ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

327
4x
D/B/ux

THIOACID SYNTHESIS AND APPLICATION IN PEPTIDE AND PROTEIN CHEMISTRY

ZHANG XIAOHONG

School of Biological Sciences

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

2009

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

To my family

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Acknowledgements

I would like to express my gratitude to all those who gave me help, support and concern during my postgraduate study. Without them, this thesis could not have been finished smoothly.

First of all, my deepest thanks go to my supervisor, Prof. Liu Chuan-Fa, with whose guidance and assistance I can finish this project. With his profound knowledge and rich research experience in both chemistry and biology, he has offered me so many interesting and valuable ideas and suggestions, and whenever available he has always been willing to discuss with me on my work. I am so grateful to his patience and encouragement which have always inspired me during the entire period of my Ph.D study. His rigorous attitude towards science will have a great impact on me for my entire research career.

I owe a great deal to our group members, Mr. Andre Wirjo, Ms. Hou Wen, Mr. Li Fupeng, Ms. Lu Xiaowei, Ms. Qin Wei, Dr. Tan Xiaohong, Mr. Yang Renliang, Dr. Yang Yi, and Dr. Zen Yun, for their help, encouragement and friendship during my study. And thanks are also extended to my postgraduate friends in SBS, especially Dr. Lu Le, who has always been there to offer me help whether it's the up or down times.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

I furthermore appreciate the School of Biological Sciences, Nanyang Technological University for giving me this opportunity and much assistance to pursue my Ph.D study.

At last, I am greatly indebted to my parents, my husband and my little daughter whose selfless love and encouragement have enabled me to persevere in and complete my Ph.D study.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table of Content

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Summary

There is an increasing need for the use of synthetic proteins in biomedical research and drug discovery. Since its introduction in 1963, solid phase peptide synthesis (SPPS) has been widely used for the synthesis of peptides and small proteins. In recent years, convergent strategies for protein synthesis have been developed including Kemp's prior thiol capture ligation, thiazolidine-capture ligation, thioester-mediated native chemical ligation, thioacid capture ligation, followed by Staudiger ligation and decarboxylative condensation. Among them, the thioacid capture ligation is probably the most efficient in terms of reaction rate. The main reason for the low use of this method is the difficulty in obtaining the thioacid building blocks.

In this project, we have developed novel methods to synthesize thioacid peptides and proteins which can be used not only in thioacid capture ligation but also for protein C-terminal labeling. These new methods make use of a simple but efficient hydrothiolysis reaction of peptide and protein thioesters in aqueous media. On one hand, a solid phase method was developed for the synthesis of peptide thioacid through hydrothiolysis of resin-bound peptide thioesters by using the total PEG-based ChemMatrix resin. On another hand, a biosynthetic method was developed for the synthesis of protein thioacids through interception by hydrosulfide ions of thioester intermediates in the intein-mediated protein splicing

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

process.

The synthetic utility of both the thioacid peptides and proteins was demonstrated in the synthesis of histone H3 analogs by thioacid capture ligation. The solid phase synthesis-derived H3 N-terminal peptide thioacids were used for ligation with the recombinant H3 C-terminal domain which contained an N-ter Cys residue modified by Npys. Similarly, H3 analogs containing C-terminal modifications were prepared by ligation of H3(1-109) thioacid with synthetic Npys-modified H3(110-135). These represent the first examples of applying thioacid capture ligation to protein synthesis.

Another utility of peptide/protein thioacids is the introduction of C-terminal modification through the use of the thioacid/azide amidation. The reaction was reported to be especially efficient with electron-deficient azides such as sulfonazides. Several functional azides containing tosyl, Dansyl, Dabsyl, biotin and PEG moieties were synthesized and used for the C-terminal labeling of peptides and proteins. Through these examples, we have demonstrated a new method for peptide and protein bioconjugation.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Abbreviations

Reagents and chemicals

Ac	Acetyl
ACN	Acetonitrile
Boc	t-Butyloxycarbonyl
Bzl	Benzyl
PyBop	Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate
Dabsyl	4-(dimethylamino)azobenzene-4'-sulfonyl
Dansyl	5-(dimethylamino)naphthalene-1-sulfonyl
DCC	Dicyclohexylcarbodiimide
DCM	Dichloromethane
DIEA	*N,N*-Diisopropylethylamine
DMF	*N,N*-Dimethylformamide
DMSO	Dimethylsulfoxide
DTNP	2,2'-Dithiobis-(5-nitropyridine)
DTNB	5,5'-Dithiobis-(2-nitrobenzoic acid)
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Fmoc	9-Fluorenylmethyloxycarbonyl
Gdn-HCl	Guanidine hydrochloride
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HOBt	Hydroxybenzotriazole
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MBHA	4-Methylbenzhydrylamine
Me	Methyl
MESNA	Sodium 2-sulfanylethanesulfonate
Mts	Mesitylene-2-sulfonyl
Npys	2-mercapto-5-nitropyridyl
OtBu	O-ter-Butyl
Pbf	2,2,4,6,7-Pentamethyldihydrobenzofuran-5-sulfonyl
SDS	Sodium dodecyl sulfate
tBu	tert-Butyl
TCEP	Tris(2-carboxylethyl)phosphine
TFA	Trifluoroacetic acid
TFMSA	Trifluoromethanesulfonic acid
TIS	Triisopropylsilane
Tos	Tosyl
Tris	Tris(hydroxymethyl)aminomethane
Trt	Trityl

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Xan	Xanthyl

Others

aa	Amino acid
ATP	Adenosine triphosphate
CBD	Chitin binding domain
CM resin	ChemMatrix resin
CT	C-terminal
Da	Dalton
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
ESI-MS	Electrospray ionization mass spectrometry
HF	Hydrogen fluoride
HPLC	High performance liquid chromatography
MW	Molecular weight
MALDI-TOF	Matrix-assisted laser desorption/ionization-time of flight
NMR	Nuclear magnetic resonance
NT	N-terminal
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
RNA	Ribonucleic acid
RP	Reverse phase
RT	Room temperature
SPPS	Solid phase peptide synthesis
TLC	Thin layer chromatography
UV	Ultraviolet ray

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 1: General Introduction

Proteins are the central players of life in all living organisms. Although an organism may have a limited number of genes, a far greater number of proteins can be derived from these genes owing to processes such as alternative splicing and posttranslational modification and each is required for a particular function. Since the hundreds of thousands of new proteins revealed by the genome sequencing projects are known only as predicted sequence data, elucidation of the biological function and its relationship with the native structure of a protein represents a great challenge and opportunity for life science research. The understanding of how a protein works to excert its function or functions in either its native or post-translationally modified form is bound not only to solve problems in basic biology but also to produce therapeutic breakthroughs in medicine. For this reason, protein structure-function study is a central topic of modern biological research. Historically, the study of proteins as well as bioactive peptides has been dependent on the various biological and chemical methods developed for peptide and protein synthesis [1-6]. Such methods have dramatically enhanced our ability to manipulate peptide or protein molecules for their biophysical and biochemical characterization and to create entirely new peptide/protein molecules with desirable properties for medical and biotechnological application.

In this introduction chapter, I will give a review on the currently available

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

biological and chemical protein synthesis methods. The advantages and limitations of the two types of approaches are discussed in view of their capability of introducing nonnatural structural elements. For the biological approaches, the discussion is focused on the methods that utilize an amber stop codon and a suppressor tRNA for the incorporation of nonnatural amino acids. For the chemical approaches, special attention is paid to chemical ligation methods that share a general two-step scheme of prior capture and proximity-driven intramolecular acyl transfer for peptide bond formation. In particular, the distinct features of thioacid-based synthetic methods are highlighted. The case of histone proteins is used to illustrate the increasing need for such protein synthesis methods.

1.1 Biological approaches

The expression of proteins by recombinant DNA technology has been a dominating force in the field of protein chemistry for over 30 years. This powerful method enables the production of large amounts of proteins and allows changes of amino acid sequences of proteins through the site-directed mutagenesis technique [1]. However, this technology has some limitations. For instance, the overexpression of some proteins is toxic to the cells and the uncontrolled processing of the nascent polypeptide can cause product heterogeneity [2]. Moreover, since living cells are used for protein production, such synthesis systems are inherently limited to the 20 genetically encoded amino acids.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Nevertheless, our understanding of the protein translation process has made it possible to make use of the ribosomal protein synthesis machinery for the introduction of modified nonnatural amino acids into proteins.

1.1.1 Biosynthesis of proteins bearing nonnatural amino acids in cell–free systems

In nature, the formation of a peptide bond during protein synthesis is achieved by transferring the C-terminal acyl group of a growing peptide chain, in the ribosome, to the α-amino group of an aminoacyl tRNA [3]. It is now known that, once charged onto a tRNA, the structure of an amino acid side chain has little influence on the codon-anticodon recognition between the messenger RNA and the transfer RNA. Based on this "adaptor hypothesis" [4] and by making use of a nonsense codon, which does not encode any amino acids but rather signals terminating polypeptide synthesis by binding release factors, and a corresponding nonsense suppressor tRNA that normally insert one of the common amino acids in response to a stop codon, nonnatural amino acids can be introduced into proteins with excellent translation efficiency and fidelity by the natural translation machinery [5,7].

In 1989, Schultz *et al.* [5] reported a new technique for site-specific incorporation of nonnatural amino acids into proteins using the extract of *E. coli* cell lysates. They used a nonsense codon UAG to replace the codon encoding the amino acid of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

interest by site-directed mutagenesis. A chemically aminoacylated supressor tRNA with the desired amino acid incorporated the amino acid into proteins at the specified site. In order to use such a system for incorporation of nonnatural amino acids, the supressor tRNA must be orthogonal to the endogenous aminoacyl-tRNA synthetases of the cell-free translation system, as proof-reading (deacylation of the noncognate amino acid) and re-aminoacylation with the cognate amino acid could occur if the suppressor tRNAs were recognized by the aminoacylation synthetases [8]. An orthogonal amber suppressor tRNA bearing the CUA anticodon derived from yeast phenylalanine-tRNA was first used for incorporation of a variety of nonnatural amino acids into specific sites into proteins in *E. coli* cell-free systems [9,10]. Although this yeast suppressor tRNA cannot be recognized by any aminoacyl tRNA synthetase in the in vitro translation system, which reduces the risk of enzymatic deacylaion and/or reacylation of undesired amino acids, it is not compatible to the *E. coli* protein biosynthesis system, resulting in poor suppression efficiency. In particular, amino acids bearing highly polar sidechains are generally incorporated in low yield. [9,10]. Based on the assumptions that tRNAs with an anticodon mutation are no longer recognized by their cognate aminoacyl-tRNA synthetases but can efficiently be accepted by the ribosomal translation system to incorporate noncognate amino acids, Schultz and co-workers generated an *E. coli* $tRNA^{ASN}$ – derived suppressor which has poor in vivo suppression efficiency [11] to improve the efficiency of incorporation of both polar and non polar amino acids [12].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Frameshift suppression of four-base codons can also be used to introduce nonnatural amino acids into proteins. Compared to the amber codon method in which competitive binding by the release factor results in premature peptide chain termination, a four-base codon strategy decreases this competition and favors to maintain the proper reading frame only when the extended codon is translated by the suppressor tRNA to obtain a full-length protein [13],[14],[15]. Sisido and co–workers used a corresponding four-base anticodon yeast $tRNA^{Phe}$ to incorporate many phenylalanine derivatives into streptavidin by using *E. coli* cell-free system. Using rarely used codons such as CGGG and GGGU as the 4-base anticodon, the efficiency of incorporation of 2-naphthylalanine and 7-amino-4-nitrobenz-2-oxa-1,3-diazole (NBD) derivative of lysine into streptavidin was improved from 9% to 64% [15].

In the techniques for site-specific incorporation of nonnatural amino acids in an *in vitro* translation system, aminoacylation of tRNA is the crucial step. Chemical misacylation of tRNA was first reported by the Hecht group [16]. In this method, tRNAs lacking the 3'-terminal dinucleotide, pCpA, were ligated with aminoacylated pCpA by T4 RNA ligase to form misaminoacylated tRNAs. Since then, several improvements have been made to facilitate the preparation of misacylated tRNA. It was found that N-protected amino acid cyanomethyl ester could selectively monoacylate the 2',3'-hydroxy groups of the dinucleotide with

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

high yield [17]. Using deoxycytidine instead of cytidine in pCpA made the synthesis easy and eliminated the reactive 2'-OH group on cytidine without affecting the biological activity. Prior to acylation of the dinucleotide, the α-amino group and any reactive side chain group of the amino acid are protected as their nitroveratryloxy carbamate, ester, or ether derivatives [18,19]. After removal of the N^{α}-protection group from the ligated aminoacyl-tRNA by photochemistry, the aminoacyl tRNA is ready for translation. Recently, the Hecht group developed a new amino protecting group, 4-pentenoyl, for aminoacylation of tRNA. This group can be removed by treatment with iodine under mild chemical conditions [20].

In spite of several methods having been made to improve the chemical aminoacylation methods, there are some limitations: mainly the poor yields of the aminoacylation reaction and multiple steps involved to prepare aminoacylated tRNA. To eliminate these limitations, a new methodology was recently developed to prepare the aminoacyl-tRNA through the catalysis of an aminoacyl-tRNA synthetase-like riboyzmes. "The RNA world" hypothesis [21] suggests that RNA has not only genomic function but catalytic function as well at an early stage of evolution of life. For example, a structural study [22,23] has revealed that ribosome is a ribozyme and may function as a catalyst in the primitive translation system. Another postulated catalytic function of ribozyme is the aminoacylation of tRNAs [24,25]. Based on the above hypothesis and discoveries, Saito and co-workers devised and *in vitro* selected several aminoacyl-tRNA synthetase-like

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ribozymes. One of these ribozymes was generated from a 5' leader sequence of a precursor tRNA capable of charging Phe onto its own 3'-end [26]. Aromatic amino acids such as phenylalanine (Phe) and tyrosine (Tyr) have been successfully charged onto the 3'-OH group of a tRNA specifically by this ribozyme [27]. Since the aminoacyl-tRNA synthetase-like ribozyme (ARS ribozyme) is a de novo catalyst, it could in principle aminoacylate a wide range of nonnatural amino acids onto any tRNAs. As a result of this idea, an improved and highly flexible tRNA aminoacylation synthetase-like RNA system called Flexizyme was devised. This artificial flexizyme can only recognize the 3'-end of tRNA, 5'-CCA-3', so that a variety of tRNA substrates can be aminoacylated [28]. Using this highly flexible acylation system, several nonnatural amino acid derivatives, such as non-natural *a-L*-aminoacids, *a-N*-methyl *L*-amino acids, *a-N*-acyl *L*-aminoacids, *a-D*-amino acids, β-amino acids, and α-hydroxyacids, were charged onto the $tRNA^{Asn}_{CUA}$.

1.1.2 Biosynthesis of proteins bearing nonnatural amino acids *in vivo*

Low sustainability is a major problem of the above *in vitro* biosynthetic methods, which limits the production yield of the mutant protein. It is therefore highly desirable to have methods that allow incorporation of nonnatural amino acids directly into proteins *in vivo*.

In nature, aminoacylation of tRNAs is catalyzed by a set of protein enzymes called

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

aminoacyl-tRNA synthetases (ARSs). These ARSs specifically assign each amino acid to the cognate tRNA to insure high fidelity in protein synthesis. The ability to incorporate nonnatural amino acid *in vivo* requires a unique tRNA-codon pair and a corresponding aminoacyl-tRNA synthetase [8] that are genetically introduced into the organisms to specify a particular nonnatural amino acid. Such a tRNA-condon pair and the cognate aminoacyl-tRNA synthetase must work orthogonally with respect to the endogenous factors of the host.

Using this principle, a number of orthogonal tRNA/synthetase pairs have been generated for use to incorporate nonnatural amino acids genetically in bacteria [29], yeast [8,30] and mammalian cells [31]. A large number of structurally diverse nonnatural amino acids with interesting physicochemical and electronic properties have been shown to be efficiently incorporated into various proteins. This method holds tremendous promise as a powerful tool for studying protein structure and function both *in vitro* and *in vivo*, and for generating proteins with new or enchanced properties [32].

1.2 chemical approaches

There are certain inherent limitations with respect to the use of biosynthetic protein production systems. For instance, many proteins have deleterious effects on the host cells and expressing such proteins with the recombinant DNA technology is

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

problematic, as is often the case with the membrane proteins. In addition, however to accommodate the ribosomal protein translation machinery, there is a limit as to the extent it can tolerate a novel amino acid. For example, if a nonnatural amino acid is too drastically different from the natural ones, it may not be accepted by the ribosomal machinery for incorporation into proteins. Furthermore, no prokaryotic or eukaryotic expression systems are yet available that allow simultaneous incorporation of two or more different nonnatural amino acids into a protein. On the other hand, it is clear that proteins in the cells often need to undergo a series of modifications to carry out their functions. Alternative methods are therefore needed for the preparation of these multi-modified proteins to facilitate structural and functional characterization by biochemical and biophysical means. Chemical approaches are much more versatile than biosynthetic methods, as they allow the synthesis of proteins with any structural elements, such as nonnatural amino acids and architectures.

Since its introduction in 1963 [33], solid phase peptide synthesis (SPPS) has been widely used for the synthesis of peptides and small proteins containing a large variety of structural units. Using SPPS, peptides can be easily synthesized in repetitive steps of amino acid assembling. However, because of the imperfection of the chemistry used for peptide synthesis and possible interchain aggregation of the long growing protected peptide chain, it is still a challenge to use SPPS for the synthesis of very large peptides or proteins. For this reason, the convergent

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

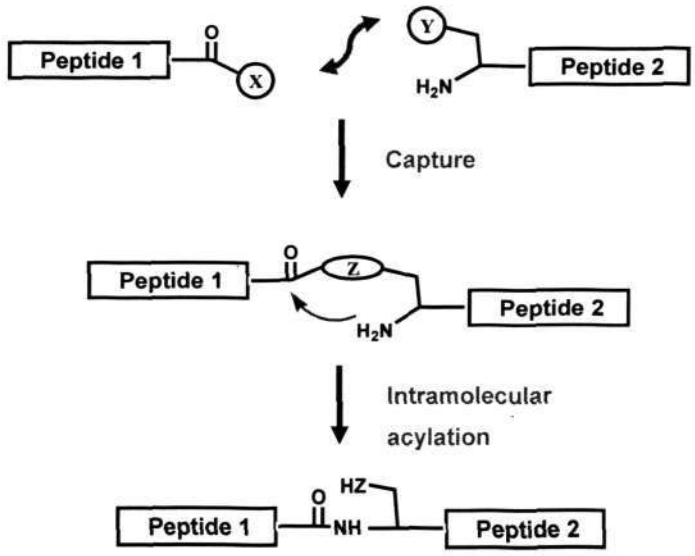
segment condensation methods appear attractive from a strategical point of view.

The method developed by Blake, et al [34] is a very useful segment condensation method, as it involves the use of a unique C-terminal thiocarboxyl group for coupling with an N-terminal amine. The condensation between two peptides is achieved through selective activation of the thiocarboxyl group by silver ions. β-lipotropin, α-inhibin-96, human pancreatic growth hormone-releasing factor and two omission analogs [35] were successfully synthesized by using this strategy. An alternative thioester-based method can be employed to replace the thiocarboxylate which is not very stable and gradually decomposes by oxidative hydrolysis [35]. A thioester can also be activated in the presence of silver ion. The activation of thiocarboxylate or thioester by silver ions is specific and works in the presence of other free carboxyl groups on the amino acid side chains [36].

However, these condensation methods still require partially protected peptides as building blocks, and the efficiency of the intermolecular coupling reaction is inherently low for large peptide segments owing to the difficulty of achieving high molar concentrations for high molecular weight compounds [37]. Nevertheless, if the local concentration of the reacting carboxyl and amino groups can be increased by bringing the two peptides together, the proximity effect may facilitate the peptide bond formation reaction. It was first found in 1953 that peptide bond formation between an amino acid thioester and a cysteine derivative could be

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

brought about via an intramolecular acyl transfer reaction [38]. Therefore, the same principle has been applied to the development of the currently known peptide ligation methods, which share a common scheme as shown in Scheme 1.1.



Scheme 1.1 Proximity-driven peptide bond formation.

The ligation-based peptide coupling strategies have some distinct advantages. As a separate and efficient capture reaction joins together the two relatively weak mutual reactivities of the C-terminal α-carboxyl of one segment and N-terminal α-amino group of the other, the subsequent peptide bond formation reaction follows first-order kinetics. Factors such as steric hindrance by a bulky α-substituent on either of the reacting amino acid residues on two peptides will less affect the acyl transfer compared with corresponding intermolecular reactions. The close proximity of the terminal amino and terminal acyl groups results in a high

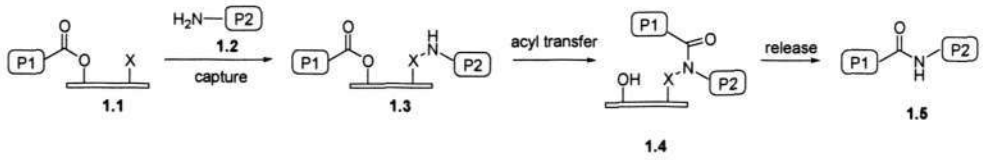
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

local concentrations of the two, offering a strong entropic advantage to the system and thereby removing the need for enthalpic activation of the acyl group [39]. Because the acyl transfer reaction involves a relatively unactivated group, side-reactions such as epimerization are minimized, and it is also possible to use unprotected peptide segments to conduct ligation reaction in aqueous solution [3]. The following reviews the various chemical ligation methods that have been developed based on this general concept.

1.2.1 Prior chemical ligation

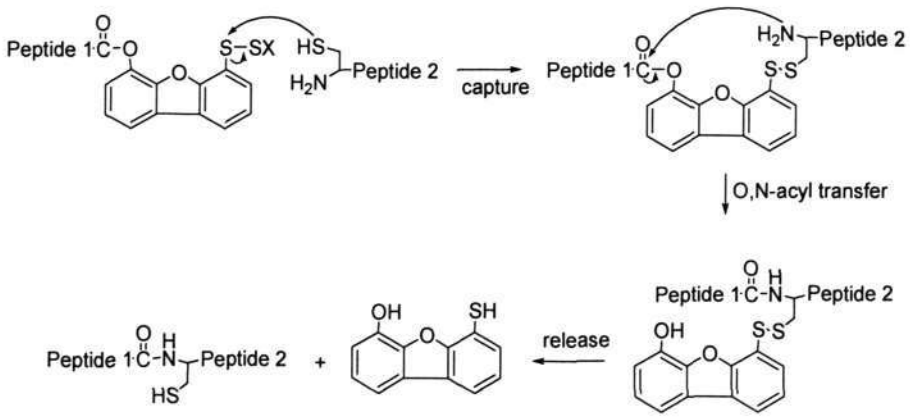
The first ligation method put into the practice of peptide synthesis was developed by Kemp *et al.* in 1975 [40] as the *Prior Amine Ligation Strategy*. The original concept is outlined in Scheme 1.2. The C-terminus of the N-terminal peptide **1.1** is a weakly activated ester. The group of X is positioned near the hydroxyl portion of the ester, providing a capture site for binding the amino component to bring the acyl and amino groups into proximity. Capture happens in the presence of C-terminal peptide **1.2** to give **1.3**. The intramolecular acyl transfer takes place to link the two peptides together to form **1.4**. Finally, removal of the capture-assisting moiety releases the full length peptide **1.5**. However, several limitations made this method to be impractical. Intramolecular acyl transfer may not be effective because of unfavorable steric hinderance. And the amino group may not be nucleophilic enough to achieve rapid capture at high dilution.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



Scheme 1.2 Prior amino capture strategy.

A modification of this initial strategy led to the *Prior Thiol Capture Strategy* [41]. In contrast to the amino group used in the prior amine capture strategy, the thiol is strongly nucleophilic and greatly enhances the capture prior to intramolecular acyl transfer, which works in high dilutions. The acyl transfer in this system might proceed through a transition state of at least a nine-membered ring (Scheme 1.3).



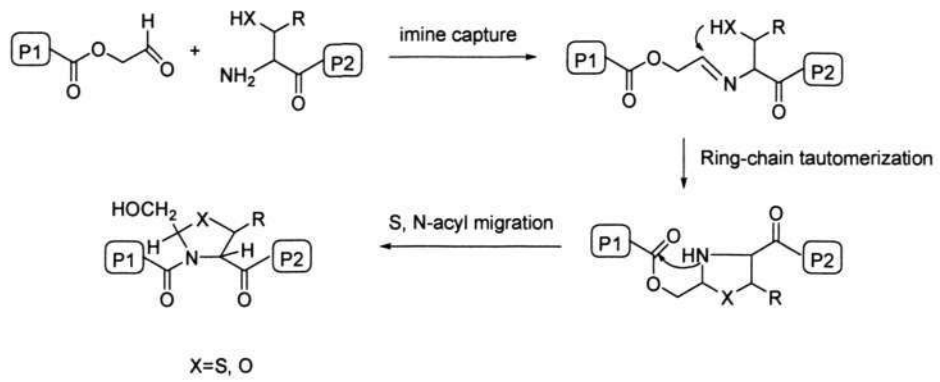
Scheme 1.3 Prior thiol capture ligation.

1.2.2 Ligation by thiazolidine/oxazolidine capture

This ligation method first developed by Liu [37] involves formation of a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

thiazolidine/oxazolidine as the capture device between a C-terminal peptide-ester aldehyde and an N-terminal 1,2-thiol/hydroxyl amine group. The capture reaction is especially efficient for the 1,2-thiol amine moiety, as the aldehyde group selectively condenses with an N-terminal cysteine residue to form a stable thiazolidine intermediate. Subsequently, the acyl transfer reaction takes place through a five-membered ring transition state to give a proline-like imidic bond (as shown in Scheme 1.4).



Scheme 1.4 Outline of ligation through thiazolidine/oxazolidine capture.

Although there are potentially many different ligation products which may generate from nucleophilic attack on the aldehyde by different amino acid side chains, only the thiazolidine is stable during ligation and present in a significant amount. Ligation at acidic pH can avoid ester hydrolysis and side reactions with other nucleophiles, such as side chains of lysine or arginine [42]. Once the ligation intermediate forms, the intramolecular acyl transfer can be facilitated by increasing the reaction pH.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

There are three ways to attach the glycoaldehyde moiety to the carboxyl terminus of the N-terminal peptide. SPPS provides a synthetic peptide which is protected at its N-terminus and cysteine side chain, and has an ester at its C-terminus. After side chain deprotection and cleavage from resin, the minimally protected peptide is subjected to trypsin catalysed coupling to the dimethoxyethyl ester of alanine. The glycoaldehyde-derivatized peptide can be obtained by TFA treatment [42]. Another approach [43] to preparing glycoaldehyde-derivatized peptides can avoid limitations imposed by substrate specificity of the enzymatic method. In this method, an unprotected thioester-derivatized peptide is treated with a large excess of an appropriate amino acid derivative carrying the protected glycoaldehyde in the presence of silver ion, and final TFA treatment gives the desired peptide with a free aldehyde. A third method is to use solid phase synthesis to prepare the 1,2-diol precusor form of the glycoaldehyde which is then oxidized to give the peptide glycoaldehyde ester [44].

1.2.3 Thioester-mediated ligation

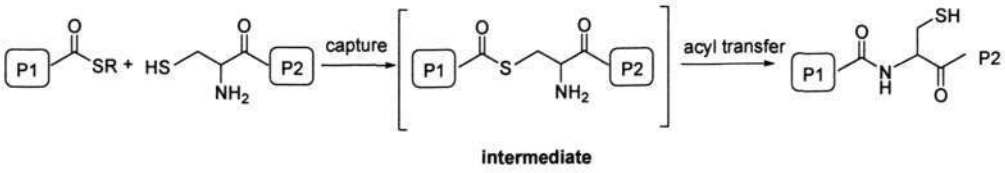
This strategy contains several ligation methods characterized by formation of a thioester as the reaction intermediate.

1.2.3.1 Native Chemical Ligation

As mentioned earlier, facile peptide bond formation was observed by Wieland [38]

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

when reacting a thioester with a cysteine compound. It was proposed that a thiol-thioester exchange reaction first took place, which formed a new thioester between the acyl group and the side-chain thiol of cysteine and which was followed by an intramolecular S-N acyl transfer for peptide bond formation. Kent *et al.* [45] adopted the reaction to devise a peptide ligation method which they named as *Native Chemical Ligation* (Scheme 1.5). So for the two components of ligation reaction, one has a cysteine residue at its N-terminus and the other has a thioester at its C-terminus. The thiol group of the cysteine residue participates in an exchange reaction with the thioester in a highly chemoselective manner to form a new thioester. This thioester intermediate spontaneously rearranges, resulting in the formation of an amide bond between the two peptides with the cysteine residue at the ligation site.



Scheme 1.5 Native chemical ligation.

Homo-cysteine can also be used for ligation, which can be further methylated to form a methionine residue at the coupling site [46]. Preparation of peptide thioalkylesters by direct SPPS further enhanced the practical value of this ligation chemistry [47]. Moreover, addition of a mixture of tris(2-carboxylethyl)phosphine (TCEP) and 3-mercaptopropanoic acid maintains the reducing environment of the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

ligation reaction to prevent N,S-bisacylated byproduct and reduced disulfide formation, improving therefore the overall ligation yield [47].

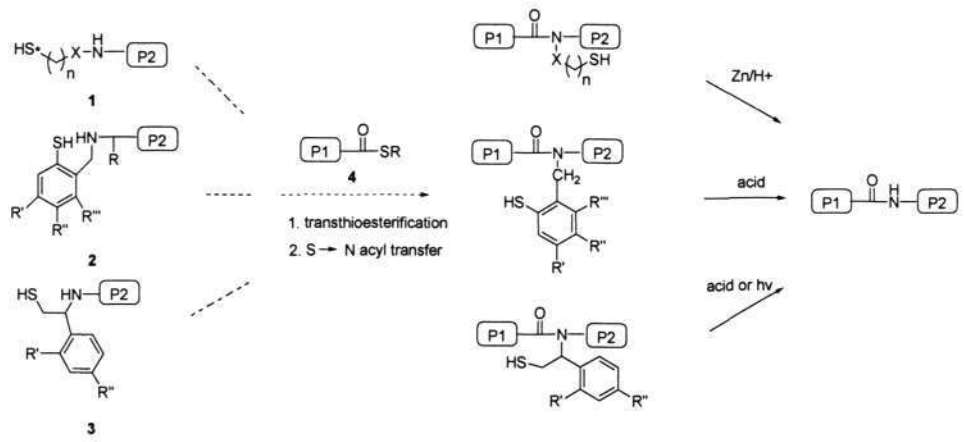
pH affects the reaction rate of ligation. At neutral or slightly basic pH, ligation proceeds faster and is more complete than at weakly acidic pH, showing that the thiolate of cysteine is the reactive nucleophile. It was also shown that the nature of the C-terminal amino acid in the peptide-thioester influences the ligation reaction. Ligation goes faster for less hindered C-terminal residues like Gly- or Ala-COSR than for sterically hindered residues, such as β-branched Ile- or Val-COSR [48].

The Kent group [49] studied the effects of thiol catalysts on the ligation rate. Among a set of different thiol catalysts, phenyl thiols are superior ligation catalysts compared to alkanethiols. Although, for alkanethiols such as MESNA and benzyl mercaptan, thiol-thioester exchange is rapid and not rate limiting, they are not good leaving groups. Therefore, transthioesterification by the N-terminal cysteine becomes rate limiting. To facilitate the reaction, an optimal phenylthiol catalyst which can undergo rapid and complete thiol exchange with the peptide-thioester while maintaining high reactivity toward cysteine transthioesterification can be used to catalyze the ligation reaction.

The scope of the ligation chemistry can be expanded by the use of removable thiol-based auxiliaries. The auxiliaries are linked synthetically to the N-terminal

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

amine of the second peptide. Once the two peptides are ligated together, the removal of the extraneous atoms of the auxiliary gives the final ligation product (as shown in Scheme 1.6).



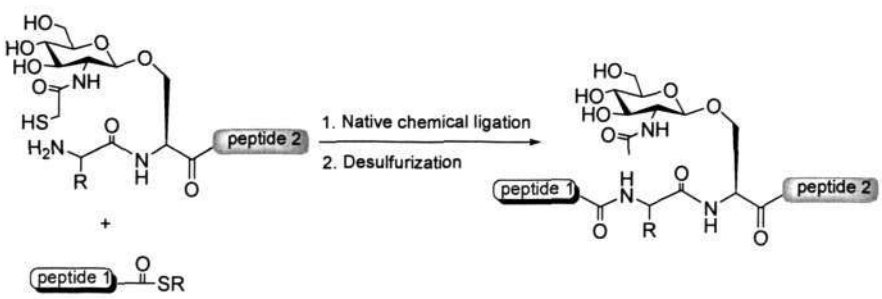
Scheme 1.6 Auxiliary-mediated peptide ligation.

N^{α}-ethanethiol and N^{α}-oxyethanethiol were the first auxiliaries used for peptide ligaton. The coupling between two glycine residues was the most efficient, while there was no ligation detected between two nonglycyl residues. After ligation, N^{α}-oxyethanethiol could be released from the peptide by reduction with zinc under acidic conditions [50]. The second class of auxiliaries is based on N^{α}-2-mercaptobenzylamine. Two peptides are linked together by the help of the auxiliaries in a manner similar to native chemical ligation, and then acyl transfer occurs via a six-membered ring intermediate. Removal of the auxiliaries can be done under acidic conditions [51,52]. The third class of auxiliaries, N^{α}-(1-phenyl-2-mercaptoethyl) group bearing a more nucleophilic alkyl thiol, can

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

enhance the transthioesterification in the capture step and mediate acyl transfer via a five-membered ring intermediate [53]. In general, the use of auxiliaries leads to increased steric hindrance at the transition state, limiting ligation mainly to the Gly-Gly junction [54].

Recently, a new method for the synthesis of cysteine-free glycopeptides using an auxiliary-assisted ligation was developed by Wong *et al.* [55]. In this ligation (see Scheme 1.7), a glycopeptide is modified at the 2nd residue with a carbohydrate (*N*-acetyl glucosamine) derivatized with a mercaptoacetate auxiliary. After linking two peptides through thiol-thioester exchange, an $S \rightarrow N$ acyl transfer proceeds to give a ligated product with a native peptide backbone. The final product can be obtained by desulfurization [56]. It has been applied to the total synthesis of the antibacterial glycoprotein diptericin.

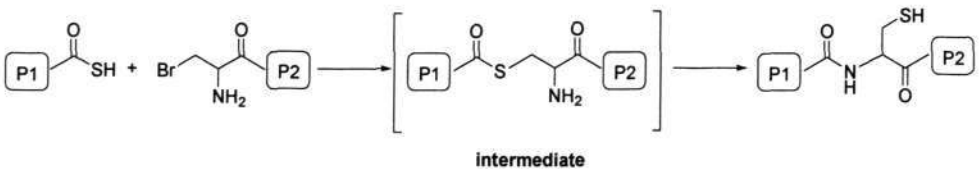


Scheme 1.7 Glycopeptide ligation through a side-chain auxiliary.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

1.2.3.2 Ligation by thioalkylation capture

Thioester formation as the capture reaction can also be achieved through thioalkylation between a C-terminal thioacid and an N-terminal bromoethylamine moiety [47]. In this case, the electrophile, N-terminal-β-bromoalanine, is *S*-alkylated by a nucleophile thioacid to form a covalent thioester intermediate. Acyl transfer proceeds rapidly and rearranges to give a ligation product with a cysteine introduced at the ligation site, at which the cysteine is not initially present. Scheme 1.8 shows the outline of this ligation method. This ligation reaction should be performed at the pH below 5 to eliminate side reactions.



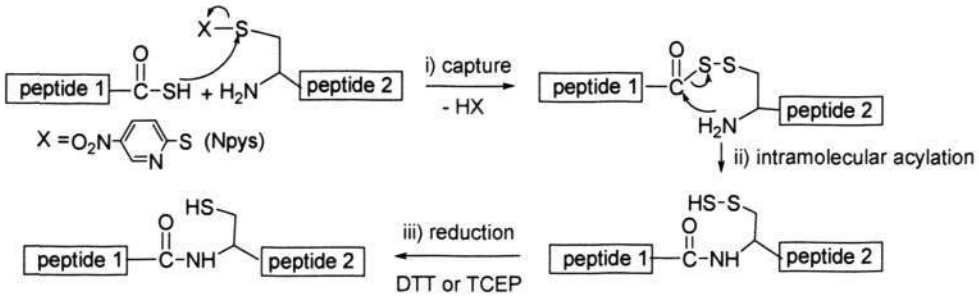
Scheme 1.8 Peptide ligation via thioalkylation capture of bromoalanine.

1.2.4 Ligation by disulfide exchange (thioacid capture ligation)

In this ligation method, as shown in Scheme 1.9, the N-terminal Cys thiol group of one peptide is activated by 2-mercapto-5-nitropyridyl (Npys), and attacked by the nucleophilic sulfur atom of a C-terminal thioacid of the other peptide. The acyl-disulfied intermediate then undergoes rearrangement through a six-membered ring transition state to give a ligation product, in which a hydrosulfide is added to the cysteine thiol. The final ligation product can be generated by removing the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

hydrosulfide using a reducing agent [57]. A 32-residue model peptide was first synthesized by this disulfide ligation method by coupling a 17-residue Npys modified C-terminal peptide and a 15-residue N-terminal peptide thioacid. The capture and acylation was performed in a mixed acetonitrile-water solution. Because of the activated nature of the Npys-S bond, strong nucleophilicity and low *p*Ka of the thioacid compared to the alkylthiol of cysteine, the capture reaction proceeds very fast and completes almost instantaneously at pH 2-3. After adjusting the pH to 5-6, efficient acyl transfer occurs through a six-membered ring intermediate, resulting in 90% of a hydrosulfide-derivatized peptide in only 5 min. The desired ligation product is obtained by treatment with dithiothreitol.



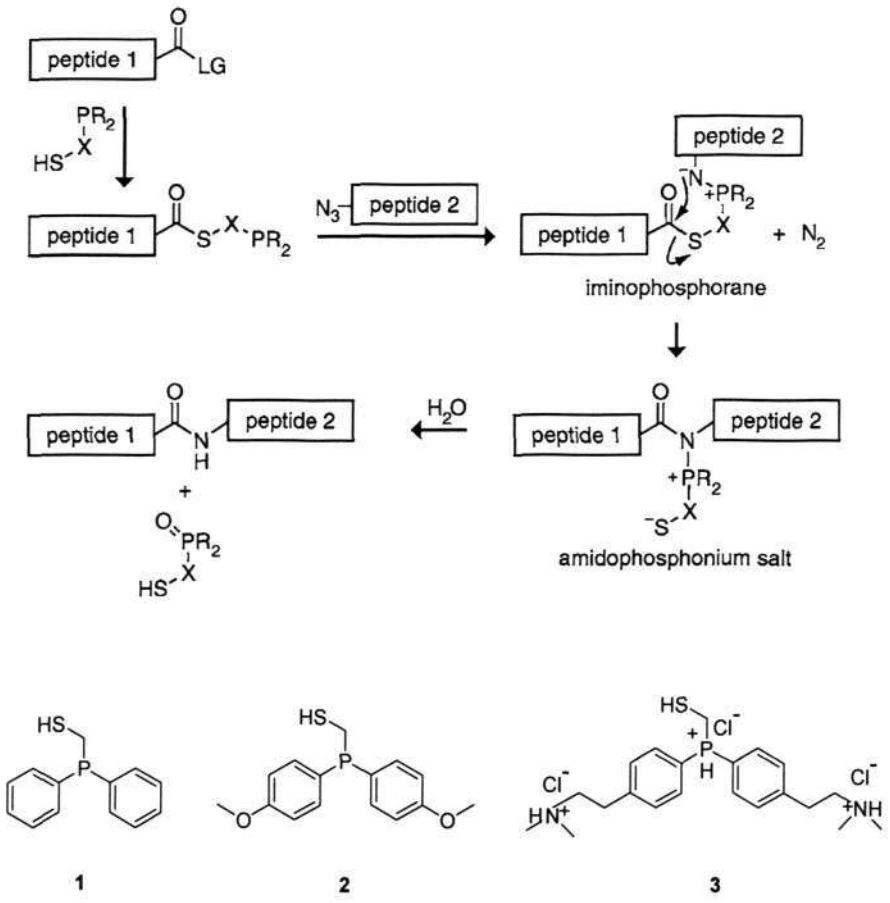
Scheme 1.9 General scheme of thioacid-capture ligation.

1.2.5 Staudinger ligation

Most of the developed ligation methods involve a thiol group to initiate the ligation reaction and keep it at the junction of the ligation product. The Staudinger reaction

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

provides a simple alternative to these methods with the promise of removing the requirement for a thiol.



Scheme 1.10. Outline of Staudinger ligation (adopted from [58]) and phosphinothiol compounds for mediation of peptide ligation.

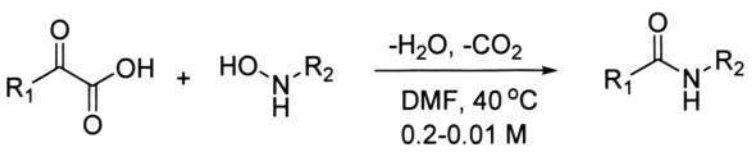
A traceless Staudinger ligation has been developed for protein synthesis based on the Staudinger reaction: $PR_3 + N_3R' + H_2O \rightarrow O{=}PR_3 + H_2NR' + N_2$ (g) [59,60]. As shown in scheme 1.10, a C-terminal phosphinothioester of one peptide reacts with an N-terminal azide of the other, giving an iminophosphorane intermediate. An amidophosphonium salt is formed by attack of the iminophosphorane nitrogen

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

on the thioester carbonyl. Hydrolysis of the amidophosphonium salt then releases phosphineoxide and results in ligation of the two peptides with an amide bond and releases a phosphine oxide.

Several phosphinothiol compounds have been studied for the ability to mediate ligation (Scheme 1.10). (Diphenylphosphino)methanethiol **1** [60] is the most efficient compound for ligation with a glycine residue at junction, while compound **2,** *p*-methoxy-substituted phosphinothiol, can give higher yield of ligation at a nonglycine residue [61]. Since these compounds are insoluble in water, the ligation can only be performed in organic solvents or organic/aqueous mixture. In order to expand the utility of the Staudinger ligation, Raines *et al* [62] developed several water-soluble phosphinothiol compounds which enable the ligation in water. The most efficacious compound is **3** bis(p-dimethylaminoethyl)phosphinomethanethiol (see Scheme 1.10) which can mediate rapid ligation in water and also perform a transthioesterification reaction with a intein-mediated protein splicing thioester.

1.2.6 Peptide bond formation by decarboxylative condensation



Scheme 1.11 Amide formation by decarboxylation.

Another chemoselective peptide bond reaction without the requirement of the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

relatively rare cysteine residue for ligation is proposed recently by Bode, *et al*. [63]. As shown in scheme 1.11, N-alkylhydroxylamines react with α-ketoacids to form a hemiaminal which further undergoes oxidative decarboxylation to give an amide product. Several ketoacids and hydroxylamines were tested for this reaction. Although the detailed reaction mechanism remains to be elucidated, nucleophilic attack of the hydroxylamines towards the ketone is likely the key capture reaction that serves to bring the two components together. This is a mechanistically novel amide bond formation reaction which takes place without any coupling reagents, catalysts and with only water and CO_2 as by products.

1.3 Objectives of this study

As a general goal of my thesis project, I aim to make an original contribution to this area of research by developing new chemistries for peptide and protein synthesis. Specifically, my work is focused on the chemistry of a unique functional group, the thioacid, and the exploitation of it in the chemical synthesis and modification of proteins.

Among all ligation methods developed so far, thioacid capture ligation is the most efficient in terms of reaction rate. However, the difficulty in obtaining the thioacid building blocks limits its practical use in peptide and protein synthesis. In this study, we set to overcome this problem by developing new methods to prepare

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

peptides and proteins containing a C-terminal thioacid functionality. The methods are built upon a simple but effective hydrothiolysis reaction of thioesters which was previously developed in our lab [64]. This part of work constitutes the first objective of this study and is the content of Chapter 2 of this thesis.

The second objective of this study is to use the thioacid peptides or proteins prepared by our new methods for the synthesis of proteins by using thioacid capture ligation. To demonstrate the synthetic value of thioacid capture ligation, histone proteins were chosen as the target molecules. Histone proteins are used in eukaryotic cells to package DNA into highly ordered chromatin structures. Chromatin activity is closely linked to the state of histone modification. These post-translational histone modifications include acetylation, methylation, phosphorylation, ubiquitynation, sumoylation, and ribosylation on specific amino acid side chains such as Lys, Arg and Ser [65]. Most of these modifications are concentrated on the flexible N-terminal tails of histones. To study whether and how a particular or a particular set of modifications affect chromatin structure and evoke chromatin-based functions, it is essential to obtain these histone analogues containing one or more such modifications in homogeneous forms which cannot be easily prepared by the recombinant DNA technology alone. So, the synthesis of histone proteins can be used as a good example to demonstrate the utility of thioacid capture ligation in protein synthesis. Chapter 3 describes the details of this part of research.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The third objective of this study, which is the subject of chapter 4, focuses on the use of thioacid functional groups for peptide and protein modifications. Thioacid is known to be able to react with certain organic azides to form an amide bond. Since it was first described in 1980, thioacid/azide amidation has only involved small organic thioacids and azide compounds. In this study, we have successfully employed this reaction to tag small organic biophysical probes or large polymers to the C-termini of peptides and proteins. As such, this represents a novel strategy for protein C-terminus-specific modification.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 2: Methods for the Synthesis of Thioacid Peptides and Proteins

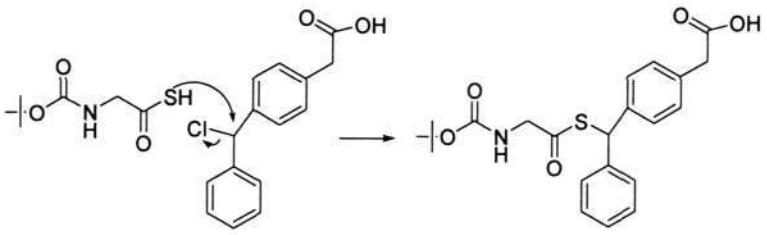
2.1 Introduction

The thioacid functionality is a very soft and powerful nucleophile which has long been exploited for peptide synthesis. Early uses of thioacids were found in peptide segment condensation where a minimal side chain protecting scheme could be employed owing to thioacid's unique ability to be activated by a heavy metal ion or as an acyl-aryl disulfide [34]. At weakly acidic condition, the thioacid (p*Ka*=3) remains a good nucleophile while other nucleophilic groups in peptides and proteins are inactive. This property makes thioacid capable of reacting specifically with electrophilic groups in unprotected peptides in aqueous solution, which was used in the development of several chemoselective conjugation or ligation methods [47,57]. Thioacids also have a unique reactivity towards organic azides, which can be used to synthesize complex amides [66].

Despite the great synthetic value of the thioacid functionality, the synthesis of peptide thioacids remains a difficult task. Traditionally, the only method for solid phase synthesis of peptide C^{α}-thioacids is based on the original Boc SPPS method developed by Blake [34]. Boc-amino thioacid, prepared from Boc-amino acid, is first reacted with a special benzydryl linker **2.1,** to form the Boc amino acid thioester linker **2.2** which is then loaded to a solid support. After standard

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

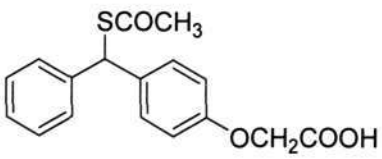
Boc-SPPS, the peptide is released from the resin as a thioacid by treatment with hydrogen fluoride. This method requires preparing the Boc-amino acid thioester linker for each amino acid in solution before loading it to the solid support.



2.1 2.2

Scheme 2.1 Preparation of Boc-aminoacid thioester benzhydryl linker.

Alternatively, the linker can be prepared as acetylated benzhydryl mercaptan **2.3** (Figure 2.1). After being loaded to the resin, the acetyl group is removed by piperidine/β-mercapoethanol and the first Boc-amino acid coupled to the exposed thiol group [67].



2.3

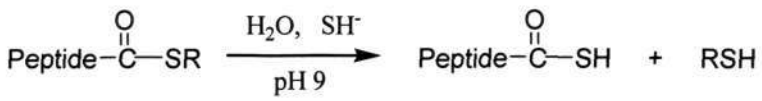
Figure 2.1 Chemical structure of 4-[α-(*S*-acetyl)mercaptobenzyl]phenoxyacetic acid.

However, broad use of this SPPS method has been limited by the need for HF cleavage which requires a special apparatus. In addition, the Blake benzhydryl

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

linker is still not commercially available. Most seriously, the yield of peptide thioacids is usually low because the cationic electophiles released from the various side-chain protecting groups tend to attack the supernucleophilic thioacid functionality.

Beside Boc-SPPS chemistry for the synthesis of peptide thioacids, a new enzymatic method was developed recently. This method makes use of a cysteine protease-like enzyme, subtiligase, to convert certain peptide glycolate esters to thioacid in aqueous solution containing HS^- ions [64]. In this enzymatic reaction, the thiol group of cysteine in subtiligase makes a nucleophilic attack on a peptide glycolate ester substrate, forming an acyl-enzyme intermediate which has a thioester bond between the peptide and the enzyme. This thioester bond is further cleaved by a hydrosulfide anion nucleophile, giving a peptide thioacid. However, the enzyme has its substrate specificity, which limits the application scope of this subtiligase-catalyzed thioacid formation reaction.



Scheme 2.2 Conversion of peptide thioester to thioacid in aqueous buffer.

To overcome this limitation, a more general and efficient method to produce peptide C^{α}-thioacids through hydrothiolysis of thioesters was also developed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

recently in our lab [64] (Scheme 2.2).

A number of small peptide thioesters were tested for the hydrothiolysis reaction in aqueous system using NaSH, Na_2S, or $(NH_4)_2S$ as the thiol nucleophiles, and all of these were found effective. Since the pK_{a1} of H_2S is 6.9, the reaction was performed under weakly basic condition to keep a high enough thiol concentration for thiolysis and avoid the hydrolysis of the thioester at high pH (>9). The reaction provides a convenient and highly efficient method to prepare peptide thioacids. Because there exists many methods by both Boc and Fmoc chemistry for preparing peptide thioesters, the hydrothiolysis reaction has the potential to overtake the traditional method of Blake [34] for the synthesis of peptide thioacids.

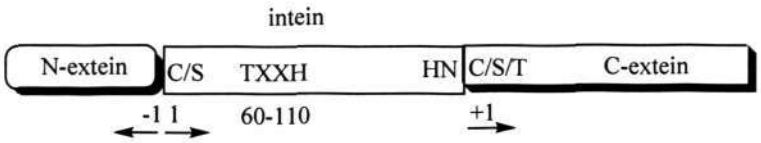
Although the above chemical methods are useful for preparing small to medium sized thioacid peptides, it is not practical to use these methods to prepare large thioacid peptides or proteins. A viable thioacid protein synthesis method would therefore rely on a biosynthetic process which can provide a protein thioester precursor. Protein splicing generates a thioester intermediate which can be intercepted or captured by a thiol nucleophile.

The first protein splicing phenomenon was described by Anraku and coworkers [68] in 1990. When the catalytic subunit of a vacuolar membrane H^+-translocation ATPase (VMA) was expressed from a *Saccharomyces cerevisiae* gene, a product of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

67 kDa was obtained which was different from the deduced molecular weight (119 kDa). This phenomenon predicted a novel protein processing mechanism refered to as "protein splicing".

Protein splicing involves the excision of an intervening polypeptide (the "intein") from a precursor protein and the linking of the flanking N-terminal and C-terminal regions (the exteins) through a peptide bond, generating an active protein. The mechanism is now relatively well understood (see Scheme 2.4).

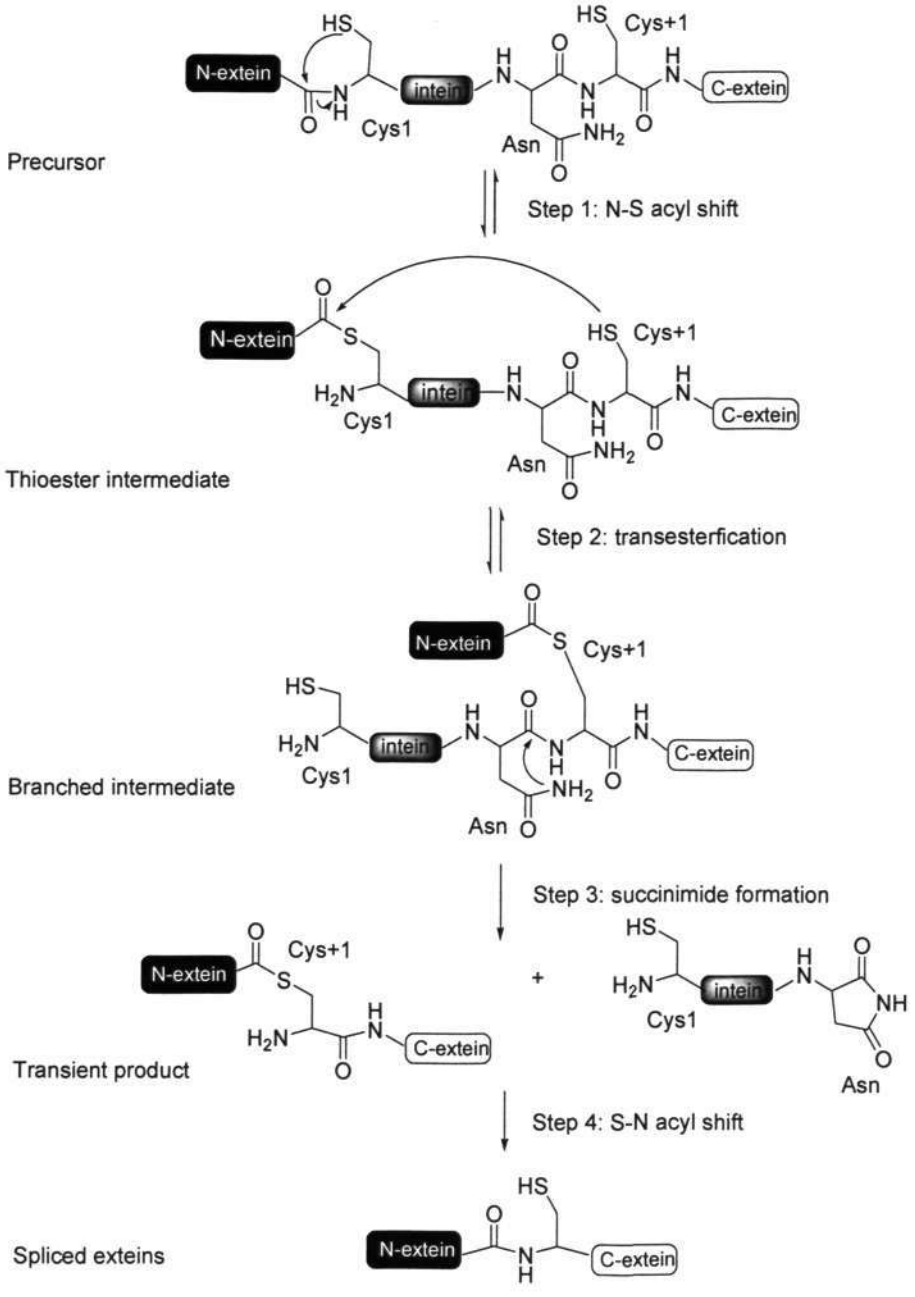


Scheme 2.3 A standard structure of a protein splicing precursor.

A standard structure of a protein splicing precursor is shown in Scheme 2.3. Conserved residues are found at both the splice junctions in standard inteins. A conserved Thr-x-x-His motif is found around 60-100 amino acid from the N-terminus of the intein. Amino acids in the intein are numbered from the N-terminus to C-terminus beginning with 1, while the residues of the N-extein are numbered from the C- to N-terminus beginning with -1. C-extein amino acids are numbered starting with the residue immediately following the intein including a plus sign. Amino acids at the N-terminus of the intein and the C-extein are always residues bearing a nucleophilic side chain such as cysteine, serine or threonine, which can provide a nucleophilic attack on an amide or ester bond. At the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

C-terminus of the intein, asparagine is conserved, which is engaged in spontaneous cyclization reactions resulting in deamidation or peptide bond cleavage.



Scheme 2.4 The proposed chemical mechanism of protein splicing.

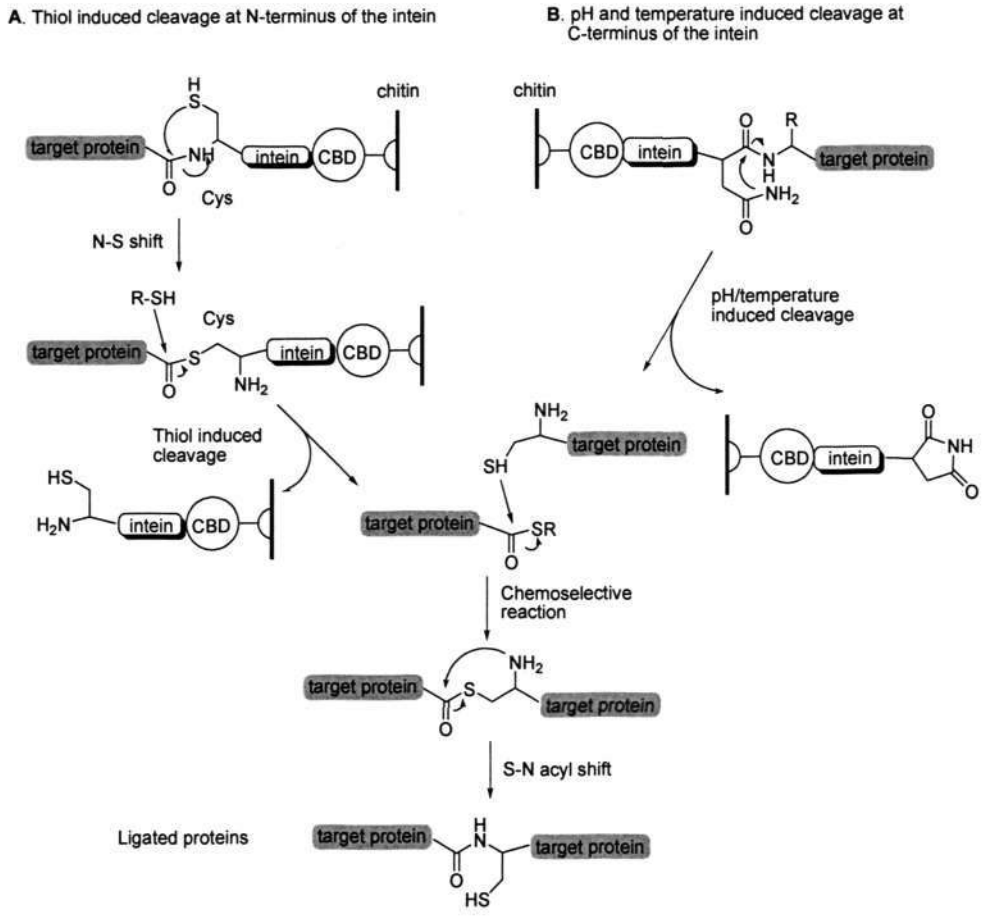
The proposed protein splicing pathway involves four reaction steps [69]. In the initial step, a linear thioester/ester intermediate is obtained by an N → S or N → O

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

acyl rearrangement (which is induced under weakly acidic condition) of the peptide bond between Cys/Ser1 and the –1 aa. This linear intermediate was verified by hydroxylamine cleavage, mass spectrometry and colorimetric assays in expressed proteins [54,70]. In step 2, the side chain of the first amino acid (Cys+1) of the C-extein launches a nucleophilic attack on this linear thioester/ester intermediate giving a branched intermediate by transthioesterification. In step 3, this branched intermediate is cleaved by the attack of amido group of the asparagine side chain to the carbonyl carbon leading to succinimide formation and excision of the intein, as well as formation of a ligated extein intermediate. Finally, in step 4, a spontaneous S → N or O → N acyl rearrangement occurs in the ligated extein intermediate, linking the spliced exteins with a stable peptide bond.

The individual steps of protein splicing can occur at the N- or C-terminus intein-extein junction independently. Mutation of intein's C-terminal residue asparagine to alanine blocked the C-terminal cleavage by preventing succinimide formation [70], while the N-S acyl shift could still take place between the extein and the intein to form a thioester intermediate. The addition of a thiol reagent such as MESNA can induce the thioester exchange resulting in the formation of a reactive thioester protein, which can be the building blocks for subsequent protein ligation reactions (Scheme 2.5).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

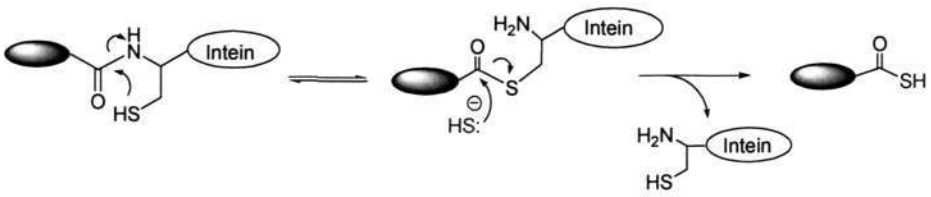


Scheme 2.5 The mechanism of intein-mediated protein ligation.

Using the modified Sce VMA intein, Xu and coworkers developed a novel protein preparation and purification system [71], known as IMPACT. In this system, a target protein is fused to the N-terminus of a modified intein, which is fused to a chitin binding domain (CBD) from *Bacillus circulans*. When the fused protein passes through a chitin resin, it is separated from other cell proteins by binding to the resin. Overnight incubation of the bound target protein-intein-CBD with a nucleophilic thiol such as R-SH results in the cleavage of the peptide bond between target protein and intein. The target protein with a C-terminal thioester finally elutes from the chitin resin.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Several thiol reagents were investigated for the ability to cleave intein fusion proteins [71]. Among them, thiophenol and 1-mercapto-2-ethanesulfonic acid showed high efficiency of cleaving the intein fusion protein, giving a thioester-ending target protein with low hydrolysis. Subsequently, several other intein systems were developed, which gave high yield of cleavage by MESNA or thiophenol. The *Mxe* GyrA intein from *Mycobacterium xenopi gyrA* gene, the *Mth* RIR1 intein from *Methanobacterium thermoautotrophicum* and a mini-intein from the *dna*B helicase gene of *Synechocysis* sp work well with MESNA and have provided key protein thioester building blocks for so-called *expressed protein ligation* [72] which uses the same chemistry as *native chemical ligation*.



Scheme 2.6 Outline of protein thioacid synthesis through hydrothiolysis reaction.

Using the same protein splicing mechanism and hydrothiolysis reaction, protein thioesters can be converted to the related protein thioacids as shown in Scheme 2.6 [73]. These protein thioacids would provide key building blocks for protein synthesis and other chemistry applications.

In this project, we first adapted the solution hydrothiolysis reaction to solid phase to convert resin-bound peptide thioesters to the corresponding thioacids in aqueous

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

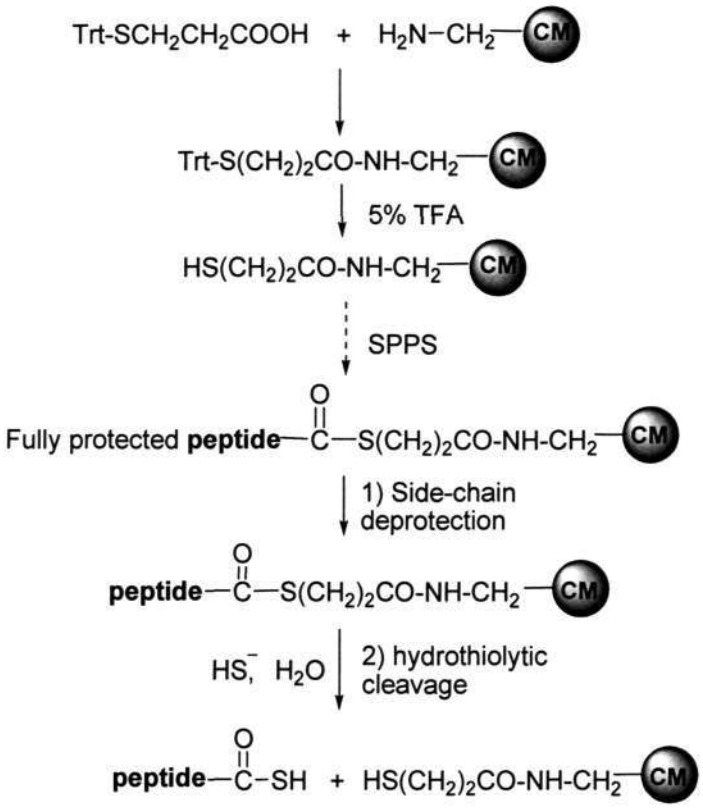
buffer by using a resin that has good water-swelling property. A solid-phase hydrothiolysis reaction would be practically more convenient to use than the solution reaction, as it eliminates the step of isolating and purifying the thioester precursor. Also it will not have the shortcomings of the Blake method with which the thioacid released during the final HF cleavage is susceptible to electrophilic attacks by the highly reactive electrophiles derived from the amino acid side chain protecting groups. In our solid-phase hydrothiolysis scheme, by making use of a non acid-labile linker, one can conduct the usual final deprotection-cleavage experiment in two steps (Scheme 2.7): 1) the acidolytic deprotection to remove all the side chain protecting groups and 2) hydrothiolytic cleavage of the resin-bound peptide thioester to give the desired peptide thioacid.

Several peptide thioacids were made by this solid-phase hydrothiolysis reaction. The solid support chosen is a total PEG-based resin, the ChemMatrix resin, which has excellent swelling property in aqueous media.

Following Scheme 2.6, the same thioester-to-thioacid conversion, *i.e.*, the hydrothiolysis reaction, was applied to the preparation of two protein thioacids, truncated histone H3 C^{α}-thioacid and ubiquitin C^{α}-thioacid. The thioesters of these two proteins were generated as intermediates of the intein-mediated protein splicing process, and were intercepted by hydrosulfide anions to give the protein thioacids. This hydrothiolysis reaction provides an effective and easy way to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

prepare large protein thioacid molecules which are essential building blocks for the semi-synthesis of proteins and are useful for some other applications as well.



Scheme 2.7 Synthesis of peptide thioacids by solid-phase hydrothiolysis reaction. CM=ChemMatrix resin.

2.2 Materials and methods

2.2.1 Materials

All amino acids and coupling reagents were purchased from GL Biochem (Shanghai, China) and Novabiochem (Germany). Aminomethyl-ChemMatrix resin was from MATRIX INNOVATION (Montreal, Canada). All chemical reagents were purchased from commercial suppliers. Plasmid pTWIN1, pTYB1 and restriction enzymes were from New England Biolabs (USA). *E. coli* strain

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

BL21(DE3) was from Stratagene (USA).

2.2.2 Instrumentation

2.2.2.1 HPLC analysis and purification

Analytical and semi-prep HPLC was performed on a Shimazu UFLC HPLC system equipped with a PDA-100 photodiode array detector. A C18 reversed-phase column (Jupiter, 250 x 4.6 mm) and a C4 reversed-phase column (Vydac, 250 x 4.6 mm) were used for analytical analysis. C8 reversed-phase column (Vydac, 250 x 10 mm) and a C4 reversed-phase column (Vadyc, 250 x 10 mm) were used for semi-prep purification. Prep RP-HPLC was carried out on a Waters 2487 HPLC system using a C18 reversed-phase column (Prosphere C18 column, 250 x 22 mm). The analytes were eluted using a gradient mixture of two solvents: Solvent A was distilled deionized water containing 0.05% TFA and solvent B was 90% ACN in distilled deionized water containing 0.05% TFA. The mobile phase flow rate was 1 mL/min for analytical analysis, 2.5 mL/min for semi-prep and 10 mL/min for prep purification. The separation temperature was room temperature.

2.2.2.2 Mass Spectrometry

ESI mass spectra were recorded using a bench-top ion trap mass spectrometer (FINNIGAN LCQ Deca XP MAX) equipped with standard ESI sources. Samples were introduced by direct injection using a syringe. The carrier buffer was

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

composed of two solvents in 1:1 v/v ratio: solvent A (ddH_2O containing 0.045% TFA) and solvent B (90% ACN in ddH_2O containing 0.045% TFA). Nitrogen served both as the nebulizer gas and the drying gas. Nitrogen was generated by a PEAK SCIENTIFIC lab gas generator. Helium served as the damping gas for the ion trap and collision gas for the MS^n experiments. Data analysis software Xcalibur 1.4 was used for processing the spectra.

MALDI-TOF mass spectra were recorded on a 4800 MALDI TOF/TOF Analyzer operating in the MS reflector positive ion mode and using α-cyano-4-hydroxycinnamic acid or sinapic acid as the matrix.

2.2.3 Synthesis of peptide thioacids

All resin-bound peptide thioesters are of the general structure: peptide-CO-SCH_2CH_2CO-CM resin. First, the thio-functionalized CM resin was prepared by coupling Trt-SCH_2CH_2COOH onto commercial aminomethyl CM resin followed by detritylation with 5% TFA. Peptides were then assembled on the thiol-derived resin using typical SPPS protocols. All amino acids were used in 4 eq of the resin, and preactivated by 4 eq of PyBop and 8 eq of DIEA. After 1 hour coupling of the first amino acid residue, the resin was washed with DCM (3 x), DFM (3 x) and DCM (3 x). Capping was performed by using 5% acetic anhydride/2.5% DIEA in DMF/DCM (1:1) for 2 x 10 min. The resin was washed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

with DCM (3 x), DFM (3 x) and DCM (3 x). After capping, the successive α-amino group deprotection steps were performed in 30% TFA/DCM 2 x 10 min between each amino acid coupling step.

Both Fmoc and Boc chemistry were used for the synthesis of solid-phase (CM)-bound peptide thioesters.

Peptide **1** (see Table 2.1 in this chapter) was synthesized by Fmoc (N^{α}-[9-fluorenylmethyloxycarbonyl]) -chemistry SPPS. The following side chain protecting groups were used: tBu for Ser, Thr; Boc for Lys; Pbf for Arg; Trt for Gln. The deprotection of Fmoc [74] was done in the mixture of 1-methylpyrrolidine (25%), hexamethylene imine (2%), HOBt (2%) in NMP-DMSO (1:1), 1 x 10 min, 1x 20 min. The resin was washed with 3 x DMF, 3 x DCM, and 3 x DMF. The couplings were carried out by using amino acids at 3-4 fold excess and the same molar amount of PyBop in DMF/DCM in the presence of DIEA. After 1 hour reaction, the qualitative ninhydrin test [75] was performed to monitor the completeness of the coupling. For attachment of Arg(Pbf), double couplings were performed. After complete assembly of the peptide chain, deprotection of all side-chain protection groups was performed in 95% TFA/ 2.5% TIS/ 2.5% H_2O.

Peptides **2-6** (see Table 2.1 in this chapter) were prepared by Boc-chemistry SPPS. The following side-chain protecting groups were used. Xan for Asn; Bzl for Thr,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Ser; OBzl for Asp; 4MeOBzl for Cys; 2-Cl-Z for Lys; 2-Br-Z, Mts for Arg. Coupling was conducted using a 3-4 fold excess of amino acids and PyBop in DCM. Wash steps were carried out with DCM and DMF. The successive deprotection of Boc on the α-amino group was performed in 30% TFA in DCM. The qualitative ninhydrin test was used to monitor the completeness of the coupling. In the case of Arg(Mts), double coupling was necessary. Final deprotection was conducted by using a cocktail of TFMSA:thioanisole:TFA/1:1.5:10 at room temperature for 1 hr [76].

Peptide **3** was synthesized on the recycled HS-CH_2CH_2-CO-$NHCH_2$-ChemMatrix resin. The used HS-CH_2CH_2-CO-chemmatrix resin was regenerated by treatment of the resin with 5% hydrazine, 1% 2-mercaptoethanol in DMF for 30 min, followed by wash with DCM and DMF. The regeneration of thiol was checked by the Ellman test [77].

After washing with TFA and ACN, the deprotected peptides bound on the resin were released from the resin by hydrothiolysis using 0.1 M Na_2S or 0.2 M $(NH_4)_2S$ in 0.3 M HEPES buffer, pH 7.6 or 8.6, with or without 6 M Gdn-HCl at room temperature . Once the peptide resins are fully deprotected, they should be cleaved by thiol reagent as soon as possible. Otherwise, the deprotected peptide resin must be soaked in TFA for some time to re-swell the resin thoroughly and washed with ACN or DMSO before hydrothiolysis.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The cleaved peptide thioacids were analyzed on an analytical C18 RP-HPLC column and were purified by semi-prep C18 RP-HPLC. After HPLC purification, the purity was over 95% for all of the demonstrated peptide thioacids based on analytical HPLC. The molecular weights of the peptide thioacids were determined by ESI-MS.

2.2.4 Preparation of histone protein H3(1-109)-COSH

2.2.4.1 Construction of the plasmid pTWIN1-H3(1-109) overexpression system

To fuse the target gene to the N-terminus of the intein-CBD reading frame, the wild type *Xenopus laevis* histone H3(1-109) gene was amplified from plasmid pET-3d-H3 by polymerase chain reaction (PCR), using a forward primer (5'-GGT GGT CAT ATG GCC CGT ACC AAG CAG ACC -3') containing a restriction site for *NdeI* and a reverse primer (5'-GGT GGT TGC TCT TCC GCA CAG GTT GGT GTC CTC AAA GAG -3') containing the *SapI* restriction site. The condition used for amplification of the target gene was: 95 ℃ for 3 min, 30 cycles of 95 ℃ 30 s, 66 ℃ 15 s, 72 ℃ 40 s, and 72 ℃ 10 min for final extension. The PCR product was purified using a PCR purification kit (Qiagene), digested with *NdeI* and *SapI* (New England Biolabs, USA) and ligated into an *NdeI-SapI* digested pTWIN1 expression vector (New England Biolabs, USA). After DNA-sequencing verification of the ligated product, the pTWIN1-H3(1-109) was transformed into *E. coli* strain BL21(DE3) (Stratagene, USA).

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.2.4.2 Expression and purification of the H3(1-109)-intein-CBD fusion protein

The recombinant *E. coli* strain was grown in LB medium supplemented with 100 mg/L ampicillin at 37 ℃ until optical density (OD) at 600 nm reached 0.5-0.7. Induction was performed by the addition of 0.3 mM IPTG. After incubation at 37 ℃ for 2 h, the cells were harvested by centrifugation at 6000 rpm for 10 min (Beckman, JA-10 rotor). The cell pellet was suspensed in the lysis buffer (20 mM phosphate buffer, pH 7.0, 0.5 M NaCl, 1 mM EDTA) and lysed by a microfluider (Microfluidics, Newton, USA) at a chamber pressure of 12K/50 psi. The lysate was then centrifuged at 20,000 x *g*, 4 ℃ for 30 min. (HERMLE, Z36HK, German). Because the expressed protein existed in the form of inclusion bodies, the pellet was resuspended in the wash buffer (20mM phosphate, pH 7.0, 0.5 M NaCl, 1 mM EDTA, 0.5% Triton X-100) and centrifuged again at 20,000 x *g* and 4 ℃ for 30 min. The wash step was repeated twice to remove the cell debris and other impurities. The washed fusion protein pellet was dissolved in 6 M Gdn-HCl in the lysis buffer and centrifuged at 20,000 x *g* and 4 ℃ for 30 min to remove any insoluble materials. The protein solution was then dialyzed gradually against 8 M urea, 4 M urea, and 2 M urea in the lysis buffer containing 0.1 mM TCEP. Each dialysis step took at least 5 h at 4 ℃. The dialyzed protein solution was centrifuged at 20,000 x *g* and 4 ℃ for 30 min to remove any remaining impurities.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.2.4.3 Cleavage of H3(1-109)-COSH from intein-CBD in solution

Typically, the inclusion body of a 1L culture was dissolved in 80 mL of 6 M Gdn-HCl/lysis buffer. After dialysis, the 2 M urea/lysis buffer was added to the volume of 120 mL. The final cleavage volume was 160 mL after adding 40 mL of concentrated cleavage buffer: 0.4 M Na_2S, 1 M HEPES, 1 mM EDTA. The final condition of hydrothiolysis to obtain the C-terminally truncated H3(1-109) thioacid was 0.1 M Na_2S, 0.25 M HEPES, 1 mM EDTA, 0.37 M NaCl, 1.5 M urea, pH 8.0. The cleavage of H3(1-109) from intein-CBD was performed overnight at 4 ℃. After cleavage, the H3(1-109)-CO-SH was formed as precipitates, which were collected by centrifugation at 20,000 x *g* and 4 ℃ for 30 min. The H3(1-109) thioacid pellet was dissolved in 6 M Gdn-HCl/lysis buffer containing 5 mM TCEP. After centrifugation, the supernatant was purified by C18 prep RP-HPLC (Waters 2487 HPLC system, Prosphere C18, 250 x 22 mm) using a gradient: 0% to 50% in 25 min, then to 80% in 30 min of buffer B (90% ACN/0.05% TFA) in buffer A (0.05% TFA/H_2O) at a flow rate of 10 mL/min. The purified H3(1-109)-CO-SH was lyophilized and the molecular weight was confirmed by MALDI-MS.

2.2.5 Preparation of the ubiquitin thioacid

2.2.5.1 Construction of the ubiquitin expression plasmid pTYB1-Ubi

The plasmid containing the human ubiquitin gene was obtained from Dr. Cheung Ching For, Peter's lab. The ubiquitin gene was amplified by PCR using the primers:

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Ubi_F: 5'-GGTGGTCATATGCAGATCTTTGTGAAG-3'

and Ubi_R:5'-GGTGGTTGCTCTTCCGCAGCCACCTCGCAGGCG-3'. The PCR condition was 95 ℃ for 3 min, 30 cycles of 95 ℃ 30 s, 54 ℃ 30 s, 72 ℃ 30 s, and 72 ℃ 10 min for final extension. The PCR product was purified and ligated into the T-easy vector (Promega). The T-Ubi vector was transformed into *E. coli* cells and amplified. The extracted and purified vector T-Ubi was digested with the *Nde*I and *Sap*I restriction enzymes. The digested segment was purified using the PCR purification kit (Qiagen) and ligated into the identically digested pTYB1 vector (New England Biolabs). The correct insert was comfirmed by DNA sequencing.

2.2.5.2 Overexpression and purification of the ubiquitin thioacid

The plasmid pTYB1-Ubi was transformed into *E. coli* BL21(DE3) $CaCl_2$ competent cells. The cells were grown in LB medium containing 100 μg/ml ampicillin at 37 ℃ with shaking at 250 rpm to an OD_{600} of 0.6-0.8. The target protein was induced by 50 μM IPTG at 15 ℃ for 18 h. After centrifugation at 6000 rpm for 10 min, cell pellets from 1 liter culture were suspended in 50 ml lysis buffer (20 mM HEPES, 0.5 M NaCl, 1 mM EDTA, pH 7.0). Cells were lysed by a microfluider (Microfluidics, Newton, USA) at a chamber pressure of 12K/50 psi. Debris was removed by centrifugation at 20,000 x *g* for 30min. The supernatant was mixed with 3 ml chitin beads (New England Biolabs) pre-equilibrated by the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

lysis buffer at 37 ℃ for 2 h. The beads were poured into a column and washed with 40 ml of the lysis buffer. The fusion protein was cleaved by adding 2.5 ml of the cleavage buffer (100 mM Na_2S, 0.2 M HEPES, 0.5 M NaCl, 1 mM EDTA, pH 8.5) and incubating at 37 ℃ overnight. The Ubiqutin thioacid was eluted by 10 ml of the lysis buffer. The cleaved product was purified by C8 semi-prep RP-HPLC.

2.3 Results and discussion

2.3.1 Solid phase synthesis of peptide thioacids by hydrothiolysis of resin-bound peptide thioesters

The difficulty in obtaining thioacid peptides has been a limiting factor in applying thioacid chemistry in protein synthesis. Traditional methods for peptide thioacid synthesis are all based on Blake's benzhydryl thioester linker chemistry [34] which must use Boc-SPPS and therefore hazardous hydrogen fluoride for final deprotection and cleavage. The linker is commercially unavailable. And the yield is usually low because the supernucleophilicy of the thioacid functionality makes it highly susceptible to electrophilic attacks by the cationic electrophiles generated from the various protecting groups during the final cleavage stage. The enzymatic method may have its limitations owing to the substrate specificity restrictions of subtiligase [78].

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Table 2.1. Synthesis of peptide thioacids through hydrothiolysis of CM resin-bound peptide thioesters

No	Sequence	Cleavage Yield (%)[a]	MW[b] Calcd	*m/z*[b] found
1	ARTKQTARKSTG[c]	68 (2 h)	1319.7	1320.8
2	GIGDPVTCLKSGAI[d]	85 (8 h)	1345.7	1347.4
3	VGLFEDTNL[d, f]	70 (3 h)	1022.5	1023.4
4	RLLLPGELA[d]	67 (5 h)	996.6	997.8
5	APKRYKANY[e]	80 (3 h)	1125.6	1127.0
6	DSARAGS[c]	80 (3 h)	678.3	679.6

[a] Cleavage yield was calculated using the quantitative ninhydrin test (absorbance at 570 nm) on the peptide-resin before and after cleavage; values in brackets are the cleavage time.

[b] Calculated isotopic molecular weight and found ($[M+H]^+$) *m/z* value of the peptide thioacid products.

Cleavage conditions: [c] 0.2 M $(NH_4)_2S$, 0.3 M HEPES buffer, pH 8.6, R.T.; [d] 0.2 M $(NH_4)_2S$, 0.3 M HEPES buffer, 6 M guanidine-HCl, pH 8.6, R.T.; [e] 0.1 M Na_2S, 0.3 M HEPES buffer, pH 7.6.

[f] Synthesized on recycled $HSCH_2CH_2CO\text{-}NHCH_2$-CM resin.

Hydrothiolysis of peptide thioesters provides an efficient way to produce peptide thioacids [64]. We reasoned that, provided with a solid support that has good swelling properties in aqueous media, it should also be possible to conduct solid-phase hydrothiolysis on resin-bound peptide thioesters. Since the hydrothiolysis reaction needs to be conducted in aqueous media, the solid support to be used must have good water-swellability. Our first attempt to use a poly(ethylene glycol)-grafted polystyrene (PEG-PS) resin, the TentaGel resin, gave very poor results, with the thiolytic cleavage yield being about 30% at best [64].

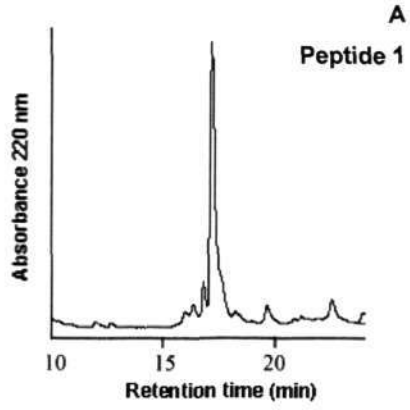
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

This unsatisfactory result is apparently due to insufficient water-swelling of the resin, of which the polystyrene core matrix of which is highly hydrophobic. We therefore turned our attention to a recently developed, totally PEG-based resin, the ChemMatrix resin (CM resin). CM resin has a high degree of cross-linking and exhibits good loading capacity and mechanical stability. Most importantly, it is well-solvated in a broad range of nonpolar and polar solvents, including water. CM resin has a swelling volume in water of ~ 11 mL/g [79]. This makes it an excellent candidate resin for solid phase-supported reactions in aqueous media. We therefore used the CM resin for the synthesis of peptide thioacids through hydrothiolysis of resin-bound peptide thioesters.

The conditions for the solid-phase hydrothiolysis reaction on the resin bound peptide thioesters were similar to those used in solution reactions [64]. Either ammonium sulphide or sodium sulphide was used as the source of hydrosulfide ions. For a typical reaction, 25-30 mg peptide-resin was suspended in 1 mL hydrothiolysis buffer (see Table 2.1) in an Eppendorf vial with gentle shaking. The reaction was stopped by acidification with 20% TFA/H_2O and with ice cooling. The mixture was subjected to HPLC purification to give the purified thioacid peptide, and the molecular weight of each peptide thioacid was determined by ESI-MS. Good hydrothiolytic cleavage yields were achieved for all the CM resin-bound peptide thioesters, as shown in Table 2.1.

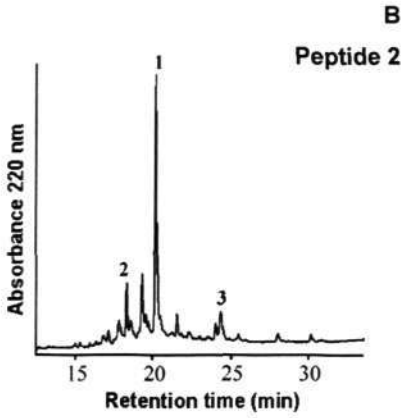
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

A



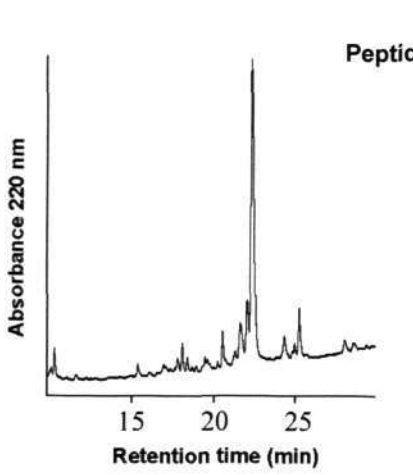
HPLC condition: 0% to 12% of buffer B in buffer A in 24 min

B



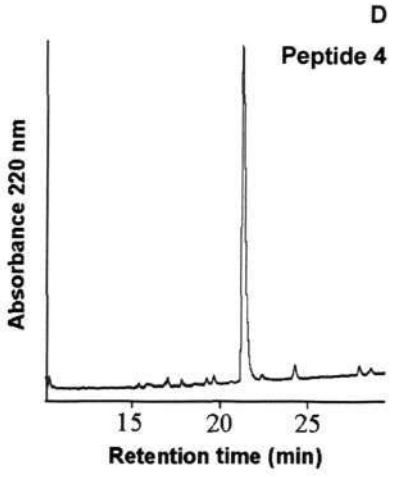
HPLC condition: 0% to 60% of buffer B in buffer A in 30 min

C



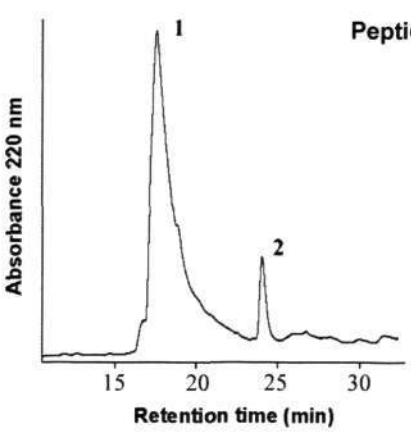
HPLC condition: 0% to 60% of buffer B in buffer A in 30 min

D



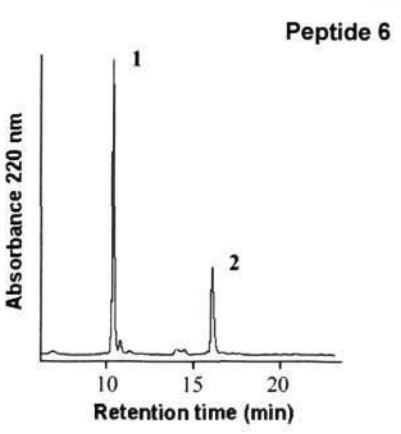
HPLC condition: 0% to 60% of buffer B in buffer A in 30 min

E



HPLC condition: 0% to 30% of buffer B in buffer A in 30 min

F



HPLC condition: 0% to 12% of buffer B in buffer A in 24 min

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Figure. 2.2 Analytical HPLC analysis of crude peptide thioacids cleaved from CM resin. (**A**) Thioacid peptide **1**, H_2N-**ARTKQTARKSTG**-COSH, after 2 h cleavage. (**B**)Thioacid peptide **2**, H_2N-**GIGDPVTCLKSGAI**-COSH, after 5 h cleavage. Peak 1, the cleaved peptide thioacid with the correct MW; peak 2, the cleaved truncated peptide thioacid H_2N-**PVTCLKSGAI**-COSH. *m/z* $[M+H]^+$ found: 1005.4, MW calcd: 1003.5; peak 3, the cleaved product, H_2N-**GIGDPVTCLKSGAI**-CO-SSSSH (*m/z* $[M+H]^+$ found: 1443.2, MW calcd: 1441.7), with a molecular weight 96 Da higher than the cleaved peptide thioacid H_2N-**GIGDPVTCLKSGAI**-COSH. (**C**) Thioacid peptide **3**, H_2N-**VGLFEDTNL**-COSH, after 5 h cleavage. This thioacid peptide was synthesized on recycled CM resin. (**D**) Thioacid peptide **4**, H_2N-**RLLLPGELA**-COSH, after 3 h cleavage. (**E**) Thioacid peptide **5**, H_2N-**APKRYKANY**-COSH, after 4 h cleavage. Peak 1, the cleaved product with correct MW; peak 2, cleaved peptide thioacid without **APK**. *m/z* $[M+H]^+$ found: 832.0, MW calcd: 829.4. (**F**) Thioacid peptide **6**, H_2N-**DSARAGS**-COSH, after 2 h cleavage. Peak 1, cleavage product with correct MW; peak 2, the cleaved product, H_2N-**DSARAGS**-CO-SSSSH (*m/z* $[M+H]^+$ found: 775.4 MW calcd: 774.3), with a molecular weight 96 Da higher than the cleaved peptide thioacid H_2N-**DSARAGS**-COSH.

As seen from Table 2.1, these peptides contain a diverse set of C-terminal residues, from the smallest Gly to the hindered Ile. For instance, thioacid peptide **2**, which has an Ile as the C-terminal residue, was obtained in excellent yield after a prolonged cleavage reaction time. For most other peptides, the reaction reached its maximum yield after about 3 h.

The cleaved crude products were in general of good quality, as shown by analytical HPLC analysis (Figure 2.2) and characterization by ESI-MS. Basically, the purity of the crude peptide depends solely on the efficiency of the SPPS. (see Figure. 2.2,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

peak 2 in (**B**) and (**E**)). Peptide thioacid **5** gave a broad peak on HPLC analysis (see Figure 2.2, peak 1 in (**E**)). This is probably because of the high basic nature of this peptide thioacid (four basic groups for a nine-residue peptide). It is known that when separating basic compounds on reversed-phse columns peak tailing may occur owing to their interactions with the acidic free silanol groups present in various amounts in all silica-based reaversed-phase column [80]. It should be noted that peptide **3** was synthesized on recycled mercaptopropionylaminomethyl CM resin. This is because the thiol-derived CM resin can be regenerated after thiolytic cleavage and therefore can be reused. Before reuse, the recycled resin was treated with 5% hydrazine, 1% mercapthanol in DMF for 30 min to ensure complete removal of any uncleaved peptide.

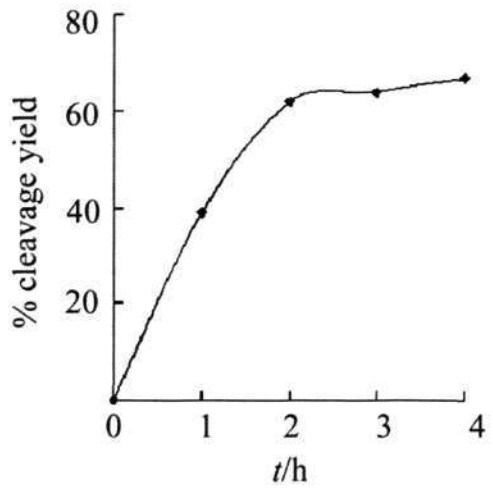


Figure 2.3 Cleavage yield of thioacid peptide **4** as a function of time.

We noticed a decrease in the cleavage yield if the peptide resin was stored for some time. This is likely due to interchain interactions and aggregation of the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

resin-bound peptide. Therefore, it is recommended that, once the peptide resins are fully deprotected, they should be cleaved with the thiol reagent as soon as possible. Otherwise, the deprotected peptide resin stored over time must be soaked in TFA for 15 – 20 min to re-swell the resin thoroughly and break up any inter-winding peptide chains, and then has to be washed with acetonitrile and DMSO before thiolytic cleavage.

Figure 2.3 shows a simple kinetic study of the cleavage of peptide **4**. The cleavage yield reached a plateau after 3 h reaction.

During cleavage, no *S*-alkylation side products were observed from HPLC and MS analysis, presumably because all the reactive species that would be derived from the side-chain protecting groups had already been removed and washed away in the previous step. We also noticed that, in some cases, there was a cleavage byproduct whose molecular weight was 96 Da higher than that of the cleaved peptide thioacid. This is likely an acyl tetra sulfide of the structure peptide-CO-SSSSH due to oxidative formation of HSSSSH. This by-product disappeared when a reducing environment was applied.

It is worth noting that the solid phase hydrothiolysis reaction can be conducted under denaturing conditions and this is highly desirable for hydrophobic peptides

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

that are prone to aggregation. In this study, thioacid peptides **2**, **3** and **4** were prepared in the presence of 6 M Gdn-HCl.

Tan et al. [64] investigated the influence of pH on the thioacid formation in solution by hydrothiolysis. It was found that the higher the pH, the faster the conversion of the thioester to thioacid occurs in solution hydrothiolysis due to the pK_{a1} of H_2S of 6.9 [81]. We first conducted the hydrothiolysis of the peptide-bound resin at pH 9. The CM resin bound deprotected peptides appeared prone to adhere to each other and for this reason, the cleavage could not proceed well at this basic pH. Therefore, a lower pH of 8.5 or 7.6 was tested. At these two pHs, the effective concentrations of HS^- were high enough to allow the hydrothiolysis reaction to occur. Moreover, hydrolysis of peptide thioesters or thioacids was hardly observed during cleavage except in the case of peptide thioacid **1** which, with a Gly at its C-terminus, has a higher propensity to hydrolyze under basic conditions.

2.3.2 Intein-mediated protein thioacid preparation

2.3.2.1 C-terminal truncated H3(1-109) thioacid preparation

Protein splicing is a self-catalyzed process without either added enzymes or cofactors, in which the protein splicing element, termed an intein, catalyzes the splicing. Since the first report of the protein splicing phenomenon, over 100 inteins have been found in eubacteria, archaebacteria, and unicellular eukaryotic

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

organisms [82]. Since extein sequences have little effect on the intein function, a foreign protein replacing the native extein can then be used to fuse to the intein for protein engineering applications.

The pTWIN1 vector available from NEB uses a modified *Ssp* DnaB (154 amino acids) as intein 1 and a modified *Mxe* GyrA (198 amino acids) as intein 2. The target gene coding histone H3(1-109) digested by *Nde*I and *Sap*I was inserted into the pTWIN1 vector which was digested with the same restriction enzymes, resulting in the fusion of the target gene to the N-terminus of the *Mxe* Gyr intein. The engineered *E. coli* strain BL21(DE3)/pTWIN-H3(1-109) overexpressed the target protein H3(1-109)-intein-CBD as shown in Fig 2.4.

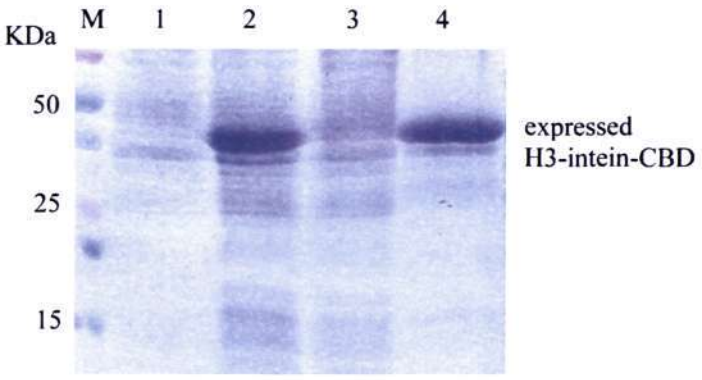


Figure. 2.4 15% SDS-PAGE analysis of the expression of the target protein from engineered *E. coli* BL21(DE3)/pTWIN1-H3(1-109). Lane 1, uninduced cell culture; lane 2, induced cell culture at 37 ºC by 0.3 mM IPTG; lane 3, supernatant of induced cell lysate after centrifugation; lane 4, pellet of induced cell lysate after centrifugation.

The fusion protein was expressed in the form of an inclusion body. We tried to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

express the fusion protein at different induction temperatures (37 ℃ for 2 h, 30 ℃ for 3 h, 15 ℃ overnight) and different concentrations of IPTG to improve the fusion protein's solubility. However, there was little variation in the expression level (data not shown). The condition we then chose to ensure a high expression yield and sacrifice of the solubility of the expressed protein was to do culture at 37 ℃ for 2 h and induction by 0.3 mM IPTG. After several washings with the wash buffer containing 0.5% Triton-100 (20 mM phosphate, pH 7.0, 0.5 M NaCl, 1 mM EDTA, 0.5% Triton X-100), the inclusion bodies were completely dissolved in 6 M Gdn-HCl in the lysis buffer (20 mM phosphate buffer, pH 7.0, 0.5 M NaCl, 1 mM EDTA). Because the intein splicing depends on the proper folding of the intein, the refolding after denaturation by Gdn-HCl is important. We adopted the protocol from NEB to successfully refold the H3(1-109)-intein-CBD fusion protein by using the following procedures: first, fully denature the fusion protein in 6 M Gdn-HCl in the lysis buffer; second, dialyze the solution stepwise in 8 M urea, 4 M urea, and 2 M urea in the lysis buffer; finally, carry out the cleavage with HS^- at a urea concentration below 2 M. We performed the cleavage reaction in solution by mixing the dialyzed solution with the Na_2S solution. Since the cleaved protein thioacid is highly insoluble, the insoluble target protein thioacid was easily separated by centrifugation from intein-CBD which is soluble in solution (Figure 2.5). Further RP-HPLC purification provided pure H3(1-109)-COSH, the identity of which was confirmed by MALDI-MS analysis, as shown in Figure 2.6. The yield of histone H3(1-109)-COSH is around 10 mg/L. Although this truncated

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

histone H3 has a size of only around 12 KDa, it is very hydrophobic. We were unable to purify it on a semi-preparative C18-RP column. After complete dissolution in 6 M Gdn-HCl, this H3 protein was purified on a C4 semi-prep column.

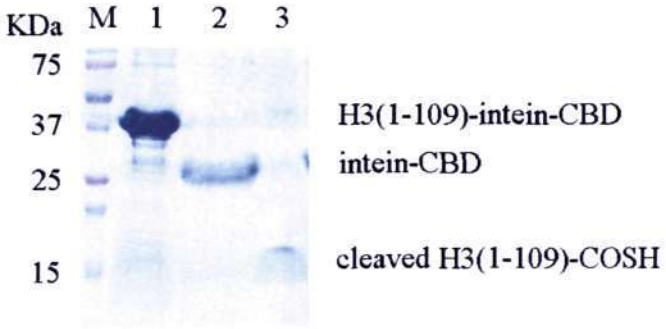


Figure 2.5 Hydrothiolysis cleavage of H3(1-109)-intein-CBD in solution. Lane 1, the refolded fusion protein, H3(1-109)-intein-CBD (around 41 KDa), after dialysis; lane 2, supernatant of the cleavage solution; lane 3, precipitate of the cleaved solution.

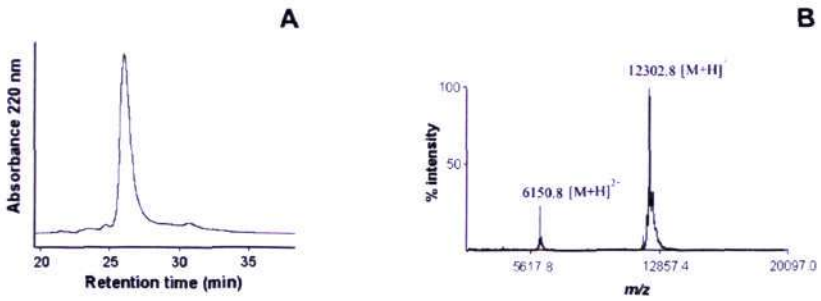


Figure. 2.6 Characterization of H3(1-109)-COSH. (**A**) HPLC profile of H3(1-109)-COSH. The protein was analysized by C4 analytic RP-HPLC using a gradient of buffer B in buffer A from 0% to 40% in 20 min, then to 50% in 10 min. (**B**) MALDI-TOF mass spectral analysis of H3(1-109)-COSH. *m/z* $[M+H]^+$ found: 12302.8, MW calcd: 12301.3.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2.3.2.2 Ubiquitin thioacid preparation

We also successfully prepared another small protein ubiquitin thioacid by using hydrothiolysis of the ubiquitin thioester formed by the intein-fusion protein system in solution. The human ubiquitin gene encoding the 76 amino acid open reading frame prepared by digestion with *Nde*I and *Sap*I was inserted into the pTYB1 plasmid. The 3' end of the target gene introduced by *Sap*I was placed immediately adjacent to the intein N-terminus, resulting in a fused ubiquitin protein which has the thiol-inducible cleavage activity. The fusion protein was expressed in BL21(DE3) cells transformed with pTYB1-ubi. The IPTG induction was performed when the culture reached an OD_{600nm} of 0.6-0.8. Since the fusion protein is prone to be produced as an inclusion body at high temperature and high inducer concentration, the cells were induced with 50 μM IPTG and grown at 15 ℃ after induction in order to obtain the fusion protein in a more soluble form (Figure 2.7).

The expressed ubiquitin-intein-CBD fusion protein was released from *E. coli* cells by microfluiderization. After the chitin beads were pre-equilibrated with the lysis buffer, the cell lysate containing the fusion protein was loaded. The fusion protein was cleaved overnight on the chitin column in the cleavage buffer (100 mM Na_2S, 0.2 M HEPES, 0.5 M NaCl, 1 mM EDTA, pH 8.5). All processes were carried out at 37 ℃. The eluate containing only the ubiquitin thioacid was collected and purified by HPLC. The molecular weight of the ubiquitin thioacid was determined by MALDI-MS (Figure 2.8). After HPLC purification, around 5 mg ubiquitin

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

thioacid were obtained from a 1 L culture.

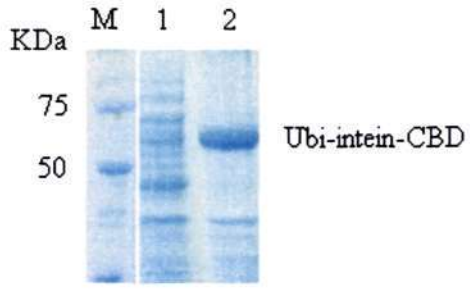


Figure. 2.7 12% SDS-PAGE gel analysis of expressed ubiquitin-intein-CBD in BL21(DE3) cells transformed with pTYB1-ubi. Lane 1, uninduced cell culture. Lane 2, IPTG induced cell culture.

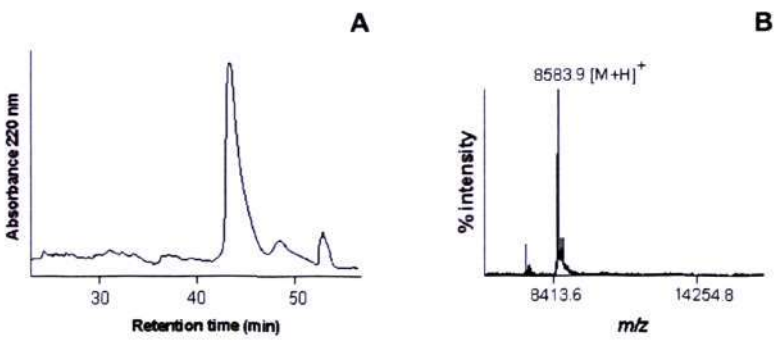


Figure. 2.8 Characterization of the ubiquitin thioacid. (**A**) HPLC profile of the C8 semi-prep purification of the ubiquitin thioacid eluate from chitin beads after hydrothiolysis. The cleaved product ubiquitin thioacid was eluted with a linear gradient of buffer B in buffer A from 0% to 60% in 60 min at a flow rate of 2.5 min/mL. (**B**) The purified ubiquitin thioacid as characterized by MALDI-MS. *m/z* $[M+H]^+$ found: 8583.9, MW calcd: 8580.8.

The folding of the ubiquitin-intein-CBD fusion protein is crucial for the splicing of

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

this fusion protein. We first tried to overexpress the fusion protein at 37 ℃ in order to obtain it in a large amount. The fusion protein was in the inclusion body. Althought we used the serial dialysis protocol to decrease urea concentrations, the splicing could not proceed, either at 4 ℃ or 37 ℃. For this reason, a temperature of 15 ℃ and a low IPTG concentration were used for protein expression. Under this condition, the fusion protein was mostly expressed in the soluble form. In this case, splicing of the fusion protein occurred readily to give the hydrothiolysis product ubiquitin thioacid after overnight incubation at 37 ℃.

2.3 Conclusion

In this chapter, we described a novel method to synthesize peptide thioacids, as shown in scheme 2.7. A total PEG-based resin was used to synthesize resin-bound peptide thioesters. The high aqueous swellability of this resin made it possible to conduct solid-phase hydrothiolysis and to convert resin-bound peptide thioesters to thioacids in aqueous buffer. Several peptide thioacids containing different C-terminal residues, ranging from the smallest Gly to the hindered Ile, were synthesized through solid-phase hydrothiolysis, indicating that there may be no C-terminal residue restriction for the synthesis of peptide thioacids by using this method. The experimental results show that high yields were obtained for all peptide thioacids synthesized by this solid-phase hydrothiolysis method and, moreover, no side reactions occurred during hydrothiolysis.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

The hydrothiolysis reaction was also applied to the synthesis of thioacid proteins: histone H3 and ubiquitin thioacids. In this case, the protein thioesters were generated through intein-mediated protein splicing and were subjected to hydrothiolysis. After refolding of the H3 fusion protein by stepwise dialysis, hydrothiolysis of the fusion protein was conducted in aqueous buffer, resulting in the desired H3 thioacid. In the case of the ubiquitin thioacid, hydrothiolysis for its release from the bead-bound ubiquitin-intein thioester was carried out. The ability of hydrothiolyzing protein thioesters prepared from the intein system either in solution or on beads makes this method a convenient and practical one for the preparation of protein thioacids.

In all, the above two methods for peptide thioacid and protein thioacid synthesis make these once difficult-to-get large thioacid compounds easily available and, therefore, stimulate the development of new thioacid-based synthetic methodologies for protein chemistry.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 3: Histone Protein Synthesis by Thioacid Capture Ligation

3.1 Introduction

As reviewed in Chapter 1, chemically based protein synthesis methods are generally more versatile as they do not have any of the limitations of the ribosome-dependent system and would therefore represent a useful alternative for protein production. In recent years, convergent strategies for protein synthesis have been developed that rely on chemoselective ligation of unprotected synthetic peptide fragments in aqueous solution. A common feature of these ligation methods is that peptide bond formation is effected by a proximity-driven intramolecular reaction which occurs spontaneously without the help of a coupling reagent. Crucial to the ligation reaction is a prior capture step to bring together the reacting C-terminus and N-terminus of the two peptides. Kemp's prior thiol capture ligation was the first such method put forth into the practice of peptide synthesis [41]. Subsequently, thiazolidine-capture ligation [37], thioester-mediated native chemical ligation [45], thioacid capture ligation [57] were developed in the mid 1990s, followed by the development in the new millennium of Staudinger ligation for peptide bond formation [59] and lately decarboxylative condensation for amide ligation [63]. So far, the method that has proved most practically useful is the thioester-mediated native chemical ligation, which has been used to prepare a large number of small to medium sized proteins over the past decade. The other

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

chemoselective ligation method that is potentially very useful but has so far received much less attention is the thioacid capture ligation (or mini-thiol capture ligation) method (Scheme 1.9). The key element of this method consists of specific capture of a C-ter thioacid of the first peptide by an activated disulfide from an Npys-modified N-ter Cys side chain of the second peptide to form an acyl disulfide intermediate which undergoes rapid intramolecular acylation to generate an amide bond. The final product with a native Cys residue at the ligation site is obtained after a simple thiolytic reduction reaction [57]. Since we have developed new methods for producing peptide and protein thioacids in solution and on solid phase, obtaining the thioacid building blocks is not the bottleneck of using thioacid capture ligation any more. These methods will therefore make thioacid capture ligation more practically useful for the synthesis of proteins.

Protein chemical synthesis is essential in preparing proteins that are not accessible by tranditional recombinant technologies, such as those that contain posttranslational modifications. A good example to illustrate protein post-translational modification and the need for protein synthesis methods is the histone proteins. Histone proteins are used in eukaryotic cells to package DNA into highly ordered chromatin structures. The basic units of chromatin are the nucleosomes formed by about 147 bp DNA wrapping around a histone octamer core which contains two copies of each of the four histones: H2A, H2B, H3, and H4. Strings of nucleosomes fold further into more compact chromatin structures.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

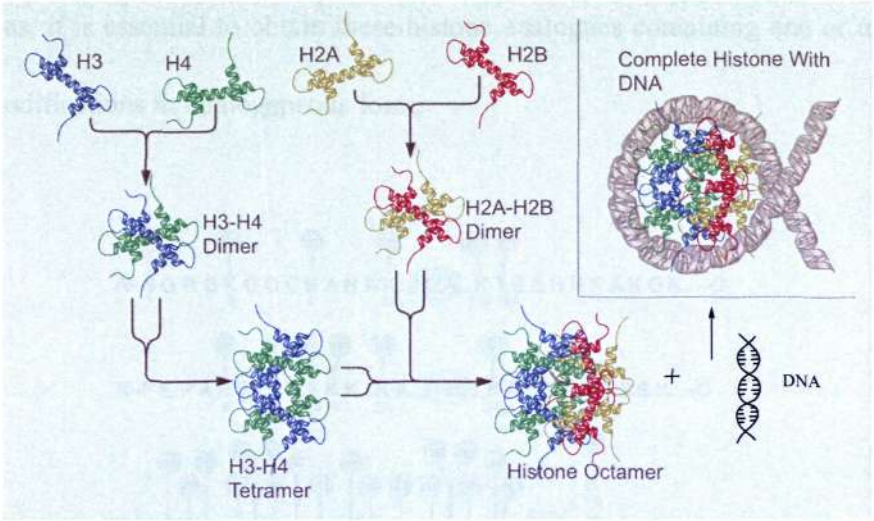


Figure 3.1 Outline of the histone octamer formation and the assambly of the core histones into the nucleosome (adopted from http://en.wikipedia.org/wiki/Histones).

Chromatin activity is closely linked to the state of histone modification [83]. These post-translational histone modifications [84] include acetylation, methylation, phosphorylation, ubiquitynation, sumoylation, and ribosylation on specific amino acid side chains. Most of these modifications are concentrated on the flexible N-terminal tails of histones, although some are also found on the C-terminal and central parts of the proteins (as shown in Figure 3.2). As suggested in the histone code hypothesis [85], distinct histone modification patterns not only directly influence chromatin structure in certain ways but also provide specific binding motives for chromatin associated proteins which, in turn, dictate dynamic transitions between different chromatin states, controlling therefore all DNA-related processes such as replication, transcription, recombination, repair or

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

silencing. To study the relationship between histone modifications and chromatin functions, it is essential to obtain these histone analogues containing one or more such modifications in homogeneous form.

Figure 3.2 Posttranslational modification of human nucleosomal histones (adopted from [86]). Ac: acetylation, me: methylation, ph: phosphorylation, ub1: ubiquitination.

However, histones obtained from natural sources are highly heterogenous due to differences in the degree, type, and positions of the posttranslational modification. Several methods have been developed for the preparation of site-specifically modified histones. One of them is based on thioester-mediated chemical ligation. In this method, an N-terminal peptide thioester of a histone bearing specifically modified amino acid residues was synthesized by standard solid phase peptide synthesis, while a C-terminal protein fragment containing an N-terminal cysteine was produced by the recombinant DNA technology [87,88]. A full length histone with modified amino acid residues in the N-terminal tail was then generated by

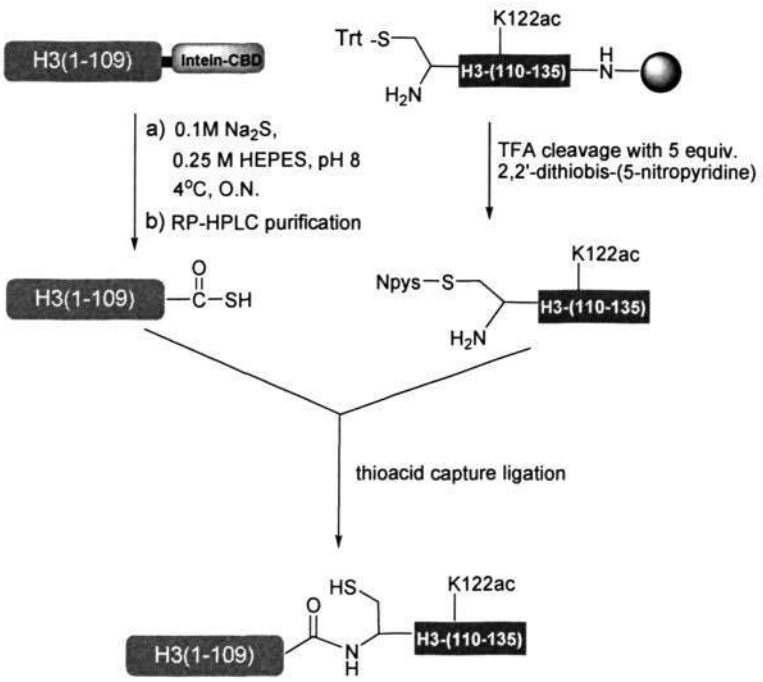
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

linking the two parts via native chemical ligation. Since histone proteins do not have Cys residues in their sequence, He, *et al.* chose Ala as the ligation site which was generated by hydrogenolytic desulfurization after ligation to change the Cys to Ala [89]. However, the yield of this reaction is usually quiet low due to the absorption of the protein to the Raney nickel catalyst. Using this method, they prepared several histone anologs, such as a diacetylated H4, a trimethylated H3, and a pentaacetylated H3/H4 chimera. Recently, another method to chemically methylate recombinant histone proteins was developed by Shokat and coworkers [90]. In this method, a traditional aminoethylation reaction capable of converting a cysteine residue to a lysine analog was used to introduce the methylated lysine analog into histone H3. Site-specific alkylation of a cysteine on H3 by (2-chloroethyl)-methylammonium chloride, (2-chloroethyl)-dimethylammonium chloride, (2-bromoethyl)-trimethylammonium bromomide installed mono-, di- or trimethylated isosteric Lys into the histone. Structural and functional studies [91,92] had shown that such alkylated cysteines were similar to their native methyl lysine counterparts with regard to the binding of effector proteins and the rates of nucleosome remodeling. This method is simple to use and can generate histones with site and degree-specific methylation.

In this study, we focused on the synthesis of histone H3 analogs through thioacid capture ligation. This ligation chemistry was used for the synthesis of histone H3 analogs with either N-terminal or C-terminal modifications. On the one hand,

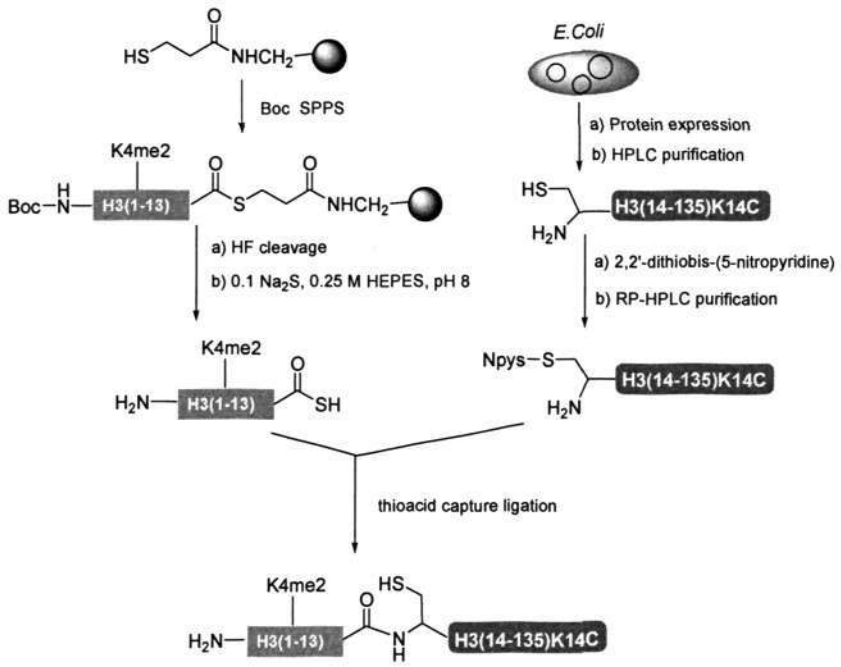
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

H3(1-109)-COSH, which was generated by cleavage of the H3(1-109)-intein fusion protein with hydrosulfide ions (see Chapter 2), was ligated with the chemically synthesized C-terminal peptide Npys-H3(110-135)K122ac, giving the full length H3 with the C-terminal modification (Scheme 3.1). On the other hand, chemically synthesized H3(1-13)K4me2 peptide thioacid was ligated through thioacid capture ligation with biosynthetic Npys-H3(14-135) to give the full length H3 containing the modified N-terminal segment (Scheme 3.2). These analogs were used to test their abilities to form histone octamers and nucleosome core particles. Our work demonstrates that the thioacid capture ligation method is simple and highly efficient, and in principle, can be used to provide any modified histones for chromatin structural and functional studies.



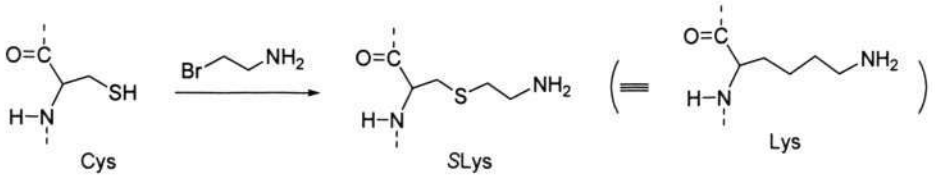
Scheme 3.1 Outline of using thioacid capture ligation to prepare H3K122ac - an H3 analog containing an acetylated C-terminal Lys122.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library



Scheme 3.2 Outline of the synthesis of H3K4me2/K14C - an H3 analog containing a dimethylated N-terminal Lys4 - through thioacid capture ligation.

The thioester-mediated native chemical ligation was also used to synthesize histone H3 analogs in this study. The ligation site was also at residue 14 which is a Lys residue in the H3 sequence. After ligation, the ligated product was then treated with bromoethylamine hydrobromide to install the *S*Lys14 as a Lys isostere (as shown in Scheme 3.3). This aminoethylation reaction is highly specific for the thiol group and has been used to modify cysteine residues in natural proteins [90,92,93]. Further experimental data showed that these alkylation products behaved normally in histone octamer formation and nucleosome core particle assembly.



Scheme 3.3 Conversion of Cys to *S*Lys for ligation at lysine sites.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.2 Materials and methods

3.2.1 Materials

All Fmoc and Boc-amino acid derivatives, coupling reagents, Wang resin, MBHA resin were purchased from GL Biochem (Shanghai, China) and Novabiochem (Germany). All chemical reagents were purchased from commercial suppliers. The size-exclusion 26/60 Sephacryl S-200 column was from GE Healthcare, USA.

3.2.2 Instrumentation

3.2.2.1 HPLC analysis and purification

HPLC system, columns, buffers and conditions were the same as described in Chapter 2.

3.2.2.2 Mass Spectrometry

ESI-MS and MALDI-MS were the same as described in Chapter 2.

3.2.3 Synthesis of Npys-H3(110-135) peptides

Preloaded Fmoc-Ala-Wang resin was used to synthesize the peptides corresponding to the C-terminal sequence of residues 110-135 of histone H3 following standard Fmoc-N^{α} protection manual solid-phase peptide synthesis protocols. The following side-chain protecting groups were used: tBu for Thr; Boc

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

for Lys; Pbf for Arg; Trt for Gln, His, Cys; OtBu for Asp, Glu. 4 eq. of Fmoc amino acids were coupled to the resin using 4 eq. of PyBop and 8 eq. of DIEA in DMF. The successive α-amino group deprotection and wash steps were carried out in 20% piperidine/DMF (1 x 5 min, 1 x 20 min) and DMF (3 x 2 min), DCM (3 x 2 min), DMF (3 x 2 min). For peptide H3(110-135)/K122ac and H3(110-135)/K115ac, Fmoc-Lys(ac)-OH was used to introduce acelytated Lys at position 122 and 115. The ninhydrin test was performed to monitor the coupling and deprotection. After assembling of peptides, the peptide was cleaved with 95% TFA/2.5% TIS/2.5% H_2O and HPLC purified. The HPLC purified peptide was Npys modified in acetic acid:water (3:1) containing 3-5 molar equiv 2,2'-dithio-bis-(5-nitropyridine). The reaction was performed for 4-6 h (method I) [94]. Or, the peptide was cleaved for 3 h with a mixture of 95% TFA/2.5% TIS/2.5% H_2O in the presence of 4 eq. of 2,2'-dithio-bis-(5-nitropyridine) which was to introduce the Npys group to Cys thiol. The cleaved peptide was precipitated by cold ethyl ether (method II) [95]. The Npys modified peptides were lyophilized and purified by RP-HPLC, and molecular weight was determined by ESI-MS.

3.2.4 Model peptide ligation study

H_2N-**RLLLPGELA**-COSH (thioacid peptide **4** in Table 2.1, Chapter 2) was synthesized on CM resin, and cleaved from solid support by hdrothiolysis (see section 2.2.3 in chapter 2). 3 molar eq. of Npys-H3(110-135)K115ac was added to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

0.5 mM of H_2N-**RLLLPGELA**-COSH which was dissolved in ACN/water (0.05% TFA). After a yellow color developed, the pH of the reaction was adjusted to 5-6 by addition of 0.2 M sodium phosphate buffer, pH 7. 10 to 15 min later, TCEP or DTT was added to the solution for thiolytic reduction. After the reduction, the reaction was analyzed by C18 RP-HPLC.

3.2.5 Ligation of the H3(1-109) thioacid with Npys-H3(110-135) peptides

H3(1-109)-COSH (from section 2.2.4 in Chapter 2) was dissolved in 6 M Gdn-HCl, 0.2 M acetic acid, pH 3, to a final concentration of 1 mM. The protein solution was transferred to an eppendof tube containing 3 eq. of the Npys modified H3 C-terminal peptide, Npys-H3(110-135). After mixing, a yellow color developed immediately. The pH was then adjusted to 6-7, and the reaction was allowed to continue for half an hour before 1 M TCEP or DTT was added to stop the reaction. The ligated protein was analyzed and purified by RP-HPLC. The molecular weight of the ligation product was determined by MALDI-MS.

3.2.6 Construction of the plasmid pET-3d-H3(14-135)/K14C overexpression system

The plasmid pET-3d containing the *Xenopus laevis* histone H3 gene was the source of the target gene. The wild type *Xenopus laevis* histone H3 gene was first mutated at Cys110 to Ala by a QuickChange Site-Directed Mutagenesis Kit (Stratagene), using a forward primer 5'-GAG GAC ACC AAC CTG *GCC* GCC ATC CAC GCC

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

AAG -3' and a reverse primer 5'-CTT GGC GTG GAT GGC *GGC* CAG GTT GGT GTC CTC -3'. The condition used was 95 ℃, 30 s, 18 cycles of 95 ℃, 30 s, 55 ℃, 1 min, and 68 ℃, 6 min. The mutated gene was transformed into XL1-Blue $CaCl_2$ competent cells (Stratagene), and amplified.

Amplified plasmid pET-3d-H3/C110A, refered to as the wild type here, was purified using a plasmid purification kit (Qiagen) and the second mutation was performed based on pET-3d-H3/C110A to delete amino acid residues 1-13. The forward and reverse primers for the second mutation were: 5'-CTT TAA GAA GGA GAT ATA CAT ATG *TGC* GCT CCC CGC AAG CAG CTG GCC ACC -3' and 5'-GGT GGC CAG CTG CTT GCG GGG AGC GCA CAT ATG TAT ATC TCC TTC TTA AAG -3', respectively. The mutagenesis condition was the same as for the first mutation. The mutated gene was transformed into $CaCl_2$-competent BL21(DE3)pLysS cells (Stratagene).

3.2.7 Overexpression and purification of recombinant Histone H3(14-135)/K14C

Cells were grown in 2 x TY medium containing 16 g Bacto Tryptone, 10 g yeast extract, 10 g NaCl, 100 mg ampicillin, 25 mg chloramphenicol in 1 L. After the OD_{600nm} of the cells reached 0.6-0.8, IPTG (isopropyl-β-D-thiogalactopyranoside) was added to a final concentration of 0.5 mM, and the culture was incubated for another 2 h. Cultured cells were harvested by centrifugation (Beckman centrifuge,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

JA-10 rotor) at 6000 rpm for 10 min at room temperature. The cells were suspended in the wash buffer (20 mM phosphate, pH 7.0, 0.5 M NaCl, 1 mM EDTA), and stored at -20 ℃.

The cell suspension was thawed in a water bath at 37 ℃. The cell lysate was sheared by a microfluider. The crude cell extract was centrifuged at 20,000 x *g* and 4 ℃ for 30 min (HERMLE, Z36HK, Germany). The pellet was washed by resuspension and centrifugation twice in the wash buffer containing 0.5% (v/v) Triton X-100 to remove any cell debris and impurities. The detergent was removed by one more washing with the wash buffer. The remaining pellet was then dissolved in 6 M Gdn-HCl in the wash buffer for 1 hour at room temperature. After centrifugation to remove all the insoluble material, the supernatant was purified by C18 prep RP-HPLC using a gradient of 0%-50% for 25 min, then to 80% for 30 min of buffer B (90% ACN/0.05% TFA) in buffer A (0.05% TFA/H_2O) at a flow rate of 10 mL/min. The purified protein was lyophilized and the molecular weight was determined by MALDI-MS.

3.2.8 Expression and purification of recombinant histone proteins (wild type)

The plasmids pET-3a containing the *Xenopus laevis* histone H2A, H2B and H4 gene and pET-3d containing H3 gene were obtained from Dr. C.A. Davey's lab. The plasmids were transformed into *E. coli* strain BL21(DE3)/pLysS

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

$CaCl_2$-competent cells. Cells were grown in 2 x TY medium containing 16 g Bacto Tryptone, 10 g yeast extract, 10 g NaCl, 100 mg ampicillin, 25 mg chloramphenicol in 1 L. The induction by 0.5 mM IPTG was at OD_{600nm} = 0.6-0.8. The induction was for 3h at 37 ℃. Cells were harvested by centrifugation at 6000 x *g* for 10 min at room temperature. The cell pellet was suspended in the wash buffer (20 mM phosphate buffer, pH 7.0, 0.5 M NaCl, 1 mM EDTA), and stored in -20 ℃. The histone proteins were extracted and purified in the same way as it was done for the truncated histone H3(14-135)/K14C (see section 3.2.7 in Chapter 3).

3.2.9 The H3(1-13)/K4me2 peptide thioester and K9me2 peptide thioester synthesis

The syntheses of the N-terminal H3 peptide thioesters were carried out following the standard protocols for Boc (*tert*-butyloxycarbonyl)-based solid phase peptide synthesis. The following side-chain protecting groups were used: Bzl for Thr, Ser; 2-Cl-Z for Lys; Tos for Arg; Trt for Gln. First, MBHA resin (0.9 mmol/g) was washed with DCM 2-3 times. 4 eq. of (Trt)-S-CH_2CH_2-COOH and of PyBop in DCM containing 8 eq. of DIEA were then used for coupling to the MBHA resin. After 2 hour reaction at room temperature and washing, the thiol protection group trityl was removed with 5% TFA/2.5% TIS/1% 2-mercaptoethanol in DCM, 3 x 10 min. The deprotected resin was washed (3 x DCM, 3 x DMF, 3 x DCM). The obtained HS-CH_2CH_2-CO-NH_2-MBHA resin was used for peptide assembly.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Loading of the 1st (C-terminal) amino acid to the thiol-functionized resin was done by using 4 eq. of Boc-amino acid preactivated by 4 eq of PyBop and 8 eq of DIEA. After 1 hour coupling of the first amino acid, the resin was washed with DCM (3 x), DMF (3 x) and DCM (3 x). Capping was performed by using 5% acetic anhydride/2.5% DIEA in DCM for 2 x 10 min. The resin was washed with DCM (3 x), DMF (3 x) and DCM (3 x). Afterwards, the successive α-amino group deprotection steps were performed with 30% TFA/DCM 2 x 10 min between each amino acid coupling step. The residue at position 4 of the peptide H3(1-13) /K4me2 (**ART*Kme2*QTARKSGG**-CO-$SCH_2CH_2CONH_2$) and position 9 of the peptide H3(1-13)/K9me2 (**ARTKQTAR*Kme2*SGG**-CO-$SCH_2CH_2CONH_2$) was dimethylated Lys, which was introduced by using Fmoc-Lys(me2)-OH. The deprotection of Fmoc was carried out with a mixture solution containing 25% 1-methylpyrrolidine, 2% hexamethylene imine, 2% HOBt in 1:1 NMP/DMSO, for 1 x 5 min, 1 x 20 min [74]. The ninhydrin test was performed to estimate the completeness of the coupling. The cleavage of the peptide thioesters from the resin was carried out by HF treatment. After washing with diethyl ether, the crude peptide was dissolved in 10% ACN/ H_2O, and lyophilized for further HPLC purification.

3.2.10 Ligation of the H3(1-13) thioester with H3(14-135)/K14C

The ligation buffer was 0.2 M HEPES, pH 8.5, containing 6 M Gdn-HCl, 0.1 M

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

NaCl, 40 mM TCEP, 2% (w/v) MESNA, and 2% (w/v) thiophenol sodium salt. Lyophilized H3(14-135)/K14C was dissolved in this ligation buffer at the final concentration of 0.5 mM. This solution was transferred to an eppendorf tube containing 3 eq. of the thioester peptide. The reaction was mixed gently, and incubated at room temperature for 24-48 hours. Gently mixing was done periodically during the reaction. The ligation reaction was monitored by C4 analytical RP-HPLC, and purification of the ligation product was performed by C8 semi-prep RP-HPLC.

3.2.11 Alkylation of the ligated protein Histone H3K14C

The HPLC purified ligation product was dissolved in the alkylation buffer containing 6 M Gdn-HCl, 1 M HEPES, 10 mM *D*/*L* methionine, 5 mM TCEP, final pH 7.8. Immediately before the reaction, DTT was added to a final concentration of 20 mM and 2-bromoethylamino hydrobromide 0.14 M. The reaction mixture was incubated at room temperature for 10 hours. The alkylation mixture was acidified to stop the reaction, analyzed and purified by RP-HPLC. The identity of the alkylated product was confirmed by MALDI-MS or ESI.

3.2.12 Histone octamer formation and purification

The four histones with equal molar amount (around 1mg each) were individually dissolved in the unfolding buffer (7 M Gdn-HCl, 10 mM Tris-HCl, pH 7.5, 10 mM

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

DTT) to a final concentration of 2 mg/mL. For histone protein H3, 20 mM DTT was added. After a 30 min unfolding period, the four proteins were mixed together. The mixed solution was dialyzed against 600 mL of refolding buffer (2 M NaCl, 10 mM Tris-HCl, pH 7.5, 1 mM Na-EDTA, 10 mM 2-mercaptoethanol) at 4 ℃. The dialysis was carried out three times, each time for at least 4 h. The precipitated material formed during dialysis was removed by centrifugation at 20,000 *g* for 10 min at room temperature. The supernatant was concentrated using an Amicon concentrator (MW cut-off of 10 kDa) and purified by size-exclusion chromatography using a 26/60 Sephacryl S-200 column which was previously equilibrated with refolding buffer. The fractions were collected and analyzed by 18% SDS-PAGE. The purified octamer solution was mixed with an equivolume of glycerol, and stored at -20 ℃.

3.2.13 Assembly of the nucleosome core particle

5'-CTGGAGAATCCCGGTGCCGAGGCCGCTCAATTGGTCGTAGACAGCTCTAGCAC CGCTTAAACGCACGTACGCGCTGTCCCCCGCGTTTAACCGCCAAGGGGATTACTCC CTAGTCTCCAGGCACGTGTCAGATATATACATCCTGT-3' was synthesized by Proligo as a template to prepare the 146 bp core DNA. 5'-CGGGATCCCGGCGCCCTGGAGAATCCCGGTGCC-3' and 5'-GTCAGATATATACATCCTGTGCATGGAAGATCTTCGCTCGAGCG-3' were used as the forward and reverse primers, respectively. The PCR condition for

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

amplification of this core DNA fragment was: 95 ℃, 30 s, 30 cycles of 95 ℃, 30 s, 54 ℃, 30s, and 72 ℃, 30 s; 72 ℃, 5 min. The amplified DNA was gel purified and ligated to T-vector (Promega). After sequencing to verify the sequence, the T-vector was transformed into *E. coli* XL1-Blue $CaCl_2$-competent cells. After replication and extraction, the T-vector containing the 185 bp core DNA was used as a template to prepare this 185 bp core DNA.

The histone octamer was mixed with the 185 bp core DNA fragment with a final 0.9 molar ratio of octamer to DNA and with a final DNA concentration of 0.2 μM. KCl was added to a concentration of 2 M, and DTT was added to a concentration of 10 μM. The mixture was let stand at 4 ℃ for 30 min, and was then stepwisely dialyzed against TCS-0.85 buffer (0.85 M KCl, 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT), TCS-0.65 buffer (0.65 M KCl, 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT), TCS-0.45 buffer (0.45 M KCl, 20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT), and TCS-0 buffer (20 mM Tris-HCl, pH 7.5, 1 mM EDTA, 1 mM DTT) [96]. Each dialysis took 2.5 h. The nucleosome core particle formation was verified by resolving on a 5% native PAGE gel (5% acrylamide, 0.15% bis-acrylamide), followed by staining with ethidium bromide.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.3 Results and discussion

3.3.1 The application of thioacid capture ligation to the synthesis of a model peptide

The key element of thioacid capture ligation consists of specific capture of a C-terminal thioacid of the first peptide by an activated disulfide from an Npys-modified N-terminal Cys side chain of the second peptide to form an acyl disulfide intermediate which undergoes rapid intramolecular acylation to generate an amide bond. We first demonstrated in a model study that thioacid peptides prepared by our new solid phase hydrothiolysis method were viable substrates for thioacid capture ligation. Peptide thioacid, H_2N-**RLLLPGELA**-COSH (peptide thioacid **4** in Table 2.1, Chapter 2), was assembled by solid-phase peptide synthesis on a CM resin, and released from the solid support by hydrothiolysis reaction. The Npys-modified peptide used was Npys-H3(110-135)K115ac, H_2N-**C(Npys)AIHA*Kac*RVTIMPKDIQLARRIRGERA**-COOH (see next section for preparation). The ligation was achieved by simply mixing the two peptides in 10% ACN/H_2O (0.05% TFA, pH 2). A yellow color developed immediately. After the color developed, the pH of the reaction was adjusted to 5-6 with 0.2 M sodium phosphate buffer, pH 7. 10 to 15 min later, TCEP or DTT was added to the reaction mixture for thiolytic reduction [57]. After the reduction, the reaction mixture was analyzed by RP-HPLC as shown in Figure 3.3.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

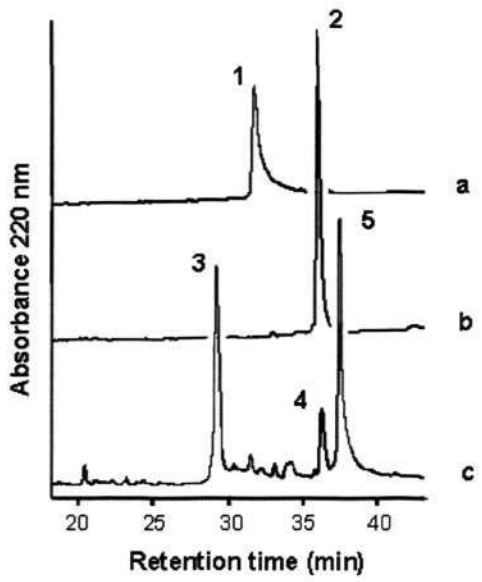


Figure 3.3 C18 analytical HPLC monitoring of ligation between H_2N-**RLLLPGELA**-COSH and Npys-H3(110-135). Trace **a**, peak 1 is Npys-H3(110-135) or H-**C(Npys)AIHA*K*(ac)RVTIMPKDIQLARRIRGERA**-COOH. *m/z* $[M+3H]^{3+}$ found: 1067.8, MW calcd: 3197.7. Trace **b**, peak 2 is H_2N-**RLLLPGELA**-COSH. Trace **c**, TCEP-reduced reaction mixture after 20 min ligation. Peak 3 is reduced H3(110-135). *m/z* $[M+2H]^{2+}$ found: 1523.6, MW calcd: 3043.7. Peak 4 is the hydrolysis product of the peptide thioacid. *m/z* $[M+H]^{+}$ found: 981.8, MW calcd: 980.6. Peak 5 is the ligation product, H_2N-**RLLLPGELACAIHA*K*(ac)RVTIMPKDIQLARRIRGERA**-COOH. *m/z* $[M+3H]^{3+}$ found: 1337.2, MW calcd: 4006.3. HPLC linear gradient: 0% to 40% of buffer B in buffer A for 40 min.

3.3.2 Synthesis of histone H3 containing modified Lys at the C-terminal tail

3.3.2.1 Synthesis of Npys-modified histone H3 C-terminal peptide fragment (Npys-H3(110-135))

Standard Fmoc solid phase peptide synthesis was used for the preparation of the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

H3 C-terminal peptide fragment (110-135), H_2N-**CAIHAKRVTIMP*K*DIQLARRIRGERA**-COOH, which contains either an ε-acetylated Lys or an unmodified lysine at position 122. These H3 peptide fragments needed to be modified by 2,2'-dithiobis-(5-nitropyridine) for thioacid capture ligation in the next step. We employed two ways to modify the peptides by Npys. In the first method (method I), the RP-HPLC purified peptide was dissolved in a minimum amount of acetic acid:water (3:1, v/v) containing 3-5 molar equivalents of 2,2'-dithiobis-(5-nitropyridine). The reaction was complete after 4 to 6 hours with vigorous shaking [94]. The Npys-modified peptide was obtained after HPLC purification. The other method (method II) used was to perform the Npys modification during the cleavage. In this case, the cocktail for cleavage was 95% TFA/2.5% water/2.5% TIS containing 3-5 equivalents of 2,2'-dithiobis-(5-nitropyridine) [95]. The pure Npys-modified peptide was obtained by ether precipitation and HPLC purification (see Figure 3.4 and 3.5). Both methods gave a high yield of the Npys-modified peptides.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

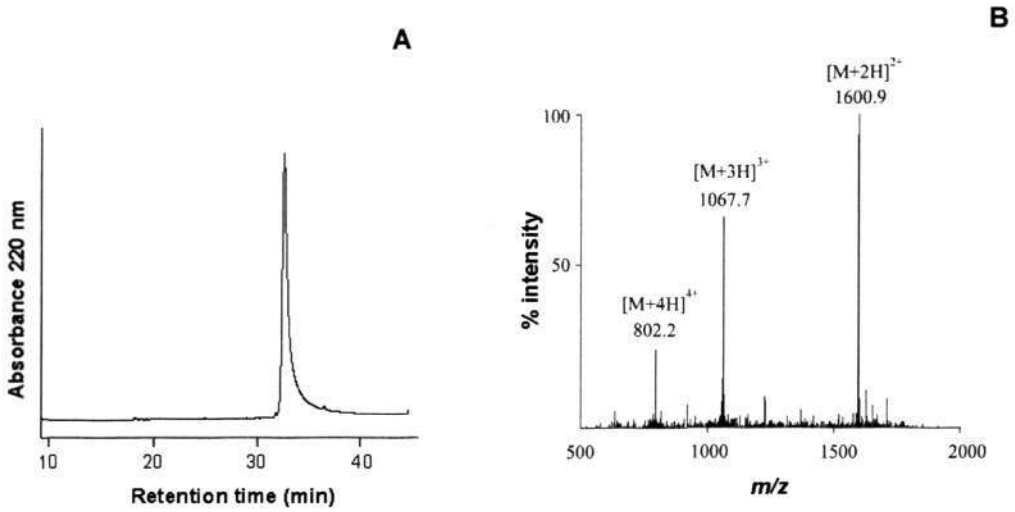


Figure 3.4 Characterization of Npys modification of H3(110-135)K122ac. (**A**) C18 analytical HPLC profile of pure H_2N-**C(Npys)AIHAKRVTIMP*Kac*DIQLARRIRGERA**-COOH prepared by method I. HPLC condition: 0-40% of Buffer B in buffer A over 40 min. (**B**) Molecular weight of the Npys-modified peptide as measured by ESI-MS. *m/z* $[M+2H]^{2+}$ found:1600.9, MW calcd: 3197.7.

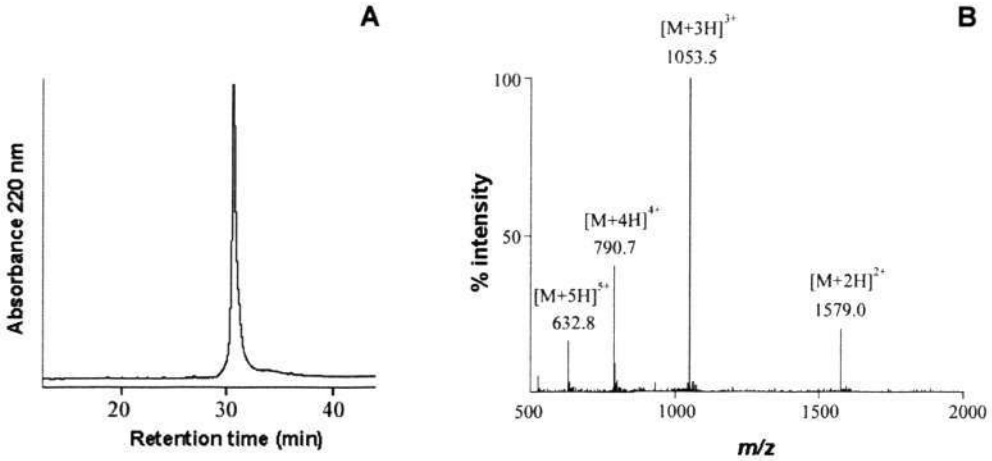


Figure 3.5 Characterization of Npys modification of H3(110-135). (A) C18 analytical HPLC profile of pure H_2N-**C(Npys)AIHAKRVTIMP*K*DIQLARRIRGERA**-COOH prepared by method II. HPLC condition: 0-40% of Buffer B in buffer A over 40 min. (B) Molecular weight of the Npys-modified peptide was measured by ESI-MS. *m/z* $[M+3H]^{3+}$ found:1053.5, MW calcd: 3156.8.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.3.2.2 Preparation of full length histone H3 by thioacid capture ligation between H3(1-109) and H3(110-135)

There are several reasons that account for the high efficiency of thioacid capture ligation [57]. The capture step is instantaneous because of the highly activated Npys-S disulfide and the supernucleophilicity and low pK_a value of the thioacid compared to a normal alkyl sulfhydryl group [36]. The intramolecular acyl transfer step is also very rapid, as the acyl disulfide activates the carbonyl group for a highly efficient nucleophilic substitution reaction with the closely positioned N^{α}-amine. As a result, the entire ligation reaction is usually complete within 5-10 min. In spite of this, the thioacid capture ligation method has not yet been used to prepare any proteins. In this study, we successfully synthesized histone H3 proteins with or without posttranslational modification marks by using thioacid capture ligation. The capture reaction was carried out under weakly acidic condition, pH 2-3. After adjusting the pH to 6, the ligation reaction was allowed to proceed for half an hour for intramolecular acylation completeness. Because the histone protein is insoluble in aqueous buffer, 6 M Gdn-HCl was included in the reaction buffer for protein solubilization. The ligation mixture was analyzed by SDS-PAGE which showed the ligation product at the expected MW size (Figure 3.6 A). After HPLC purification, the pure ligation product was confirmed by MALDI-TOF (Figure 3.6 B, C)

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

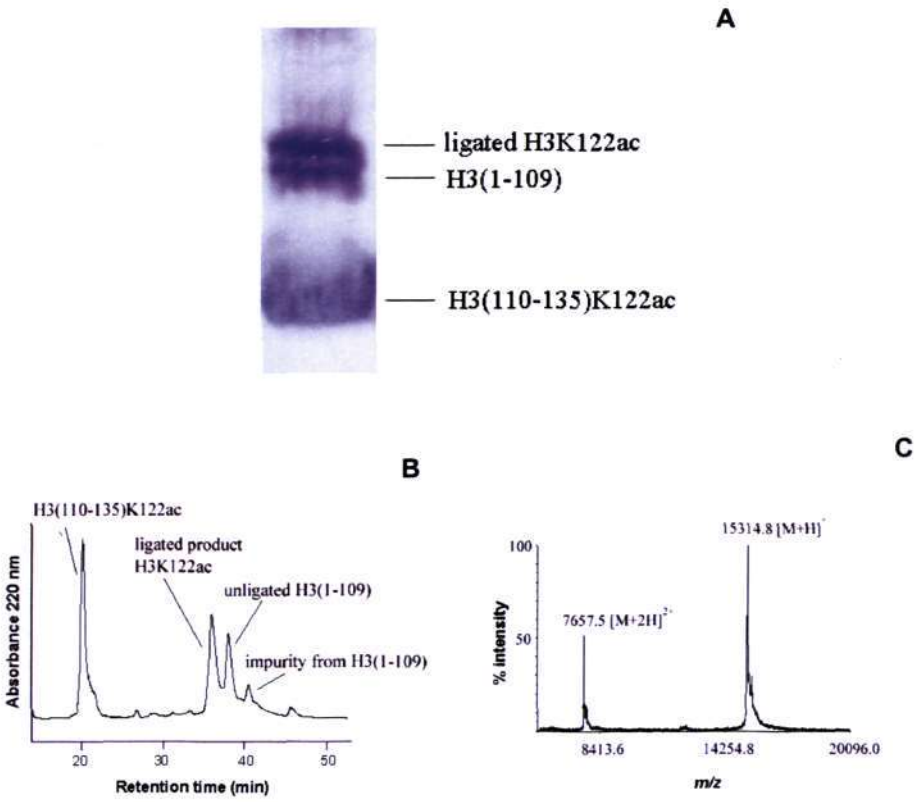


Figure 3.6 Thioacid capture ligation between histone H3(1-109)-COSH and H_2N-**C(Npys)AIHAKRVTIMP*Kac*DIQLARRIRGERA**-COOH (**A**) 15% SDS-PAGE gel analysis of the ligation mixture. (**B**) C8 semi-prep RP-HPLC purification of the ligation mixture. The ligation product was eluted by using a gradient: 0% to 50% in 25 min, then to 60% in 20 min of Buffer B in buffer A. (**C**) MALDI-TOF mass spectral analysis of HPLC purified ligation product. *m/z* $[M+H]^+$ found:15314.8, MW calcd: 15312.2.

We also synthesized the full length wild type H3 by using thioacid capture ligation between H3(1-109)-COSH and Npys-H3(110-135) (Figure 3.7). The ligation product and unligated H3(1-109) were not well separated. The separation of proteins on RP-HPLC column is mainly governed by hydrophobic interactions. Addition of a relatively hydrophilic and small C-terminal peptide compared to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

H3(1-109) may not contribute enough hydrophobicity to make the full length H3 protein well separable from H3(1-109). Similar results can also be seen in Figure 3.13, 3.14, 3.15, 3.16 and 3.17. In those cases, the H3 N-terminal peptides are much more hydrophilic than the H3 C-terminal peptide, therefore, it was even harder to separate the ligation products from the unligated H3(14-135)K14C. The separations were not significantly improved after changing the HPLC gradient. We also used a cation exchange HPLC column-CM300 (250 x 4.6 mm, Eprogen, USA) to separate the product of the ligation reaction of H3(1-109) and H3(110-135). The result was worse than using a RP-HPLC column (data not shown).

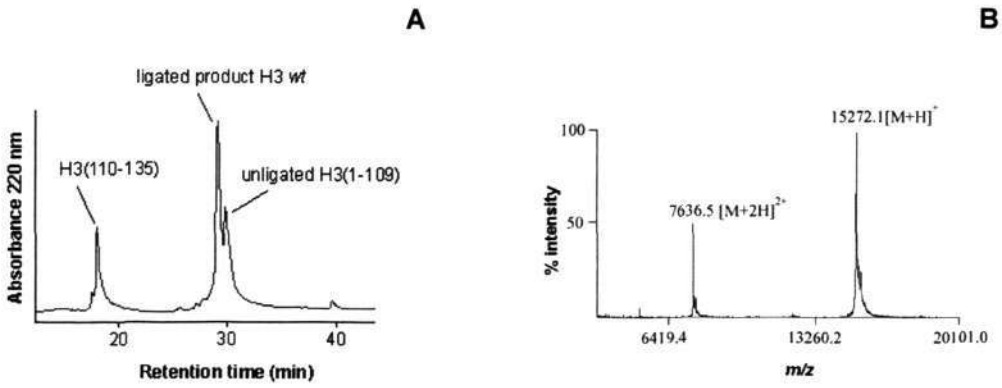


Figure 3.7 Wild type histone H3 produced by thioacid capture ligation. (**A**) C4 analytical RP-HPLC analyis of ligation reaction. The ligation product was eluted by using a linear gradient of 0% to 60% Buffer B in buffer A over 30 min. (**B**) MALDI-TOF mass spectral analysis of HPLC purified ligation product. *m/z* $[M+H]^+$ found:15272.1, MW calcd: 15270.2.

We observed that the ligation yield could be affected adversely by traces of H_2S remaining in protein thioacids which would reduce the Npys-activated disulfide bond. Therefore, after hydrothiolysis, the protein thioacids should be carefully

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

purified and lyophilized repeatedly if necessary to remove all H_2S from the samples. The above work demonstrated that thioacid capture ligation provides a highly efficient, simple and practical way to semi-synthesize proteins containing a modified amino acid in the C-terminal region.

3.3.3 Synthesis of histone H3 containing N-terminal posttranslational modification

Since most H3 posttranslational modifications occur at the N-terminal tail, we also used the thioacid capture ligation and native chemical ligation to synthesize such modified H3. In this case, chemically synthesized H3 N-terminal peptide thioacids were ligated with a recombinantly produced, N-terminally-truncated H3 protein fragment.

3.3.3.1 Chemical synthesis of histone H3 N-terminal peptide thioesters and thioacid

H3 N-terminal peptide thioester, H3(1-13) thioester, H_2N-**ARTKQTARKSTGG**-CO-$SCH_2CH_2CONH_2$ with N^{ε}- dimethylated Lys4 or Lys9 was synthesized by standard Boc SPPS. The thioester bond was directly generated on the MBHA resin by using *S*-trityl mercaptopropionic acid [48]. The amino acids were assembled onto the mercaptopropionyl MBHA resin following standard SPPS protocols. After HF deprotection and cleavage, the peptide thioester

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

was released from the resin. The peptide thioester was purified by RP-HPLC and the molecular weight was determined by ESI-MS (see Figure 3.8 and 3.9).

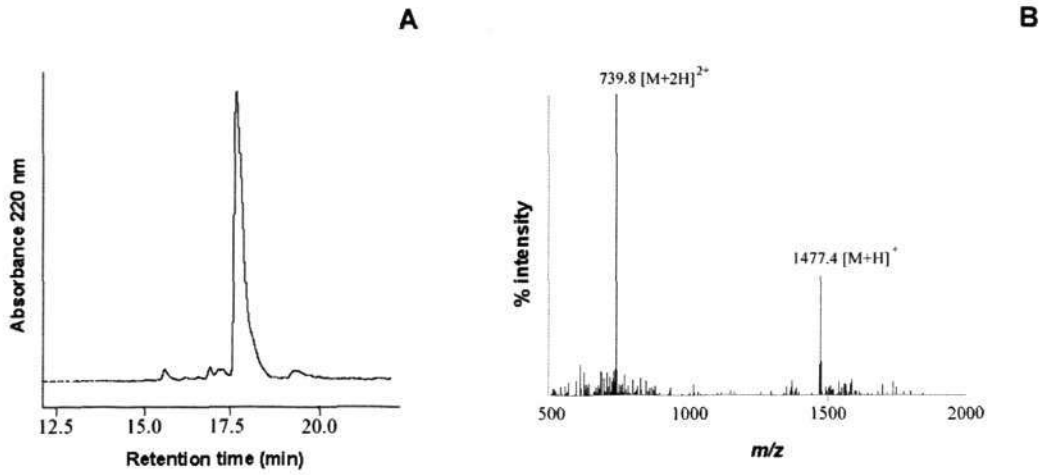


Figure 3.8 Characterization of the H3(1-13)/K4me2 thioester. (**A**) C18 analytical HPLC profile of H_2N-**ART*Kme2*QTARKSTGG**-CO-$SCH_2CH_2CONH_2$. HPLC condition: 0% to 30% of buffer B in buffer A over 30 min. (**B**) Mass spectrum of this peptide thioester determined by ESI-MS. *m/z* $[M+2H]^{2+}$ found: 739.8, MW calcd: 1477.8.

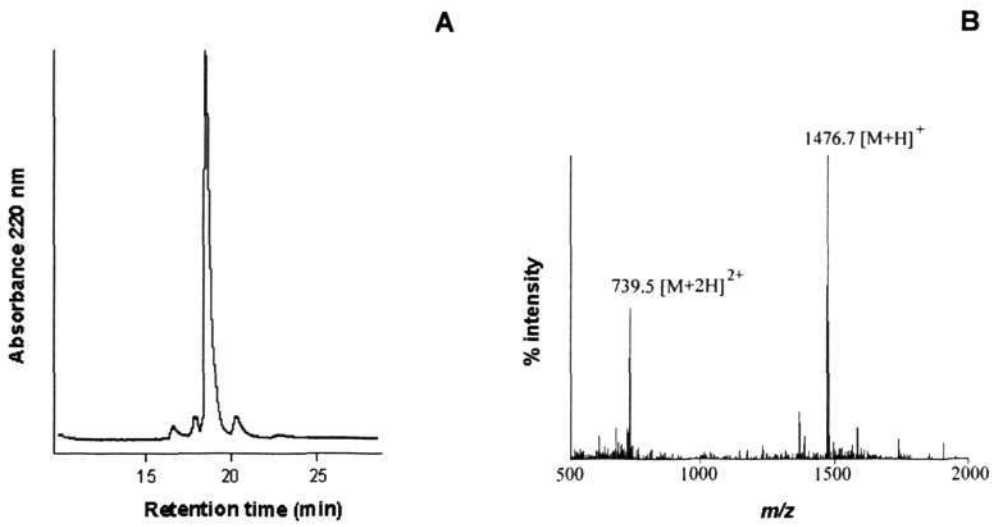


Figure 3.9 Characterization of H3(1-13)/K9me2 thioester. (**A**) C18 analytical HPLC profile of H_2N-**ARTKQTAR*Kme2*STGG**-CO-$SCH_2CH_2CONH_2$. HPLC condition: 0% to 30% of buffer B in buffer A over 30 min. (**B**) Mass spectrum of this peptide thioacid as determined by ESI-MS. *m/z* $[M+2H]^{2+}$ found: 739.5, MW calcd: 1477.8.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

To make the peptide thioacid, the peptide thioester was hydrothiolysed with 0.1 M Na_2S in 0.25 M HEPES, pH 8.0, at room temperature for 4-5 hours. After HPLC purification, the pure peptide thioacid was obtained. The purity of the peptide thioacid was checked by analytic HPLC and the identity was confirmed by electrospray mass spectrometry (Figure 3.10)

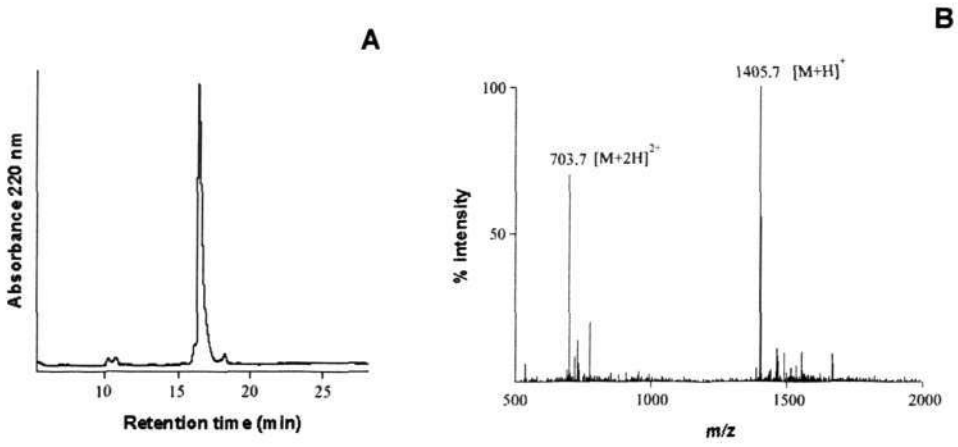


Figure 3.10 Characterization of H3(1-13)/K4me2 thioacid. (**A**) C18 analytical HPLC profile of H_2N-**ART*Kme2*QTARKSTGG**-COSH. HPLC condition: 0% to 30% of Buffer B in buffer A over 30 min. (**B**) Mass spectrum of this peptide thioacid determined by ESI-MS. *m/z* $[M+2H]^{2+}$ found: 703.7, MW calcd: 1406.7.b

3.3.3.2 Preparation of Npys-H3(14-135)K14C

Since a Cys residue is required for the chemical ligation reaction, we mutated the Lys14 to Cys by site-directed mutagenesis of a *Xenopus* histone H3. The truncated H3, H3(14-135)K14C was expressed in *E. coli* strain BL21(DE3)/pLysS, and the N-terminal methionine of the expressed protein was auto-removed during protein expression. This was confirmed by MALDI-TOF mass spectral analysis (Figure

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.11).

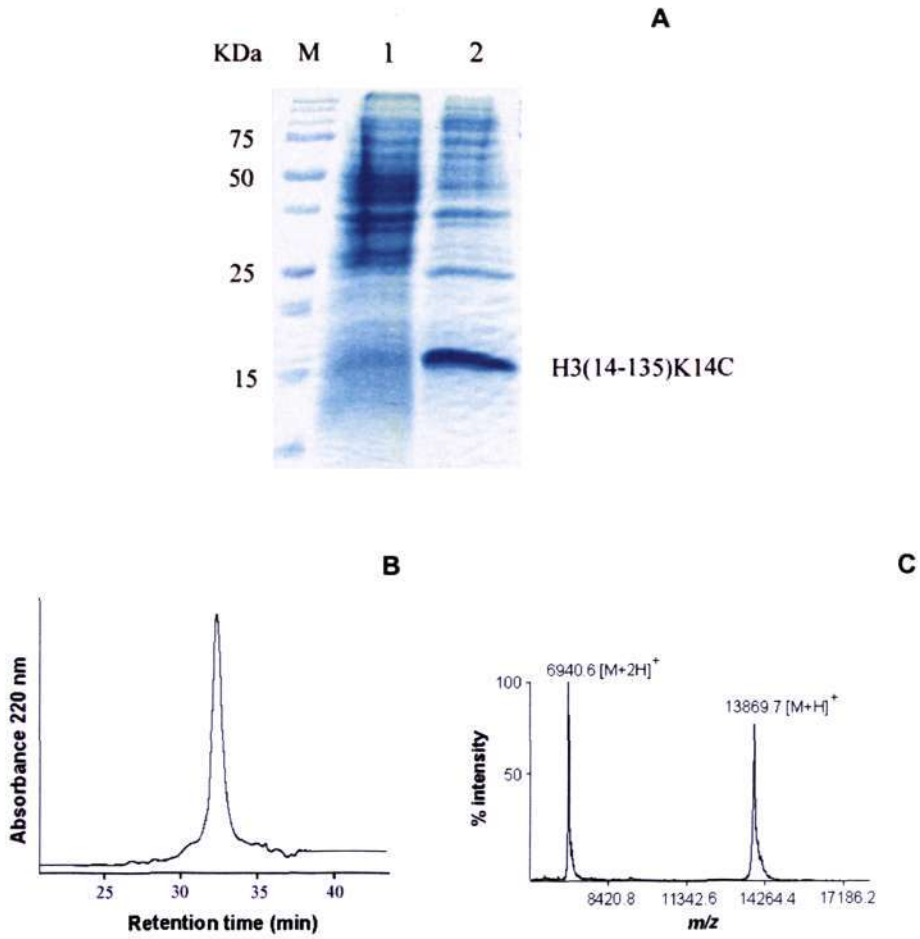


Figure 3.11 Expression and characterization of H3(14-135)K14C. (**A**) 15% SDS-PAGE gel analysis of H3(14-135)K14C expression in *E. coli*. Lane 1, uninduced cell culture; lane 2, cell culture induced by IPTG. The expressed protein is in the inclusion body. (**B**) C8 semi-prep RP-HPLC purification of H3(14-135)K14C. HPLC condition: 0% to 50% in 25 min, then to 60% in 20 min of buffer B in buffer A. (**C**) MALDI-TOF MS of H3(14-135)K14C, which shows that the expressed protein has no methionine at its N-terminus. *m/z* $[M+H]^+$ found:13869.7, MW calcd: 13870.3.

The HPLC purified H3(14-135)K14C was reacted with 3-5 eq. of 2,2'-dithiobis-(5-nitropyridine) in 3:1 of acetic acid/water. The reaction was highly

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

efficient, and almost all of the H3(14-135)K14C was converted to Npys-H3(14-135)K14C. Semi-prep HPLC and ESI-MS results (Figure 3.12) showed the purity and the correct MW of the product.

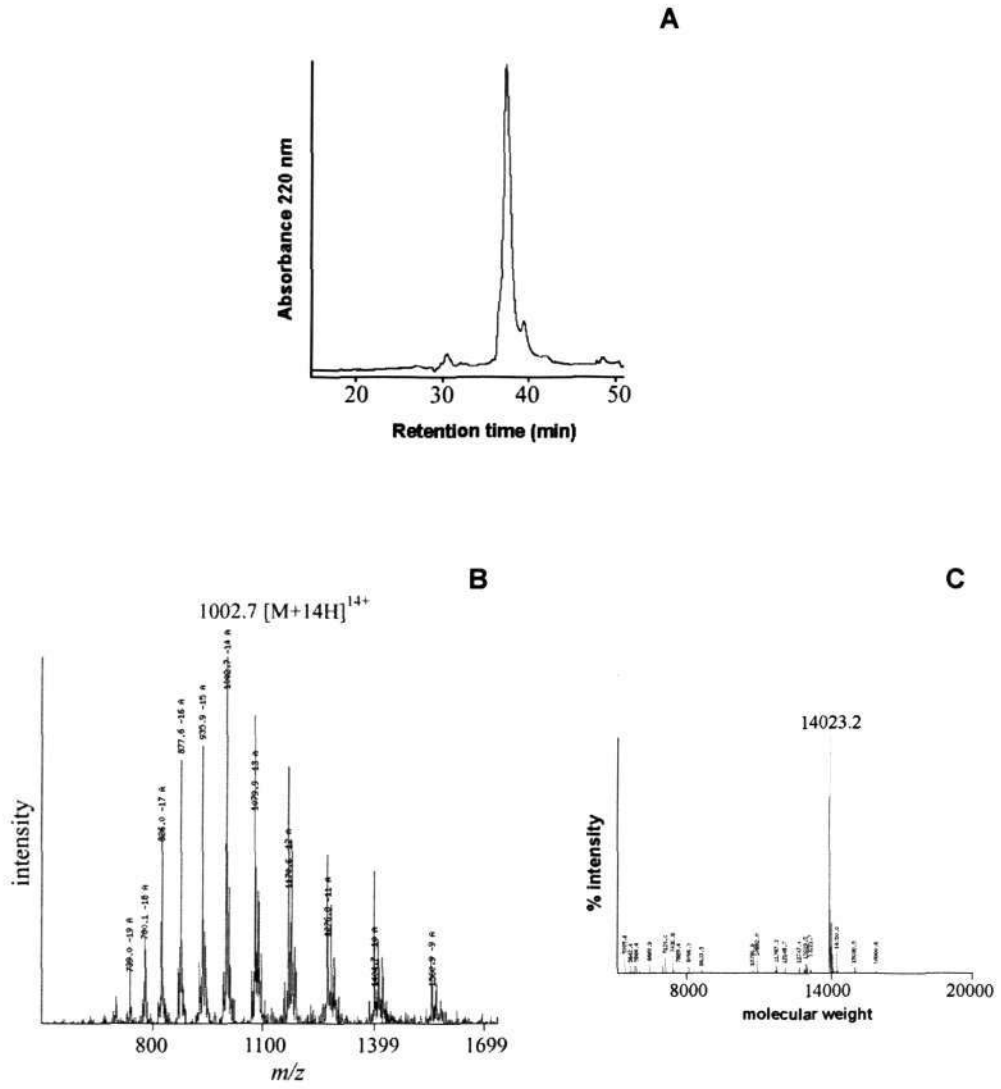


Figure 3.12 Characterization of Npys-modified H3(14-135)K14C. (**A**) C8 semi-prep purification of H3(14-135)K14C after Npys-modification reaction. HPLC condition: 0% to 50% in 25 min, then to 60% in 20 min of buffer B in buffer A. (**B**) ESI-MS mass spectrum, and (**C**) Deconvoluted ESI mass spectrum of Npys-H3(14-135)K14C. MW found: 14023.2, MW calcd: 14024.5.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

3.3.3.3 Preparation of full length H3K14C containing K4me2 by thioacid capture ligation

Since the Npys group has been successfully added to the N-terminal Cys of H3(14-135)K14C, we can ligate the Npys-modified protein with the chemically synthesized N-terminal peptide thioacid, H3(1-13)-COSH, containing N^{ε}-dimethylated Lys4 to obtain the full length H3. Fig 3.13 shows the ligation reaction analysed by HPLC and the ligation product was confirmed by MALDI-MS. Again, the ligation was carried out in 6 M Gdn-HCl to keep the protein denatured.

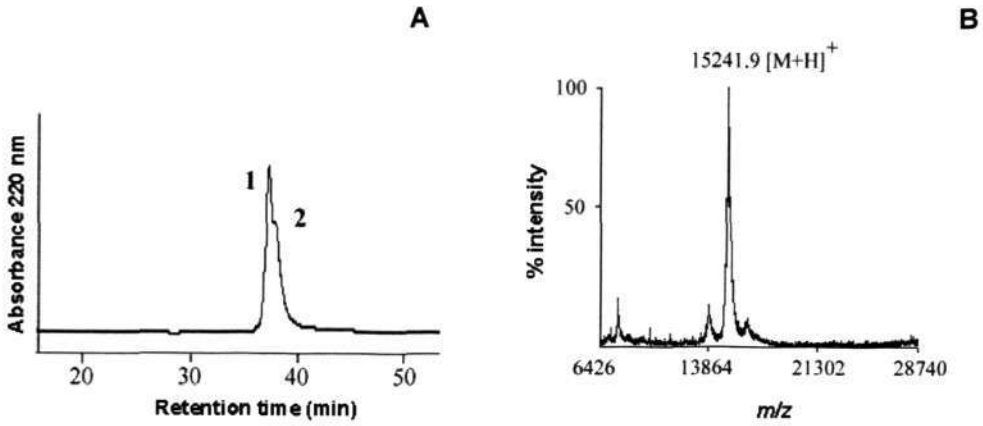


Figure 3.13 Characterization of the ligation between H3(1-13)-COSH and Npys-H3(14-135)K14C. (**A**) C8 semi-prep HPLC analysis of the ligation reaction. Peak 1, is the ligated product, H3K4me2/K14C, as confirmed by MALDI-MS. Peak 2, is the unligated H3(14-135)K14C. HPLC condition: 0% to 50% in 25 min, then to 60% in 20 min of buffer B in buffer A. (**B**) MALDI-MS spectrum of the ligation product. *m/z* $[M+H]^+$ found: 15241.9, MW calcd: 15241.7.

This result of the performed ligation indicates that the thioacid capture ligation method can also be applied to the coupling of peptide thioacids with

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Npys-modified proteins with reasonably good yield in the short reaction time of half an hour.

3.3.4 Alkylation of H3 K14C

So far, most chemical ligation methods involve a cysteine residue at the ligation site and retain it after ligation. Cysteine is the second least common amino acid comprising only 1.7% of all amino acid residues in proteins [97]. In histone proteins, there is only one cysteine residue present in the C-terminal part of H3, which was used for the ligation described in the previous section. The N-terminal portions of histone proteins contain no Cys at all but a large number of lysines, many of which are modified posttranslationally. Shokat and coworkers [90] used a traditional aminoethylation method to convert a cysteine to a lysine analog to install different lysine methylation patterns on histones. In this study, we mutated a lysine residue in the N-terminal region to a cysteine residue for use as the ligation site. This mutation allowed us to ligate a chemical synthesized N-terminal peptide bearing any modification with the remaining C-terminal domain by using either native chemical ligation or thioacid capture ligation which is a more effective ligation method. The cysteine residue was finally converted to *S*Lys by reaction with bromoethylamine hydrobromide. *S*Lys is a close isostere of Lys. The two have comparable steric and electronic properties, with *S*Lys having slightly less basicity due to the electron withdrawing effect of the thioether. The cleavage of a protein

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

containing aminoethylcysteine by lysine-directed proteases such as trypsin indicated that the aminoethylcysteine is similar to lysine [91]. We have prepared two H3 analogs by using alkylation of H3K14C after ligation of H3(14-135) and a H3 N-terminal peptide fragment containing either K4me2 or K9me2 through native chemical ligation which took 40 h (Figure 3.14 and 3.15).

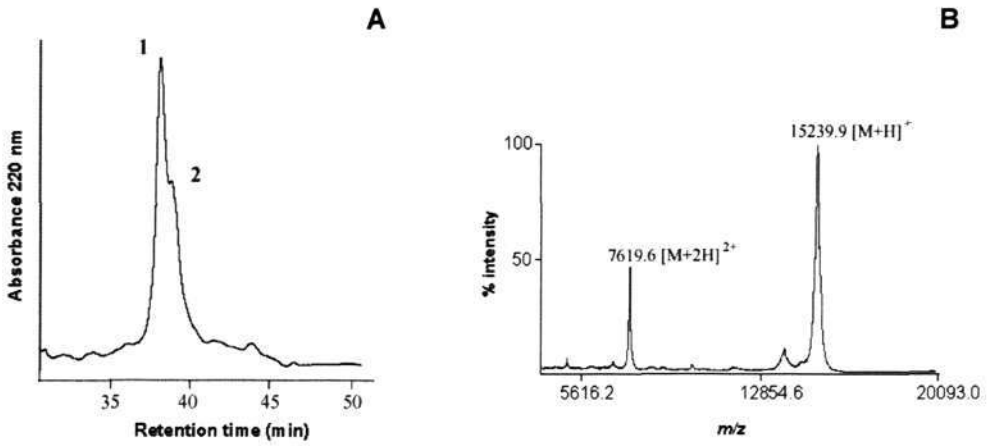


Figure 3.14 Ligation reaction between H3(1-13)/K4me2-CO-$SCH_2CH_2CONH_2$ and H3(14-135)K14C. 0.5 mM H3(14-135)K14C and 3 eq. of H3 N-terminal peptide thioester was reacted in 0.2 M HEPES, pH 8.5, 6 M Gdn-HCl, 0.1 M NaCl, 40 mM TCEP, 2% (w/v) MESNA, and 2% (w/v) thiophenol sodium salt. (**A**) After 40 hours, the ligation product was purified by C8 semi-prep RP-HPLC column. Peak 1, the ligated product, H3K4me2K14C; peak 2, unligated H3(14-135)K14C. HPLC condition: 0% to 50% in 25 min, then to 60% in 20 min of buffer B in buffer A. (**B**) MALDI-TOF analysis of the ligation product, H3K4me2/K14C. *m/z* $[M+H]^+$ found: 15239.9, MW calcd: 15241.7.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

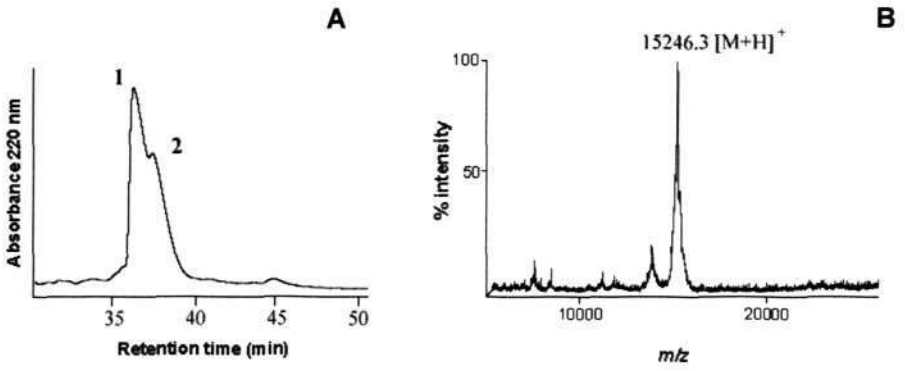


Figure 3.15 Ligation reaction between H3(1-13)/K9me2-CO-$SCH_2CH_2CONH_2$ and H3(14-135)K14C Reaction condition is the same as described in the legend of Figure 3.14. (**A**) After 40 hours, the ligation product was purified by C8 semi-prep RP-HPLC column. Peak 1, the ligated product, H3K9me2K14C; peak 2, unligated H3(14-135)K14C. HPLC condition: the same as described in the legend of Figure 3.14. (**B**) MALDI-TOF analysis of ligation products, H3K9me2/K14C. *m/z* $[M+H]^+$ found: 15246.3, MW calcd: 15241.7.

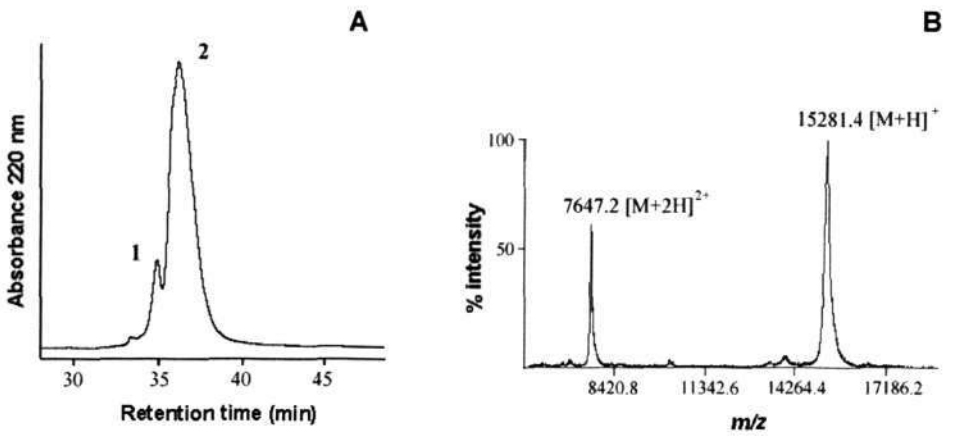


Figure 3.16 Converstion of Cys to *S*Lys in the ligation product H3K4me2/K14C. Reaction condition: 0.5 mM H3K4me2/K14C was alkylated in 6 M Gdn-HCl, 1 M HEPES, 20 mM D/L methionine, 5 mM TCEP, final pH 7.8, and 0.14 M 2-bromoethylamino hydrobromide, for 10 h. (**A**) The alkylated product H3K4me2/*S*Lys14 was isolated by C8 semi-prep RP-HPLC column. Peak 1 is the unligated H3(14-135)K14C; peak 2 is the alkylated product. HPLC condition: 0% to 50% in 25 min, then to 60% in 20 min of buffer B in buffer A. (**B**) MALDI-TOF spectrum of the alkylated product, H3K4me2/*S*Lys14. *m/z* $[M+H]^+$ found: 15281.4, MW calcd: 15282.7.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

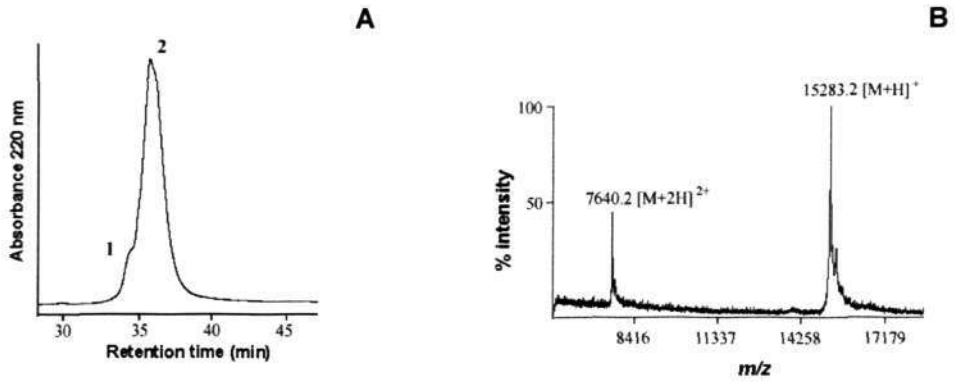


Figure 3.17 Converstion of Cys to *S*Lys in the ligation product, H3K9me2/K14C. Reaction condition is the same as in Figure 3.16. (**A**) The alkylated product H3K9me2/*S*Lys14 was isolated by C8 semi-prep RP-HPLC column. Peak 1 is the unligated H3(14-135)K14C; peak 2, the alkylated product. HPLC condition is the same as in Figure 3.16. (**B**) MALDI-TOF spectrum of the alkylated product, H3K9me2/*S*Lys14, *m/z* $[M+H]^+$ found: 15283.2, MW calcd: 15282.7

Since there is only one conserved Cys residue (H3 C110) in all four histones H2A, H2B, H3 and H4 and the mutation of this Cys to Ala does not disrupt the protein functions, it is possible to introduce a unique Cys at any position of H3 (in conjuction with C110A mutation). By using the alkylation method, the Cys residue can be converted to *S*Lys, a close analog of Lys. The pH was crucial for the alkylation reaction and should be optimized to get most Cys deprotonated and avoid side reactions at other sites. Different pHs were tested on model peptides and a pH value of 7.8 was found to be optimal (data not shown). The reaction was performed for about 10 h and little di-alkylated product was obtained (Figure 3.16 and 3.17). The oxidation of Met was prevented by the addition of free methionine to the alkylation solution. This method provides an efficient and economic access

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

to large amount of modified proteins with a Lys at the ligation site.

3.3.5 Octamer formation and nucleosome core particle assembly

Wild type H3 or modified H3 - H3K122ac, H3K4me2/SLys14, H3K9me2/*S*Lys14 were used respectively to form the octameric histone complex with recombinant H2A, H2B and H4 in equal amount. The histones were individually dissovled in denaturing buffers and then mixed together. The formation of histone octamers was achieved by serial dialysis from 7 M guanidinium-HCl to 2 M NaCl. Gel filtration was used to isolate the octamers from high molecular mass aggregates and excess free histone proteins. After purified by gel filtration, the octamers were analyzed by 15% SDS-PAGE, which showed equal molar distribution of the four histone proteins in the purified octamer sample (Figure 3.18). The results showed that there were no significant differences in the efficiency of octamer formation among expressed wild type H3, synthesized wild type H3, and modified H3 proteins. Furthermore, the octamer complexes were stable in 2 M NaCl solution.

We then tested the binding ability of these histone octamers towards a 185 bp DNA template. The results were analyzed by 5% native PAGE gel stained by ethidium bromide (Figure 3.19). Most of the template DNA were assembled with octamers to form mononucleosomes. There are no detectable differences in the ability of forming mononuclesomes between the different octamers.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

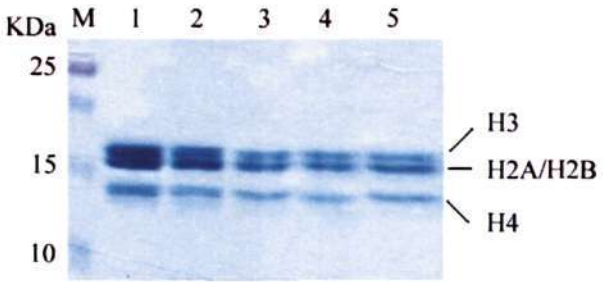


Figure 3.18 15% SDS-PAGE gel analysis of histone octamer formation. Lane 1, histone octamer with all expressed histone proteins; lane 2, with synthesized wild type H3; lane 3, with H3K122ac; lane 4, with H3K4me2/*S*Lys; lane 5, with H3K9me2/*S*Lys.

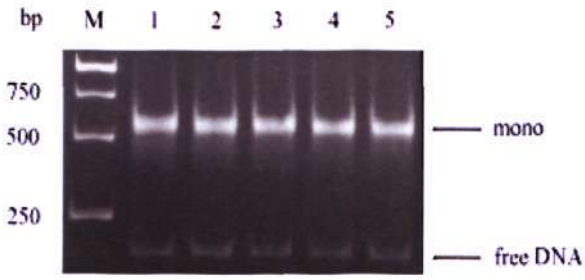


Figure 3.19 Analysis of nucleosome core particle assembly on 5% native PAGE gel. Lane 1, with expressed wild type H3; lane 2, with synthesized wild type H3; lane 3, with H3K122ac; lane 4, with H3K4me2/SLys; lane 5, with H3K9me2/SLys. **Mono** indicates the mononuclesome formed by the octamer and the 185 bp DNA template. **Free DNA** indicates the unbound 185 bp DNA.

3.4 Conclusion

Up to date, the thioacid capture ligation has received relatively little attention mainly due to the difficulty in obtaining the thioacid building blocks. The solid phase and solution hydrothiolysis reactions make peptide and protein thioacids

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

easily obtainable and therefore provide key building block materials for thioacid capture ligation. We used the solution phase synthesis derived H3 N-terminal peptide thioacids (H3(1-13)-COSH) for ligation with the recombinant H3 C-terminal domain which contained an N-terminal Cys residue modified by Npys (Npys-H3(14-135)K14C). Similarly, H3 analogs containing C-terminal modifications were prepared by ligation of H3(1-109)-COSH with the synthetic, Npys-modified H3 N-terminal peptide H3(110-135). These represented the first examples of thioacid capture ligation applied to protein synthesis, demonstrating that this method is very efficient and easy to use.

Besides using thioacid capture ligation for H3 synthesis, we also prepared H3 analogs by native chemical ligation. We found that the two methods gave comparable yields of ligation products, but with the thioacid capture ligation requiring much less time. The Cys residue at the ligation site in the ligation products could be converted to *S*Lys after ligation. This method is very useful for histone protein synthesis since there is no Cys but many Lys residues in the N-terminal parts of the four histone proteins. Therefore, any of the Lys residues could be mutated to a Cys residue to generate a suitable ligation function and could then be converted back to SLys after ligation. This expands the application scope of both thioacid capture ligation and native chemical ligation. This coupled ligation-alkylation scheme provides an efficient and economical way to obtain large amounts of modified histone proteins.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Chapter 4: Thioacid/Azide Coupling as a Novel Site-Selective Method for Peptide and Protein C-terminal Tagging

4.1 Introduction

Site-selective protein modification techniques play important roles in protein structure and function studies and in drug discovery. Such techniques can be used to label proteins with a special isotope tag or biophysical probe to facilitate proteomic analysis or structural and biochemical studies [98,99], or with a biocompatible polymer for the development of protein-based therapeutic agents [100,101]. Furthermore, they are also useful as immobilizing techniques [102] to tether proteins to solid supports for the fabrication of protein chips.

Traditional methods for protein modification rely mainly on electrophilic acylation or alkylation reactions on the side chain nucleophiles of Lys [103] or Cys residues [104,105]. However, the use of these methods is often limited by the fact that proteins typically carry multiple copies of the targeted residue, especially the relatively abundant Lys, which makes it difficult to control the degree and site of modification. This commonly results in product mixtures with considerable structural heterogeneity. Other methods that target the electron-rich aromatic rings on Trp or Tyr side chains [106,107] have also been developed to functionalize a protein, but again the same problem would exist if one wishes to introduce only

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

one modification at a single site.

Biosynthetic methods that make use of the cell's protein synthesis machinery for protein tagging offer powerful alternatives to chemically based methods. A recent exciting development is the methodology devised by Schultz's group and others [12,108]. This approach relys on an amber stop codon and a corresponding suppressor tRNA that is charged with a desired nonnatural amino acid which may already carry the biophysical probe of interest or contain a special chemical/photochemical reactivity for later use. This concept has been applied to both cell-free and *in vivo* protein synthesis systems which are based on either prokaryotic or eukaryotic organisms [8,109].

Another way for attaching a special probe to a protein at a single site is by using protein splicing *in trans* as performed in the study by Kurpiers *et al.* [110]. Cysteine residues are widely used as the anchor points to introduce protein modifications because of the unique chemical reactivity of the sulfhydryl group and the relatively low abundance of cysteines in proteins. The modifications typically involve thioether or disulfide linkages with different functional groups such as maleimides, haloacetamides or thiols on the tags to be introduced. However, as indicated above, the modification reaction will not be site-selective if the target protein contains several cysteine residues. The method of Kurpiers *et al.* used protein splicing *in trans* to overcome this limitation. First, a peptide bearing a

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

single cysteine is fused to the C-terminus of a split intein's C-terminal fragment and the fusion protein is expressed. Then the cysteine residue of the produced fusion protein is chemically modified. The target protein is fused to the N-terminus of the split intein's N-terminal fragment and expressed. Finally, the labeled Cys-peptide is linked to the target protein by protein splicing *in trans*. By using this technique, regioselectively fluorescein-labeled and catalytically active β-lactamase and thioredoxin were prepared. Furthermore, this approach can also be used to selectively label proteins in cell lysates.

The above methods modify proteins on the side chains of amino acid residues. The choice of the modification site is often critically important as modification at a wrong site may result in a decrease or even complete loss of biological activity of the modified protein. Compared to internal residues, the N- and C-terminal segments of a protein are relatively flexible and represent more desirable sites to introduce a modification as they are usually solvent-exposed and distant from the active site.

There exist many methods to introduce a modification in the terminal segments of a protein. Chemically, periodate oxidation of an N-terminal Ser or Thr residue produces an aldehyde which is useful for tagging a weak base nucleophile [111]. An N-terminal Cys can react with an aldehyde or ketone compound for the formation of a thioazolidine ring [37]. Enzymatically, C-terminal specific labeling

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

can be carried out by catalytic transacylation reactions using carboxypeptidase Y due to its broad specificity and stability [98,112]. Biologically, green fluorescent protein (GFP) is commonly used as a fluorescent reporter fused to the N- or C-terminus of a target gene to trace the synthesis, localization, distribution and biological function of the protein product *in vivo* [113]. A puromycin-based technique provides another way to attach fluorescent or biotin moieties, or other functional groups to the C-termini of expressed proteins at the translation step. Puromycin is a structural analog of aminoacylated-tRNA and can take part in peptide bond formation with the nascent polypeptide chain [114]. Many puromycin derivatives were shown to be incorporated into proteins during translation *in vitro* and *in vivo* [115,116].

Intein-mediated protein splicing [70] generates a thioester intermediate which is reactive towards an appropriate nucleophile and can be used to introduce C-terminal specific modifications. Therefore, a protein can be modified with a cysteinyl-tag through the thioester-mediated peptide ligation chemistry. Nitrogen-derived nucleophiles can also be directly used to react with the intein-derived thioester, which gives rise to an inert amide linkage. For instance, Kalia *et al.* [117] developed a bifunctional reagent which could attack the thioester through its hydrazino functionality and append the other azido group to the target protein. The azido group can then be further modified in a 1,3-dipolar azide-alkyne cycloaddition reaction.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

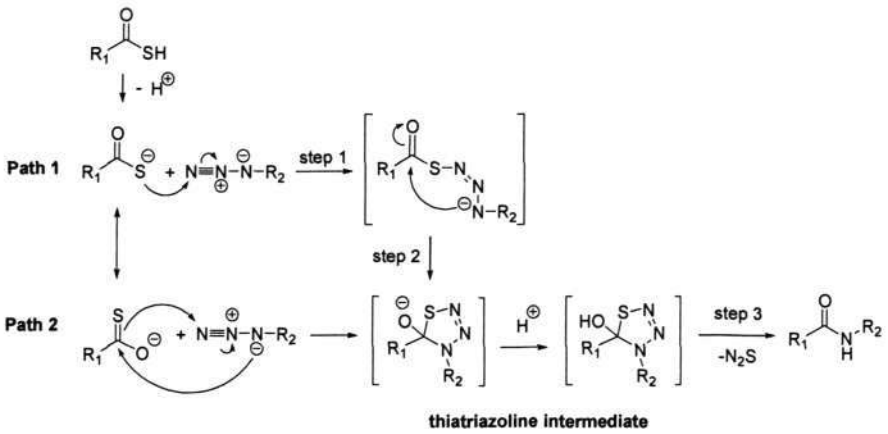
Since we have successfully developed a novel method which is able to convert peptide and protein thioesters to their corresponding C-terminal thioacids through hydrothiolysis reaction with high efficacy and specificity, as shown in Chapter 2, the difficulty in obtaining peptide and protein thioacids does not exist any more. Therefore, it is possible to make use of the thioacid functional group to achieve C-terminal specific peptide and protein modification.

It is known that the thioacid functionality is a very soft and powerful nucleophile, which makes it uniquely useful in organic synthesis. For example, the conjugate addition reactions of sulfur nucleophiles to α, β-unsaturated carbonyl groups of 2(5H)-turanones can generate 4-thio-4,5-dihydro-2(3H)-furanones. The reaction yield is very high with thioacids and such nucleophiles raise the possibility of selective hydrolysis of a thioester to deliver a free thiol functionality at the β-position of the lactone [118]. Thioacid as a powerful nucleophile was also used for the synthesis of β-hydroxy thioesters with high synthetic potential by regioselective ring opening of epoxides in the presence of β-cyclodextrin in water [119]. As a powerful nucleophile, thioacid reacts readily with a bromoacetyl moiety, which has been used to prepare protein analogs with an isosteric thioester linkage [47].

However, the most interesting property of the thioacid functionality useful for bioconjugation resides in its selective reactivity towards certain organic azide

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

compounds, which leads to the formation of amides. The first thioacid/azide amidation was reported in 1980 [120]. Simple acetamides were prepared from the reaction of thioacetic acid with different organic azides [121]. However, only until recently was this thioacid/azide amidation reaction repaid attention by chemists. Williams and coworkers reevaluated this reaction in the development of new synthetic methods for the preparation of complex amides. In a thorough mechanistic study, they have proposed that the reaction proceeds through a thiatriazoline intermediate as shown in Scheme 4.1, and not via an amine intermediate formed by reduction of the azide, as suggested originally [66].



Scheme 4.1 Proposed mechanism of thioacid/azide amidation reaction.

However, electron-poor azides and electron-rich azides seem to take two distinct paths leading to the formation of the 5-membered ring thiatriazoline, although both paths are catalyzed by a base. Highly electron-deficient azides react with thioacids in a prior thiol capture type of reaction whereby nulceophilic addition of the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

thiocarboxylate to the azide links the sulfur to the end nitrogen of the azide group to form a linear intermediate which then rapidly cyclizes to form the thiatriazoline via a five-membered ring transition state. For electron-rich azides, the thiatriazoline intermediate is formed in a single step by concerted anion-accelerated [3+2] cycloaddition. In both paths, the amide product is formed from the thiatriazoline as it decomposes via retro [3+2] cycloaddition with the concurrent loss of nitrogen and sulfur [122]. The reaction is very rapid in the case of highly electron-deficient azides and occurs readily at room temperature and at low concentrations of both reactants. But in the case of electron-rich azides, the reaction is much slower and often requires prolonged heating at elevated temperature and at high concentrations of both reactants.

Conventional methods for amide synthesis always involve activation of the carboxylic acid moiety for its intermolecular acylation onto the amine nucleophile, which requires an appropriate protection group scheme. The distinctly different mechanism of thioacid/azide coupling and the uniqueness of the thioacid and azide functionalities make this reaction a particularly appealing method for the synthesis of complex amide products. An interesting application was the loading of an amino acid onto a sulfonamide resin, compatible with Boc and Fmoc SPPS, by reacting the N-protected amino thioacid with a resin-bound sulfonyl azide. After amino acid assembling, the N-peptidyl sulfonamide resin was activated by alkylation with alkyl bromide. Release of the peptide from the resin by nucleophiles such as a thiol

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

gave peptides with a thioester at their C-terminus which can be used in chemoselective peptide ligation [123].

The coupling reaction of thioacids with electron-deficient azides is particularly appealing, as it possesses all the desirable features of a site-specific bioconjugation method, such as chemo-orthogonality and high efficiency. It is also a complement to the Staudinger ligation method which is very effective for electron-rich azides [59]. Although only small organic thioacids and azides have been used so far in the coupling reaction, we reasoned that the unique reaction properties would also make thioacid/azide amidation applicable to large and complex compounds such as peptides and proteins. In this project, we have succeeded in making use of the thioacid/azide amidation reaction to tag a variety of groups including tosyl, Dabsyl and Dansyl groups, biotin and the PEG moiety to the C-terminus of a peptide or protein. The experimental results show that the use of the thioacid/azide amidation reaction provides a new method for specific C-terminal modification of peptides and proteins.

4.2 Materials and methods

4.2.1 Materials

All amino acid derivatives and chemical reagents were purchased from commercial suppliers. All solvents used were of analytical or HPLC grade.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.2.2 Instrumentation

4.2.2.1 TLC Chromatography

Thin layer chromatography (TLC) was carried out on aluminum sheets coated with silica [silica 60 F254; layer thickness 0.2 mm (Merk)]. Analytes were visualized under UV light (λ = 254 or 365 nm).

4.2.2.2 HPLC

The HPLC system, columns, buffers and conditions were the same as described in chapter 2.

4.2.2.3 Mass Spectrometry

ESI-MS and MALDI-MS analyses were performed in the same way as described in chapter 2.

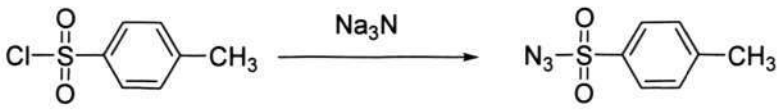
4.2.2.4 NMR Spectroscopy

^{1}H and ^{13}C NMR spectra were recorded on a Bruker DRX-400 (400MHz) NMR spectrometer. ^{13}C spectra were recorded in the broad band decoupled mode. ^{1}H and ^{13}C chemical shifts are quoted in ppm relative to the internal standard tetramethylsilane (TMS) or the residual solvent proton signals.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.2.3 The synthesis of azide compounds

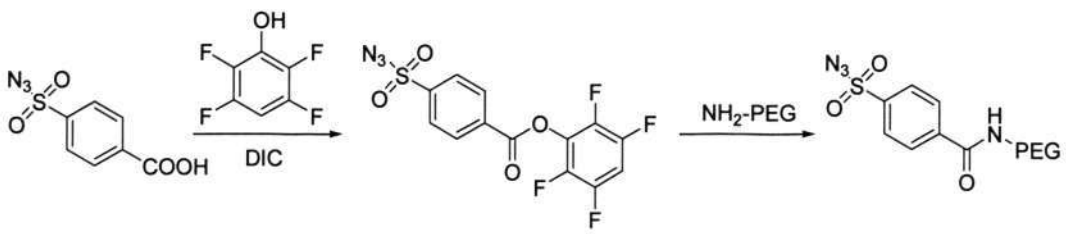
4.2.3.1 Tosyl azide (4.1)



4.1

Tosyl chloride (95 mg, 0.5 mmol) and NaN_3 (65 mg, 1 mmol) were dissolved in acetone and water (1:1) at 0 ℃. The mixture was warmed to room temperature slowly and allowed to stir overnight. Acetone was removed under reduced pressure. The residue was extracted with DCM (30 mL x 3). The combined organic layer was washed with brine, dried under anhydrous sodium sulphate and concentrated under reduced pressure to give a colorless oil, 94 mg, and yield 95%. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.791(d, 2H), 7.378(d, 2H), 2.436 (s, 3H); ^{13}C NMR (400 MHz, $CDCl_3$) δ 146.407, 135.375, 130.345, 127.441, 21.629. (See Appendix 1 for spectum).

4.2.3.2 PEG azide (4.2)



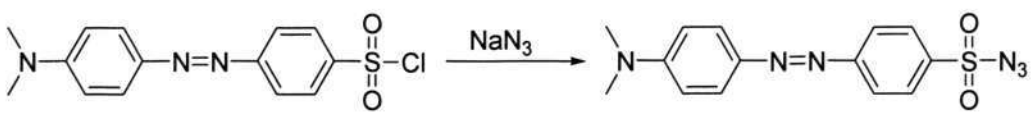
4.2

4-Carboxybenzensulphonylazide (113.6 g, 0.5 mmol), 2,3,4,6-tetrafluorophenol

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

(91.4 mg, 0.55 mmol), DIC (117 μL, 0.75 mmol) were dissolved in dry DCM (2 mL) at 0 ℃, and two drops of TEA were added. The reaction was allowed to occur with stirring at room temperature overnight. H_2O was added to quench the reaction. The mixture was extracted with DCM and the combined DCM extracts were washed with brine, dried and concentrated to afford a light yellow solid 180 mg, yield 96%. 20 mg of this product without any further purification was dissolved in dry DCM (2 mL), and aminopolyethylene glycol 5,000 monomethylether (266 mg, 1 eq.) and DIEA (27 μL) were added to this solution at 0 ℃. The mixture was allowed to stir at room temperature for 2 h. DCM was removed and the residue was washed with cold ether and centrifuged to collect the white precipitate. The crude product was purified by preparative HPLC. MALDI-TOF was used to determine the molecular weight of the product, average *m/z* $[M+H]^+$ found: 5000, average MW calcd: 5000. ^1H NMR (400 MHz, $CDCl_3$) δ 8.41 (d, 2H), 8.02 (d, 2H), 7.75 (s, 1H), 3.36 (s, 651H), 2.03 (s, 33H). (See Appendix 2 for pectrum).

4.2.3.3 Dabsyl azide (4.3)



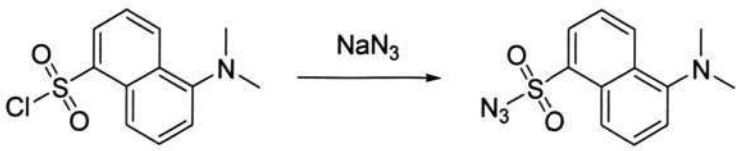
4.3

Dabsyl chloride (0.1 g, 0.31 mmol) and NaN_3 (36 mg, 0.56 mmol) were dissolved in acetone and water (1:1) at 0 ℃. The mixture was warmed to room temperature

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

slowly and allowed to stir overnight. Acetone was removed under reduced pressure. The residue was extracted with DCM (30 mL x 3). The combined organic extracts were washed with brine, dried under anhydrous sodium sulphate and concentrated under reduced pressure to give a yellow-red product, 0.1 g, yield 97%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 7.84-7.98 (m, 6H), 6.68 (d, 2H), 3.07 (s, 6H). ^{13}C NMR (400 MHz, $CDCl_3$) δ 156.88, 153.51, 143.63, 137.28, 128.61, 126.25, 122.97, 111.59, 40.33. (See Appendix 3 for pectrum). ESI-MS, *m/z* $[M+H]^+$ found:331.0, MW calcd: 330.1.

4.2.3.4 Dansyl azide (4.4)



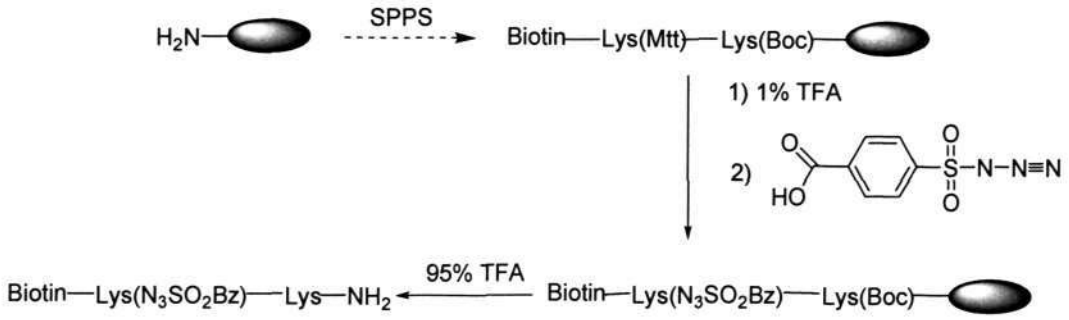
4.4

Dansyl chloride (0.1 g, 0.37 mmol) and NaN_3 (36 mg, 0.56 mmol) were dissolved in an acetone-and-water (1:1) cosolvent system at 0 ℃. The mixture was warmed to room temperature slowly and allowed to stir overnight. Acetone was removed under reduced pressure. The residue was extracted with DCM (30 mL x 3). The combined organic extracts were washed with brine, dried under anhydrous sodium sulphate and concentrated under reduced pressure to give a yellow green oil, 0.1 g, yield 97%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 8.55 (d, 1H, J = 8.4 Hz), 8.22 (d, 1H,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

J = 7.2 Hz), 8.1 (d, 1H, J = 8.8 Hz), 7.52 (t, 1H, J = 8 Hz), 7.47 (t, 1H, J = 8 Hz), 7.12 (d, 1H, J = 7.6 Hz), 2.78 (s, 6H). ^{13}C NMR (400 MHz, $CDCl_3$): δ =152.16, 133.71, 132.72, 130.13, 129.65, 129.28, 122.97, 118.76, 115.87, 45.40. (See Appendix 4 for pectrum). ESI-MS, *m/z* $[M+H]^+$ found: 276.9, MW calcd: 276.1

4.2.3.5 Biotin-Lys(N_3SO_2Bz)-Lys-NH_2 (4.5)



4.5

4 eq. of Fmoc-Lys(Boc)-OH was used to load to the Rink amide resin. The second Lys was coupled using Fmoc-Lys(Mtt)-OH. After Fmoc deprotection, 4 eq. of *D*-biotin was used to couple to the N-terminal amine. The protecting group Mtt was removed by 1% TFA in DCM (1 x 10min, 1 x 20min). Then, 4 eq. of 4-carboxybenzensulphonylazide, PyBop and DIEA were mixed and added to the resin. The reaction was performed at room temperature for 2 h. The biotinyl-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5** was cleaved from the resin in 95% TFA/ 5% H_2O, and purified by RP-HPLC. The molecular weight was determined by ESI-MS. *m/z* $[M+H]^+$ found: 709.3, MW calcd: 708.3.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.2.4 Trypsin substrate Dansyl-DSARAGS-Dabsyl

The peptide Dansyl-**DSA*R*AGS**-COSH was synthesized by Boc-SPPS on $HSCH_2CH_2CO$-NH_2CH_2-CM resin through a thioester linkage. After deprotection of the Boc group from Asp, 5-(dimethylamino)naphthalene-1-sulfonyl chloride (Dansyl chloride) was coupled onto the resin-bound peptide. 2 eq. of Dansyl chloride were used in the coupling reaction. Double coupling was performed when needed. Before cleavage by HS^-, all the protection groups were removed by using a cocktail of TFMSA:thioanisole:TFA/1:2.5:10 at room temperature for 1 hr [76]. The deprotected peptide resin was then cleaved by 0.2 M $(NH_4)_2S$ in 0.3 M HEPES buffer, pH 8.5. The reaction was carried out at room temperature for 3-4 h, and the cleavage solution was purified by semi-preparative RP-HPLC and lyophilized.

0.8 mg Dansyl-**DSA*R*AGS**-COSH was dissolved in 600 µl of 90% ACN/H_2O containing 1 mg of 2,6-lutidine. 3 mg of 4-((4-(dimethylamino)phenyl)diazenyl)benzene-1-sulfonyl azide (Dabsyl azide) **4.3** was added into the solution. The Dabsyl azide was not completely dissolved initially; however, the solution was getting clearer as the reaction progressed. This reaction took 24 h at room temperature to complete.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.2.5 Enzymatic assay on the trypsin substrate Dansyl-DSARAGS-Dabsyl

A stock solution of Dansyl-**DSA*R*AGS**-Dabsyl was prepared in ACN/H_2O 1:1, and stored at 4 ℃. Trypsin cleavage was performed in 50 mM sodium phosphate buffer, pH 8.0. The final concentrations of substrate and trypsin were 50 μm and 0.1 μM, respectively. Fluorescence monitoring of tryptic hydrolysis of Dansyl-**DSA*R*AGS**-Dabsyl peptide was performed on a fluorescence spectrophotometer (Cary Eclipse, Varian) using an excitation wavelength of 245 nm and an emission wavelength of 560 nm. Analytical RP-HPLC was also used to monitor the trypsin digestion of this substrate.

4.2.6 thioacid/azide amidation for peptide or protein C-terminal labeling

4.2.6.1 Peptide C-terminal tagging

Peptide thioacids, H_2N-**RLLLPGELA**-COSH and H_2N-**APKRYKANY**-COSH (see Table 4.1; see also section 2.2.3 in Chapter 2 for preparation) were used for demonstrating C-terminal tagging by using the thioacid/azide reaction. All azide compounds were 5-10 molar eq. in excess of peptide thioacids. The reaction was conducted in 90% ACN/H_2O or MeOH/H_2O containing 2,6-lutidine in the same molar amount as the azide [66]. The final concentration of peptide thioacids were ~0.2 mM. Analytical RP-HPLC was used to monitor the reactions.

The reaction with PEG azide **4.2** was conducted in aqueous buffer without

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

2,6-lutidine: 20 mM HEPES, pH 7.5.

4.2.6.2 Protein C-terminal tagging

For demonstrating the C-terminal labeling of proteins by a biophysical or other functional group, ubiquitin thioacid (see section 2.3.2.2 in Chapter 2 for preparation) was subjected to the amidation reaction with either the biotin-azide **4.5** or PEG azide **4.2** compound prepared earlier. For reaction with the PEG azide, 0.3 mM ubiquitin thioacid, 5-6 eq. of PEG azide **4.2** and 7-8 eq. of 2,6-lutidine were mixed in 6 M Gdn-HCl. For reaction with biotin-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5,** 0.3 mM ubiquitin thioacid, 5-10 eq. of biotin-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5** and 3 mM 2,6-lutidine were dissolved in 6 M Gdn-HCl. The reactions were conducted at room temperature for 12 h. Analytical RP-HPLC was used to monitor the reactions.

4.3 Results and discussion

4.3.1 Peptide and protein C-terminal labeling by thioacid/azide coupling

One application of the thioacid/azide coupling reaction in protein chemistry is to conjugate a special biophysical or other functional group to the C-terminus of a peptide or protein.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

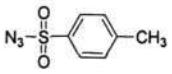
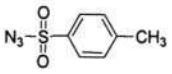
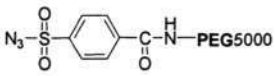
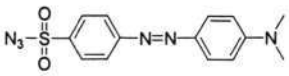
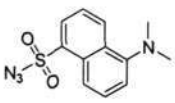
The thioacid/azide amidation reaction first proceeds through a prior capture step leading to the formation of a thiatriazoline intermediate, and the intermediate then decomposes to give the amide product with nitrosulfur as byproducts. Although thioacid/azide coupling has only been demonstrated with small thioacids such as thioacetic acid and N-protected amino thioacid compounds, its unique reaction features imply that it should have the required chemical orthorgonal properties to work in a chemoselective manner with large and complex thioacid compounds such as peptide and protein thioacids. To demonstrate this application, three peptide thioacids were subjected to thioacid/azide coupling with several kinds of azide compounds, with MW ranging from 197 to >5000. All azide compounds used in these reactions are electron-deficient sulfonazides, since amidations of these azides are much faster than those of electron-rich azides [66].

Table 4.1 shows examples of the reaction of peptide thioacids with the different azide compounds. Peptide thioacids H_2N-**RLLLPGELA**-COSH and H_2N-**APKRYKANY**-COSH were both used to react with tosyl azide **4.1**. HPLC analyses showed clean and near quantitative conversions (>95%) of the peptide thioacids to the desired *N*-peptidyl sulfonamide products after 2 h at room temperature (Figure 4.1 and 4.2). Similarly, a high yield (~90%) in the amidation of the peptide thioacid, H_2N-**APKRYKANY**-COSH, with Dansyl azide **4.4** was also obtained (Figure 4.3). In spite of the presence of some minor impurity in the original Dansyl chloride, no impact on the amidation reaction was observed. All

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

amidation reactions were quite efficient and rapid, and were completed within a few hours. The only exception was the reaction with Dabsyl azide **4.3**, which was due to the low solubility of Dabsyl azide **4.3,** even in organic solvents (see section 4.3.2).

Table 4.1 Peptide C-terminal tagging with different biofunctional groups through thioacid/azide amidation (Yields are estimated based on HPLC analysis).

No.	Peptide thioacid	Tag	Yield (%)	Calcd MW	Obsd MS *m/z* $[M+H]^+$
1	H_2N-**RLLLPGELA**-COSH	 **4.1**	95 (2 h)	1135.8	1134.3
2	H_2N-**APKRYKANY**-COSH	 **4.1**	95 (2 h)	1263.2	1263.5
3	H_2N-**APKRYKANY**-COSH	 **4.2**	80 (5 h)	~6300	av. 6300
4	Dansyl-**DSARAGS**-COSH	 **4.3**	80 (24 h)	1181.4	1182.7
5	H_2N-**APKRYKANY**-COSH	 **4.4**	90 (5 h)	1341.7	1342.7

As shown in Figure 4.1-4.3, these reactions occurred with high chemoselectivity despite the presence of a number of side-chain functional groups and a free N-terminal amine in these peptides. The thioacid/azide amidations between different peptide thioacids and azide compounds are all of high yield and clean,

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

and no significant by-product indicated by RP-HPLC analysis.

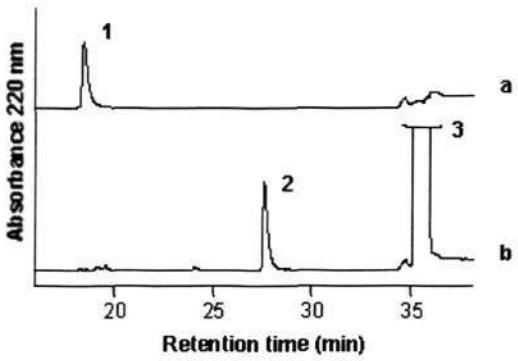


Figure 4.1 HPLC monitoring of the reaction of peptide thioacid, H_2N-**APKRYKANY**-COSH, with tosyl azide **4.1**. Trace **a**, the peptide thioacid. Peak 1, *m/z* $[M+H]^+$ found:1127.0, MW calcd: 1125.6. Trace **b**, reaction at 2 h. Peak 2, the reaction product H_2N-**APKRYKANY**-CONH-SO_2-Ph-CH_3. *m/z* $[M+H]^+$ found:1263.5, MW calcd: 1263.2. Peak 3, tosyl azide **4.1** in excess. HPLC gradient: 0% to 70% of buffer B in buffer A for 35 min.

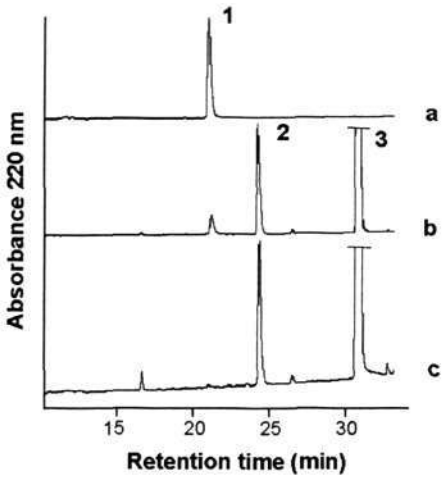


Figure 4.2 HPLC monitoring of the reaction of peptide thioacid, H_2N-**RLLLPGELA**-COSH, with tosyl azide **4.1**. Trace **a**, the thioacid peptide. Peak 1, *m/z* $[M+H]^+$ found: 997.7; MW calcd: 996.6. Trace **b**, reaction mixture at 1 h. Peak 2, the reaction product H_2N-**RLLLPGELA**-CONH-SO_2-Ph-CH_3, *m/z* $[M+H]^+$found: 1135.8, MW calcd:1133.6. Peak 3, tosyl azide **4.1** in excess. Trace **c**, reaction mixture at 2 h. HPLC gradient: 0% to 80% of buffer B in buffer A for 40 min.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

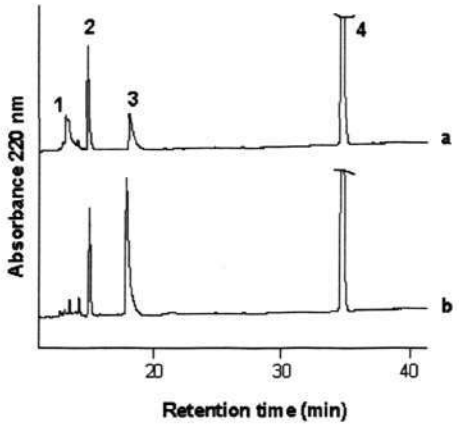


Figure 4.3 Analytical RP-HPLC profile of the reaction of peptide thioacid, H_2N-**APKRYKANY**-COSH, with Dansyl azide **4.4**. Trace **a**, reaction at 1 h. Peak 1, the peptide thioacid. *m/z* $[M+H]^+$ found:1127.0, MW calcd: 1125.6. Peak 2, the decomposed product of Dansyl azide **4.4**. *m/z* $[M+H]^+$ found: 251.2, MW calcd: 250.3. Peak 3, the reaction product H_2N-**APKRYKANY**-Dansyl. *m/z* $[M+H]^+$ found:1342.7, MW calcd: 1341.7. Peak 4, Dansyl azide **4.4.** *m/z* $[M+H]^+$ found: 276.9, MW calcd: 276.1. Trace **b**, reaction at 5 h was almost completed. HPLC gradient: 0-80% of buffer B in buffer A in 40 min.

PEG azide **4.2** was synthesized to test the application of the thioacid/azide amidation using large and complex azides. PEG has been used as a biofunctional group since 1977 [124]. Covalent attachment of PEG to therapeutic proteins often increases their *in vivo* stability and solubility [125]. Traditional PEGylation methods are non-specific and introduce PEG nonspecically onto nucleophilic amino acids in proteins, giving a mixture of PEGylated isomers [126]. In contract to this nonspecific PEGylation, the thioacid/azide amidation reaction provides a new method to site-specifically attach PEG to peptides or proteins because of its unique reaction mechanism. The amidation reaction of the peptide thioacid H-**APKRYKANY**-COSH with the PEG azide **4.2** was performed in an aqueous buffer: 20 mM HEPES, pH 7.5, instead of an organic-water cosolvent, because

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

PEG azide **4.2** is highly water-soluble. A relatively low peptide thioacid concentration (~0.2 mM) was used. After 5 h reaction, about 80% of the amidation product was obtained. As shown in Figure 4.4 and 4.5, there was only one product formed in the reaction, indicating that coupling of PEG to the peptide is highly chemoselective despite the presence of other side-chain nucleophilic groups in the peptide. The MS results also confirmed the correct amidation product. The slightly lower yield and slower reaction rate may be due to steric hindrance in the PEG azide compound **4.2** in which the relatively hydrophobic sulfonylazide may be made less accessible by the surrounding large hydrophilic PEG chain.

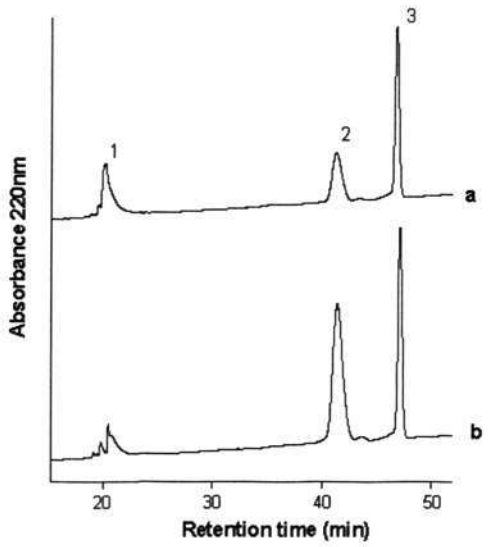


Figure 4.4 HPLC monitoring of the reaction of peptide thioacid, H_2N-**APKRYKANY**-COSH, with PEG azide **4.2**. Trace **a**, reaction at 1 h. Peak 1, the peptide thioacid. *m/z* $[M+H]^+$ found: 1127.0, MW calcd: 1125.6; peak 2, reaction product H_2N-**APKRYKANY**-CONH-PEG; peak 3, PEG-azide **4.2**. Trace **b**, reaction at 3 h, showing almost complete disappearance of the peptide thioacid. HPLC gradient: 0-50% of buffer B in buffer A in 50 min.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

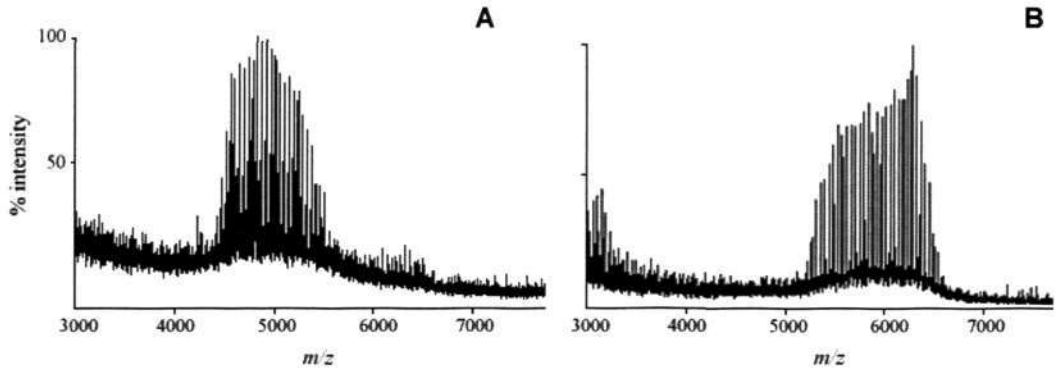


Figure 4.5 Characterization of PEG azide and its amidation product with peptide thioacid, H_2N-**APKRYKANY**-COSH, by MALDI-TOF. (**A**) Molecular weight of PEG azide **4.2**, average *m/z* $[M+H]^+$ found 5000, average MW calcd: 5000. (**B**) Molecular weight of the amidation product, H_2N-**APKRYKANY**-CONH-PEG, average *m/z* $[M+H]^+$ found: ~6000, average MW calcd: ~6100.

A protein thioacid, ubiquitin thioacid, obtained by hydrothiolysis of the biosynthetic intermediate of protein splicing thioester ubiquitin-intein-CBD (see section 2.2.5 in chapter 2 for preparation), was used to perform a protein thioacid/azide amidation reaction (Figure 4.6 and 4.7). Compared to amidation between peptide thioacids and azides, the reaction efficiency was lower. For the reaction with PEG azide in 6 M Gdn-HCl containing 2,6-lutidine, it took almost 30 hours to complete the reaction, which was accompanied by some hydrolysis. When the reaction was conducted in aqueous buffer, pH 7-8, without using 2,6-lutidine, both the amidation and the hydrolysis took place immediately after the protein thioacid had been mixed with the azide. When 2,6-lutidine was used instead of a weakly basic aqueous buffer, the hydrolysis did not occur at the beginning of the amidation, but still some protein thioacid was hydrolyzed as the reaction proceeded.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

However, based on HPLC analysis, the yield of the amidation was about 70%, which is quite satisfactory. The long reaction time and some degree of hydrolysis may be caused by the steric hindrance of the PEG azide **4.2,** making it less accessible to the ubiquitin thioacid and the C-terminal Gly residue of ubiquitin–a small amino acid-may be more easily hydrolyzed.

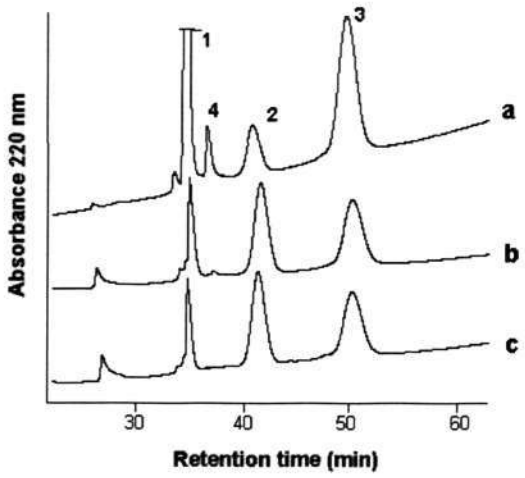


Figure 4.6 C4 analytical RP-HPLC monitoring of the amidation of ubiquitin thioacid with the PEG azide. The reaction was conducted in 6 M Gdn-HCl containing 0.3 mM ubiquitin thioacid, 2 mM PEG azide and 3 mM 2,6-lutidine. Trace **a**, the beginning of the reaction. Peak 1, ubiquitin thioacid; peak 2, the amidation product; peak 3, PEG azide **4.2**; peak 4, an unidentified component which disapeared as the reaction progressed. Trace **b**, reaction at 12 h. Trace **c**, reaction at 36 hours. HPLC gradient: 0 to 60% of buffer B in buffer A in 60 min.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

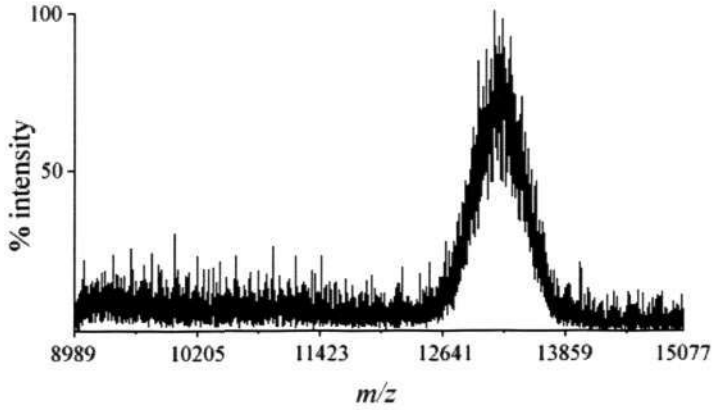


Figure 4.7 Characterization of the amidation product of ubiquitin thioacid with PEG azide **4.2** by MALDI-TOF. The amidation product ubiquitin-CONH-PEG, average *m/z* $[M+H]^+$ found: ~13200, average MW calcd: ~13500);.

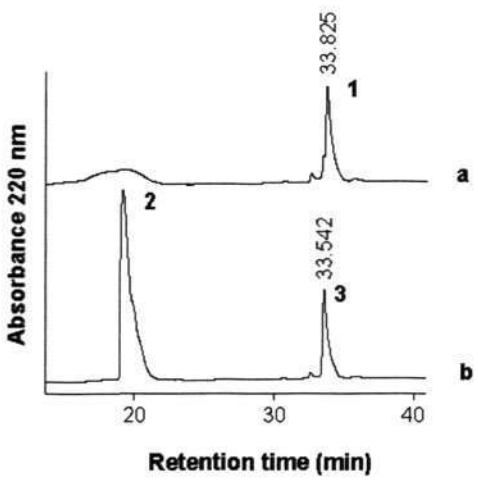


Figure 4.8 C4 analytical RP-HPLC monitoring of the amidation of ubiquitin thioacid with biotinyl-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5**. The reaction was conducted in 6 M Gdn-HCl containing 0.3 mM ubiquitin thioacid, 5-10 eq. of biotinyl-Lys(N_3SO_2Bz)-Lys-NH_2 and 3 mM 2,6-lutidine. Trace **a**, peak 1, ubiquitin thioacid. Trace **b**, the reaction at 12 h. Peak 2, biotinyl-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5**; peak 3, the amidation product, ubiquitin-CONH-biotinyl-Lys(SO_2Bz)-Lys-NH_2, containing some minor amount of hydrolyzed ubiquitin. HPLC gradient: 0-40% of buffer B in buffer A in 40 min.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

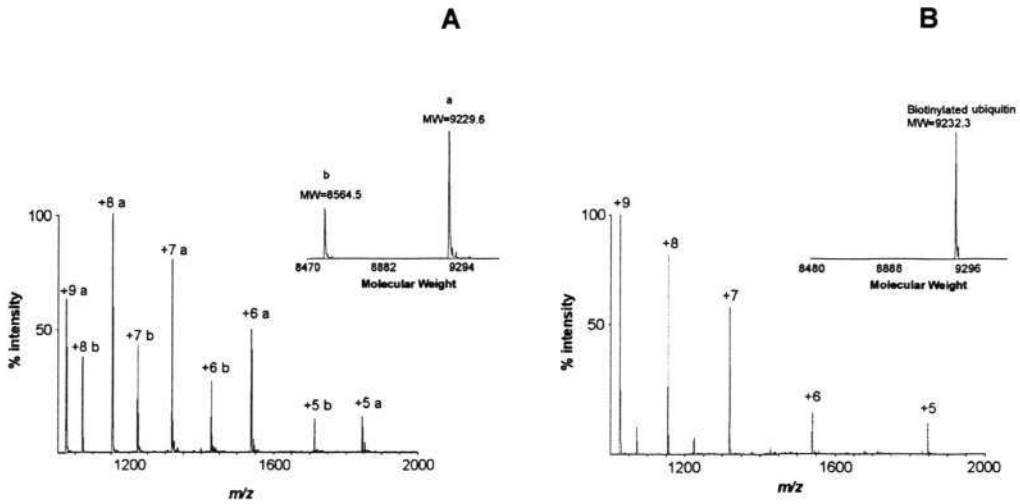


Figure 4.9 ESI mass spectra and deconvoluted MS profile of the amidation of biotinyl-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5** with ubiquitin thioacid. (**A**) The first quarter of peak 3 in Figure 4.8. **a**, the amidation product; **b**, the hydrolyzed ubiquitin. (**B**) The last 3 quarters of peak 3 in Figure 4.8. The amidation product, ubiquitin-CONH-biotinyl-Lys(SO_2Bz)-Lys-NH_2, MW found: 9230.9 (as shown in **A-b** and **B**), MW calcd: 9228.5. Hydrolyzed ubiquitin: MW found: 8564.5 (as shown in **A-a**), MW calcd: 8564.8.

The ubiquitin thioacid was also used to couple with the biotin-azide **4.5** which is much smaller than the PEG azide **4.2**. In biochemistry, it is very common to covalently link or tag biotin to a biomolecule such as a protein for use in biochemical assays. Such processes refered to as biotinylation modify primary amines, sulfhydryl, carboxyl groups *etc.* (http://en.wikipedia.org/wiki/Biotinylation). However, similar to PEGylation, all of these are non site-specific tagging. In this study, based on the mechanism of the thioacid/azide amidation reaction, we have attached a biotin compound site-specifically to ubiquitin. Biotinyl-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5** was designed

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

for this reaction. In this compound, a biotin and a tosyl azide were linked to the α-amine and ε-amine on one Lys residue, respectively. The other Lys residue is introduced for increasing the solubility of the biotinylation reagent. In the coupling experiment, this biotin azide **4.5** was successfully tagged to ubiquitin's C-terminus through the thioacid/azide amidation reaction. The correct amidation product was confirmed by ESI-MS (MW found: 9230.9, MW calcd: 9228.5, as shown in Figure 4.9). Compared to PEG azide **4.2**, the smaller molecule Biotinyl-Lys(N_3SO_2Bz)-Lys-NH_2 **4.5** reacted with ubiquitin thioacid much faster, although hydrolysis of the ubiquitin thioacid still occurred to a minor extent during amidation (as shown in Figure 4.8 and 4.9). According to the MS results, the amidation yield is estimated to be ~80%. Notably, the amidation of ubiquitin with both the PEG and the biotin azide compounds was conducted under denatured conditions, which is highly desirable for hydrophobic peptides and proteins.

Taken together, it is seen that the peptide and protein thioacids used in this study contain a diverse set of C-terminal residues, from the simplest Gly to the aromatic Tyr. All the amidation reactions performed with these different thioacids were highly efficient and specific, indicating that this reaction is largely independent of the nature of the peptide's or protein's C-terminal residue. All the azide compounds used in this study were electron-deficient, because, as mentioned by Kolakowski, the amidation rate is much slower for electron-rich azides due to the large amount of entropy cost for activation. In fact, in trying to apply the amidation reaction to

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

peptide ligation, we synthesized a peptide containing an N-terminal azidoacetyl group. This N_3–peptide was then reacted with a peptide thioacid. Unfortunately, the amidation did not occur even after overnight reaction, indicating that this azide group is not electron-deficient enough. This result is in accordance with the mechanism of amidation proposed by Kolakowski, *et al* [66].

It is noted that all reactions with the highly electron-deficient azides were conducted in diluted solutions of peptide or protein thioacids (0.2-0.3 mM). The high efficiency in such dilute solutions can only be explained by the thioacid/azide amidation mechanism of the prior capture type proposed by Williams and coworkers [66].

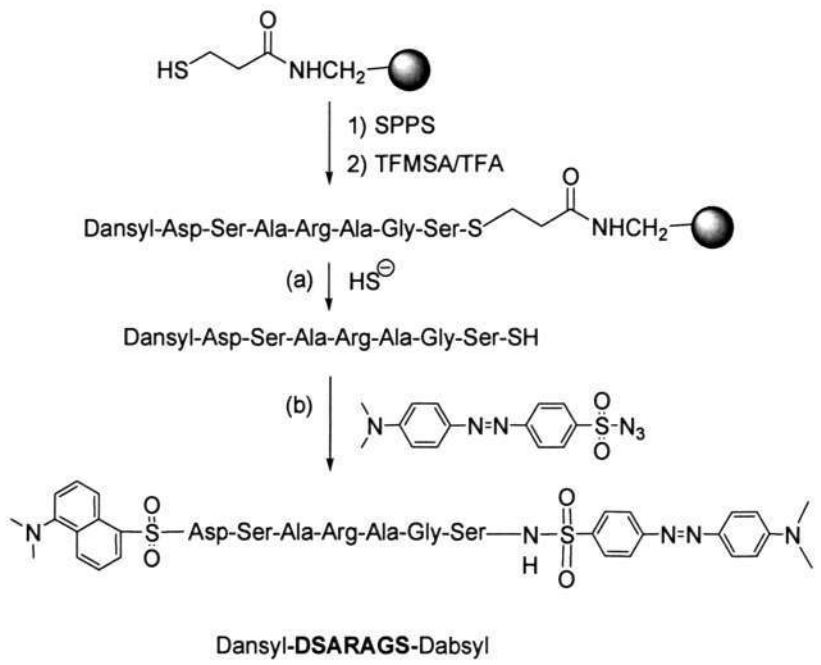
4.3.2 The application of thioacid/azide reaction in the synthesis of an internally quenched fluorogenic trypsin substrate

We also used the thioacid/azide coupling reaction to synthesize a FRET (Fluorescence resonance energy transfer) substrate of trypsin, which contains the fluorescent Dansyl group at the N-terminus and the fluorescent quenching group Dabsyl at the C-terminus (Scheme 4.2)

The peptide thioacid Dansyl-**DSA*R*AGS**-COSH was synthesized by SPPS on CM resin, and released from the solid phase by hydrothiolysis. The peptide thioacid was then subjected to the amidation reaction with

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4-((4-(dimethylamino)phenyl)diazenyl)benzene-1-sulfonyl azide (Dabsyl azide **4.3**) in wet methonal in the presence of 2,6-lutidine to give an internally quenced fluorogenic trypsin substrate Dansyl-**DSA*R*AGS**-Dabsyl. Since the Dabsyl contains a double bond which is fagile in peptide synthesis, we chose to introduce the Dansyl group at the N-terminus during SPPS, while the Dabsyl group was later attached at the C-terminus of Dansyl-**DSA*R*AGS**-COSH by using the thioacid/azide amidation reaction.



Scheme 4.2 The synthesis of trypsin substrate Dansyl-**DSA*R*AGS**-Dabsyl

HPLC analysis showed that about 80% of the desired product was formed after 24 hours reaction (see Figure 4.10). The reason for the relatively slow reaction compared to the reactions with other azide compounds (which require normally 2-4 hour reaction time) may be due to the very low solubility of the Dabsyl azide

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

4.3 even in the organic solvents used for the reaction.

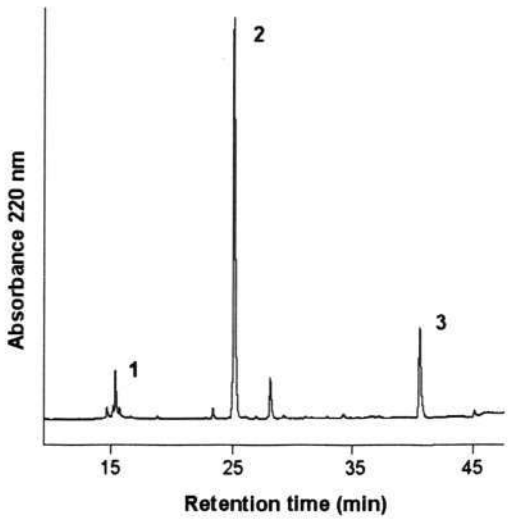


Figure 4.10 Analytical HPLC profile of the amidation of Dansyl-**DSA*R*AGS**-COSH with Dabsyl azide **4.3** at room temperature for 24 hours. Peak 1, Dansyl-**DSA*R*AGS**-COSH left after 24 hour reaction; peak 2, the amidation product, Dansyl-**DSA*R*AGS**-Dabsyl, *m/z* $[M+H]^+$ found:1182.4, MW calcd: 1181.6; peak 3, Dabsyl azide **4.3**. HPLC condition: 0% to 80% of buffer B in buffer A over 40 min.

The FRET peptide, Dansyl-**DSA*R*AGS**-Dabsyl was then used in an enzymatic assay to test its hydrolysis activity. Fluorescence resonance energy transfer (FRET) was first introduced for the detection of proteolytic activity of carboxypeptidase A by Latt *et al.* [127]. The excitation energy from an excited fluorescent donor is transferred to a quenching acceptor. 5-(dimethylamino)naphthalene-1-sulfonyl (Dansyl) and 4-((4-(dimethylamino)phenyl)diazenyl)benzene-1-sulfonyl (Dabsyl) are a donor/acceptor pair with the required spectral overlap permitting FRET to occur [128]. Labeling with Dabsyl at the C-terminus of Dansyl-**DSA*R*AGS**-COSH

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

through thioacid/azide reaction makes it an internally quenched fluorogenic substrate which can be used for the detection of proteolytic activity by recording the increase in fluorosence intensity. Figure 4.11 shows the change in fluoroscence emission intensity in the course of the trypsin digestion reaction of Dansyl-**DSA*R*AGS**-Dabsyl, reaching a steady state after 300 sec of digestion. Because the cleavage of the substrate by trypsin occurred very fast, the recording of the fluorosence intensity was impossible to start from zero. The results shown in this experiment indicates a new way to prepare the FRET substrates in which the quencher is introduced onto the C-terminus, whereas in the traditional methods such groups are introduced on the side chain of an amino acid.

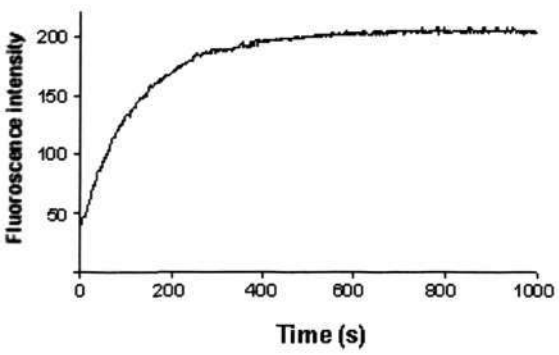


Figure 4.11 Trypsin digestion reaction of Dansyl-**DSA*R*AGS**-Dabsyl as monitored by the change in fluorescence intensity. Assay conditions: 50 μM peptide substrate, 0.1 μM trypsin in 50 mM phosphate buffer, pH 8.0. λ_{ex}= 245 nm, λ_{em}= 560 nm.

We also used RP-HPLC to monitor the trypsin-catalyzed cleavage of the Dansyl-**DSA*R*AGS**-Dabsyl substrate (see Figure 4.12). The result showed the

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

complete digestion of the Dansyl-**DSA*R*AGS**-Dabsyl by trypsin. Taken together, the thioacid/azide amidation reaction is a practical method for the synthesis of internally quenched fluorogenic peptide substrates.

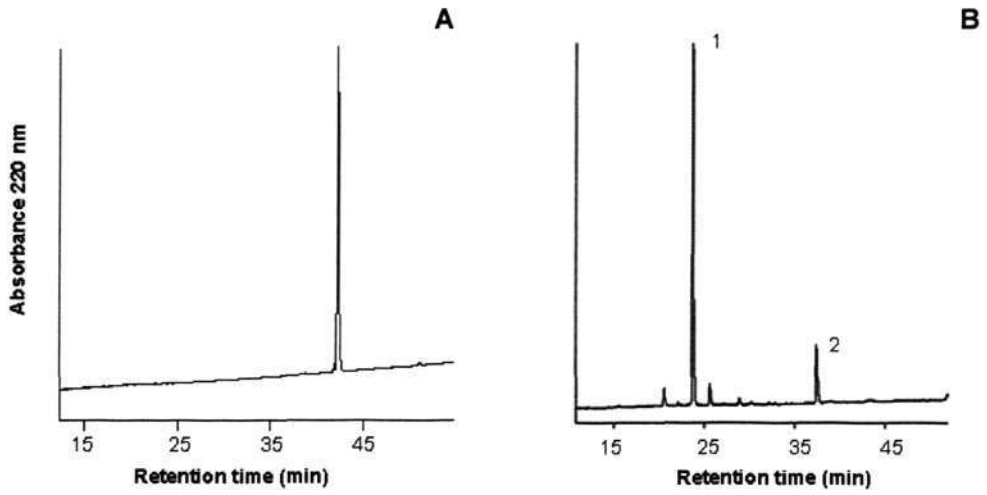


Figure 4.12 HPLC monitoring of the trypsin digestion reaction of Dansyl-**DSA*R*AGS**-Dabsyl. (**A**) Analytical RP-HPLC analysis of Dansyl-**DSA*R*AGS**-Dabsyl before trypsin digestion. (**B**) Analytical RP-HPLC profile of the trypsin digestion after 1h. Cleavage assay conditions: 50 μM peptide substrate, 0.1 μM trypsin in 50 mM phosphate buffer, pH 8.0. HPLC condition: 0% to 60% of buffer B in buffer A over 60 min. Peak 1, Dansyl-**DSAR**-OH; peak 2, H-**AGS**-Dabsyl.

4.4 Conclusion

In this study, the thioacid/azide amidation reaction was applied to couple different azide compounds with peptide and protein thioacids. All azide compounds were electron-deficient sulfonylazides that contain a tosyl, Dansyl or Dabsyl group, or a biotin or PEG moiety. The amidation reactions were fast and efficient, with the exception of Dabsyl azide with which the amidation was much slower, probably

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

due to its low solubility. The high efficiency in the amidation reactions with these peptides and proteins containing different C-terminal residues suggests that the amidation reaction is independent of the nature of C-terminal residue. HPLC and MS analyses of these amidation reactions showed that there were no significant by-products formed, indicating that the amidation reaction is highly chemoselective despite the presence of a large number of side-chain functional groups and a free N-terminal amine in these peptides and proteins. Moreover, the fact that these reactions were conducted in relatively dilute solution is in agreement with the prior capture mechanism proposed by Williams and coworkers.

Using the amidation reaction, we synthesized an internally quenched fluorogenic peptide trypsin substrate Dansyl-**DSA*R*AGS**-Dabsyl. The proteolytic experiment showed that the bioactivity of this substrate was retained, indicating that thioacid/azide amidation had has no adverse effect on the recognition and binding of the substrate by the enzyme.

In all, the experiments demonstrate that thioacid/azide amidation is a simple and efficient method for peptide or protein C-terminal tagging, capable of providing key reagents and materials for biophysical and biochemical research and for drug discovery.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

General Conclusion

A novel and efficient method to prepare peptide thioacids by solid-phase synthesis through hydrothiolysis of resin-bound peptide thioesters was developed in this study. Peptide thioacids are an important class of compounds with increasingly recognized value as synthetic intermediates. Different from the traditional approach of using the benzhydryl linker chemistry originally developed by Blake *et al.*, peptide thioesters were first synthesized by Boc or Fmoc SPPS on a ChemMatrix resin which has good swelling ability in water, and hydrothiolysis was then conducted in aqueous buffer resulting in the formation of peptide thioacids. We prepared several peptide thioacids with different amino acid residues at the C-terminus. The reaction can also be performed in an aqueous buffer containing a denaturing reagent, such as 6 M Gdn-HCl, to prevent aggregation. This study demonstrates that the hydrothiolysis reaction on resin-bound peptide thioesters is a simple and economic route for peptide thioacid synthesis. We also showed that the hydrothiolysis reaction can be applied to the preparation of protein thioacids from the thioester intermediates of protein splicing.

The truncated histone H3 protein fragment 1-109 was biosynthesized as a fusion protein using an intein expression system. The truncated H3 thioacid was obtained by hydrothiolysis of the fused protein thioester in a hydrogen sulfide solution.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Synthetic peptide and protein thioacids serve as essential building blocks for thioacid capture ligation. The thioacid peptides and proteins prepared by hydrothiolysis were used for ligation with their respective ligation partners which had an N-terminal H-Cys(Npys) residue to give several H3 analogs containing an acetylated or methylated Lys residue in either the N-terminal or C-terminal segment. These semi-synthetic H3 proteins were incorporated into histone octamers which in turn were assembled to nucleosomes with a double stranded DNA. Our work demonstrates that thioacid capture ligation is a viable protein synthesis method and that synthetic proteins prepared by such a ligation chemistry are invaluable reagents for the study of the most basic problems in biology, such as the regulatory effects of histone posttranslational modification on chromatin structure and function.

The importance of the thioacid functionality is further seen in its unique reactivity with certain organic azides. We conducted the thioacid/azide amidation reaction between several thioacids and compounds containing an electron-deficient sulfonyl azide. All reactions gave good yields, although the reaction times were different, depending on the solubility of the azide compounds and their compatibility with the solvents used. Using the amidation of Dansyl-**DSA*R*AGS**-SH with Dabsyl azide, we prepared an internally quenched fluorogenic trypsin substrate with good FRET activity. To demonstrate the applicability of this reaction in site-specific protein modification, we prepared ubiquitin thioacid and labeled the protein at its

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

C-terminus with a biotin or PEG moiety through thioacid/azide coupling. The results we obtained provide strong evidence to support the utility of this previously unexplored reaction in bioconjugation chemistry.

In conclusion, the work presented here demonstrates that peptide and protein thioacids, a class of compounds previously thought to be difficult to synthesize, can be conveniently prepared by a very simple hydrothiolysis reaction. Furthermore, we showed that the unique property of the thioacid functionality can make it uniquely useful in peptide and protein chemistry, as demonstrated by the successful application of thioacid capture ligation in the synthesis of H3 analogs and of thioacid/azide coupling in peptide/protein C-terminal modification.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

References

1. Smith M: **Synthetic DNA and Biology**. *Angew. Chem. Intl. Ed* 1994, **33**: 1214-1221.

2. Dawson PE, Kent SB: **Synthesis of native proteins by chemical ligation**. *Annu Rev Biochem* 2000, **69**:923-960.

3. Coltart DM: **Peptide Segment Coupling by Prior Ligation and Proximity-Induced Intramolecular Acyl Transfer**. *Tetrahedron Lett.* 2000, **56**:3449-3491.

4. Chapeville F, Lipmann F, Ehrenstein GV, Weisblum B, Ray WJ, Benzer S: **On the Role of Soluble Ribonucleic Acid in Coding for Amino Acids**. *Proc. Natl. Acad. Sci. USA* 1962, **48**:1086-1092.

5. Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG: **A general method for site-specific incorporation of unnatural amino acids into proteins**. *Science* 1989, **244**:182-188.

6. Tam JP, Yu Q, Miao Z: **Orthogonal ligation strategies for peptide and protein**. *Biopolymers* 1999, **51**:311-332.

7. Ellman JA, Mendel D, Schultz PG: **Site-specific incorporation of novel backbone structures into proteins**. *Science* 1992, **255**:197-200.

8. Wang L, Schultz PG: **Expanding the Genetic Code**. *Angew. Chem. Intl. Ed* 2005, **44**:34-66.

9. Mendel D, Cornish VW, Schultz PG: **Site-directed mutagenesis with an expanded genetic code**. *Annu Rev Biophys Biomol Struct* 1995, **24**:435-462.

10. Kwok Y, Wong JT: **Evolutionary relationship between Halobacterium cutirubrum and eukaryotes determined by use of aminoacyl-tRNA synthetases as phylogenetic probes**. *Can J Biochem* 1980, **58**:213-218.

11. Kleina LG, Masson JM, Normanly J, Abelson J, Miller JH: **Construction of Escherichia coli amber suppressor tRNA genes. II. Synthesis of additional tRNA genes and improvement of suppressor efficiency**. *J Mol Biol* 1990, **213**:705-717.

12. Cload ST, Liu DR, Froland WA, Schultz PG: **Development of improved tRNAs for in vitro biosynthesis of proteins containing unnatural amino acids**. *Chem Biol* 1996, **3**:1033-1038.

13. Hohsaka T, Ashizuka Y, Murakami H, Sisido M: **Incorporation of Nonnatural Amino Acids into Streptavidin through In Vitro Frame-Shift Suppression**. *J Am Chem Soc* 1996, **118**:9778-9779.

14. Hohsaka T, Kajihara D, Ashizuka Y, Murakami H, Sisido M: **Efficient Incorporation of Nonnatural Amino Acids with Large Aromatic Groups into Streptavidin in In Vitro Protein Synthesizing Systems**. *J Am Chem Soc* 1999, **121**:34-40.

15. Hohsaka T, Ashizuka Y, Taira H, Murakami H, Sisido M: **Incorporation of nonnatural amino acids into**

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

proteins by using various four-base codons in an Escherichia coli in vitro translation system. *Biochemistry* 2001, **40**:11060-11064.

16. Hecht SM, Alford BL, Kuroda Y, Kitano S: **"Chemical aminoacylation" of tRNA's**. *J Biol Chem* 1978, **253**:4517-4520.

17. Robertson SA, Ellman JA, Schultz PG: **A general and efficient route for chemical aminoacylation of transfer RNAs**. *J Am Chem Soc* 1991, **113**:2722-2729.

18. Patchornik A, Amit B, Woodward RB: **Photosensitive protecting groups**. *J Am Chem Soc* 1970, **92**:6333-6335.

19. Adams SR, Kao JPY, Tsien RY: **Biologically useful chelators that take up calcium(2+) upon illumination**. *J Am Chem Soc* 1989, **111**:7957-7968.

20. Lodder M, Wang B, Hecht SM: **The N-pentenoyl protecting group for aminoacyl-tRNAs**. *Methods* 2005, **36**:245-251.

21. Gilbert W: **Origin of life:The RNA world**. *Nature* 1986, **319**:618.

22. Nissen P, Hansen J, Ban N, Moore PB, Steitz TA: **The structural basis of ribosome activity in peptide bond synthesis**. *Science* 2000, **289**:920-930.

23. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF: **Crystal structure of the ribosome at 5.5 A resolution**. *Science* 2001, **292**:883-896.

24. Illangasekare M, Sanchez G, Nickles T, Yarus M: **Aminoacyl-RNA synthesis catalyzed by an RNA**. *Science* 1995, **267**:643-647.

25. Lee N, Bessho Y, Wei K, Szostak JW, Suga H: **Ribozyme-catalyzed tRNA aminoacylation**. *Nat Struct Biol* 2000, **7**:28-33.

26. Saito H, Kourouklis D, Suga H: **An in vitro evolved precursor tRNA with aminoacylation activity**. *EMBO J* 2001, **20**:1797-1806.

27. Saito H, Suga H: **A Ribozyme Exclusively Aminoacylates the 3'-Hydroxyl Group of the tRNA Terminal Adenosine**. *J Am Chem Soc* 2001, **123**:7178-7179.

28. Murakami H, Ohta A, Ashigai H, Suga H: **A highly flexible tRNA acylation method for non-natural polypeptide synthesis**. *Nat Methods* 2006, **3**:357-359.

29. Anderson JC, Schultz PG: **Adaptation of an orthogonal archaeal leucyl-tRNA and synthetase pair for four-base, amber, and opal suppression**. *Biochemistry* 2003, **42**:9598-9608.

30. Edwards H, Schimmel P: **A bacterial amber suppressor in Saccharomyces cerevisiae is selectively recognized by a bacterial aminoacyl-tRNA synthetase**. *Mol Cell Biol* 1990, **10**:1633-1641.

31. Sakamoto K, Hayashi A, Sakamoto A, Kiga D, Nakayama H, Soma A, Kobayashi T, Kitabatake M, Takio

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

K, Saito K, et al.: **Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells**. *Nucleic Acids Res* 2002, **30**:4692-4699.

32. Wang L, Xie J, Schultz PG: **Expanding the genetic code**. *Annu Rev Biophys Biomol Struct* 2006, **35**:225-249.

33. Merrifield RB: **Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide**. *J Am Chem Soc* 1963, **85**:2149-2154.

34. Blake J, Li CH: **New segment-coupling method for peptide synthesis in aqueous solution: Application to synthesis of human [Glyl7]-,8-endorphin**. *Proc. Natl. Acad. Sci. USA* 1981, **78**:4055-4058.

35. Saburo A: **Contemporary Methods for Peptide and Protein Synthesis** *Current Organic Chemistry* 2001, **5**:45-87.

36. Lloyd-Williams P, Albericio F, Giralt E (Ed): **Chemical approaches to the synthesis of peptides and proteins** *Barcelona: CRC Press* 1997.

37. Liu C-F, Tam JP: **Chemical Ligation Approach To Form a Peptide Bond between Unprotected Peptide Segments. Concept and Model Study**. *J Am Chem Soc* 1994, **116**:4149-4153.

38. Wieland T, Bokelmann E, Bauer L, Lang HU, Lau H: *Justus Liebigs Ann. Chem.* 1953, **583**:129-149.

39. Proudfoot AE, Rose K, Wallace CJ: **Conformation-directed recombination of enzyme-activated peptide fragments: a simple and efficient means to protein engineering. Its use in the creation of cytochrome c analogues for structure-function studies**. *J Biol Chem* 1989, **264**:8764-8770.

40. Kemp DS: **The amine capture strategy for peptide bond formation - an outline of progress**. *Biopolymers* 1981, **20**:1793-1804.

41. Kemp DS, Leung S-L, Kerkman DJ: **Models that demonstrate peptide bond formation by prior thiol capture I. Capture by disulfide formation**. *Tetrahedron Lett.* 1981, **22**:181-184.

42. LiU C-F, TAM JP: **Peptide Segment Ligation Strategy without Use of Protecting Groups** *Proc. Natl. Acad. Sci. USA* 1994, **91**:6584-6588.

43. Liu C-F, Rao C, Tam JP: **Orthogonal Ligation of Unprotected Peptide Segments through Pseudoproline Formation for the Synthesis of HIV-1 Protease Analogs**. *J Am Chem Soc* 1996, **118**:307-312.

44. Botti P, Pallin TD, Tam JP: **Cyclic Peptides from Linear Unprotected Peptide Precursors through Thiazolidine Formation**. *J Am Chem Soc* 1996, **118**:10018 - 10024.

45. Dawson PE, Muir TW, Clark-Lewis I, Kent SB: **Synthesis of proteins by native chemical ligation**. *Science* 1994, **266**:776-779.

46. Tam JP, Yu Q: **Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones**. *Biopolymers: Peptide Science Section* 1998, **46**:319-327.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

47. TAM JP, Lu Y-A, Liu C-F, SHAO J: **Peptide Synthesis Using Unprotected Peptides through Orthogonal Coupling Methods** *Proc. Natl. Acad. Sci. USA* 1995, **92**:12485-12489.

48. Hackeng TM, Griffin JH, Dawson PE: **Protein synthesis by native chemical ligation: Expanded scope by using straightforward methodology**. *Proc. Natl. Acad. Sci. USA* 1999, **96**:10068-10073.

49. Johnson EC, Kent SB: **Insights into the mechanism and catalysis of the native chemical ligation reaction**. *J Am Chem Soc* 2006, **128**:6640-6646.

50. Shao Y, Lu W, Kent SBH: **A novel method to synthesize cyclic peptides**. *Tetrahedron Lett.* 1998, **39**:3911–3914.

51. Offer J, Boddy CNC, Dawson PE: **Extending synthetic access to proteins with a removable acyl transfer auxiliary**. *J. Am. Chem. Soc.* 2002, **124**:4642-4646.

52. Kawakami T, Akaji K, Aimoto S: **Peptide Bond Formation Mediated by 4,5-Dimethoxy-2-mercaptobenzylamine after Periodate Oxidation of the N-Terminal Serine Residue**. *Org. Lett.* 2001, **3**:1403-1405.

53. Marinzi C, Bark SJ, Offer J, Dawson PE: **A new scaffold for amide ligation**. *Bioorg. Med. Chem.* 2001, **9**:2323-2328.

54. Shao Y, Xu MQ, Paulus H: **Protein splicing: evidence for an N-O acyl rearrangement as the initial step in the splicing process**. *Biochemistry* 1996, **35**:3810-3815.

55. Brik A, Yang YY, Ficht S, Wong CH: **Sugar-assisted glycopeptide ligation**. *J Am Chem Soc* 2006, **128**:5626-5627.

56. Lutsky M-Y, Nepomniaschiy N, Brik A: **Peptide ligation via side-chain auxiliary**. *Chem. Commun* 2008:1229-1231.

57. Liu C-F, Rao C, Tam JP: **Acyl disulfide-mediated intramolecular acylation for orthogonal coupling between unprotected peptide segments. Mechanism and application**. *Tetrahedron Lett.* 1996, **37**:933-936.

58. Nilsson BL, Soellner MB, Raines RT: **Chemical synthesis of proteins**. *Annu Rev Biophys Biomol Struct* 2005, **34**:91-118.

59. Nilsson BL, Kiessling LL, Raines RT: **Staudinger ligation: a peptide from a thioester and azide**. *Org. Lett.* 2000, **2**:1939-1941.

60. Nilsson BL, Kiessling LL, Raines RT: **High-yielding Staudinger ligation of a phosphinothioester and azide to form a peptide.** *Org. Lett.* 2001, **3**:9-12.

61. Soellner MB, Tam A, Raines RT: **Staudinger Ligation of Peptides at Non-Glycyl Residues** *J. Org. Chem.* 2006, **71**:9824-9830.

62. Tam A, Soellner MB, Raines RT: **Water-soluble phosphinothiols for traceless staudinger ligation and**

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

integration with expressed protein ligation. *J Am Chem Soc* 2007, **129**:11421-11430.

63. Bode JW, F. RM, Kyle DB: **Chemoselective amide ligations by decarboxylative condensations of N-alkylhydroxylamine and a-ketoacids**. *Angew. Chem. Intl. Ed* 2006, **45**:1248-1252.

64. Tan XH, Zhang X, Yang R, Liu CF: **A simple method for preparing peptide C-terminal thioacids and their application in sequential chemoenzymatic ligation**. *Chembiochem* 2008, **9**:1052-1056.

65. Ito T: **Role of histone modification in chromatin dynamics**. *J Biochem* 2007, **141**:609-614.

66. Kolakowski RV, Shangguan N, Sauers RR, Williams LJ: **Mechanism of thio acid/azide amidation**. *J Am Chem Soc* 2006, **128**:5695-5702.

67. Gaertner H, Villain M, Botti P, Canne L: **Synthesis of a thioester linker precursor for a general preparation of peptide C-terminal thioacids**. *Tetrahedron Lett.* 2004, **45**:2239-2241.

68. Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y: **Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of Saccharomyces cerevisiae**. *J Biol Chem* 1990, **265**:6726-6733.

69. Mills KV, Perler FB: **The mechanism of intein-mediated protein splicing: variations on a theme**. *Protein Pept Lett* 2005, **12**:751-755.

70. Xu MQ, Perler FB: **The mechanism of protein splicing and its modulation by mutation**. *EMBO J* 1996, **15**:5146-5153.

71. Xu MQ, Evans TC, Jr.: **Intein-mediated ligation and cyclization of expressed proteins**. *Methods* 2001, **24**:257-277.

72. Muir TW, Sondhi D, Cole PA: **Expressed protein ligation: A general method for protein engineering** *Proc. Natl. Acad. Sci. USA* 1998, **95**:6705-6710.

73. Kinsland C, Taylor SV, Kelleher NL, McLafferty FW, Begley TP: **Overexpression of recombinant proteins with a C-terminal thiocarboxylate: implications for protein semisynthesis and thiamin biosynthesis**. *Protein Sci* 1998, **7**:1839-1842.

74. Li X, Kawakami T, Aimoto S: **Direct preparation of peptide thioesters using an Fmoc solid-phase method**. *Tetrahedron Lett.* 1998, **39**:8669-8672.

75. Sarin VK, Kent SB, Tam JP, Merrifield RB: **Quantitative monitoring of solid-phase peptide synthesis by the ninhydrin reaction**. *Anal Biochem* 1981, **117**:147-157.

76. **Novabiochem Catalog**. *Merck Biosciences* 2006/2007.

77. Ellman GL, Courtney KD, Andres V, Jr., Feather-Stone RM: **A new and rapid colorimetric determination of acetylcholinesterase activity**. *Biochem Pharmacol* 1961, **7**:88-95.

78. Tan XH, Wirjo A, Liu CF: **An enzymatic approach to the synthesis of peptide thioesters: mechanism**

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

and scope. *Chembiochem* 2007, **8**:1512-1515.

79. García-Martín F, Quintanar-Audelo M, García-Ramos Y, Cruz LJ, Gravel C, Furic R, Côté S, Tulla-Puche J, Albericio F: **ChemMatrix, a Poly(ethylene glycol)-Based Support for the Solid-Phase Synthesis of Complex Peptides**. *Journal of Combinatorial Chemistry* 2006, **8**:213-220.

80. Schlüter H: **Reversed-Phase Chromatography**. In *protein liquid chromatography*. Edited by Kastner M: Elsevier; 2000:147-217.

81. Licht S, Forouzan F, Longo K: **Differential densometric analysis of equilibria in highly concentrated media: determination of the aqueous second acid dissociation constant of H2S**. *Anal. Chem.* 1990, **62**:1356-1360.

82. Perler FB: **InBase, the Intein Database**. *Nucleic Acids Res* 2000, **28**:344-345.

83. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ: **Crystal structure of the nucleosome core particle at 2.8 A resolution**. *Nature* 1997, **389**:251-260.

84. Kouzarides T: **Chromatin modifications and their function**. *Cell* 2007, **128**:693-705.

85. Strahl BD, Allis CD: **The language of covalent histone modifications**. *Nature* 2000, **403**:41-45.

86. Bhaumik SR, Smith E, Shilatifard A: **Covalent modifications of histones during development and disease pathogenesis**. *Nat Struct Mol Biol* 2007, **14**:1008-1016.

87. Shogren-Knaak MA, Peterson CL: **Creating designer histones by native chemical ligation**. *Methods Enzymol* 2004, **375**:62-76.

88. Shogren-Knaak M, Ishii H, Sun JM, Pazin MJ, Davie JR, Peterson CL: **Histone H4-K16 acetylation controls chromatin structure and protein interactions**. *Science* 2006, **311**:844-847.

89. He S, Bauman D, Davis JS, Loyola A, Nishioka K, Gronlund JL: **Facile synthesis of site-specifically acetylated and methylated histone proteins: Reagents for evaluation of the histone code hypothesis**. *Proc. Natl. Acad. Sci. USA* 2003, **100**:12033-12038.

90. Simon MD, Chu F, Racki LR, de la Cruz CC, Burlingame AL, Panning B, Narlikar GJ, Shokat KM: **The site-specific installation of methyl-lysine analogs into recombinant histones**. *Cell* 2007, **128**:1003-1012.

91. Gloss LM, Kirsch JF: **Decreasing the basicity of the active site base, Lys-258, of Escherichia coli aspartate aminotransferase by replacement with gamma-thialysine**. *Biochemistry* 1995, **34**:3990-3998.

92. Raftery MA, Cole RD: **On the aminoethylation of proteins**. *J Biol Chem* 1966, **241**:3457-3461.

93. Kenyon GL, Bruice TW: **Novel sulfhydryl reagents**. *Methods Enzymol* 1977, **47**:407-430.

94. Rabanal F, DeGrado WF, Dutton PL: **Use of 2,2′ -dithiobis(5-nitropyridine) for the heterodimerization**

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

of cysteine containing peptides. Introduction of the 5-nitro-2-pyridinesulfenyl group. *Tetrahedron Lett.* 1991, **37**:1347-1350.

95. Ghosh AK, Fan E: **A novel method for sequence independent incorporation of activated/protected cysteine in Fmoc solid phase peptide synthesis**. *Tetrahedron Lett.* 2000, **41**:165-168.

96. Richmond TJ, Searles MA, Simpson RT: **Crystals of a nucleosome core particle containing defined sequence DNA**. *J Mol Biol* 1988, **199**:161-170.

97. McCaldon P, Argos P: **Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences**. *Proteins* 1988, **4**:99-122.

98. Buckler DR, Haas E, Scheraga HA: **C-terminal labeling of ribonuclease A with an extrinsic fluorescent probe by carboxypeptidase Y-catalyzed transpeptidation in the presence of urea**. *Anal Biochem* 1993, **209**:20-31.

99. Carles C, Gueguen P, Ribadeau-Dumas B: **C-terminal labelling of beta-casein**. *FEBS Lett* 1987, **212**:163-167.

100. Vinther A, Adelhorst K, Kirk O: **Using capillary electrophoresis in the optimization of a carboxypeptidase Y catalyzed transpeptidation reaction**. *Electrophoresis* 1993, **14**:486-491.

101. Henriksen DB, Rolland M, Jakobsen MH, Buchardt O, Breddam K: **C-terminal amidation of calcitonin by carboxypeptidase Y catalyzed transpeptidation with a photocleavable nucleophile**. *Pept Res* 1992, **5**:321-324.

102. Wood RJ, Pascoe DD, Brown ZK, Medlicott EM, Kriek M, Neylon C, Roach PL: **Optimized conjugation of a fluorescent label to proteins via intein-mediated activation and ligation**. *Bioconjug Chem* 2004, **15**:366-372.

103. Tominaga J, Kemori Y, Tanaka Y, Maruyama T, Kamiya N, Goto M: **An enzymatic method for site-specific labeling of recombinant proteins with oligonucleotides**. *Chem. Commun.* 2007:401-403.

104. Kao MW, Yang LL, Lin JC, Lim TS, Fann W, Chen RP: **Strategy for efficient site-specific FRET-dye labeling of ubiquitin**. *Bioconjug Chem* 2008, **19**:1124-1126.

105. Ha T, Ting AY, Liang J, Caldwell WB, Deniz AA, Chemla DS, Schultz PG, Weiss S: **Single-molecule fluorescence spectroscopy of enzyme conformational dynamics and cleavage mechanism**. *Proc. Natl. Acad. Sci. USA* 1999, **96**:893-898.

106. Antos JM, Francis MB: **Selective Tryptophan Modification with Rhodium Carbenoids in Aqueous Solution**. *J Am Chem Soc* 2004, **126**:10256-10257.

107. Wilchek M, Ben-Hur H, Bayer EA: **p-Diazobenzoyl biocytin--a new biotinylating reagent for the labeling of tyrosines and histidines in proteins**. *Biochem Biophys Res Commun* 1986, **138**:872-879.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

108. Zhang Z, Smith BA, Wang L, Brock A, Cho C, Schultz PG: **A new strategy for the site-specific modification of proteins in vivo**. *Biochemistry* 2003, **42**:6735-6746.

109. Wang L, Magliery TJ, Liu DR, Schultz PG: **A New Functional Suppressor tRNA/Aminoacyl-tRNA Synthetase Pair for the in Vivo Incorporation of Unnatural Amino Acids into Proteins**. *J Am Chem Soc* 2000, **122**:5010.

110. Kurpiers T, Mootz HD: **Regioselective Cysteine Bioconjugation by Appending a Labeled Cystein Tag to a Protein by Using Protein Splicing in trans**. *Angew. Chem. Intl. Ed* 2007, **46**:5234-5237.

111. Geoghegan KF, Stroh JG: **Site-directed conjugation of nonpeptide groups to peptides and proteins via periodate oxidation of a 2-amino alcohol. Application to modification at N-terminal serine**. *Bioconjug Chem* 1992, **3**:138-146.

112. Stennicke HR, Olesen K, Sorensen SB, Breddam K: **C-terminal incorporation of fluorogenic and affinity labels using wild-type and mutagenized carboxypeptidase Y**. *Anal Biochem* 1997, **248**:141-148.

113. Zhang J, Campbell RE, Ting AY, Tsien RY: **Creating new fluorescent probes for cell biology**. *Nat Rev Mol Cell Biol* 2002, **3**:906-918.

114. Yarmolinsky MB, Haba GLDL: **Inhibition by Puromycin of Amino Acid Incorporation into Protein**. *Proc. Natl. Acad. Sci. USA* 1959, **45**:721-729.

115. Agafonov DE, Rabe KS, Grote M, Voertler CS, Sprinzl M: **C-terminal modifications of a protein by UAG-encoded incorporation of puromycin during in vitro protein synthesis in the absence of release factor 1**. *Chembiochem* 2006, **7**:330-336.

116. Chattopadhaya S, Tan LP, Yao SQ: **Strategies for site-specific protein biotinylation using in vitro, in vivo and cell-free systems: toward functional protein arrays**. *Nat Protoc* 2006, **1**:2386-2398.

117. Kalia J, Raines RT: **Reactivity of intein thioesters: appending a functional group to a protein**. *Chembiochem* 2006, **7**:1375-1383.

118. Busque´ F, March Pd, Figueredo M, Font J, Gonza´lez Ls: **A Study of the Conjugate Addition of Thionucleophiles to 2(5H)-Furanones**. *Eur. J. Org. Chem.* 2004, **2004**:1492-1499.

119. Srinivas B, Sridhar R, Surendra K, Krishnaveni NS, Kumar P, Nageswar YVD, Rao KR: **Efficient Regioselective Ring Opening of Oxiranes with Thioacids in the Presence of β-Cyclodextrin in Water** *Synthetic Communications* 2006, **36**:3455-3459.

120. Hakimelahi GH, Just G: **A simple synthesis of 2,2-disubstituted tetrahydrothiophenes**. *Tetrahedron Lett.* 1980, **21**:2119-2122.

121. Rosen T, Lico IM, Chu DTW: **A convenient and highly chemoselective method for the reductive acetylation of azides**. *J. Org. Chem.* 1988, **53**:1580-1582.

122. Shangguan N, Katukojvala S, Greenberg R, Williams LJ: **The Reaction of Thio Acids with Azides: A**

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

New Mechanism and New Synthetic Applications. *J Am Chem Soc* 2003, **125**:7754-7755.

123. Remco Merkx, Haren MJv, Rijkers DTS, Liskamp RMJ: **Resin-Bound Sulfonyl Azides: Efficient Loading and Activation Strategy for the Preparation of the N-Acyl Sulfonamide Linker**. *J. Org. Chem.* 2007, **72**:4574-4577.

124. Abuchowski A, McCoy JR, Palczuk NC, van Es T, Davis FF: **Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase**. *J Biol Chem* 1977, **252**:3582-3586.

125. Harris JM, Chess RB: **Effect of pegylation on pharmaceuticals**. *Nat Rev Drug Discov* 2003, **2**:214-221.

126. Grace MJ, Lee S, Bradshaw S, Chapman J, Spond J, Cox S, Delorenzo M, Brassard D, Wylie D, Cannon-Carlson S, et al.: **Site of pegylation and polyethylene glycol molecule size attenuate interferon-alpha antiviral and antiproliferative activities through the JAK/STAT signaling pathway**. *J Biol Chem* 2005, **280**:6327-6336.

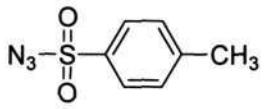
127. Latt SA, Auld DS, Vallee BL: **Fluorescende determination of carboxypeptidase A activity based on electronic energy transfer**. *Anal Biochem* 1972, **50**:56-62.

128. Matayoshi ED, Wang GT, Krafft GA, Erickson J: **Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer**. *Science* 1990, **247**:954-958.

ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

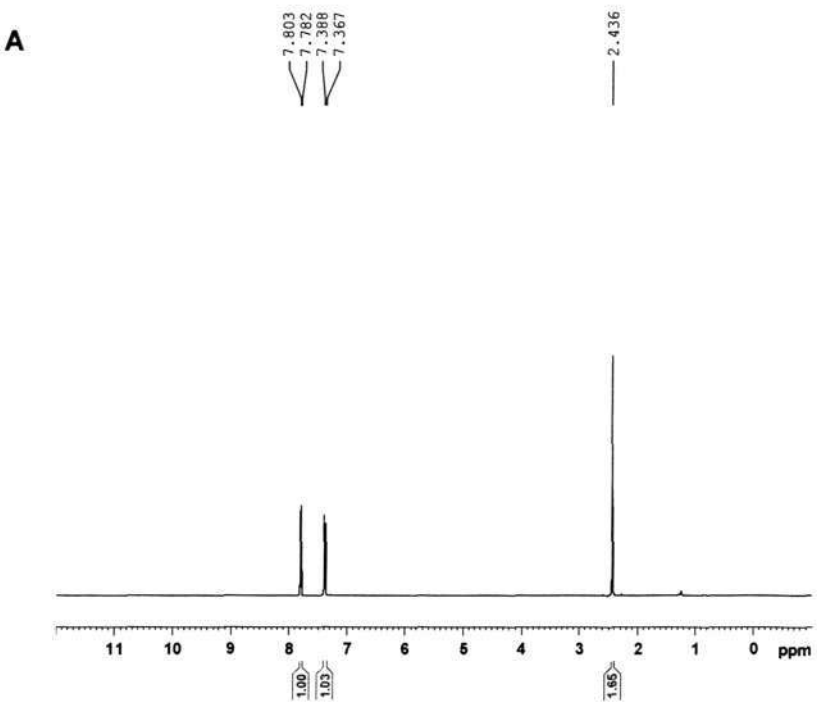
Appendix: NMR spectra of the azide compounds prepared in this study

Appendix 1. (A) ^{1}H and (B) ^{13}C NMR spectra of tosyl azide (**4.1**)

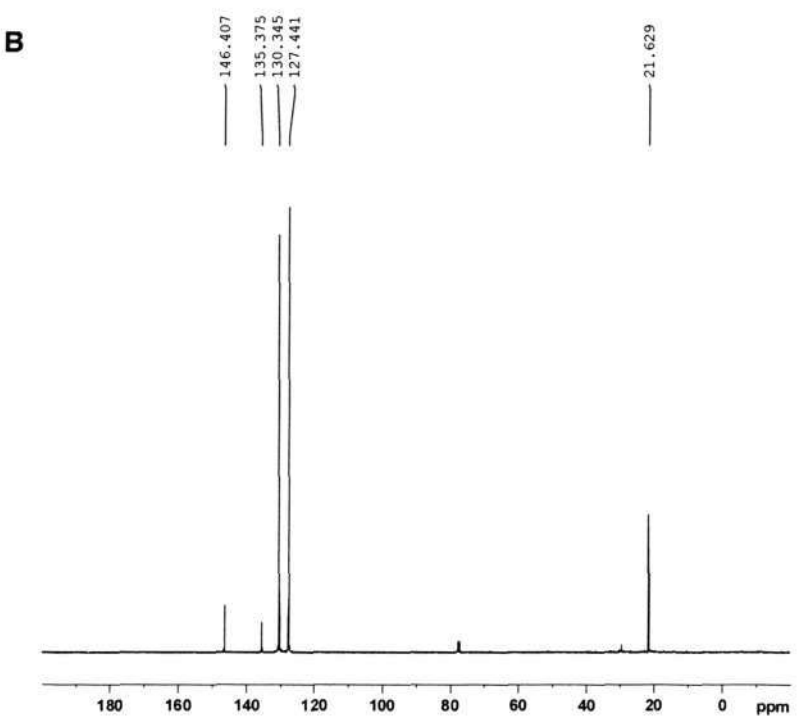


4.1

A

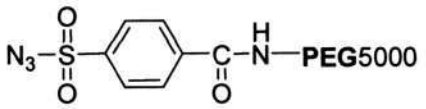


B



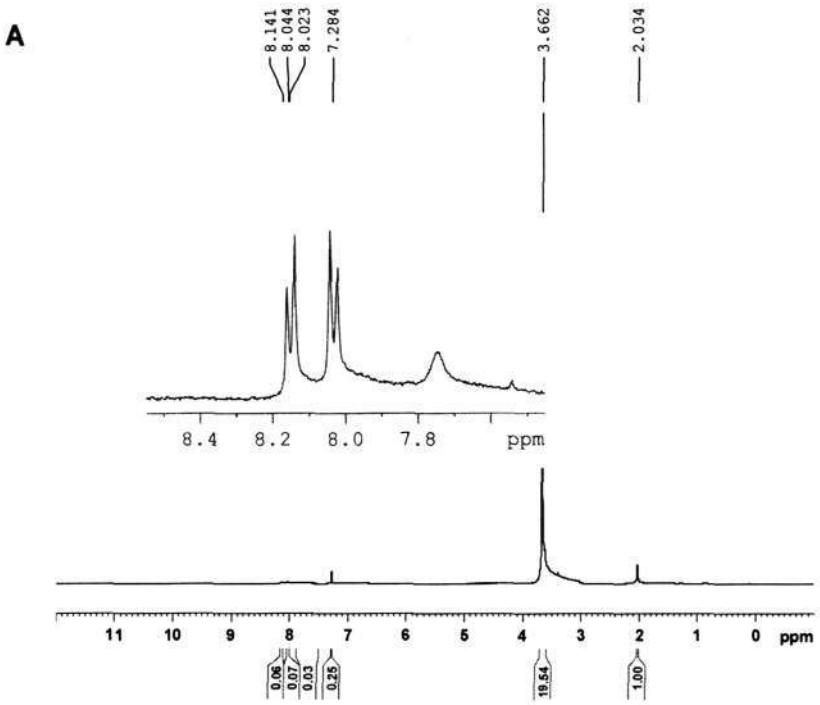
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Appendix 2: (A) ^{1}H and (B) ^{13}C NMR spectra of PEG azide (**4.2**).

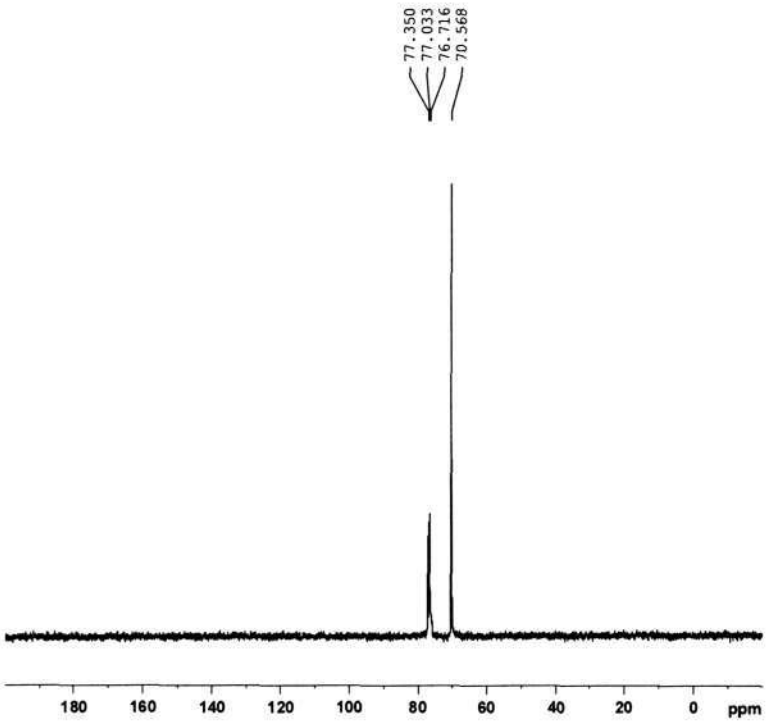


4.2

A

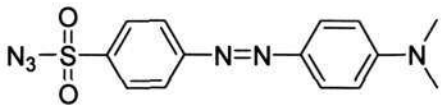


B



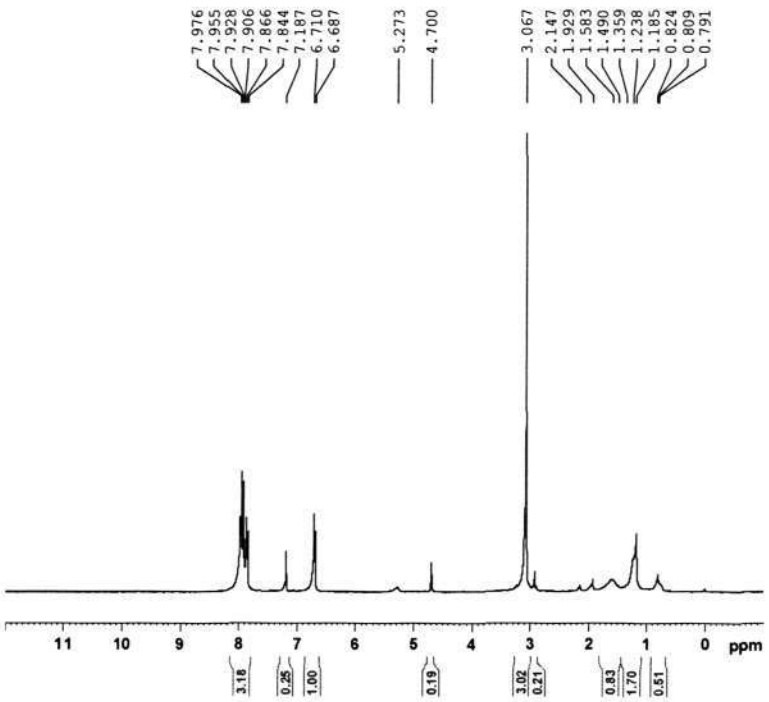
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Appendix 3: (A) ^{1}H and (B) ^{13}C NMR spectra of Dabsyl azide (**4.3**).

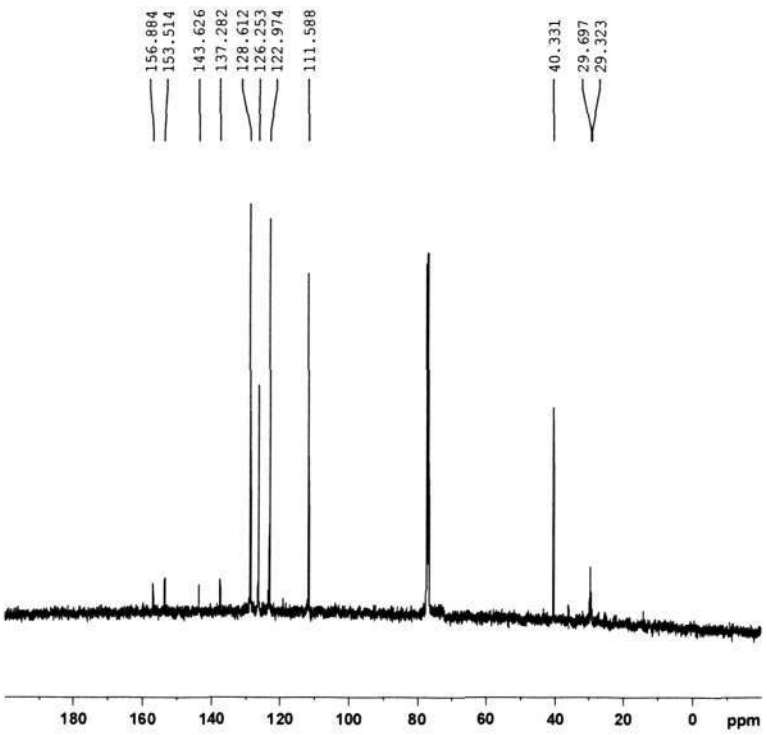


4.3

A

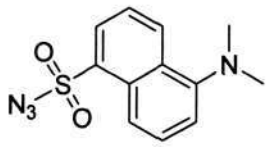


B



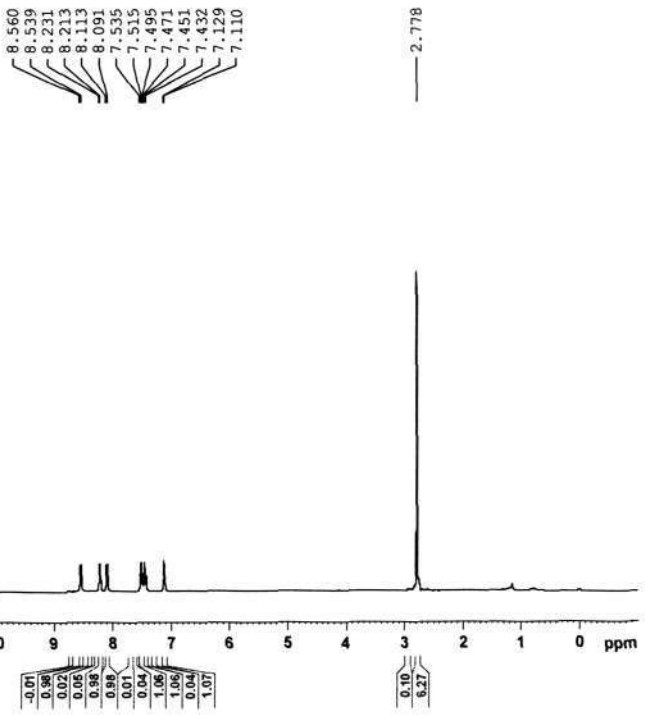
ATTENTION: The Singapore Copyright Act applies to the use of this document. Nanyang Technological University Library

Appendix 4: (A) ^{1}H and (B) ^{13}C NMR spectra of Dansyl azide (**4.4**)



4.4

A



B