynthetic stategy of BMMT. B) Total synthesis route of BMMT : (a) DIEA, MeOH; (b) Boc_2O, DIEA, DCM; (c) $LiAlH_4$, dry THF; (d) MsCl, DCM, TEA; (e) KSAc, dry DMF; (f) NaOH, MeOH; (g) 30% TFA, 1,2-Ethanedithiol.

2.2.2.1 Formation of the thiazolidine ring

The overall synthesis began with the reaction between tritylthioacetal aldehyde and H-Cys-OMe to generate the scaffold – thiazolidine ring. It is a very efficient and fast ring-formation reaction. The aldehyde initially condenses with the amine of cysteine to form an imine which is unstable in aqueous solution. Subsequently, the nucleophilic mercapto group on the β position of cysteine would attack the carbon of imine, forming

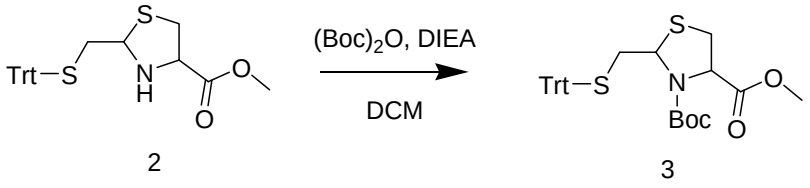
a stable five-member ring **(Scheme 2-9)**. After this step, a new chiral center is formed, creating a pair of *cis*/*trans* isomers.



Scheme 2-9. Mechanism of thiazolidine ring formation.

2.2.2.2 Protection of compound 2 with Boc protecting group

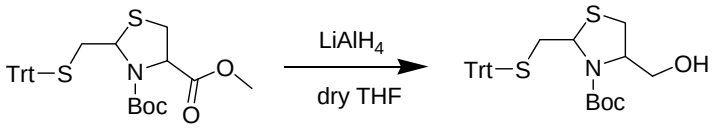
The secondary amine on compound **2** was a potential site of side reactions, so it had to be temporarily protected. Base-stable protecting group Boc instead of acid-stable protecting group Fmoc was chosen because sodium hydroxide was used in one of the later steps. The yield of this protecting reaction was very low. It might be attributed by two reasons. The first one is the electron-withdrawing effect of the methyl ester group. The second reason might be the bulky trityl group on the mercaptan group. The steric hindrance could make the adding of Boc difficult.



Scheme 2-10. Boc protection of thiazolidine ring.

2.2.2.3 Reduction of compound 3 to alcohol 4

In this step, the methyl ester group was reduced to hydroxyl group which could be used as the precursor of thiol group. Lithium aluminum hydride ($LiAlH_4$) was used for this purpose.

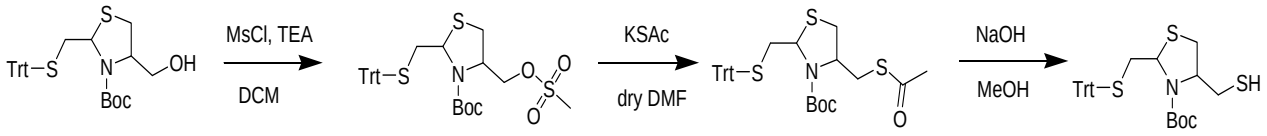


3 4

Scheme 2-11. Reduction of ester 3 to alcohol 4.

2.2.2.4 Conversion of alcohol 4 to thiol 7

The next three steps were to convert the hydroxyl group to thiol group. This conversion was done in three steps as direct conversion by Lawensson's reagent was not successful after many trials.

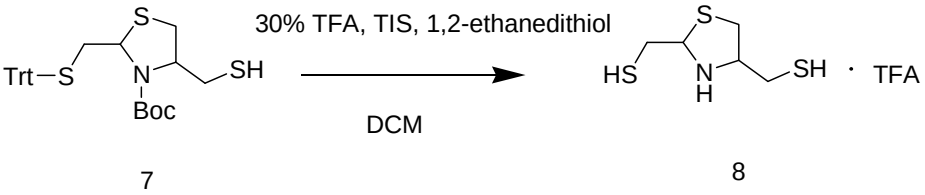


4 5 6 7

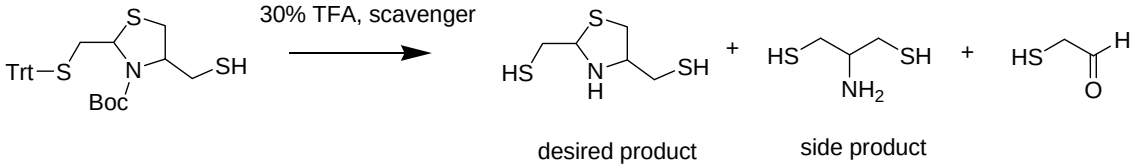
Scheme 2-12. Conversion of hydroxyl group to thiol group.

2.2.2.5 Liberation of desired BMMT linker from the protecting groups

The end of the synthesis was the global deprotection. Both acid-labile protecting groups Boc and trityl were kicked out using 30% TFA **(Scheme 2-13)**. It was also found that during this acid treatment, a ring-open side reaction occurred **(Scheme 2-14)**.



Scheme 2-13. Liberation of the desired product BMMT.



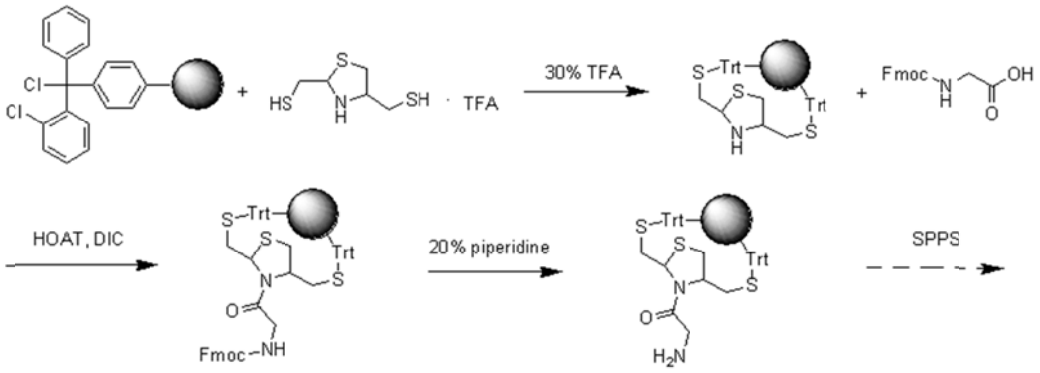
Scheme 2-14. Ring-open side reaction in the acid treatment.

2.2.3 Preparation of peptidyl-BMMT as the thioester precursor

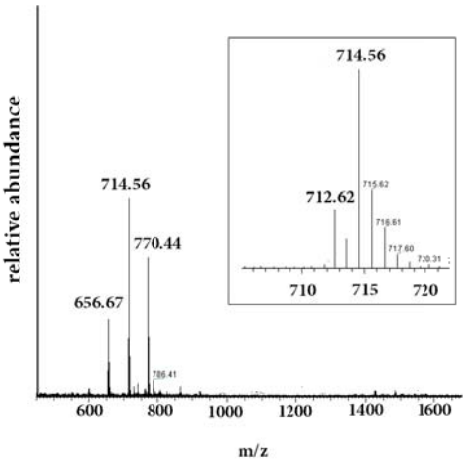
The BMMT linker was then loaded to 2-chlorotrityl chloride resin using the free thiol groups. A model peptide with the sequence LKSFG was assembled on the BMMT resin **(Figure 2-1 A)**. After cleavage from resin, besides the desired product, some other side protecting groups were also found. The ESI profile of the cleavage crude showed mainly four molecular weights: 714.56, 770.44, 656.67, 712.62 **(Figure 2-1 B)** with proportion of 6:5:3:2. **Figure 2-1 C** shows the structures corresponding to those molecular weights. The desired BMMT-peptide has the $[M + H]^+$ molecular ion of

714.56. The species with a m/z of 770.44 is its t-butyl alkylated adduct. This +56 adduct is the result of the alkylation reaction on one of the thiol groups which is a commonly encountered side reaction involving t-butyl and Boc protecting groups. In order to suppress this side reaction, a substantial amount of thiol scavenger like EDT (1,2-ethanedithiol) would be necessary. 656.67 and 712.62 are associated with instability of BMMT towards acid treatment.

A)



B)



C)

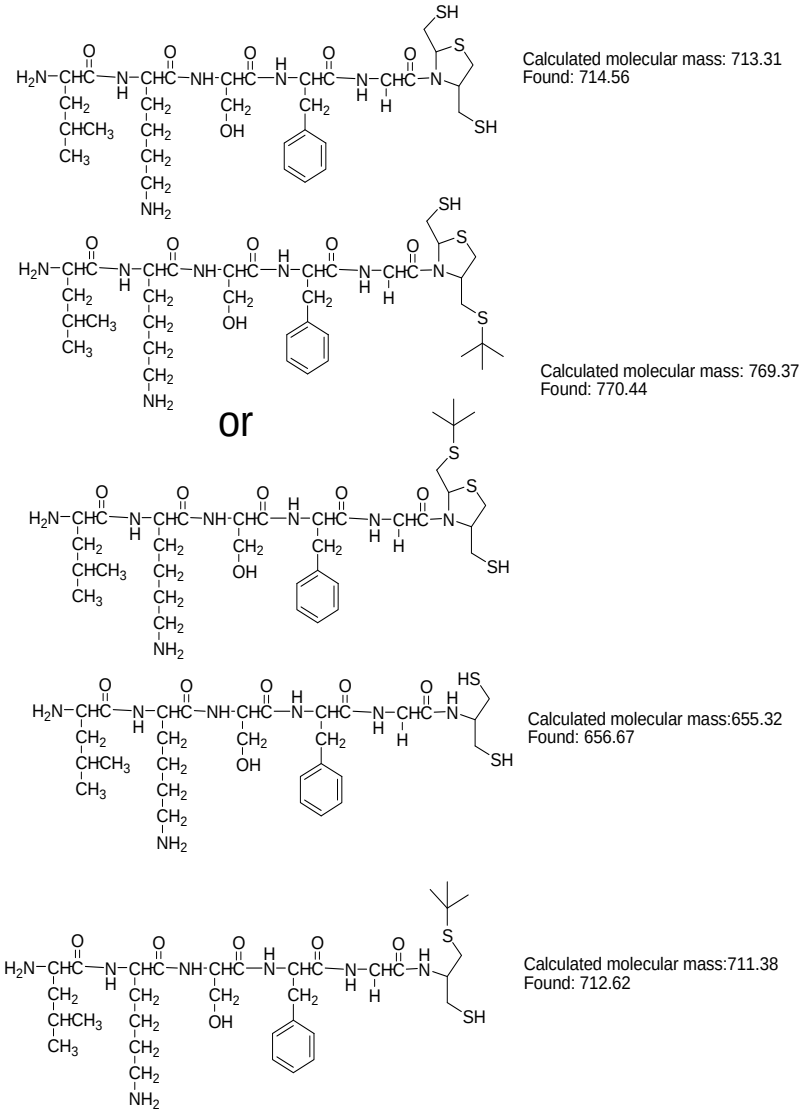


Figure 2-15. Analysis of the cleavage crude of peptide-BMMT.

A) Loading of the BMMT liker to resin and assembly of peptide. B) The raw ESI-MS profile of cleavage solution. C) The possible structures relative to the molecular weights detected in MS.

2.2.4 Ligation between thioester precursor and a cysteinyl peptide

After purification of the peptide H-LKSFG-BMMT (peak b and peak b*) by HPLC, we used them to ligate with a cysteinyl peptide H-CAKAFA-NH_2.

Ligation product peak c was observed in a very short time. However, the product peak remained unchanged after the total consumption of one starting material isomer b*. With prolonging incubation, another starting material isomer b was transformed into peak d which was 2 dalton less. This -2 species was believed to be the disulfide form of starting material peak b **(Figure 2-2)**. As peak b can undergo such oxidation reaction, we hypothesized that peak b was the *cis* isomer with two thiol groups at the same side of the ring while peak b* was its *trans* counterpart.

LKSFG-BMMT + CAKAFA → LKSFGCAKAFA

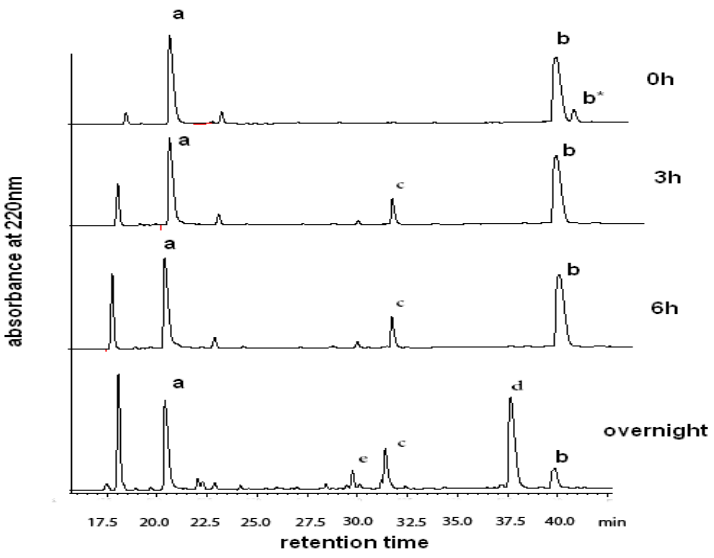


Figure 2-2. Ligation reaction between H-LKSFG-BMMT and H-CAKAFA-NH_2.

Reaction condition: 3 mM thioester precursor, 6 mM cysteinyl peptide, 20 mM methoxyl amine, 50mM TCEP, 2% N-methyl mercaptoacetamide, pH 6, 42^{o}C incubation

This experiment indicated two important facts: 1 two cis/*trans* isomers of H-LKSFG-BMMT differ greatly in relative abundance 2 only the *trans* isomer is productive in ligation reaction.

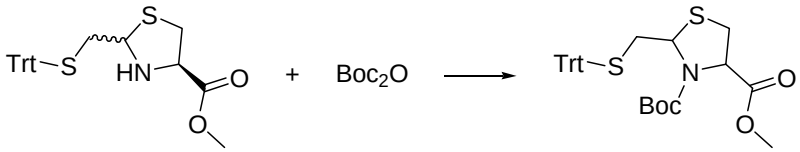
2.2.5 Analysis of the *cis/trans* issue

Since only the *trans* isomer is productive in ligation, effort to get it as major product is worth pursuing.

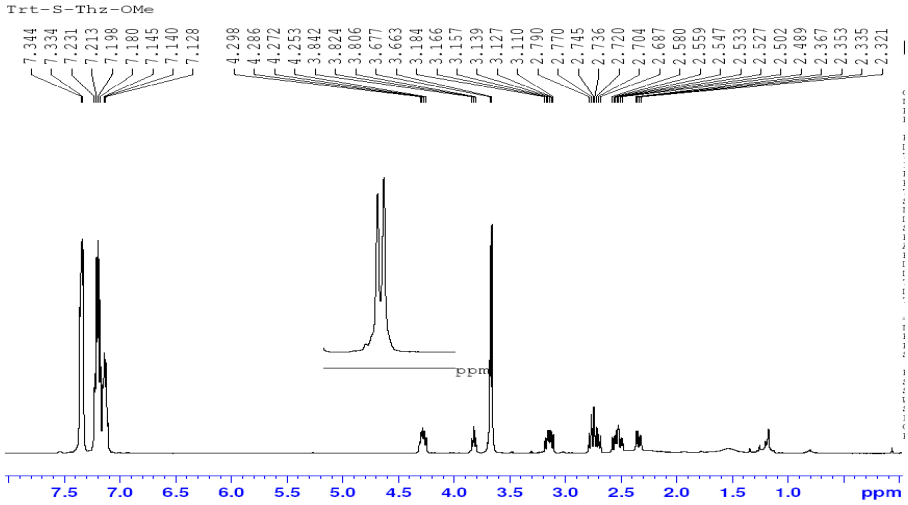
To investigate why two isomers of the final product differed that much in relative abundance, we started with thorough examination of every synthesis step. The Boc protecting step was found to deserve special attention. We used NMR as our tool to study the reaction. The methyl ester acted as a good indicator of the constitution of stereoisomers (detail described in the legend). From the profile, it was clear that the ring-formation step in the synthesis afforded two *cis*/*trans* isomers with a ratio close to 1:1. However after it was protected with Boc, there was merely one isomer left **(Figure 2-3 B and C)**.

Given the fact that the yield of this protection was below 50%, we envisioned that if we could find a stronger reaction condition to improve the yield we could get both *cis*/*trans* isomer in the final product after the whole synthesis..

A)



B)



C)

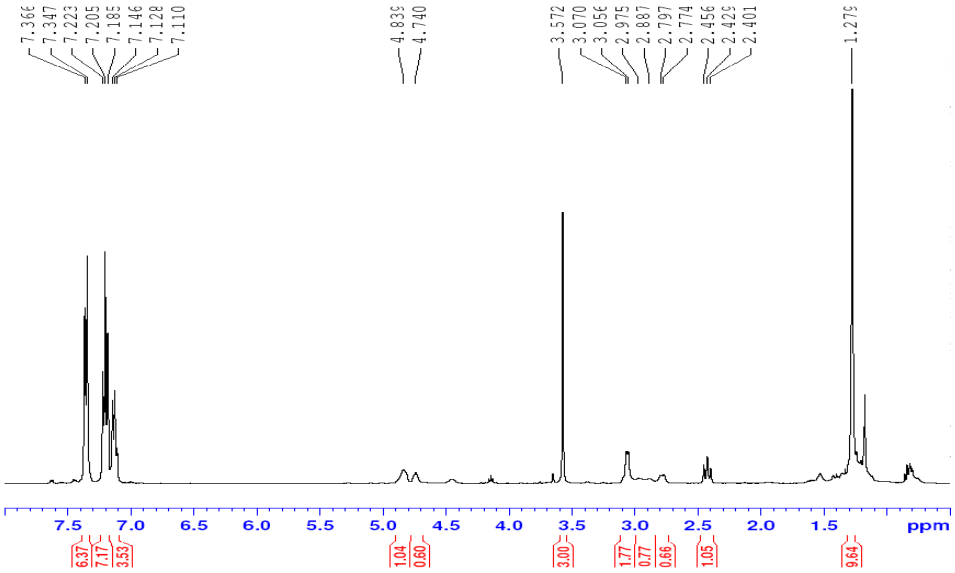


Figure 2-3. Study on the stereochemistry of thiazolidine.

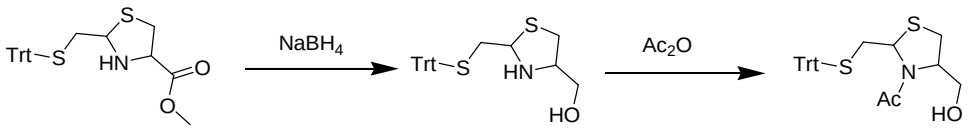
A) Boc protection of thiaolidine ring. B) NMR profile of unprotected thiazolidine ring. The doublet with chemical shifts 3.663 and 3.677 correspond to the three hydrogens of the methyl ester group in two cis/trans isomers. The area ratio of the two peaks is close to 1:1 which means the equality of the two isomer's relative abundance. C) NMR profile of thiazolidine ring after Boc protection. The singlet with chemical shift 3.572 corresponds to the three hydrogens of the methyl ester group in one single isomer.

We found two conditions that significantly improved the yield. One was to use ACN as the solvent, DIEA as the base, heating the reaction at 90 ^{o}C. The other one was to use water and dioxane mixture as solvent, sodium bicarbonate as the base, room temperature. Representative HPLC profile of reaction analysis clearly showed that most of the starting material was converted to the product. NMR analysis was done on the

purified product. Disappointingly, it still showed that there was only one isomer in the Boc-protected product just as previous.

An occasional finding provided us the clue to this problem. There was one time we changed the synthesis strategy in which we reduced the methyl ester first with excessive reducing reagent and then protected it with acetyl anhydride without prior purification **(Figure 2-4 A)**. In the HPLC analysis profile, we found mainly four products, one with one acetyl group added, two with two acetyl groups and one with three acetyl groups added. This was totally unexpected because there should be, at most, two functional groups, secondary amine and hydroxyl group, available for acetylation **(Figure 2-4 B)**.

A)



B)

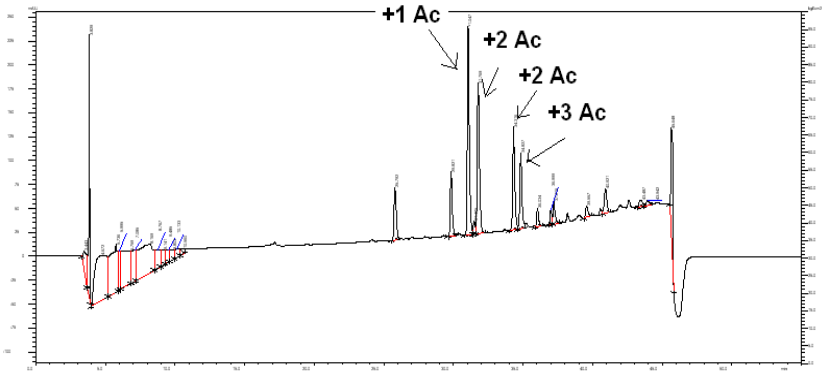


Figure 2-4. Acetylation of thiazolidine ring.

A) Two steps of reaction to protect thiazolidine with acetyl group B) HPLC analysis profile after the two steps of reaction.

Careful check on the reduction product showed that besides the desired product (408.05) there was a +2 species (410.08) in it **(Figure 2-5)**. This unexpected finding drove us to study the nature of thiazolidine ring more deeply. Literature indicated that the two *cis/trans* isomers of thiazolidine ring are epimerizable through a linear tautomer **(Scheme 2-15)**. Based on this equilibrium, the acetylation question could be explained: the +2 species should be due to the overreduction on the imine group of the linear tautomer **(Scheme 2-16)**. The overreducted product had three functional groups for acetylation: hydroxyl group, a secondary amine group and also the thiol group.

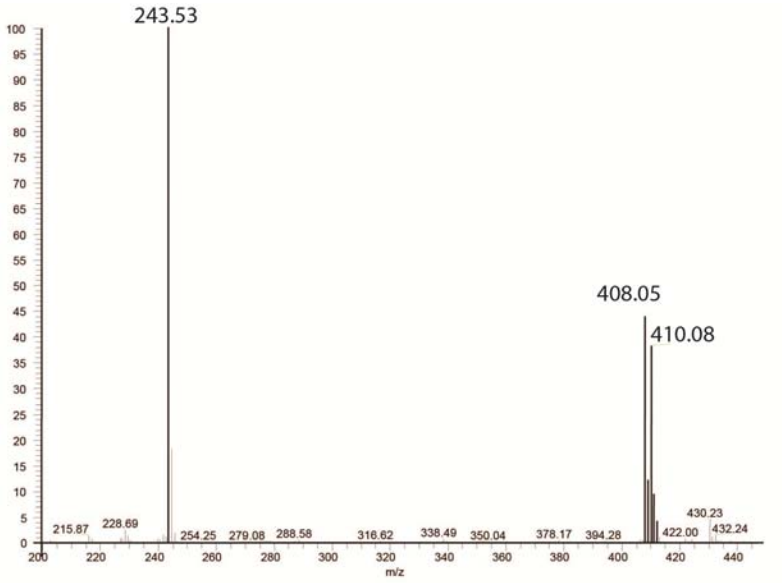
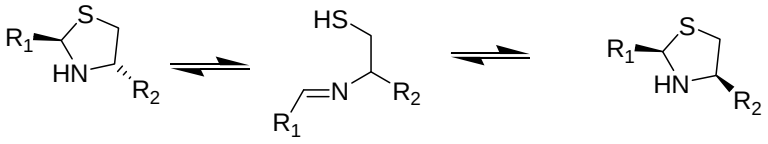
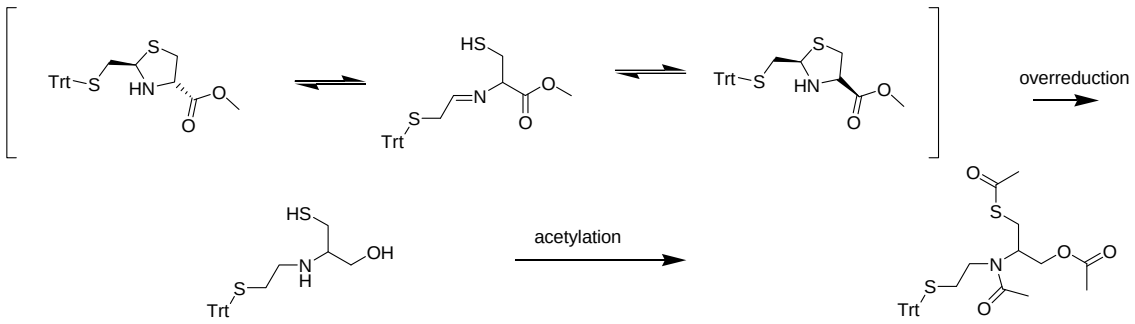


Figure 2-5. ESI profile of the crude reduced product. $[M + H]^+$ 408.05 corresponds to the desired product, 243.53 is the trityl group on the compound. $[M + H]^+$ 410.08 indicates there is a +2 side product.



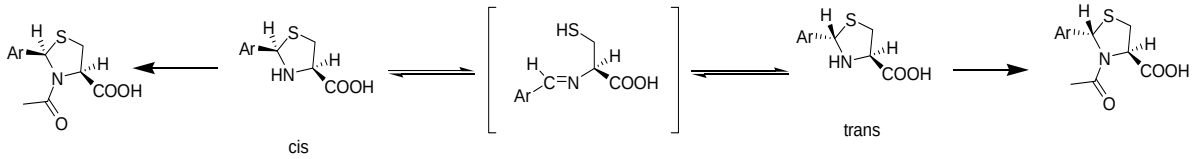
Scheme 2-15. Two *cis/trans* isomers of thiazolidine are epimerizable through a linear tautomer.



Scheme 2-16. Overreduction and multiacetylation of thiazolidine.

We hypothesized that the great difference in the relative abundance of the two isomers in the final product could be due to the condition used in the acetylation step which greatly drove the *cis/trans* equilibrium to one side.

A 1978 JACS paper reported a very similar case.[116] In this paper, it was detailed described that upon acetylation, different conditions especially the solvent used could lead to product with different conformation. In their case, if water was used and the reaction was heated at 100 ^{o}C, almost all the product formed would adopt *cis* form, however, if pyridine was used instead, the reaction done at room temperature would give product exclusively in *trans* conformation **(Scheme 2-17)**.

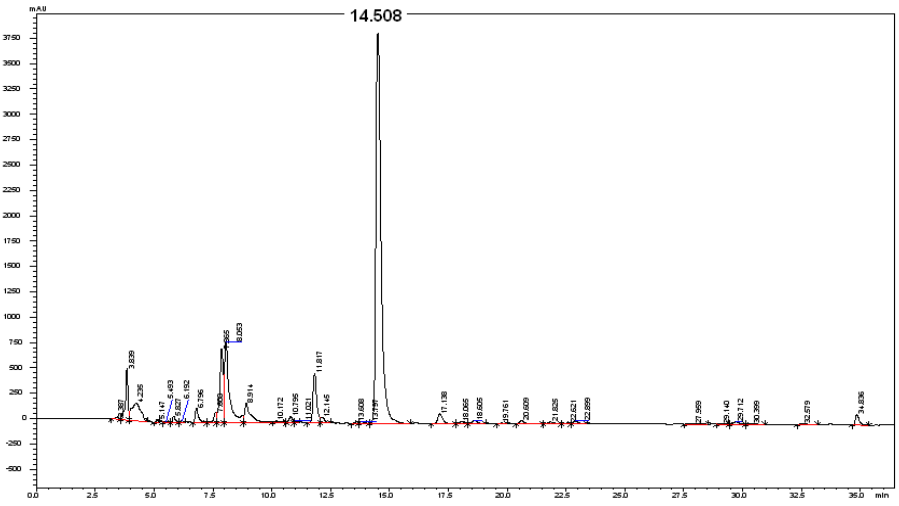


Scheme 2-17. Effects of reaction conditions on the thiazolidine stereo conformations.

Inspired by this report, we did a test to see whether the conformation in our case could also be decided by the acetylation reaction condition or not. We tried two different conditions to do the Boc protection on the thiazolidine ring: one was to use water/dioxane mixture as solvent, sodium bicarbonate as base, room temperature; the other one was to replace water with pyridine as solvent, no base added and room temperature. HPLC was used to monitor the reaction. The former condition gave product in exclusively one single peak (retention time = 14.508 min) while the latter one yielded two products both with desired product's molecular weight, eluting out at 12.9 and 13.3 minutes **(Figure 2-6)**.

This experiment gave us the direction of future work: in order to get desired trans peptidyl-BMMT isomer, proper reaction condition needed to be found for the first amino acid loading to the thiazolidine ring.

A)



B)

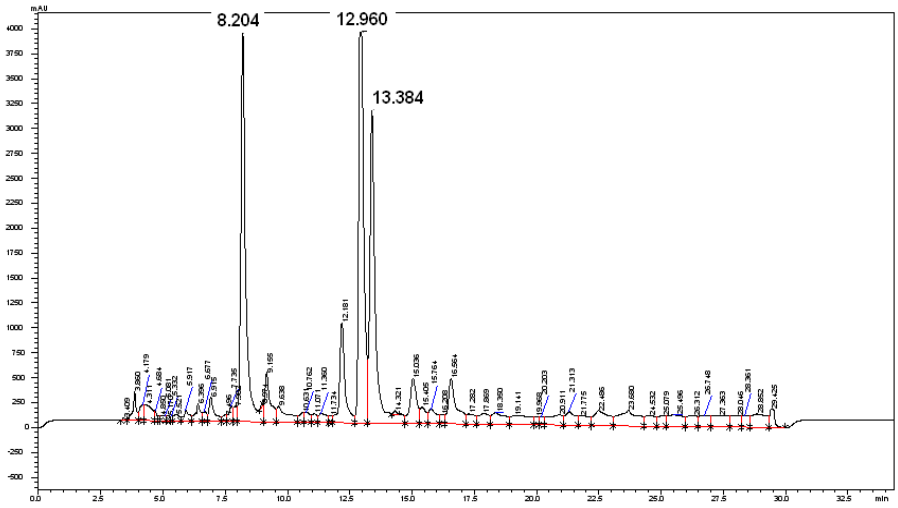


Figure 2-6. Acetylation of thiazolidine using two different conditions.

A) HPLC profile of the reaction using water/dioxane mix as solvent. Peak with retention time 14.508 min is the product. B) HPLC profile of the reaction using pyridine as solvent. Peak with retention time 8.204 min is the unconsumed starting material and the two peaks with retention time 12.960 min and 13.384 min are products formed.

2.2.6 Optimization of the synthesis route

Another problem yet to solve was the poor overall yield (< 10%) of the synthesis. To make our method practically useful, optimization of the synthesis strategy needed to be done. After trials and errors, a more suitable synthesis route was found **(Scheme 2-18)**. We mainly made changes in three aspects of the synthesis: one aspect is the Boc protection step which appeared to have very low yield, the second aspect is the hydroxyl-to-thiol transformation which comprised three steps in our first design and the

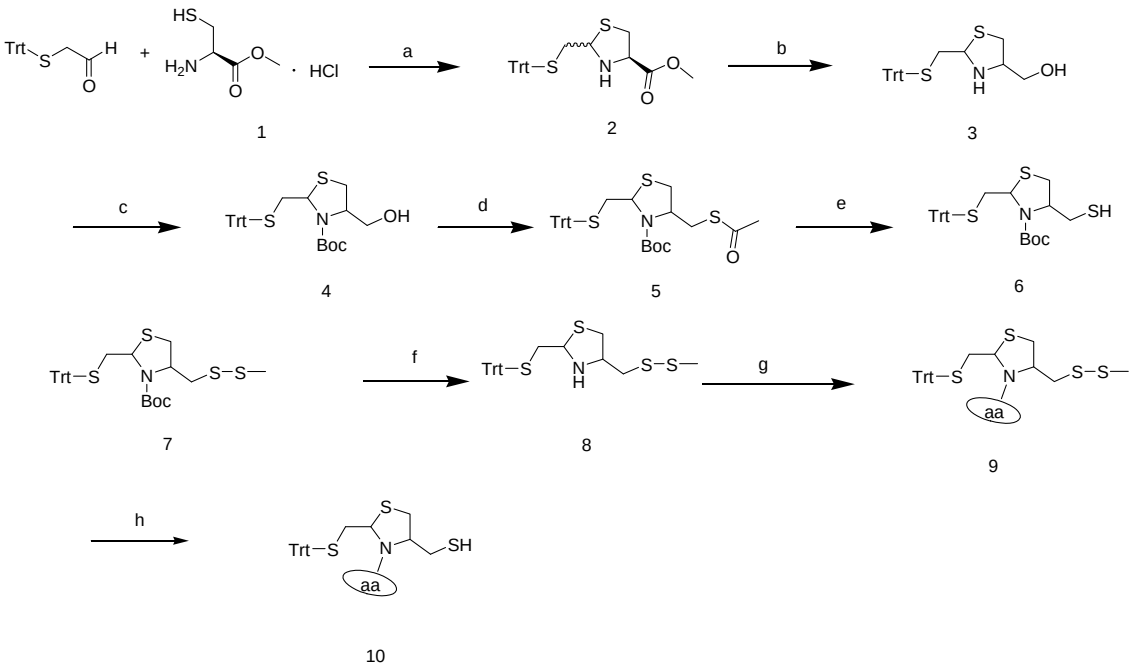
third one is the Boc deprotection as TFA treatment in our initial attempt led to the ring-open side reaction.

For Boc protection, since the electron-withdrawing effect of the methyl ester prevents efficient Boc adding, we decided to reduce the methyl ester first and then protect the secondary amine. Normal reducing condition would be problematic as overreduction of the linear tautomer would happen which was described above. We found that a mixture of $Ca(BH_4)_2/NaBH_4$ could reduce the methyl ester to hydroxyl group while leaving thiazolidine ring intact. After reduction, the secondary amine was protected with Boc with a much higher yield than previous. This validated our previous hypothesis about the electron-withdrawing effect of methyl ester group contributing to the difficult Boc-adding.

For the conversion of hydroxyl group to thiol group, we managed to reduce the previous three steps to two steps by using Mitsunobu reaction. This would lead to the overall yield improvement.

As for the Boc removal, 3N HCl in methanol was used to replace TFA. This acid appeared to largely prevent the ring-opening side reaction.

Subsequent loading of the first amino acid onto the thiazolidine ring is still under investigation. This is the crucial step of the whole synthesis as its reaction condition decides the isomer constitution to be obtained in the final peptide product.



Scheme 2-18. Optimized synthesis route (under working). Conditions: (a) DIEA, MeOH; (b)$NaBH_4$, $CaCl_2$, EtOH, -10 ^oC to rt; (c) Boc_2O, H_2O/dioxane, $NaHCO_3$; (d) Ph_3P, DIAD, HSAc; (e) NaOH, MeOH; (f) 3 N HCl in MeOH; (g) reaction under investigation; (h) $NaBH_4$, THF.

2.3 Conclusion

In this project, we have designed a new molecule BMMT to prepare peptide thioesters by Fmoc chemistry utilizing the N-S acyl shift mechanism. Two stereoisomers were obtained in the final BMMT-peptide product due to the chiral centers on the thiazolidine ring. According to our experiment, it was found that only *trans* isomer was productive in ligation. Through analysis of the molecule synthesis, we discovered that the stereoisomer constitution of the final peptide-BMMT product was very likely to be determined by the reaction condition, especially the solvent, in the acylation step of the thiazolidine. In order to make our method more practically useful, proper investigation on the reaction condition for first amino acid loading is worth our further efforts. Also, we have optimized the chemical synthesis route leading to a higher yield.

2.4 Materials and Methods

2.4.1 Materials

Amino acids, coupling reagents and resins were obtained from chemimpex, Novabiochem and GL biochem. All other chemical reagents were of analytical grade, obtained from Sigma Aldrich, alfa aesar and Acros Organics. All solvents and chemicals were used as received without purification unless otherwise indicated.

2.4.2 Instruments

The HPLC runs were carried out on a Shimadzu HPLC system with a SPD-M20A prominence diode array detector. The analytical analysis was done with a C18 reversed-phase column (Jupiter C18, 5 µM, 4.6 x 250 mm). The purification was performed using a semi-preparative HPLC column (Jupiter C18, 5 µM, 10 x 250 mm). All the HPLC runs were done using the mixture of two solvents, A and B, as the mobile phase. Solution A was H_2O containing 0.045% TFA and solution B was acetonitrile/H_2O (90:10) containing 0.039% TFA. The flow rate was 0.3 ml/ min for analysis and 2.5 ml/min for purification. UV detection was carried out at 220 nm.

Peptides ESI mass spectra data were obtained on a Thermo Finnigan LCQ DECA XP MAX with ESI ion source.

NMR spectra were recorded at room temperature on a 400 MHz Bruker DPX 400.

2.4.3 Solid phase peptide synthesis

Standard Fmoc chemistry is performed on rink-amide-MBHA (for synthesis of peptides with C-terminal amide) or 2-chlorotrityl chloride resin (for synthesis of peptides with C-terminal acid).

Specific steps are as follows:

1 Swell the resin in DCM for 10 min before use.

2 For Fmoc removal, 20% piperidine in DMF was added. Deprotection was done twice for 2 and 20 min each time. After deprotection, the resin was washed with DMF, DCM, DMF alternatively.

3 For coupling, solution of 4eq Fmoc amino acids, 4eq Pybop and 8eq DIEA dissolved in DMF/DCM mixture was added to the resin. Coupling reaction was carried out for 60-90 min. Coupling efficiency could be checked using Kaiser test. After the coupling was done, the resin was washed with DMF, DCM, DMF alternately.

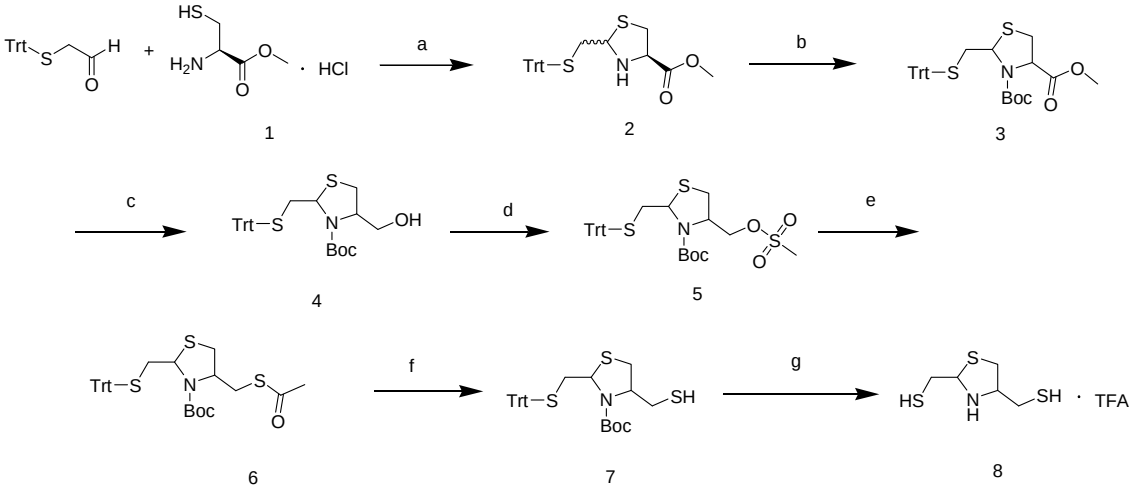
4 After the whole sequence was assembled, the peptide was cleaved from the solid support using a cleavage cocktail containing 95% TFA, 2.5% water and 2.5% TIS for 1 hour. Precipitated by cold ether, the crude peptides were purified by HPLC. After lypholization, peptides were obtained in powder form, which could be used freshly or stored at -20 ^{o}C for further use.

For the synthesis of peptide BMMT, 2-chlorotrityl chloride resin in vessel was first treated with 30% TFA dissolved in DCM for 2 to 3 mins. Acid was then filtrated and resin was washed with anhydrous DCM. The above procedure was repeated three times.

BMMT was dissolved in anhydrous DCM and added to the treated resin. Vessel was gently shaken for 30 mins. After that, solution was filtrated. The residual resins in vessel were ready for SPPS.

2.4.4 Chemical synthesis of BMMT

initial synthesis route



General procedure to prepare tritylthioethyl aldehyde

One of the building block of the thiazolidine ring – tritylthioethyl aldehyde is prepared from the commercially available compound 3-mercapto-1,2-propanediol. To a solution of 3-mercapto-1,2-propanediol (60 mmol, 6.472 g) in THF (50 ml) was added triphenyl methane chloride (90 mmol, 25.08g). The mixture was stirred at room temperature (rt) for 15 hours. The organic solvent was evaporated, and the residue was purified by silica-gel column chromatography using hexane and EtOAc as the eluent to give 15.5 g (74%) of 3-tritylthio-1,2-propanediol. 3-tritylthio-1,2-propanediol (11.7 mmol, 4.1 g)

was then dissolved in 30 ml THF/H_2O (5:1) followed by the addition of sodium periodate (12.86 mmol, 2.75 g) in one portion. The mixture was sonicated in normal ultra sonic for 15 minutes and then stirred for 30 minutes at room temperature. The solid was then removed by filter, the filtration was concentrated under reduced pressure, the residue was extracted by DCM/H_2O. The organic phase was washed with brine, dried over sodium sulfate, and was concentrated under reduced pressure. The product was obtained in the form of oil.

General procedure to prepare **2**

Compound **1** (27 mmol, 4.6g) was first dissolved in methanol (50 ml) and carefully neutralized with equimolar amounts of DIEA. Freshly prepared tritylthioethyl aldehyde (17 mmol) was added to the above solution. The reaction was stirred for 24 hours and the generation of the product b could be checked by ESI. Methanol was evaporated and the crude product was purified over silica gel chromatography using n-hexane/EtOAc as eluent. The amount of the final product was 5.81g in the form of yellow oil (yield ~ 78.6%).

General procedure to prepare **3**

To a solution of **2** (13 mmol, 5.81 g) in DCM was added di-tert-butyl dicarbonate anhydride (25 mmol, 5.45 g). A few drops of DIEA were added to adjust the pH of reaction to slightly basic (~ PH 8). Let the reaction stir for two days. The organic solvent was removed under reduced pressure and crude product was purified over silica gel chromatography. Yield after silica gel chromatography (in the form of transparent oil): 4.7 g, 8.8 mmol, ~68%. ^{1}H NMR (400MHz, $CDCl_3$): δ □7.25 - 7.65 (m, 15H), 4.90 - 5.01 (m, 2H), 3.74 (s, 3H), 3.30 (m, 2H), 2.97 (broad s, 1H), 2.60 (t, 1H), 1.41 (s,

9H); ^{13}C NMR (400 MHz, $CDCl_3$): δ 171.15, 152.65, 144.74, 129.67, 127.93, 126.70, 81.61, 66.86, 63.03, 52.52, 38.49, 32.51, 30.07, 22.71, 14.15.

General procedure to prepare **4**

Compound **3** (2.4 mmol, 1.3 g) was dissolved using dry THF (30 ml) at 0 ^{o}C. Reducing agent Lithium aluminium hydride (3.375 mmol, 0.135 g) was slowly added in. The reaction was stirred for three hours when TLC showed completion. Quenching reagent $KHSO_4$ solution was added dropwise into the reaction on ice until no more bubbles could be seen. The reaction was concentrated. The concentrate was dissolved in ethyl acetate and washed with water and brine, dried over sodium sulfate. Yield after silica gel chromatography (in the form of oil): 730 mg, 1.44 mmol, 45%. ^{1}H NMR (400MHz, $CDCl_3$): δ 7.27 – 7.49 (m, 15H), 5.36 (s, 1H), 5.23 (s, 1H), 4.36 (s, 1H), 3.57 – 3.78 (m, 2H), 2.91 – 3.16(m, 2H), 1.48 (s, 9H).

General procedure to prepare **5**

To a solution of **4** (0.8 mmol, 400 mg) and Et_3N (243 µl, 1.73 mmol) at 0 ^{o}C in DCM (15 ml) was added MsCl (108 µl, 1.37 mmol). The reaction mixture was stirred for one hour when TLC analysis showed completion. The crude product was purified over silica gel affording the mesylate **5** as white foam. Yield: > 90%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.27 - 7.51 (m, 15H), 5.11 (s, 1H), 4.55 (s, 1H), 4.30 (s, 2H), 3.12 - 3.15 (m, 2H), 2.96 – 3.03 (m, 3H), 2.46 – 2.96 (d, 2H), 1.47 (s, 9H); ^{13}C NMR (400MHz, $CDCl_3$): δ 152.97, 144.58, 129.59, 128.05, 126.85, 81.72, 68.14, 66.85, 62.56, 60.18, 39.10, 37.23, 31.92, 28.29.

General procedure to prepare **6**

Potassium thioacetate (208 mg, 1.82 mmol) was added to a solution of the mesylate **5** (610 mg, 1.04 mmol) in dry DMF (30 ml). The reaction was stirred at 65 ^{o}C for 3 hours when TLC analysis showed completion. The DMF was evaporated and the concentrate was dissolved in EtOAc, washed with water and brine, dried, and concentrated. Yield after silica gel chromatography (in the form of oil): 240 mg, 0.447 mmol, 43 %. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.26 – 7.54 (m, 15H), 5.15 (s, 1H), 4.49 (s, 1H), 3.22 – 3.24 (d, 2H), 3.07 – 3.22 (m, 1H), 2.79 – 2.83 (m, 2H), 2.45 – 2.50 (t, 1H), 2.35 (s, 3H), 1.50 (s, 9H); ^{13}C NMR (400MHz, $CDCl_3$): δ 194.50, 153.10, 144.673, 129.65, 127.498, 126.74, 81.16, 66.79, 62.64, 60.83, 34.20, 32.31, 31.65, 28.35, 22.72.

General procedure to prepare **7**

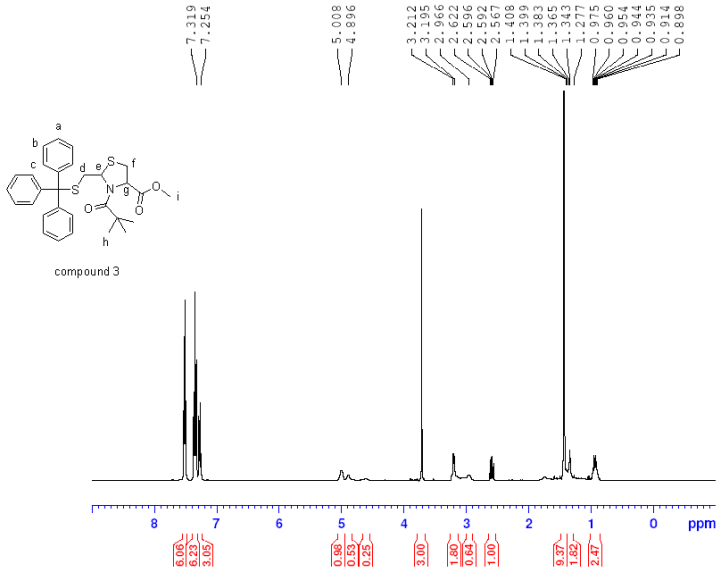
Thioacetate **6** (220 mg, 0.41 mmol) was dissolved in methanol (10 ml) and treated with 1N NaOH solution (0.492 ml) for 15 min at 0^{o}C. The reaction mixture was carefully neutralized with equimolar amounts of HOAc and concentrated. The concentrate was dissolved in ethyl acetate and washed with water and brine, dried and concentrated, affording the crude thiol as oil. Yield after silica gel chromatography: 130 mg, 0.25 mmol, 61%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.26 - 7.55 (m, 15H), 5.19 (s, 1H), 4.33 – 4.36 (t, 1H), 3.12 – 3.17 (m, 1H), 3.05 (broad s, 1H), 2.69 – 2.81 (m, 3H), 2.49 (broad s, 1H), 1.50 (s, 9H); ^{13}C NMR (400MHz, $CDCl_3$): δ 153.09, 144.64, 129.63, 128.03, 126.79, 81.17, 68.16, 66.77, 64.22, 31.80, 28.40, 23.85, 23.08.

General procedure to prepare **8**

Thiol **7** (130 mg, 0.25 mmol) was treated with 30% TFA in DCM. Thiol scavenger t-butyl thiol (150 µl) and TIS were added to prevent side reactions. Let the reaction stir for 20 min. Solvents and reagents were removed under reduced pressure. Lyopholization was performed to remove as much thiol scavenger as possible. The compound was obtained in the form of foam.

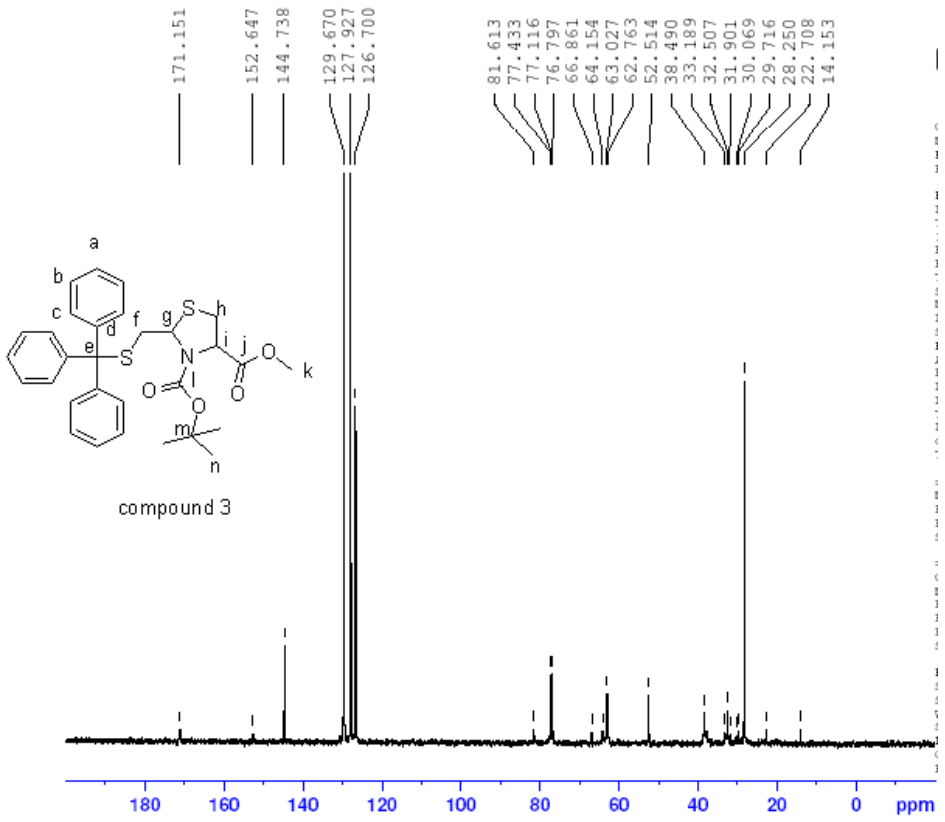
2.4.5 NMR spectra

^{1}H NMR-spectrum of compound **3**



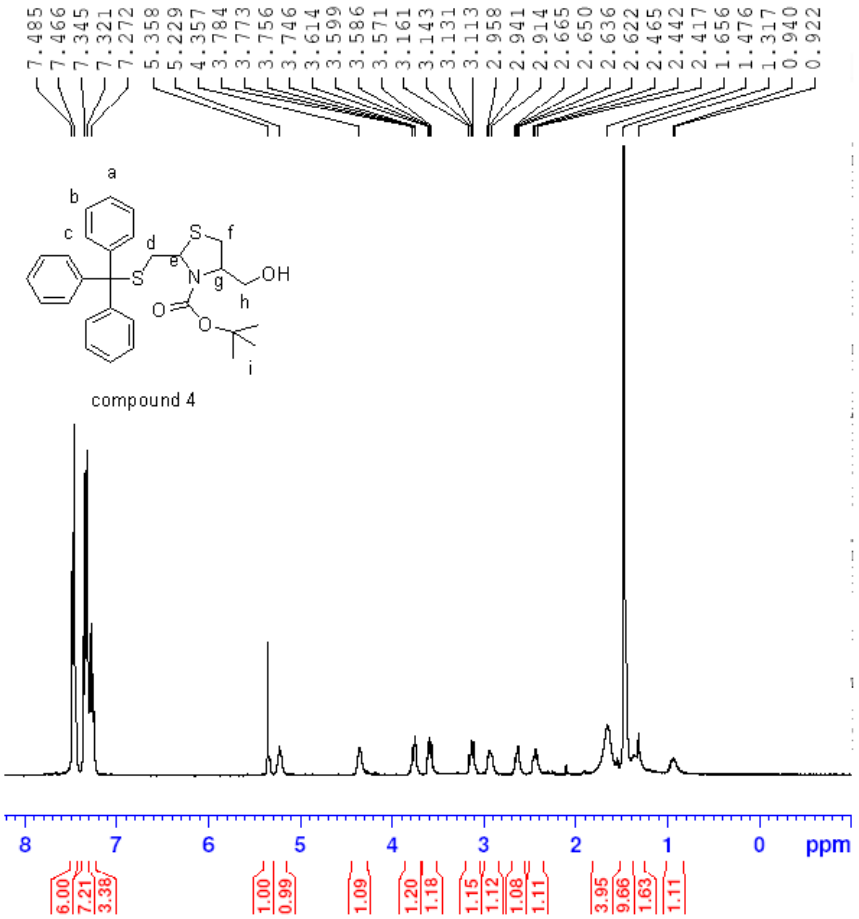
categories of ^{1}H	chemical shift
a, b, c (- trityl)	7-8
d, f (-CH_2-)	2.5-3.5
e, g (-CH-)	4.5-5.2
h (- tBu)	~1.4
impurities	0.8-1.1, ~1.3

^{13}C NMR-spectrum of compound **3**



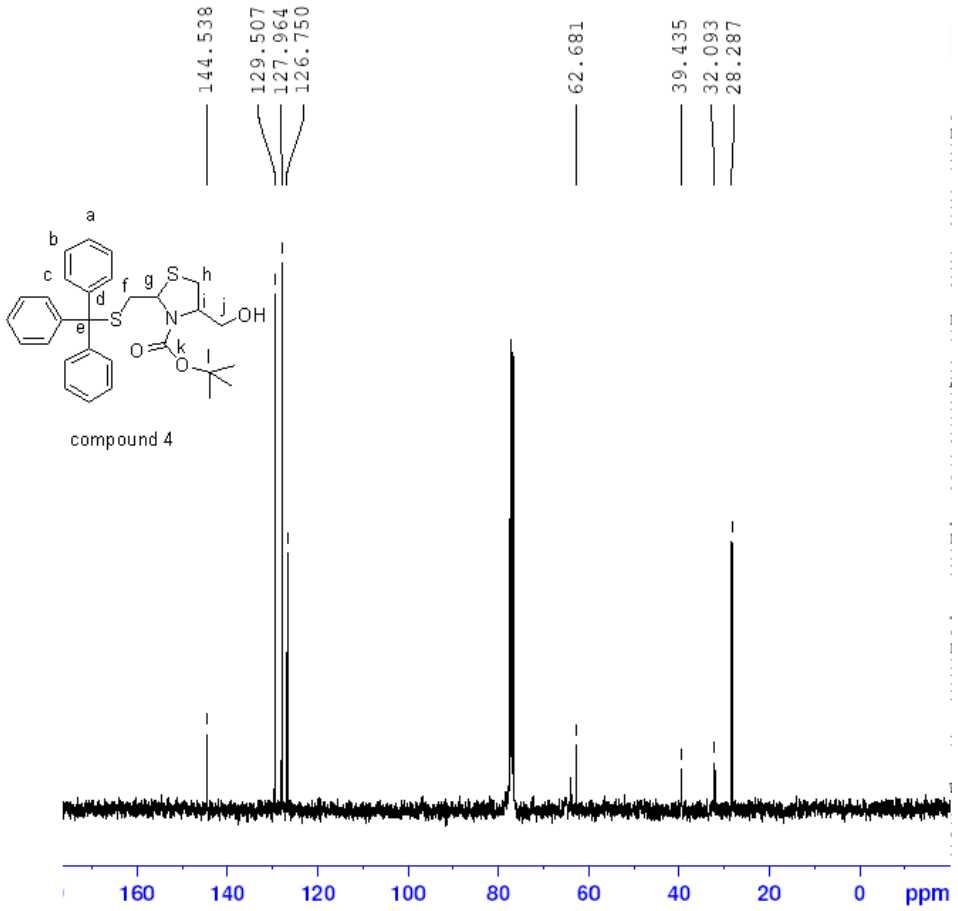
categories of ^{13}C	chemical shift
a, b, c	126-130
d	144.7
e, g, i	60-70
f, h	30-40
j	171
k	52.5
l	152.6
m	81.6
n	28.2
impurities	22.7, 14.1

^{1}H NMR-spectrum of compound **4**



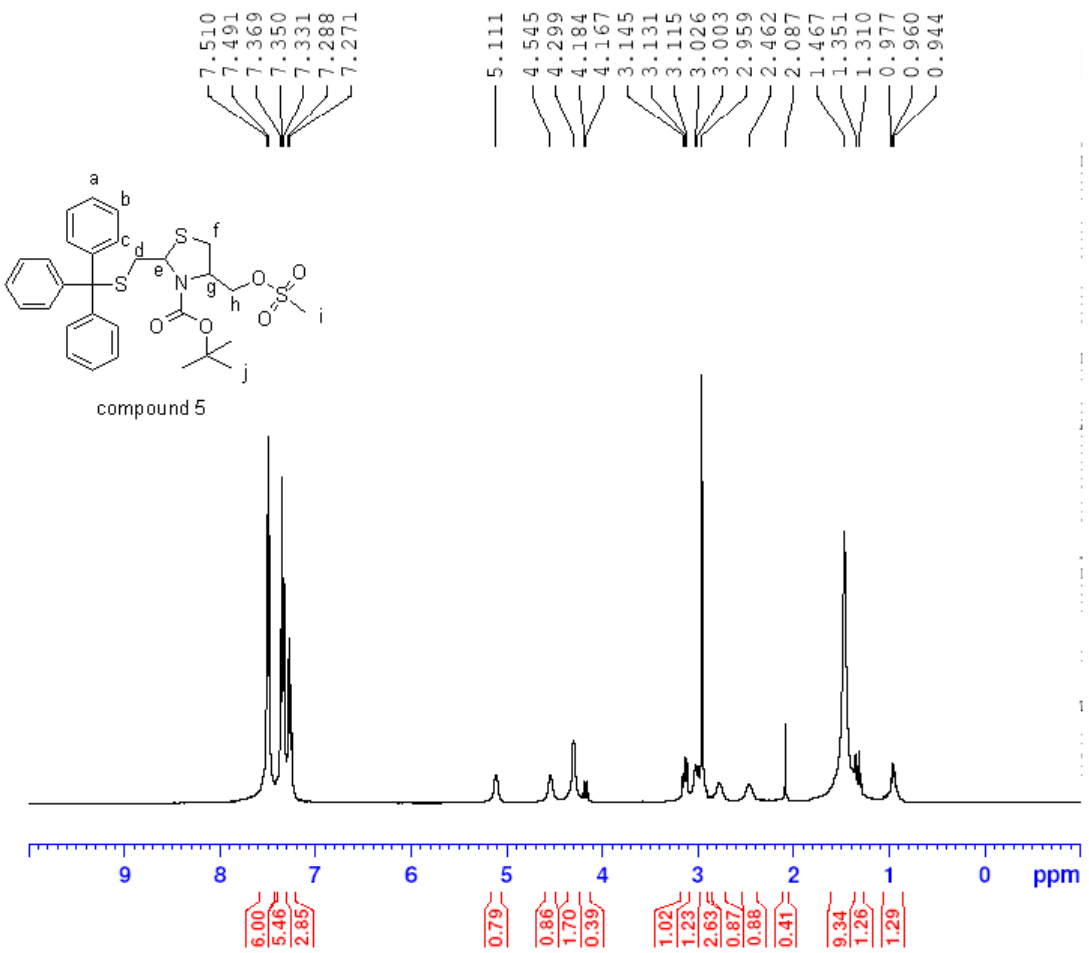
categories of ^{1}H	chemical shift
a, b, c (-trityl)	7-8
d	2.91-3.16
e	5.1-5.3
f	2.41-2.66
g	4.3-4.5
h	3.57-3.78
i (-tBu)	1.47
impurities	0.8-1, 1.2-1.4, 1.6-1.8

^{13}C NMR-spectrum of compound **4**



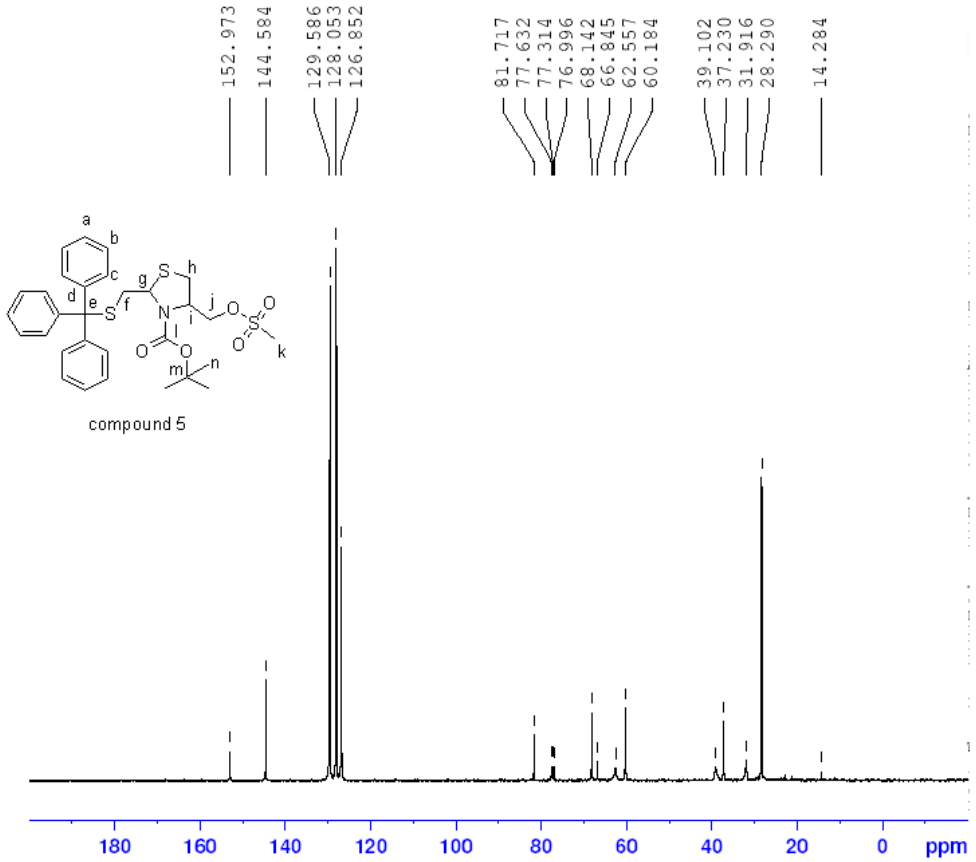
categories of ^{13}C	chemical shift
a, b, c	126.7-129.5
d	144.5
e, g, i, j	60-70
f	32.0
h	39.4
l	28.2

^{1}H NMR-spectrum of compound **5**



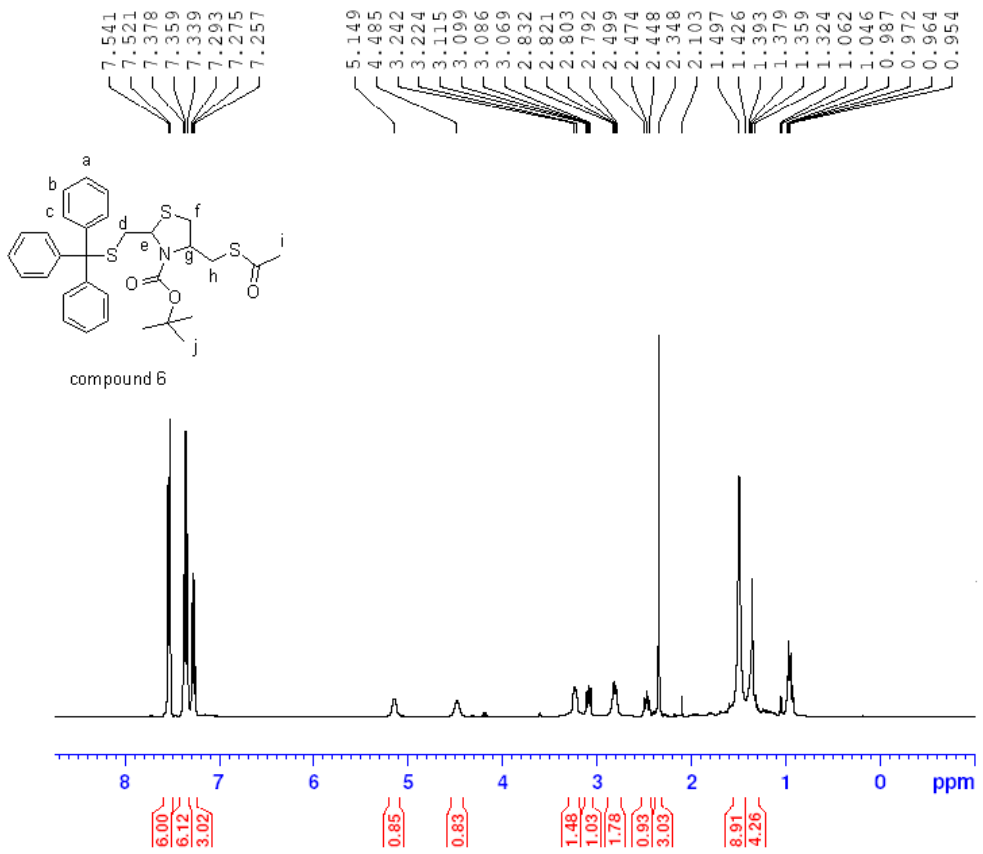
categories of ^{1}H	chemical shift
a, b, c (-trityl)	7.2-7.5
d	3.11-3.14
e	5.11
f	2.46-2.95
g	4.54
h	4.16-4.29
i (-tBu)	3.0
j	1.35
impurities	2.08, 0.94-0.97, 1.31-1.35

^{13}C NMR-spectrum of compound **5**



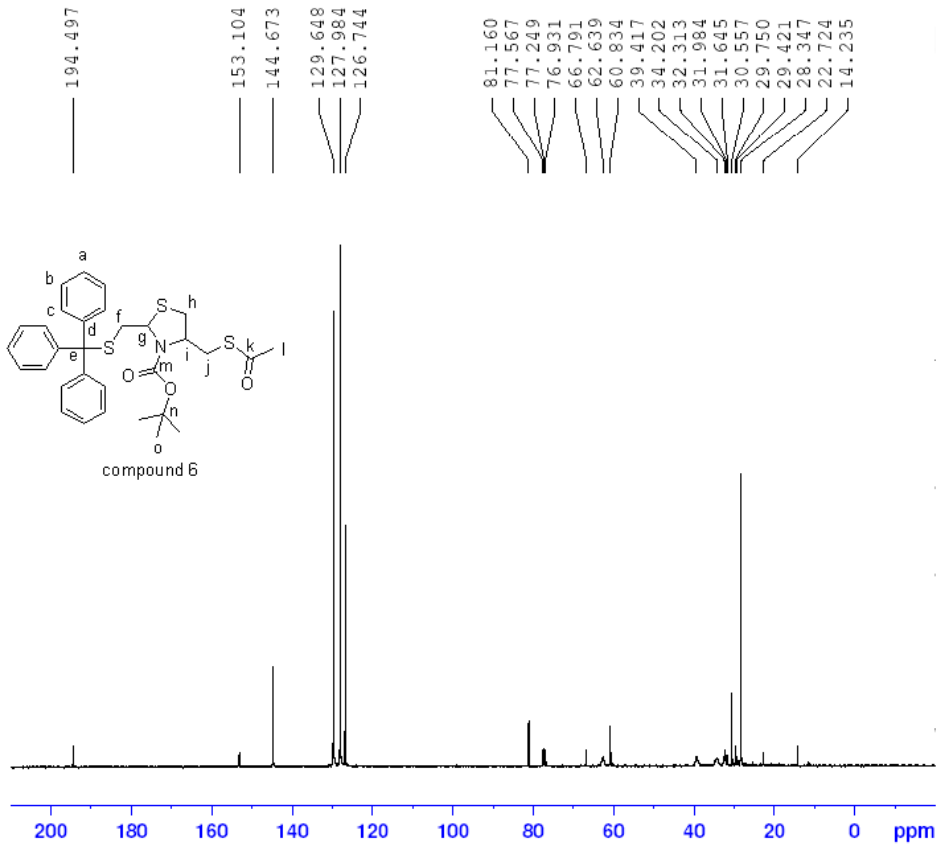
categories of ^{13}C	chemical shift
a, b, c	126.8-129.5
d	144.5
l	152.9
m	81.7
e	68.1
g	66.8
j	62.5
i	59.9
k	39.2
f	37.2
h	31.9
n	28.2
impurities	14.2

^{1}H NMR-spectrum of compound **6**



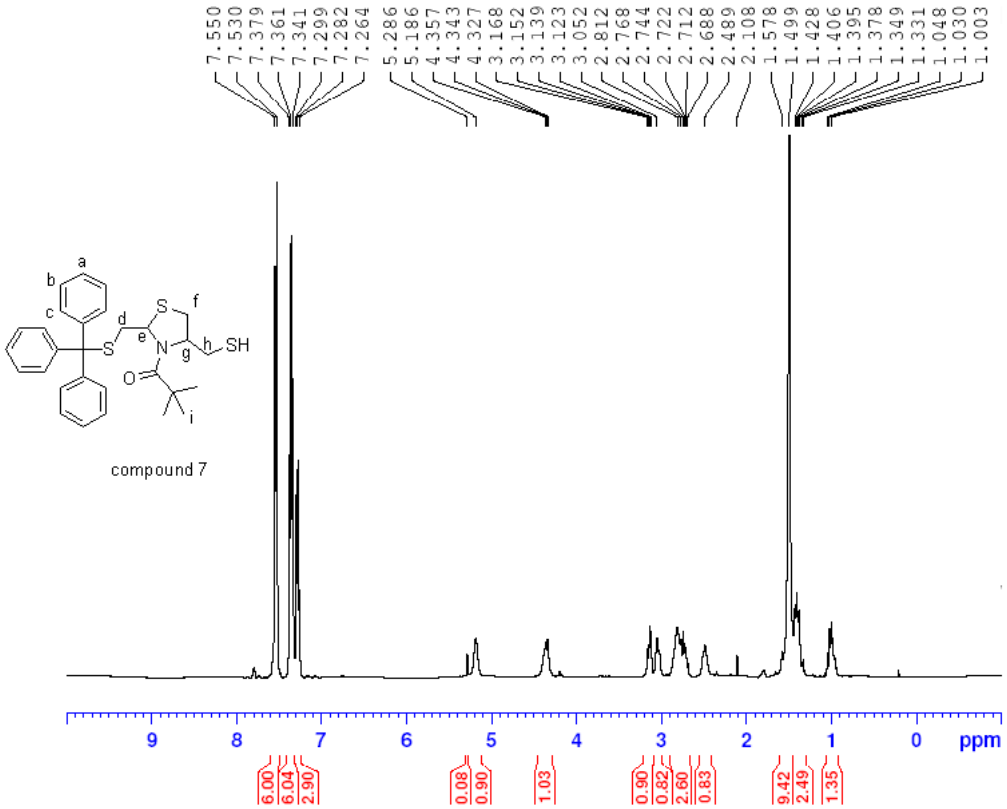
categories of ^{1}H	chemical shift
a,b,c (-trityl)	7-8
e	5.14
g	4.48
h, d	2.79-3.24
i	2.47
j (-tBu)	1.49
impurities	2.10, 1.35-1.42, 0.95-1.06

^{13}C NMR-spectrum of compound **6**



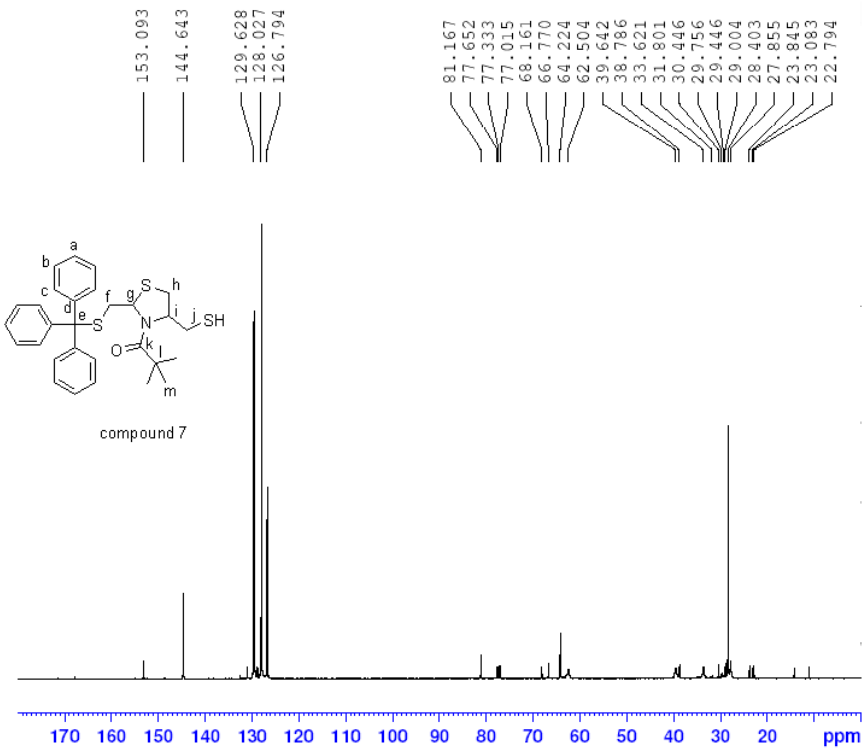
categories of ^{13}C	chemical shift
a, b, c	126.74-129.64
k	194.49
m	153.10
d	144.67
h	81.16
e	66.79
g	62.63
i	60.2
f, h, j, l	29.42-39.41
o	28.34
impurities	22.72, 14.23

^{1}H NMR-spectrum of compound **7**



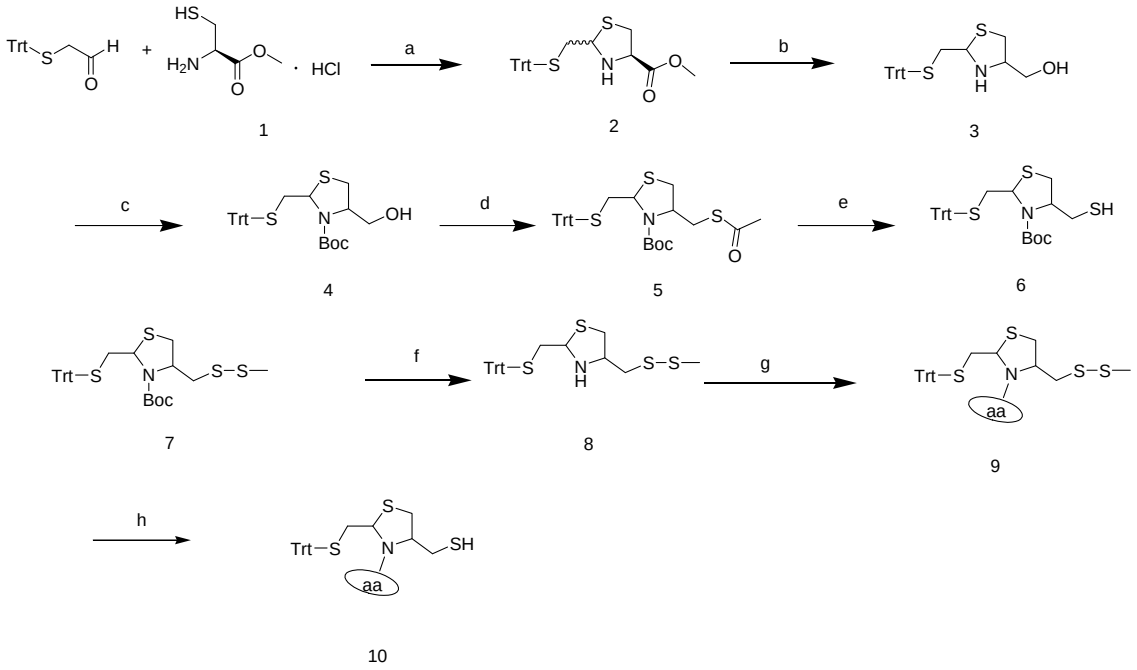
categories of ^{1}H	chemical shift
a, b, c (-trityl)	7.26-7.55
e	5.18
g	4.32-4.35
d	3.05-3.16
f, h	2.48-2.81
i (- tBu)	1.49
impurities	2.108, 1.33-1.42, 1.00-1.04

^{13}C NMR-spectrum of compound **7**



categories of ^{13}C	chemical shift
k	153.093
d	144.643
a, b, c	126.794-129.628
e	68.16
g	66.77
i	64.22
l	39.64
h	39.78
f	33.62
m	29.44
j	31.80
impurities	81.16, 62.504, other small peaks

2.4.6 Procedures of the optimized synthesis route



General procedure to prepare **3**

At – 10 ^{o}C, a solution of sodium borohydride (133 mg, 3 mmol) in ethanol (11 ml) was added dropwide to calcium dehydrate (165 mg, 1.5 mmol) dissolved in ethanol (11 ml). The mixture was stirred at this temperature for 30 min. Compound **2** dissolved in ethanol (25 ml) was subsequently added to the former solution. The mixture was stirred for another 5 hours at -10 ^{o}C. The reaction was concentrated using rotary evaporator and then extracted using EA/water. The organic layer was washed with brine and dried over anhydrous sodium sulfate. The desired pure product was obtained by silica gel chromatography purification.

General procedure to prepare **4**

Compound **3** (20 mg, 0.05 mmol) suspended in a mixture of water/dioxane (20 ml, 1:1) was added $NaHCO_3$ (23 mg, 0.25 mmol). Boc anhydride (43.6 mg, 0.2 mmol) dissolved in dioxane (5 ml) was subsequently added in a dropwise fashion. The reaction was then stirred at room temperature. Upon reaction completion indicated by TLC analysis, dioxane was removed under reduced pressure. The crude reaction was purified by silica gel chromatography.

General procedure to prepare **5**

Diisopropyl azodicarboxylate (2.0825 g, 10 mmol) was added to vigorously stirred solution of triphenylphosphine (2.625 g, 10 mmol) in 25 ml THF at 0 ^{o}C. The mixture was stirred for 30 min. Compound **5** (2.825 g, 5 mmol) and thioacetic acid (0.85 g , 10 mmol) in 10 ml THF was added slowly into it. First keep the reaction stirring at 0 ^{o}C for one hour and another one hour at room temperature. The reaction was concentrated and then purified using silica gel chromatography. The final product was obtained in oil form and the NMR data was identical to the above shown.

Chapter 3 Efficient butelase 1–mediated protein N-terminal labeling enabled by thiodepsipeptides

3.1 Introduction

Modifying the proteins with site-specific covalent linkage offers tool for not only in biomedical research but also in industries like food industry and leather producing. Many chemical reactions have been developed to functionalize specific residue, however homogeneous product is difficult to obtain when there are multiple copies of the target amino acid. To deal with this site specificity issue, other strategies are adopted, mainly taking advantage of two particular functionalities on proteins: mercaptan group of the cysteine side chain and the amino group on N-terminus.

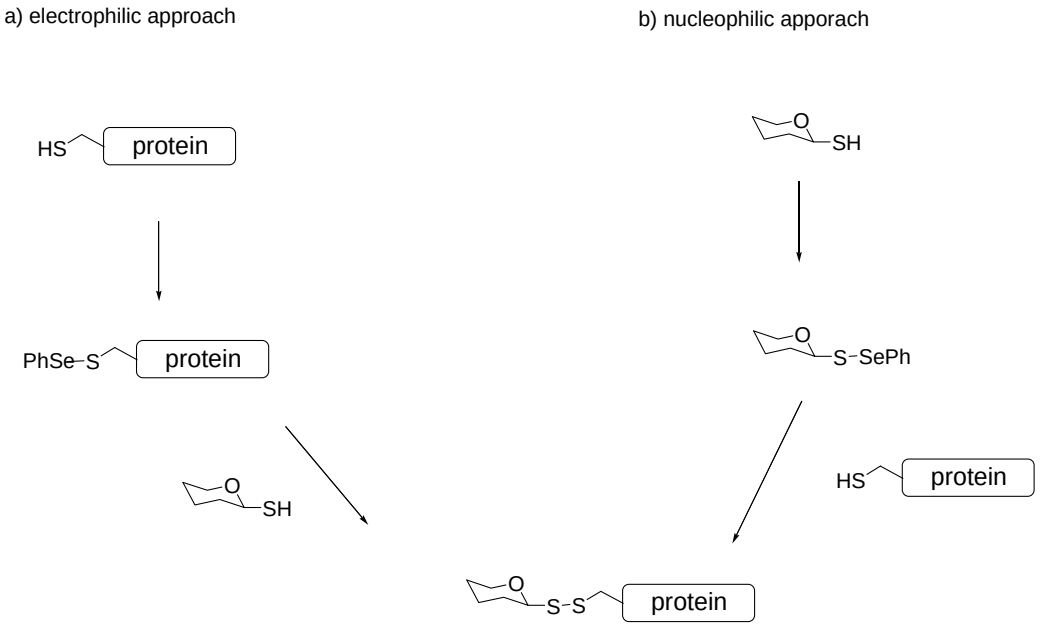
3.1.1 Chemical approach for protein application

3.1.1.1 Cysteine-directed derivatization

Cysteine residue bears thiol group on its side chain which is a very good nucleophile and this property makes it a popular candidate for various conjugation reactions. Being one of the rare occurrences in proteins, cysteine is commonly utilized for site-specific modification on proteins. One good example comes from Davis's group.[117] They used a chemoselective selenylsulfide-mediated reaction for protein glysoylation. In their report, two contrasting strategies were adopted in which cysteine residue could act as either an electrophile or nucleophile. In the first case, the cysteine residue within the protein is transformed to phenylselenenylsulfide at the beginning; the electrophilic sulfur atom in the S-Se would be susceptible to nucleophilc attack by thiol-containing saccharides. As

a result, saccharides are attached to cysteine through a disulfide bond. In another strategy, thiol-containing saccharide instead of the cysteine is converted into the phenylseleneylsulfide. The nucleophilc cysteine could attack the S-Se bond to generate the disulfide-connected ligation product.

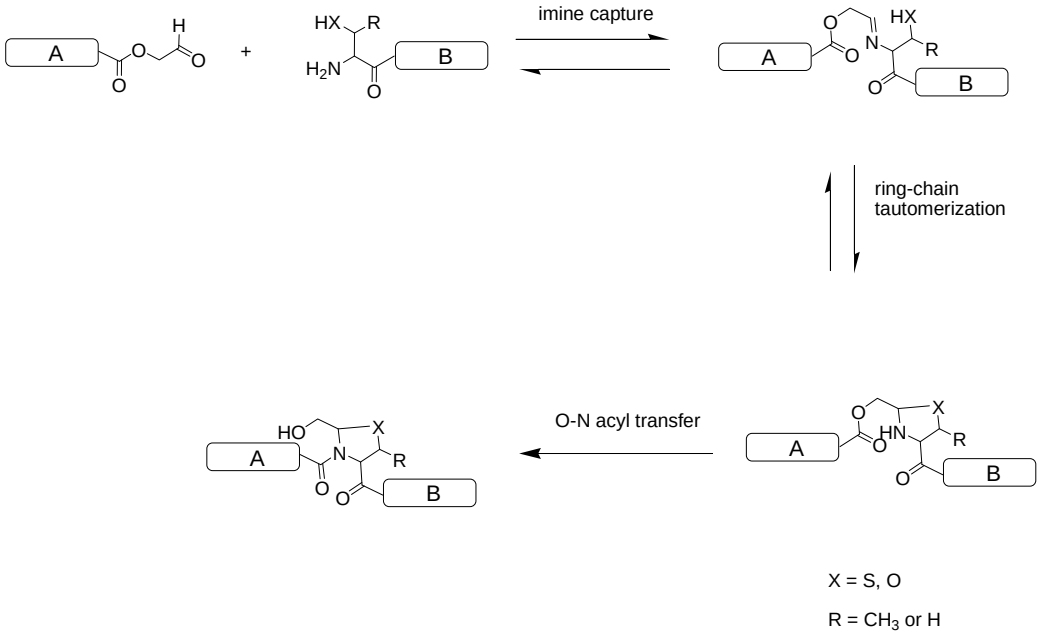
It is worth mentioning that this cysteine-directed derivatization strategy has limitation to its application scope since there are still many situations in which modification of one particular cysteine residue is difficult.



Scheme 3-1. Selenylsulfide-mediated protein glycosylation.

3.1.1.2 N-terminal chemical modification

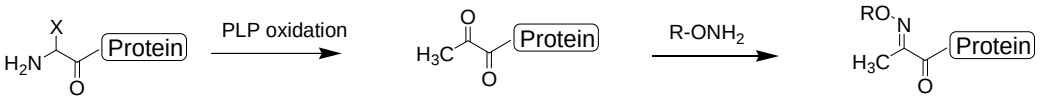
The chemoselective property of the N-terminus makes it a popular site for selective protein modification. One strategy is to take advantage of an amino acid's side-chain functional group in combination of the N^{α}-amine. The most popular reaction of this category could be the native chemical ligation in which a thioester moiety selectively reacts with an N-terminal cysteine residue to form an amide junction.[29,30,28] N-terminal cysteine could also be modified through the efficient thiaozlidine-formation reaction with an aldehyde **(Scheme 3-2)**.[118]



Scheme 3-2. Protein conjugation through thiazolidine formation.

There are also approaches independent of amino acid side chains. It has been reported that through careful control of pH, N-terminal amino groups could be selectively modified in the presence of other lysine side chains taking advantage of the two's

different pK_a values.[119] However, in the presence of many competing lysine residues within one protein, good selectivity via this method is still beyond the reach. An alternative route is reported by Francis's group who achieved N-terminal modification thourgh PLP-mediated oxidation. In this design, a protein is first treated with pyridoxal-5-phosphate (PLP) to afford ketone which could subsequently conjugate with a labeling reagent R-ONH_2 to generate an oxime linkage.[120] Since the ketone-formation selectively takes place at the N-terminus of the peptide/protein, the internal lysine side chains would not involve in the modification. This method has some restrictions on the N-terminal residue: serine, threonine, cysteine, proline and tryptophan are not compatible with this method.



Scheme 3-3. PLP-mediated oxime ligation.

3.1.2 Enzymatic approach for protein labeling

Besides the chemical ligation methods, enzyme acts as a good complement to modify proteins with the merits of substrate specificity and ease of use. In the past decade, Srt A from *Staphylococus aureus* is probably the hottest research target for protein modification. Its advantages include the following aspects: 1 The active and water-soluble recombinant Srt A is easy to prepare. 2 The enzyme has a very strong tolerance on the structure and sequence of the substrate.[76,77,79] 3 The LPXTG recognition motif

does not hamper the folding and expression of the recombinant protein. Srt A-mediated ligation has been so far successfully used in many aspects of protein study.

3.1.2.1 Protein/peptide – protein/peptide ligation

Pollok's group first demonstrated the feasibility of using Srt A as a tool for site-specific peptide/protein–peptide/protein ligation.[121] They determined the number of glycines required at the N-terminus of the nucleophile part. The conclusion was made that one single glycine was already enough for effective Srt A-mediated ligation, though two or more glycines would slightly accelerate the reaction. One example of successful cases of full size protein ligation is from Boder's group.[122] They made a dimeric eGFP by ligation of sorting motif-bearing eGFP with another eGFP with three glycines at its N-terminus.

3.1.2.2 Peptide-nucleic acid ligatoin

Peptide-nucleic acid (PNA) conjugates, as analogs of DNA, have therapeutic value. Though traditional solid phase synthesis offers a direct way to prepare these substrates,[123] enzymatic methods are also employed. Using Srt A, Pritz's group successfully made a nucleic acid-peptide fusion by ligating an 18-mer nucleic acid comprising a LPKTG sorting motif with an amphipathic peptide having a triglycine motif at its N-terminus.[124] This PNA conjugate has been shown to display improvement in cell-penetrating ability. It is worth mentioning that they also proved that the reversibility problem associated with the ligase's nature could be solved when this

reaction was done in dialysis. This design could greatly improve the ligation yield from 38% to 94%.

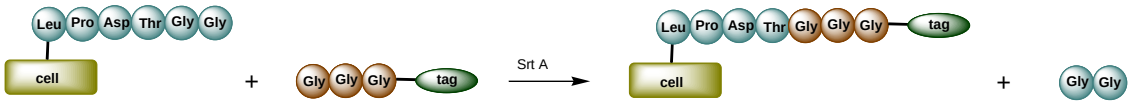
3.1.2.3 Preparation of peptide/proteins with posttranslational modification

Glycoproteins play crucial roles in many important biological events, thus the preparation of the homogeneous samples is in great need for biomedical research. Though there has been great advancement in chemical synthesis of glycoproteins, their preparation still remains a challenge. A combination of chemical and enzymatic method has been a trend now. It has been found that sugars containing a $–CH_2\text{-}NH_2$ motif could be possibly recognized by Srt A. Several 6-aminohexose-based sugar nucleophiles have already been successfully used to ligate with various peptides and proteins harboring LPXTG motif.[125]

For lipidation which is another important posttranslational modification, Srt A has also been used to prepare lipoproteins. One such example is from Ploegh's group.[126] They labeled a glycine-containing lipid substrate to a LPETG-tagged eGFP.

3.1.2.4 Labeling of living cells

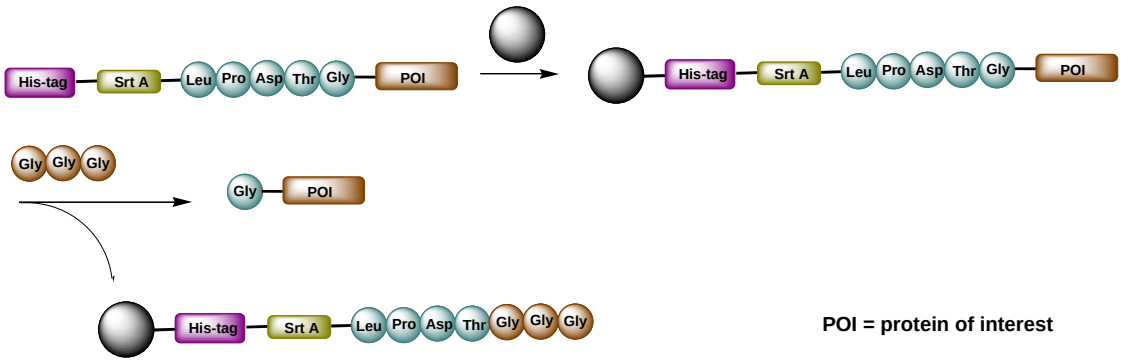
With the assistance of Srt A, surface proteins on living cells were site-specifically labeled with small molecule probes and full size proteins at both their N-terminus and C-terminus. This application allows detailed study of protein functions.[127-130]



Scheme 3-4. Labeling the C-terminus of cell surface proteins.

3.1.2.5 Protein purification

Sortase can even be used to facilitate protein purification. This strategy was reported by Mao. In this work, proteins with a His_6-Srt A-LPXTG tag are expressed and could be separated from other cell proteins by binding to the histag affinity beads. For the release of the proteins of interest, triglycine was added so that the target proteins were cleaved from the beads, tagged with one extra glycine residual at its N-terminus.[131]

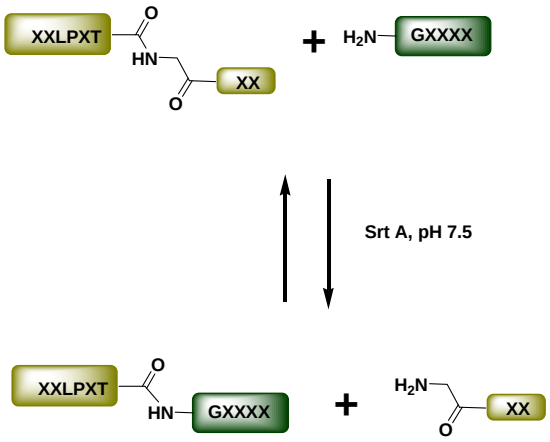


Scheme 3-5. Srt A-facilitated protein purification.

3.1.3 Reversibility issue

Despite its versatility, the catalytic efficiency of Srt A is usually not high. A satisfactory ligation yield usually requires a long incubation time of up to 24 h and high

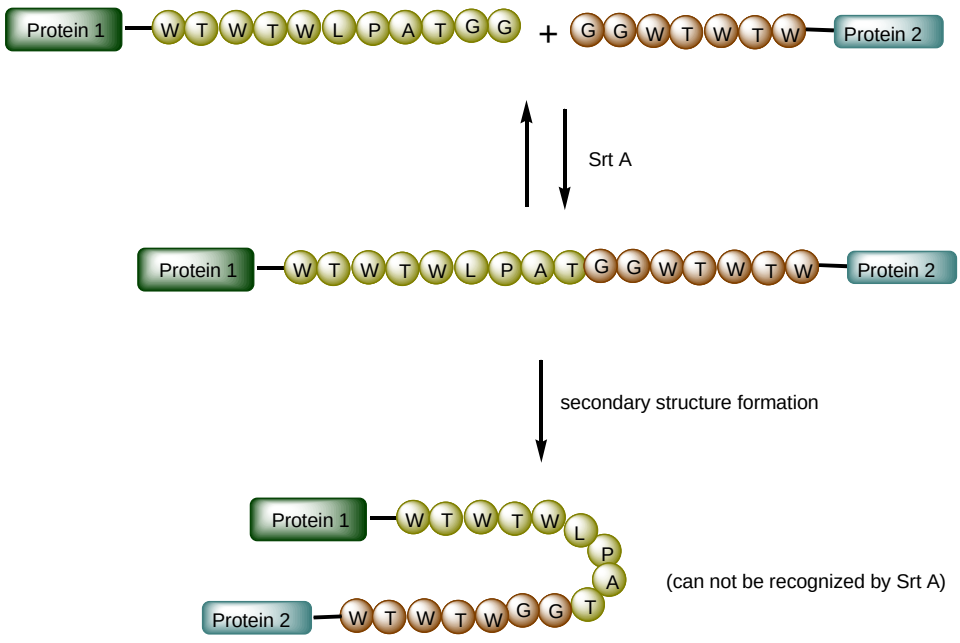
equivalent of both enzyme and substrate. The major factor that impedes the high ligation yield is the reversibility of the Srt A-catalyzed reaction. This issue originates from the fact that the ligation products could also be recognized by the enzyme just like the starting materials. Another downside of Srt A is its strict recognition requirement which would leave an unnecessary tag containing LPXTG after the ligation. This feature would probably alter the biological behaviors of the proteins, being a deterring element to its application in biomedical research.



Scheme 3-6. Reversibility of the Srt A-mediated ligation.

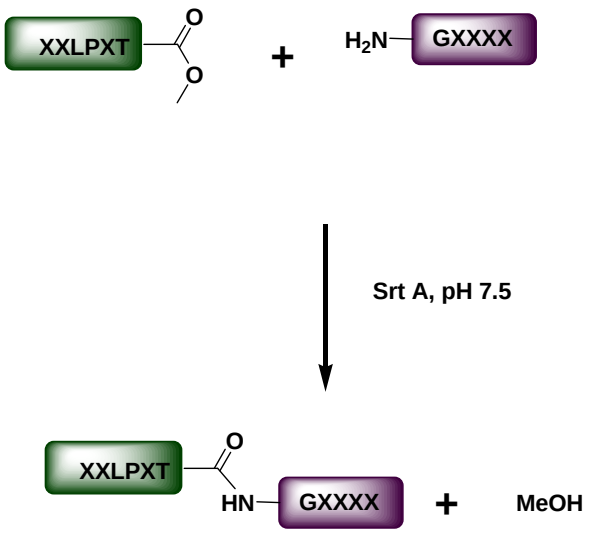
To deal with the reversibility of Srt A, several methods have been developed and achieved with some success.

Nagamune's group reported a method in which a rigid secondary structure like β hairpin was introduced around the ligation site in the ligation product. The rationale behind this design is that the secondary structure cannot be recognized by Srt A anymore, making the ligation product a dead end of the reaction.[132]



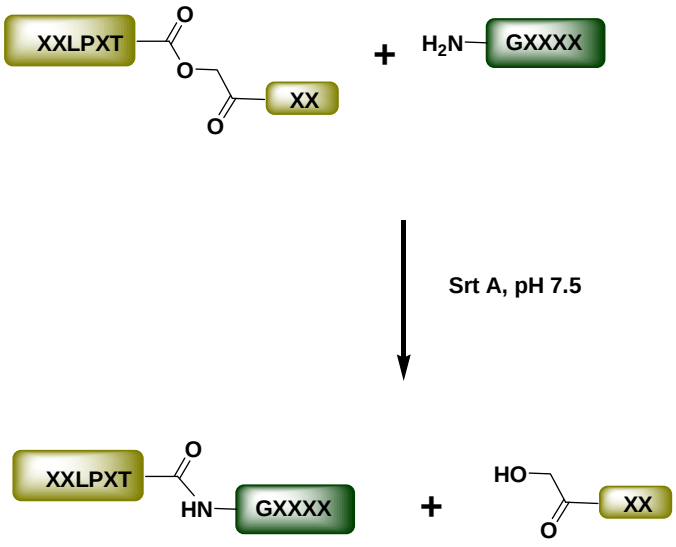
Scheme 3-7. Improvement of yield by introducing a secondary structure in ligation product.

Another strategy came from Ploegh's group. Instead of using the standard LPXTG motif in the electrophile labeling peptide, they used a C-terminal methyl ester (LPXT-OMe). MeOH rather than a glycine peptide was obtained as the side product after the ligation. Since MeOH was not a substrate of Srt A, the reversibility problem could be avoided.[133]



Scheme 3-8. Use of C-terminal methyl esters for improvement of ligation yields.

A similar strategy was adopted by Turnbull's group. Depsipeptide was used as the electrophile and the hydroxyl side product generated after the ligation was unrecognizable by sortase. In this way, the ligation efficiency could be improved.[134]



Scheme 3-9. Sortase A ligation using depsipeptide substrates.

Liu lei's group also developed an enzymatic method to irreversibly hydrazinolize proteins by coupling proteins with hydrazine and hydrazine derivatives in the catalysis of Srt A.[135] The description of this work would be placed in chapter 4.

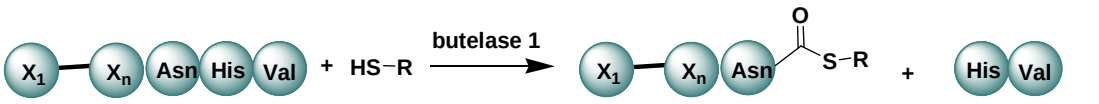
Although the reversibility could be solved by artificial designs, the "unnecessary tag" issue is much more difficult to deal with since it is associated with the nature of Srt A. Another limitation of Srt A is the low catalytic efficiency. Usually a long incubation time (up to 24 h) and a high molar ratio of enzyme is needed (typically from 0.1 to 1). Thus a ligase with broad substrate specificity and high catalytic efficiency would be highly desirable. In this aspect, we found that butelase 1 was the best candidate so far since it only leaves one residue Asx (Asn/Asp) after ligation.

Similar to Srt A, the butelase 1–mediated ligation is also reversible although butelase 1 has a much higher kinetics efficiency. In order to drive the reaction to completion, usually 10-fold excess of substrate needs to be added. This would be undesirable when the substrate is valuable. A technology that enables us to improve the butelase 1 ligation yield would be practically useful in research and worth our pursuing.

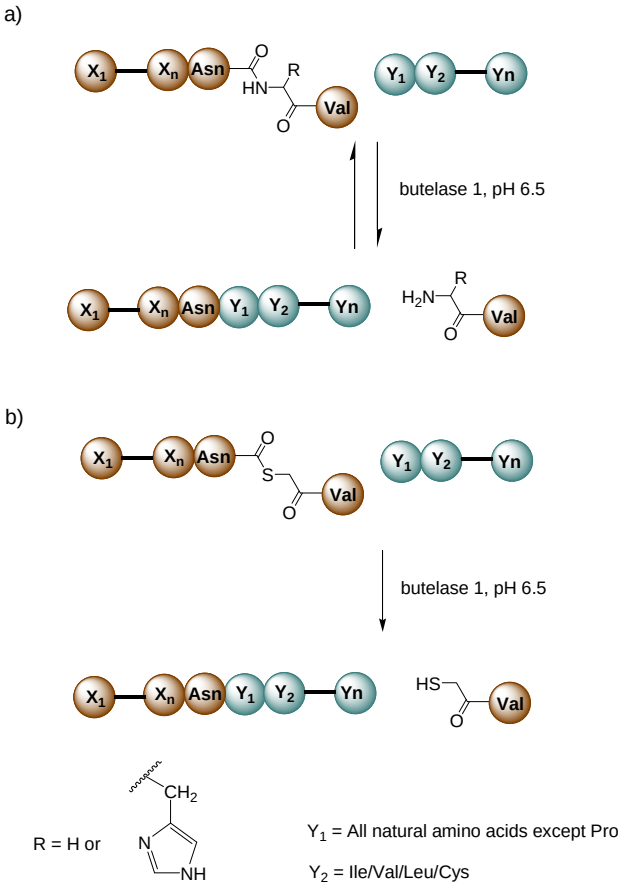
3.2 Results and discussion

3.2.1 Design rationale

In this chapter, a new method is presented which enables great improvement of butelase 1 ligation efficiency by using thiodepsipeptides. This design is based on our previous work that incubation of a model peptide KALVINHV with different kinds of thiols in the presence of butelase 1 did not lead to any peptide thioester product as we expected **(Scheme 3-10)**. This suggested that the mercaptan group might not be a recognizable acceptor nucleophile by butelase 1. We envisioned that if the scissile asparaginyl amide bond is replaced by the thioester bond, the reaction could be made irreversible since after the ligation a thioacetyl byproduct instead of a native dipeptide was generated **(Scheme 3-11)**.



Scheme 3-10. Initial proposed scheme for butelase-mediated thioesterification.



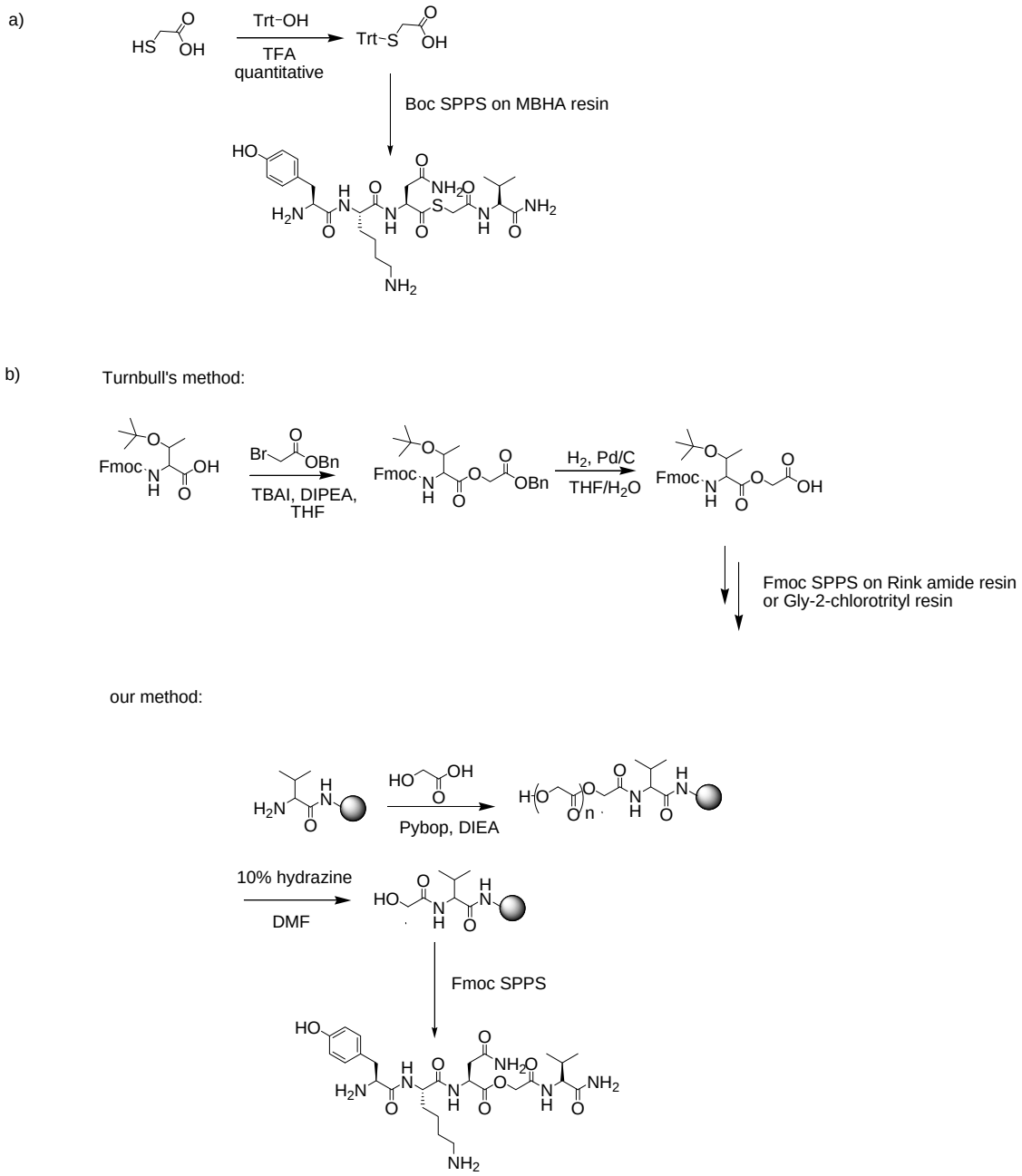
Scheme 3-11. Butelase 1 ligation using a) natural peptide and b) thiodepsipeptide.

3.2.2 Synthesis of thiodepsipeptides and depsipeptides

To support our hypothesis, we synthesized four different types of peptide substrates with the common sequence YKNXV (X = His, amino-, thio- or hydroxyacetic acid): a thiodepsipeptide 1, a depsipeptide 2 and two peptides 3 and 4 **(Table 3-1)**. For comparison we synthesized the depsipeptide 2 because Turnbull's group has reported good improvement on Srt A-mediated ligation by using a depsipeptide substrate.

The synthesis of thiodepesipeptide lies on the preparation of the building block - trityl-protected thioglycolic acid. First, mixture of thioglycolic acid with equal molar amount

of triphenylmethanol in the presence of neat trifluroacetic acid yielded trityl-protected thioglycolic acid after solvent removal. This compound was further used in SPPS without further purification. Thiodepsipeptides were synthesized on MBHA resin with Boc-protected amino acids and the pre-made building block by standard Boc chemistry. For depsipeptide 2, unlike the previously reported method by Turnbull's group which required a nonstandard dipeptide inaccessible by solid phase, we developed a fully solid-phase compatible and straightforward approach. The depsipeptides were synthesized on Rink amide MBHA resin with Fmoc amino acids and glycolic acid. Hydrazinolysis after coupling the glycolic acid cleaved all over-coupled glycolic acids linked through susceptible ester bonds while leaving only one glycolic acid attached to the previous residue with stable amide bond. Our synthesis scheme provides a new convenient access to depsipeptides.



Scheme 3-12. Synthesis of thiodepsipeptides and depsipeptdes.

a) Synthetic strategy for thiodepsipeptide. b) Preparation of depsipeptide by Turnbull's method and our method.

Table 3-1. Number of the peptide substrates used in this project and the corresponding molecular weights.

Number	Sequence	MW (M + H) $^+$	
		calcd	obsd
1	YKN-thioglc-V-NH_2	596.28	596.28
2	YKN-glc-V-NH_2	580.3	580.19
3	YKNHV-NH_2	659.36	659.27
4	YKNGV-NH_2	579.32	579.16
5	GIGGIR-NH_2	571.36	571.3
6	biotin-TYKN-thioglc-V-NH_2	923.4	923.31

3.2.3 Time-course of butelase-mediated ligation of four substrates with a model peptide GIGGIR

We compared the ligation efficiency of each of the four substrates with a short model peptide GIGGIR 5. The reactions were performed in the presence of 100 μM GIGGIR, one or two molar equivalents of each peptide substrate and 50 nM of butelase 1 (0.0005 molar equivalent). Time-course analysis showed that thiodepsispeptide 1 was the most reactive substrate followed by depsipeptide 2. Peptide 3 and peptide 4 were the least reactive. All the ligation reactions reached equilibrium after 60 min with marginal improvement after 90 min. At two molar equivalent of 1, more than 95% of peptide 5 was converted into the desired ligation product within one hour. The ligation yields only reached about 31%, 59% and 68% for substrate 4, 3 and 2 in the same reaction

condition. Even at one molar equivalent of thiodepsipeptide 1, a satisfactory 73% ligation yield could be obtained. This confirmed that thiodepsipeptide was the best substrate of the four for butelase 1-mediated ligation.

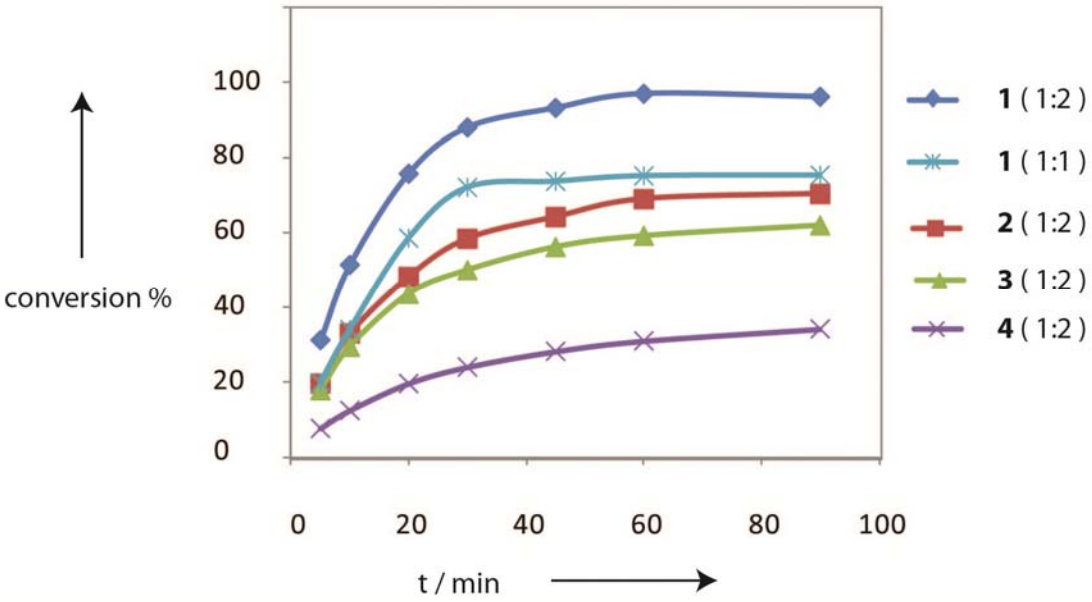


Figure 3-1. Time-course of butelase-mediated ligation of four substrates and 5. The reaction was monitored by HPLC at conditions: 50 nM butelase 1, 100 μM peptide 5, corresponding ratios of different nucleophilic acyl donor, 20 mM phosphate buffer 1 mM EDTA pH 6.5, 42 ^{o}C.

3.2.4 Kinetics parameters measurement of the four substrates

We then conducted a kinetics study to quantify the difference among these four different substrates. The result is consistent with the time-course experiment showing that thiodepsipeptide 1 has the highest catalytic efficiency (k_{cat}/K_m) and peptide 4 has the lowest value. The efficiency of an enzyme to one particular substrate is generally related with the K_m value which indicates the binding affinity of the enzyme-substrate

and the k_{cat} value measuring the conversion of the acyl-enzyme intermediate to the final product. Thiodepsipeptide 1 has a higher K_m value than peptide 3, indicating its weaker binding with the enzyme. This could be rationalized since peptide 3 has the natural recognition sequence Asn-His-Val while the thiodepsipeptide 1 has an unnatural motif. The K_m values for depsipeptide 2 and peptide 4 are also higher than that of peptide 3. This suggests that His-Val is the best recognizable motif by butelase 1 and also explains its high sequence conservation in the natural-occuring cyclotide precursors. Although thiodepsipeptide 1 has a weak binding affinity, this is compensated by a high turnover rate k_{cat} which is about 14 times of peptide 3. The higher k_{cat} value of thiodepsipeptide could be reasoned that the thiol byproduct released after the ligation cannot react with the acyl-enzyme intermediate, while for the peptides 3 and 4, the dipeptides (His-Val and Gly-Val) released are still recognizable nucleophiles in the reverse reaction direction.

peptide	sequence	$k_{cat}[s^{-1}]$	K_m[mM]	$k_{cat}/K_m[M^{-1}s^{-1}]$
1	YKN-thioglc-V	56.5±7.5	2.76±0.0075	20480
2	YKN-glc-V	12.4±0.32	0.74±0.01	16840
3	YKNHV	4.1±0.65	0.3±0.01	13490
4	YKNGV	7.9±0.72	0.88±0.08	8900

Table 3-2. Kinetic parameters of different peptide substrates.

3.2.5 Ligation with the poorly recognized nucleophiles using thiodepsipeptide

Given the fact that thiodepsipeptde gave the best performance, we went on to investigate whether its use could facilitate the ligation with some poor nucleophiles.

We prepared 8 different peptides which, in our previous study, ligated poorly with the model peptide KALVINHV **(Table 3-3)**. The result showed thiodepsipeptides gave no obvious improvement compared with normal peptides. This probably indicates that in these ligations, step two (acylation of the acyl-enzyme intermediate by the nucleophile) is the rate-limiting step. The poor ligations are associated with the poor recognitions of these nucleophiles by the enzyme.

Table 3-3. The sequences of the eight "difficult" nuclophiles.

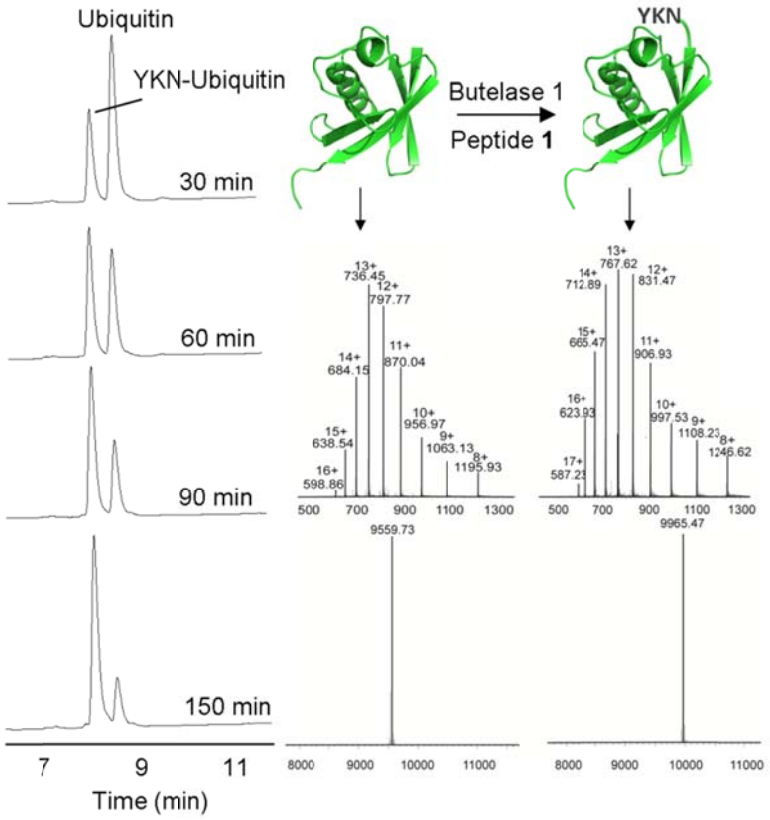
entry	peptide sequence
1	DIGGIR
2	EIGGIR
3	TIGGIR
4	VIGGIR
5	LAGGIR
6	LFGGIR
7	LMGGIR
8	LTGGIR

3.2.6 Protein labeling using thiodepsipeptide

To show that our method could also be applied to protein labeling at the N-terminus, we prepared two recombinant proteins as our study models: ubiquitin and green fluorescent protein (GFP). To make the proteins recognizable by butelase 1, we added a dipeptide motif Gly-Ile at the N-terminus of ubiquitin and modified GFP with an N-terminal Met-Ile motif. For ease of purification, the recombinant ubiquitin contains a His-tag at the C-terminus and the GFP has a His-tag following the N-terminal dipeptide

motif. Both proteins were labeled with the model thiodepsipeptide 1 catalyzed by 0.001 molar equivalent of butelase 1. The labeling of proteins was slower than peptide. This is most likely to be caused by two factors: 1 the big difference between the size of proteins and that of peptide 2 the N-terminus of protein is much more steric hindered than peptide thus being less accessible to butelase 1. Also, we found that the half life of thiodepsipeptide was quite short mainly due to hydrolysis and aspartimide formation which was associated with the nature of the asparagines residue. This feature poses a problem when the ligation proceeds slowly. In fact, a moderate yield (~60%) could be obtained after 5 equivalents of thiodepsipeptide 1 being completely consumed. Due to the instability of thiodepsipeptides, instead of adding all labeling reagent at once in the beginning, we added one molar equivalent every thirty minutes. This strategy enabled 82% of ubiquitin to be labeled with five molar equivalents of thiodepsipeptide after 2.5 h as indicated by HPLC profile **(Figure 3-2 a)**. The same strategy was applied to GFP. Since the labeling reaction created no shift on HPLC, ESI was used to monitor the reaction. An estimated 70% yield was obtained with four equivalents after 2.5 h **(Figure 3-2 b)**. It is worth mentioning that butelase 1 has been reported for peptide ligation so far and we set the first example to use butelase 1 for protein labeling.

a)



b)

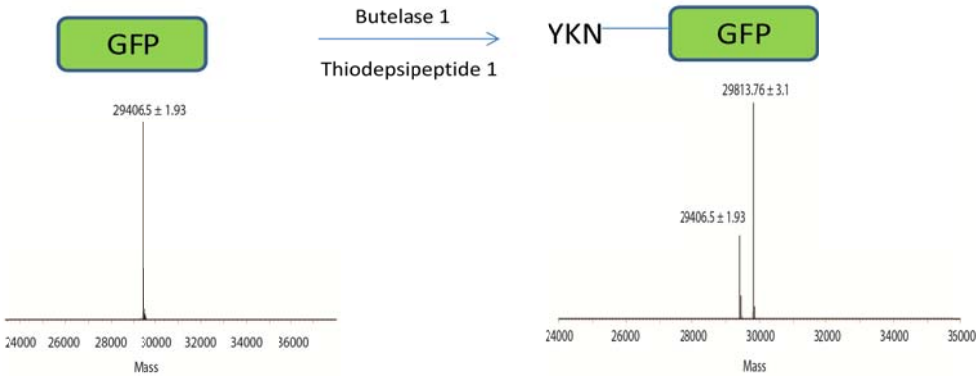


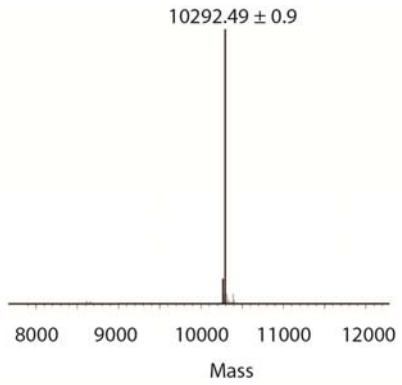
Figure 3-2. Labeling of ubiquitin and GFP.

a) Labeling of ubiquitin by thiodepsipeptide 1. Ubiquitin before labeling: calculated (M + H): 9559.91, observed (M + H): 9559.73. Ubiquitin after labeling: calculated (M + H): 9965.3, observed (M + H): 9965.47. b) Labeling of GFP with thiodepsipeptide 1.

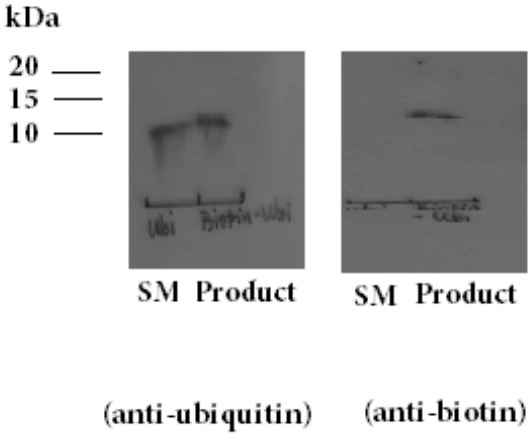
GFP before labeling: calculated (M + H): 29404.59, observed (M + H): 29406.5 ± 1.93.
GFP after labeling: calculated (M + H): 29809.79, observed (M + H): 29813.76 ± 3.1.

With the success on the model thiodepsipeptide, we investigated conjugation of proteins with biological functional groups. A small thiodepsipeptide 6 carrying a biotin at its N-terminus **(Table 3-1)** was synthesized by standard Boc chemistry. Labeling was successfully done on both ubiquitin and GFP using the same strategy described above. Western blot and ESI data demonstrated the identity of the biotin-labeled protein products.

a)



b)



c)

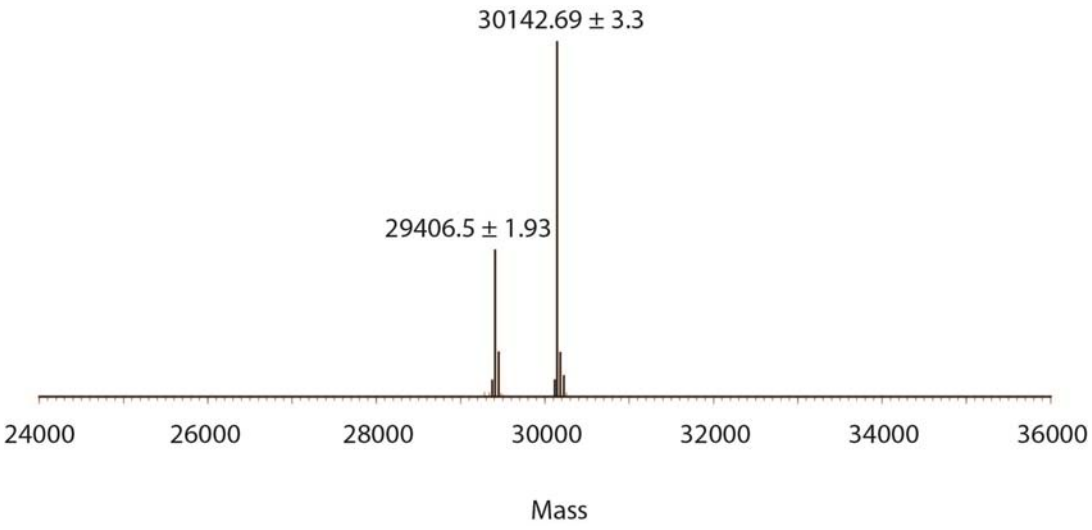


Figure 3-3. Labeling of both ubiquitin and GFP with biotin-bearing substrate.

a) The deconvoluted MW of biotin-labeled ubiquitin (calculated (M + H): 10292.43, observed (M + H): 10292.49 ± 0.9). b) Western blot of the unlabeled ubiquitin and biotin-labeled ubiquitin. Left panel: treatment with mouse-generated anti-ubiquitin antibody. Right panel: treatment with anti-biotin, HRP-coupled antibody. c) The deconvoluted MW after biotin-labeling of GFP. biotin-GFP (calculated (M + H): 30136.92, observed (M + H): 30142.69 ± 3.3)

3.3 Conclusion

We have developed a new method to improve the efficiency of butelase 1-catalyzed protein N-terminus labeling by using a thiodepsipeptide as the labeling reagent. This new method has two main advantages. First is that the preparation of the thiodepsipeptide is operationally easy. The second advantage is that our method leaves only one amino acid residue Asn which is much shorter than that by Sortase A-mediated technology ("LPXTG"), there would be less alterations on the structure of the labeled proteins. We believe that this method can act as a good tool to label proteins of interest in biological research.

3.4 Materials and methods

3.4.1 Materials

The sources of chemicals are the same as stated in chapter 2. For molecular biology part, all the restriction enzymes for plasmid construction were purchased from New England Biolabs (NEB).

3.4.2 Instrument

The analytical HPLC analyses were carried out on NexraX_2 LC-30AD with an analytical column (Aeris peptide XB-C18, 4.6 x 250 mm). The purification was performed using Shimadzu HPLC equipped with a semi-preparative HPLC column (Jupiter C18, 5 μM, 10 x 250 mm). All the HPLC runs were done using the mixture of two solvents, A and B, as the mobile phase. Solution A was H_2O containing 0.045% TFA and solution B was acetonitrile/H_2O (90:10) containing 0.039% TFA. The flow rate was 0.3 ml/ min for analysis and 2.5 ml/min for purification. UV detection was carried out at 220 nm.

Small peptides ESI mass spectra data were obtained on a Thermo Finnigan LCQ DECA XP MAX with ESI ion source. Protein ESI mass spectra were measured on Q Exative hybrid Quadrupole-Orbitrap Mass Spectrometer. The deconvoluted data was obtained using the software of MegTran 1.03 and ESIProt 1.0.

3.4.3 Synthesis of trityl-protected thioglycolic acid

Triphenylmethanol (805 mg, 3 mmol) was mixed with thioglycolic acid (208.4 μl, 3 mmol) in the presence of neat TFA (4 mL). The reaction was stirred at r.t. for 30 min. After removal of TFA in vacuo, tolune was added to removal the residual water. After complete removal of solution in the reaction mixture, desired product was obtained as white powder in quantitative yield and it was used in the following SPPS without further purification.

3.4.4 Solid phase peptide synthesis

General procedure for synthesis of thiodepsipeptides

To prepare thiodepsipeptides, standard Boc chemistry was used. MBHA resin was used for the synthesis. For the coupling of amino acids, the Boc protected amino acid (4 eq.), pybop (4 eq.) was dissolved in a mixture of DCM and DMF. The solution was first added to the resin followed by the addition of 9 equivalent of DIEA. The reaction is monitored by ninhydrin test and normally takes 1.5 hour to finish. To remove the Boc protecting group, 30% TFA in DCM is employed (2 min, 20 min). Between each coupling/deprotection, repeated DCM/DMF washes are needed. The pre-synthesized trityl-protected glycolic acid was loaded to resin in the same fashion as normal Boc amino acids. To remove the trityl protecting group, TIS is added along with 30% TFA in DCM. The deprotection process could be monitored by the color change of the solution. The disappearance of yellow color indicates the removal of trityl group. For cleavage (scale of 250 mg resin), first add 750μl thioanisole/ethanedithiol mixture (2:1) on the ice, then add 5 ml TFA followed by 500 μl TFMSA dropwise adding. Cleavage

usually takes one hour at room temperature. The crude peptide was further purified by HPLC. The HPLC collections were lypholized to white powder form for stock or reaction use.

Synthesis of depsipeptide 2

The first amino acid Val was coupled to rink amide MBHA resin using standard Fmoc chemistry. After removal of Fmoc protecting group with 20% piperidine, a mixture of 4eq glycolic acid, 4eq pybop and 8eq DIEA was added. The reaction was shaked for two hours when Kaiser test showed complete coupling. 10% hydrazine in DMF was then added to the resin to remove the over-coupled glycolic acid. The reaction was undertaken for 30 min. The next amino acid Asn was coupled following the below steps:

1 Suspend the resin in DCM/DMF mixture (9:1 v/v)

2Dissolve 5eq Fmoc amino acid in a minimum amount of DMF. Add 5eq HOBT to the solution. Vortex and sonicate the mixture until all the reagents were totally dissolved. Add this solution to the resin.

3 Add 5eq DIC to the resin

4 Dissolve 0.1eq DMAP in a minimum amount of DMF and then add to the reaction. Let the reaction shake for overnight.

5 Block the uncoupled hydroxyl groups using 5% acetic anhydride dissolved in pyridine, shake the reaction for 20 min.

For the following sequence, use the standard Fmoc chemistry as described above.

3.4.5 Preparation OF the engineered proteins

Protein sequences of engineered ubiquitin and GFP

GI-Ubi-His_6:

GIMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGR
TLSDYNIQKESTLHLVLRLRGGHHHHHH

MI-His_6-GFP:

MIHHHHHHSGVDLGTENLYFQSMSKGEELFTGVVPILVELDGDVNGHKFSVSG
EGEGDATYGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFF
KSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILG
HKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDG
PVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGMDELYK

Genetic sequences of both proteins

GI-Ubi-His_6:

ATGGGCATTATGCAGATCTTCGTCAAGACGTTAACCGGTAAAACCATAACT
CTGGAAGTTGAACCATCCGATACCATCGAAAACGTTAAGGCTAAAATTCAA
GACAAGGAAGGAATTCCACCTGATCAACAAAGATTGATCTTTGCCGGTAAG
CAGCTCGAGGACGGTAGAACGCTGTCTGATTACAACATTCAGAAGGAGTCG
ACCTTACATCTTGTCTTAAGACTAAGAGGTGGTCATCATCATCATCATCATT
GA

The plasmid of engineered GFP was received as a gift from Dr. Tobias Cornvik.

As for engineered ubiquitin, three residues MGI were added to its N terminus and six histidines were added to the C terminus by using the primers: 5’CGCCATATGGGCATTATGCAGATCTTCGTCAAGAC3’ (forward) and

5’CGCGGATCCTCAATGATGATGATGATGATGACCACCTCTTAGTCTTAAGA CAAG3’ (reverse). It was then subcloned into plasmid pET3b. The plasmid and the insert amplified by PCR all contained restriction endonuclease sites for Nde I and BamH I. They were digested by both of these two enzymes (New England Biolab) at37 °C and then ligated using T4 DNA ligase (New England Biolab) at 37 °C for 1 hour. The ligation mixture was then transformed into *E. coli* DH5α strain competent cells. Positive clone was screened by digestion with Nde I-BamH I. The sequence was confirmed by DNA sequencing.

General procedure for expression and purification of soluble histagged-protein

1 µl of constructed plasmids was transformed into competent *E. coli* strain BL21 by 90 seconds heat shock in 42 ^{o}C water bath and a subsequent 5 min cool-down on ice. Around 1 ml LB was added into the transformed cells. The mixture was shaked at 37 ^{o}C for one hour. Centrifuge down the cells and discard ~900 µl LB supernatant. Apply the mixture of the cells with the residual 100 µl LB onto a prewarmed LB/ampicillin (final concentration of 100 mg/L) plate. The plate was incubated in 37 ^{o}C for overnight. Colonies grown on the plate were picked and inoculated in LB/Amp culture for overnight primary amplification. 1 ml culture was added to 100 ml LB/Amp in a 2 L flask and was shaked at 37 ^{o}C (4 hours, 220 rpm). A final concentration of 0.3 mM IPTG was added to induce the expression of proteins. The culture was then shaked for another 4 hours. Cells were harvested from the culture by centrifuge (10 min, 10000 rpm). Discard the supernatant and resuspend the cells with PBS. The cell pellets were lysed using microfluidizer. The lysate was spinned down (30 min, 15000 rpm), leaving the proteins in the supernatant. The supernatant was taken out and mixed with Nickel

nitrilotriacetic acid agarose beads (Qiagen) pre-equilibrated in PBS loading buffer. The target protein would bind to the resin while the other unwanted proteins would be simply removed by filtration. The protein of interest was then eluted out using the elution buffer containing 250 mM immidazole. The eluent was dialyzed in the desired stock buffer for overnight at 4 ^{o}C and concentrated to the desired stock concentration with centrifugal filter.

3.4.6 Procedure of kinetics parameters determination

Kinetics constants for all peptide substrates were obtained from initial reaction rate measurements of the ligation reaction.

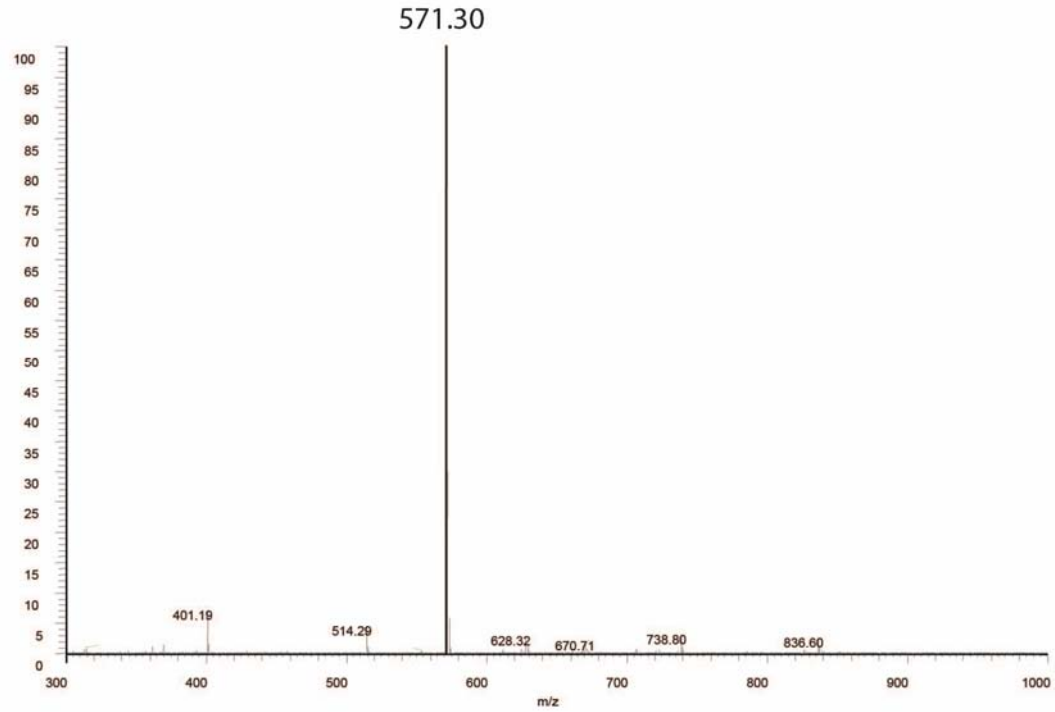
Ligation assays were performed in 50 μl mixtures containing reaction buffer (1 mM EDTA, 20 mM phosphate buffer, pH 6.5), 0.1 μM butelase 1, 1 mM GIGGIR and varying concentrations of peptides **1-4** (25 to 400 μM). The reactions were performed in triplicate at 42 ^{o}C and quenched by adding 5 μl of 1 M HCl solution. The peptides were separated by using a reversed-phase C18 analytical column (150 x 2.1 mm, Vydac) with a linear gradient from 5% to 40% acetonitrile over 15 min on a Nexera UHPLC system (Shimadzu). The ligation velocities were calculated by converting the HPLC-peak areas of remained linear precursors or the cyclized products into concentrations. The identity of each HPLC peak was analyzed by MALDI-TOF MS (ABI 4800 MALDI TOF/TOF). The kinetic parameters (*kcat* and *Km*) for each peptide were obtained by the Lineweaver–Burk plot.

3.4.7 Procedure of protein labeling with thiodepsipeptide

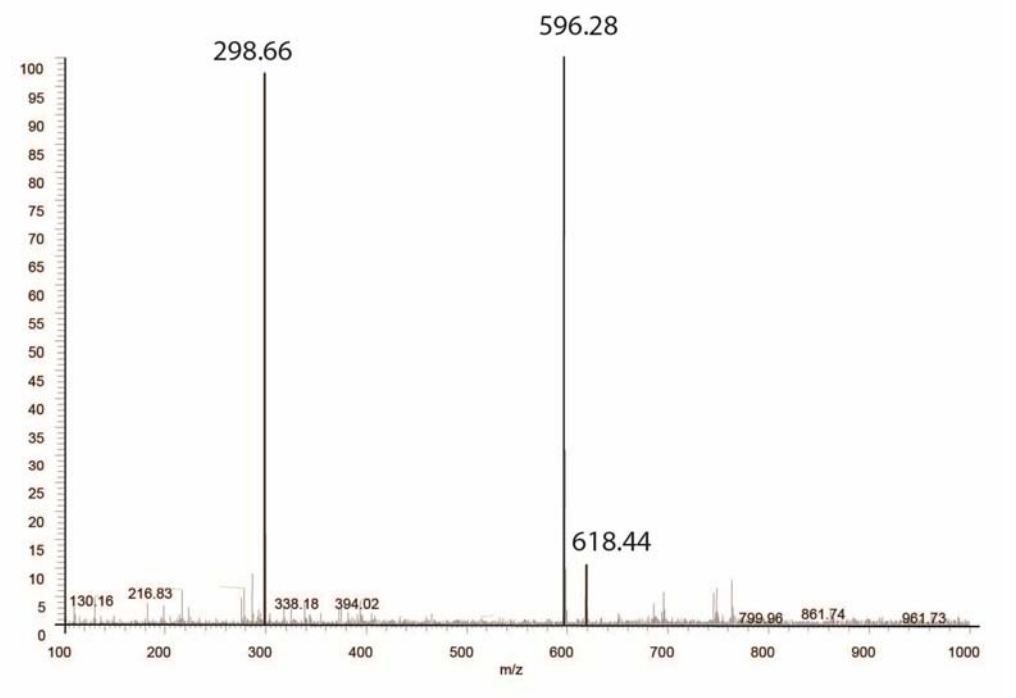
The reaction buffers (pH 6.5, 20 mM phosphate buffer for ubiquitin and PBS for GFP) were first preheated at 42 ^{o}C for 3 min. First mix 100 μM protein and equal molar amount of labeling reagent. Add 0.001 molar equivalent of butelase 1 at last. One equivalent of labeling reagent was added every 30 min. For ubiquitin, 5 equivalents of thiodepsipeptide were added in total, as for GFP, 4 equivalents were added. The process of ubiquitin labeling was monitored using HPLC while for GFP, ESI was employed since there was no shift on HPLC.

3.4.8 Mass spectra of peptides

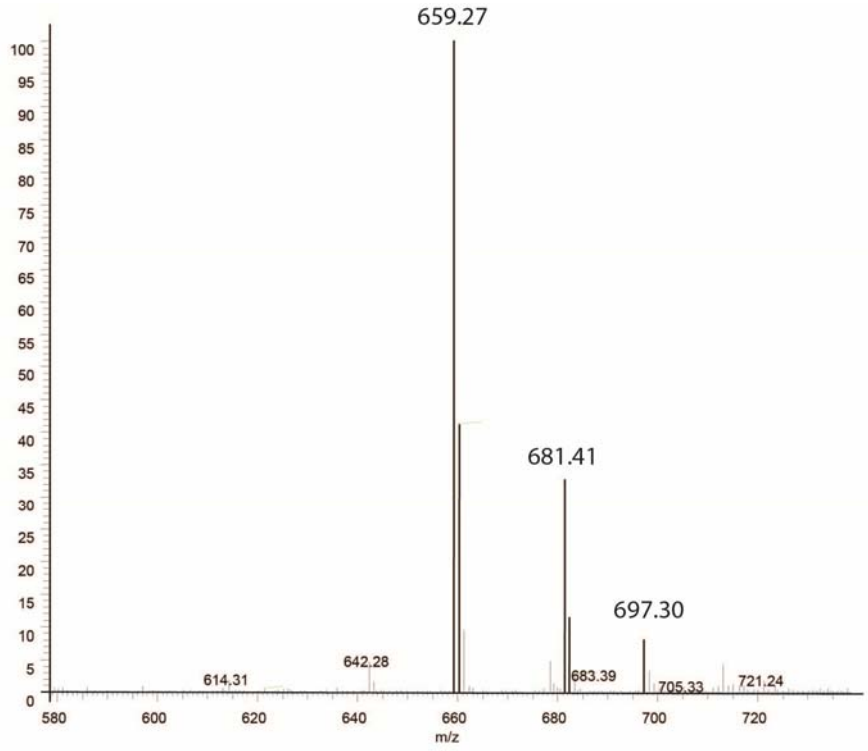
H-**GIGGIR**-NH_2 (*m/z* $[M + H]^+$ 571.30, calcd isotopic M_W 570.69)



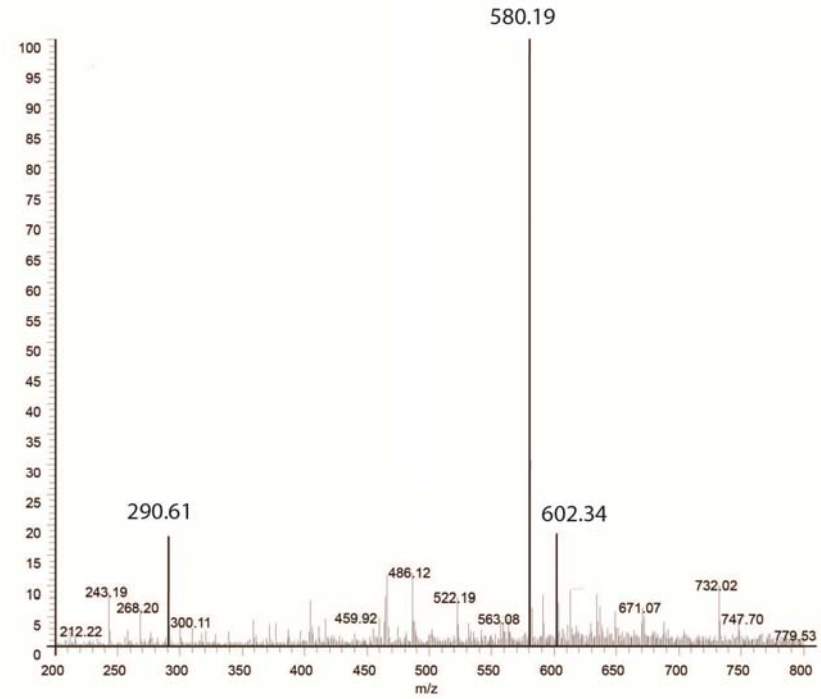
H-**YKN-thioglc-V**-NH_2 (*m/z* $[M + H]^+$ 596.28, $[M + Na]^+$ 618.44, $[M + 2H]^{2+}$ 298.66, calcd isotopic M_W 595.71)



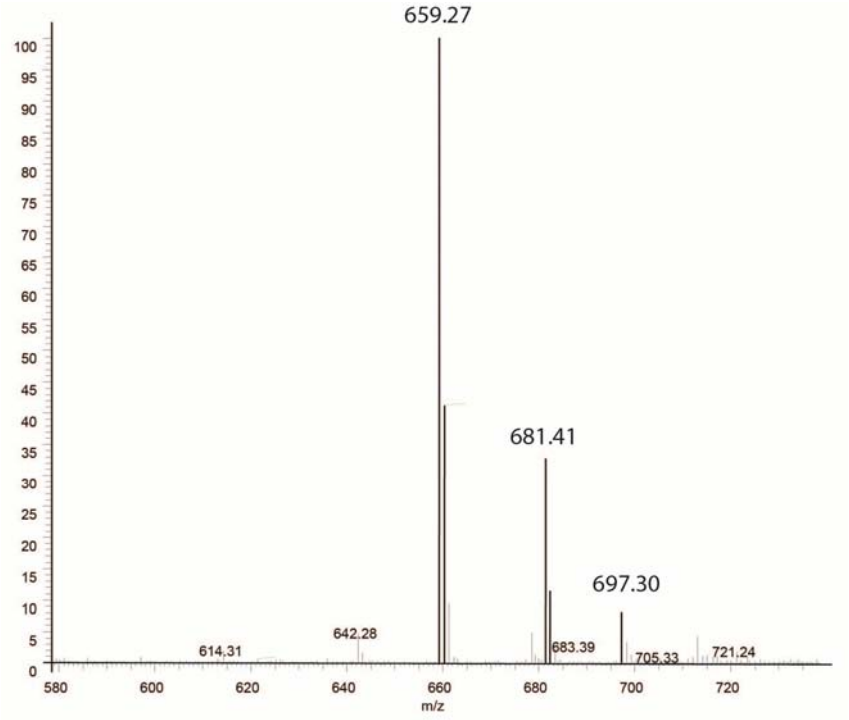
H-**YKNGV**-NH_2 (*m/z* $[M + H]^+$ 579.16, $[M + 2H]^{2+}$ 290.08, calcd isotopic M_W 578.66)



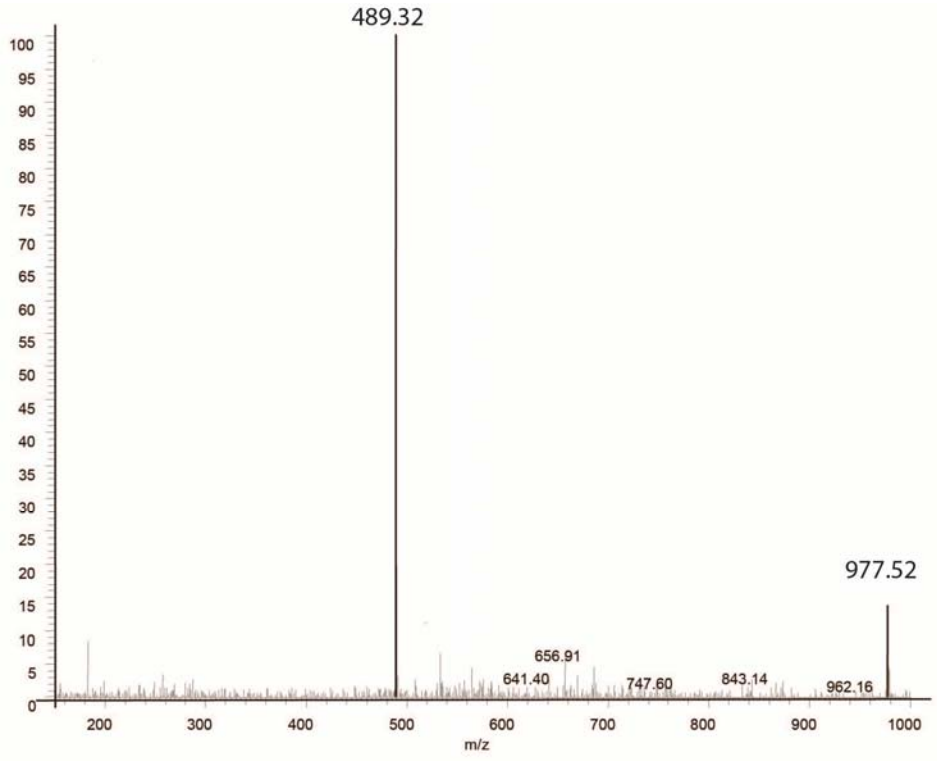
H-**YKN-glc-V**-NH_2 (*m/z* $[M + H]^+$ 580.19, $[M + Na]^+$ 602.34, $[M + 2H]^{2+}$ 290.61, calcd isotopic M_W 579.30)



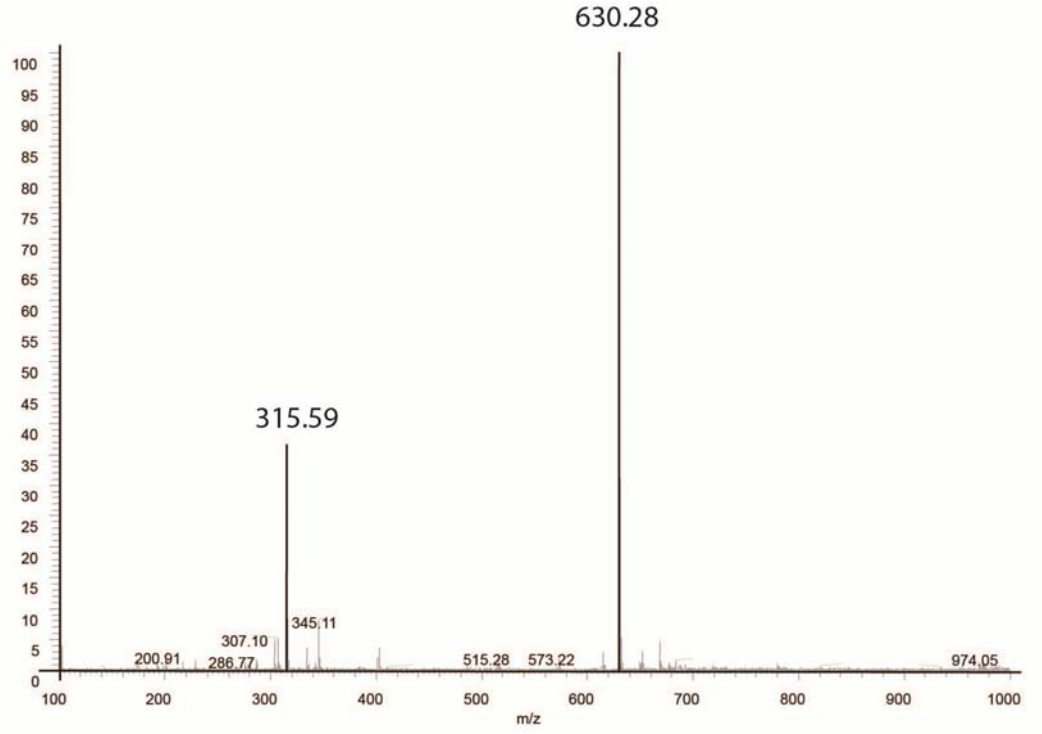
H-**YKNHV**-NH_2 (*m/z* $[M + H]^+$ 659.27, $[M + Na]^+$ 681.41, $[M + K]^+$ 697.30, calcd isotopic M_W 658.35)



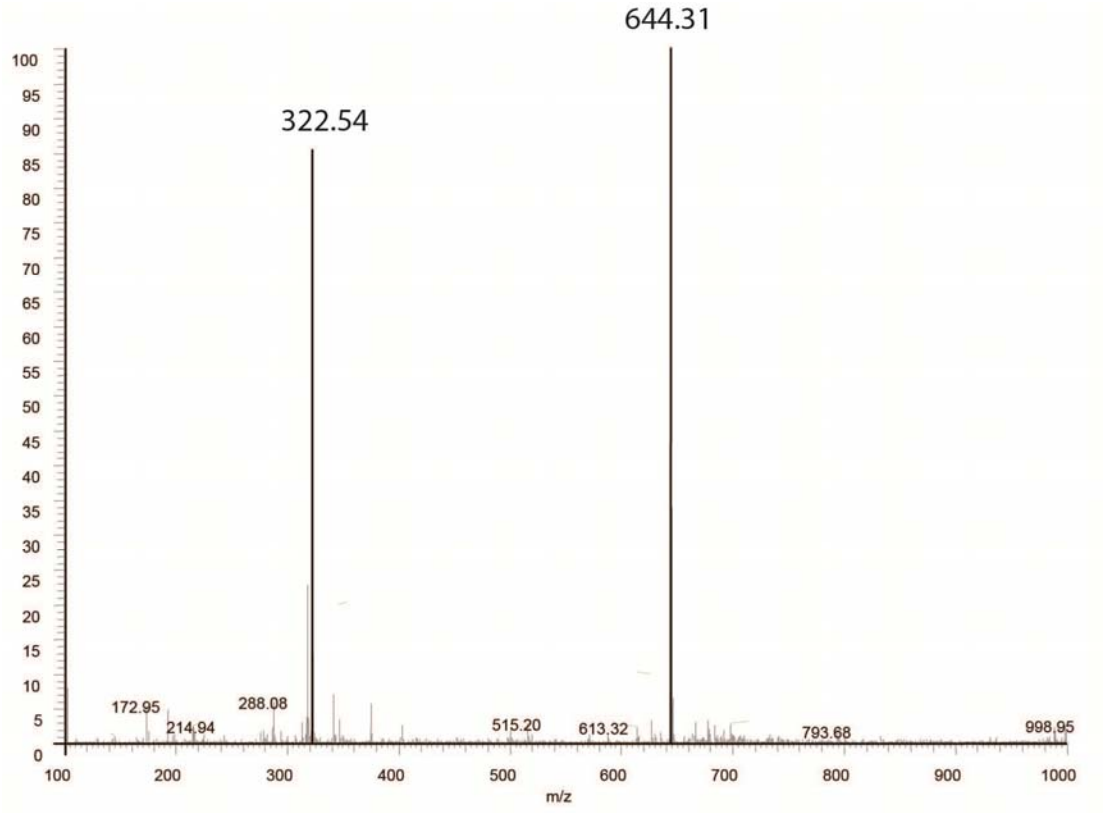
H-**YKNGIGGIR**-NH_2 (*m/z* $[M + H]^+$ 977.52, $[M + 2H]^{2+}$ 489.32, calcd isotopic M_W 976.56)



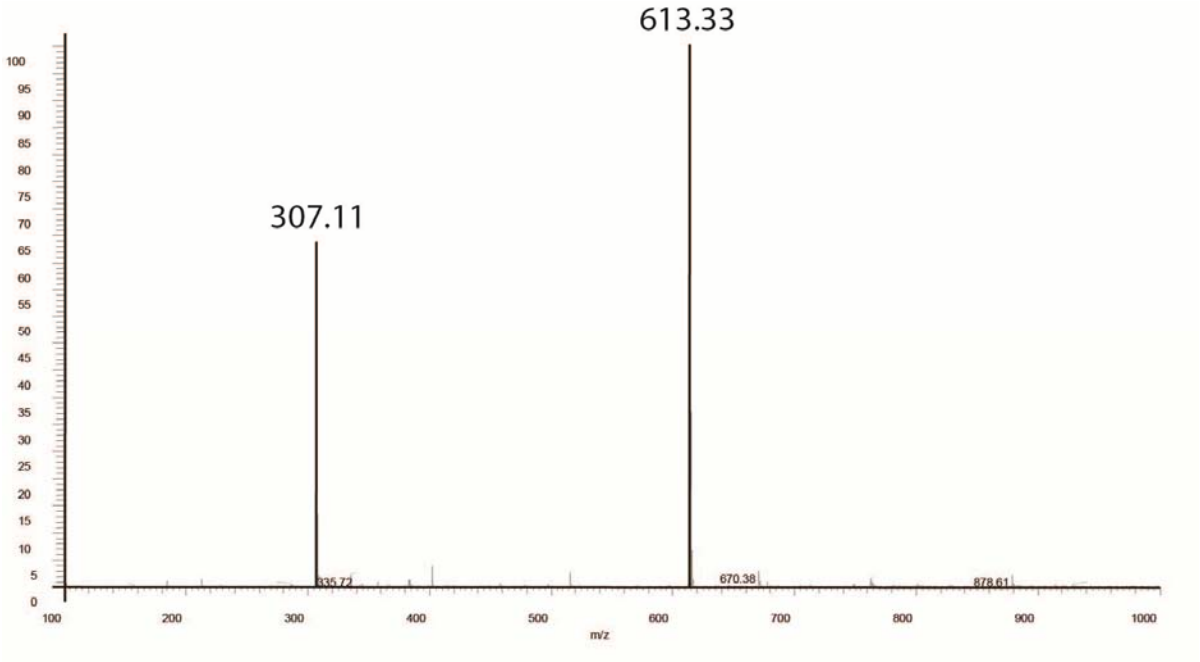
H-**DIGGIR**-OH (*m/z* $[M + H]^+$ 630.28, $[M + H]^{2+}$ 315.59, calcd isotopic M_W 629.71)



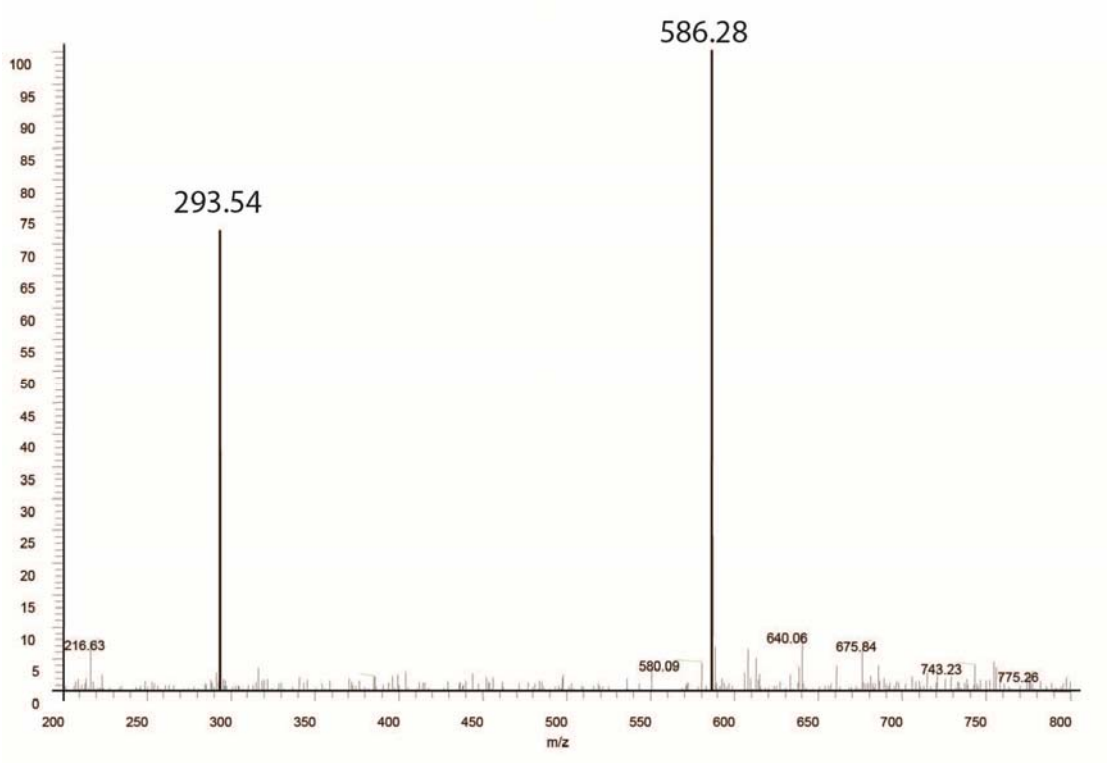
H-**EIGGIR**-OH (*m/z* $[M + H]^+$ 644.31, $[M + 2H]^{2+}$ 322.54, calcd isotopic M_W 643.73)



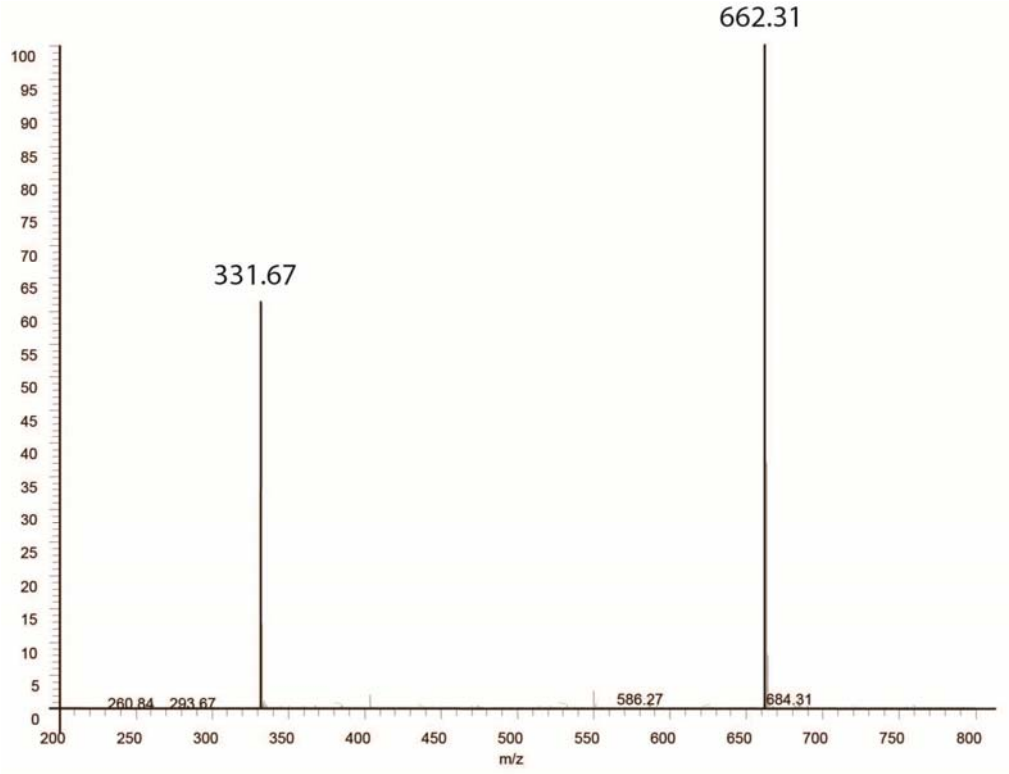
H-**VIGGIR**-NH_2 (*m/z* $[M + H]^+$ 613.33, $[M + 2H]^{2+}$ 307.11, calcd isotopic M_W 612.77)



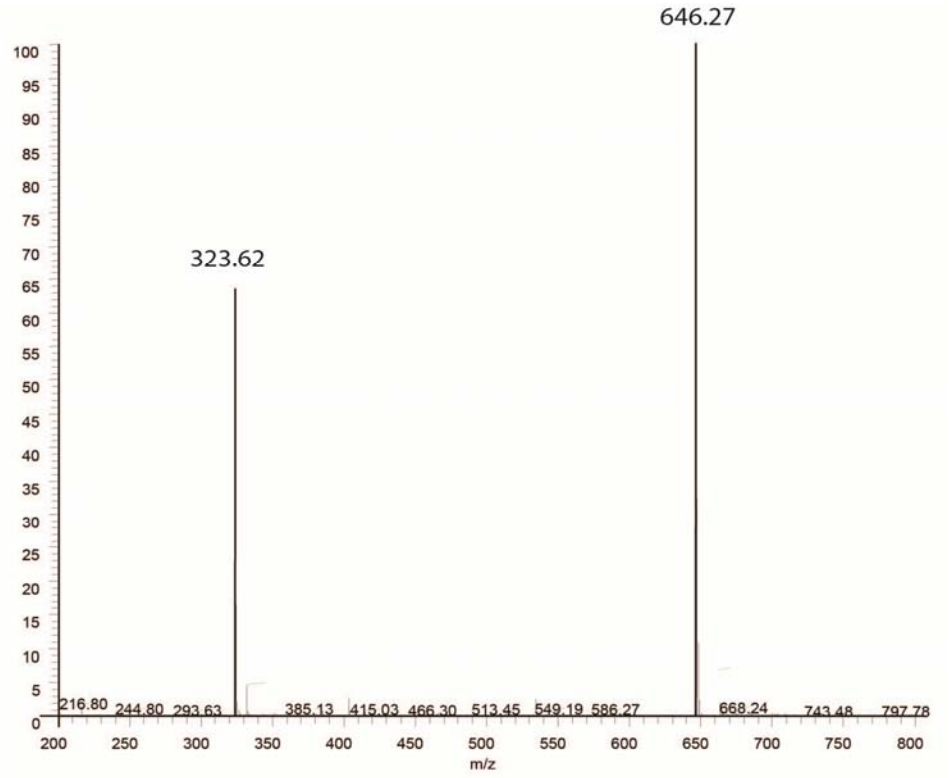
H-**LAGGIR**-OH (*m/z* $[M + H]^+$ 586.28, $[M + 2H]^{2+}$ 293.54, calcd isotopic M_W 585.70)



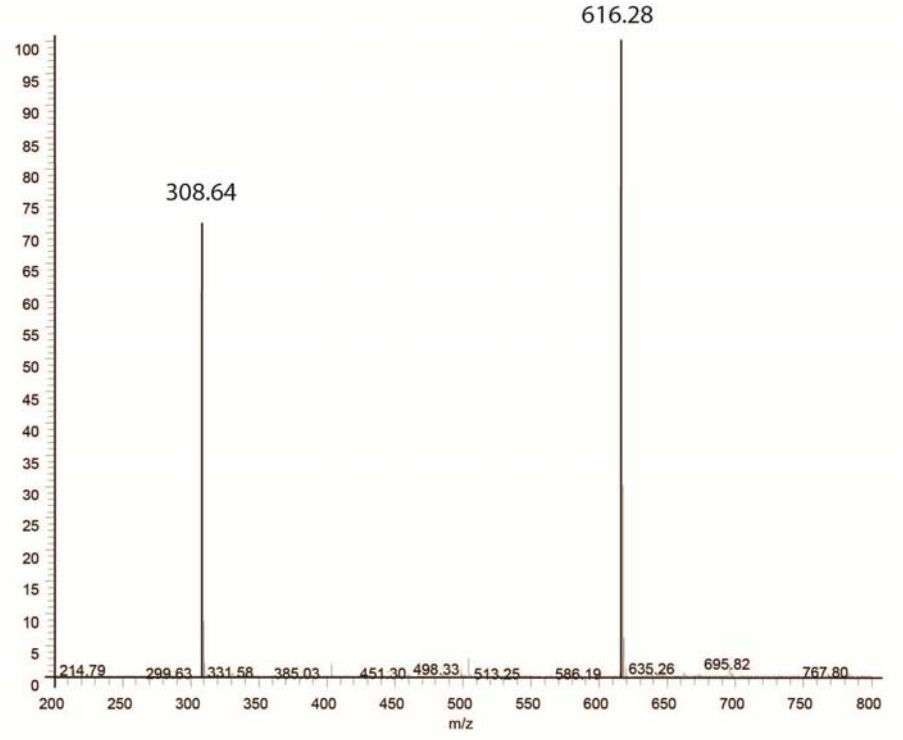
H-**LFGGIR**-OH (*m/z* $[M + H]^+$ 662.31, $[M + 2H]^{2+}$ 331.67, calcd isotopic M_W 661.79)



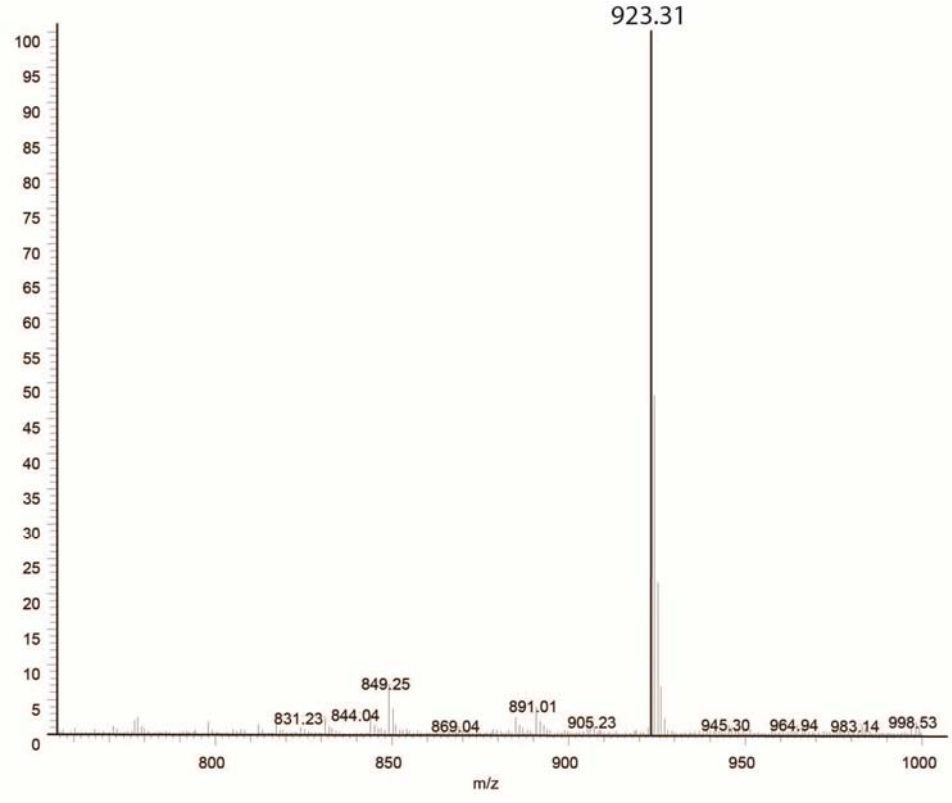
H-**LMGGIR**-OH (*m/z* $[M + H]^+$ 646.27, $[M + 2H]^{2+}$ 323.62, calcd isotopic M_W 645.81)



H-**LTGGIR**-OH (*m/z* $[M + H]^+$ 616.28, $[M + 2H]^{2+}$ 308.64, calcd isotopic M_W 615.72)



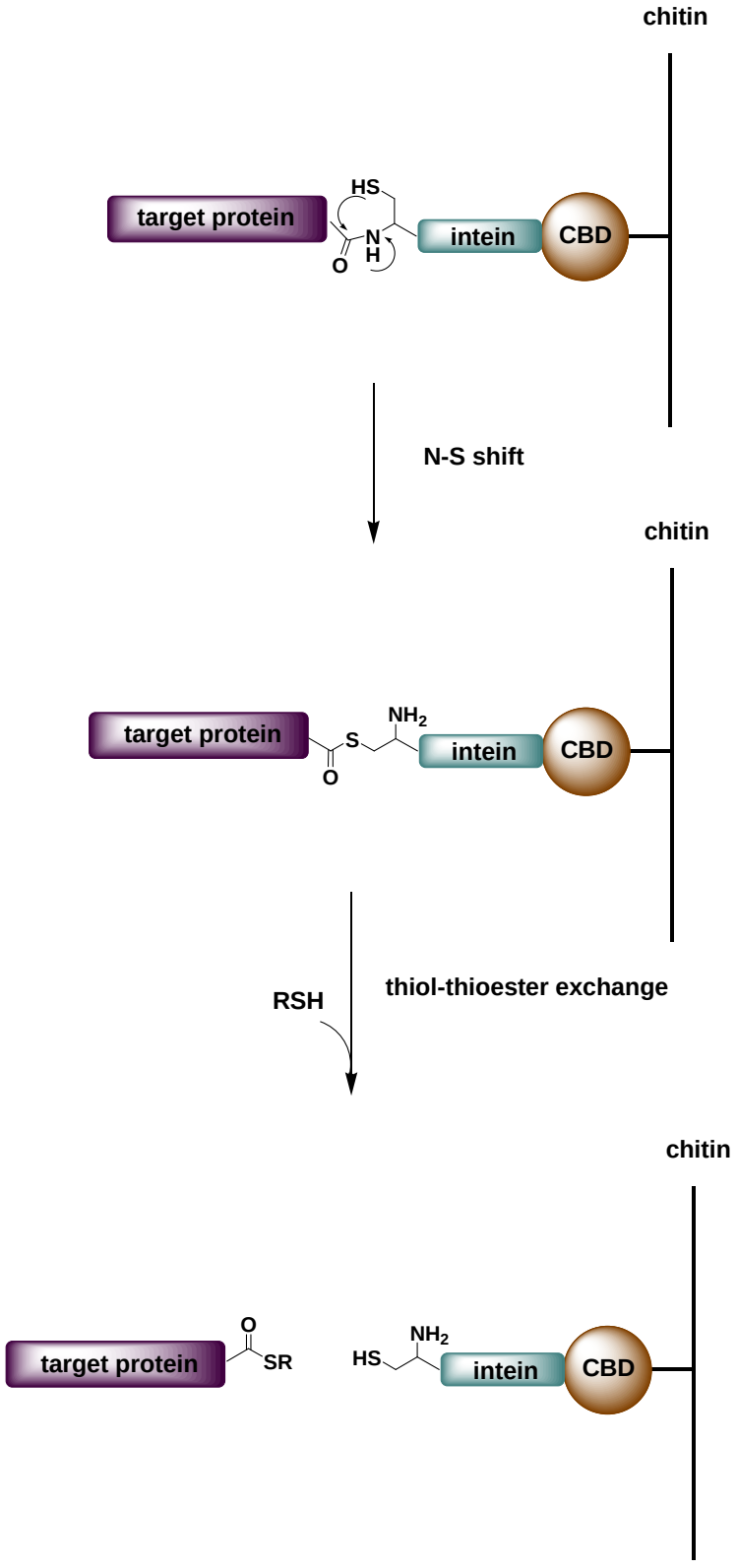
biotin-**TYKN-thioglc-V**-NH_2 (*m/z* $[M + H]^+$ 923.31, calcd isotopic M_W 922.11)



Chapter 4 A novel enzymatic approach to prepare protein thioesters and tandem chemoenzymatic ligation

4.1 Introduction

In chapter 2, a chemical access to peptide thioester was described. However, direct access to protein thioester via this method is not practical because of the size limitation of SPPS. Typically, protein thioesters are produced by genetic fusion of the target protein with an engineered intein comprising an Asn-Ala mutation. This mutation allows only the first step of protein splicing mechanism involving N to S acyl shift to happen at proper pH which affords a thioester linkage between the protein of interest and the intein. The following succinimide formation step would be prevented due to the missing of asparagine residue. For the convenience of purification, the engineered intein is associated with an affinity tag like a chitin binding domain (CBD). After the N to S acyl shift, addition of external thiol would release the protein from intein in the form of protein thioester by thiol-thioester exchange **(Scheme 4-1)**.[136]

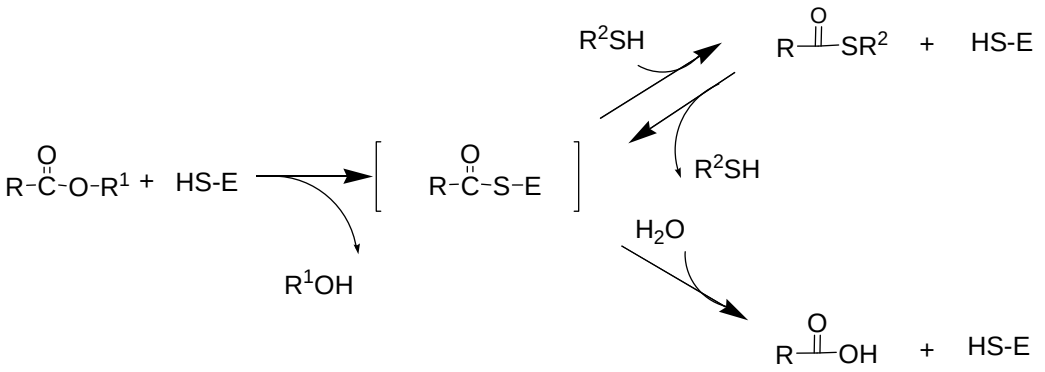


Scheme 4-1. Intein-mediated protein thioesterification.

Protease has a long history to catalyze various organic transformations.[137-140] Compared with chemical approach, this enzymatic strategy bears the merits of site-specificity and operational simplicity. Synthesis of peptides[141-150] and peptide esters[151,152] has been actualized by this strategy. The value of protease in synthesis is associated with its ability to engage a nucleophile other than water. Although thiolysis mediated by any serine or cysteine protease appears feasible in a manner analogous to aminolysis leading to amide formation, peptide thioester synthesis has not been achieved by using protease. The major reason associated with this challenge is that thioester bond formed after thiolysis is very susceptible towards protease hydrolysis. Due to this reason, direct thioesterification by natural cysteine or serine protease is not suitable.

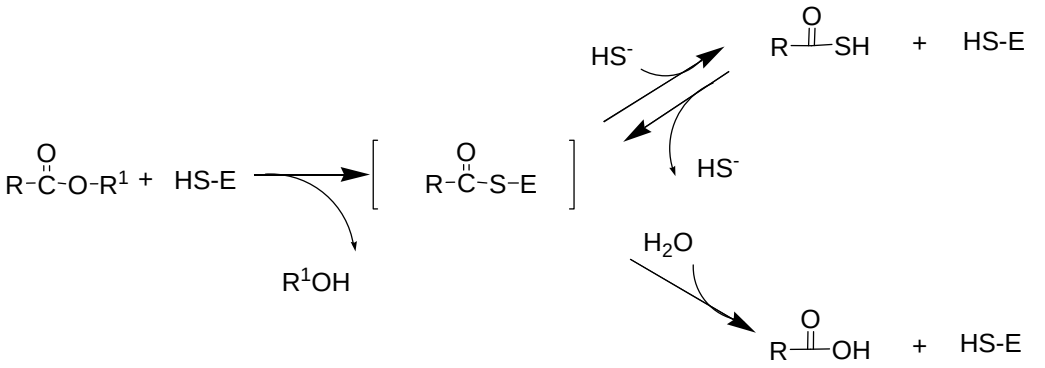
Ligases which catalyze the reverse reaction of proteases have ideal synthetic value. However although proteases are ubiquitous, ligases are rare-occurrence in nature. Subtiligase, a double mutant of protease subtilisin BPN, is a successful example of protein engineering. Though rational design, the original destructive protease is transformed to a ligase with practical synthetic value. Subtiligase retains most of the esterase activity of subtilisin but significantly reduces the original amidase activity.[153] This feature enables a longer life of the acyl-enzyme intermediate, making its capture by a thiol nucleophile possible.

Our group actualized this concept by converting various peptide esters to their thioester form. Using optimized the reaction condition, we managed in maintaining the thiolysis/hydrolysis ratio to an acceptable value **(Scheme 4-2)**. [153]



Scheme 4-2. Peptide ester-to-thioester conversion mediated by subtiligase.

Using a similar mechanism, we also prepared peptide thioacids which is a building block in the efficient thioacid-capture ligation **(Scheme 4-3)**.[154]

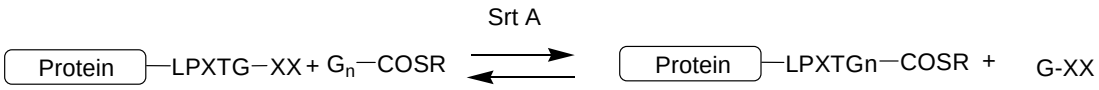


Scheme 4-3. Peptide ester-to-thioacid conversion mediated by subtiligase.

An obvious limitation of subtiligase is that its substrate is peptide ester instead of a normal peptide.

Sortase, one of the few found ligases in nature, is practically useful because of its relative high catalytic efficiency. Since its discovery, Sortase A (Srt A), one subcategory of Sortase, has been used very intensively in protein engineering. Some organic transformations have been achieved by Srt A.

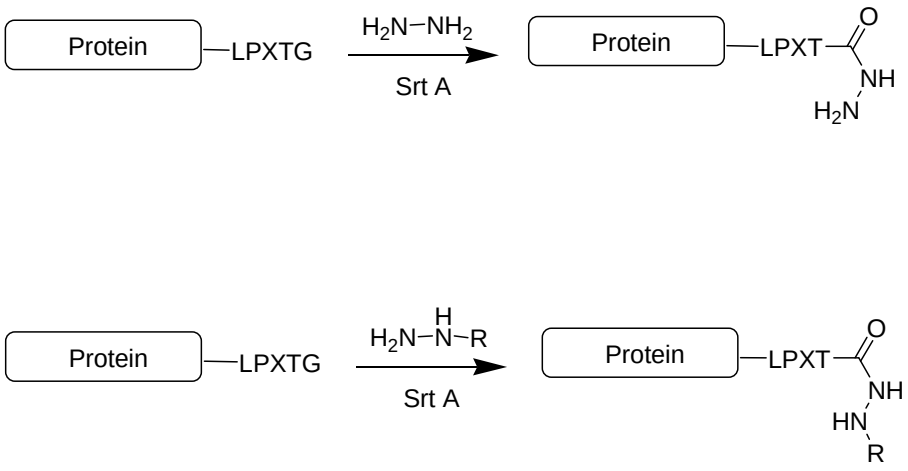
Using Srt A as the biocatalyst, protein thioester preparation from normal protein was achieved by Pentelute's group. They adopted a straightforward strategy in which Srt A was used to attach synthetic oligoglycine thioesters to various proteins. In this report, Sortase A-mediated ligation (SML) was successfully coupled with NCL to perform protein-protein ligation. Using this strategy, the translocation behavior of LF_N-DTA was investigated **(Scheme 4-4)**.[155]



Scheme 4-4. Srt A mediated protein thioester preparation.

Protein hydrazide as protein thioester surrogate has also been sought after due to its value in protein semisynthesis. Liu's group found that Srt A could recognize high concentration of hydrazine and hydrazine derivatives. They could be attached to enzyme motif-containing proteins, generating protein hydrazide its derivative as ligation products. Since the product protein hydrazide is no longer a recognizable substrate of Srt A, the reversibility problem of the ligation reaction is also solved, leading to high yield of ligation reaction. For application, protein hydrazide could be used for protein C-terminal modification or semisynthesis with either cysteine-

containing proteins, aldehydes or ketones. The application scope could be further expanded by using hydrazine derivatives. Functional groups on hydrazine derivatives allow direct C-terminal modification or convergent ligation. Ubiquitin-azide and ubiquitin alkyne could be obtained by reacting each ubiquitin with azide-bearing and alkyne-bearing hydrazide derivatives, respectively. Using click chemistry, a C-to-C ubiquitin dimer was conjugated between the two substrates.[135]



Scheme 4-5. Protein hydrazinolysis by use of Srt A.

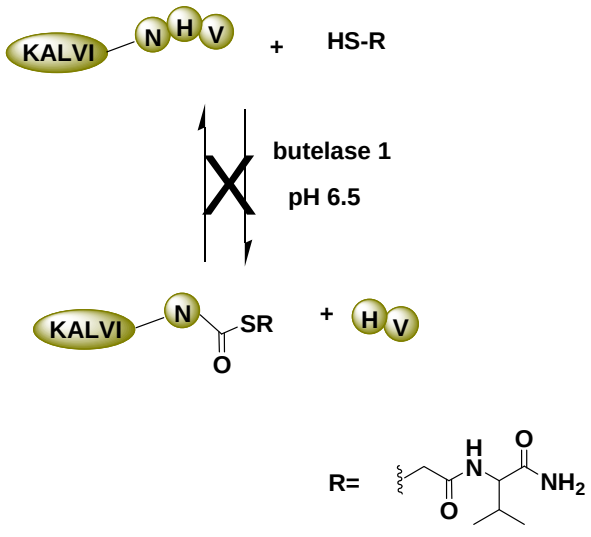
In this chapter, Butelase 1, an enzyme with much higher catalytic efficiency, was examined for protein thioester preparation. Also butelase 1 leaves only a dipeptide Asx-Xaa linkage after the ligation and it has very broad specificity of the Xaa. Considering both efficiency and tag issue, butelase 1 could act as a better tool to bring in chemical functional groups on proteins.

4.2 Results and discussion

In this chapter, an enzymatic method to prepare peptide/protein thioesters would be presented as a complementary method to the chemical method, sortase A-mediated technology and intein-mediated technology. To achieve this goal, the newly discovered robust ligase - butelase is employed.

4.2.1 Peptide/protein thioester preparation by thiolysis

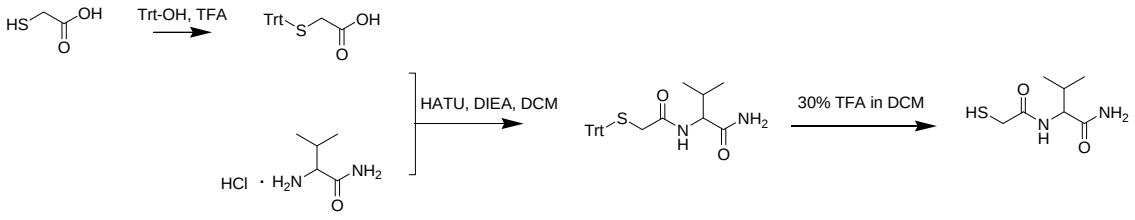
Our first attempt to prepare peptide thioester starts with the incubation of a model peptide KALVINHV with an external thiol in the presence of butelase 1. We hypothesized that this enzymatic amide-to-thioester transformation could be straightforward, as thiolysis of the acyl-enzyme intermediate by an external thiol would lead to a thioester bond in analogy to aminolysis by which a new amide bond is formed **(Scheme 4-6)**.



Scheme 4-6. Peptide thioester preparation by direct thiolysis.

We chose thioglc-Val-NH_2 as our thiol substrate. Thioglycolic acid was used because of its commercial availability and valine residue was selected because in our previous study it was found that butelase 1 preferred a hydrophobic amino acid (Ile/Leu/Val) at the P2" position of the acyl acceptor.

The overall synthesis of the thiol-containing molecule requires three steps: 1 trityl protection of the thioglycolic acid 2 coupling between protected glycolic acid and valine amide 3 removal of trityl group **(Scheme 4-7)**.



Scheme 4-7. Synthesis route of thioglc-Val-NH_2.

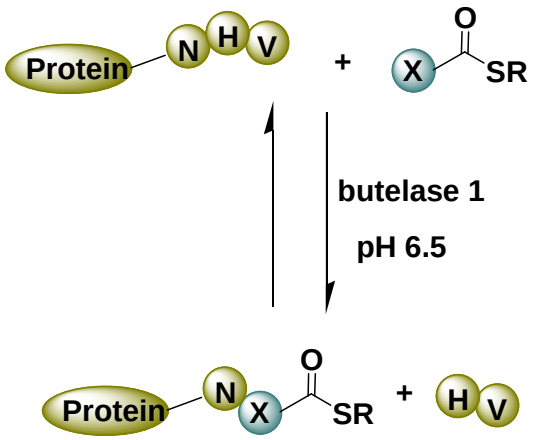
To test our hypothesis, we performed an enzymatic thiolysis reaction of peptide KALVINHV by thiol thioglc-Val-NH_2. Unfortunately, no detectable reaction product KALVIN-thioglc-V was found and intact peptide substrate was recovered after several hours of incubation.

This indicated that the mercaptan group might not act as a butelase-recognizable acceptor nucleophile.

4.2.2 Peptide/protein thioester preparation using amino acid thioester

4.2.2.1 Design rationale

Since the mercaptan group could not be recognized by butelase 1, a natural amino acid at the P1" position would be necessary for enzyme recognition. We envisioned that an amino acid carrying a thioester functionality could be effectively recognized by butelase 1 and ligated with a protein comprising the butelase 1 recognition motif. According to this design, the target protein could be tagged with the amino acid thioester to form an N + 1 protein analog with a C-terminal thioester after the ligation **(Scheme 4-8)**.

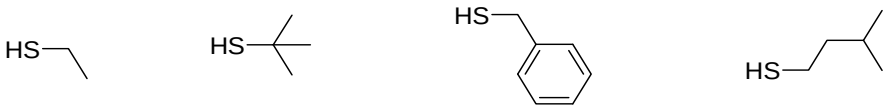


Scheme 4-8. Protein thioester preparation using amino thioester.

4.2.2.2 Screen of thiols

We started with screening thiols that were to be incorporated with the amino acid. We tended to look for those thiols that have structures which, more or less, mimicked the

side chains of natural amino acids. We found four commercially available thiols met our requirement **(Figure 4-1)**.



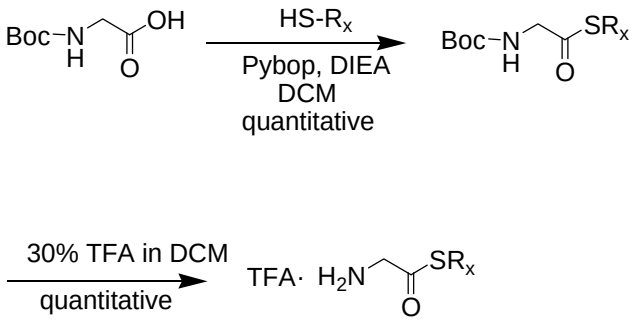
(ethanethiol) (t-butyl thiol) (benzyl mercaptan) (3-methyl-1-butanethiol)

Figure 4-1. Four commercially available thiols for our screening.

The above four thiols have side chains that are similar to those of Alanine, Valine Phenylalanine and Leucine.

4.2.2.3 Synthesis of the amino thioesters

For the P1" amino acid we chose glycine for our initial screening as it has been found to be a common P1" residue in the BML. The preparation of the desired glycine thioesters was straightforward using standard Boc chemistry. First, mixture of Boc-Gly-OH and thiols in the presence of coupling reagent Pybop could afford the Boc protected glycine thioester. TFA treatment would remove the Boc protecting group to give the final desired thioesters in the form of TFA salt.



Scheme 4-9. Synthesis route of glycine thioesters.

4.2.2.4 Model enzymatic ligation with the model peptide YKNHV

With these four glycine thioesters at hand, we used them for ligation test with a model peptide YKNHV. To our delight, under the standard ligation condition (200 μM YKNHV, 5 mM amino thioester, 200 nM butelase 1, 42 ^{o}C), all the four glycine thioesters could be efficiently recognized by the enzyme and gave good ligation yields after 2 hours when the reaction equilibrium was reached (from 64.95% to 88.88%). Since thiol HS-R_2 gave the best ligation performance, we decided to work exclusively with it in our following work.

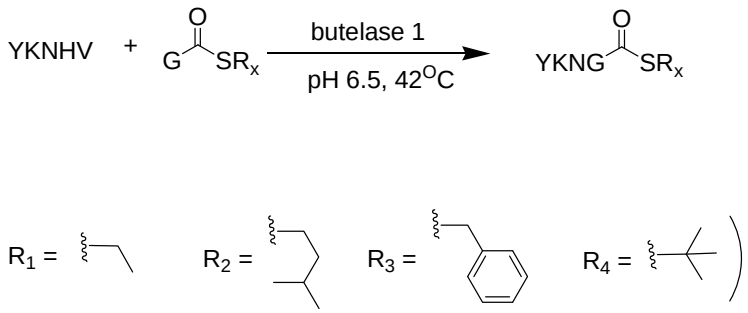


Table 4-1. Ligation efficiencies of different glycine thioesters.

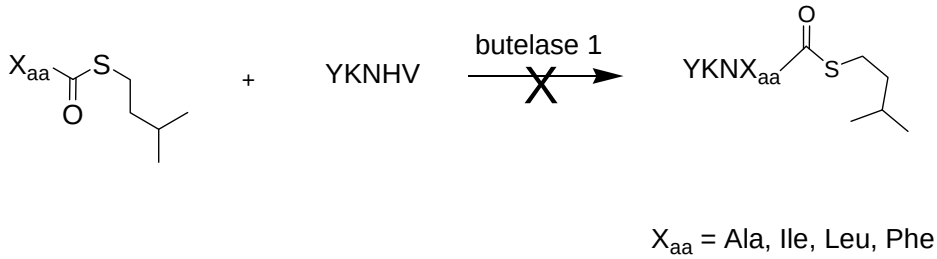
Entry	Rx	Yield(%)
1	R_1	75.12
2	R_2	88.88
3	R_3	87.85
4	R_4	64.95

BML between model peptide and various glycine thioesters. All reactions were carried out in standard reaction condition: 200 μM peptide, 5 mM glycine thioester, 200 nM butelase, pH 6.5 20 mM phosphate buffer, 42 ^{o}C. All the yields were measured after two hours.

4.2.2.5 Substitution of glycine with other amino acids

We then continued to explore whether the glycine residue could be replaced by other amino acids. Using the same chemistry and thiol molecule HS-R_2, we synthesized four other amino thioesters with Ala, Leu, Ile and Phe residues **(Scheme 4-10)**. However incubating each of these four amino thioesters with our model peptide YKNHV could hardly afford any ligation product. This was unexpected to us as according to our previous study on peptide-peptide ligation, butelase 1 showed very high tolerance

(almost all 20 natural amino acids) to the N-terminal residue of the acyl acceptor part. This phenomenon remains to be further investigated. However, the P1" limitation can be easily overcome by replacing the amino acid thioester by a short peptide thioester.



Scheme 4-10. BML between model peptide YKNHV and four other amino thioesters.

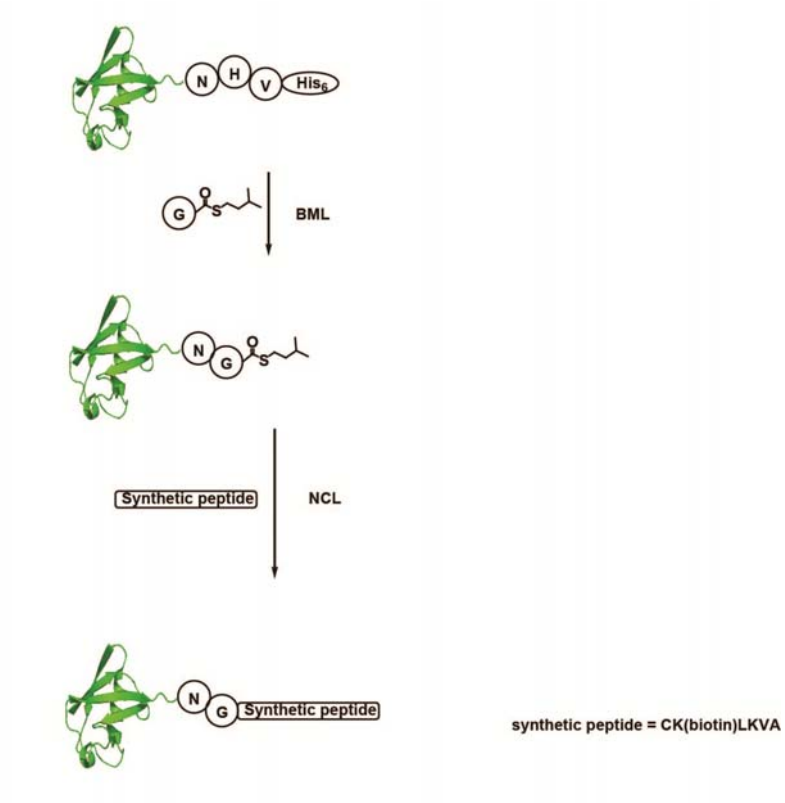
4.2.2.6 Application on proteins

To test our method's suitability on proteins, we engineered an ubiquitin with a -NHV recognition motif followed by a histag at its C-terminus. Incubation of the ubiquitin (100 μM) with the glycine thioester (2 mM) in the presence of butelase 1 (200 nM) afforded the thioester product with a >90% HPLC yield **(Figure 4-2b)**. Another bigger engineered protein ERK-DARPin (~30 k) was also tested. Using a similar reaction condition, we obtained a ~70% HPLC conversion of the target protein to its N + 1 thioester form after 1 hour **(Figure 4-3)**. Thus our method could be used to anchor a thioester group onto the C-terminus of proteins with good efficiency.

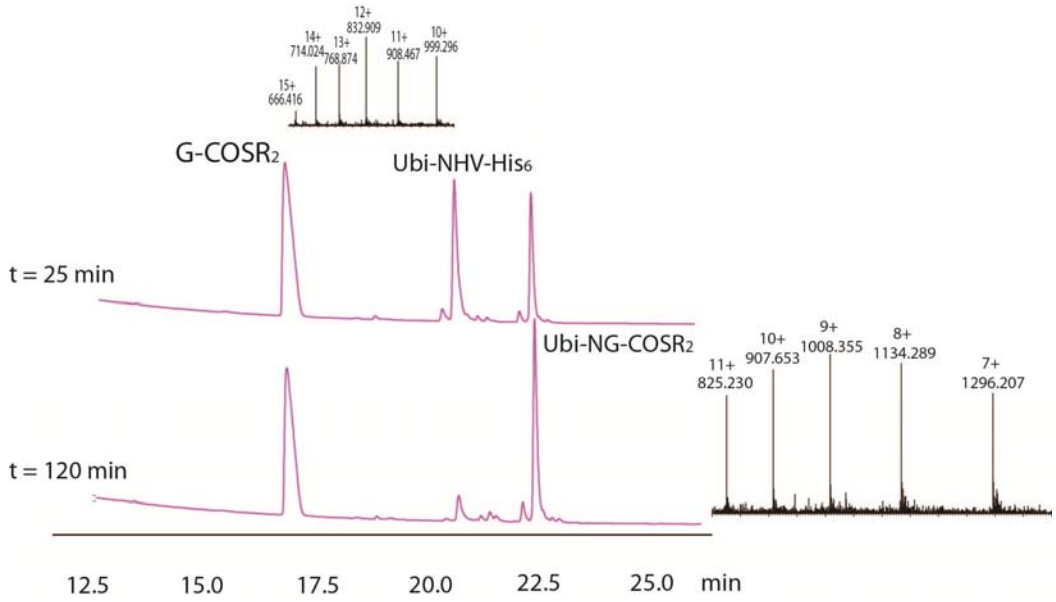
A subsequent NCL following the above process could be used for protein C-terminal labeling. To demonstrate this idea, a cysteinyl peptide bearing a biofuntional group,

biotin, was then synthesized. Standard native chemical ligation was performed between the ubiquitin thioester and the biotin-bearing cysteinyl peptide. After 1 hour, almost complete transformation to the desired product was observed **(Figure 4-2c)**, proving our method's utility in practical research.

a)



b)



c)

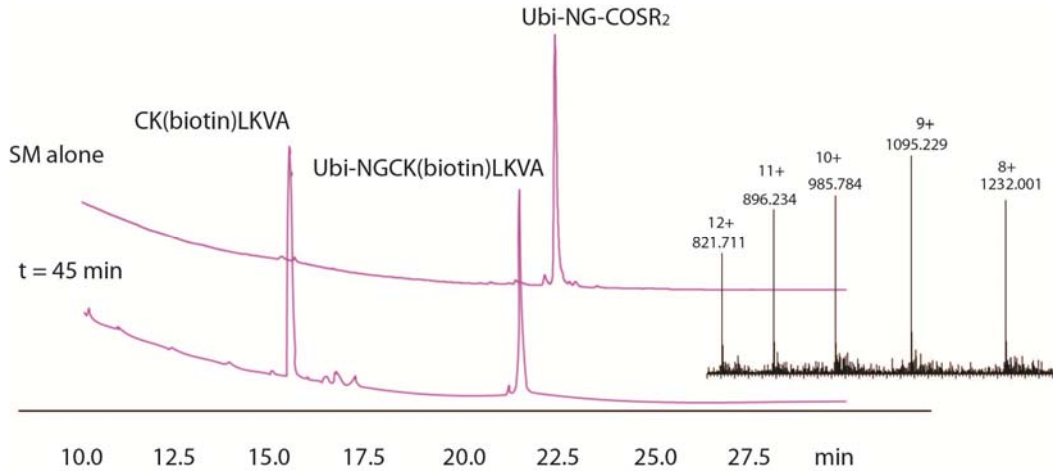


Figure 4-2. C-terminal modification of ubiquitin with a biotin-bearing peptide.

a) Coupling BML and NCL.for C-terminal modification of protein b) HPLC traces for butelase 1 mediated thioesterificaiton of ubiquitin for t = 25 and 120 min. Reaction conditions: 100 μM ubiquitin, 2 mM Gly-$COSR_2$, 200 nM butelase, pH 6.5 phosphate

buffer (20 mM), 42 °C. c) HPLC traces for NCL between ubiquitin thioester and cysteinyl peptide CK(biotin)LKVA. Reaction conditions: ubiquitin thioester: 50 µM, CK(biotin)LKVA: 1 mM, TCEP: 20 mM, MESNa: 20 mM, 42 °C, pH 7.4 PBS. ESI-MS mass for starting material for ubiquitin-NHV-His_6 is 9983 Da (calc. 9983.37 Da). ESI-MS mass for ubiquitin-NG-$COSR_2$ is 9067 Da (calc. 9067.15 Da). ESI-MS mass for ubiquitin-NGCK(biotin)LKVA is 9849 Da (calcd. 9849.48 Da).

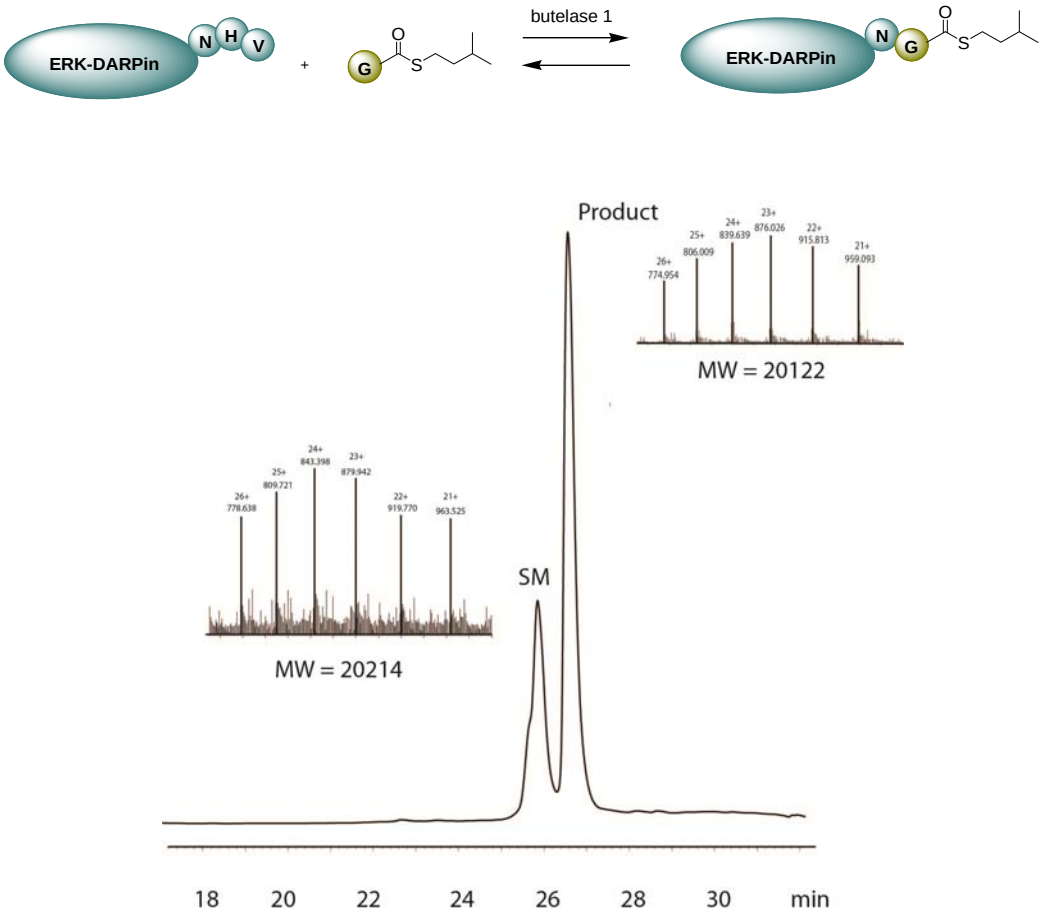


Figure 4-3. Thioesterification of ERK-DARPin.

HPLC profile of thioesterification of Erk-DARPin (1 hour). Reaction condition: protein 100 µM, Glycine thioester 1 mM, butelase 200 nM, pH 6.5 20 mM phosphate buffer,

42 ^{o}C. ESI-MS mass for starting material is 20214 Da (calc. 20210). ESI-MS mass for the thioester product is 20122 (calc. 20117).

4.2.2.7 Tandem ligation using combination of butelase 1, Srt A and NCL

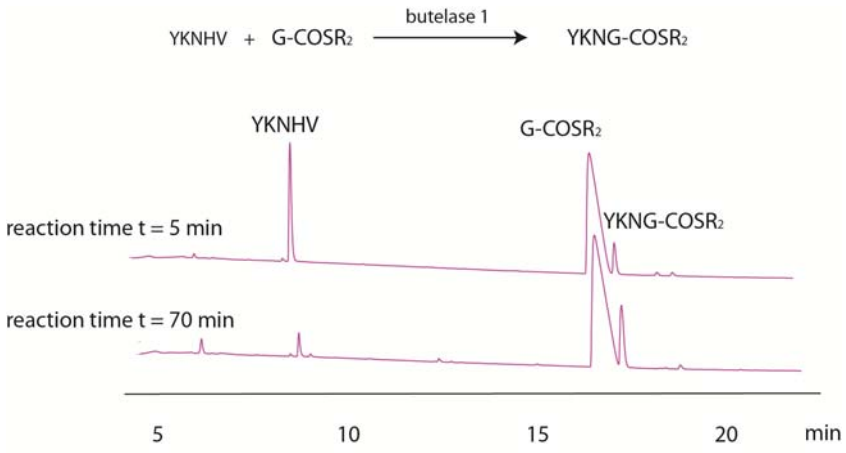
Another usage of our method coupling with NCL could be tandem ligation which could conjugate several peptide segments to afford a longer peptide or protein. To validate this idea, we first tested on small peptide. The model peptide YKNHV was transformed to its corresponding thioester form YKNG-$COSR_2$ by reacting with the glycine thioester via BML **(Figure 4-4a)**. RP-HPLC was used to facilitate the purification of this ligation product. We prepared a cysteinyl peptide CGYKNHV for the native chemical ligation with the peptide thioester. A C-terminal –NHV motif was intentionally included in the sequence because we planned a second BML after NCL.

NCL was then performed between YKNG-$COSR_2$ and the cysteinyl peptide using MESNa as thiol additive. After one hour of reaction, HPLC monitor showed peptide thioester had a quantitative transformation into the desired ligation product YKNGCGYKNHV **(Figure 4-4 b)**.

The NCL product was purified and used for another BML with a model peptide GIGGIR. The ligation did afford the desired ligation product YKNGCGYKNGIGGIR. However a side product YKNGIGGIR was also found, indicating that not only the C-terminal –NHV motif was recognizable but the internal sequence could also be recognized by butelase 1. This was not totally unexpected since the linkage NG was

originally formed by butelase 1, it could also be recognized and cleaved during the second BML.

a)



b)

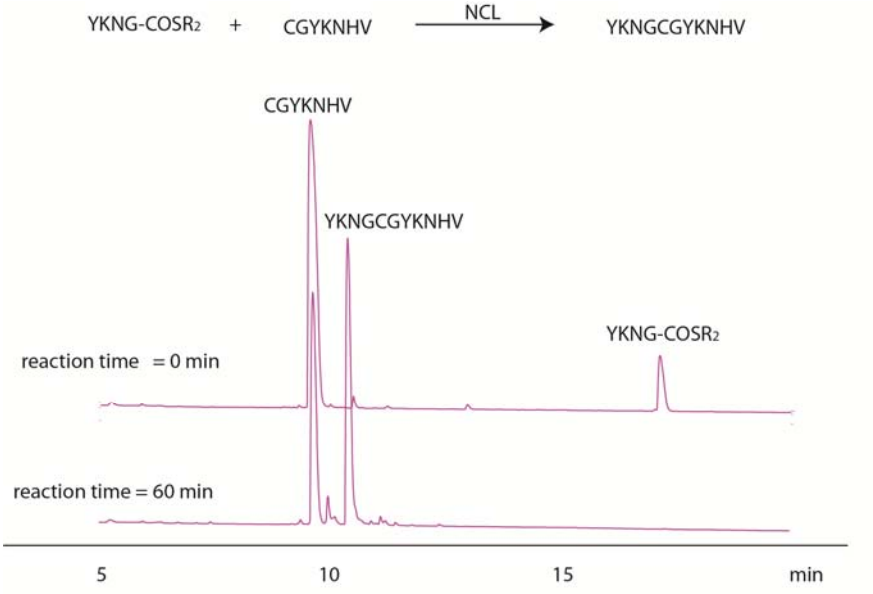


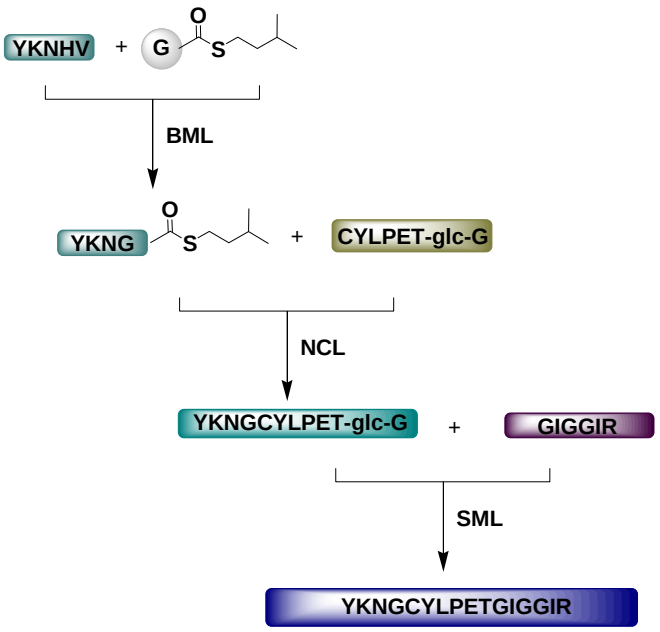
Figure 4-4. Coupling BML and NCL on model peptides.

a) HPLC profiles of BML between model peptide YKNHV and glycine thioester. Reaction conditions: 200 μM peptide, 5 mM glycine thioester, 200 nM butelase, pH 6.5 20 mM phosphate buffer, 42 ^{o}C. b) HPLC profiles of native chemical ligation (NCL)

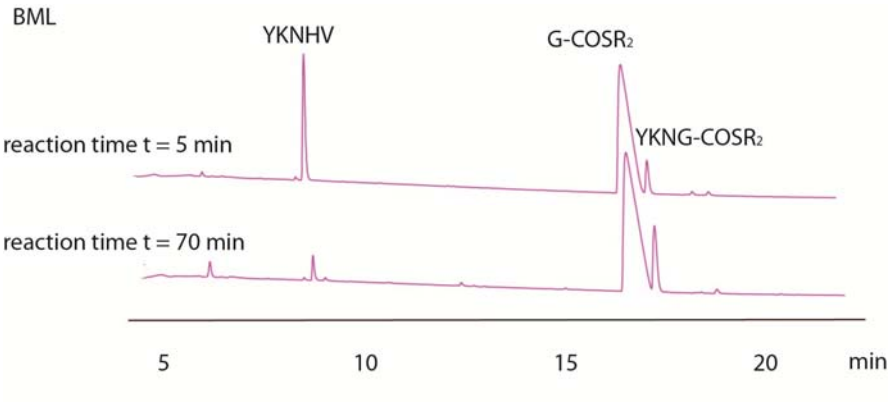
between the peptide thioester product and the cysteinyl peptide CGYKNHV. Reaction conditions: peptide thioester: 1 mM, cysteinyl peptide: 3 mM, MESNa: 50 mM, TCEP 25 Mm, pH 7.4 PBS buffer, 42 ^{o}C.

Srt A could be used to replace butelase 1 for tandem ligation since the two enzymatic methods are orthogonal to each other. Thus we redesigned the sequential ligaiton strategy by including three orthogonal ligation technologies: butelase 1-mediated ligation (BML), native chemical ligation (NCL) and sortase A-mediated ligation (SML) **(Figure 4-5a)**. To demonstrate this concept, the model peptide YKNHV was used, which was first efficiently installed with a thioester group using our method **(Figure 4-5b)**. A second ligation via NCL was then performed between peptide thioester YKNG-$COSR_2$ and a cysteinyl depsipeptide CYLPET-glc-G. The sortase A-recognizable depsipeptide instead of native peptide was used as it has been reported to improve ligation efficiency by solving the reversibility problem of sortase A. Using standard NCL condition, the thioester YKNG-$COSR_2$ was quantitatively transformed into the desired product YKNGCYLPET-glc-G **(Figure 4-5c)**. Finally SML was performed to conjugate YKNGCYLPET-glc-G with GIGGIR to afford the final full-length ligation product YKNGCYLPETGIGGIR **(Figure 4-5d)**. The conversion was above 90% based on HPLC.

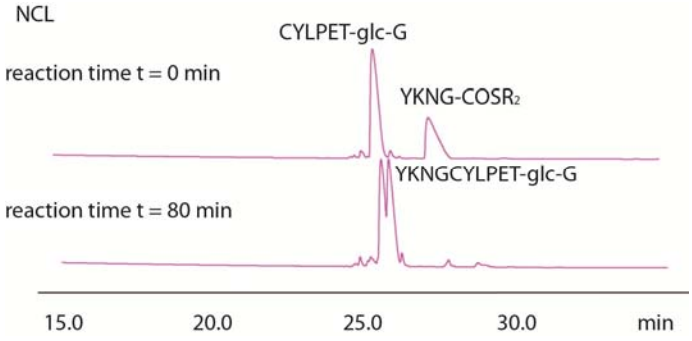
a)


YKNHV
+
G
O
S
BML
YKNG
O
S
+
CYLPET-glc-G
NCL
YKNGCYLPET-glc-G
+
GIGGIR
SML
YKNGCYLPETGIGGIR

b)


BML
YKNHV
G-COSR2
YKNG-COSR2
reaction time t = 5 min
reaction time t = 70 min
5
10
15
20
min

c)



d)

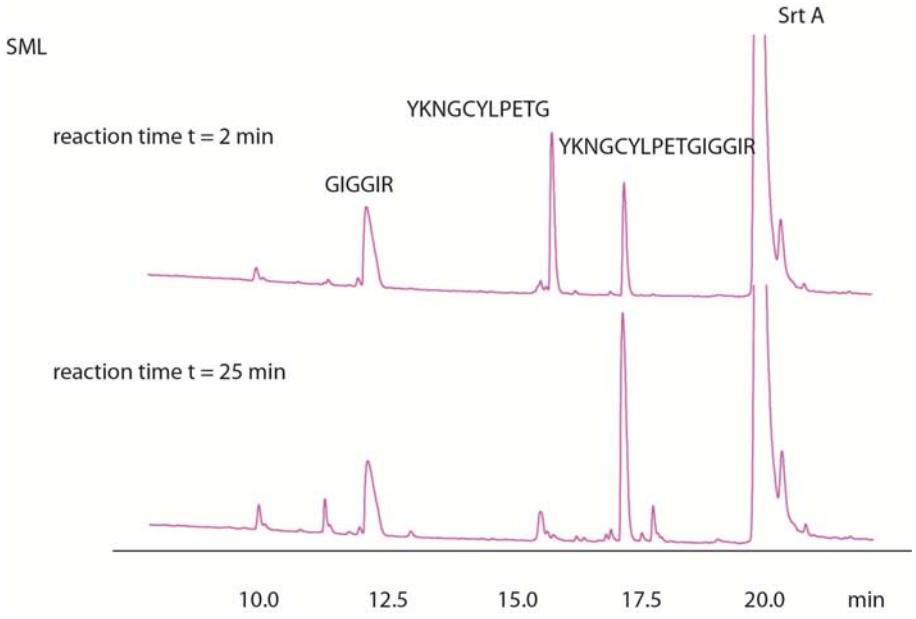


Figure 4-5.Sequential ligations using BML, NCL and SML. a) Schematic demonstration of sequential ligations b) HPLC profile of BML for t = 5 min and t = 70 min. Reaction conditions: 200 μM YKNHV, 5 mM glycine thioester, 200 nM butelase, pH 6.5 20 mM phosphate buffer, 42 ^{o}C. c) HPLC profile of NCL Reaction conditions: 500 μM YKNG-COSR$_2$, 1.5 mM CYLPET-glc-G, 40 mM MESNa, 20 mM TCEP, pH 7.4 PBS buffer, 42 ^{o}C. d) HPLC profile of the purified SML product. Reaction

conditions: 100 μM YKNGCYLPET-glc-G, 500 μM GIGGIR, 50 μM Srt A, 5 mM TCEP, Srt A buffer (150 mM NaCl, 50 mM Tris, 10 mM NaCl,), 37 ^{o}C.

4.3 Conclusion

In this chapter, we developed a novel method to prepare protein thioester by the use of butelase 1 and an amino thioester X_{aa}-COSR_x. According to our screening study, it was found that butelase 1 had a high tolerance on the structure R_x of the thiol molecule but a very strict requirement on the P1" residue for the amino acid thioester. This unexpected phenomenon needs further investigation since in our previous peptide-peptide ligation study, butelase 1 displayed loose requirement on P1''. However, this residue constrain does not pose a big problem as we could overcome the limitation by replacing the amino acid thioester by a peptide thioester. We showed that our method could be used on proteins with high efficiency and could also be coupled with NCL for protein C-terminal labeling. It has been shown that our method was quite efficient and operationally easy. It is also worth mentioning our method leaves only a two amino acid tag "NX_{aa}" after ligation so alterations of the modified proteins would be very small. We believe that our design could serve as a good complement to the classic intein-mediated protein thioesterification technology. Also by using small peptide models we demonstrated that BML, NCL and SML could be coupled together to do tandem ligation which would be useful for protein total synthesis. The success achieved with the small model encourages us to apply the tandem ligation strategy to protein substrates in practice as future work.

4.4 Materials and methods

4.4.1 Materials

All the chemicals used in this chapter are the same as chapter 3.

4.4.2 Instrument

Instrument used in this chapter are the same as chapter 3.

4.4.3 General procedure for synthesis of thioglc-Val-NH_2

Trityl-protected glycolic acid was first prepared in the wayn as described in chapter 3 supporting information. To a mixture of trityl-protected glycolic acid (1 mmol, 334 mg) in DCM was added the H-Val-NH_2·HCl (1.2 mmol, 182.4 mg), DIEA (3 mmol, 519 μl) and HATU (1.2 mmol, 456.276 mg). The mixure was stirring at RT and monitored by TLC. Upon reaction completion, reaction was diluted with EA and sat aq $NaHCO_3$. The layers were separated and the organic layer was extracted with EA, dried over $MgSO_4$, concentrated and purified using silica gel chromatography.

30% TFA and some TIS in DCM was applied to the above purified compound. The reaction stirred at RT for 30 min. Rotary evaporator was used to get rid of the TFA and DCM in the reaction. Lypholization was performed overnight to get the final desired compound.

4.4.4 General procedure for synthesis of amino thioesters (Glycine thioester Gly-$COSR_2$ as example)

Boc-Gly-OH (0.5 mmol, 88.57 mg) and Pybop (0.6 mmol, 312 mg) were first dissolved in DCM. Equal equivalent of 3-methybutane-1-thiol (0.5 mmol, 53.6 mg) was added to the above stirred reaction followed by the addition of 1.5 equivalent of DIEA (0.75 mmol, 127.5 µl). After one hour of reaction, the reaction was concentrated, extracted with EA/H_2O. The organic layer was further washed with brine, dried over $MgSO_4$ and purified.

The purified compound was treated with 30% TFA in DCM for 30 min. The concentrated residue was dissolved in water/ACN mixture and lypholized for overnight to get the final desired product.

4.4.5 Solid phase peptide synthesis

H-**YKNHV**-NH_2, H-**CGYKNHV**-NH_2, H-**GIGGIR**-NH_2 and H-**CK**(biotin)**LKVA**-NH_2 were synthesized on rink amide MBHA resin using strandard Fmoc chemistry as described in chapter 2. The depsipeptide CYLPET-glc-G was synthesized as described in chapter 3.

4.4.6 Preparation of the plasmids for ubiquitin and ERK-DARPin

Protein sequences of the engineered ubiquitin and ERK-DARPin

ERK-DARPin:

MHHHHHHSSGVDLGTENLYFQSMGSDLGKKLLEAARAGQDDEVRILMANGA
DVNAHDDQGSTPLHLAAWIGHPEIVEVLLKHGADVNARDTDGWTPLHLAAD
NGHLEIVEVLLKYGADVNAQDAYGLTPLHLAADRGHLEIVEVLLKHGADVNA
QDKFGKTAFDISIDNGNEDLAEILQKLNKNHV

Ubiquitin:

MGGMQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLED
GRTLSDYNIQKESTLHLVLRLRGGNHVHHHHHH

Genetic sequences of the engineered ubiquitin and ERK-DARPin

ERK-DARPin:

ATGCACCATCATCATCATCATTCTTCTGGTGTAGATCTGGGTACCGAGAACC
TGTACTTCCAATCCATGGGCTCCGACCTGGGTAAAAAACTGCTGGAAGCTG
CTCGTGCTGGTCAAGATGATGAAGTGCGTATCCTGATGGCTAATGGTGCCG
ACGTGAACGCGCATGATGACCAGGGCAGCACCCCGCTGCATCTGGCAGCCT
GGATTGGTCACCCGGAAATCGTGGAAGTTCTGCTGAAACACGGCGCGGATG
TTAACGCCCGTGATACCGACGGTTGGACGCCGCTGCATCTGGCAGCTGACA
ATGGCCACCTGGAAATTGTCGAAGTGCTGCTGAAGTATGGCGCGGATGTCA
ACGCACAGGACGCTTACGGTCTGACCCCGCTGCATCTGGCGGCCGATCGTG
GTCACCTGGAAATCGTTGAAGTCCTGCTGAAACATGGCGCAGATGTTAATG
CTCAAGACAAATTTGGCAAGACGGCCTTCGACATCTCTATTGATAACGGTA
ATGAAGACCTGGCTGAAATCCTGCAAAAACTGAATTGA

ubiquitin:

ATGGGTGGTATGCAGATCTTCGTCAAGACGTTAACCGGTAAAACCATAACT

CTGGAAGTTGAACCATCCGATACCATCGAAAACGTTAAGGCTAAAATTCAA

GACAAGGAAGGAATTCCACCTGATCAACAAAGATTGATCTTTGCCGGTAAG

CAGCTCGAGGACGGTAGAACGCTGTCTGATTACAACATTCAGAAGGAGTCG

ACCTTACATCTTGTCTTAAGACTAAGAGGTGGTAACCATGTTCATCATCATC

ATCATCATTGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGC

The plasmid of ERK-DARPin was received as a gift from Dr. Tobias.

The primers used for ubiquitin plasmid construction are as follows:

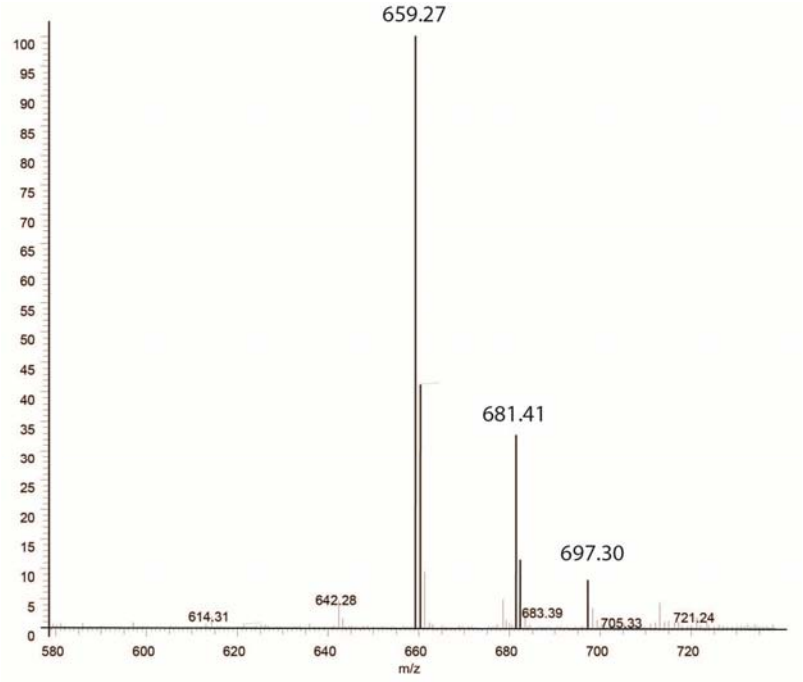
Forward: 5CGCCATATG GGTGGT ATGCAGATCT TCGTC3

Reverse: 5ATGATGAT GATGATG AACATGGTTA CCACCTCTTA GTC3

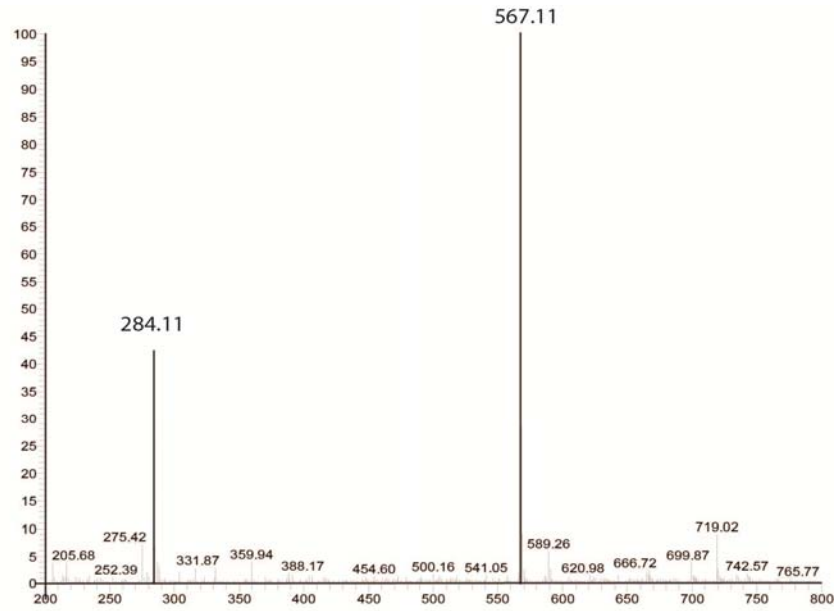
The procedures for the plasmid construction and protein expression are the same described in chapter 3 supporting information.

4.4.7 Mass spectra

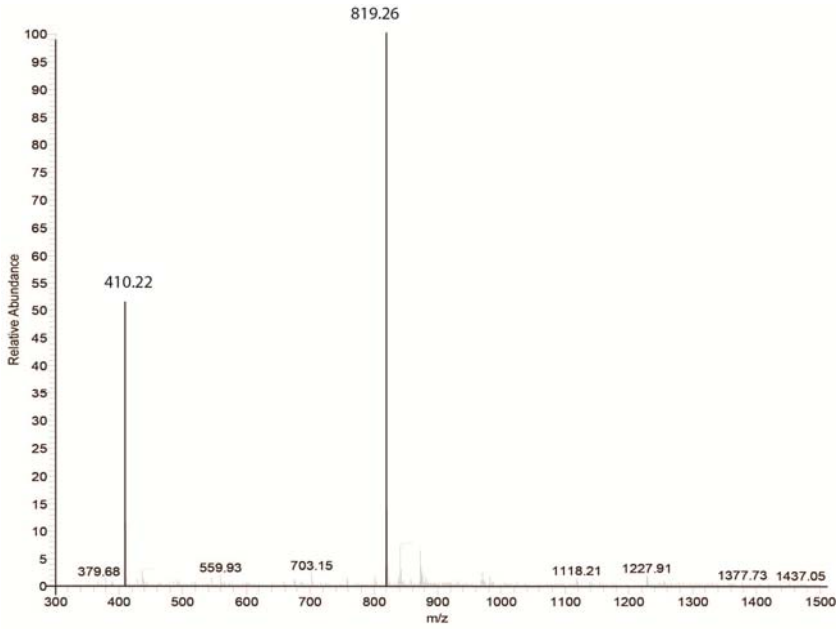
H-**YKNHV**-NH_2 (*m/z* $[M + H]^+$ 659.27, $[M + Na]^+$ 681.41, $[M + K]^+$ 697.30, calcd isotopic M_W 658.35)



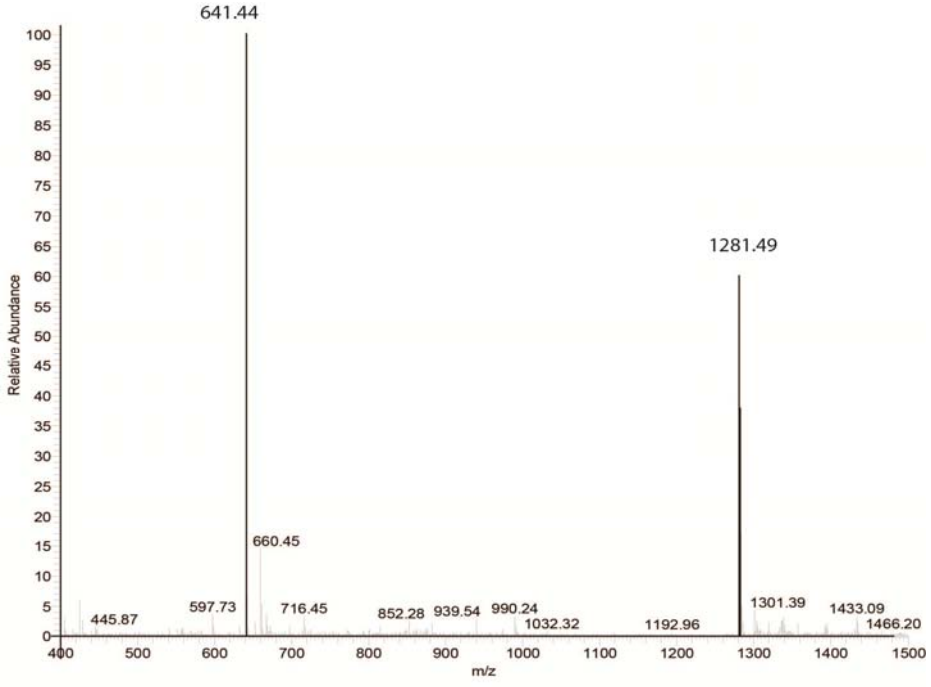
H-**YKNG**-$COSR_2$ (*m/z* $[M + H]^+$ 567.11, $[M + 2H]^{2+}$ 284.11, calcd isotopic M_W 566.28)



H-**CGYKNHV**-NH_2 (*m/z* $[M + H]^+$ 819.26, $[M + 2H]^{2+}$ 410.22, calcd isotopic M_W 818.38)

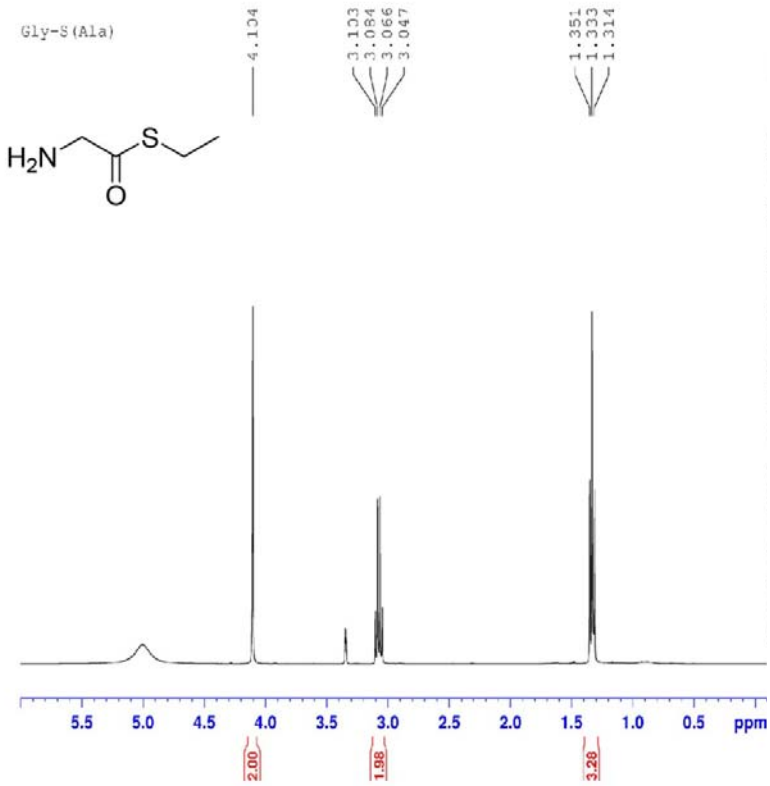


H-**YKNGCGYKNHV**-NH_2 (*m/z* $[M + H]^+$ 1281.49, $[M + 2H]^{2+}$ 641.44, calcd isotopic M_W 1280.60)

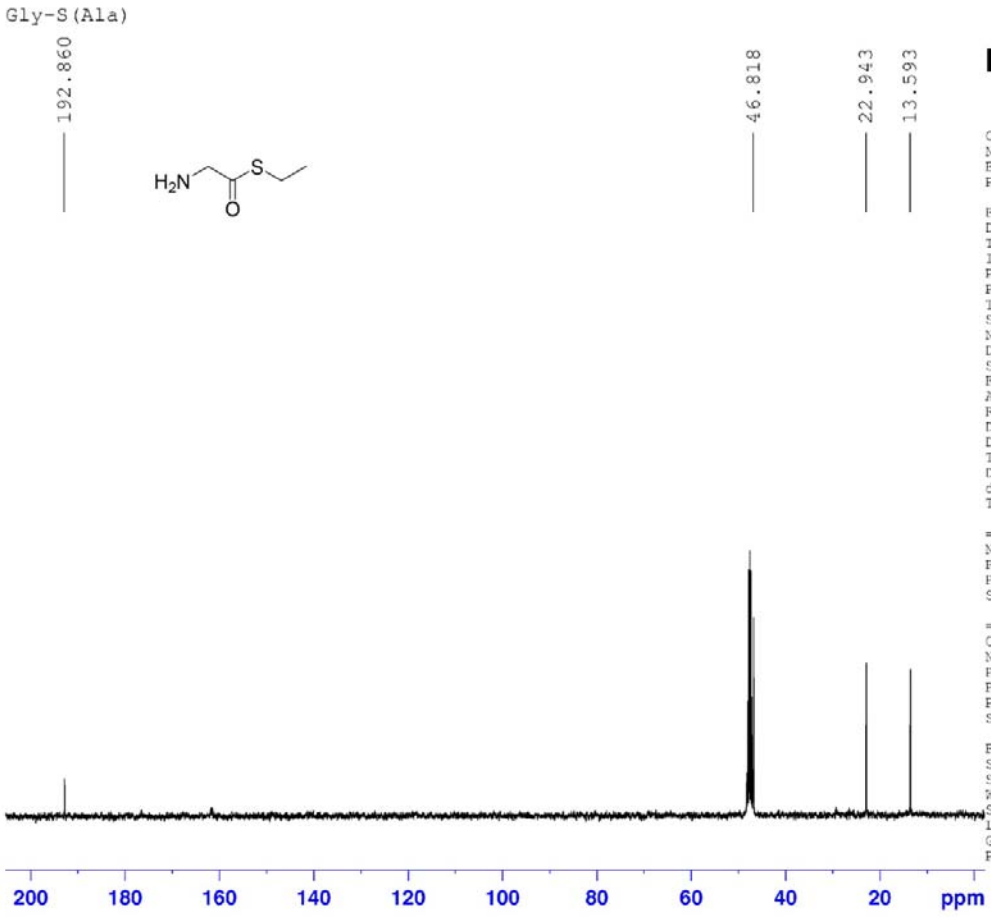


4.4.8 NMR spectra

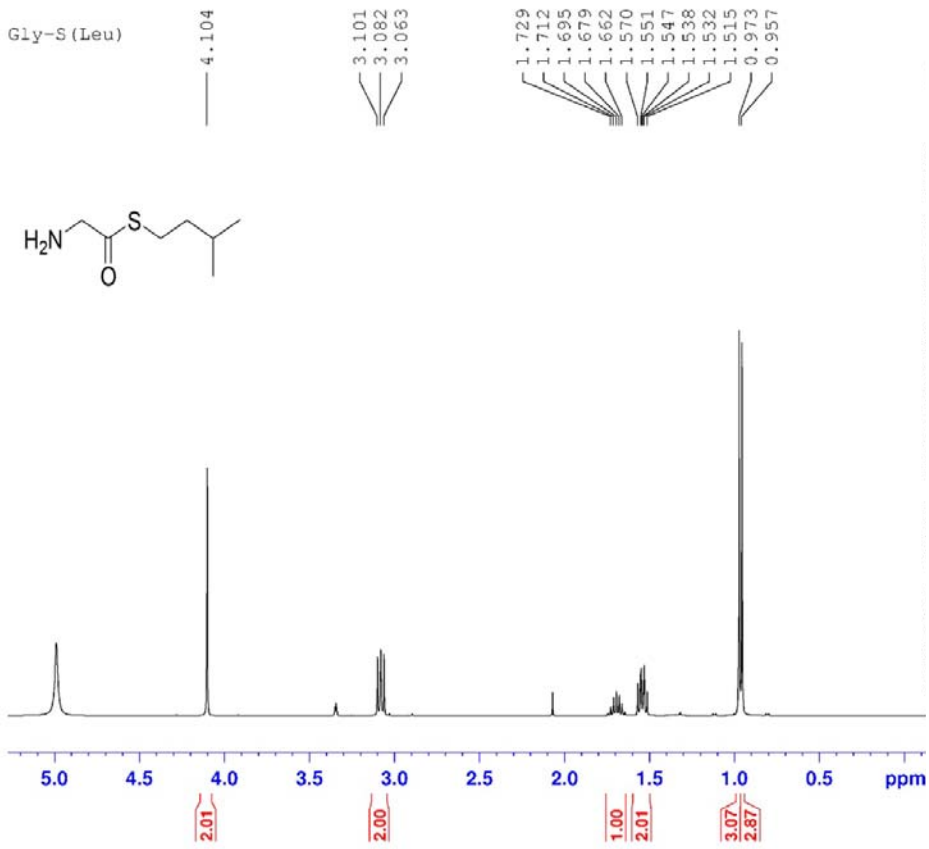
Gly-COSR$_x$



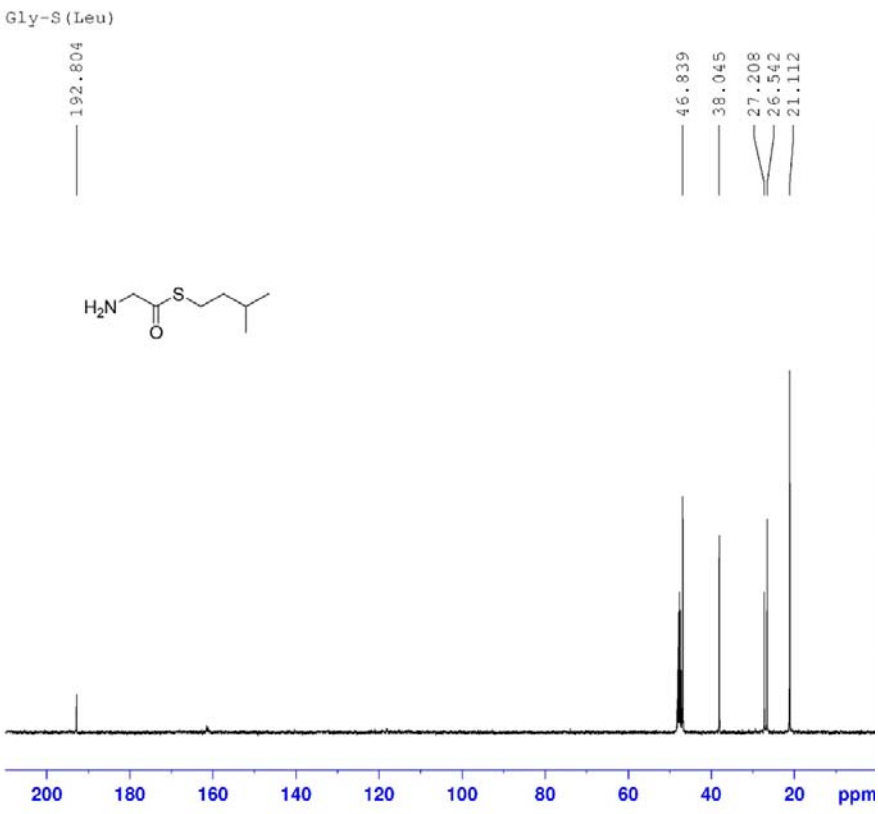
^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 4.104 (s, 2H, NH_2-C<u>H$_2$</u>), 3.103-3.047 (q, 2H, S-CH_2), 1.351-1.314 (t, 3H, CH_3)



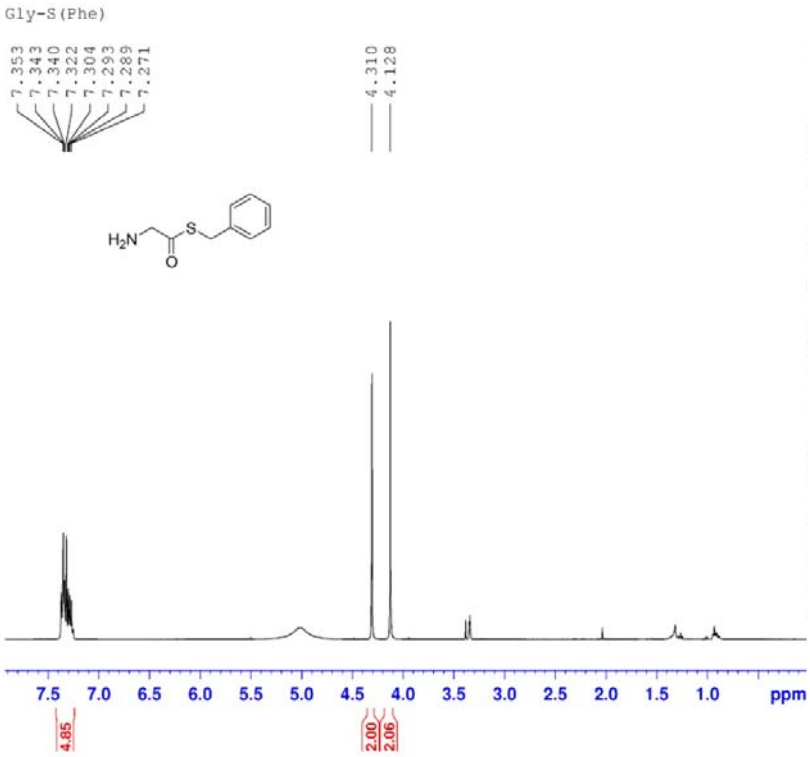
^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 192.860 (C=O), 46.818 (CH_2), 22.943(CH_2), 13.593 (CH_3)



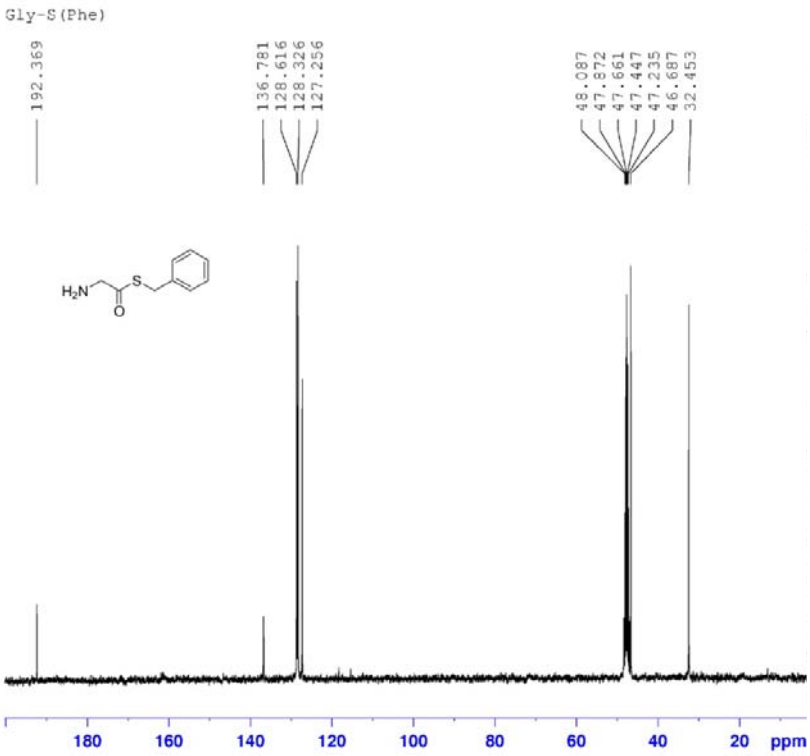
^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 4.104 (s, 2H, NH_2-C<u>H$_2$</u>), 3.101-3.063 (t, 2H, S-CH_2), 1.729-1.662 (m, 1H, CH), 1.570-1.515 (m, 2H, CH_2), 0.973 (s, 3H, CH_3), 0.957 (s, 3H, CH_3)



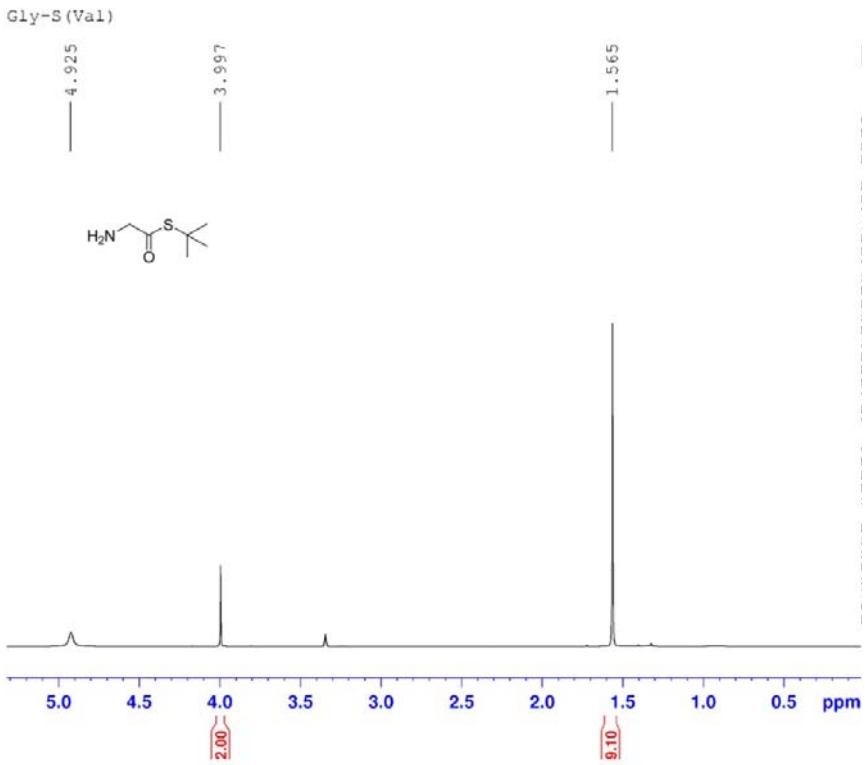
^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 192.804 (C=O), 46.839 (NH_2-CH_2), 38.045 (S-CH_2), 27.208 (CH_2), 26.542 (CH), 21.112 (CH_3)



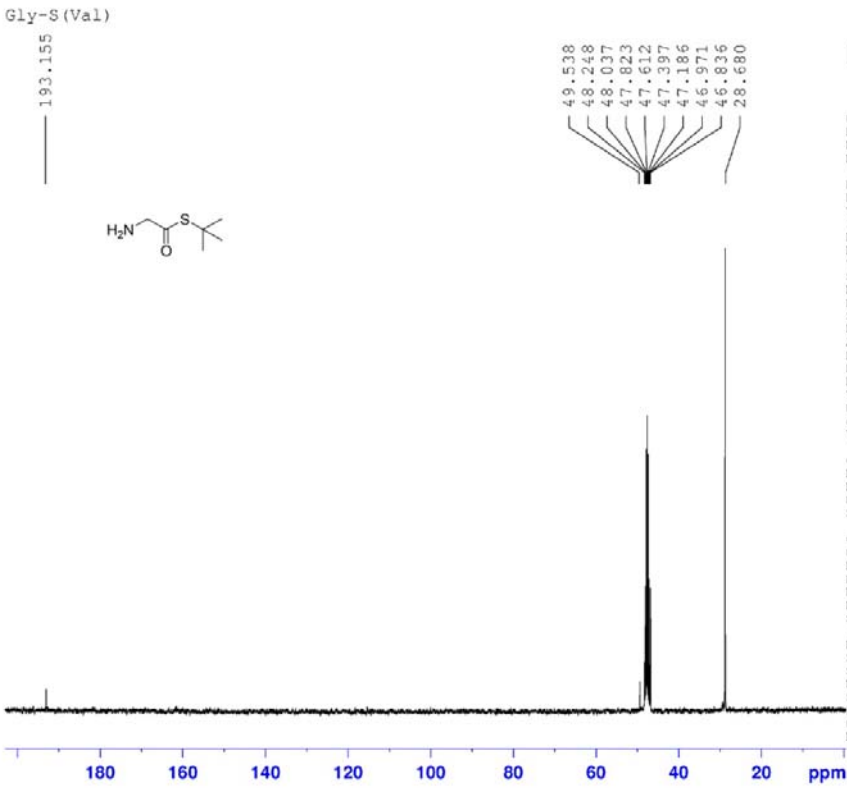
^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 7.353-7.271 (m, 5H, Ph), 4.310 (s, 2H, CH_2-Ph), 4.128 (s, 2H, NH_2-CH_2)



^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 192.369 (C=O), 136.781 (Ph), 128.161(Ph), 128.326 (Ph), 127.256 (Ph), 46.687 (S-CH_2), 32.453 (NH_2-CH_2)



^{1}H NMR (400 MHz, $CDCl_3$): δ in ppm 3.997 (s, 2H, CH_2), 1.565 (s, 9H, CH_3)



^{13}C NMR (400 MHz, $CDCl_3$): δ in ppm 193.155 (C=O), 28.680 (CH_3)

References

1. Dawson, P. E.; Kent, S. B. *Annual Review of Biochemistry* **2000,** *69*, 923-60.
2. Hackenberger, C. P.; Schwarzer, D. *Angew Chem Int Ed Engl* **2008,** *47*, 10030-74.
3. Bianco, A.; Townsley, F. M.; Greiss, S.; Lang, K.; Chin, J. W. *Nature Chemical Biology* **2012,** *8*, 748-50.
4. Chin, J. W.; Cropp, T. A.; Anderson, J. C.; Mukherji, M.; Zhang, Z.; Schultz, P. G. *Science* **2003,** *301*, 964-7.
5. Furter, R. *Protein Sci* **1998,** *7*, 419-26.
6. Greiss, S.; Chin, J. W. *J Am Chem Soc* **2011,** *133*, 14196-9.
7. Liu, C. C.; Schultz, P. G. *Annual Review of Biochemistry* **2010,** *79*, 413-44.
8. Liu, W.; Brock, A.; Chen, S.; Schultz, P. G. *Nature Methods* **2007,** *4*, 239-44.
9. Parrish, A. R.; She, X.; Xiang, Z.; Coin, I.; Shen, Z.; Briggs, S. P.; Dillin, A.; Wang, L. *ACS Chem Biol* **2012,** *7*, 1292-302.
10. Sakamoto, K.; Hayashi, A.; Sakamoto, A.; Kiga, D.; Nakayama, H.; Soma, A.; Kobayashi, T.; Kitabatake, M.; Takio, K.; Saito, K.; Shirouzu, M.; Hirao, I.; Yokoyama, S. *Nucleic Acids Res* **2002,** *30*, 4692-9.
11. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. *Science* **2001,** *292*, 498-500.
12. Merrifield, R. B. *Journal of the American Chemical Society* **1963,** *85*, 2149-&.
13. Coin, I.; Dolling, R.; Krause, E.; Bienert, M.; Beyermann, M.; Sferdean, C. D.; Carpino, L. A. *Journal of Organic Chemistry* **2006,** *71*, 6171-7.
14. Mutter, M.; Nefzi, A.; Sato, T.; Sun, X.; Wahl, F.; Wohr, T. *Pept Res* **1995,** *8*, 145-53.
15. Yoshiya, T.; Taniguchi, A.; Sohma, Y.; Fukao, F.; Nakamura, S.; Abe, N.; Ito, N.; Skwarczynski, M.; Kimura, T.; Hayashi, Y.; Kiso, Y. *Organic & Biomolecular Chemistry* **2007,** *5*, 1720-30.
16. Kent, S. B. *Chem Soc Rev* **2009,** *38*, 338-51.
17. Kiyama, S.; Fujii, N.; Yajima, H.; Moriga, M.; Takagi, A. *Int J Pept Protein Res* **1984,** *23*, 174-86.
18. Sakakibara, S. *Biopolymers* **1995,** *37*, 17-28.
19. Baca, M.; Kent, S. B. *Proc Natl Acad Sci U S A* **1993,** *90*, 11638-42.
20. Fisch, I.; Kunzi, G.; Rose, K.; Offord, R. E. *Bioconjugate Chemistry* **1992,** *3*, 147-153.
21. King, T. P.; Zhao, S. W.; Lam, T. *Biochemistry* **1986,** *25*, 5774-5779.
22. Liu, C. F.; Tam, J. P. *Proc Natl Acad Sci U S A* **1994,** *91*, 6584-8.
23. Rose, K. *Journal of the American Chemical Society* **1994,** *116*, 30-33.
24. Schnolzer, M.; Kent, S. B. *Science* **1992,** *256*, 221-5.
25. Segawa, Y.; Higashihara, T.; Ueda, M. *J Am Chem Soc* **2010,** *132*, 11000-1.
26. Kemp, D. S.; Carey, R. I. *Journal of Organic Chemistry* **1989,** *54*, 3640-3646.
27. Kemp, D. S.; Fotouhi, N. *Tetrahedron Letters* **1987,** *28*, 4637-4640.
28. Wieland, T.; Bokelmann, E.; Bauer, L.; Lang, H. U.; Lau, H. *Annalen Der Chemie-Justus Liebig* **1953,** *583*, 129-149.
29. Dawson, P. E.; Muir, T. W.; Clarklewis, I.; Kent, S. B. H. *Science* **1994,** *266*, 776-779.
30. Tam, J. P.; Lu, Y. A.; Liu, C. F.; Shao, J. *Proceedings of the National Academy of Sciences of the United States of America* **1995,** *92*, 12485-12489.
31. Aimoto, S. *Biopolymers* **1999,** *51*, 247-65.
32. Johnson, E. C.; Kent, S. B. *J Am Chem Soc* **2006,** *128*, 6640-6.
33. Tam, J. P.; Yu, Q. *Biopolymers* **1998,** *46*, 319-27.
34. Chen, G.; Warren, J. D.; Chen, J.; Wu, B.; Wan, Q.; Danishefsky, S. J. *J Am Chem Soc* **2006,** *128*, 7460-2.
35. Low, D. W.; Hill, M. G.; Carrasco, M. R.; Kent, S. B.; Botti, P. *Proc Natl Acad Sci U S A* **2001,** *98*, 6554-9.

36. Offer, J. *Biopolymers* **2010,** *94*, 530-41.
37. Offer, J.; Boddy, C. N.; Dawson, P. E. *J Am Chem Soc* **2002,** *124*, 4642-6.
38. Wu, B.; Chen, J.; Warren, J. D.; Chen, G.; Hua, Z.; Danishefsky, S. J. *Angew Chem Int Ed Engl* **2006,** *45*, 4116-25.
39. Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J. *Angew Chem Int Ed Engl* **2008,** *47*, 8521-4.
40. Chen, J.; Wang, P.; Zhu, J.; Wan, Q.; Danishefsky, S. J. *Tetrahedron* **2010,** *66*, 2277-2283.
41. Crich, D.; Banerjee, A. *J Am Chem Soc* **2007,** *129*, 10064-5.
42. Haase, C.; Rohde, H.; Seitz, O. *Angew Chem Int Ed Engl* **2008,** *47*, 6807-10.
43. Harpaz, Z.; Siman, P.; Kumar, K. S.; Brik, A. *ChemBioChem* **2010,** *11*, 1232-5.
44. Tan, Z.; Shang, S.; Danishefsky, S. J. *Angew Chem Int Ed Engl* **2010,** *49*, 9500-3.
45. Yan, L. Z.; Dawson, P. E. *J Am Chem Soc* **2001,** *123*, 526-33.
46. Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. *J Am Chem Soc* **2009,** *131*, 13592-3.
47. Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. *Chem Commun (Camb)* **2010,** *46*, 7199-201.
48. Liu, C. F.; Rao, C.; Tam, J. P. *Tetrahedron Letters* **1996,** *37*, 933-936.
49. Amitai, G.; Dassa, B.; Pietrokovski, S. *J Biol Chem* **2004,** *279*, 3121-31.
50. David, R.; Richter, M. P.; Beck-Sickinger, A. G. *Eur J Biochem* **2004,** *271*, 663-77.
51. Ding, Y.; Xu, M. Q.; Ghosh, I.; Chen, X.; Ferrandon, S.; Lesage, G.; Rao, Z. *J Biol Chem* **2003,** *278*, 39133-42.
52. Evans, T. C., Jr.; Benner, J.; Xu, M. Q. *J Biol Chem* **1999,** *274*, 18359-63.
53. Giriat, I.; Muir, T. W.; Perler, F. B. *Genet Eng (N Y)* **2001,** *23*, 171-99.
54. Klabunde, T.; Sharma, S.; Telenti, A.; Jacobs, W. R., Jr.; Sacchettini, J. C. *Nature Structural Biology* **1998,** *5*, 31-6.
55. Lovrinovic, M.; Seidel, R.; Wacker, R.; Schroeder, H.; Seitz, O.; Engelhard, M.; Goody, R. S.; Niemeyer, C. M. *Chem Commun (Camb)* **2003**, 822-3.
56. Mills, K. V.; Manning, J. S.; Garcia, A. M.; Wuerdeman, L. A. *J Biol Chem* **2004,** *279*, 20685-91.
57. Muir, T. W. *Annual Review of Biochemistry* **2003,** *72*, 249-89.
58. Noren, C. J.; Wang, J.; Perler, F. B. *Angew Chem Int Ed Engl* **2000,** *39*, 450-466.
59. Paulus, H. *Bioorg Chem* **2001,** *29*, 119-29.
60. Perler, F. B. *Nucleic Acids Res* **2002,** *30*, 383-4.
61. Perler, F. B.; Olsen, G. J.; Adam, E. *Nucleic Acids Res* **1997,** *25*, 1087-93.
62. Pietrokovski, S. *Curr Biol* **1998,** *8*, R634-5.
63. Pietrokovski, S. *Trends Genet* **2001,** *17*, 465-72.
64. Poland, B. W.; Xu, M. Q.; Quiocho, F. A. *J Biol Chem* **2000,** *275*, 16408-13.
65. Romanelli, A.; Shekhtman, A.; Cowburn, D.; Muir, T. W. *Proc Natl Acad Sci U S A* **2004,** *101*, 6397-402.
66. Southworth, M. W.; Benner, J.; Perler, F. B. *EMBO J* **2000,** *19*, 5019-26.
67. Xu, M. Q.; Evans, T. C., Jr. *Methods Mol Biol* **2003,** *205*, 43-68.
68. Xu, M. Q.; Southworth, M. W.; Mersha, F. B.; Hornstra, L. J.; Perler, F. B. *Cell* **1993,** *75*, 1371-7.
69. Mills, K. V.; Johnson, M. A.; Perler, F. B. *Journal of Biological Chemistry* **2014,** *289*, 14498-14505.
70. Shah, N. H.; Muir, T. W. *Chemical Science* **2014,** *5*, 446-461.
71. Wood, D. W.; Camarero, J. A. *Journal of Biological Chemistry* **2014,** *289*, 14512-14519.
72. Connolly, K. M.; Smith, B. T.; Pilpa, R.; Ilangovan, U.; Jung, M. E.; Clubb, R. T. *J Biol Chem* **2003,** *278*, 34061-5.
73. Frankel, B. A.; Kruger, R. G.; Robinson, D. E.; Kelleher, N. L.; McCafferty, D. G. *Biochemistry* **2005,** *44*, 11188-200.

74. Huang, X.; Aulabaugh, A.; Ding, W.; Kapoor, B.; Alksne, L.; Tabei, K.; Ellestad, G. *Biochemistry* **2003,** *42*, 11307-15.
75. Popp, M. W. L.; Ploegh, H. L. *Angewandte Chemie-International Edition* **2011,** *50*, 5024-5032.
76. Pritz, S. *Mini-Reviews in Organic Chemistry* **2008,** *5*, 47-52.
77. Proft, T. *Biotechnology Letters* **2010,** *32*, 1-10.
78. Tian, B. X.; Eriksson, L. A. *J Phys Chem B* **2011,** *115*, 13003-11.
79. Tsukiji, S.; Nagamune, T. *ChemBioChem* **2009,** *10*, 787-798.
80. Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *J Am Chem Soc* **2008,** *130*, 16338-43.
81. Guo, X.; Wang, Q.; Swarts, B. M.; Guo, Z. *J Am Chem Soc* **2009,** *131*, 9878-9.
82. Matsunaga, S.; Matsuoka, K.; Shimizu, K.; Endo, Y.; Sawasaki, T. *BMC Biotechnol* **2010,** *10*, 42.
83. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nature Chemical Biology* **2007,** *3*, 707-8.
84. Pritz, S.; Wolf, Y.; Kraetke, O.; Klose, J.; Bienert, M.; Beyermann, M. *Journal of Organic Chemistry* **2007,** *72*, 3909-12.
85. Samantaray, S.; Marathe, U.; Dasgupta, S.; Nandicoori, V. K.; Roy, R. P. *J Am Chem Soc* **2008,** *130*, 2132-3.
86. Tanaka, T.; Yamamoto, T.; Tsukiji, S.; Nagamune, T. *ChemBioChem* **2008,** *9*, 802-7.
87. Wu, Z.; Guo, X.; Guo, Z. *Chem Commun (Camb)* **2010,** *46*, 5773-4.
88. Wu, Z.; Guo, X.; Wang, Q.; Swarts, B. M.; Guo, Z. *J Am Chem Soc* **2010,** *132*, 1567-71.
89. Yamamoto, T.; Nagamune, T. *Chem Commun (Camb)* **2009**, 1022-4.
90. Nguyen, G. K.; Wang, S.; Qiu, Y.; Hemu, X.; Lian, Y.; Tam, J. P. *Nature Chemical Biology* **2014,** *10*, 732-8.
91. Hojo, H.; Aimoto, S. *Bulletin of the Chemical Society of Japan* **1991,** *64*, 111-117.
92. Bu, X. Z.; Xie, G. Y.; Law, C. W.; Guo, Z. H. *Tetrahedron Letters* **2002,** *43*, 2419-2422.
93. Clippingdale, A. B.; Barrow, C. J.; Wade, J. D. *Journal of Peptide Science* **2000,** *6*, 225-234.
94. Li, X. Q.; Kawakami, T.; Aimoto, S. *Tetrahedron Letters* **1998,** *39*, 8669-8672.
95. Aboye, T. L.; Clark, R. J.; Craik, D. J.; Goransson, U. *Journal of Peptide Science* **2008,** *14*, 144-144.
96. Mende, F.; Seitz, O. *Angewandte Chemie-International Edition* **2011,** *50*, 1232-1240.
97. Mezo, A. R.; Cheng, R. P.; Imperiali, B. *Journal of the American Chemical Society* **2001,** *123*, 3885-3891.
98. Nagalingam, A. C.; Radford, S. E.; Warriner, S. L. *Synlett* **2007**, 2517-2520.
99. Erlich, L. A.; Kumar, K. S.; Haj-Yahya, M.; Dawson, P. E.; Brik, A. *Organic & Biomolecular Chemistry* **2010,** *8*, 2392-6.
100. Hojo, H.; Murasawa, Y.; Katayama, H.; Ohira, T.; Nakahara, Y. *Organic & Biomolecular Chemistry* **2008,** *6*, 1808-13.
101. Hojo, H.; Onuma, Y.; Akimoto, Y.; Nakahara, Y.; Nakahara, Y. *Tetrahedron Letters* **2007,** *48*, 1299-1299.
102. Kang, J.; Reynolds, N. L.; Tyrrell, C.; Dorin, J. R.; Macmillan, D. *Organic & Biomolecular Chemistry* **2009,** *7*, 4918-23.
103. Kang, J.; Richardson, J. P.; Macmillan, D. *Chem Commun (Camb)* **2009**, 407-9.
104. Katayama, H.; Hojo, H.; Shimizu, I.; Nakahara, Y. *Organic & Biomolecular Chemistry* **2010,** *8*, 1966-72.
105. Nagaike, F.; Onuma, Y.; Kanazawa, C.; Hojo, H.; Ueki, A.; Nakahara, Y. *Organic Letters* **2006,** *8*, 4465-8.
106. Nagaike, F.; Onuma, Y.; Kanazawa, C.; Hojo, H.; Ueki, A.; Nakahara, Y.; Nakahara, Y. *Organic Letters* **2006,** *8*, 4465-4468.

107. Nakamura, K.; Kanao, T.; Uesugi, T.; Hara, T.; Sato, T.; Kawakami, T.; Aimoto, S. *Journal of Peptide Science* **2009,** *15*, 731-7.
108. Ohta, Y.; Itoh, S.; Shigenaga, A.; Shintaku, S.; Fujii, N.; Otaka, A. *Organic Letters* **2006,** *8*, 467-70.
109. Ozawa, C.; Katayama, H.; Hojo, H.; Nakahara, Y. *Organic Letters* **2008,** *10*, 3531-3.
110. Richardson, J. P.; Chan, C. H.; Blanc, J.; Saadi, M.; Macmillan, D. *Organic & Biomolecular Chemistry* **2010,** *8*, 1351-60.
111. Sharma, R. K.; Tam, J. P. *Organic Letters* **2011,** *13*, 5176-9.
112. Tsuda, S.; Shigenaga, A.; Bando, K.; Otaka, A. *Organic Letters* **2009,** *11*, 823-6.
113. Vizzavona, J.; Dick, F.; Vorherr, T. *Bioorganic & Medicinal Chemistry Letters* **2002,** *12*, 1963-5.
114. Hou, W.; Zhang, X. H.; Li, F. P.; Liu, C. F. *Organic Letters* **2011,** *13*, 386-389.
115. Ollivier, N.; Dheur, J.; Mhidia, R.; Blanpain, A.; Melnyk, O. *Organic Letters* **2010,** *12*, 5238-5241.
116. Szilagyi, L.; Gyorgydeak, Z. *Journal of the American Chemical Society* **1979,** *101*, 427-432.
117. Gamblin, D. P.; Garnier, P.; van Kasteren, S.; Oldham, N. J.; Fairbanks, A. J.; Davis, B. G. *Angew Chem Int Ed Engl* **2004,** *43*, 828-33.
118. Tam, J. P.; Yu, Q.; Miao, Z. *Biopolymers* **1999,** *51*, 311-32.
119. Baker, D. P.; Lin, E. Y.; Lin, K.; Pellegrini, M.; Petter, R. C.; Chen, L. L.; Arduini, R. M.; Brickelmaier, M.; Wen, D.; Hess, D. M.; Chen, L.; Grant, D.; Whitty, A.; Gill, A.; Lindner, D. J.; Pepinsky, R. B. *Bioconjug Chem* **2006,** *17*, 179-88.
120. Gilmore, J. M.; Scheck, R. A.; Esser-Kahn, A. P.; Joshi, N. S.; Francis, M. B. *Angew Chem Int Ed Engl* **2006,** *45*, 5307-11.
121. Mao, H. Y.; Hart, S. A.; Schink, A.; Pollok, B. A. *Journal of the American Chemical Society* **2004,** *126*, 2670-2671.
122. Parthasarathy, R.; Subramanian, S.; Boder, E. T. *Bioconjugate Chemistry* **2007,** *18*, 469-476.
123. de Koning, M. C.; van der Marel, G. A.; Overhand, M. *Current Opinion in Chemical Biology* **2003,** *7*, 734-740.
124. Pritz, S.; Wolf, Y.; Kraetke, O.; Klose, J.; Bienert, M.; Beyermann, M. *Journal of Organic Chemistry* **2007,** *72*, 3909-3912.
125. Samantaray, S.; Marathe, U.; Dasgupta, S.; Nandicoori, V. K.; Roy, R. P. *Journal of the American Chemical Society* **2008,** *130*, 2132-+.
126. Antos, J. M.; Miller, G. M.; Grotenbreg, G. M.; Ploegh, H. L. *Journal of the American Chemical Society* **2008,** *130*, 16338-16343.
127. Johnsson, N.; George, N.; Johnsson, K. *ChemBioChem* **2005,** *6*, 47-52.
128. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L. *Nature Chemical Biology* **2007,** *3*, 707-708.
129. Tanaka, T.; Yamamoto, T.; Tsukiji, S.; Nagamune, T. *ChemBioChem* **2008,** *9*, 802-807.
130. Yamamoto, T.; Nagamune, T. *Chemical Communications* **2009**, 1022-1024.
131. Mao, H. Y. *Protein Expression and Purification* **2004,** *37*, 253-263.
132. Yamamura, Y.; Hirakawa, H.; Yamaguchi, S.; Nagamune, T. *Chem Commun (Camb)* **2011,** *47*, 4742-4.
133. Antos, J. M.; Chew, G. L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W.; Ploegh, H. L. *J Am Chem Soc* **2009,** *131*, 10800-1.
134. Williamson, D. J.; Fascione, M. A.; Webb, M. E.; Turnbull, W. B. *Angew Chem Int Ed Engl* **2012,** *51*, 9377-80.
135. Li, Y. M.; Li, Y. T.; Pan, M.; Kong, X. Q.; Huang, Y. C.; Hong, Z. Y.; Liu, L. *Angew Chem Int Ed Engl* **2014,** *53*, 2198-202.
136. Xu, M. Q.; Evans, T. C., Jr. *Methods* **2001,** *24*, 257-77.

137. Bordusa, F. *Chem Rev* **2002,** *102*, 4817-68.
138. Kise, H.; Hayakawa, A.; Noritomi, H. *J Biotechnol* **1990,** *14*, 239-54.
139. Lombard, C.; Saulnier, J.; Wallach, J. M. *Protein Pept Lett* **2005,** *12*, 621-9.
140. Wong, C. H.; Wang, K. T. *Experientia* **1991,** *47*, 1123-9.
141. Abrahmsen, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. *Biochemistry* **1991,** *30*, 4151-9.
142. Barbas, C. F.; Matos, J. R.; West, J. B.; Wong, C. H. *Journal of the American Chemical Society* **1988,** *110*, 5162-5166.
143. Bordusa, F. *Curr Protein Pept Sci* **2002,** *3*, 159-80.
144. Braisted, A. C.; Judice, J. K.; Wells, J. A. *Methods Enzymol* **1997,** *289*, 298-313.
145. Chang, T. K.; Jackson, D. Y.; Burnier, J. P.; Wells, J. A. *Proc Natl Acad Sci U S A* **1994,** *91*, 12544-8.
146. Hansler, M.; Jakubke, H. D. *Journal of Peptide Science* **1996,** *2*, 279-89.
147. Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A. *Science* **1994,** *266*, 243-7.
148. Sears, P.; Schuster, M.; Wang, P.; Witte, K.; Wong, C.-H. *Journal of the American Chemical Society* **1994,** *116*, 6521-6530.
149. Wehofsky, N.; Koglin, N.; Thust, S.; Bordusa, F. *J Am Chem Soc* **2003,** *125*, 6126-33.
150. Zaks, A.; Klibanov, A. M. *J Biol Chem* **1988,** *263*, 3194-201.
151. Kise, H.; Shirato, H. *Tetrahedron Letters* **1985,** *26*, 6081-6084.
152. Liu, C. F.; Tam, J. P. *Organic Letters* **2001,** *3*, 4157-9.
153. Tan, X. H.; Wirjo, A.; Liu, C. F. *ChemBioChem* **2007,** *8*, 1512-5.
154. Tan, X. H.; Zhang, X.; Yang, R.; Liu, C. F. *ChemBioChem* **2008,** *9*, 1052-6.
155. Ling, J. J.; Policarpo, R. L.; Rabideau, A. E.; Liao, X.; Pentelute, B. L. *J Am Chem Soc* **2012,** *134*, 10749-52.