Development of novel chemical ligation methods for the synthesis of posttranslationally modified proteins

Yang, Renliang

2012

Development of Novel Chemical Ligation Methods for the Synthesis of Posttranslationally Modified Proteins

Yang Renliang

School of Biological Sciences

A thesis submitted to the Nanyang Technological University

in partial fulfillment of the requirement for the degree of

Doctor of Philosophy

2012

Acknowledgment

I would like to express my gratitude to all those who helped, supported and encouraged me during my postgraduate study. Without them, this thesis could not have been finished smoothly.

First, I would like to thank my supervisor, Dr. Liu Chuan-Fa, for his guidance and assistance during the four years. With his profound knowledge and research experience in chemical biology, he offered me many interesting research topics in the research field. He created a free but motivative research atmosphere in which I could freely use my imagination and explore the mystery of science. He always created the conditions I needed to complete my research. When I needed collaborations, he always helped me to get the suitable collaborators. He always supported me during my research. As my first mentor in my scientific research career, he will have a great impact on me for my entire research career.

I want to thank my collaborators: Prof. Lars Nordenskiöld, Dr. Abdollah Allahverdi, Dr. Liu Xue-Wei and Mr. Kalyan Kumar Pasunooti. I had a happy time during my collaboration with them. Without their work, my projects could not be completed.

I want to thank my group members, Dr. Zhang Xiaohong, Dr. Lu Xiaowei, Dr. Tan Xiaohong, Dr. Liu Yanling and Dr. Zhao Junfeng, Mr Li Fupeng, Ms Hou Wen, and Ms Ding Yingjie for their help, encouragement and friendship during my study. Especially, I want to Dr. Tan Xiaohong for his help during my final year project and the beginning period my Ph.D study. Many of my skills and techniques in peptide synthesis and HPLC were acquired from him. I also want to thank my friends in SBS, especially Mr Wei Lei, Miss Chen Di and Mr Zheng Shichun and so on.

I would like to thank my parents for taking care of me for so many years and their encourangement during my Ph.D study.

Finally, I would like to thank the School of Biological Sciences, Nanyang Technological University and Minstery of Education, Singapore for financial support.

Table of Content

Abstract

Proteins are the central molecules of life. They play important functional roles in vitually every biological process. Peptide and protein chemical synthesis is an enabling tool for the study of the structure and function of proteins, especially those with posttranslational modifications (PTMs), which are difficult to access by the traditional recombinant techniques.

Since the development of solid-phase peptide synthesis (SPPS) and many chemical ligation methods, such as thioester-mediated native chemical ligation, the chemical synthesis of peptides and proteins are revolutionized. Proteins with PTMs remain challenging for biochemists to synthesize utilizing currently available approaches. During my Ph.D study, my work focused on the development of novel chemical methods to overcome these challenges.

In this thesis, I am going to present some of my achievements in methodology development for the synthesis of some typical protein PTMs. In chapter 1, a brief introduction to the background of PTMs, the development of modern peptide chemistry and the methods for the preparation of proteins with PTMs were given.

In chapter 2, we present the ligation/S-alkylation approach which we successfully applied to the synthesis N-terminal tail acetylated histone H4. Three different H4 variants with combinations of acetylation(s) at lysine 5, 8, 12 and 16 were synthesized using the ligation/S-alkylation approach.

In chapter 3 and 4, we focus on the development of methods for the site-specific chemical ubiquitination of peptides and proteins. In chapter 3, we introduce a novel chemical ligation method named dual native chemical ligation at lysine. The dual ligation refers to the ligation at both α- and ε-amine of lysine mediated by the thiol group of 4-thiolysine. This dual

ligation approach does not only expand the scope of chemical ligation to lysine residue but also provides an efficient chemical approach for site-specific ubiquitination. In chapter 4, we modify and improve our dual chemical ligation approach and apply it in the synthesis of a K48-linked diubiquitin. This demonstrated the feasibility of our methodology in synthesizing ubiquitinated proteins.

In chapter 5, we developed a novel N- to C-terminus sequential chemical ligation approach for protein synthesis. This approach works through the combination of native chemical ligation and peptidyl *N,N*-bis(2-mercaptoethyl)-amide (BMEA) mediated ligation. We first demonstrated the feasibility of the approach using small model peptides and then applied the approach to the chemical synthesis of ubiquitin.

In the concluding chapter, I review and discuss the current achievements in chemical synthesis of acetylated and ubiquitinated proteins since these two PTMs are important but challenging to be accessed by chemical approaches.

Abbreviations (listed in alphabetical order)

ACN: acetonitrile

BMEA: peptidyl *N,N*-bis(2-mercaptoethyl)-amide

Boc: *tert*-butyloxylcarbonyl

Cbz: benzyloxylcarbonyl

CD: circular dichroism

DCM: dichloromethane

DIEA/DIPEA: N-ethyldiisopropylamine

DMF: *N,N*-dimethylformamide

DMSO: dimethyl sulfoxide

DTT: 1,4-dithio-D-threitol

E. coli: *Escherichia coli*

EDTA: Ethylenediaminetetraacetic acid

EPL: expressed protein ligation

ESI-MS: electrospray ionization mass spectrometry

Fmoc: 9-Fluorenylmethyloxycarbonyl

FPLC: fast purification liquid chromatography

FT-MS: Fourier transform-mass spectrometry

Gdn•HCl: guanidinium hydrochloride

HEPES: 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid

HF: hydrogen fluoride

HOBt: hydroxylbenzotriazole

HOObt: 3-Hydroxy-1,2,3-benzotriazin-4(3H)-one

HOSu: N-hydroxysuccinamide

HPLC: high performance liquid chromatography

HRMS: high resolution mass spectrometry

IPTG: isopropyl β-D-1-thiogalactopyranoside

MALDI-TOF: matrix-assisted laser desorption and ionization-time of flight

MBHA: 4-Methylbenzhydrylamine

MESNa: sodium 2-mercaptoethanesulfonate

MMTS: S-methyl methanethiosulfonate

MPAA: 4-mercaptophenyl acetate

NCL: native chemical ligation

Npys: 5-nitro-2-pyridinesulfenyl

NVOC: *o*-nitroveratryloxycarbonyl

PCR: polymerase chain reaction

PTMs: posttranslational modifications

PyBOP: Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate

SPPS: solid-phase peptide synthesis

TCEP: *tris*(2-carboxylethyl)phosphine

TFA: trifluoroacetic acid

TFMSA: trifluoromethanesulfonic acid

Thz: thiazolidine

TIS: triisopropylsilane

Trt: trityl

Ub: ubiquitin

VA-044:   2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride

Chapter 1: Introduction

1.1 Overview of protein posttranslational modifications

Proteins are a class of important molecules in biological systems. They are the major executors of almost all aspects of biological functions. Many proteins play structural roles. Proteins like actin and tubulin form an intracellular protein network called cytoskeleton, which is important for the maintenance of cell shape, cell movement and the facilitation of cell division. Many proteins are cell surface receptors and intracellular signaling proteins. These proteins form a complicate signaling network to regulate cellular functions. Many proteins are classified as enzymes. They catalyze chemical reactions which are essential for the biological systems. Although many researchers have devoted their effort to the structural and functional studies of proteins, the structure and function of the majority of the proteins even in the simple organism such as *Escherichia coli (E. coli)* remain unknown.

The basic building blocks of proteins are the 20 natural amino acids. These amino acids join together through amide bonds and form polypeptide chains. The side chain functional groups of these amino acids, such as hydroxyl, thiol, imidazole, amine and amide, endue proteins with diverse functions. Besides these existing functional groups, proteins often undergo extensive posttranslational modifications (PTMs), which add more functional diversity to proteins. These PTMs include methylation, acetylation, acylation (lipidation), phosphophorylation, glycosylation, oxidation, ubiquitination and sumoylation, occurring at the N-terminus of proteins or the side chains of variety of amino acid residues includine lysine, arginine, cysteine, serine, threonine, tyrosine.[1, 2] These modifications play important and diversal roles in modulating protein functions. (Table 1)

Table 1. The list of protein PTMs (The table is revised on based on table 1 of reference 1):

PTM type	Modified amino acid residue	Position	Remarks
Acetylation	S K	N-term anywhere	Reversible, protein stability, regulation of protein function
Phosphorylation	Y, S, T, H, D	anywhere	Reversible, regulation of protein activity, signaling
Cys oxidation disulfide bond glutathionylation sulfenic acid sulfinic acid	 C C C C	anywhere	Reversible, oxidative regulation of proteins
Acylation farnesylation myristoylation palmitoylation	 C G K C (S, T, K)	 anywhere N-term anywhere anywhere	Reversible, cellular localization to membrane
Glycosylation O-linked (O-Glc-NAc) N-linked	 S, T N	anywhere	Reversible, cell-cell interaction and regulation of proteins
Deamidation	N, Q	anywhere	N to D, Q to E
Methylation monomethylation dimethylation trimethylation	 K K K	anywhere	Regulation of gene expression, protein stability
Nitration S-Nitrosylation	Y C		Oxidative damage
Ubiquitination Sumoylation	K K	anywhere [ILFV]K.D	Reversible/irreversible
Hydroxyproline Pyroglutamic acid	P Q	 N-term	Protein stability

Among these modifications, some are extensively studied while others are poorly understood. Core histones (H2A, H2B, H3 and H4) are well known examples undergoing extensive PTMs. Histone PTMs include mono-, di- and tri-methylation of lysine, mono-, symmetric di- and asymmetric di-methylation of arginine, acetylation of lysine, phosphorylation of serine and theronine and ubiquitination of lysine and so on.[3] Many of these PTMs occur at the N-terminal tails of histones, such as the methylation at K4, 9 and 27 of H3, and the acetylation of K5, 8, 12 and 16 of H4. A few important modifications are also

found in the middle of the histone, such as the acetylation of K56 and methylation of K79 of H3. Ubiquitination is found at the C-termini of H2A and H2B. [3]

Different histone PTMs play different roles in chromatin organization, gene expression and epigenetic regulation.[3-6] For instance, histone H3 lysine 4 methylation is associated with transcription activation[7] and the acetylation at histone H4 lysine 16 (H4 K16Ac) modulates both higher chromatin structure and functional interactions between a nonhistone protein and the chromatin fiber.[8] In addition, different histone modifications may be interdependent on each other. The ubiquitination at K120 of human H2B has been shown to activate the intranucleosomal methylation of H3 lysine 79 mediated by K79-specific methyltransferase Dot1 (KMT4).[9]

1.2 Preparation of proteins with PTMs

To elucidate the physiological roles of different modifications, it is important to isolate large amount of homogeneously modified proteins for in vitro studies. The modification of histones using histone modifying enzymes is not satisfying due to the requirement for the identification and isolation of the enzyme. Many of these enzymes are either not specific or cannot drive the reaction to completion. For example, the acetylation of H4 K16 can only reach a maximum yield of 30 % by the *Drosophila* histone acetyltransferase MOF complex which has a highly specificity towards H4 K16.[10] The modified histone and unmodified counterpart are usually very difficult to separate from each other. The study of the modification with the mixture will render the results less conclusive. Several biochemical approaches have been developed to prepare proteins with site-specific PTMs. These approaches are i) genetic incorporation of non-natural amino acids; ii) site-specific modification of recombinant proteins and iii) chemical synthesis of proteins with PTMs through chemical ligation methods.

1.2.1 Genetic incorporation of non-natural amino acids

Many proteins consisting of the 20 natural amino acids can be produced by recombinant technology. A vector containing the gene coding for the desired protein is transformed into the specific expression system. The protein can be expressed and purified. However, the traditional technique can only produce proteins with the natural amino acids and it is difficult to use it to introduce those PTMs site-specifically into proteins.

Both in vitro (cell-free) and in vivo methods have been developed to incorporate non-natural amino acids into peptides and proteins. As the in vitro approaches are usually less efficient and have low fidelity, here only the more successful in vivo mutagenesis methods are introduced.

The first method for the genetic incorporation of non-natural amino acids such as those with PTMs is through the suppression of amber stop codon (UAG).[11] With this approach, an exogeneous orthogonal pair of amino acyl tRNA synthetase (RS)/tRNA$_{CUA}$ pair together with the gene coding for the desired protein is co-transformed into the expression system. The desired codon for the non-natural amino acid is mutated to amber stop codon. The orthogonal aminoacyl tRNA can specifically acylate the tRNA with the non-natural amino acid but not the 20 natural amino acids. The amino acyl tRNA can read through the amber stop and incorporate the non-natural amino acid site-specifically. Until now, mainly two pairs of exogeneous orthogonal RS/tRNA$_{CUA}$ pair are employed for non-natural amino acid incorporation. The first pair is *Mehanococcus jannaschiii* MjTyrRS/tRNA$^{Tyr}_{CUA}$ which enables the incorporation of aromatic amino acids which are analogs of phenylalanine or tyrosine.[12, 13] The second one is *Methanosarcina barkeri* pyrrolysine PylRS/ tRNA$^{Pyl}_{CUA}$ which enables the incorporation of side chain protected lysine analogs.[14-17] With this

technique, many non-natural functional groups have been introduced into proteins, including alkyne, alkene and azide.

The genetic incorporation of non-natural amino acids through amber stop codon suppression with exogeneous RS/tRNA pair is a very innovative approach for the introduction of non-natural elements as well as PTMs into proteins. However, this approach is sophisticated and the systems are not readily available to many research labs. In addition, many non-natural amino acids cannot be incorporated directly with the wild type tRNA synthetase. Frequently, multiple rounds of selection or evolution of tRNA synthetase are required to incorporate the new non-natural amino acids.[18] Moreover, the efficiency of non-natural amino acid incorporation is usually much lower than the incorporation of the native amino acid, which results in the lower yield of protein expression.

1.2.2 Site-specific chemical modification of recombinant proteins

Another method for the generation of proteins with site-specific PTMs is through site-selective chemical reactions, such as alkylation. As mentioned, a number of functional groups, such as hydroxyl, thiol, amine and imidazole, are present in proteins. Among these functional groups, the thiol of cysteine has very unique chemical properties as compared to others. It can i) form disulfide bond with another thiol;[19] ii) form thiol ether bond through S-alkylation;[19, 21-25] iii) can be removed by desulfurization;[26-28] iv) can be eliminated and form dehydroalanine[29] which can be further modified through multiple different chemical reactions;[30-33] v) undergo free radical mediated thiol-ene click chemistry. (Scheme 1.1) Many of these thiol-involving chemistries can be employed for the site-specific installation of protein PTMs.[33, 34] (Scheme 1.1) Several PTMs, such as glyocosylation, methylation, acetylation and ubiquitination have been site-specifically installed based on the unique chemical reactivity of the cysteine thiol. It should be noted that the strategies for the site-

specific introduction of PTMs through the modification of cysteine residue are limited to proteins displaying no other cysteine residues except those involved in modifications.
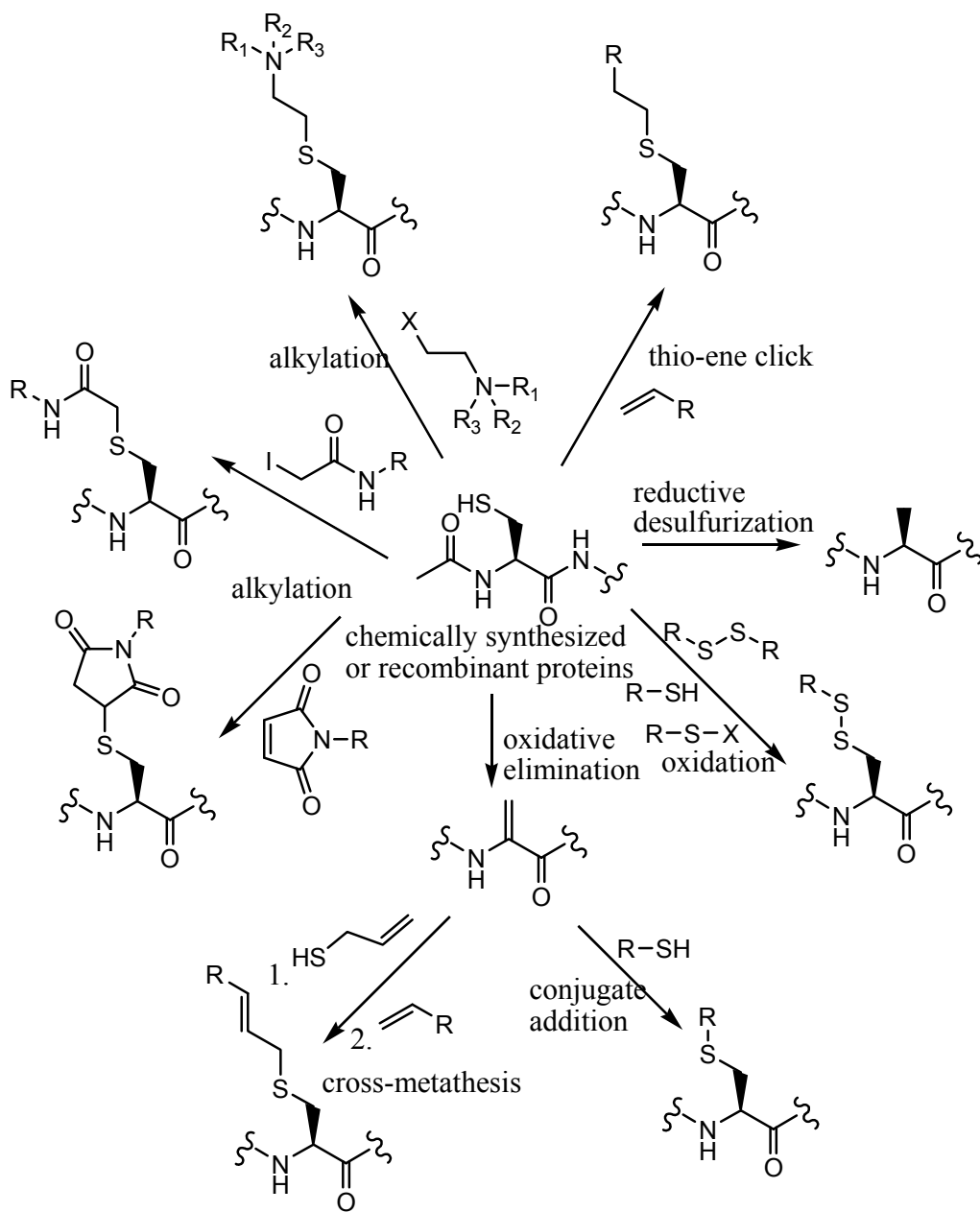
Scheme 1.1 Chemical modifications of proteins at cysteine.

1.2.3 Chemical synthesis of proteins with PTMs

Another robust way for the production of posttranslationally modified proteins is peptide and protein chemical synthesis. Since the research interest of our lab is to develop novel methods for the chemical synthesis of proteins as well as to apply these methods to the synthesis of posttranslationally modified proteins, the following part of the introduction is going to focus on the development of peptide and protein chemistry.

1.2.3.1 Solid-phase peptide synthesis (SPPS)

The monumental innovation of modern peptide chemistry is the development of solid-phase peptide synthesis (SPPS) by Robert Bruce Merrifield in 1963.[35] In his strategy, the Nα-protected C-terminal amino acid was first loaded on a functionalized solid support (resin) which is porous to allow the reactants to access to the resin-bound functional groups but insoluble in the solvent. The Nα-amino group was then deprotected and the second last amino acid was coupled. By using excessive amount of reactant, the coupling reaction can be driven to >99% completion. After coupling, the unreacted reactant can be washed away while the assembled peptide sequence remains attached to the resin. The peptide sequence was assembled by repeating the cycle of deprotection and coupling. After sequence assembly, the global deprotection and detachment from the resin were performed simultaneously by treating the resin with a strong acid. (Scheme 1.2) The advantage of SPPS compared to the traditional solution-phase peptide synthesis is that the former approach avoids the purification of intermediates. For the solution-phase peptide synthesis, work-up and purification are required after each step of reactions. Since the introduction of the concept of SPPS, many different kinds of resins with different functional groups have been developed which enable the synthesis of peptides with different C-terminal functionalities, such as peptide acid, amide, alcohol and hydrazine.
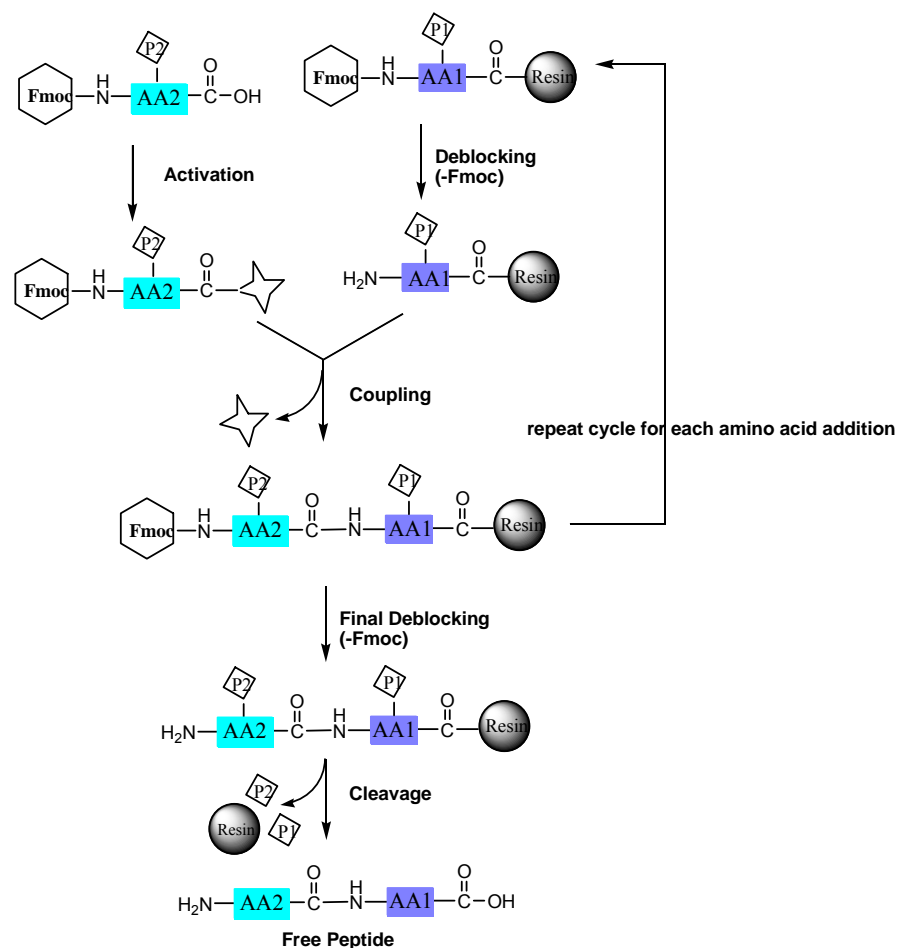
Scheme 1.2 General scheme for Fmoc-SPPS

The length of polypeptides which can be prepared by SPPS is limited to about 50 residues. As the peptide chain length grows, the efficiency of deprotection and coupling are affected due to interchain aggregations. When very long peptides are synthesized, the overall yield is low and the desired product is very difficult to separate from the deletion side products. Many strategies have been developed to overcome this issue. One of the strategy is to use aggregation breaker such as pseudoprolines (Ser($\psi^{Me,Me}$Pro)-OH, Thr(ψMe,MePro)-OH),[36, 37] Nα-dimethoxylbenzyl (Dmb)[38] protected or Nα-hydroxymethoxybenzyl (Hmb)[39] protected amino acids, and isoacyl dipeptides[40-42] instead of the common amino acids to break those secondary structure and minimize the aggregation, therefore increasing the

18

overall yield. Another strategy to synthesize long chain peptide using SPPS is the solid phase peptide segment condensation method.[43, 44] Instead of a single amino acids or dipeptides, the side chain fully protected short peptides with about ten amino acids are used as building blocks for the SPPS. Due to the lower efficiency of this solid phase segmental condensation strategy, the strategy works better when less hindered amino acids such as glycine is present at the condensation junction. In addition, due to the problems of racemization upon activation, only epimerization free amino acids, such as glycine, proline, the pseudoproline derivatives of serine and theronine, are selected as the C-terminal residue of the condensation segments.[43, 44]

Another strategy for the synthesis of long chain polypeptide is through chemical ligation which refers to the solution phase joining of two or more peptides through covalent bond formation. The more general definition of chemical ligations includes the ligation between either partially protected or fully deprotected peptide segments, while the more stringent definition only refers to the later one. In the following part of introduction, I am going to introduce these two different categories of ligations, separately.

1.2.3.2 Ligation between partially protected peptides

The amide bond formation between carboxylate and amine is unlikely without activating reagents or enzymatic catalysis. For peptide ligation to occur, we need to seek more reactive peptide carboxyl derivatives to promote the amide bond formation.

Peptide thioesters have been widely used for amide bond formation between partially protected peptide segments.[45, 46] They are mild reactive peptide C-terminal functionalities. To accelerate the reaction, further activation of alkyl thioester by Ag^+ and HOObt/HOSu is

often required. Recently, it was reported that peptide aryl thioester like S-phenylacetate or S-phenyl thioesters do not require Ag^+ activation for condensation with amine. [47, 48]

In recent years, thioacids also draw more and more attentions to peptide chemists. The condensation between peptide thioacid and amine was also reported.[49, 50] The reaction works in the presence of HOBt, 4Å molecular shieves with DMSO as solvent.[49, 50]

The disadvantage of these thioester and thioacid mediated condensations is that the condensations do not have chemoselectivity and all the amino groups which are not involved in condensation must be protected to avoid undesired reactions.[45-50]

1.2.3.3 Ligation between fully deprotected peptides

The development of methods for chemoselective chemical ligation between fully deprotected peptide segments is one of the important and fundamental research areas in modern peptide chemistry. These methods rely on the unique reactivity between two functional groups at each side of the ligation junction. Many of these ligation methods, such as hydrazone ligation[51, 54] (the reaction between aldehyde/ketone and hydrazine), oxime ligation[52, 54] (the reaction between aldehyde/ketone and aminooxy group) and triazole ligation[53] (the click chemistry between alkyne and azide), donot generate peptide bond after ligation. Even though many of these ligation methods donot generate native peptide bonds after ligation, these methods are useful tools for protein modification and biomolecular labeling.[55]

The development of peptide bond forming chemoselective ligation methods is one of essential topics in peptide research. The principle of these ligation methods is based on proximity-driven peptide bond formation. When the C-terminal functionality **X** of the peptide acyl donor reacts with the N-terminal functionality **Y** of acyl acceptor, the acyl donor was

captured by the acceptor. The newly formed bond inbetween brings the two parts in proximity. The local concentration of the reacting carboxyl and amine is dramatically increased and the proximity-driven effect promotes the amide bond formation. The reaction between **X** and **Y** is chemoselective, which enables the chemoselective ligation between two fully deprotected peptides. By varying the pair of **X** and **Y**, many different chemical ligation methods have been and are being developed. Almost all the modern chemoselective ligation methods are developed based on this proximity-driven principle and they all share the common scheme as shown in Scheme 1.3. In the following part the introduction, I am going to illustrate some typical examples of peptide bond forming chemoselective peptide ligations.
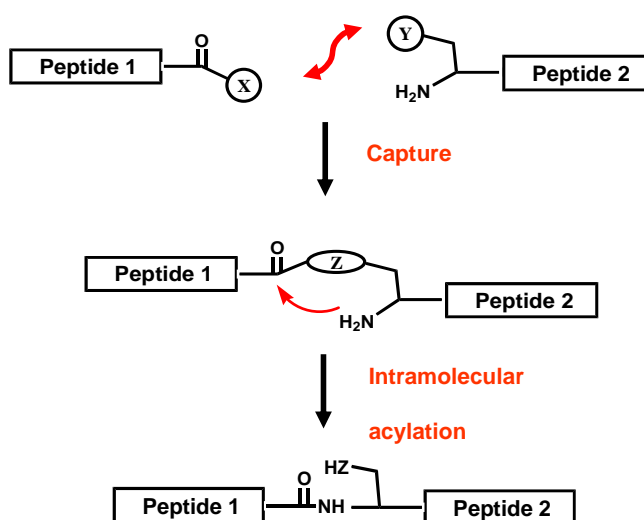
Scheme 1.3 General scheme for proximity-driven chemoselective peptide ligation.

1.2.3.3.1 Thiazolidine/oxazolidine capture ligation

The thiazolidine/oxazolidine capture ligation was first developed by Liu C.-F. et al.[54, 56-58] (Scheme 1.4) The ligation is performed between peptide glycoaldehyde esters and peptides with an N-terminal cysteine/serine/threonine. The ligation is initiated by the

thiazolidine/oxazolidine formation between the aldehyde and the cysteine/serine/theronine. After this capture step, O to N acyl transfer occurs through a five-member ring intermediate and forms an amide bond. This ligation is not traceless as the thiazolidine/oxazolidine remains after ligation, which is perhaps the only disadvantage of this method.
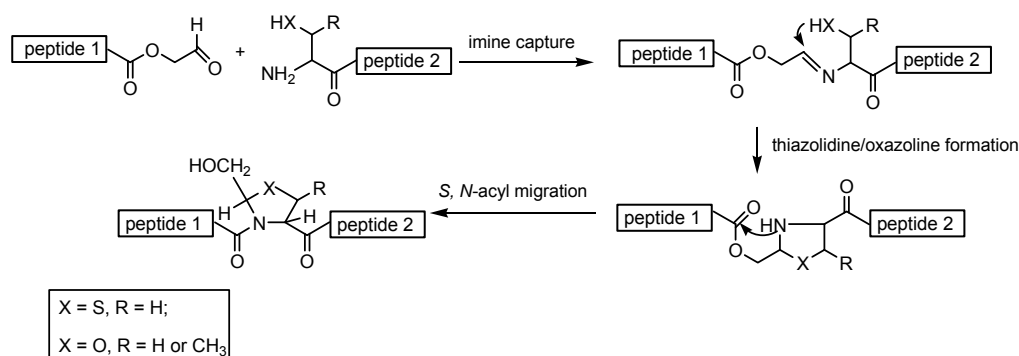
Scheme 1.4 General cheme for thiazolidine/oxazoline capture ligation between peptide glycoaldehydes and peptides with N-terminal cysteine/serine/theronine.

Recently, Li et al. revised the ligation strategy and developed a traceless oxazolidine capture ligation strategy.[59] Instead of alkyl aldehyde, they used an aromatic aldehyde. The ligation is performed between peptide O-benzyl aldehyde and seryl/theronyl peptide. After ligation the oxazolidine can be removed by trifluoacetic acid (TFA) treatment and native peptide bond is formed at Xaa-Ser or Xaa-Thr junctions.

1.2.3.3.2 Thioester-mediated ligation (NCL) and its extentions:

The concept of thioester-mediated ligation can be traced back as early as 1953. In 1953, Wieland et al. reported the amide bond formation between a thioester and a cysteine compound.[60] In 1994, Kent and co-workers first applied this concept to peptide ligation and synthesized a protein through the ligation of peptide thioesters and cysteinyl peptide.[61] They named the ligation as native chemical ligation (NCL) as the ligation is traceless and generates

the native peptide bond at Xaa-Cys site. The ligation begins with the transthioesterification between thioester and the thiol group of cysteine residue and a new thioester intermediate which links the two peptide segments is formed. S to N acyl transfter occurs through a five-member ring transition state and forms the native peptide bond (Scheme 1.5).
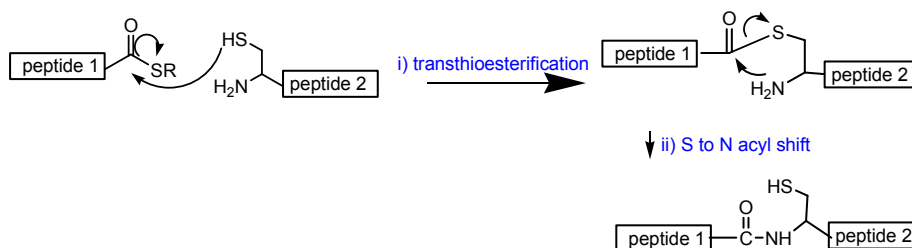
Scheme 1.5 General scheme for NCL.

Since the first development in 1994, NCL has been extensively studied. The relative easy Boc SPPS method for the preparation of peptide alkyl thioesters enhances the practical value of this method.[62] The reaction condition is also standardized. The reaction works more efficiently under mild alkaline condition. Tris(2-carboxylethyl)phosphine (TCEP) serves as a reducing agent to keep the thiol of cysteine from oxidation. An excess of thiol additive is added to generate more reactive thioster in situ and avoid S-acylation of cysteine residues of the ligation product. The effect of different thiol to NCL was examined.[63] It was found 4-mercaptophenyl acetate (MPAA) is one of the best thiol additives.[63] It is soluble in alkaline condition and can generate more reactive aryl thioesters by exchange with alkyl thioster.

The ligation efficiency of the thioesters with different C-terminal residues was also examined.[64] As expected, thioesters with less steric hindered amino acids, such as Gly, Ala, at C-termini ligates more efficiently with Cys-peptide, while those with β-branched side chains (Val, Ile) work less efficiently. Prolyl thioesters is extremely unreactive towards cysteinyl peptides.

To synthesized large-sized proteins, two peptide segments are usually not sufficient to cover the full length. Therefore, the sequential ligation of multiple peptide fragments is required.[65-68] To do this, the protecting group for the N-terminal cysteinyl residue of the middle cyteinyl peptide thioesters was developed.[69-71] Thiazoline was used as protected N-terminal cysteine residue of the middle thioester, which enables the C to N sequential ligations to synthesize large proteins.[69, 70] (Scheme 1.6) The thiazolidine derivative of cysteine is very easy to synthesize and also commercially available. The deprotection of Thz is achieved relatively easily by using methoxylamine. All these make the sequential ligation strategy widely used.
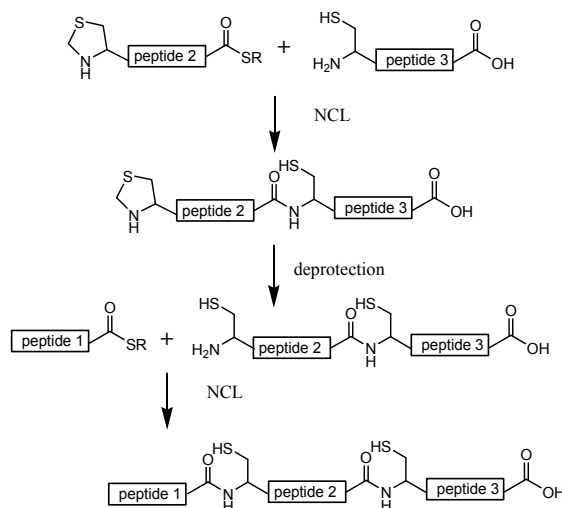
Scheme 1.6 General scheme for C to N sequential native chemical ligation with Thz as internal segment protecting group.

The development of NCL brings about two issues of peptide chemistry which are related to the two ligation partners. The first one is the preparation of peptide thioesters. The second one is to overcome cysteine restriction. As mentioned earlier, peptide alkyl thioester can be directly synthesized using tert-butyloxylcarbonyl- (Boc-) based SPPS chemistry. However, they cannot be directly synthesized using 9-fluorenylmethyloxycarbonyl- (Fmoc-) chemistry

due to the lability of thioester bond to nucleophilic attack of piperidine during Fmoc removal. Several strategies have been developed for the indirect synthesis of thioesters using Fmoc chemistry.[72] Some typical strategies include i) the revise of Fmoc deprotection protocol which is safer to thioester bond;[73, 74] ii) the O to S[75-77] or N to S[78-90] acyl transfer and iii) thiolysis of CO-X bond which is stable to piperidine.[91-97] One strategy contributed from our laboroty is the Fmoc synthesis of thioester derived peptidyl *N,N*-bis(2-mercaptoethyl)-amide (BMEA).[89] The same strategy was also independently developed by Melnyk's group[90, 91].

Different from peptide thioester which can be prepared by SPPS, the C-terminal thioesters of large-sized proteins cannot be prepared through this way due to the length of peptides that can be synthesized by SPPS. The way to prepare protein thioesters is through the thiolysis of protein-intein fusions. Intein is a class of proteins which can catalyze protein splicing and link the two peptide segments flanking their N-terminus and C-terminus (termed *exteins*) together and form a new protein. Based on this, when a protein is C-terminally fused with intein, its thioester can be generated by thiolysis of the fusion with an alkyl thiol, like sodium 2-mercaptoethanesulfonate (MESNa). The thioester prepared this way can be ligated with either chemically synthesized cysteinyl peptides or recombinant proteins through NCL. This is the so called expressed protein ligation (EPL).[98] The strategy is especially useful for protein C-terminal labeling and the semisynthesis of proteins with C-terminal modifications.

The second issue regarding NCL is that it requires a cysteine residue at the ligation junction. Cysteine is not a common amino acid and has low frequency in naturally-occurring proteins. Many strategies have been developed to expand the scope of NCL to non-cysteinyl residues.

Ligation at Xaa-Met site was performed by homocysteine (Hcy) mediated ligation followed by S-methylation.[99] The ligation between thioesters and homocysteine is through a

six-membered ring intermediate rather than the five-membered ring for cysteine mediated ligations. After ligation, the homocysteine can be methylated and converted to the native methionine residue at the ligation junction. The Hcy mediated ligation works effectively under pH 7-8. Under lower pH, the N-terminal Hcy is protonated which slows down the transthioesterification and S→N acyl migration process. When higher pH (>8) is used for ligation, degradation of Hcy peptide occurs due to the lability of Hcy-X amide bond, which can undergo a five-member ring thiolactone cyclization.[99]

Another strategy to expand thioester-mediated ligation to non-cysteinyl amino acids is Nα-auxiliary mediated ligations.[100-105] A thiol-containing removable auxiliary was placed at the N-terminal amino group of acyl receptor peptide. After ligation with a thioester, the auxiliary can be removed by physical (such as UV irradiation) or chemical (such as strong acid) means and the native amide bond is generated at the ligation site. As listed in Scheme 1.7, many different groups have been selected as the auxiliary. Due to the attachment of the auxiliary, the N-terminal amine becomes a secondary amine and sterically hindered, which significantly affects the ligation efficiency. Therefore, the approach only works when less hindered amino acids such as Gly are involved at the ligation site.
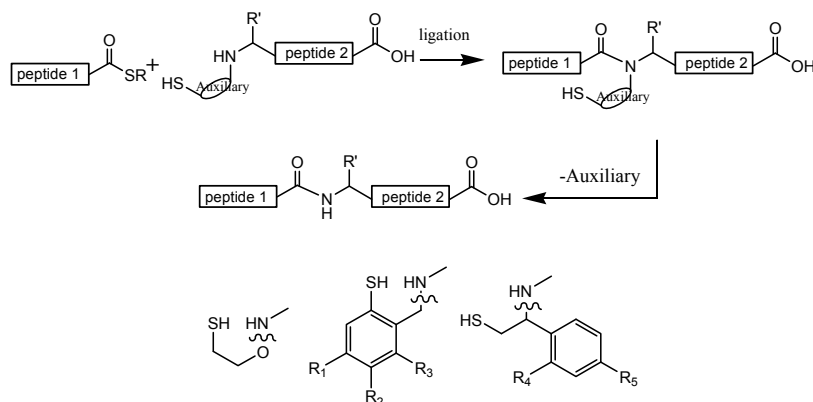
Scheme 1.7 General scheme for Nα-auxiliary mediated ligation and the examples of different auxiliaries.

26

The third strategy to expand the scope of NCL is ligation/desulfurization.[106-115] In contrast to the relatively large auxiliary used in Nα-auxiliary mediated ligation, a small thiol group (-SH) can also serve as an auxiliary. Instead of putting the auxiliary at N-terminal amine, the thiol group can be placed at the right position (β and γ position) of the side chain of the N-terminal amino acid to facilitate the chemical ligation. After ligation, the thiol group can be removed tracelessly through metal-mediated or free-radical mediated desulfurization. This ligation/desulfurization strategy has enabled the chemical ligation at Ala[106], Phe[107], Val[108, 109], Thr[110], Leu[111, 112] and Lys[113, 114] (Scheme 1.8). It has to be noted that all the thiols which are not involved in desulfurization should be protected to avoid undesired desulfurization.
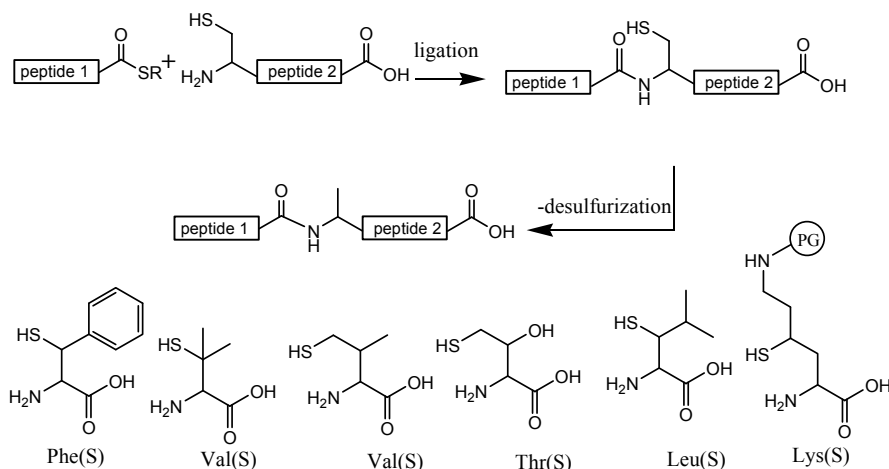
Scheme 1.8 General scheme for Ala ligation through ligation/desulfurization. The similar strategy also used for phe, val, thr, leu and lysine ligations. The below figures shows the S-containing amino acid surrogates which have been used for ligation.

1.2.3.3.3. Thioacid mediated ligations

The thioacid functionality is a very soft and powerful neucleophile which has long been exploited for peptide synthesis. It is involved in several kinds of typical peptide ligation reactions which form native peptide bonds. One of them is thioalkylation capture ligation.[62]

(Scheme 1.9) When a peptide thioacid is reacted with a peptide bearing an N-terminal β-bromoalanine, the thiol can be alkylated by the 2-bromoethylamine functionality. S to N acyl transfer then occurs and forms the native amide bond. After ligation, a cysteine residue is generated at the ligation site. This ligation should be performed at pH below 5 to eliminate side reactions. When the reaction is performed at pH 5.0, the thioesterification of the thioacid through direct S-alkylation by β-bromoalanine will be dominant and the formation of aziridine of the N-terminal β-bromoalanine is minimized. Under higher pH (6.0), the aziridine formation is favored and provides two options (either at α or β position of the ring) for ring opening of thioacid. While the S-attack of the β position of the ring results in the formation of the desired ligation product, the S-attack of α position results in the formation of a side product containing a β-amino acid rather than cysteine after ligation.
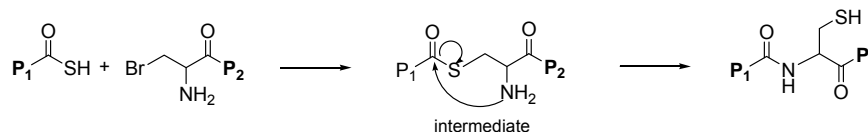
Scheme1.9 General scheme for thioalkylation capture ligation. **P₁**, **P₂**, peptide 1 and peptide 2.

Another chemoselective chemical ligation involving thioacid is thiol capture ligation[116]. The ligation is performed between a peptide thioacid and a peptide bearing an N-terminal NPys protected cysteine residue. (Scheme 1.10) Under mild acidic condition, the thiol of thioacid can attack the activated disulfide bond and form new disulfide bond with cysteine residue. S to N acyl transfer then occurs via a six-member ring intermediate and forms amide bond. After ligation, reducing agent, like 1,4-dithio-D-threitol (DTT) or TCEP, is added to reduce the disulfide and generate native cysteine at the ligation site.
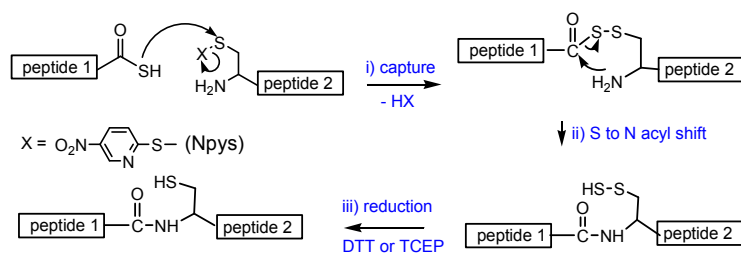
Scheme 1.10 General scheme of thioacid-capture ligation.

1.2.3.3.4 Traceless Staudinger ligation:

Staudinger reaction or reduction, which is invented by Hermann Staudinger, refers to the chemical reaction between an azide and a phosphine. The reaction produces an iminophosphane intermediate which is then hydrolyzed and produce an amine and a phosphane oxide.[117]

Staudinger ligation [118] is developed by Saxon and Bertozzi in 2000 through the modification of the classical Staudinger ligation. In Staudinger ligation, an electrophilic trap is properly placed on the triaryl phosphane. The iminophosphane intermediate undergoes rearrangement and forms an amidophosphonium salt through the attack of the acyl group by the nitrogen. After hydrolysis, a phosphine oxide is formed and linked with the newly formed amide.

A traceless Staudinger ligation strategy[119-122] was developed for peptide ligation and protein modifications through the reaction between a peptide-phosphinothioester and an azido-peptide. After hydrolysis, the phosphane oxide moiety is released from the backbone of the ligation product. (Scheme 1.11)
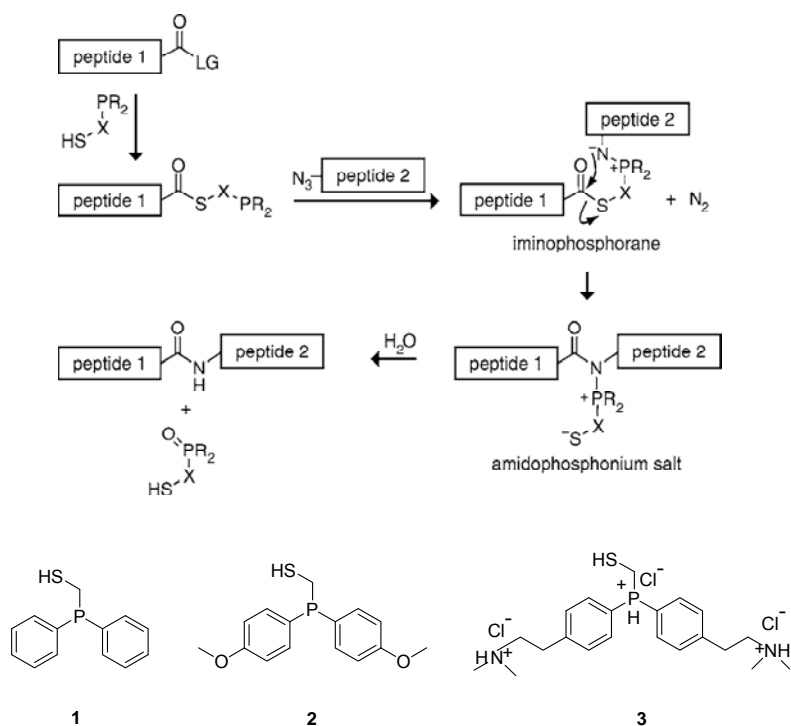
Scheme 1.11 Outline of traceless Staudinger ligation and phosphinothiol compounds for mediation of peptide ligation.

The reaction mechanism and kinetics of the traceless Staudinger ligation have been extensively studied.[123] Intial studies performed for Staudinger ligation between glycyl residues showed that (diphenylphosphino)methanethiol (compound **1** in Scheme 1.11) was judged to be the most efficacious known reagent for the traceless Staudinger ligation.[123] Later, the ligation was extended to non-glycyl residues by installing *p*-methoxy groups on the phenyl groups of (diphenylphosphino)methanethiol (compound **2**, Scheme 1.11) and using a solvent of low polarity (such as toluene or dioxane).[124] Both of the two phosphinothiols developed are water insoluble, which limits the biological applications. To overcome the limitation, water-soluble phosphinothiols (such as compound **3**, Scheme 1.11) were developed for traceless Staudinger ligation. The development of these water-soluble phosphinothiols enabled the ease production of protein phosphinothioesters through thiolysis of protein-intein fusions.[125]

One of the advantages of traceless Staudinger ligation compared to native chemical ligation is that the former approach doesnot require the cysteine resiude at the ligation junction. However, although extensive efforts have been done to optimize the phosphinothiols for traceless Staudinger ligation, the ligation approach was only demonstrated for the ligation of very short peptides. Recently, the ligation was successfully applied for the synthesis of 11-amino acid cyclized peptides.[126]

1.2.3.3.5 Decarboxylative condensation

Decarboxylative condensation[127] is an innovative peptide ligation approach which does not use common peptide carboxyl derivatives as peptide acyl donors. The condensation is between a peptide α-ketoacid and a Nα-hydroxypeptide. (Scheme 1.12) It was proposed that during the ligation, N-alkylhydroxylamines reacted with α-ketoacids to produce a hemiaminal poised for oxidative decarboxylation to give amide products.
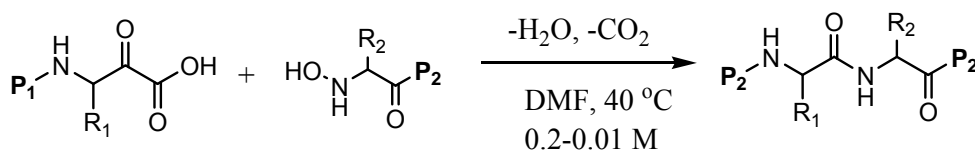
Scheme 1.12 Peptide bond formation through decarboxylative condensation

As mentioned, this method is an innovative ligation approach which is not based on the knowledge of thioester-mediated ligations. However, although extensive research has been performed on the preparation of the two ligation partners (α-ketoacids and Nα-hydroxypeptides),[128-132] the synthesis is still not as straight-forward as peptide thioesters and cysteinyl peptides. This significantly limits the application of this ligation strategy.

Previously, ligation with this decarboxylative condensation method was demonstrated with short peptides. Recently, this method was succesfully applied to the synthesis of a relatively long peptide human GLP-1 (7-36) through an one-step ligation.[133]

1.2.3.3.6 Ligation with peptide hydrazine

The chemoselective ligation between peptide hydrazines and cysteinyl peptides was recently developed by Liu et al.[134] The ligation was performed with unprotected peptides in aqueous solution. The peptide hydrazine was first oxidized by sodium nitrite to peptide carboxylic azide which is a reactive carboxylic derivative at pH 3.0. The peptide azide was then exchanged with excess of thiol (for example, MPAA) and formed peptide thioester under neutral pH. The following reaction was thioester-mediated ligation. The method not only applies to small peptide ligation, it can also synthesize large proteins as the protein C-terminal hydrazines can be easily generated through the hydrazinolysis of protein-intein fusions.
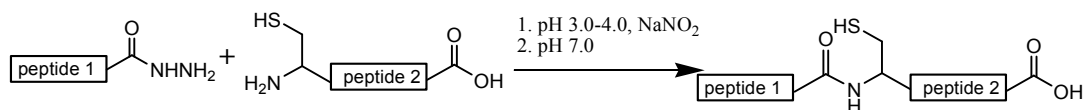
Scheme 1.13 General scheme for the chemical ligation between peptide hydrazines and cysteinyl peptides.

1.2.3.4 Synthesis of posttranslationally modified proteins with chemical ligation methods

Chemical ligation methods, especially NCL, have been widely used as methods for the synthesis of proteins with PTMs as many of these modifications can be directly incorporated during the SPPS into the segments for the ligation. With these chemical ligation methods, the PTMs close to protein N-terminus can be easily accessed by semisynthesis through the ligation of chemically synthesized N-terminal thioesters with desired modifications and

recombinant C-terminal domain bearing an N-terminal cysteine residue. With these semisynthetic approach, histone H3 with N-terminal trimethylated K9 or pentaacetylations at K4, 9, 14, 18, 23 were chemically synthesized with Ala25Cys as ligation junction.[135] H4 with combination of acetylations at K5, 8, 12 have also been synthesized with Ala15Cys as ligation site.[135] After ligation, the cysteine residue at the ligation junction was converted to alanine through metal-mediated desulfurization and generated the native sequence.[135] With a similar ligation/desulfurization approach, H2A with N-terminal phosphorylated serine was also synthesized through semisynthesis.[136] In addition, H4 with acetyllysine 16 was also synthesized by usine Arg23Cys as ligation site.[137, 138] After ligation, this point mutation remained. Although, it has been shown that this mutation did not affect the K16Ac mediated nucleosome unfolding and array aggregation, the H4 synthesized this way differs significantly from the native protein.

Proteins with PTMs close to their C-termini can also been synthesized through semisynthetic approach. With the expressed protein ligation (EPL) approach, the thioester of the N-terminal domain of a protein bearing no PTMs can be generated through the thiolysis of its intein fusion with thiols like MESNa. While the remaining C-terminal peptide containing desired PTMs can be synthesized by SPPS. The two parts can then be ligation through thioester-mediated chemical ligation. With EPL approach, H3 with C-terminal acetylations[139] and H2B with C-terminal ubiquitination[140] were synthesized. Besides the EPL approach, the semisynthesis of protein with C-terminal PTMs can also be done with other chemistries. Recently, our group reported the semisynthesis of C-terminally acetylated H3 through thioacid capture ligation.[141] An H3 N-terminal domain thioacid was generated through the hydrothiolysis of H3 N-domain-intein fusion with HS⁻. The full length H3 was generated through the ligation between this thioacid and Npys activated C-terminal cysteinyl

peptide. It was reported that the semisynthesis of H3 employing thioacid capture ligation (at Leu-Cys site) was much more efficient than thioester-mediated ligation.

It is difficult to access those PTMs located in the central region of protein sequences using semisynthetic approaches. For the synthesis of proteins with PTMs at their central region, total chemical synthesis through the ligation of multiple peptide fragments is often required. For example, H3 with acetylation at K56 was chemically synthesized through the C-to-N sequential ligation of three segments.[142]

The PTMs which can be accessed by chemical ligation approaches are not limited to those low molecular weight modifications. Peptides and proteins with large modifications, such as ubiquitination, were also chemically synthesized. The chemical synthesis of site-specifically ubiquitinated peptides was first reported through Nα-auxiliary mediated ligation.[143] In the approach, the terminal glycine 76 of ubiquitin with a photoremovable auxiliary at its α-amine was chemically predeposited on a lysine side chain which was involved in ubiquitination. A ubiquitin (1-75)-MES generated through the thiolysis of ubiquitin-intein fusion with MESNa was ligation with the Gly76 through auxiliary mediated ligation. The auxiliary was removed to generate the native ubiquitin tag through UV irradiation after ligation. Although the ligation was performed at Gly-Gly site, the ligation efficiency was slow as reported. Other chemical approaches which enable efficient site-specific ubiquitination need to be developed.

1.3 Objectives of this study

The objectives of my study are to develop novel chemical ligation methods for peptide and protein synthesis. The ultimate goal of doing this is to apply the newly developed methods to the chemical synthesis of proteins with those PTMs which are difficult to be synthesized using the existing methodologies.

The first objective is to develop new method for the synthesis of H4 with N-terminal modifications, especially H4 K16Ac. Although, many methods to synthesize N-terminal acetylated H4 have been reported. The ligation/desulfurization approach with Ala15Cys as the ligation junction is only suitable for the synthesis of H4 with acetylation(s) at K5, 8 and/or 12.[123] The only reported method for the synthesis of H4 K16Ac before we initiated the project created a significant point mutation (R23C) after NCL.[137, 138] We set to develop a new method which could synthesize H4 K16Ac as close to the native sequence as possible. The part of the work regarding this objective is presented in chapter 2.

The second objective of my study is to develop a new method for peptide/protein ubiquitination. Ubiquitination is an important PTMs. As this modification is the linking of a relatively large protein to the lysine side chain of the modified protein through isopeptide bond. The majority of the traditional chemical ligation methods developed cannot be directly applied to attach the ubiquitin to lysine side chain. Therefore, chemical ubiquitination is extremely challenging. As mentioned, Muir's team reported the first method for site-specific peptide ubiquitination by using Nα-auxiliary mediated chemical ligation.[131] However, the kinetics of the ligation is slow. Therefore, we set to develop more efficient methods for the chemical synthesis of ubiquinated peptides. The work on the development of new chemical ligation method for peptide ubiquination is presented in chapter 3.

After the achieving of the goal of efficient site-specific chemical ubiquination of small peptide, our next objective is to apply this method for protein ubiquination to verify the feasibility of our methodology. We choose to synthesize K48-linked diubiquitin as diubiquitins are the fundemental repeative units of polyubiquitin. The study of diubiquitins will be extremely helpful for the elucidation of structural and functional effects of

polyubiquitins. In chapter 4, we have revised the methodology for peptide ubiquitination and applied it to the synthesis of K48-linked diubiquitin.

As mentioned earlier, it is difficult to access the PTMs in the middle of the protein sequence using the semisynthetic approach. To do this, the total syntheis of the protein through repetitive ligation of multiple chemically synthesized peptide fragments are required. The ligations can be performed in a convergent way, from C to N direction, or in the reverse direction. While many C-to-N sequential ligation methods have been reported using different chemistries, only a few N-to-C seqential ligation methods are reported. The final objective of my study is to develop a new method for protein synthesis through N-to-C sequential ligation. The part of this work is presented in chapter 5.

References

[1] J. Seo, K.-J. Lee, *J. Biochem. Mol. Biol.* **2004**, 37, 35.

[2] C. T. Walsh, S. Garneau-Tsodikova, G. J. Gatto, Jr., *Angew. Chem., Int. Ed.* **2005**, 44, 7342.

[3] C. L. Peterson, M. A. Laniel, *Curr. Biol.* **2004**, 14, R546.

[4] S. L. Berger, *Nature* **2007**, 447. 407.

[5] J. C. Rice, C. D. Allis, *Curr. Opin. Cell. Biol.* **2001**, 13, 263.

[6] B. M. Turner, *Bioassays* **2000**, 22, 836.

[7] B. D. Strahl, R. Ohba, R. G. Cook, C. D. Allis, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 14967.

[8] M. Shogren-Knaak, H. Ishii, J. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson, *Science* **2006**, 311, 844.

[9] R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder, T. W. Muir, *Nature* **2008**, 453, 812.

[10] P. J. J. Robinson, W. An, A. Routh, F. Martino, L. Chapman, R. G. Roeder, D. Rhodes, *J. Mol. Biol.* **2008**, 381, 816.

[11] L. Wang, P. G. Schultz, *Angew. Chem., Int. Ed.* **2005**, 44, 34.

[12] L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, *J. Am. Chem. Soc.* **2000**, 122, 5010.

[13] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, 124, 9026.

[14] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, 296, 1459.

[15] S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Green-Church, M. K. Chan, J. A. Krzycki, *Nature* **2004**, 431, 333.

[16] C. R. Polycarpo, S. Herring, A. Berube, J. L. Wood, D. Söll, A. Ambrogelly, *FEBS Lett.* **2006**, 580, 6695.

[17] H. Neumann, S. Y. Peak-Chewand, J. W. Chin, *Nat. Chem. Biol.* **2008**, 4, 232.

[18] S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotech.* **2002**, 20, 1044.

[19] M. W. Crankshaw, G. A. Grant, *Modification at cysteine,* Wiley, Unit 15.1, **1996**.

[20] D. R. Goddard, L. Michaelis, *J. Biol. Chem.* **1935**, 112, 361.

[21] G. T. Hermanson, *Bioconjugation Techniques,* 2nd edition, Academic Press, Inc., **2008**.

[22] M. S. Masri, M. Friedman, *J. Protein Chem.* **1988**, 7, 49.

[23] H. Lindley, *Nature* **1956**, 178, 647.

[24] H. B. Smith, F. C. Hartman, *J. Biol. Chem.* **1988**, 263, 4921.

[25] M. D. Simon, F. Chu, L. R. Racki, C. C. De La Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, K. M. Shokat, *Cell* **2007**, 128, 1003.

[26] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, 123, 526.

[27] Q. Wan, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2007**, 46, 9248.

[28] C. Haase, H. Rohde, O. Seitz, *Angew. Chem., Int. Ed.* **2008**, 47, 6807.

[29] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. L. Bernardes, L. Griffin, H. Hailu, C. J. Schofield, B. G. Davis, *Chem. Sci.* **2011**, DOI: 10.1039/C1SC00185J.

[30] T. J. Holmes, R. G. Lawton, *J. Am. Chem. Soc.* **1977**, 99, 1984.

[31] D. H. Rich, J. Tam, P. Mathiaparanam, J. A. Grant, C. Mabuni, *J. Chem. Soc. Chem. Commun.* **1974**, 897.

[32] G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, *J. Am. Chem. Soc.* **2008**, 130, 5052.

[33] J. M. Chalker, G. J. L. Bernardes, Y. A. Lin, B. G. Davis, *Chem. Asian J.* **2009**, 4, 630.

[34] C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem., Int. Ed.* **2008**, 47, 10030.

[35] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, 85, 2149.

[36] T. Haack, M. Mutter, *Tetrahedron Lett.* 1992, 33, 1589.

[37] M. Mutter, A. Nefzi, T. Sato, X. Sun, F. Wahl, T. Wohr, *Pept. Res.* **1995**, 8, 145.

[38] T. Johnson, L. C. Packman, C. B. Hyde, D. Owen, M. Quibell, *J. chem. soc., Perkin trans. 1.* **1996**, 719.

[39] T. Johnson, M. Quibell, D. Owen, R. C. Sheppard, *J. chem. Soc., chem. Commun.* **1993**, 369.

[40] Y. Sohma, A. Taniguchi, M. Skwarczynski, T. Yoshiya, F. Fukao, T. Kimura, Y. Hayashi, Y. Kiso, *Tetrahedron Lett.* **2006**, 47, 3013.

[41] I. Coin, R. Dölling, E. Krause, M. Bienert, M. Beyermann, C. D. Sferdean, L. A. Carpino, *J. Org. Chem.* **2006**, 71, 6171.

[42] T. Yoshiya, A. Taniguchi, Y. Sohma, F. Fukao, S. Nakamura, N. Abe, N. Ito, M. Skwarczynski, T. Kimura, Y. Hayashi, Y. Kiso, *Org. Biomol. Chem.* **2007**, 5, 1720.

[43] P. Lloyd-Williams, F. Albericio, E. Giralt, *Tetrahedron* **1993**, 49, 11065.

[44] C. Heinlein, D. V. Silva, A. Troster, J. Schmidt, A. Gross, C. Unverzagt, *Angew. Chem., Int. Ed.* **2011**, 50, 6406.

[45] H. Hojo, S. Aimoto, *Bull. Chem. Soc. Jpn.* **1991**, 64, 111.

[46] S. Aimoto, *Biopolymers* **1999**, 51, 247.

[47] H. Hojo, Y. Murasawa, H. Katayama, T. Ohira, Y. Nakahara, Y. Nakahara, *Org. Biomol. Chem.* **2008**, 6, 1808.

[48] R. J. Payne, S. Ficht, W. A. Greenberg, C.-H. Wong, *Angew. Chem., Int. Ed.* **2008**, 47, 4411.

[49] P. Wang, S. J. Danishefsky, *J. Am. Chem. Soc.* **2010**, 132, 17045.

[50] P. Wang, X. Li, J. Zhu, J. Chen, Y. Yuan, X. Wu, S. J. Danishefsky, *J. Am. Chem. Soc.* **2011**, 133, 1597.

[51] T. P. King, S. W. Zhao, T. Lam, *Biochemistry* **1986**, 25, 5774.

[52] K. Rose, *J. Am. Chem. Soc.* **1994**, 116, 30.

[53] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, 40, 2004.

[54] J. Shao, J. P. Tam, *J. Am. Chem. Soc.* **1995**, 117, 3893

[55] J. A. Prescher, C. R. Bertozzi, *Nat. Chem. Biol.* **2005**, 1, 13.

[56] C.-F. Liu, J. P. Tam, *J. Am. Chem. Soc.* **1994**, 116, 4149.

[57] C.-F. Liu, J. P. Tam, *Proc. Natl. Acad. Sci. U. S.A.* **1994**, 91, 6584.

[58] C.-F. Liu, C. Rao, J. P. Tam, *J. Am. Chem. Soc.* **1996**, 118, 307.

[59] X. Li, H. Y. Lam, Y. Zhang, C. K. Chan, *Org. Lett.* **2010**, 12, 1724.

[60] T. Wieland, E. Bokelmann, L. Bauer, H. U. Lang, H. Lau, *Justus Liebigs Ann. Chem.* **1953**, 583, 129.

[61] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B, H. Kent, *Science* **1994**, 266, 776.

[62] J. P. Tam, Y. Lu, C.-F. Liu, J. Shao, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 12485.

[63] Johnson, E. C. B. Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, 128, 6640.

[64] T. M. Hackeng, J. H. Griffin, P. E. Dawson, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 10068.

[65] T. Durek, V. Y. Torbeev, S. B. H. Kent, *Proc. Natl. Acad. Sci. U.S.A.* **2007**, 104, 4846.

[66] J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier, J. J. Ottesen, *J. Mol. Biol.* **2011**, 408, 187.

[67] C. F. W. Becker, C. L. Hunter, R. Seidel, S. B. H. Kent, R. S. Goody, M. Engelhard, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 5075.

[68] S. Lahiri, M. Brehs, D. Olschewski, C. F. W. Becker, *Angew. Chem., Int. Ed.* **2011**, 50, 3988.

[69] D. Bang, S. B. H. Kent, *Angew. Chem., Int. Ed.* **2004**, 43, 2534.

[70] D. Bang, S. B. H. Kent, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 102, 5014.

[71] S. Ueda, M. Fujita, H. Tamamura, N. Fujii, A. Otaka, *ChemBiochem* **2005**, 6, 1983.

[72] F. Mende, O. Seitz, *Angew. Chem., Int. Ed.* **2011**, 50, 1232

[73] X. Q. Li, T. Kawakami, S. Aimoto, *Tetrahedron Lett.* **1998**, 39, 8669.

[74] X. Z. Bu, G. Y. Xie, C. W. Law, Z. H. Guo, *Tetrahedron Lett.* **2002**, 43, 2419.

[75] J. D. Warren, J. S. Miller, S. J. Keding, S. J. Danishefsky, *J. Am. Chem. Soc.* **2004**, 126, 6576.

[76] P. Botti, M. Villain, S. Manganiello, H. Gaertner, *Org. Lett.* **2004**, 6, 4861.

[77] J. S. Zheng, H. K. Cui, G. M. Fang, W. X. Xi, L. Liu, *ChemBioChem* **2010**, 11, 511.

[78] T. Kawakami, M. Sumida, K. Nakamura, T. Vorherr, S. Aimoto, *Tetrahedron Lett.* **2005**, 46, 8805.

[79] K. Nakamura, M. Sumida, T. Kawakami, T. Vorherr, S. Aimoto, *Bull. Chem. Soc. Jpn.* **2006**, 79, 1773.

[80] K. Nakamura, H. Mori, T. Kawakami, H. Hojo, Y. Nakahara, S. Aimoto, *Int. J. Protein Pept. Res.* **2007**, 13, 191.

[81] F. Nagaike, Y. Onuma, C. Kanazawa, H. Hojo, A. Ueki, Y. Nakahara, *Org. Lett.* **2006**, 8, 4465.

[82] Y. Ohta, S. Itoh, A. Shigenaga, S. Shintaku, N. Fujii, A. Otaka, *Org. Lett.* **2006**, 8, 467.

[83] S. Tsuda, A. Shigenaga, K. Bando, A. Otaka, *Org. Lett.* **2009**, 11, 823.

[84] H. Hojo, *Tetrahedron Lett.* **2007**, 48, 25.

[85] G. Zanotti, F. Pinnen, G. Lucente, *Tetrahedron Lett.* **1985**, 26, 5481.

[86] T. Kawakami, S. Aimoto, *Tetrahedron Lett.* **2007**, 48, 1903.

[87] T. Kawakami, S. Aimoto, *Tetrahedron* **2009**, 65, 3871.

[88] J. Kang, J. P. Richardson, D. Macmillan, *Chem. Commun.* **2009**, 4, 407.

[89] W. Hou, X. Zhang, F. Li, C.-F. Liu, *Org. Lett.* **2011**, 13, 386.

[90] N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain, O. Melnyk, *Org. Lett.* **2010**, 12, 5238.

[91] J. Dheur, N. Ollivier, A. Vallin, O. Melnyk, *J. Org. Chem.* **2011**, 76, 3194.

[92] R. Ingenito, E. Bianchi, D. Fattori, A. Pessi, *J. Am. Chem. Soc.* **1999**, 121, 11369.

[93] Y. Shin, K. A. Winans, B. J. Backes, S. B. H. Kent, J. A. Ellman, C. R. Bertozzi, *J. Am. Chem. Soc.* **1999**, 121, 11684.

[94] N. Ollivier, J. B. Behr, O. El-Mahdi, A. Blanpain, O. Melnyk, *Org. Lett.* **2005**, 7, 2647.

[95] F. Mende, O. Seitz, *Angew. Chem., Int. Ed.* **2007**, 46, 4577.

[96] J. B. Blanco-Canosa, P. E. Dawson, *Angew. Chem., Int. Ed.* **2008**, 47, 6851.

[97] A. P. Tofteng, K. K. Sorensen, K.W. Conde-Frieboes, T. Hoeg-Jensen, K. J. Jensen, *Angew. Chem., Int. Ed.* **2009**, 48, 7411.

[98] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. U.S.A.* **1998**, 95, 6705.

[99] J. P, Tam, Q. Yu, *Biopolymers* **1998**, 46, 319.

[100] J. Offer, *Biopolymers* **2010**, 94, 530.

[101] L. E. Canne, S. J. Bark, S. B. H. Kent, *J. Am. Chem. Soc.* **1996**, 118, 5891.

[102] D. W. Low, M. G. Hill, M. R. Carrasco, S. B. H. Kent, P. Botti, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98, 6554.

[103] J. Offer, C. N. C. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, 124, 4642.

[104] B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2006**, 45, 4116.

[105] G. Chen, J. D. Warren, J. Chen, B. Wu, Q. Wan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, 128, 7460.

[106] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, 123, 526.

[107] D. Crich, A. Banerjee, *J. Am. Chem. Soc.* **2007**, 129, 10064.

[108] C. Haase, H. Rohde, O. Seitz, *Angew. Chem., Int. Ed.* **2008**, 47, 6807.

[109] J. Chen, Q. Wan, Y. Yuan, J. Zhu, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2008**, 47, 8521.

[110] J. Chen, P. Wang, J. Zhu, Q. Wan, S. J. Danishefsky, *Tetrahedron* 2010, 66, 2277.

[111] Z. Harpaz, P. Siman, K. S. Ajish Kumar, A. Brik, *ChemBioChem* **2010**, 11, 1232.

[112] Z. Tan, S. Shang, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2010**, 49, 9500.

[113] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *J. Am. Chem. Soc.* **2009**, 131, 13592.

[114] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *Chem. Commun.* **2010**, 46, 7199.

[115] H. Rodhe, O. Seitz, *Biopolymers* **2010**, 94, 551.

[116] C.-F. Liu, C. Rao, J. P. Tam, *Tetrahedron Lett.* **1996**, 37, 933.

[117] H. Staudinger, J. Meyer, *Helv. Chim. Acta* 1919, 2 635, doi:10.1002/hlca.19190020164.

[118] E. Saxon, C. R. Bertozzi, *Science* **2000**, 287, 2007.

[119] E. Saxon, J. I. Armstrong, C. R. Bertozzi, *Org. Lett.* **2000**, 2, 2141.

[120] B. L. Nilsson, L. L. Kiesling, R. T. Raines, *Org. Lett.* **2000**, 2, 1939.

[121] B. L. Nilsson, L. L. Kiesling, R. T. Raines, *Org. Lett.* **2001**, 3, 9.

[122] B. L. Nilsson, R. J. Hondal, M. B. Soellner, R. T. Raines, *J. Am. Chem. Soc.* **2003**, 125, 5268.

[123] M. B. Soellner, B. L. Nilsson, R. T. Raines, *J. Am. Chem. Soc.* **2006**, 128, 8820.

[124] M. B. Soellner, A. Tam, R. T. Raines, *J. Org. Chem.* **2006**, 71, 9824.

[125] A. Tam, M. B. Soellner, R. T. Raines, *J. Am. Chem. Soc.* **2007**, 129, 11421.

[126] R. Kleineweischede, C. P. R. Hackenberger, *Angew. Chem., Int. Ed.* **2008**, 47, 5984.

[127] J. W. Bode, R. M. Fox, K. D. Baucom, *Angew. Chem., Int. Ed.* **2006**, 45, 1248.

[128] M. A. Flores, J. W. Bode, *Org. Lett.* **2010**, 12, 1924.

[129] L. Ju, J. W. Bode, *Org. & Biomol. Chem.* **2009**, 7**,** 2259.

[130] L. Ju, A. R. Lippert, J. W. Bode, *J. Am. Chem. Soc.* **2008**, 130, 4253.

[131] S. I. Medina, J. Wu, J. W. Bode, *Org. & Biomol. Chem.* **2010**, 8, 3405.

[132] T. Fukuzumi, J. W. Bode. *J. Am. Chem. Soc.* **2009,** 131, 3864.

[133] J. wu, J. Ruiz-Rodriguez, J. M. Comstock, J. Z. Dong, J. W. Bode, Chem. Sci. 2011, DOI: 10.1039/c1sc00398d.

[134] G.-M. Fang, Y.-M. Li, F. Shen, Y.-C. Huang, J.-B. Li, Y. Lin, H.-K. Cui, L. Liu, *Angew. Chem., Int. Ed.* DOI: 10.1002/anie.201100996.

[135] S. He, D. Bauman, J. S. Davis, A. Loyola, K. Nishioka, J. L. Gronlund, D. Reinberg, F. Meng, N. Kelleher, D. G. McCafferty, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 12033.

[136] K. P. Chiang, M. S. Jensen, R. K. McGinty, T. W. Muir, *ChemBioChem* **2009**, 10, 2182.

[137] M. Shogren-Knaak, H. Ishii, J. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson, *Science* **2006**, 311, 844.

[138] M. A. Shogren-Knaak, C. J. Fry, C. L. Peterson, *J. Biol. Chem.* **2003**, 278, 15744.

[139] M. Manohar, A. M. Mooney, J. A. North, R. J. Nakkula, J. W. Picking, A. Edon, R. Fishel, M. G. Poirier, J. J. Ottesen, *J. Biol. Chem.* **2009**, 284, 23312.

[140] R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder, T. W. Muir, *Nature* **2008**, 453, 812.

[141] X. Zhang, F. Li, C.-F. Liu, *Chem. Commun.* **2011**, 47, 1746.

[142] J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier, J. J. Ottesen, *J. Mol. Biol.* **2011**, 408, 187.

[143] C. Chatterjee, R. K. McGinty, J.-P. Pellois, T. W. Muir, *Angew. Chem., Int. Ed.* **2007**, 46, 2814.

Chapter 2: Semisynthesis of N-terminal Tail Acetylated Histone H4

2.1 Introduction

Eukaryotic chromatin is assembled with DNA, histones and chromatin-associated proteins. The basic building blocks of eukaryotic chromatin are nucleosomes which are made up of 147 bp of DNA wrapping around histone octamers. The nucleosomal arrays further fold into different levels of chromatin structures. Each octamer contains two copies of each of the core histones H2A, H2B, H3 and H4. [1-3] The assembly of nucleosome from individual histones is shown in Fig. 2.1. First, dimers of H3-H4 as well as H2A-H2B form, respectively. Two copies of H3-H4 dimer then form a tetramer. Two H2A-H2B dimers associate with the tetramer of H3 and H4 to form the octamer. The DNA then wraps around the octamer to form the final nucleosome.
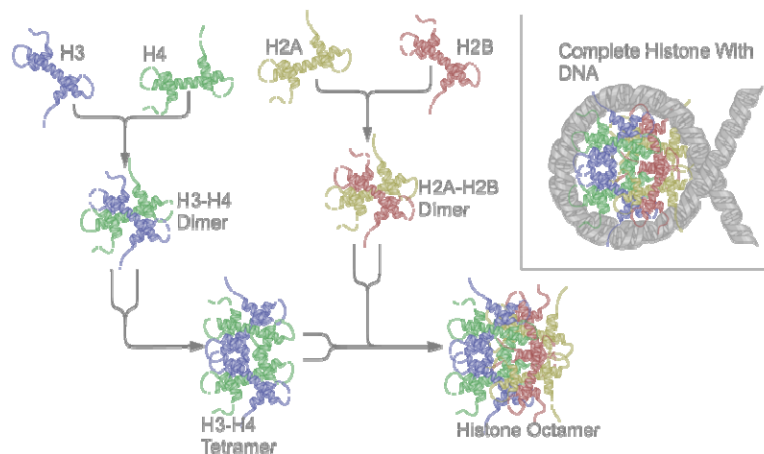
Fig. 2.1 The assembly of nucleosome.

The core histones undergo extensive posttranslational modifications, especially on their N-terminal tails. These modifications include phosphorylation at serine and theronine, methylation at lysine and arginine, acetylation or ubiquitination at lysine (Fig. 2.2).[4] It has

been shown that different modification or combination of modifications play different roles. The histone code hypothesis states that "multiple histone modifications, acting in a combinatorial or sequential fashion on one or multiple histone tails, specify unique downstream functions".[8] The methylation and acetylation of lysine residues in core histones play an important role in chromatin organization, gene expression and epigenetic regulation.[4-7] For instance, histone H3 lysine 4 methylation is associated with transcription activation[9] and the acetylation at histone H4 lysine 16 (H4 K16Ac) modulates both higher chromatin structure and functional interactions between a nonhistone protein and the chromatin fiber.[10] In addition, different histone modifications may be interdependent on each other. The ubiquitination at lysine120 of human H2B has been shown to activate the intranucleosomal methylation of H3 lysine 79 mediated by K79-specific methyltransferase Dot1 (KMT4).[11]
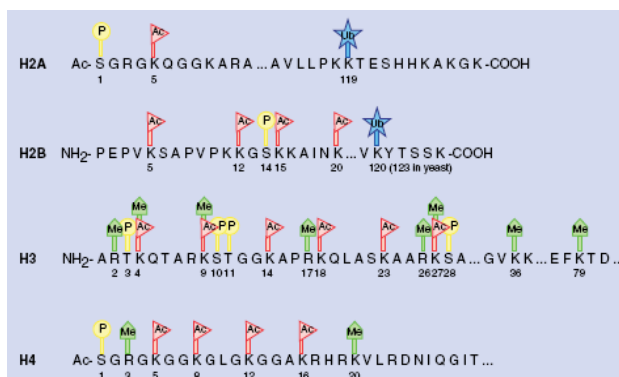
Fig. 2.2 Posttranslational modification map of core histones (Adopted from reference 4).

The different histone modifications form a complicate epigenetic signaling network. To better understand the roles of these modifications, it is important to generate a large amount of the modified proteins in homogeneous form. Until now, several strategies have been developed to achieve the goal.[12] Chemical synthesis of proteins through peptide ligations is

a fundamental approach to synthesize posttranslationally modified proteins, especially histones. Semisynthetic methods employing native chemical ligation (NCL)[13, 14] have been developed to access the modifications at the N-terminal tails of histones.[10, 15-17] This method involves the ligation between a peptide thioester corresponding to the modified histone N-terminal tail and a recombinant histone C-terminal domain with a free N-terminal cysteine (Scheme 2.1A). The modifications at the C-terminal tails of the histones can also been accessed by semisynthetic approach.[11, 18] In this approach, a C-terminal truncated histone thioester was generated through the thiolysis of recombinant histone-intein fusion protein. Then the thioester can ligate with a chemically synthesized C-terminal peptide bearing the desired modifications to generate the full length histone (Scheme 2.1B1). Recently, our group also reported the semisynthesis of C-terminally modified histones employing thiol capture ligation which is different from the thioester mediated native chemical ligation.[19] (Scheme 2.1B2) With this thiol capture ligation approach, histone H3 with C-terminal acetylation(s) was synthesized through the ligation between H3 N-terminal thioacid generated by thiolysis of H3-intein fusion with sodium sulfide or sodium hydrogen sulfide and 5-nitro-2-pyridinesulfenyl (Npys) activated C-terminal cysteine peptide bearing the desired acetylation(s).[19] To access the modifications in the middle of histone, such as the acetylation of lysine 56 of H3, total chemical synthesis through the multiple ligations of chemically synthesized segments was also performed.[20]
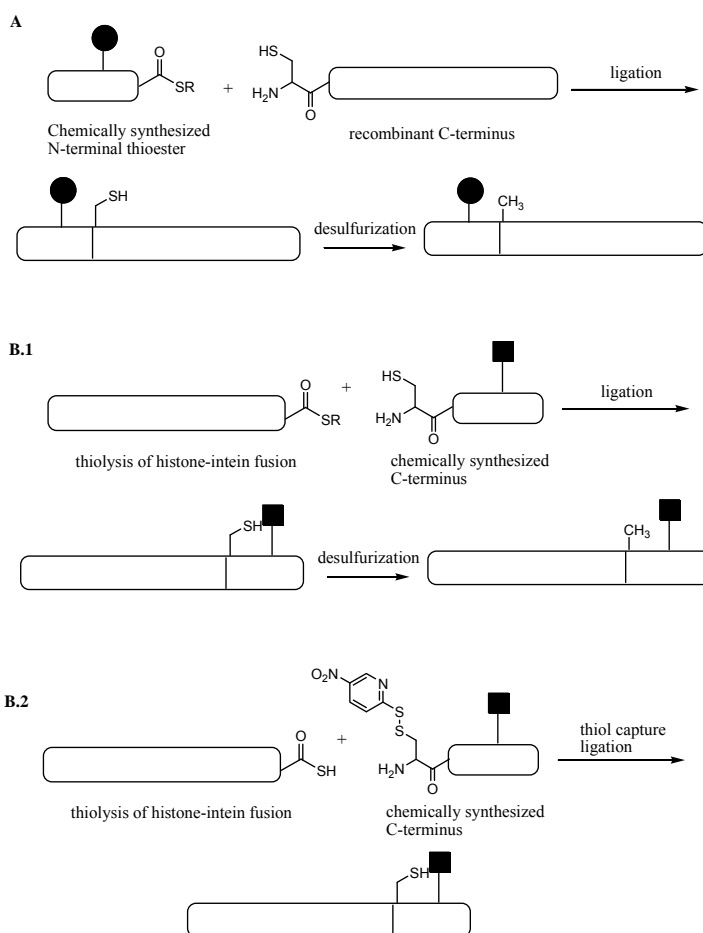
Scheme 2.1. General scheme for the semisynthesis of N-terminal tail (scheme A) or C-terminal tail (scheme B) modified histones using chemical ligation approaches. (● and □ indicate the desired modifications). Sheme A and B.1 show the semisynthetic approach employing thioester mediated ligation. Sheme B.2 shows the synthesis of C-terminal modified histones employing thiol capture ligation.

Other approaches different from chemical ligations were also developed for the preparation of modified histones. Methyllysine analogs can be installed site-specifically into recombinant histones by directly alkylating cysteine residue with haloethylmethyamine.[21] (Scheme 2.2). The method is useful for histone proteins as there are no native cysteine residues in histone H2A, H2B and H4. There is only one cysteine residue close to the C-

terminal tail of H3. The native cysteine residue is mutated to alanine before installing

methyllysine analogs in H3 using the S-alkylation approach. Methyl- and acetyl- lysine

analogs can also been incorporated site-specifically into histones through oxidizing

phenylselenocysteine (to dehydroalanine) containing recombinant histones followed by

Michael addition to dehydroalanine with N-methylcysteamine or N-acetylcysteamine

(Scheme 2.3).[22] Finally, It has been reported that *Nε*-methyllysine or *Nε*-acetyllysine can be

genetically introduced into histones in *E. coli* in response to an amber nonsense mutation.[23-

27] (Scheme 2.4)

Scheme 2.2: General scheme for the site-specific installation of methyllysine analogs into

recombinant histones.

Scheme 2.3: General scheme for the site-specific installation of methyllysine or acetyllysine analogs

into recombinant histones.

Scheme 2.4: The Nε-acetyl- and Nε-methyl-lysine which have been genetically incorporated into proteins through the suppression of amber stop codon. Since Nε-methyllysine cannot be directly incorporated. Different side-chain protected methyllysine derivatives were used instead. After incorporation, the protecting groups were removed to generate the native monomethylated lysine.

Histone H4 has 102 amino acid residues and is the smallest core histone. Posttranslational modifications mainly occur at the N-terminal tail (about the first 20 residues) of H4. These modifications include the phosphorylation at serine 1, the acetylation at lysine 5, 8, 12, 16 and the methylation at lysine 20. It has been shown that acetylations at different locations play different roles.

The first project of my study was to develop a method for the synthesis of H4 analogs with various acetylation(s) at its N-terminal tails for the study of differential effects of these modifications. Various NCL approaches have been developed to synthesize N-terminal tail-acetylated H4. McCafferty's group has synthesized H4 with combination of acetylations at lysine 5, 8 and 12 using a semisynthetic approach with Ala15Cys as the ligation junction.[16] After ligation, the Cys was converted to Ala through desulfurization. The shortcoming of this ligation-desulfurization approach is that the method cannot access the acetylation at lysine 16 due to the lack of Cys or Ala at the right position for the mediation of ligation. Peterson's group has synthesized H4 with acetylation at lysine 16 by using Arg23Cys as the ligation

junction. After ligation, no further manipulation was performed with Cys remaining at position 23.[10] Although in vitro biophysical studies have shown that the Arg23Cys point mutation does not cause significant deviation from the wild type, the mutation does cause significant change in protein sequence and isoelectric property. Recently, Chin's group has generated H4 K16Ac through the site-specific incorporation of non-natural amino acid acetyllysine through the suppression of amber stop codon using orthogonal tRNA synthetase/tRNA pair.[23] However, the approach is sophisticated and the system is not readily available to many labs. In addition, although the approach works for the efficient incorporation of single acetyllysine, the efficiency of incorporation of multiple acetyllysines for the synthesis of multi-acetylated H4 might not be acceptable. The purpose of my study is to develop a relative simple but effective method for the preparation of H4 with combinations of acetylations at lysine 5, 8, 12 and 16.

The semisynthesis of N-terminal tail acetylated H4 using NCL is challenging due to the lack of cysteine at the tail. Although the "Alanine" ligation enables the access of acetylations at lysine 5, 8 and 12, the synthesis of H4 K16Ac is extremely challenging. Although many other chemical ligation methods have been developed based on a similar principle as NCL, all these ligation strategies are more suitable for peptide/protein chemical synthesis and not applicable to the semisynthesis of N-terminal tail-modified H4.

To extend the scope of NCL and its application in the chemical synthesis of modified histones, we were inspired by the fact that a cysteine residue in a peptide sequence can be specifically alkylated with 2-bromoethylamine and converted to aminoethylcysteine (pseudolysine, **sK**). Aminoethylcysteine is a close lysine analog (Scheme 2.5).[28] It differs from native lysine only at γ position (-S- versus -CH$_2$-). It has been shown that **sK** has close physiochemical properties as native lysine. Based on these facts, we realized that the lysine

residue can be selected as a ligation junction for NCL. To do this, the lysine at the ligation site was first mutated to cysteine to facilitate the ligation. After ligation, the cysteine at the ligation site was converted to **sK** through S-alkylation with 2-bromoethylamine. This NCL/S-alkylation approach is particularly suitable for the semisynthesis of N-terminal tail modified histones as there are plenty of lysines at their N-terminal tails. Here, in this chapter, we report the successful synthesis of the H4 variants containing combinations of acetylation(s) at lysine 5, 8, 12 and 16 with the newly developed NCL/S-alkylation approach.
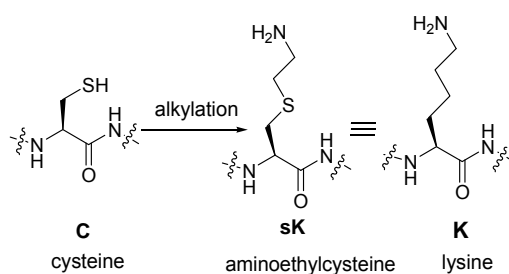
Scheme 2.5. S-alkylation of cysteine residue with 2-bromoethylamine to create aminoethylcysteine, a lysine analog.

To access all of the four acetyllysines at the N-terminal tail of H4, we chose lysine 20 as the ligation junction. Three thioester peptides corresponding to the residue 1 to 19 of H4 and containing the desired acetyllysine(s) were chemically synthesized using SPPS. The C-terminal domain of H4 with lysine 20 mutated to cysteine was generated as a recombinant protein. The two parts were then ligated together through NCL to generate the full length H4 with desired acetyllysine(s). The cysteine at the ligation junction was then converted to **sK** by S-alkylation with 2-bromoethylamine. The general scheme for the synthesis is shown in scheme 2.6. With this two-step method, I synthesized three variants of H4, H4 K16Ac, H4 K(5, 8, 12) Ac_3 and H4 K(5, 8, 12, 16)Ac_4.
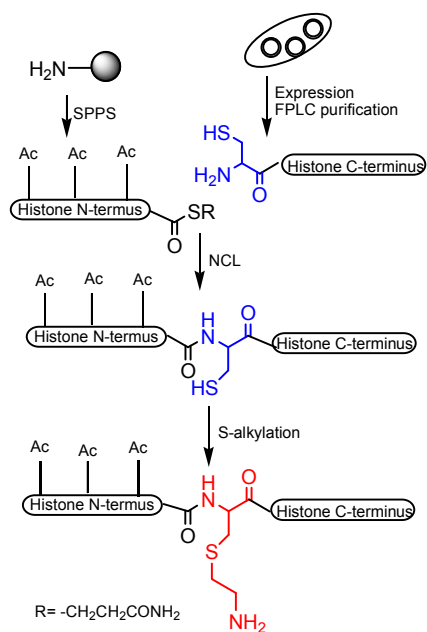
Scheme 2.6. General scheme for the semisynthesis of histones with acetyllysine(s) at their N-terminal tails through NCL/S-alkylation approach.

2.2. Results

2.2.1 Model study using small peptides to test the feasibility of the NCL/S-alkylation approach

Since NCL has been very well demonstrated in peptide synthesis, the remaining concern of the approach is the specificity and degree of S-alkylation. To test the method, we first synthesized two model peptides, thioester peptide **1** and Cys-peptide **2**. As expected, the two peptides ligated efficiently (Fig. 2.3A). To test the S-alkylation step, 0.5 mM (final concentration) of the ligation product **3** was dissolved in alkylation buffer containing 4 M Guanidinium hydrochloride (Gdn•HCl), 1 M 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid (HEPES), 10 mM D/L methionine, 5 mM TCEP, 20 mM DTT, 160 mM 2-bromoethylamine, pH 7.8. After 11 h at dark, **3** was quantitatively alkylated by 2-bromoethylamine (Fig. 2.3B). Only a single peak corresponding to the monoalkylated product was observed. It has been reported that **sK** can direct proteolytic cleavage by lysine-specific protease such as trypsin.[14] To confirm that the S-alkylation was specific at cysteine residue, peptide **3** and its alkylated product peptide **4** were digested with trypsin. Without S-alkylation, trypsin could only cleave after the C-terminal lysine and converted the peptide amide to peptide acid (Fig. 2.3C). However, after S-alkylation, beside the C-terminal lysine, trypsin could also cleave after **sK**. This resulted in two new peaks shown in analytical HPLC (Fig. 2.3D). These observations indicated that the peptide **3** was site- and degree-specifically alkylated at cysteine despite the presence of other functional groups.

1: H-**Leu-Ser-Thr-Glu-Ala**-SCH$_2$CH$_2$CONH$_2$

2: H-**Cys-Ala-Asn-Tyr-Phe-Lys**-NH$_2$

3: H-**Leu-Ser-Thr-Glu-Ala-Cys-Ala-Asn-Tyr-Phe-Lys**-NH$_2$

4: H-**Leu-Ser-Thr-Glu-Ala-sLys-Ala-Asn-Tyr-Phe-Lys**-NH$_2$

5: H-**Leu-Ser-Thr-Glu-Ala-Cys-Ala-Asn-Tyr-Phe-Lys**-OH

6: H-**Leu-Ser-Thr-Glu-Ala-sLys**-OH

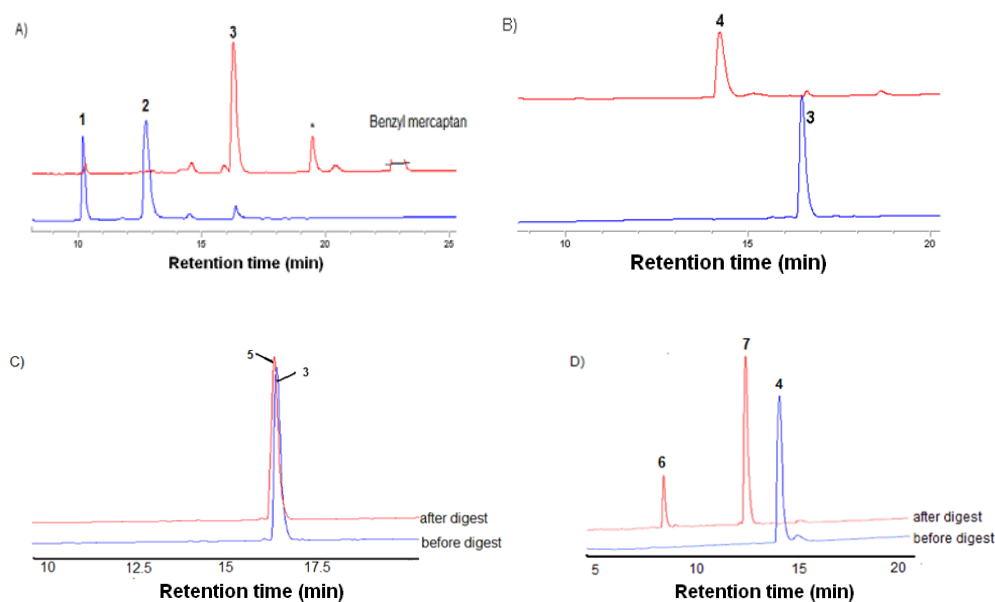**7**: H-**Ala-Asn-Tyr-Phe-Lys**-OH

Figure 2.3. A). Native chemical ligation between peptide **1** and **2**. The asterisk peak is the peptide **1** benzyl mercaptan exchange product. Peak **3** is the ligation product ([M+H]$^+$ m/z calc: 1245.4, found: 1245.5). Conditions: 16.5 mM **1,** 12 mM **2**, 6 M Gdn·HCl, 200 mM phosphate, 20 mM TCEP, 2.5% benzyl mercaptan, pH 8.0, 24 h. B). S-alkylation of peptide **3** with 2-bromoethylamine. Peak **4** is the S-alkylated product ([M+H]$^+$ m/z calc: 1288.5; found: 1288.6). C). Tryptic digest of peptide **3**. ([M+H]$^+$ m/z calc: 1246.4; found: 1246.5). D). Tryptic digest of peptide **4**. Peptide **6**, M+H]$^+$ m/z calc: 665.7; found: 666.3. Peptide **7**, [M+H]$^+$ m/z calc: 641.7, found: 642.4. Conditions for tryptic digest: 0.5 mM **3** or **4**, 100 mM tricine, 2.5 mM TCEP, 1.6 μM of trypsin, pH 8.0, 37 °C, 5 h.

2.2.2 Synthesis of N-terminal tail acetylated H4 variants

2.2.2.1 Synthesis of H4 K16Ac

H4 K16Ac is a prevalent and reversible posttranslational chromatin modification in eukaryotes. It has been shown that H4 K16Ac could inhibit the formation of compact chromatin fibre and inhibit the nucleosomal sliding catalyzed by ACF.[3] As mentioned

earlier, it is challenging to synthesize H4 K16Ac using currently available NCL-based methods. Herein, with our NCL/S-alkylation approach, we successfully synthesized H4 K16Ac. To synthesize H4 K16Ac, we selected lysine 20, the sole lysine at H4 N-terminal tail after position 16, as ligation site. A peptide thioester corresponding to residues 1 to 19 of H4 with acetylated lysine 16, **H4(1-19, K16Ac)-COSCH$_2$CH$_2$CONH$_2$ (8a)**, was manually synthesized using Boc-based SPPS. Multiple side chain protected Boc amino acids were used during the synthesis. Acetyllysine 16 was introduced directly using Boc-Lys(Ac)-OH. Other non-acetylated lysines were introduced using Boc-Lys(2-Cl-Z)-OH. The other side chain protected Boc amino acids used were Boc-Ser(Bzl)-OH, Boc-His(Dnp)-OH and Boc-Arg(Tos)-OH. We also tried different side chain protected derivatives for His and Arg. It turned out that His(Dnp) and Arg(Tos) were the better ones. After synthesis, the Dnp group was first removed by multiple times of treatment with excess of *p*-thiocresol/N-ethyldiisopropylamine (DIEA) in *N,N*-dimethylformamide (DMF). The Boc group of the final amino acid was removed by 30% TFA in DCM before HF cleavage. After HF cleavage, the crude peptide thioester was immediately purified with preparative HPLC. The purified product was analyzed with C18 analytical HPLC and electrospray ionization mass spectrometry (ESI-MS) (Fig. 2.4). The desired product was lyophilized.
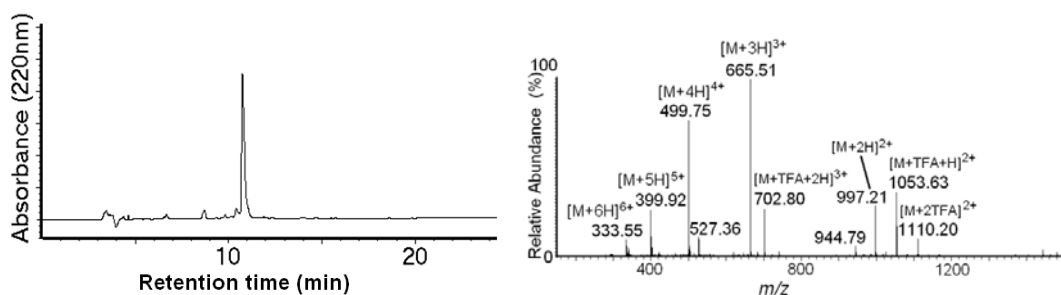

Fig. 2.4 The C18 analytical HPLC (left panel) and ESI-MS spectrum (right panel) of peptide thioester **8a**. HPLC gradient: 0-40% buffer B in Buffer A for 20 min. ESI-MS: calculated mass 1992.10 Da, observed [M+2H]$^{2+}$ m/z 997.21. The peak with m/z 944.79 is a double charged species corresponding to a 104.52 Da less than the calculated mass. The peak is possible due to the

formation of lactum between δ-amine of arginine side chain and C-terminal carbonyl group. However, since analytical HPLC showed a shark peak, the -104 Da peak is possibly formed during ESI measurement.

Next, the H4 C-terminal domain with residue 20 to 102 and lysine 20 mutated to cysteine (**H4(20-102)K20C,** peptide **9**) was expressed in *E. Coli* and purified by gel filtration chromatography. In the expression vector we constructed, the Cys20 residue immediately follows the initiating methionine without any tags in between. Previous report has shown that the majority of the initiating methionine can be removed after the expression of histones in *E. Coli* when the initiating methionine is followed by cysteine.[16] The matrix-assisted laser desorption and ionization-time of flight (MALDI-TOF) mass of our purified protein was 27 Da larger than the expected molecular weight of the Met removed product. This was due to adduct resulted from acetaldehyde reacted with the N-terminal free cysteine in *E. Coli*. To free the cysteine, the purified protein was dissolved in a buffer containing 6 M Gdn·HCl, 0.2 M phosphate, 0.4 M methoxylamine, pH 4.0. After 1 h at 37 °C, the protein was purified by dialysis against water. The MALDI-TOF mass of the protein showed the correct molecular weight of the expected product (Fig 2.5).
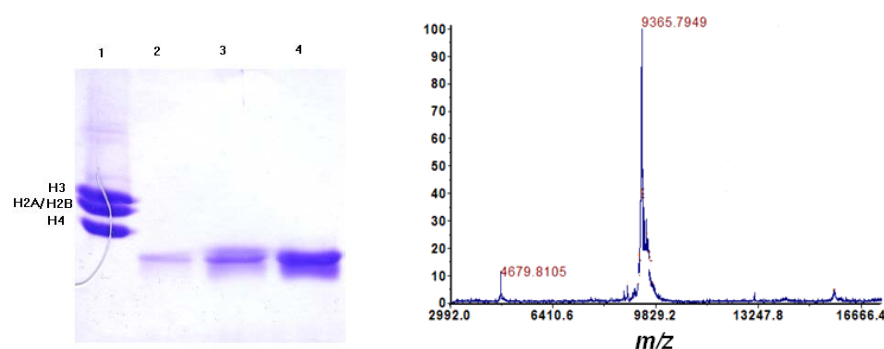

Fig. 2.5 Characterization of purified **9**. Left panel: 18% SDS-PAGE of gel-filtration purified **9**. Lane 1, wild-type octamer as protein marker; Lane 2, 3 and 4, different elution fractions with pure **9**. Right panel: MALDI-TOF MS of methoxylamine treated **9**. Observed mass 9365.8 Da, calculated mass 9365 Da.

To generate the full-length H4, thioester **8a** was ligated with truncated protein **9** under the condition for native chemical ligation. In a typical ligation reaction, about 20 mg of the thioester and 20 mg of the truncated protein were dissolved in 1.5 mL buffer containing 6 M Gdn·HCl, 0.2 M phosphate, 40 mM TCEP, pH 8.5. 1 % v/v of benzyl mercaptan was then added as the thiol additive. The ligation reaction was monitored with C4 analytical HPLC. The reaction proceeded efficiently. Within 15 min, the peak corresponding to ligation product which was eluted slightly earlier could be visualized on the HPLC profile. After overnight reaction, the majority of the truncated protein was ligated with the thioester and formed the full length product. (Fig. 2.6) The ligation product was separated from the unreacted protein by C4 semi-preparative HPLC. The ligation product was analyzed with MALDI-TOF mass which showed the correct molecular weight and the ligation product was free of truncated protein (Fig. 2.7). After lyophilization, 10 mg of ligation product **H4(K16Ac, K20C)** (**10a**) was isolated (yield 42%).
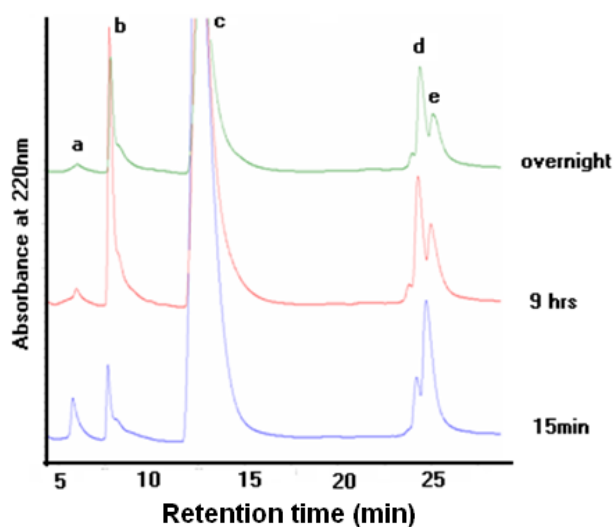

Fig. 2.6 Time course of the ligation between **8a** and **9** monitored with C4 analytical HPLC. Gradient: 0-60% buffer B in buffer A for 30 min. peak a, **8a**; b, H4(1-19, K16Ac)-COSBz; c, benzyl mercaptan; d, ligation product, **10a**; e, H4 C-terminus, **9**.
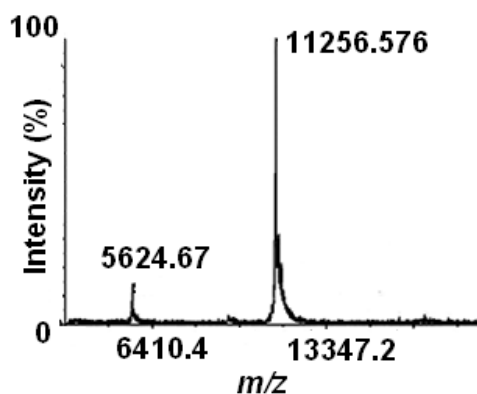
Fig. 2.7 MALDI-TOF MS of the purified full length **10a**. Observed mass 11256.58 Da, calculated mass 11253.02 Da.

To convert the cysteine residue at the ligation site to **sK**, the ligation product **10a** was alkylated with 2-bromoethylamine under denaturing conditions for 12 h. The alkylation product **H4(K16Ac)** (**11a**) was purified by dialysis against water. The molecular weight of the final product was confirmed with MALDI-TOF mass (data not shown). After lyophilization, 8 mg of **11a** was obtained. Due to the low resolution of MALDI-TOF, it is difficult to resolve the monoalkylated product with starting material as well as any possible overalkylated products. To determine whether the alkylation was finished and no overalkylation occurred, the ligation product and alkylation product were analyzed with high resolution FT-MS. The deconvoluted FT-MS spectrum indicated that almost all the **10a** was monoalkylated (Fig. 2.8).
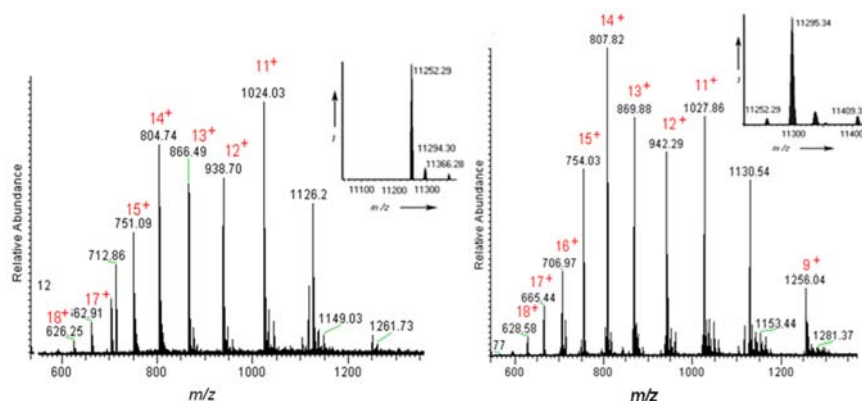
Fig. 2.8 FT-MS of unalkylated (left) and S-alkylated (right) **10a**. The inserts showed the deconvoluted ESI spectrum. ESI-MS, before S-alkylation, observed mass 11252.29 Da, calculated mass 11253.02 Da; After S-alkylation, observed mass 11295.34 Da, calculated mass 11296.09 Da.

2.2.2.2 Synthesis of H4(K5, 8, 12Ac$_3$)

The method for the synthesis of H4 with triacetylation at K5, 8 and 12 was similar to that of H4 K16Ac. We also chose lysine 20 as ligation junction and synthesized a peptide thioester corresponding to the residue 1 to 19 of H4 with acetylation at K5, 8 and 12, **H4(1-19, K5, 8, 12Ac$_3$)-COSCH$_2$CH$_2$CONH$_2$ (8b)**. The SPPS and post assembly manipulations of this thioester were similar to **8a**. The crude was immediately purified after HF cleavage. The purity and identity of the thioester were confirmed with C18 analytical HPLC and ESI-MS (Fig. 2.9).
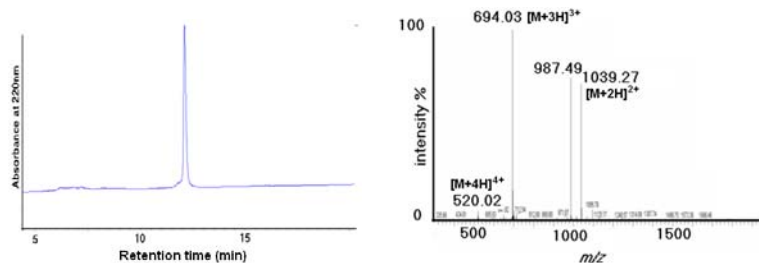
Fig. 2.9 Characterization of thioester **8b**. Left panel, C18 analytical HPLC; gradient, 0-40% buffer B in buffer A for 20 min. Right panel, ESI-MS of purified thioester. Observed [M+2H]$^{2+}$ m/z 1039.27, calculated mass 2076.12 Da. The peak with m/z 987.49 is a double charged species corresponding to

a 103.14 Da less than the calculated mass. The peak is possible due to the formation of lactum between δ-amine of arginine side chain and C-terminal carbonyl group. However, since analytical HPLC showed a shark peak, the 987.49 peak is possibly formed during ESI measurement.

The full length **H4(K5, 8, 12Ac₃, K20C)** (**10b**) was generated by NCL between the thioester **8b** and the H4 C-terminal domain **9** under denaturated conditions in the presence of a thiol catalyst (benzyl mercaptan). The reaction progress was monitored with C4 analytical HPLC. The analytical HPLC showed that the ligation product in this reaction is not very well separated from the unreacted **9** compared with the ligation reaction for the generation of H4 K16Ac (Fig. 2.10). This is due to the more hydrophobic nature of the tri-acetylated H4 compared to the mono-acetylated one. For the better separation between the ligation product and the unreacted **9**, the ligation yield must be high. In our reaction, we could get some fractions which contain only the ligation product (Fig. 2.11, left panel) after purification with C4 semi-preparative HPLC.
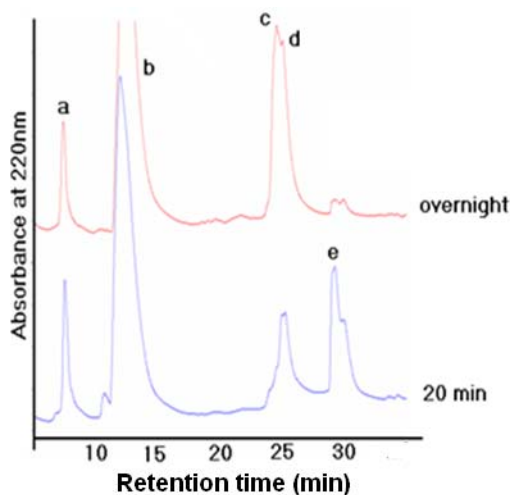
Fig. 2.10 Time course of the ligation between **8b** and **9** monitored with C4 analytical HPLC. Gradient: 0-60% buffer B in buffer A for 30 min. Peak a, **8b**; b, benzyl mercaptan; c, ligation product, **10b**; d, **9**. Peak e could not be identified by MS and was reducible, possible the dimer of benzyl mercaptan due to oxidation.

To convert the Cys 20 to **sK**, **10b** was S-alkylated with 2-bromoethylamine for 12 h. After S-alkylation, the alkylated product **H4(5, 8, 12Ac₃) (11b)** was purified through dialysis. The MALDI-TOF data showed that the protein was mono-alkylated (Fig. 2.11).
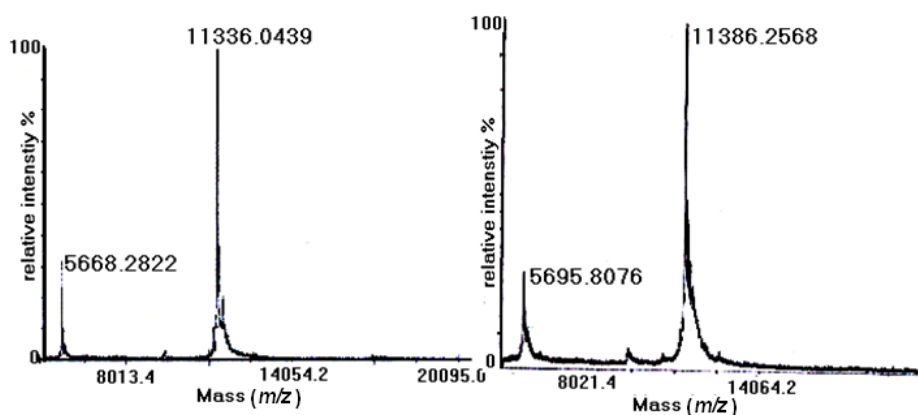
Fig. 2.11 MALDI-TOF MS of **10b** (left panel, observed mass 11336.04 Da, calculated mass 11337.16 Da) and **11b** (right panel, observed mass 11386.27 Da, calculated mass 11380.29 Da).

2.2.2.3 Synthesis of H4 K(5, 8, 12, 16)Ac₄

Besides the mono- and tri-acetylated H4, we also synthesized the tetra-acetylated H4 variant, **H4 K(5, 8, 12, 16)Ac₄** (**11c**). To synthesize **11c**, a thioester peptide, **H4(1-19, K5, 8, 12, 16Ac₄)-COSCH₂CH₂CONH₂ (8c)**, with residue 1 to 19 and tetra-acetyllysine at residue 5, 8, 12 and 16 was first synthesized by SPPS. The synthesis and post-assembly manipulations were the same as the synthesis of the previous two H4 N-terminal thioesters. The C18 analytical HPLC and ESI-MS of the purified thioester were shown in Fig. 2.12.
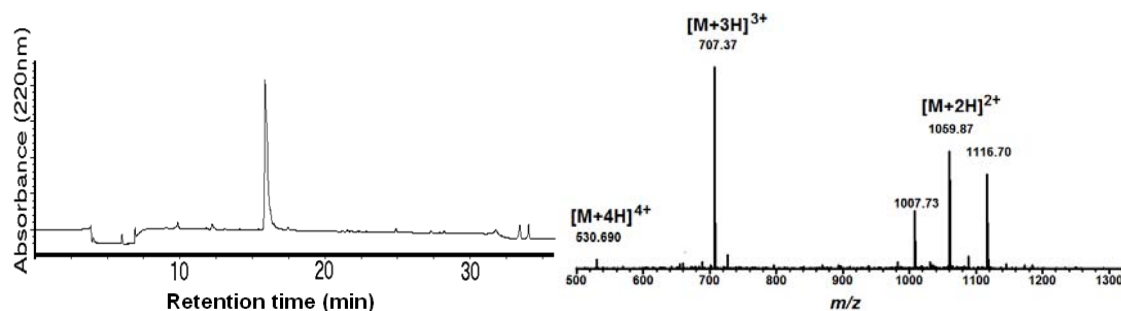
Fig. 2.12 C18 analytical HPLC (left panel) and ESI-MS (right panel) of purified thioester **8c**. HPLC gradient: 0-40 % of buffer B in buffer A for 20 min. Observed $[M+2H]^{2+}$ m/z 1059,87, calculated mass 2118.13 Da. The peak with m/z 1007.73 is a double charged species corresponding to a 104.67 Da less than the calculated mass. The peak is possible due to the formation of lactum between δ-amine of arginine side chain and C-terminal carbonyl group. However, since analytical HPLC showed a shark peak, the 1007.73 peak is possibly formed during ESI measurement.

To synthesize full length tetraacetylated H4, the ligation was performed between the thioester **8c** and the H4 C-terminal domain **9** under denaturated conditions in the presence of a thiol catalyst (benzyl mercaptan). In a typical ligation reaction, about 20 mg of thioester **8c** and 20 mg of **9** were dissolved in 1.5 to 2 mL of ligation buffer (6 M Gdn·HCl, 0.2 M phosphate, 40 mM TCEP, 1 % v/v of benzyl mercaptan, pH 8.5). The ligation product **H4 (5, 8, 12, 16Ac₄, K20C)** (**10c**) and unreacted **9** could not be separated on C4 analytical HPLC (data not shown). After overnight reaction, the full length H4 was separated from the truncated one with C4 semi-preparative HPLC. The collected fractions were analyzed by MALDI-TOF. The pure fractions (Fig. 2.13, left panel) were combined and lyophilized. About 10 mg of pure ligation product was obtained.

To convert the Cys 20 to **sK**, **10c** was S-alkylated with 2-bromoethylamine for 12 h. After S-alkylation, the alkylated product **H4(5, 8, 12, 16Ac₄)** (**11c**) was purified through

dialysis. About 8 mg of final alkylated product was obtained after lyophilization. The MALDI-TOF data of the final alkylated product was shown in Fig. 2.13, right panel.
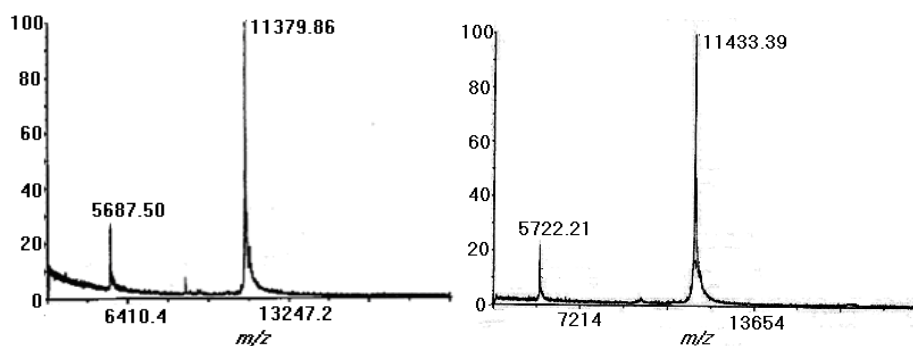
Fig. 2.13 Left panel: MALDI-TOF MS of **10c**. [M+H]$^+$ m/z observed: 11379.86, calculated: 11380. Right panel: MALDI-TOF MS of **11c**. [M+H]$^+$ m/z observed: 11433.39, calculated: 11423.

2.3. Materials, Methods and Experimental Procedures

2.3.1 Materials and general methods

Benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP), S-trityl-mercaptopropionic acid, 4-methylbenzhydrylamine (MBHA) resin, Rink Amide-MBHA resin, Boc-Lys(Ac)-OH, Boc-Arg(Mts)-OH, Boc-His(Dnp)-OH.IPA, Boc-Ser(Bzl)-OH, Fmoc-Cys(Trt)-OH, Fmoc-Lys(Boc)-OH were obtained from GL Biochems (Shanghai, China). Other Boc amino acids and Fmoc amino acids were purchased from Novabiochem. The side-chain protection groups used were as follows: Boc-Ser(Bzl)-OH, Boc-Thr(Bzl)-OH, Boc-Glu(OBzl)-OH, Boc-Lys(2-Cl-Z)-OH, Boc-His(Dnp)-OH·IPA, Boc-Arg(Tos)-OH for Boc chemistry and Fmoc-Cys(Trt)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Lys(Boc)-OH for Fmoc chemistry. DMF, dichloromethane (DCM), diethyl ether, and HPLC-grade acetonitrile (ACN) were purchased from Fisher. TFA, DIEA, benzyl mercaptan, piperidine, triisopropylsilane

(TIS), trifluoromethanesulfonic acid (TFMSA) were obtained from Alpha Alster (NJ, USA). TCEP•HCl was purchased from Fluka.

All the analytical HPLC were done using an Agilent 1100 series HPLC equipped with a Jupiter C18 (5 μm, 4.6 x 250 mm) or Vydac C4 (5 μm, 4.6 x 250 mm) reverse-phase column at a flow rate of 1.0 ml/min. The purifications involving semi-preparative HPLC column were done with a Shimadzu system equipped with a Vydac C18 (5 μm, 10 x 250 mm) or C4 (10 μm, 10 x 250 mm) at a flow rate of 2.5 ml/min. Preparative purifications were done using a Waters system equipped with an Altech prosphere C18 column (10 μm, 22 x 250 mm) at a flow rate of 10 ml/min. The buffer system for all the analysis was buffer A - H_2O (containing 0.045% TFA) and buffer B - 90% acetonitrile in H_2O (containing 0.045% TFA).

The ESI masses of peptides were measured using Thermo FINNIGAN LCQ Deca XP MAX equipped with ESI ion source. MALDI-TOF MS spectra were acquired on an Applied biosystems 4800 MALDI-TOF/TOF analyzer in linear mode. MALDI-TOF MS analysis of proteins was performed by mixing a sample of 0.5 μL HPLC eluted proteins and 0.5 μL sinapinic acid solution (saturated in aqueous acenitrile containing 0.1% TFA). High resolution FT-MS was performed on a Thermo LTQ FT ULTRA system.

2.3.2 Solid-phase peptide synthesis

Peptide thioester **1** (H-**Leu-Ser-Thr-Glu-Ala**-$SCH_2CH_2CONH_2$) was synthesized manually according to Boc chemistry on MBHA resin on a 0.2 mmol scale. S-trityl-mercaptopropionic acid was loaded onto the resin and served as the thiol moiety. The trityl group was removed by repeated treatment with TFA/β-mercaptoethonal/TIS/DCM (5/2.5/2.5/90) until no yellow color appeared when treated with TFA/β-mercaptoethonal/DCM (5/2.5/92.5). For the coupling of amino acids, the Boc protected

amino acid (4 eq.), pyBOP (4 eq.) was dissolved in DCM. DIEA (12 eq.) was added to the solution. After 2 min of preactivation, the solution was added into the resin. It takes about one and a half hours for each coupling. The coupling efficiency was monitored with Kaiser test. The Boc group was removed by 30% TFA in DCM. After assembly, the peptide was cleaved from the resin with TFMSA/TFA/p-cresol/methylphenyl sulfide (1:7:1:1). The crude peptide was purified with C18 semi-preprative HPLC column with a gradient of 50% buffer B in buffer A for 50 min.

Peptide **2** (H-**Cys-Ala-Asn-Tyr-Phe-Lys**-NH$_2$) was assembled manually according to the standard 9-fluorenylmethoxycarbonyl- (Fmoc) based SPPS on a rink amide MBHA resin with a 0.2 mmol scale. Before coupling, the Fmoc group was removed using 20% piperidine in DMF for 2 min followed by 20 min. The resin was washed with 2xDMF, 2xDCM and 2xDMF. For the coupling of amino acids, the Fmoc protected amino acid (4 eq.), pyBOP (4 eq.) were dissolved in DMF. DIEA (8 eq.) was then added to the solution. After 2 min of preactivation, the solution was added into the resin. The coupling reaction was undertaken for about one and a half hours. The coupling efficiency was monitored with Kaiser test. After the assembly was finished, the Fmoc protection group was removed and the peptide was cleaved from the resin with TFA/TIS/H$_2$O (95/2.5/2.5) for 1 h. The crude peptide was precipitated with ether and purified with C18 semi-preparative HPLC with a gradient of 50% buffer B in buffer A for 50 min.

The H4 N-terminal thioesters **H4 (1-19, K16Ac)-COSCH$_2$CH$_2$CONH$_2$ (8a)**, **H4(1-19, K5, 8, 12Ac$_3$)-COSCH$_2$CH$_2$CONH$_2$ (8b)** and **H4(1-19, K5, 8, 12, 16Ac$_4$)-COSCH$_2$CH$_2$CONH$_2$ (8c)** were synthesized manually using Boc-based SPPS. MBHA resin was used. Trt-SCH$_2$CH$_2$COOH was loaded onto the resin and severed as the thiol linker. Acetyllysine(s) was introduced by direct loading with commercially available Boc-Lys(Ac)-

OH. The other side chain protected amino acids used in the synthesis were Boc-Arg(Tos)-OH, Boc-His(Dnp)-OH·IPA, Boc-Lys(2-Cl-Z)-OH, Boc-Ser(Bzl)-OH. After assembly, the Dnp protection group was removed with excess thiolanisole/DIEA in DMF. The peptide thioester was then cleaved from the resin with HF/p-cresol/anisole (9/0.5/0.5) for 1.5 h at 0°C. The crude peptide was purified with HPLC immediately after cleavage and its identity was confirmed by ESI-MS.

2.3.3 Model study for the demonstration of NCL/S-alkylation approach

NCL between peptide **1** and **2**. 1 mg of peptide **1** (1.65 μmol, 16.5 mM) and 0.9 mg of peptide **2** (1.2 μmol, 12 mM) were dissolved in 100 μL ligation buffer (6 M Gdn·HCl, 200 mM phosphate, 20 mM TCEP, 2.5 % benzyl mercaptan, pH 8.0). The ligation reaction was undertaken at room temperature for 24 h and checked with C18 analytical HPLC using the gradient 0-40% buffer B in buffer A for 20 min. The ligation product (peptide **3**, H-**Leu-Ser-Thr-Glu-Ala-Cys-Ala-Asn-Tyr-Phe-Lys**-NH$_2$) was purified with C18 semi-preparative HPLC with a gradient of 0-50% buffer B in buffer A for 50 min.

S-alkylation of peptide **3**. Conditions: 0.5 mM of peptide **3** was dissolved in S-alkylation buffer (4 M Gdn·HCl, 1 M HEPES, 10 mM D/L methionine, 5 mM TCEP, pH 7.8). 20 mM DTT and 160 mM 2-bromoethylamine were then added into the solution. DTT and 2-bromoethylamine were dissolved immediately before use. The S-alkylation was undertaken in dark for 11 h. The reaction was monitored with analytical HPLC. At the end, the S-alkylated product (peptide **4**, H-**Leu-Ser-Thr-Glu-Ala-sLys-Ala-Asn-Tyr-Phe-Lys**-NH$_2$) was purified with C18 semi-preparative HPLC.

Typtic digest of peptide **3** and **4**. 0.5 mM of peptide was dissolved in digest buffer (100 mM tricine, 2.5 mM TCEP, pH 8.0). 1.6 μM of trypsin was added into the mixture. The

digestion was undertaken at 37 °C for 5 h. The process of the digestion was monitored with C18 analytical HPLC and ESI-MS.

2.3.4 Synthesis of H4 variants

Generation of H4 C-terminal domain **9**. *Xenopus laevis* histone H4 gene was amplified by polymerase chain reaction (PCR) using the forward primer 5'-GGAATTCCATATGTGCGGTCTGCGTGACAAC-3' and reverse primer 5'-CCCGGATCCTTAACCACCGAAACCGTACAGGGT-3'. The amplified segment was digested with NdeI and BamHI and incorporated into plasmid pET-3a which has been digested with the same restriction enzymes. BL21(DE3) pLysS cells harboring the expression plasmid were grown in 7.5 L 2xTY media containing 100 mg/L ampicillin and 25 mg/L chloroamphenical at 37 °C. The cells were induced with 0.4 mM Isopropyl β-D-1-thiogalactopyranoside IPTG (final concentration) when OD600 = 0.6. After 3 h, the cells were harvested by centrifugation at 7,000 g for 7 min. the following steps was done at 4 °C. Harvested cells were resuspended in washing buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM β-mercaptoethanol, pH 7.5) and lyzed by sonication. The cell lysate was centrifuged at 23, 000g for 20 min. the pellet was washing twice with washing containing 1% Triton X-100 and once with washing buffer. 1 mL of DMSO was added into the pellet. The pellet was stirred and periodically votexed for 30 min at room temperature. 50 ml of SAU-200 buffer was added and stirred for 3h. The solution was centrifuged at 23000g for 20 min, room temperature. The supernatant was collected and subjected to gel filtration FPLC purification. The eluted fractions were checked with 18% SDS-PAGE. The fractions containing pure H4(20-102)K20C were combined and dialyzed against Milli-Q water containing 4 mM beta-mercaptoethanol. The identity of the purified protein was confirmed with MALDI-TOF MS and the protein was lyophilized.

Ligation between H4 N-terminal thioesters and **9**. In a typical ligation, 10.5 mg (5.3 μmol, 5.3 mM) of **8a** and 15 mg (1.1 μmol, 1.1 mM) of **9** were dissolved in 1mL ligation buffer (6 M Gdn•HCl, 0.2 M phosphate, 20 mM TCEP, 1.5% benzyl mercaptan, pH 8.0). The reaction was monitored with analytical HPLC and stopped at 12 h. The full length **10a** was purified with semi-preparative HPLC. The molecular weight of the product was confirmed with MALDI-TOF MS.

For the synthesis of **10b**, 12.7 mg (6.1 μmol, 6.1 mM) of **8b** and 25 mg (2.67 μmol, 2.67 mM) of **9** were dissolved in 1mL ligation buffer (6M Gdn•HCl, 0.2M phosphate, 20mM TCEP, 1.5% benzyl mercaptan, pH 8.0). The reaction was monitored with analytical HPLC and stopped at about 12 hours. The full length **10b** was purified with semi-preparative HPLC. The molecular weight of the product was confirmed with MALDI-TOF MS. The synthesis of **10c** was similar to the synthesis of **10a** and **10b**.

S-alkylation of **10a**, **10b** and **10c**. The S-alkylation of histone H4 was similar to that of peptide **3**. After 12 hrs of reaction, the alkylation product **11a**, **11b** and **11c** were purified through dialysis against water. Their identities were confirmed with MALDI-TOF MS. The purified H4(K16Ac) was lyophilized and redissolved in H_2O/ACN/HCOOH (50/49.9/0.1) and subjected to high-resolution FT-MS.

References:

[1] K. Luger, A. W. Mader, R. K. Richmond, D. F. Sargent, T. J. Richmond, *Nature* **1997**, 389, 251.
[2] J. M. Harp, B. L. Hanson, D. E. Timm, G. J. Bunick, *Acta Cryst. Sect. D* **2000**, 56, 1513.
[3] C. A. Davey, D. F. Sargent, K. Luger, A. W. Maeder, T. J. Richmond, *J. Mol. Biol.* **2002**, 319, 1097.
[4] C. L. Peterson, M. A. Laniel, *Curr. Biol.* **2004**, 14, R546.
[5] S. L. Berger, *Nature* **2007**, 447, 407.
[6] J. C. Rice, C. D. Allis, *Curr. Opin. Cell. Biol.* **2001**, 13, 263.
[7] B. M. Turner, *Bioassays* **2000**, 22, 836.
[8] T. Jenuwein, C. D. Allis, *Science* **2001**, 293, 1074.
[9] B. D. Strahl, R. Ohba, R. G. Cook, C. D. Allis, *Proc. Natl. Acad. Sci. U.S.A.* **1999**, 96, 14967.

[10] M. Shogren-Knaak, H. Ishii, J. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson, *Science* **2006**, 311, 844.

[11] R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder, T. W. Muir, *Nature* **2008**, 453, 812.

[12] C. Chatterjee, T. W. Muir, *J. Biol. Chem.* **2010**, 11045.

[13] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B, H. Kent, *Science* **1994**, 266, 776.

[14] J. P. Tam, Y. Lu, C.-F. Liu, J. Shao, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 12485.

[15] M. A. Shogren-Knaak, C. J. Fry, C. L. Peterson, *J. Biol. Chem.* **2003**, 278, 15744.

[16] S. He, D. Bauman, J. S. Davis, A. Loyola, K. Nishioka, J. L. Gronlund, D. Reinberg, F. Meng, N. Kelleher, D. G. McCafferty, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 12033.

[17] K. P. Chiang, M. S. Jensen, R. K. McGinty, T. W. Muir, *ChemBioChem* **2009**, 10, 2182.

[18] M. Manohar, A. M. Mooney, J. A. North, R. J. Nakkula, J. W. Picking, A. Edon, R. Fishel, M. G. Poirier, J. J. Ottesen, *J. Biol. Chem.* **2009**, 284, 23312.

[19] X. Zhang, F. Li, C.-F. Liu, *Chem. Commun.* **2011**, 47, 1746.

[20] J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier, J. J. Ottesen, *J. Mol. Biol.* **2011**, 408, 187.

[21] M. D. Simon, F. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, K. M. Shokat, *Cell* **2007**, 128, 1003.

[22] J. Guo, J. Wang, J. S. Lee, P. G. Schultz, *Angew. Chem., Int. Ed.* **2008**, 47, 6399.

[23] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, 4, 232.

[24] D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *J. Am. Chem. Soc.*, **2009**, 131, 14194.

[25] Y.-S. Wang, B. Wu, Z. Wang, Y. Huang, W. Wan, W. K. Russell, P.-J. Pai, Y. N. Moe, D. H. Russell, W. R. Liu, *Mol. BioSystems* **2010**, 6, 1557.

[26] H.-W. Ai, J. W. Lee, P. G. Schultz, *Chem. Commun.* **2010**, 46, 5506.

[27] D. Groff, P. R. Chen, F. B. Peters, P. G. Schultz, *ChemBioChem* **2010**, 11, 1066.

[28] G. L. Kenyon, T. W. Bruice, *Methods. Enzymol.* **1977**, 47, 407.

Chapter 3: Dual Native Chemical Ligation at Lysine: the Method for the Synthesis of Ubiquitinated Peptides

3.1 Introduction

Ubiquitin is a 76 amino acid protein, which is highly conserved in eukaryotes (Fig. 3.1). Ubiquitination is one of the most important protein posttranslational modifications.[1] It refers to the linking of the C-terminal glycine of the ubiquitin protein to the lysine side chain of the modified proteins. Proteins can undergo monoubiquitination or polyubiquitination.[1] As a well-known example of monoubiquitination, the modification at the C-terminal region of histone H2A is associated with gene silencing.[2,3] The monoubiquitination at the C-terminus of H2B was shown to be associated with transcription elongation.[4] In the case of polyubiquitination, a chain consisting of multiple ubiquitins with defined linkage is attached to the lysine side chain of a modified protein target.[5] Ubiquitin contains seven lysine residues (K6, 11, 27, 29, 33, 48, 63) and each of these lysine residues can be employed for ubiquitin chain formation. The biological roles of polyubiquitin may be related to its linkage specificity.[5] K48-linked polyubiquitin is the best understood polyubiquitination and is shown to serve as a signal for targeting the protein to proteasomal degradation.[6] In contrast, K63-linked polyubiquitination is involved in a signaling pathway and associated with kinase activation.[7]

MQIFV **K**[6]TLTG **K**[11]TITL EVEPS DTIEN V**K**[27]A**K**[29]I
QD**K**[33]EG IPPDQ QRLIF AG**K**[48]QL EDGRT LSDYN
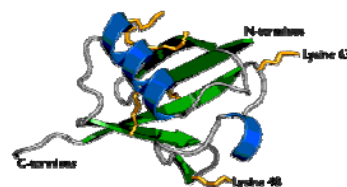IQ**K**[63]ES TLHLV LRLRGG

Fig. 3.1 The amino acid sequence and structure of ubiquitin. The seven lysine residues in the sequence are highlighted.

To elucidate the biological roles of ubiquitination, it is important to isolate sufficient amount of ubiquitinated proteins for in vitro studies. In eukaryotes, ubiquitination is catalyzed by three classes of enzymes, ubiquitin-activating enzymes E1, ubiquitin-conjugating enzymes E2 and ubiquitin ligase E3.[1] (Fig. 3.2) In vitro enzymatic synthesis of ubiquitinated proteins is limited by the requirement for identifying the ligases and the availability of the ligase.
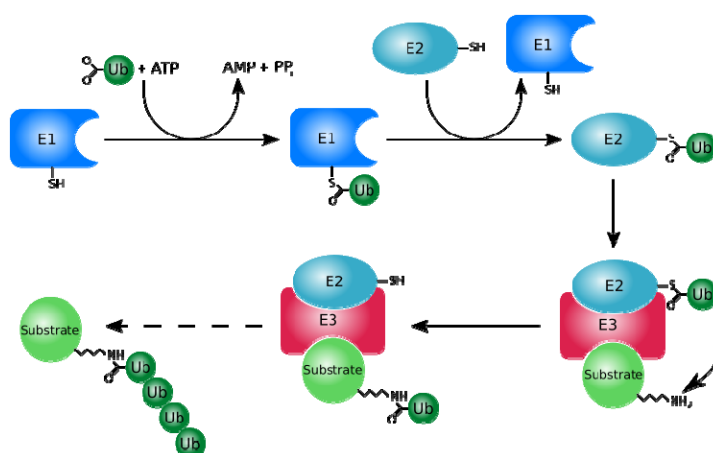
Fig. 3.2 Schematic diagram showing the enzymatic ubiquitination catalyzed by E1, E2 and E3. The first step is the ATP-dependent activation of ubiquitin by ubiquitin-activating enzyme E1. Ubiquitin C-terminus is linked to the cysteine residue of active site of E1 through a thioester linkage. Ubiquitin is then transfer to cysteine residue at the active site of ubiquitin-conjugating enzyme E2. Ubiquitin ligase E3 is responsible for substrate recognition. E2 with ubiquitin linked to its active site interacts with substrate bonded E3 and E3 catalyzed the linkage of the ubiquitin to the specific lysine of substrate protein. After multiple cycles occurred, a polyubiquitin chain can be formed on the substrate protein.

Due to the limitations of in vitro enzymatic ubiquitination, artificial ubiquitination through chemical approach has drawn more and more attention to chemical biologists in recent years. Unlike enzymatic ubiquitination, chemical ubiquitination is not restricted by substrate specificity. However, initially it was extremely challenging when chemical ubiquitination

started to emerge as an interesting research field. In 2007, Muir T. W.'s team reported the first chemical ubiquitination approach through auxiliary-mediated chemical ligation.[8] (Scheme 3.1) In their approach, the Gly 76 of ubiquitin with a *Nα*-photoremovable auxiliary was installed on the side chain of lysine residue where the ubiquitin was going to attached. Auxiliary-mediated chemical ligation between the ubiquitin thioester containing remaining 75 amino acids and Gly 76 was then performed to link the ubiquitin to the peptide. However, the ligation process is extremely slow due to the steric nature of secondary amino group where auxiliary is attached. It is urged to develop more efficient chemical ubiquitination methods.
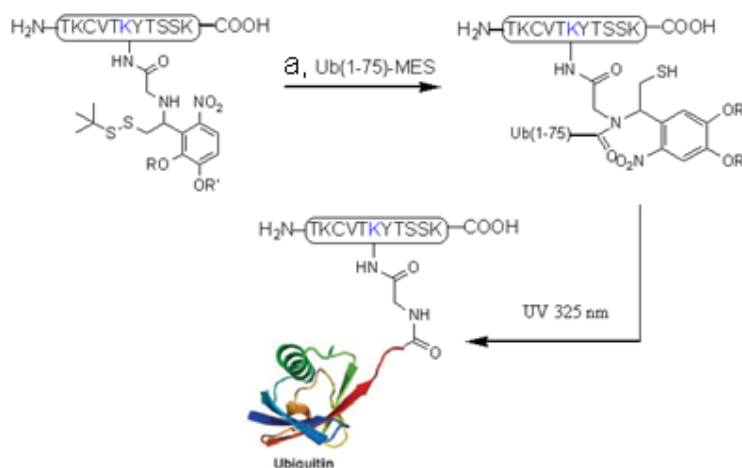
Scheme 3.1 General scheme of the Nα-auxiliary mediated site-specific peptide ubiquitination.

Condition: a, 300 mM NaPi, 50 mM MESNa, 25 mM TCEP, 3 M Gdn•HCl, pH 7.5.

Chemical ligation methods, especially the notable native chemical ligation[9, 10], have been shown to be promising methods for the synthesis of proteins, especially those with posttranslational modifications. Chemical ligations are also options for chemical ubiquitination. In the past two decades, many chemical ligation methods have been developed. However, non-native peptide bonds are formed with many of these ligation methods. Native chemical ligation which refers to the reaction between a peptide thioester

and another peptide with N-terminal cysteine enables the formation of native peptide bonds between two peptides. The method requires a cysteine residue at the ligation junction. Unfortunately, cysteine is not a common amino acid in naturally occurring proteins. To overcome the cysteine restriction, Nα-auxiliary mediated ligations were developed.[11-16] A removable thiol containing auxiliary attached to the α-amino group of the N-terminal amino acid was employed to facilitate the ligation with thioester. After ligation, the auxiliary group was removed by physiochemical approaches, such as acid treatment or UV irradiation.[11-16] However, the steric hindrance resulted from the auxiliary attachment significantly lowered the ligaiton rate. So far, the methods were only reported to be effective when less bulky residues such as glycine are present at the ligation junction. Another approach to extend the scope of chemical ligation is ligation-desulfurization.[17-24] An amino acid surrogate with a thiol group at β- or γ-position of side chain was used to replace the native part. The thiol can be used to facilitate the chemical ligation. After ligation, the thiol group can be removed by desulfurization. The ligation-desulfurization approach is traceless. With this strategy, the chemical ligation approach has been extended to many non-cysteine residues, such as alanine[17,] phenylalanine[18], valine[19, 20], theronine[21], leucine[22, 23], and so on.

Although many chemical ligation methods have been developed, not too many of them can be applied to the chemical synthesis of ubiquitinated peptides and proteins. As mentioned earlier, Muir T. W. reported the first way of chemical ubiquitination through auxiliary-mediated ligation.[8] However, the approach is not efficient. All the other available methods at that time were not applicable for the synthesis of ubiquitinated peptides or proteins with native Gly-Lys(ε) isopeptidic linkage.

The objective of my research project was to develop new method for efficient chemical ubiquitination. We were inspired by Muir T. W.'s strategy and the ligation-desulfurization

strategy that others have applied for expanding the scope of chemical ligation. Since the auxiliary attached at amine of Gly 76 produced a secondary amine and slowed down the ligation, we realized that we can put the auxiliary within the lysine side chain and free the amino group involved in ligation. In addition, the very large auxiliary employed to facilitate the ligation can be replaced by a small thiol group. Based on these two points, we designed a thiolysine derivative with a thiol group at the γ-position of lysine side chain. The γ-thiol is located at the right position of the lysine side chain where it can facilitate the ligation of both the α- and ε-amino group of lysine with thioesters, both through a six-member ring intermediate (Scheme 3.2). With this novel strategy, we can expand the scope of native chemical ligation to lysine, which is a very common residue in naturally occurring proteins. Most importantly, the strategy can be applied for chemical ubiquitination. When an ubiquitin thioester is reacted with the ε-amine of the 4-thiolysine, an isopeptide bond can be formed to link the ubiquitin to lysine side chain. Therefore, the approach can serve as a method for chemical ubiquitination.
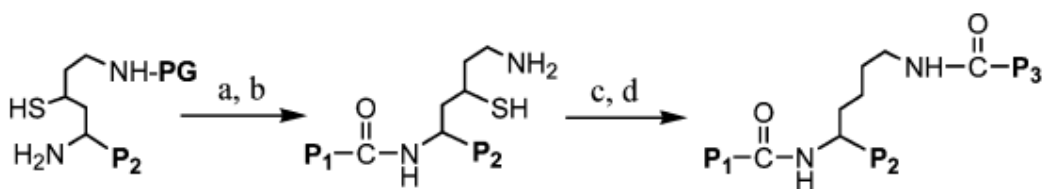

Scheme 3.2 General scheme for dual native chemical ligation at lysine: (a) First ligation with **P**1-COSR; (b) Deprotection; (c) Second ligation with **P**3-COSR; (d) Desulfurization.

3.2 Results

3.2.1 Chemical synthesis of 4-thiolysine derivative

4-thiolysine derivatives are novel lysine derivatives and are not commercial available. Therefore, we have to chemically synthesize them. The chemical synthesis of 4-thiolysine derivative **3-6** was done through the collaboration with Mr. Kalyan Kumar Pasunooti from Dr. Liu Xue-wei's lab. The chemical synthesis of **3-6** was mainly performed by him. Scheme 3.3 shows the synthesis of the protected 4-thiolysine analogue **3-6**. Essentially, the method of Guichard and co-workers[25] was adopted to prepare a 4-hydroxylysine derivative of which the 4-hydroxyl group was subsequently converted to a thiol.
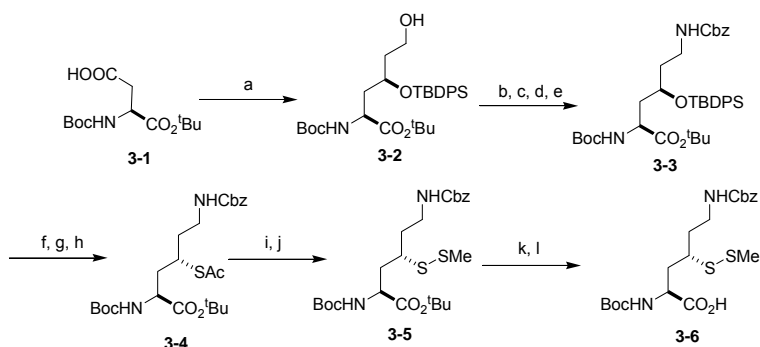
Scheme 3.3 Reagents and conditions: (a) **Ref. 25**; (b) Ms-Cl, DIPEA, 0 °C; (c) NaN$_3$, DMF, 80 °C, two steps 83% yield; (d) H$_2$, Pd/C, ethyl acetate, rt, quantitative yield; (e) Cbz-Cl, NaHCO$_3$, dioxane:water (2:1), 0 °C, 81%; (f) TBAF, THF, 0 °C, 77%; (g) Ms-Cl, DIPEA, 0 °C; (h) CH$_3$COSK, DMF, 40 °C, 70% over two steps; (i) NaOH, MeOH, rt; (j) *S*-methyl methanethiosulfonate (MMTS), triethylamine, CH$_2$Cl$_2$, rt, 50% over two steps; (k) 95% TFA, H$_2$O, rt; (l) Boc$_2$O/TEA, MeOH, rt, 78% over two steps.

Starting from (*S*)-Boc-Asp-O^{t}Bu **3-1**, the O-silyl-protected 1,3-diol **3-2** was prepared in enantiopure form through a key δ-lactam intermediate using known literature procedures. The terminal alcohol in **3-2** was mesylated and further converted to an azide. The azide was then

reduced by catalytic hydrogenation to an amine, which was protected using Cbz-Cl to afford **3-3**. After deprotection of the silyl ether in **3-3** with TBAF, mesylation of the secondary alcohol followed by nucleophilic substitution of thioacetate afforded **3-4** in good yields. Following saponification, the thiol was protected in disulfide form using S-methyl methanethiosulfonate (MMTS) to furnish **3-5** according to literature methods. Acidolytic deprotection with TFA and further protection with Boc anhydride yielded **3-6**, the protected form of 4-thiolysine ready for use in peptide synthesis.

3.2.2 Demonstration of dual NCL at lysine using small peptides

The 4-thiolysine derivative **3-6** was introduced as the N-terminal residue of the small peptide **12** with Cbz remaining on its ε-amine. The analytical HPLC and ESI-MS spectra of purified peptide **12** were shown in Fig. 3.3.
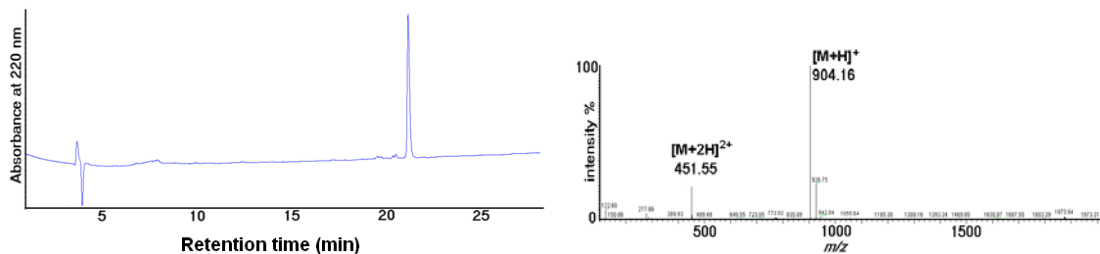
Fig. 3.3 C18 analytical HPLC and ESI-MS analysis of purified peptide **12**. Gradient: Buffer B 0-60% in 30 min. ESI-MS: [M+H]$^+$ m/z observed: 904.16, calculated: 903.42.

Peptide **12** was directly used in ligation reactions, as the free-thiol form **13** could be readily generated *in situ* under the reducing conditions of the ligation reaction (Scheme 3.4). Scheme 3.4 shows the model studies we have performed with short peptides to verify the feasibility of the dual ligation at lysine approach. At the end, we applied the approach to the synthesis of ubiquitinated peptides.
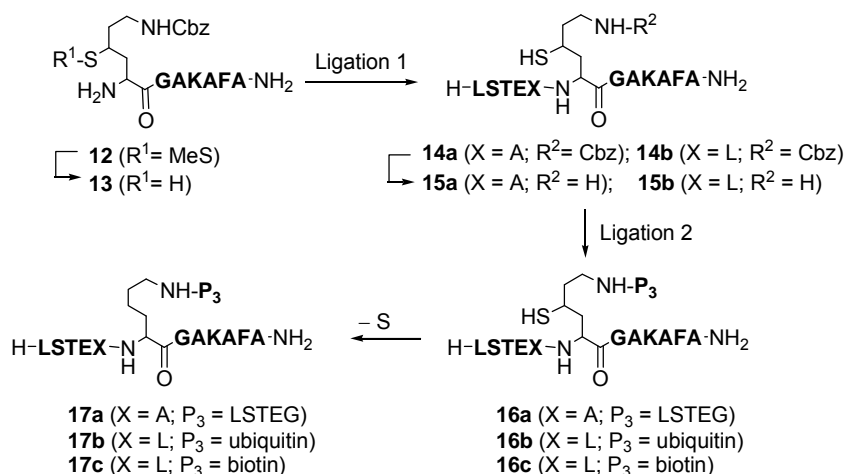
Scheme 3.4 Demonstration of peptide dual ligation.

With the ε-amine protected, we first tested the γ-thiol mediated ligation at the α-amine of 4-thiolysine. Robust ligation was observed when **12** was subjected to reaction with the thioester peptide H-**LSTEA**-COSR at pH 8. After reaction for 1 h at 37 °C, the 4-thiolysyl peptide limiting reactant **13** was completely consumed, and the ligation product **14a** was obtained in ~90% yield based on HPLC analysis. (Fig. 3.4)
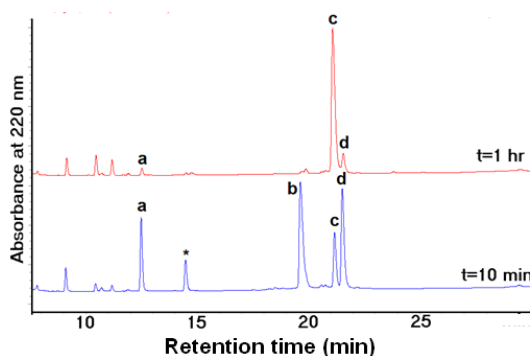
Fig. 3.4 C18 analytical HPLC monitored ligation between peptide thioester H-**LSTEA**-SCH$_2$CH$_2$CONH$_2$ and peptide **12** at 10 min and 1 h, respectively. Gradient: Buffer B 0-60 % in 30 min. Peak a, H-**LSTEA**-SCH$_2$CH$_2$CONH$_2$, ESI-MS: [M+H]$^+$ m/z observed: 607.3, calculated: 607.28; peak b, reduced peptide **12**, K(γ-SH, ε-Cbz)GAKAFA-NH$_2$, ESI-MS: [M+H]$^+$ m/z observed: 858.1, calculated: 857.44; peak c, ligation product **14a**, ESI-MS: [M+H]$^+$ m/z observed: 1359.2, calculated: 1358.68; peak d, H-**LSTEA**-SBz, ESI-MS: [M+H]$^+$ m/z observed: 626.6, calculated: 626.29; peak *, H-**LSTEA**-SMe, ESI-MS: [M+H]$^+$ m/z observed: 550.2, calculated: 550.26.

Ligation of **12** with another thioester peptide with a larger C-terminal Leu residue, H-**LSTEL**-COSR, also proceeded efficiently (Fig. 3.5), yielding the ligation product **14b** in 92% yield by HPLC analysis after reaction for 1 h at 37 °C, indicating that the ligation reaction was not affected by the bulkiness of Leu in either reaction rate or yield.
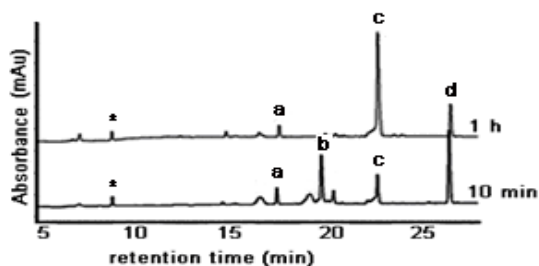
Fig. 3.5 HPLC monitoring of ligation between H-**LSTEL**-COSR and peptide **12** at 10 min and 1 h, respectively. Gradient: 0-60% buffer B in 30 min. Peak a, H-**LSTEL**-COSR; peak b, peptide **13**; peak c, ligation product peptide **14b**; peak d, H-**LSTEL**-COSBz; peak *, H-**LSTEL**-OH. Ligation conditions: 28 mM H-**LSTEL**-COSR, 20 mM **13**, 6 M Gdn·HCl, 0.2 M phosphate, 60 mM TCEP, 1% benzyl mercaptan, pH 8.0, 37 °C.

The γ-thiol mediated 1st step of ligation (ligation at α-amine) was performed at pH 8 and 37 °C. The reaction under different pH and temperature was also studied. (Table 3.1 and Fig. 3.6) We have also performed the ligation between H-**LSTEA**-SCH$_2$CH$_2$CONH$_2$ and peptide **12** at different pH and/or temperature. As mentioned earlier, at pH 8 and 37°C, the ligation was finished within 1 h and the yield was about 90%. At pH 8 and at 23 °C, the reaction gave *ca.* 78% of ligation product after 1.5 h; At pH 7 and at 37 °C, the reaction gave *ca.* 70% ligation product after 1.5 h; At pH 7 and 23°C, the yield of the ligation product was *ca.* 57 % after 1.5 h reaction. After comparing the ligation under different pH and temperatures, we found that increase in either pH (from 7 to 8) or temperature (from 23 °C to 37 °C) could increase the ligation rate and shorten the reaction time (see Fig. 3.6). pH 8 and 37 °C gave the fastest ligation rate. That is the reason we used pH 8 and 37 °C as the conditions in our study.

We also compared the rate of lysine ligation with Cys ligation under pH 7 and 23 ^{o}C. When H-**LSTEA**-SCH$_2$CH$_2$CONH$_2$ was reacted with the cysteinyl peptide H-**CGAKAFA**-NH$_2$ at pH 7 and 23°C, a ligation yield of *ca.* 73% was obtained after 1.5 h reaction. We found that the ligation rate at Cys was faster than Lys ligation under the same conditions. This is reasonable because it is a 1° –SH mediating the ligation for Cys and a 2° –SH for Lys. The more steric –SH at Lys side chain gave slower reaction rate. The ligation rate of Cys ligation under pH 7 and 23 ^{o}C was comparable to that of Lys ligation under pH 8 and 23 ^{o}C.

Time (h)	pH	Temperature (°C)	Yield (%)
1	8	37	90
1.5	8	23	78
1.5	7	37	70
1.5	7	23	57

Table 3.1. Comparison of efficiency of Lys ligation under different pH (7 or 8) and temperatures (23 or 37 ^{o}C).
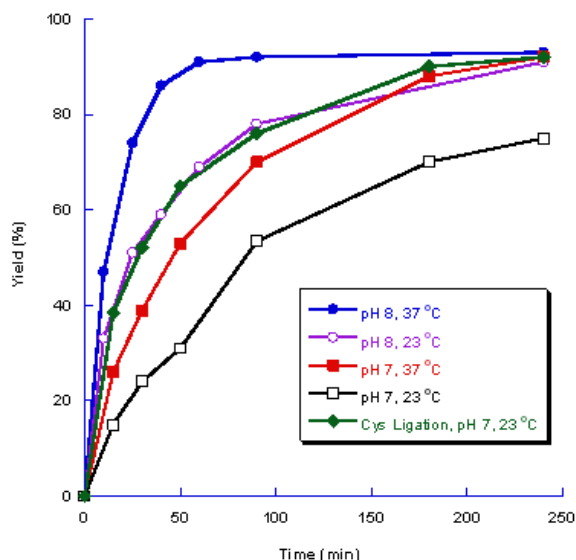
Fig. 3.6 Comparison of efficiency of Lys ligation under different pH (7 or 8) and temperatures (23 or 37 ^{o}C). Experimental conditions: For Lys ligation: 16.5 mM H-**LSTEA**-SCH$_2$CH$_2$CONH$_2$, 11 mM H-**K**(SSMe, Cbz)**GAKAFA**-NH$_2$, 6 M Gdn·HCl, 0.2 M phosphate, 60 mM TCEP, 1 % v/v benzyl

mercaptan; For Cys ligation: 16.5 mM H-**LSTEA**-SCH$_2$CH$_2$CONH$_2$, 11 mM H-**CGAKAFA**-NH$_2$, 6 M

Gdn·HCl, 0.2 M phosphate, 60 mM TCEP, 1 % v/v benzyl mercaptan, pH 7, 23 °C.

Before the 2nd ligation step (the γ-thiol mediated ligation at ε-amine) was performed, the

Cbz group in **14** was first removed using a cocktail containing TFMSA in order to expose the

ε-NH$_2$ of 4-thiolysine. Fig. 3.7 and 3.8 shows the removal of Cbz for peptide **14a** and **14b,**

respectively.
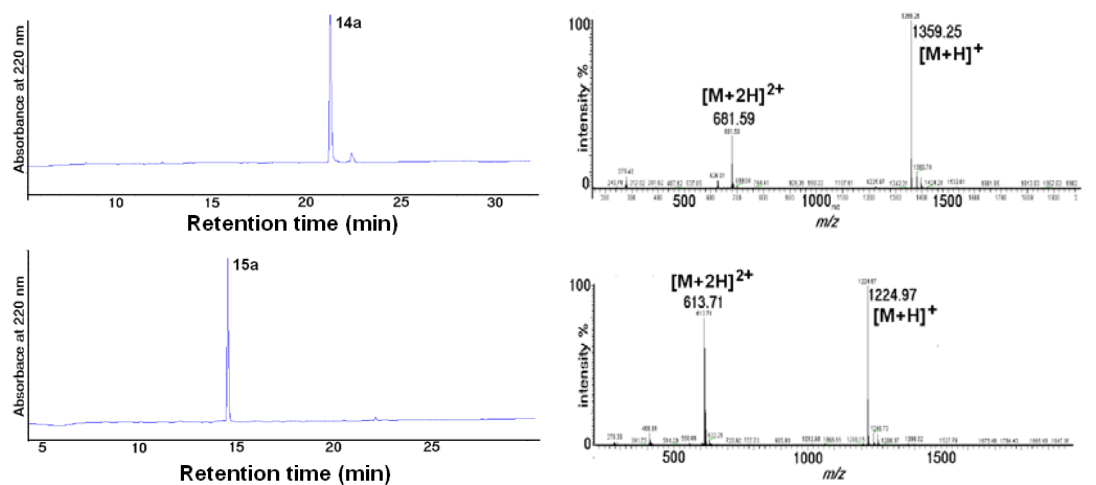
Fig. 3.7 C18 analytical HPLC analysis and ESI-MS of purified peptide **14a** (upper trace) and its

purified Cbz removed product **15a** (bottom trace). Gradient: Buffer B 0-60% in 30 min. ESI-MS for

peptide **14a,** observed [M+H]$^+$ m/z 1359.25, calculated 1358.68; ESI-MS for peptide **15a,** observed

[M+H]$^+$ m/z 1224.97, calculated 1224.64.
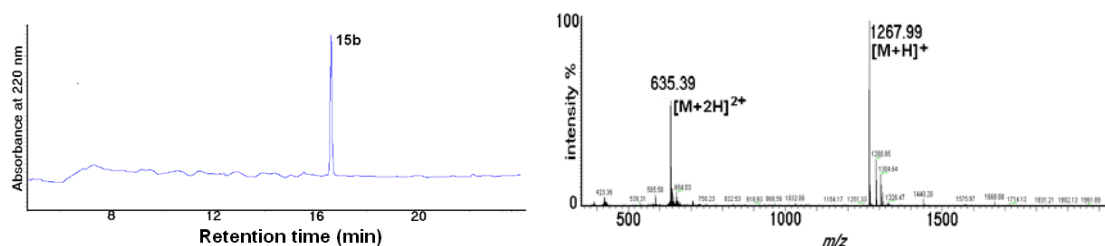
Fig. 3.8 C18 analytical HPLC analysis and ESI-MS of purified peptide **14b** (upper trace) and its purified Cbz removed product peptide **15b** (bottom trace). Gradient: Buffer B 0-60% in 30 min. ESI-MS for peptide **14b**, observed $[M+H]^+$ m/z 1400.79, calculated 1400.73; ESI-MS for peptide **15b**, observed $[M+H]^+$ m/z 1267.99, calculated 1266.69.
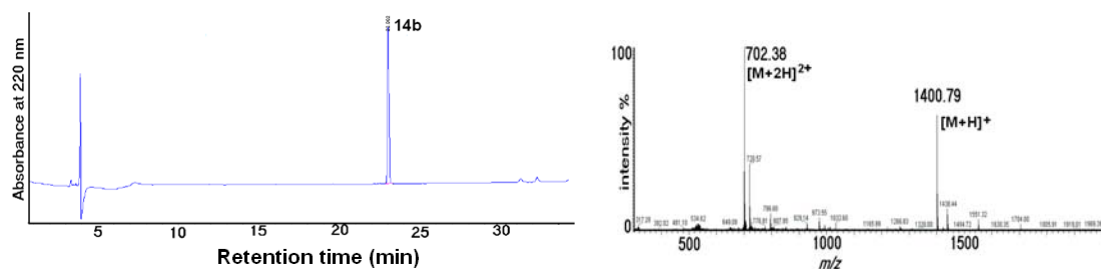
For the 2nd ligation, the purified Cbz-free peptide **15a** was reacted with a small peptide thioester, H-**LSTEG**-COSR. Similar to the first ligation at the α-amine, the second ligation proceeded very efficiently, reaching completion in 1.5 h to give the product **16a** in 92% yield by HPLC analysis. (Fig. 3.9) A control experiment performed with thioester peptide H-**LSTEA**-COSR and lysylpeptide H-**KGAKAFA**-NH$_2$ for 2 h showed no detectable amount (<2%) of direct aminolysis at either the α- or ε-amine. This indicated the vital role of the γ-SH group in mediating ligation at both the *α*- and *ε*-amino groups of the 4-thiolysine residue.
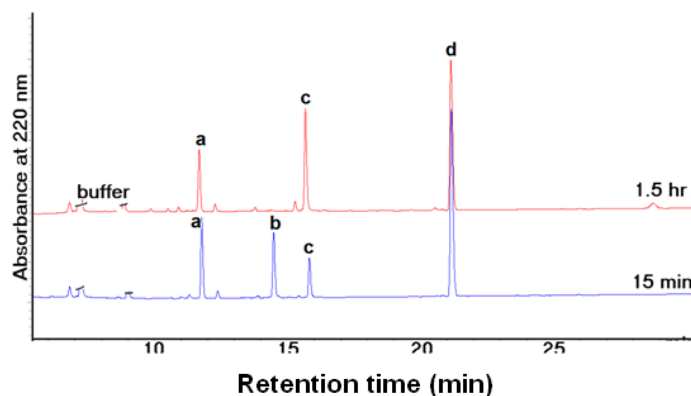
Fig. 3.9 C18 analytical HPLC monitored ligation between peptide **15a** and H-**LSTEG**-SCH$_2$CH$_2$CONH$_2$ at 15 min and 1.5 h, respectively. Gradient: Buffer B 0-60% in 30 min. Peak a, H-**LSTEG**-SCH$_2$CH$_2$CONH$_2$, ESI-MS observed $[M+H]^+$ m/z 593.6, calculated 593.26; peak b, peptide **10a**, ESI-MS observed $[M+H]^+$ m/z 1225.0, calculated 1224.64; peak c, ligation product **16a**, ESI-MS observed $[M+H]^+$ m/z 1711.8, calculated 1711.87; peak d, H-**LSTEG**-SBz, ESI-MS observed $[M+H]^+$ m/z 612.3, calculated 612.27.

To generate native Lys at the ligation junction, we first tried the Raney nickel-mediated desulfurization method.[17] Desulfurization of **16a** gave a moderate yield of 44% after 8 h reaction based on quantitative HPLC analysis. (Fig. 3.10) The yield of desulfurization was determined by comparing the integrated area of the expected product peak and the starting material peak by loading the same sample volume for analytical HPLC analysis before and after desulfurization. The lower of the yield was possibly due to the adsorption of some of the peptide materials to the Raney-nickel.
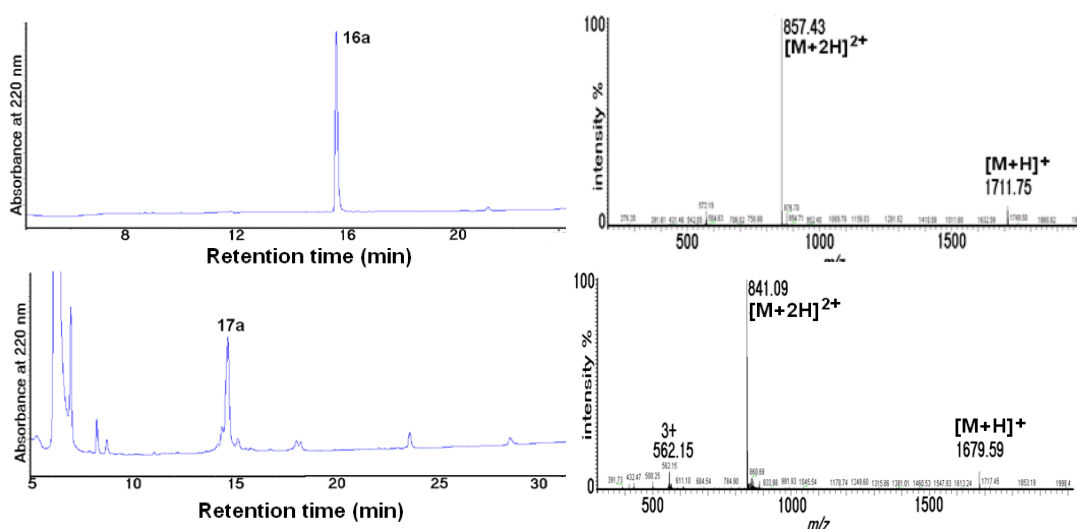
Fig. 3.10 C18 analytical HPLC and ESI-MS of peptide **16a** (upper trace) and its crude desulfurization product **17a** (lower trace). Gradient: Buffer B 0-60% in 30 min. ESI-MS for peptide **16a**: observed [M+H]+ m/z 1711.75, calculated [M+H]+ m/z 1711.87; ESI-MS for peptide **17a**: observed [M+H]+ m/z 1679.59, calculated [M+H]+ m/z 1679.90.

3.2.3 Dual NCL at lysine as an approach for peptide ubiquitination

Next, to verify whether our method can be used for site-specific peptide ubiquitination, peptide **15b** was reacted with ubiquitin thioester, ub(1-76)-COSCH$_2$CH$_2$SO$_3^-$ [ub(1-76)-MES], generated by thiolysis of a ubiquitin-intein fusion protein with MESNa.[26] The analytical HPLC and ESI-MS of purified ub(1-76)-MES were shown in Fig. 3.11.
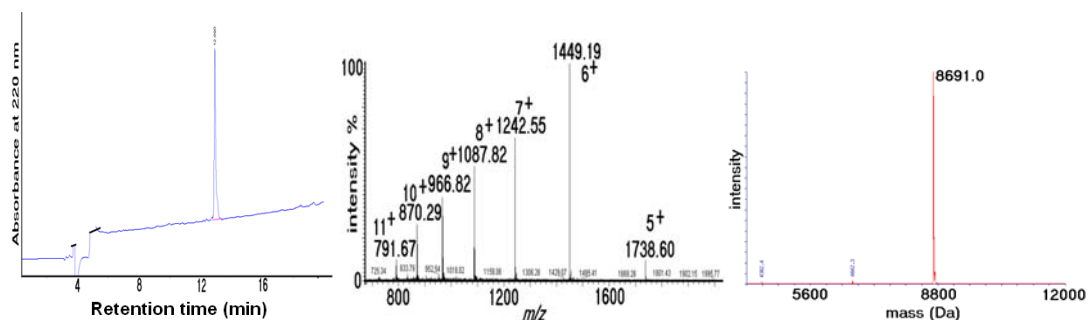
Fig. 3.11 C18 analytical HPLC and ESI-MS of purified ub(1-76)-MES. Gradient: 20-60% buffer B in buffer A for 20 min. Deconvoluted ESI-MS: observed mass 8691.0 Da, calculated mass 8689 Da.

As shown in Fig. 3.12, ligation between **15b** and the ubiquitin thioester was completed within only 45 min, giving a very clean product **16b** in > 90 % yield based on HPLC analysis. The efficiency of the reaction in our dual NCL scheme makes it a particularly viable method for the synthesis of complex protein conjugates such as ubiquitinated proteins for the functional elucidation of such posttranslational modifications on lysine.
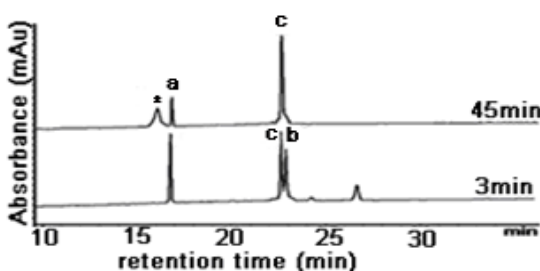
Fig. 3.12 HPLC monitored ligation between peptide **15b** and ub(1-76)-MES at 3 min and 45 min, respectively. Gradient: 0-36% buffer B in 18 min, 36-45% in 18 min. Peak a, peptide **15b**; peak b, ub(1-76)-MES; peak c, ligation product **16b**; peak *, unidentified. Ligation conditions: 6.6 mM **15b**, 1.2 mM ub(1-76)-MES, 6 M Gdn·HCl, 0.2 M phosphate, 60 mM TCEP, pH 8.0, 37 °C.

To generate the native ubiquitin linkage, desulfurization of **16b** was performed. First we tried Raney-Ni mediated desulfurization. However, this metal-based method did not work on **16b** which is much larger and more complex than **16a**. We then tried the recently developed free radical desulfurization approach.[27,19] Desulfurization of **16b** using 2,2'-Azobis[2-(2-

imidazolin-2-yl)propane]dihydrochloride (VA-044) as free radical initiator was completed within 5 h under UV irradiation (365 nm) to give the final ubquitinated peptide **17b** and the conversion was near quantitative based on HPLC and MS analysis. (Fig. 3.13)

Fig. 3.13 C18 analytical HPLC and ESI-MS monitored free radical-mediated desulfurization of peptide **16b**. The lower trace was the purified peptide **16b** dissolved in desulfurization buffer. The upper trace showed the crude desulfurization after 5 h. Peak **17b** was the desulfurization product. HPLC gradient: buffer B 20 - 60% in 20 min. Peptide **16b**: deconvoluted ESI-MS observed mass 9812.0 Da, calculated mass 9811.6 Da; peptide **17b**: deconvoluted ESI-MS observed mass 9780.3 Da, calculated mass 9779.6 Da.

3.2.4 Dual NCL at lysine as an approach for peptide side chain biotinylation

To further test whether our method can be used for site-specific biotinylation on the lysine side chain, peptide **15b** was reacted with a biotin thioester, Biotinyl-SCH$_2$CH$_2$CO-**GFRA**-NH$_2$. The reason we chose a biotinyl thioester linked to a peptide rather than directly using Biotinyl-SCH$_2$CH$_2$CONH$_2$ was for the convenience of solid-phase peptide synthesis and characterization of product by HPLC and ESI MS analysis. The reaction was completed in 3 h to give the biotinylated ligation product **16c** in 90% yield. (Fig. 3.14) Free radical desulfurization afforded the final biotinylated peptide **17c** in 80% yield. (Fig. 3.15)
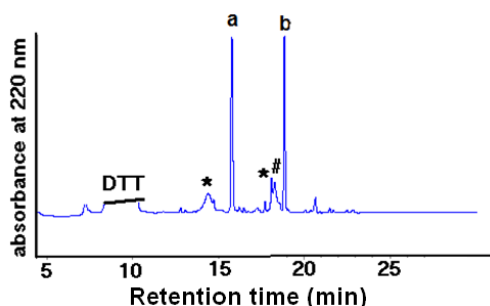
Fig. 3.14 C18 analytical HPLC analysis of ligation between peptide **15b** and Biotinyl-SCH$_2$CH$_2$CO-**GFRA**-NH$_2$ after 3 h reaction. Gradient: Buffer B 0-60% in 30 min. Peak a, HS-CH$_2$CH$_2$CO-**GFRA**-NH$_2$, ESI-MS: observed [M+H]$^+$ m/z 537.6, calculated [M+H]$^+$ m/z 537.26; peak b, peptide **16c**, ESI-MS: observed [M+H]$^+$ m/z 1492.94, calculated [M+H]$^+$ m/z 1492.77; peaks *, unidentified; peak #, incomplete reduction of oxidative adduct between HS-CH$_2$CH$_2$CO-**GFRA**-NH$_2$ and peptide **16c**.

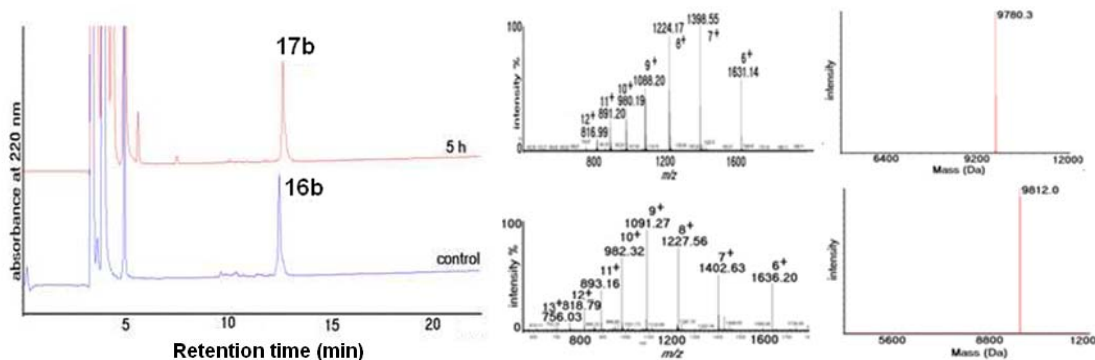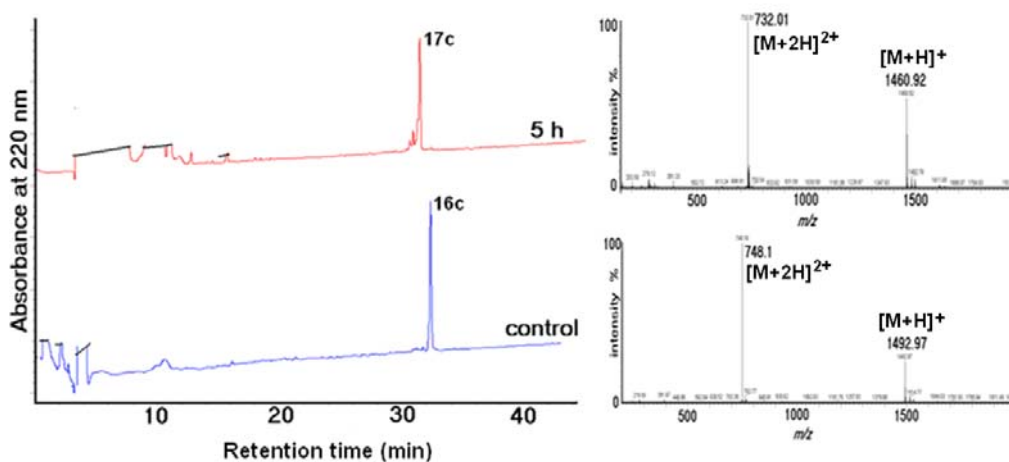**Fig. 3.15** C18 analytical HPLC monitored free radical-mediated desulfurization of peptide **16c**. The lower trace was the purified peptide **16c**. The upper trace showed the desulfurization reaction mixture after 5 h. Peak 17c was the desulfurization product. HPLC gradient: buffer B 0-40% in 40 min. ESI-MS for peptide **16c**: observed [M+H]$^+$ m/z 1492.97, calculated [M+H]$^+$ m/z 1492.77; ESI-MS for peptide **17c**: observed [M+H]$^+$ m/z 1460.92, calculated [M+H]$^+$ m/z 1460.80.

3.3 Conclusion

The above results show that, in a unique one-stone-two-birds fashion, a γ-SH group on an N-terminal lysine mediates facile chemical ligation on both its α- and ε-amine. The

unhindered nature of the two primary amines likely accounts for the robustness of the ligation reactions. If used without the second ligation step, our method would allow conventional linear NCL at Lys, a notable expansion of NCL's application scope considering the abundance of lysine in proteins. When using the dual ligation scheme, it is possible to synthesize complex protein molecules which are acylated on specific lysine side chains. We believe that our method will be particularly useful for the chemical synthesis of lysine-rich and lysine-modified proteins.

3.4 Materials and Methods

3.4.1 General methods:

Unless otherwise noted, all reactions were carried out in oven dried glassware under an atmosphere of nitrogen and distilled solvents were transferred by syringe. Solvents and reagents were purified according to the standard procedure prior to use. Evaporation of organic solutions was achieved by rotary evaporation with a water bath temperature below 40 °C. Product purification by flash column chromatography was accomplished using silica gel 60 (0.010-0.063 nm). Technical grade solvents were used for chromatography and distilled prior to use. NMR spectra were recorded at room temperature on a 300 MHz Bruker ACF 300, 400 MHz Bruker DPX 400 and 500 MHz Bruker AMX 500 NMR spectrometers, respectively. The residual solvent signals were taken as the reference (7.26 ppm for ^{1}H NMR spectroscopy and 77.0 ppm for ^{13}C NMR spectroscopy). Chemical shift (δ) is referred in terms of ppm, coupling constants (J) are given in Hz. Following abbreviations classify the multiplicity: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet or unresolved, br = broad signal. Infrared spectra were recorded on a Bio-RAD FTS 165 FT-IR Spectrometer and reported in cm^{-1}. Samples were prepared in thin film technique. HRMS (ESI) spectra were

recorded on a Finnigan/MAT LCQ quadrupole ion trap mass spectrometer, coupled with the TSP4000 HPLC system and the Crystal 310 CE system.

Amino acid derivatives, coupling reagents and resins were purchased from GL Biochem (Shanghai, China) or Novabiochem. All the other chemical reagents were purchased from Alfa Aesar, Sigma-Aldrich Chemical Company, Fisher Scientific, Acros Organics. PCR kit, restriction enzymes, and chitin beads for construction of ubiquitin-intein expression vector and the purification of intein-CBD fusion proteins were obtained from New England Biolabs. All the analytical HPLC analyses were performed by using an Agilent 1100 series instrument equipped with a Jupiter C18 (5 um, 4.6 x 250 mm) reverse-phase column with a flow rate of 1.0 mL/min. Detection was achieved with a UV-VIS-detector at wavelength $\lambda =$ 220 nm. A typical gradient used for analysis was buffer B rising 2% every minute starting from 0%. The purification was performed using a semi-preparative HPLC column on a Shimadzu system equipped with a vydac C18 column (5 um, 10 x 250 mm) with a flow rate of 2.5 mL/min. The buffer system for all the analysis was buffer A -- H_2O (containing 0.045% TFA) and buffer B -- 90% acetonitrile in H_2O (containing 0.045% TFA). Peptide masses were measured using a Thermo FINNIGAN LCQ Deca XP MAX equipped with ESI ion source.

(2S,4S)-*tert*-Butyl 2-(*tert*-butoxycarbonyl)-4-(tert-butyl diphnylsilyloxy) 6-hydroxyhexanoate (3-2):

3-2
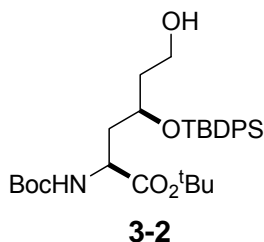
Synthesis and detailed experimental procedure for above compound has been described in reference 25.

^{1}**H NMR (500 MHz, CDCl$_3$):** δ in ppm = 7.70-7.69 (m, 4H, Ph), 7.44-7.36 (m, 6H, Ph), 5.31 (d, J = 7.4 Hz, 1H, NH), 4.25 (t, J = 5.9 Hz, 1H, C$\underline{H}$-NH), 4.13 (dd, J = 13.0 & 8.2 Hz, 1H, C$\underline{H}$-OH), 3.49 (s, 3H, C$\underline{H_3}$), 2.48 (d, J = 6.3 Hz, 2H, C$\underline{H_2}$-CO$_2$Me), 1.99-1.73 (m, 2H, C$\underline{H_2}$), 1.42 (s, 9H, 3C$\underline{H_3}$), 1.37 (s, 9H, 3C$\underline{H_3}$), 1.03 (s, 9H). ^{13}**C NMR (100 MHz, CDCl$_3$):** δ in ppm = 171.7 (C=O), 155.3 (C=O), 81.6 (C), 79.5 (C), 69.6 ($\underline{C}$H-OTBDPS), 59.0 ($\underline{C}$H$_3$), 51.9 ($\underline{C}$H-NH), 38.4 (CH$_2$), 38.3 ($\underline{C}$H$_2$-OH), 28.3 (3$\underline{C}$H$_3$), 27.8 (3CH$_3$), 26.9 (3CH$_3$), 19.2 (C-Si). **IR (CHCl$_3$):** $\bar{v}$ = 3427, 3018, 2981, 2399, 1714, 1215 cm^{-1}. **HRMS (ESI):** m/z: calcd for C$_{31}$H$_{47}$NO$_6$SiNa: 580.3070; [M+Na]$^+$ found: 580.3088.

(2S,4R)-*tert*-Butyl 6-(benzyloxycarbonyl)-2-(*tert*-butoxycarbonylamino)-4-(tert-butyl diphnyl silyloxy) hexanedioate (3-3):

3-3
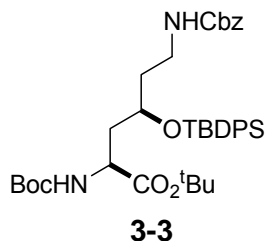
0.13 mL (0.72 mmol) of diisopropylethylamine (DIPEA) was added to a solution of 0.20 g (0.36 mmol) of compound **3-2** in 3.5 mL (0.1M) of dichloromethane at room temperature. The solution was cooled to 0 °C, and 0.04 mL (0.54 mmol) of methanesulfonyl chloride was added. The reaction mixture was allowed to reach room temperature and stirred for 2 h. The reaction mixture was quenched with saturated ammonium chloride solution at 0 °C and diluted with dichloromethane. The layers were separated, the aqueous layer was extracted with dichlormethane two times, and the combined organic extract was washed with water and brine. Drying over anhydrous sodium sulphate and removal of solvent under vacuum resulted

in a colorless oily residue which was dissolved in DMF (4 mL, 0.1M). A total of 0.07 g (1.08 mmol) of sodium azide was added to the solution, which was heated to 80 °C for 5 h. After allowed to reach room temperature, water was added to the solution, which was extracted twice with ethyl acetate. The combined organic layer was washed with water and brine, dried over sodium sulphate and evoporated. The crude product was purified by column chromatography on silica gel to give disired product (0.174 g, 83 %) as a colorless oil. **^{1}H NMR (500 MHz, CDCl$_3$):** δ in ppm = 7.69 (d, J = 6.8 Hz, 4H, Ph), 7.45-7.37 (m, 6H, Ph), 5.22 (d, J = 7.3 Hz, 1H, NH), 4.12 (d, J = 5.7 Hz, 1H, C$\underline{H}$-NH), 3.95 (d, J = 5.4 Hz, 1H, C$\underline{H}$-OH), 3.16 (t, J = 5.7 Hz, 2H, C$\underline{H}_2$-CO$_2$Me), 1.96-1.90 (m, 1H, C$\underline{H}_2$), 1.85 (d, J = 6.6 Hz, 1H, C$\underline{H}_2$), 1.74-1.69 (m, 2H, CH$_2$), 1.43 (s, 9H, 3C$\underline{H}_3$), 1.37 (s, 9H, 3C$\underline{H}_3$), 1.06 (s, 9H). **^{13}C NMR (100 MHz, CDCl$_3$):** δ in ppm = 171.6 (C=O), 155.2 (C=O), 135.8 (Ph), 135.7 (Ph), 133.6 (Ph), 133.1 (Ph), 129.9 (Ph), 129.8 (Ph), 127.7 (Ph), 127.7 (Ph), 81.6 (C), 79.5 (C), 69.0 (C$\underline{H}$-OH), 51.8 (C$\underline{H}$-NH), 47.5 (C$\underline{H}_2$-N$_3$), 38.6 (CH$_2$), 35.2 (CH$_2$) 28.3 (3C$\underline{H}_3$), 27.9 (3CH$_3$), 26.9 (3CH$_3$), 19.3 (C-Si). **IR (CHCl$_3$):** v^- = 3427, 3018, 2399, 2098, 1647, 1215 cm^{-1}. **HRMS (ESI):** m/z: calcd for C$_{31}$H$_{47}$N$_4$O$_5$Si: 583.3316; [M]$^+$ found: 583.3317.

To a solution of azide (0.1 g) in ethyl acetate was added 10% Pd/C (0.03 g) at room temperature. The reaction mixure was stirred for 6 h under an H$_2$ atmosphere at room temperature then filtered through a short pad of celite and washed with chloroform-methanol (1:1). The solvent removed under reduced pressure (quantitative yield). This compound was direct taken for the next step without further purification.

Amine (0.50 g, 0.89 mmol) was dissolved in 2:1 ratio of dioxane-water (10 mL:5 mL) and cooled to 0 °C. Sodium bicarbonate (0.19 g, 2.22 mmol) was introduced portion-wise at the same temperature. Then CbzCl (0.19 mL, 1.3 mmol) was added slowly to the reaction mixture. The reaction mixture was allowed to reach room temperature and stirred for 4 h.

After being cooled to 0 °C, the reaction mixture was quenched by the addition of 0.5 N HCl (2 mL) and stirred for a further 5 min. The solution was dissolved in ethyl acetate and washed with saturated sodium bicarbonate solution, water, and brine. The organic layer was dried over sodium sulphate and concentrated in vacuo, which was purified by column chromatography on silica gel to give disired product **3-3** (0.5 g, 81 %) as a colorless oil. **^{1}H NMR (500 MHz, CDCl$_3$):** δ in ppm = 7.69 (d, J = 6.2 Hz, 4H, Ph), 7.39-7.31 (m, 11H, Ph), 5.46 (d, J = 6.3 Hz, 1H, NH), 5.04-4.97 (m, 2H, CH$_2$-Ph), 4.35 (bs, 1H, C$\underline{H}$-NH), 4.19 (bs, NH), 3.90 (bs, 1H, C$\underline{H}$-OH), 3.12 (t, J = 6.5 Hz, 2H, C$\underline{H}_2$-NH), 2.96 (m, 1H, C$\underline{H}_2$-NH), 1.94 (bs, 2H, C$\underline{H}_2$), 1.62-1.60 (m, 2H, C$\underline{H}_2$), 1.44 (s, 9H, 3C$\underline{H}_3$), 1.40 (s, 9H, 3C$\underline{H}_3$), 1.06 (s, 9H). **^{13}C NMR (100 MHz, CDCl$_3$):** δ in ppm = 171.9 (C=O), 156.3 (C=O), 155.5 (C=O), 136.6 (Ph), 135.9 (Ph), 133.9 (Ph), 133.1 (Ph), 129.9 (Ph), 129.8 (Ph), 128.4 (Ph), 128.0 (Ph), 127.7 (Ph), 127.6 (Ph), 81.4 (C), 79.3 (C), 69.0 (C$\underline{H}$-OH), 66.5 (CH$_2$-Ph), 51.9 (C$\underline{H}$-NH), 38.0 (C$\underline{H}_2$-NH), 37.0 (CH$_2$), 36.5 (CH$_2$), 28.4 (3C$\underline{H}_3$), 27.9 (3CH$_3$), 26.9 (3CH$_3$), 19.3 (C-Si). **IR (CHCl$_3$):** $\bar{v}$ = 3392, 3018, 2399, 2088, 1645, 1215 cm^{-1}. **HRMS (ESI):** *m/z*: calcd for C$_{39}$H$_{55}$N$_2$O$_7$Si: 691.3779; [*M*]$^+$ found: 691.3785.

(2S,4S)-*tert*-Butyl 4-(acetylthio)-6-(benzyloxycarbonyl)-2-(*tert*-butoxycarbonyl) hexanedioate (3-4) :

3-4
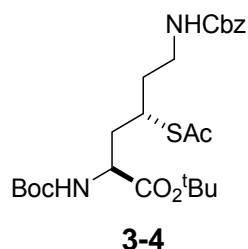
Compound **3-3** (0.05 g, 0.07 mmol) was dissolved in dry THF (1.5 mL, 0.05M) and cooled to 0 °C. 1M solution of TBAF (0.11 ml, 0.11 mmol) was introduced via syringe at the

same temperature. The reaction mixture was stirred for 8 h at 0 °C and quenched with saturated ammonium chloride solution. Evaporation of the solvent gave a residue, which was dissolved in ethyl acetate. The solution was washed with water, and brine. The organic layer was dried over sodium sulphate and concentrated *in vacuo*, which was purified by column chromatography on silica gel to give disired product (0.025 g, 77 %) as a colorless oil. **^{1}H NMR (500 MHz, CDCl$_3$):** δ in ppm = 7.34-7.27 (m, 5H, Ph), 5.46-5.44 (m, 2H, NH & NH), 5.07 (dd, J = 18.0 & 12.3 Hz, 2H, CH$_2$-Ph), 4.47 (bs, 1H, OH), 4.35 (dt, J = 8.2 & 2.4 Hz, 1H, CH-NH), 3.68 (bs, 1H, CH-OH), 3.48 (dt, J = 13.1 & 6.2 Hz, 1H, CH$_2$-NH), 3.23 (dd, J = 12.5 & 5.9 Hz, 1H, CH$_2$-NH), 1.86 (dt, J = 10.9 & 2.7 Hz, 1H, CH$_2$), 1.63-1.47 (m, 3H, CH$_2$, CH), 1.45 (s, 9H, 3CH$_3$), 1.43 (s, 9H, 3CH$_3$). **^{13}C NMR (100 MHz, CDCl$_3$):** δ in ppm = 171.7 (C=O), 156.3 (C=O), 156.7 (C=O), 136.6 (Ph), 128.4 (Ph), 128.0 (Ph), 128.0 (Ph), 82.4 (C), 80.5 (C), 66.5 (CH$_2$-Ph), 66.3 (CH-OH), 51.1 (CH-NH), 41.7 (CH$_2$-NH), 38.9 (CH$_2$), 36.0 (CH$_2$), 28.2 (3CH$_3$), 27.9 (3CH$_3$). **IR (CHCl$_3$):** v^- = 3423, 3018, 2399, 1645, 1215 cm^{-1}. **HRMS (ESI):** *m/z*: calcd for C$_{23}$H$_{37}$N$_2$O$_7$: 453.2601; [*M*]$^+$ found: 453.2590.

0.08 mL (0.44 mmol) of DIPEA was added to a solution of 0.10 g (0.22 mmol) of alcohol in 2.2 mL (0.1M) of dichloromethane at room temperature. The solution was cooled to 0 °C, and 0.02 mL (0.27 mmol) of methanesulfonyl chloride was added. The reaction mixture was allowed to reach room temperature and stirred for 2 h. The reaction mixture was quenched with saturated ammonium chloride solution at 0 °C and diluted with dichloromethane. The layers were separated, the aqueous layer was extracted with dichlormethane two times, and the combined organic extract was washed with water and brine. Drying over anhydrous sodium sulphate and removal of solvent under vacuum resulted in a colorless oily residue which was dissolved in DMF (4.4 mL, 0.05 M). A total of 0.75 g (0.66 mmol) of potassium thioacetate was added to the solution, which was heated to 40 °C for 8 h. After allowed to reach room temperature, water was added to the solution, which was

extracted twice with ethyl acetate. The combined organic layer was washed with water and brine, dried over sodium sulphate and evoporated. The crude product was purified by column chromatography on silica gel to give disired product **3-4** (0.04 g, 70 %) as a pale brown oil.

^{1}H NMR (400 MHz, CDCl$_3$): δ in ppm = 7.34-7.28 (m, 5H, Ph), 5.46 (bs, 1H, NH), 5.23 (d, J = 8.0 Hz, 1H, NH), 5.07 (dd, J = 14.8 & 12.5 Hz, 2H, CH$_2$-Ph), 4.26 (dd, J = 13.6 & 7.8 Hz, 1H, CH-NH), 3.66-3.59 (m, 1H, CH-S), 3.39 (dd, J = 13.0 & 5.4 Hz, 1H, CH$_2$-NH), 3.15-3.10 (m, 1H, CH$_2$-NH), 2.31 (s, 3H, CH$_3$), 2.03 (dd, J = 13.4 & 6.5 Hz, 2H, CH$_2$), 1.90-1.82 (m, 1H, CH$_2$), 1.68-1.59 (m, 1H, CH$_2$) 1.44 (s, 9H, 3CH$_3$), 1.40 (s, 9H, 3CH$_3$). **^{13}C NMR (100 MHz, CDCl$_3$):** δ in ppm = 195.9 (COCH$_3$), 171.1 (C=O), 156.4 (C=O), 155.6 (C=O), 136.7 (Ph), 128.4 (Ph), 128.1 (Ph), 127.9 (Ph), 82.4 (C), 80.1 (C), 66.5 (CH$_2$-Ph), 51.9 (CH-NH), 39.7 (CH$_2$-NH), 38.2 (CH$_2$), 38.2 (CH$_2$), 33.7 (CH-S), 33.7 (CH$_3$) 28.2 (3CH$_3$), 27.9 (3CH$_3$). **IR (CHCl$_3$):** ν = 3431, 3018, 2399, 1645, 2088, 1637, 1215 cm^{-1}. **HRMS (ESI):** *m/z*: calcd for C$_{25}$H$_{39}$N$_2$O$_7$S: 511.2478; [*M*]$^+$ found: 511.2480.

(2S,4S)-*tert*-Butyl-6-(benzyloxycarbonyl)-2-(*tert*-butoxycarbonyl)-4-(2-methyldisulfanyl) hexanedioate (3-5) :

3-5
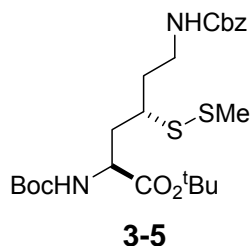
0.020 g of Boc-lys(SAc, Cbz)-OtBu (**3-4**) was dissolved in 0.6 mL of methanol. 0.175 mL of 1 N NaOH was added. After 30 min at room temperature, the mixture was neutralized carefully with 1 N HCl at 0 °C. The mixture was concentrated and extracted with ethyl

acetate three times. The combined organic layer was washed with water and brine, dried over sodium sulphate and evoporated.

The residue was dissolved in 0.4 mL of DCM and added dropwise to a mixture containing 0.4 mL of DCM, 0.006 mL of Et₃N and 0.015 mL of S-methyl methanethiosulfonate (MMTS) while stirring at room temperature. The mixture was continuously stirred for 30 min. The solvent was removed and the residue was dissolved in methanol and purified by C18 semi-preparative HPLC followed by lyophilization to give desired product Boc-lys(SSMe, Cbz)-OtBu **3-5** (10 mg) as a white solid. **¹H NMR (400 MHz, CDCl₃):** δ in ppm = 7.36-7.28 (m, 5H, Ph), 5.47 (bs, 1H, NH), 5.25 (d, J = 7.6 Hz, 1H, NH), 5.10 (s, 2H, CH₂-Ph), 4.26 (s, 1H, C$\underline{H}$-NH), 3.47 (m, 1H, C$\underline{H}$-S), 3.37 (d, J = 4.5 Hz, 1H, C$\underline{H}$₂-NH), 2.82 (bs, 1H, C$\underline{H}$), 2.40 (s, 3H, CH₃), 2.07-1.97 (m, 2H, C$\underline{H}$₂), 1.85-1.80 (m, 2H, C$\underline{H}$₂) 1.49 (s, 9H, 3C$\underline{H}$₃), 1.39 (s, 9H, 3C$\underline{H}$₃). **¹³C NMR (100 MHz, CDCl₃):** δ in ppm = 171.2 (C=O), 156.5 (C=O), 155.9 (C=O), 136.7 (Ph), 128.4 (Ph), 128.2 (Ph), 127.9 (Ph), 82.6 (C), 80.3 (C), 66.5 (CH₂-Ph), 51.9 (C$\underline{H}$-NH), 44.7 (C$\underline{H}$₂-NH), 40.4 (C$\underline{H}$₃-S), 38.4 (C$\underline{H}$₂), 32.0 (CH₂), 28.2 (3C$\underline{H}$₃), 28.0 (3C$\underline{H}$₃), 24.0 (CH-S). **IR (CHCl₃):** v^- = 3425, 3025, 1640, 2090, 1650 cm⁻¹. **HRMS (ESI):** m/z: calcd for C₂₄H₃₈N₂O₆S₂Na: 537.2069; $[M+Na]^+$ found: 537.2066.

(2S,4S)-*tert*-Butyl-6-(benzyloxycarbonyl)-2-(*tert*-butoxycarbonyl)-4-(2-methyldisulfanyl) hexanoic acid (3-6) :

3-6
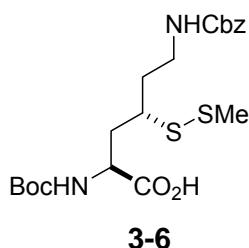
The white powder of Boc-lys(SSMe, Cbz)-OtBu was desolved in 0.5 mL 95% TFA. After 1 h at room temperature, the TFA was removed by evaporation. The residue was dissolved in

1 mL methanol/H$_2$O mixture (3:1). Adjust pH to about 8 with Et$_3$N. 0.02 mL of Boc$_2$O was added. After 3 h at room temperature, the sample was subjected to C18 semi-preparative HPLC. After lyophilization 7 mg of Boc-Lys(SSMe, Cbz)-OH **3-6** was isolated. **^{1}H NMR (400 MHz, CDCl$_3$):** δ in ppm = 7.34-7.33 (m, 5H, Ph), 5.33 (bs, 1H, NH), 5.29 (d, J = 7.2 Hz, 1H, NH), 5.08 (s, 2H, CH$_2$-Ph), 4.43 (b, 1H, C$\underline{H}$-NH), 3.38 (b, 2H, C$\underline{H}_2$), 2.92-2.86 (m, 1H, C$\underline{H}$), 2.38 (s, 3H, CH$_3$), 2.17-2.03 (m, 2H, C$\underline{H}_2$), 1.87-1.83 (m, 2H, C$\underline{H}_2$) 1.39 (s, 9H, 3C$\underline{H}_3$). **^{13}C NMR (100 MHz, CDCl$_3$):** δ in ppm = 175.3 (C=O), 156.7 (C=O), 156.0 (C=O), 136.4 (Ph), 128.4 (Ph), 128.2 (Ph), 128.1 (Ph), 80.8 (C), 66.8 (CH$_2$-Ph), 51.4 ($\underline{C}$H-NH), 44.7 ($\underline{C}$H$_2$-NH), 39.1 ($\underline{C}$H$_3$-S), 38.5 ($\underline{C}$H$_2$), 32.6 (CH$_2$), 28.2 (3$\underline{C}$H$_3$), 24.0 (CH-S). **IR (CHCl$_3$):** v^- = 3440, 3022, 2399, 1650, 1640 cm^{-1}. **HRMS (ESI):** m/z: calcd for C$_{20}$H$_{30}$N$_2$O$_6$S$_2$Na: 481.1443; [M+Na]$^+$; found 481.1446.

3.4.2 Solid-phase peptide synthesis

<u>Synthesis of thioester peptides.</u> H-**LSTEA**-SCH$_2$CH$_2$CONH$_2$, H-**LSTEG**-SCH$_2$CH$_2$CONH$_2$ and H-**LSTEL**-SCH$_2$CH$_2$CONH$_2$ were manually synthesized employing standard t-Boc chemistry. First, Trt-SCH$_2$CH$_2$COOH was coupled onto MBHA resin. The trityl group was removed by treatment with a cocktail containing TFA/TIS/β-mercaptoethanol/DCM (5:2.5:2.5:90). For the coupling of amino acids, Boc-amino acid (4 eq.) and PyBOP (4 eq.) were dissolved in DCM. DIEA (12 eq.) was added in the solution. After 2 min of activation, the mixture was mixed with resin. The reaction was undertaken for 1.5 h. The coupling efficiency was checked with Kaiser test. The Boc group was removed by treatment with 30% TFA in DCM for 10 min, followed by 15 min. The side chain protected amino acid derivatives used were Boc-Ser(Bzl)-OH, Boc-Thr(Bzl)-OH, Boc-Glu(OBzl)-OH. After sequence assembly, peptide thioesters were cleaved from the resin with a cocktail

consisting TFMSA/TFA/*p*-cresol/methyl phenyl sulfide (1:7:1:1) for 1 h. The crude peptides were purified with C18 semi-preparative HPLC.

Synthesis of Biotinyl-SCH$_2$CH$_2$CO-**GFRA**-NH$_2$. The amino acid sequence **GFRA** in the biotin thioester was assembled manually on rink-amide MBHA resin employing standard Fmoc chemistry, followed by the coupling of Trt-SCH$_2$CH$_2$COOH. After removal of trityl by repeated treatment with TFA/TIS/β-mercaptoethanol/DCM (2.5:2.5:2.5:92.5), biotin was coupled onto the thiol-derived resin (4 eq. of biotin, 4 eq. of PyBOP, 12 eq. of DIEA in DMF). The biotin thioester was cleaved from the resin with a cocktail consisting TFA/TIS/H$_2$O (95:2.5:2.5) for 2.5 h. The crude thioester was purified with C18 semi-preparative HPLC.

Synthesis of H-**K**(γ-SSMe, ε-Cbz)**GAKAFA**-NH$_2$ (peptide **12**). All the amino acids except Boc-Lys(γ-SSMe, ε-Cbz)-OH in peptide **12** were coupled employing standard Fmoc-based chemistry on rink amide MBHA resin. After coupling Boc-Lys(γ-SSMe, ε-Cbz)-OH, the peptide was cleaved from the resin with a cocktail consisting of TFA/TIS/H$_2$O (95:2.5:2.5) for 20 min. The crude peptide was purified with C18 semi-preparative HPLC.

3.4.3 Preparation of ub(1-76)-MES

Construction of ubiquitin expression plasmid pTYB1-Ubi. The plasmid containing human ubiquitin gene was obtained from Dr. Cheung Ching For, Peter's lab in Nanyang Technological University. Ubiquitin gene was amplified by PCR using the primers: Ubi_F: 5'-GGTGGTCATATGCAGATCTTTGTGAAG-3' and Ubi_R:5'-GGTGGTTGCTCTTCCGCAGCCACCTCGCAGGCG-3'. The PCR condition was 95 $^{\circ}$C for 3 min, 30 cycles of 95 $^{\circ}$C 30 s, 54 $^{\circ}$C 30 s, 72 $^{\circ}$C 30 s, and 72 $^{\circ}$C 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* cell and amplified. The extracted and purified T-Ubi vector was digested with *Nde*I and *Sap*I restriction enzymes. The digested segment was purified by PCR purification kit (Qiagen) and ligated into the identically digested pTYB1 vector (New England Biolabs). The correct insert was confirmed by DNA sequencing.

Overexpression and purification of ub(1-76)-MES. 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 °C with shaking at 250 rpm to an OD$_{600}$ of 0.6-0.8. The cells were induced by 50 μM IPTG at 15 °C for 18 h. After centrifugation at 6000 rpm for 10 min, cell pellets from 1 L culture were suspended in 50 mL lysis buffer (20 mM HEPES, 0.5 M NaCl, 1 mM EDTA, pH 7.0). Cells were lysed with a microfluider (Microfluidics, Newton, USA) at a chamber pressure of 12K/50 psi. Debris was removed by centrifugation at 20,000*g* for 30 min. The supernatant was mixed with 3 mL chitin beads (New England Biolabs) pre-equilibrated by the lysis buffer at 37 °C 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 MESNa, 20 mM HEPES, 0.5 M NaCl, 1 mM EDTA, pH 7.0) and incubating at 37 °C overnight. The Ub(1-76)-MES was eluted with 10 mL of the lysis buffer. The cleaved product was purified by C8 semi-prep RP-HPLC. The identity of the purified Ub(1-76)-MES was identified with C18 analytical HPLC and ESI-MS.

3.4.5 General procedures for reactions involved in dual NCL at lysine

General procedures for NCL at the 4-thiolysine residue. 0.2 – 2 mg of peptides involved in ligation were dissolved in 50 – 100 μL ligation buffer containing 6 M Gdn•HCl, 0.2 M phosphate, 60 mM TCEP, pH 8.0. Except the ligation involved ub(1-76)-MES, 1% v/v benzyl

mercaptan was added into the ligation mixture. All the reactions were performed at 37 °C. The time course of the ligation was monitored with C18 analytical HPLC.

General procedures for Cbz removal. 50 or 100 μL of chilled cocktail containing TFMSA/TFA/p-cresol/methyl phenyl sulfide (1:7:1:1) was added into 1 ~ 2 mg of Cbz containing peptide on ice. The mixture was kept on ice for 2 min and in room temperature for another 20 min. 0.5 ~ 1 mL diethyl ether was added and the mixture was centrifuged. The pallet was dissolved in 50 % ACN/H_2O and purified with C18 semi-preparative HPLC.

Raney nickel-mediated desulfurization of peptide **16a** to give peptide **17a**. 0.5 mg of peptide **16a** was dissolved in 200 μL of buffer containing 6 M Gdn•HCl, 0.2 M phosphate, 100 mM TCEP, pH 6.0. Raney nickel was prepared by dissolving 90 mg of $NiSO_4•6H_2O$ in 2 mL water. 5 mg of $NaBH_4$ was added into the solution slowly. After 5 min, the pellet was collected by centrifugation and washed extensively with water. The peptide solution was added into the pellet. The reaction was stirred at room temperature. After 2 h, fresh raney nickel prepared from 180 mg of $NiSO_4•6H_2O$ and 20 mg of $NaBH_4$ was added into the mixture. After another 6 h, analytical HPLC showed that the desulfurization was completed.

General procedure for free radical-mediated desulfurization. All the following steps were performed under nitrogen. All the solutions were prepared immediately before being used. 0.2 mg peptide was dissolved in 150 μL buffer (6 M Gdn•HCl, 0.1 M phosphate, pH ~6.5). 100 μL of 0.5 M TCEP (pH adjusted to 6-7 with 5 M NaOH) was added in the solution. 25 μL of 10 mM glutathione was added. 10 μL 0.2 M 2,2'-Azobis[2-(2-imidazolin-2-yl)propane]dihydrochloride (VA-044) was added. The mixture was stirred at 37 °C for 5 h.

Ligation between H-**LSTEA**-$SCH_2CH_2CONH_2$ and peptide **12**. 2 mg of H-**LSTEA**-$SCH_2CH_2CONH_2$ and 2 mg of peptide **12** were dissolved in 100 μL of ligation buffer (6 M Gdn•HCl, 0.2 M phosphate, 60 mM TCEP, pH 8.0, 1% v/v benzyl mercaptan) and incubated

at 37 °C. The ligation progression was monitored with C18 analytical HPLC. After 1 h reaction, the ligation product was purified with C18 semi-preparative HPLC.

Removal of Cbz from peptide **14a** to give peptide **15a**. 100 μL of chilled cocktail containing TFMSA/TFA/*p*-cresol/methyl phenyl sulfide (1:7:1:1) was added into 0.9 mg of peptide **14a** on ice. The mixture was kept on ice for 2 min and in room temperature for another 20 min. 1 mL diethyl ether was added and the mixture was centrifuged. The pallet was dissolved in 50% ACN/H_2O and purified with C18 semi-preparative HPLC.

Ligation between peptide **15a** and H-**LSTEG**-$SCH_2CH_2CONH_2$. 0.4 mg of peptide **15a** and 0.4 mg of H-**LSTEG**-$SCH_2CH_2CONH_2$ were dissolved in 50 μL of ligation buffer and incubated at 37 °C. The reaction was monitored with C18 analytical HPLC. After 1.5 h, the ligation product was purified with multiple runs of analytical HPLC.

Ligation between H-**LSTEL**-$SCH_2CH_2CONH_2$ and peptide **12**. 0.9 mg of H-**LSTEL**-$SCH_2CH_2CONH_2$ and 0.9 mg of peptide **12** were dissolved in 50 μL of ligation buffer and incubated at 37°C. After 1 h, the ligation product was purified with C18 semi-preparative HPLC.

Removal of Cbz from peptide **14b**. 50 μL of chilled cocktail containing TFMSA/TFA/*p*-cresol/methyl phenyl sulfide (1:7:1:1) was added into 1.2 mg of peptide **14b** on ice. The mixture was kept on ice for 2 min and in room temperature for another 20 min. 0.5 mL diethyl ether was added and the mixture was centrifuged. The pallet was dissolved in 50% ACN/H_2O and purified with C18 semi-preparative HPLC.

Free readical desulfurization of peptide **16b** to give peptide **17b**. 0.2 mg peptide **16b** was desulfurized according to the general procedure of free radical-mediated desulfurization described previously. The yield was quantitative based on HPLC and MS analysis.

Ligation of peptide **15b** with Biotinyl-SCH$_2$CH$_2$CO-**GFRA**-NH$_2$. 0.3 mg of peptide **15b** and 0.3 mg of Biotinyl-SCH$_2$CH$_2$CO-**GFRA**-NH$_2$ were dissolved in 50 μL ligation buffer and incubated at 37°C. After 3 h, excess of DTT was added into the mixture. Let the mixture be reduced for 10 min. An aliquot was taken out for C18 analytical HPLC analysis. The yield was about 90% based on analytical HPLC. The rest was purified by 2 runs of C18 analytical HPLC.

Free radical desulfurization of **16c**. 0.3 mg peptide **16c** was desulfurized according to the general procedure of free radical-mediated desulfurization. The yield was 80% based on quantitative HPLC analysis.

References

[1] O. Kerscher, R. Felberbaum, M. Hochstrasser, *Annu. Rev. Cell Dev. Biol.* **2006**, 22, 159.

[2] H. Wang, L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones, Y. Zhang, *Nature* **2004**, 431, 873.

[3] W. Zhou, P. Zhu, J. Wang, G. Pascual, K. A. Ohgi, J. Lozach, C. K. Glass, M. G. Rosenfeld, *Mol. Cell* **2008**, 29, 69.

[4] R. Pavri, B. Zhu, G. Li, P. Trojer, S. Mandal, A. Shilatifard, D. Reinberg, *Cell* **2006**, 125, 703.

[5] W. Li, Y. Ye, *Cell Mol. Life Sci.* **2008**, 65, 2397.

[6] M. Hochstrasser, *Curr. Opin. Cell Biol.* **1995**, 7, 215.

[7] C. Wang, L. Deng, M. Hong, G. R. Akkaraju, J.-I. Inoue, Z. J. Chen, *Nature* **2001**, 412, 346.

[8] C. Chatterjee, R. K. McGinty, J.-P. Pellois, T. W. Muir, *Angew. Chem., Int. Ed.* **2007**, 46, 2814.

[9] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B, H. Kent, *Science* **1994**, 266, 776.

[10] J. P. Tam, Y. Lu, C.-F. Liu, J. Shao, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 12485.

[11] L. E. Canne, S. J. Bark, S. B. H. Kent, *J. Am .Chem. Soc.* **1996**, *118*, 5891.

[12] D. W. Low, M. G. Hill, M. R. Carrasco, S. B. Kent, P. Botti, *Proc. Natl. Acad. Sci. U.S.A.* **2001**, 98, 6554.

[13] J. Offer, C. N. Boddy, P. E. Dawson, *J. Am. Chem. Soc.* **2002**, 124, 4642.

[14] B. Wu, J. Chen, J. D. Warren, G. Chen, Z. Hua, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2006**, *45*, 4116.

[15] G. Chen, J. D. Warren, J. Chen, B. Wu, B. Q. Wan, S. J. Danishefsky, *J. Am. Chem. Soc.* **2006**, *128*, 7460.

[16] J. Offer, *Biopolymers* **2010**, 94, 530.

[17] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, 123, 526.

[18] D. Crich, A. Banerjee, *J. Am. Chem. Soc.* **2007**, *129*, 10064.

[19] C. Haase, H. Rohde, O. Seitz, *Angew. Chem., Int. Ed.* **2008**, *47*, 6807.

[20] J. Chen, Q. Wan, Y. Yuan, J. Zhu, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2008**, 47, 8521.

[21] J. Chen, P. Wang, J. Zhu, Q. Wan, S. J. Danishefsky, *Tetrahedron* **2010**, 66, 2277.

[22] Z. Harpaz, P. Siman, K. S. Ajish Kumar, A. Brik, *ChemBioChem* **2010**, 11, 1232.

[23] Z. Tan, S. Shang, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2010**, 49, 9500.

[24] H. Rohde, O. Seitz, *Biopolymers* **2010**, 94, 551.

[25] J. Martin, C. Didierjean, A. Aubry, J.- R. Casmir, J. P. Briand, G. Guichand, *J. Org. Chem.* **2004**, *69*, 130.

[26] A. Borodovsky, H. Ovaa, N. Kolli, T. Gan-Erdene, K. D. Wilkinson, H. L. Ploegh, B. M. Kessler, *Chem. Biol.* **2002**, 9, 1149.

[27] Q. Wan, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2007**, 46, 9248.

Chapter 4: Synthesis of K48-linked Diubiquitin Using Dual Native Chemical Ligation at Lysine

4.1 Introduction

As mentioned in the previous chapter, ubiquitination is one of the most important protein posttranslational modifications.[1] Either one or a chain of multiple ubiquitin molecules can be linked to the modified protein target, which are termed monoubiquitination and polyubiquitination, respectively.[1] It has been reported that these two types of modifications play different roles in the biological systems. The two well-known examples of monoubiquitin are the the modification at the C-termini of histone H2A and H2B. H2A ubiquitination is associated with gene silencing[2,3] and H2B ubiquitination was shown to be associated with transcription elongation.[4] There are seven lysine residues (K6, 11, 27, 29, 33, 48, 63) in the ubiquitin sequence. All of them can be involved in polyubiquitination. Different forms of polyubiquitin chains may have different conformations although they are built by the same monomer. K48-linked polyubiquitin forms a compact structure[5-8] whereas K63-linked polyubiquitin forms extended chains lacking intersubunit interfaces[9-11]. The biological roles of polyubiquitin may be linkage-dependent.[4] K48-linked polyubiquitin is the best understood polyubiquitination so far and is shown to serve as a signal for targeting the protein to proteasomal degradation.[12] In contrary, K63-linked polyubiquitination is involved in signaling pathway and associated with kinase activation.[13] Preliminary reports have shown that K11-linked polyubiquitin may be involved in cell cycle control.[14] Compared to K48- and K63-polyubiquitin, the structure and function of other forms of polyubiquitins are poorly understood.

To elucidate the biological role of polyubiquitination, it is important to isolate

enough amount of ubiquitinated proteins for in vitro study. In eukaryotes, ubiquitination is catalyzed by three classes of enzymes, ubiquitin-activating enzymes E1, ubiquitin-conjugating enzymes E2 and ubiquitin ligase E3.[1] In vitro enzymatic synthesis of ubiquitinated proteins is limited by the requirement for identifying the ligases and the availability of the ligase. So far, only homogeneous K11- K48-, and K63-linked polyubiquitins can be assembled using the conjugating enzymes Ube2S, E2-25K, and Ubc13/Mms2, respectively.[15-17] The enzymatic synthesis of other forms of polyubiquitins have not been achieved due to the lack of the identified ubiquitin-conjugating enzymes E2 responsible for the specific linkages.

Besides enzymatic polyubiquitination, chemical ubiquitination is another option. Different from other small molecule posttranslational modifications, such as methylation, acetylation and so on, ubiquitination is the modification of proteins by a very large tag, which makes chemical ubiquitination extremely challenging. Due to the importance of ubiquination and the difficulty of synthesis, chemical ubiquitination has drawn more attentions from the chemical biologists. In recent years, starting from the chemical synthesis of ubiquitinated small peptide to the large proteins, chemical ubiquitination has made dramatic progress. Peptide ubiquitination was succeeded by either $N\alpha$-auxiliary-mediated[18] or thiolysine-mediated[19, 20] peptide ubiquitination. These methods generate the native Gly76-Lys(ε) isopeptide bond which is typical for ubiquitination. Cysteine mediated ligation/desulfurization enables the synthesis of Gly76Ala monoubiquitinatinated proteins.[21, 22] Essentially, all these methods are based on the native chemical ligation (NCL) approach.[23, 24] Recently, chemical ubiquitination employing a disulfide linkage was also reported.[25, 26] These two forms of non-native monoubiquitination were shown to exert similar effect as the native form.

After the successful achievments in monoubiquitination, the next step is to develop

chemical approaches for the study of polyubiquitinations. Polyubiquitin chains are built by ubiquitins through the repeat of the same linkage. Diubiquitins are the repeating units of polyubiquitins chains. Therefore, the study of diubiquitins may reveal many aspects of polyubiquitins. The conformation of the two linked monomers of the diubiquitin may give us some clue of the conformation of polyubiquitin. In addition, diubiquitins with different linkages are important antigens for the generation of linkage specific antibodies.[14] These antibodies are important tools for the study of the function of polyubiquitination using molecular and cellular approaches.[14]

The objective of this project was to deleveop an approach for the chemical synthesis of diubiquitins. In the previous chapter, we reported the synthesis of a monoubiquitinated peptide through dual native chemical ligation at lysine.[19, 27] A lysine derivative with a thiol group at C-4 of lysine side chain was used to mediate the chemical ligation at both the α- and ε- amino group of lysine. While the ligation at the α-amino group enables the chemical synthesis of the target protein itself, the side chain ligation can be used for the site-specific installation of the ubiquitin tag. In the previous study, we employed benzyloxycarbonyl (Cbz) as the orthogonal protecting group for the ε-amino group of 4-thiolysine to make sure that the ligation first selectively occur at the α-amine. Then the Cbz group was removed with a strong acid, such as TFMSA, to free the ε-amine for the side chain ligation.[19] We realize that the harsh condition for Cbz removal may limit the application of our methodology for large protein ubiquitination. In this project, we revised our orthogonal protection strategy by replacing Cbz with a photolabile protection group, *o*-nitroveratryloxycarbonyl (NVOC) (Fig. 4.1). We then apply our revised approach to the synthesis of K48-linked diubiquitin.
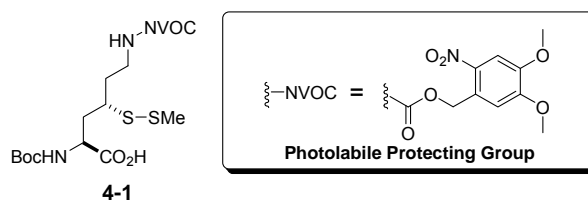
Fig. 4.1 *Nε*-NVOC protected 4-mercaptolysine derivative.

4.2 Results

The *Nε*-NVOC protected 4-thiolysine derivative **4-1** was synthesized by Kalyan Kumar Pasunooti from Dr Liu Xue-Wei's lab in a similar way as previously reported[19] by using NVOC-Cl instead of Cbz-Cl as the side chain protecting reagent.

4.1.1 Preparation of peptide segments for the synthesis of K48(SH, NVOC) containing monoubiquitin

The overall strategy of synthesizing the diubiquitin is shown in Scheme 4.1. To synthesize K48-linked diubiquitin (**24**) using dual NCL at lysine method, the crucial step is the installation of 4-thiolysine at position 48 of the monoubiquitin. The synthesis of ubiquitin and its mutants through stepwise synthesis or sequential chemical ligation/desulfurization has been reported.[28, 29] Here, we synthesized the K48(4-SH)-containing ubiquitin (**22**) using C-to-N sequential ligation with Ala28Cys and K48(4-SH) as the ligation junctions. After we got **22**, the NVOC group was removed by 365 nm UV irradiation. Monoubiquitin **23** with free ε-NH$_2$ on K48 was then reacted with ubiquitin thioester Ub(1-76)-MES which was generated by thiolysis of ubiquitin-intein fusion protein with sodium mercaptoethanesulfonate (MESNa). To generate the native K48-linked diubiquitin, free radical mediated desulfurization[30] was performed to the ligation product to convert Cys28 to Ala and K48(4-SH) to Lys.

Peptide **18**: H-**LQIFVK**[6]**TLTGK**[11]**TITLEVEPSDTIENVK**[27]-SCH$_2$CH$_2$CONH$_2$

Peptide **19**: H-**ThzK**[29]**IQDK**[33]**EGIPPDQQRLIFAG**-SCH$_2$CH$_2$CONH$_2$

Peptide **20:** H-**K**[48]**(S-SMe, NVOC)QLEDGRTLSDYNIQK**[63]**ESTLHLVLRLRGG**-OH
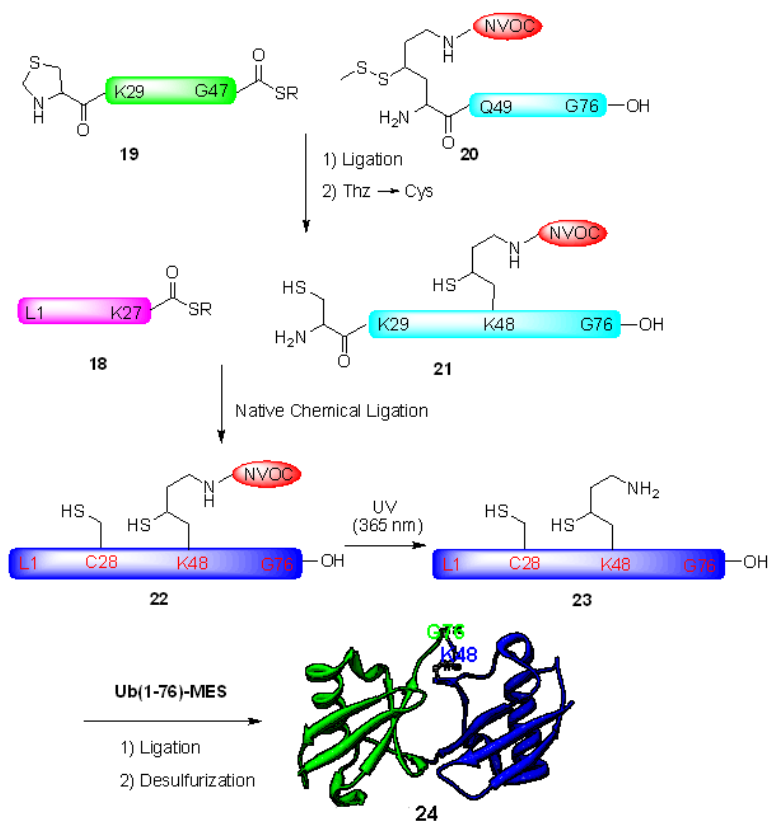
Scheme 4.1: The strategy used for the synthesis of K48-linked diubiquitin. Note: the Met1 of the synthesized ubiquitin was changed to Leu to avoid the oxidation. Thz: thiazolidine.

To synthesize K48(SH, NVOC) containing ubiquitin, a three-segmental ligation was performed. First, the three segments were prepared by solid-phase peptide synthesis. The sequences of the three segments are shown in Scheme 4.1. The two thioesters **18**, **19** were prepared by Boc chemistry. The N-terminal cysteine residue of middle thioester **19** was protected as thiazolidine (Thz) to avoid self cyclization and ligation during its reaction with the C-terminal peptide **20**. Peptide **20** with an N-terminal thiolysine derivative was prepared using Fmoc chemistry. The analytical HPLC and ESI-MS of the three purified segments were shown in Fig. 4.2-4.4
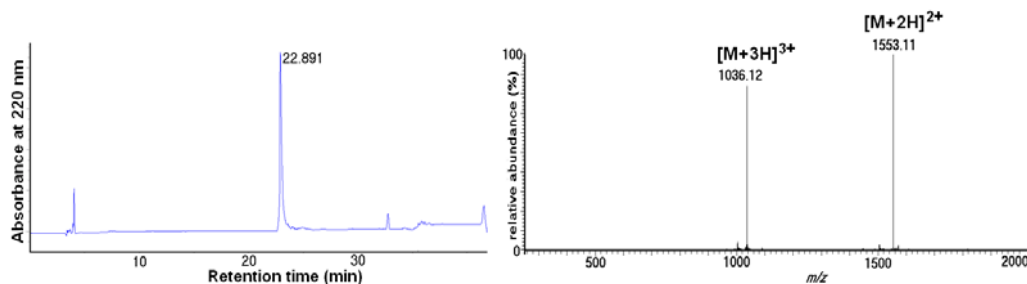
Fig. 4.2 C18 analytical HPLC and ESI-MS of purified Ub(L1-K27)-COSCH$_2$CH$_2$CONH$_2$ (**18**). HPLC gradient: 0-60% B in 30 min. Calculated [M+H]$^+$ = 3105.6, found [M+2H]$^{2+}$ = 1553.11, [M+3H]$^{3+}$ = 1036.12.
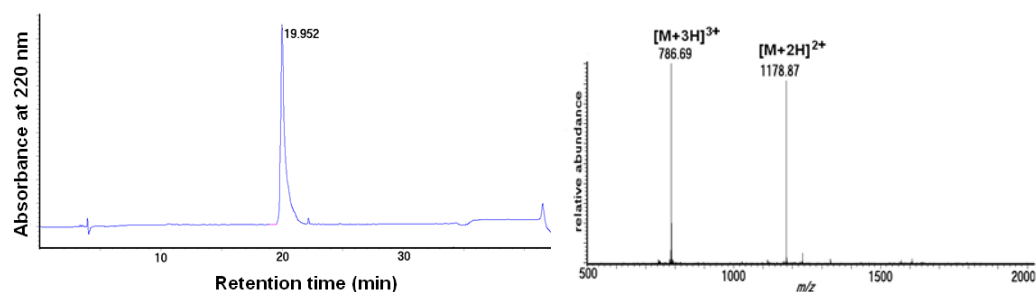

Fig. 4.3 C18 analytical HPLC and ESI-MS of purified Ub(Thz28-G47)-COSCH$_2$CH$_2$CONH$_2$ (**19**). HPLC gradient: 0-60% B in 30 min. Calculated [M+H]$^+$ m/z 2356.7, observed [M+2H]$^{2+}$ m/z 1178.87, [M+3H]$^{3+}$ m/z 786.69.
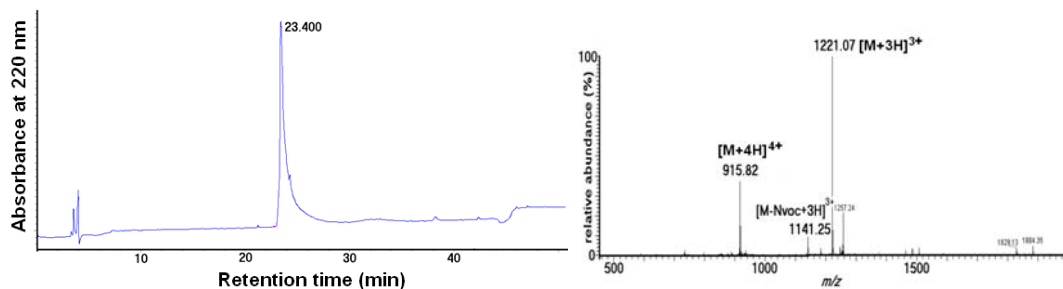

Fig. 4.4 C18 analytical HPLC and ESI-MS of purified K48(4-SSMe, NVOC)-G76-OH (**20**). HPLC gradient: 0-80% B in 40 min. Calculated [M+H]$^+$ m/z: 3659.0, observed [M+3H]$^{3+}$ m/z 1221.07, [M+4H]$^{4+}$ m/z 915.82, [M-NVOC+3H]$^{3+}$ m/z 1141.25.

4.2.2 Synthesis of monoubiquitin 22

HPLC data for the synthesis of monoubiquitin **22** was shown in Fig. 4.5. With the three segments at hand, we first performed the 1st step of ligation, the 4-thiolysine mediated ligation between peptide thioester **19** and peptide **20**. 8 mg of **19** and 10 mg of **20** were dissolved in ligation buffer (6 M Gdn•HCl, 0.2 M phosphate, 20 mM TCEP, pH 7.5). In the presence of TCEP, the disulfide in **20** was immediately reduced and gave the free thiol. Upon addition of the thiol additive MESNa (0.2 M), **19** was converted to its MES thioester. The ligation proceeded efficiently. After 12 h, the reaction was completed. (Fig. 4.5, trace A and B). The ligation product was verified by ESI-MS (Fig. 4.6). The convertion of Thz (1,3-thiazolidine-4-carboxo group) to Cys was directly performed in the ligation mixture without isolation of the ligation product. 0.4 M MeONH$_2$•HCl (final concentration) was added into the mixture and pH was adjusted to 4.0. After 4.5 h at room temperature, the deprotection was completed (Fig. 4.5C). The product **21** was purified by C18 semi-preparative HPLC and analyzed by ESI-MS (Fig. 4.7). About 10 mg of ligation product was obtained after lyophilization. For the next ligation step, 7.3 mg of peptide thioester **18** and 10 mg of **21** were dissolved in ligation mixture (6 M Gdn•HCl, 0.2 M phosphate, 0.2 M MESNa, 20 mM TCEP, pH 7.5). The reaction was completed overnight (Fig. 4.5E). The major side reaction was the self-cyclization of **18** resulting from the nucleophilic attack of the C-terminal thioester by the ε-amine of the C-terminal lysine residue. The full-length ubiquitin **18** was purified by C18 semi-preparative HPLC. The molecular weight of **22** was confirmed by ESI-MS (Fig. 4.8). About 5.1 mg of **22** was obtained after lyophilization.
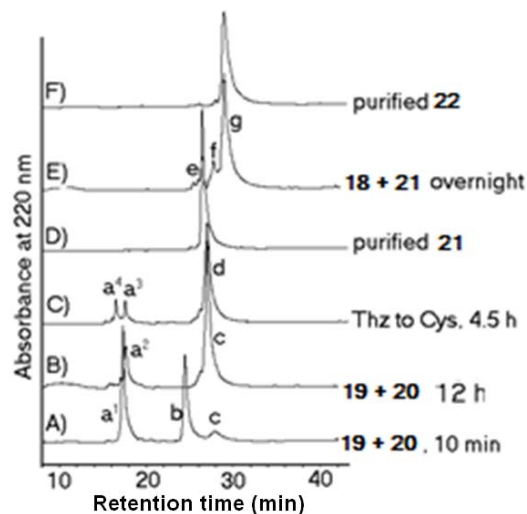
Fig. 4.5 C8 analytical HPLC data for the synthesis of **22**. A) and B): Ligation between **19** and **20** at 10 min and 12 h, respectively. Peak a[1], the 1:10 mixture of **19** and its MES thioester; peak b, reduced **20**; peak c, ligation product **Thz28-K48(4-SH, Nε-NVOC)-G76**-OH; peak a[2], about 1:1 mixture of **19** and its MES thioester. C): The *in situ* conversion of Thz to Cys for 4.5 h. peak a[3], peptide **19**; peak a[4], Cys28-G47-NHOCH₃; peak d, product **22**. D): Purified **21**. E): Ligation between **18** and **21** overnight. Peak e, the MES thioester of **18**; peak f, cyclization product of **18**; peak g, ligation product **22**. F): Purified **22**.
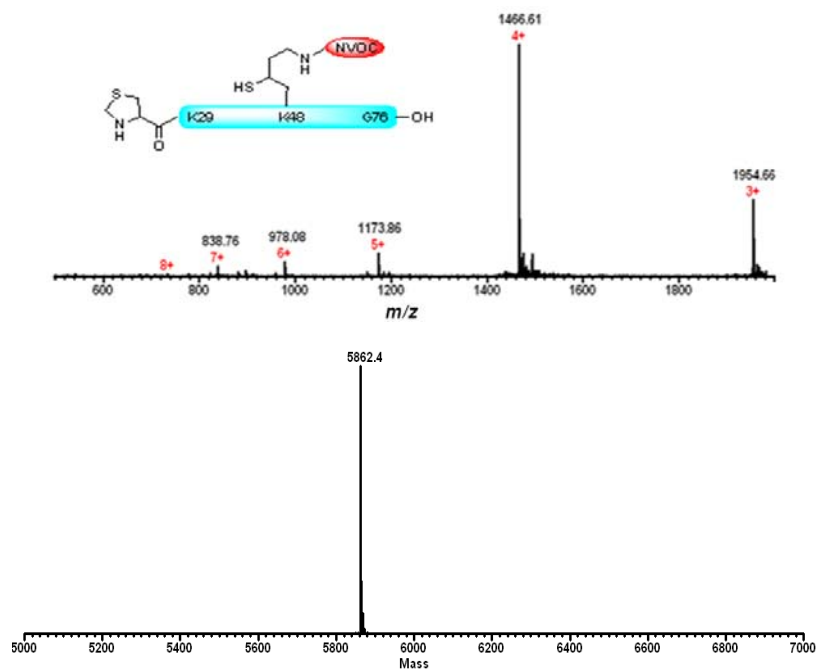

Fig. 4.6 ESI-MS of Thz28-K48(4-SH, NVOC)-G76-OH, the ligation product of **19** and **20**. Observed mass 5862.4 Da, calculated mass 5861.6 Da.
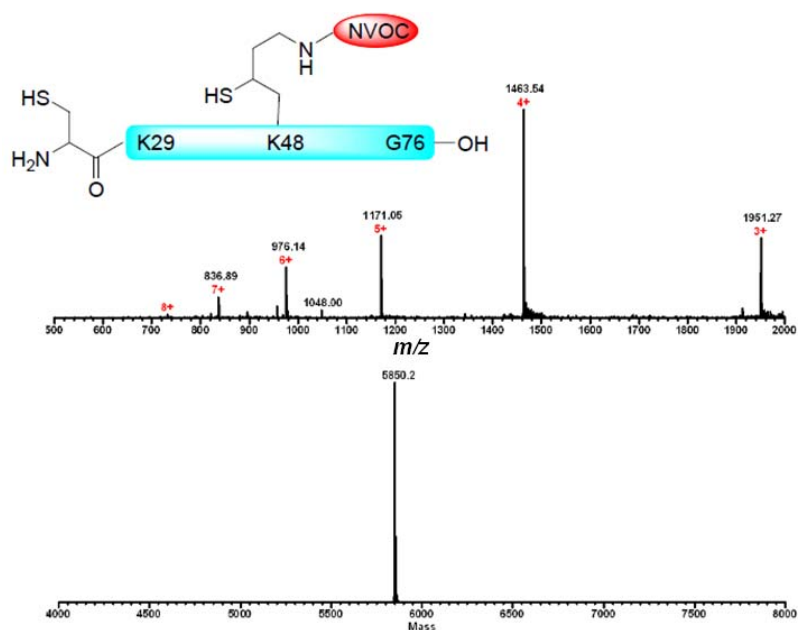
Fig. 4.7 ESI-MS of C28-K48(4-SH, NVOC)-G76-OH (**21**), the Thz deprotection product. Observed mass 5850.2 Da, calculated M = 5849.6 Da.
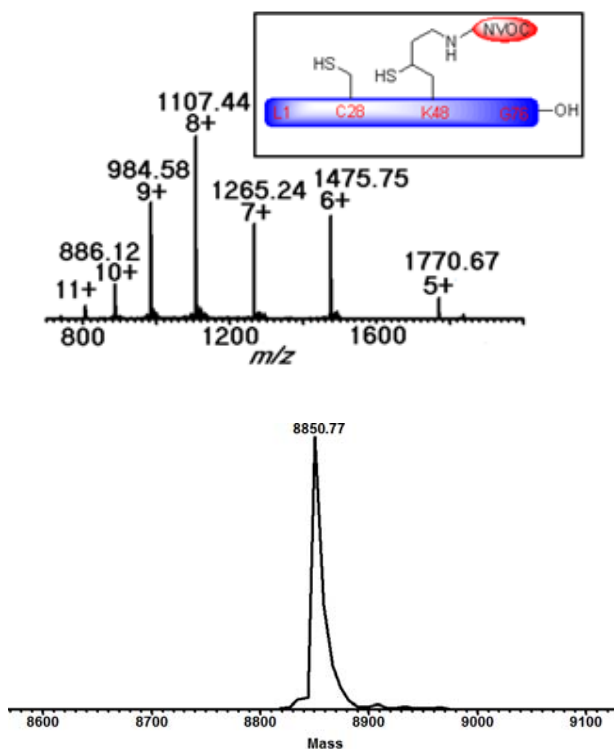

Fig. 4.8 The deconvoluted ESI-MS spectrum of **22**. Observed mass 8850.8 Da, calculated mass 8850.0 Da.

4.2.3 Removal of the photolabile protection group

To remove the photolabile protection group, 1.7 mg of **22** was dissolved in 60% acetonitrile aqueous solution (containing 0.045% TFA) at a concentration of 1 mg/mL. The solution was irradiated with 365 nm UV light. The deprotection process was monitored with ESI-MS. After 2 h, ESI-MS confirmed that the NVOC group was completely removed to afford the product **23** (Fig. 4.9). The solution was diluted and subjected to C18 semi-preparative HPLC purification. 1.3 mg of **23** was obtained after lyophilization.
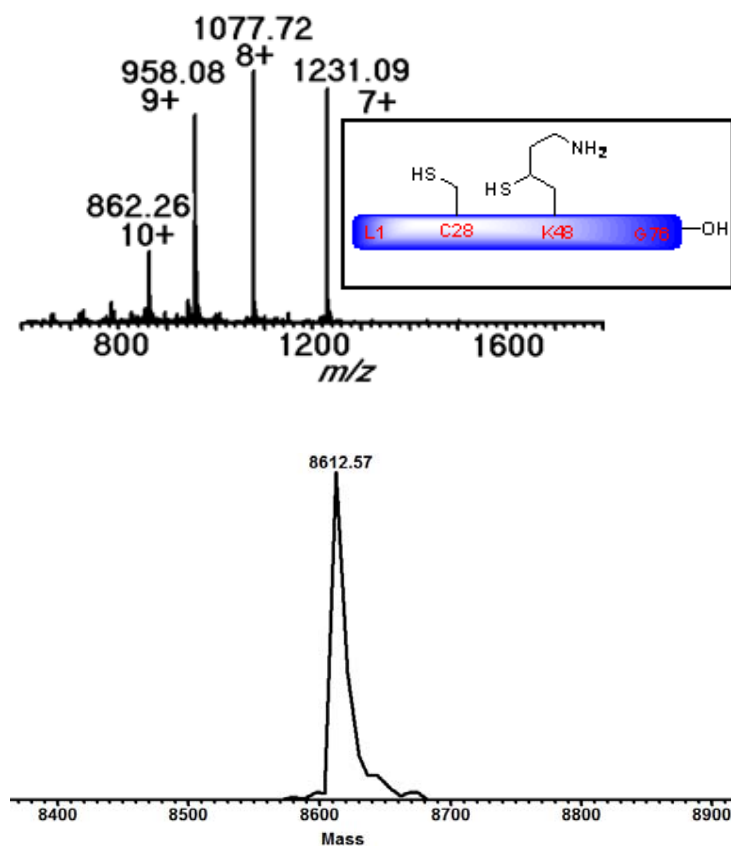

Fig. 4.9 The deconvoluted ESI-MS spectrum of **23**. Observed mass 8612.6 Da, calculated mass 8610.8 Da.

4.2.4 Synthesis of K48-linked diubiquitin 24

To synthesize K48-linked diubiquitin, ubiquitin **23** was reacted with Ub(1-76)-MES (Fig. 4.10). About 1.3 mg of **23** and 1.5 mg of Ub(1-76)-MES were dissolved in 150 μL ligation buffer (6 M Gdn•HCl, 0.1 M phosphate, 40 mM TCEP, 1 % v/v benzyl mercaptan, pH 8.0). After 6 h, the ligation product was formed with a yield of about 65 % based on analytical HPLC. After another 4 h, the yield of ligation product increased slightly. 0.8 mg of ligation product was isolated after purification by HPLC. The purified ligation product was verified by ESI-MS. The raw and deconvoluted ESI-MS was shown in Fig. 4.11.
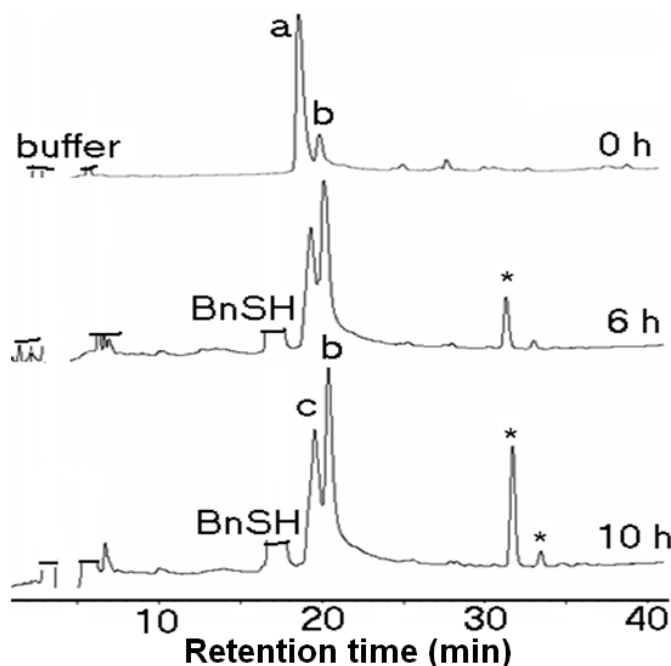

Fig. 4.10 C4 analytical HPLC monitored ligation between **23** and Ub(1-76)-MES at 0 h, 6 h and 10 h, respectively. Peak a: mixture of **23** and ubi(1-76)-MES. Peak b: ligation product. Peak c: mixture of ubi(1-76)-OH and small amount of remaining **23** and ubi(1-76)-SBn. Peak *: nonproteinous product. Note: 0 h HPLC was run before adding benzyl mercaptan. Bn = benzyl.
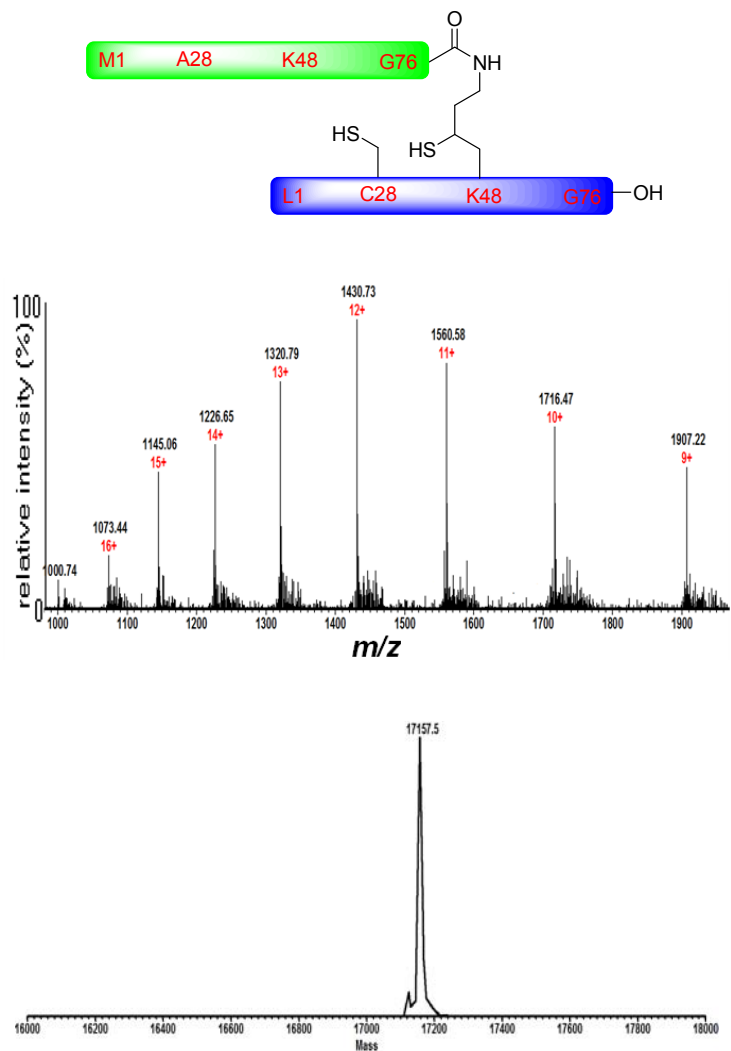
Fig. 4.11 ESI-MS of sulfur containing diubiquitin (before desulfurization). Observed mass 17157.5 Da, calculated mass 17157.6 Da.

To get the native K48-linked diubiquitin, free radical mediated desulfurization[30] was performed to convert cysteine 28 and 4-thiolysine 48 at the ligation junctions to alanine and lysine, respectively. The desulfurization was performed with VA-044 as free radical initiator[30] and glutathione as the hydrogen source[31]. After 9 h of treatment, both sulphur atoms on the two residues were removed, as confirmed by ESI-MS (Fig. 4.12). Deconvoluted ESI-MS showed the correct molecular weight of the final native K48-linked diubiquitin.
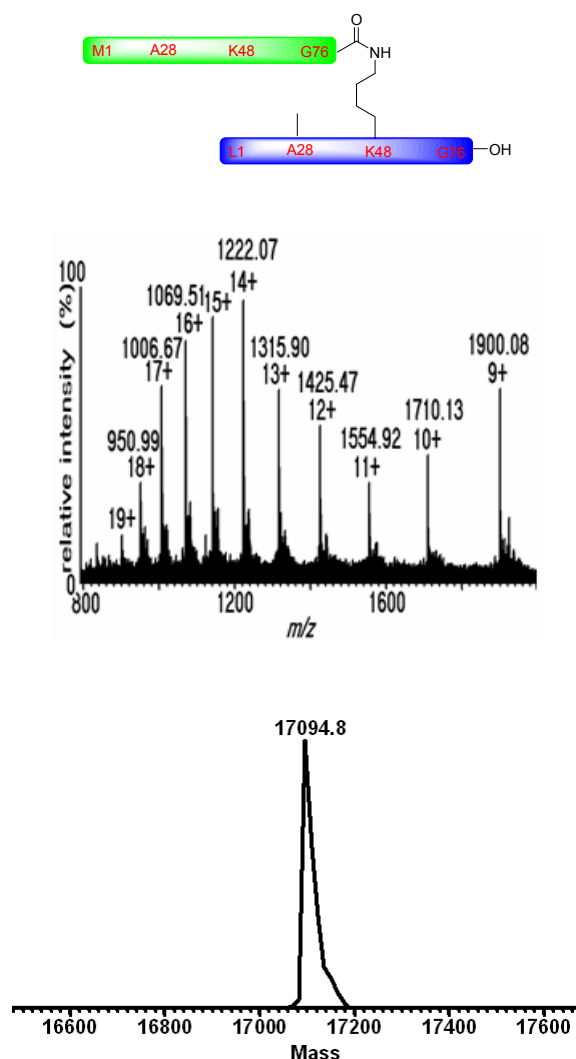
Fig. 4.12 The deconvoluted ESI-MS of K48-linked diubiquitin **24**. Observed mass 17094.8 Da, calculated mass 17093.4 Da.

For the characterization of the final product, the diubiquitin was checked with C4 analytical HPLC. HPLC profile showed a sharp and homogeneous peak (Fig. 4.13A). The diubiquitin was also analyzed with 18% SDS-PAGE. Coomassie blue staining showed a single band. Western blot with antibody FK2H, which is an HRP-conjugated antibody against mono- and polyubiquited conjugates but not free ubiquitin, detected the same band as the one in coomassie blue staining (Fig. 4.13B). To test whether the synthesized diubiquitin can be

folded to its native form, circular dichroism (CD) was measured with refolded diubiquitin **24**. CD spectrum indicated that the diubiquitin was well folded after dialysis (Fig. 4.13C).
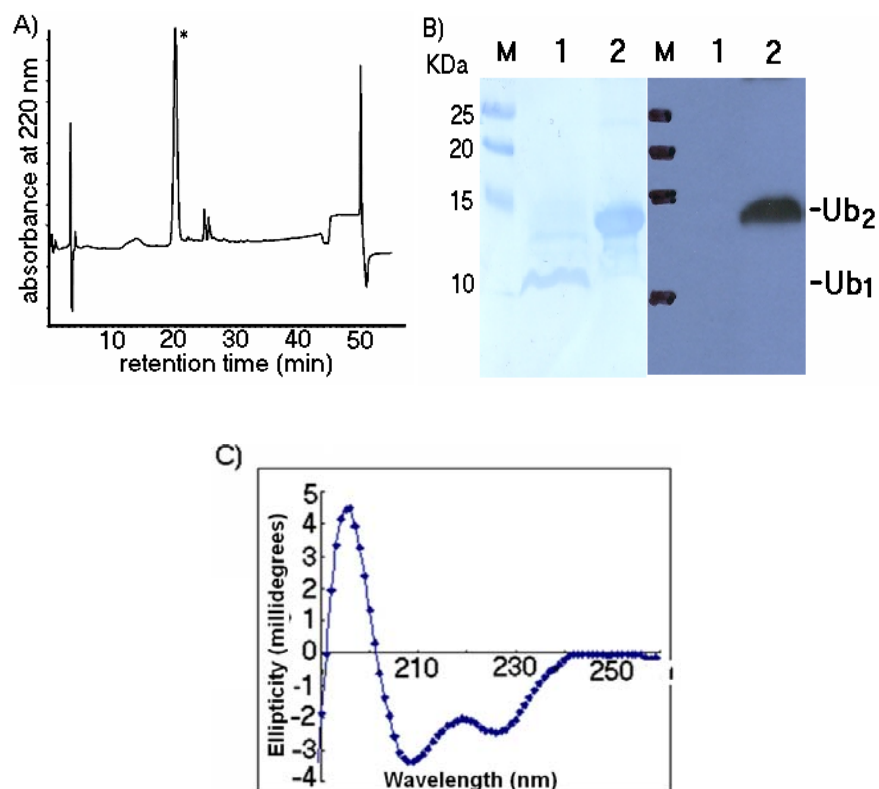
Fig. 4.13 Characterization of the synthesized K48-linked diubiquitin. A) C4 analytical HPLC. B) 18% SDS-PAGE analysis stained with coomassie blue (left) and western blot (right). M, protein marker; Lane 1, monoubiquitin; Lane 2, K48-linked diubiquitin **24**. C) CD spectrum of folded **24**.

4.3 Conclusion

In summary, herein we have optimized the dual native chemical ligation at lysine approach by replacing the strong acid-labile Cbz with the photolabile NVOC protecting group for the side-chain amine of 4-thiolysine. The synthesis of K48-linked diubiquitin using the improved protocol demonstrates the practical utility of this ligation strategy in synthetic protein chemistry.

4.4 Experimental procedures

4.4.1 General methods

Unless otherwise noted, all reactions were carried out in oven dried glassware under an atmosphere of nitrogen and distilled solvents were transferred by syringe. Solvents and reagents were purified according to the standard procedure prior to use. Evaporation of organic solutions was achieved by rotary evaporation with a water bath temperature below 40 °C. Product purification by flash column chromatography was accomplished using silica gel 60 (0.010-0.063 nm). Technical grade solvents were used for chromatography and distilled prior to use. NMR spectra were recorded at room temperature on a 300 MHz Bruker ACF 300, 400 MHz Bruker DPX 400 and 500 MHz Bruker AMX 500 NMR spectrometers, respectively. The residual solvent signals were taken as the reference (7.26 ppm for ^{1}H NMR spectroscopy and 77.0 ppm for ^{13}C NMR spectroscopy). Chemical shift (δ) is referred in terms of ppm, coupling constants (J) are given in Hz. Following abbreviations classify the multiplicity: s = singlet, d = doublet, t = triplet, q = quartet, m = multiplet or unresolved, br = broad signal. Infrared spectra were recorded on a Bio-RAD FTS 165 FT-IR Spectrometer and reported in cm[-1]. Samples were prepared in thin film technique. HRMS (ESI) spectra were recorded on a Finnigan/MAT LCQ quadrupole ion trap mass spectrometer, coupled with the TSP4000 HPLC system and the Crystal 310 CE system.

Amino acid derivatives, coupling reagents and resins were purchased from GL Biochem (Shanghai, China), Novabiochem and Chem-impex. All the other chemical reagents were purchased from Alfa Aesar, Sigma-Aldrich Chemical Company, Fisher Scientific, Acros Organics. PCR kit, restriction enzymes, and chitin beads for construction of ubiquitin-intein expression vector and the purification of intein-CBD fusion proteins were obtained from New

England Biolabs. The antibody FK2H was purchased from Enzo life sciences. C18 and C8 analytical HPLC analyses were performed by using an Agilent 1100 series instrument equipped with a Jupiter C18 (5 microm, 4.6 x 250 mm) and a Vydac MS C8 (5 microm, 4.6 x 250 mm) reverse-phase column with a flow rate of 1.0 mL/min. C4 analytical HPLC analyses were performed by using an Shimadzu UFLC system equipped with a Vydac 214MS C4 (5 microm, 4.6 x 250 mm) reverse-phase column with a flow rate of 1.0 mL/min. Detection was achieved with a UV-VIS-detector at wavelength λ = 220 nm. Semi-preparative purifications were performed with a Shimadzu system equipped with a Jupitar C18 column (5 microm, 10 x 250 mm) with a flow rate of 2.5 mL/min. Preparative purifications were performed with a Waters system equipped with a Prosphere C18 column (10 microm, 22 x 250 mm) with a flow rate of 10 mL/min. The buffer system for all the analysis was buffer A − H_2O (containing 0.045% TFA) and buffer B − 90% acetonitrile in H_2O (containing 0.04% TFA). Peptide masses were measured using a Thermo FINNIGAN LCQ Deca XP MAX equipped with ESI ion source.
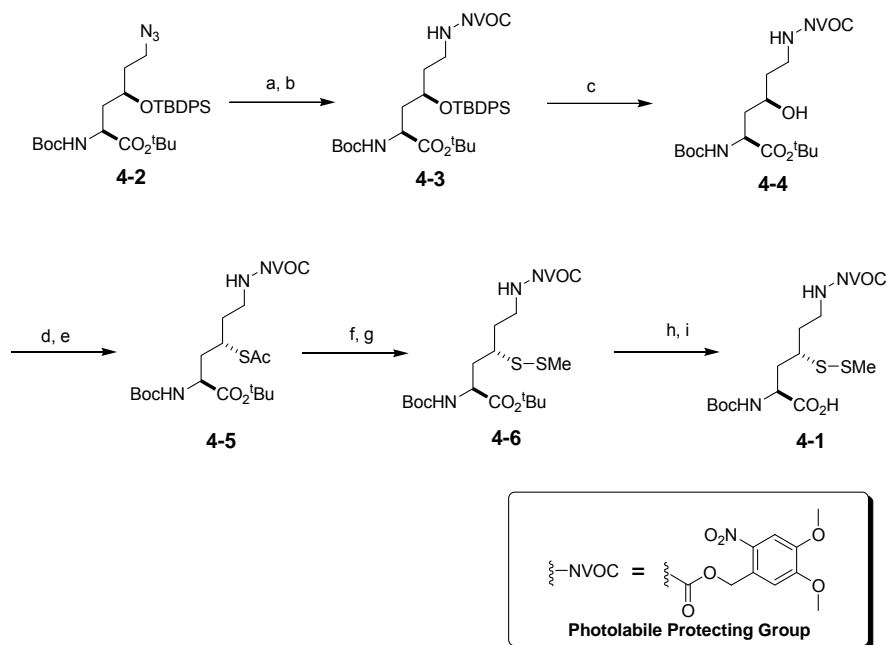
General scheme for the synthesis of Nε-NVOC protected 4-mercaptolysine derivative **4-1**. For the synthesis of **4-1**, Boc-Asp-OtBu was used as the starting material. All the steps before **4-2** were the same as previously reported (*J. Am. Chem. Soc.*, 2009, **131**, 13592). Reagents and conditions: (a) H$_2$, Pd/C, ethyl acetate, rt , 95%; (b) NVOC-Cl, Na$_2$CO$_3$, dioxane:water (2:1), 0 °C, 85%; (c) TBAF, THF, 0 °C, 75%; (d) Ms-Cl, DIPEA, 0 °C; (e) CH$_3$COSK, DMF, 40 °C, two steps 70% yield; (f) NaOH, MeOH, rt; (g) *S*-methyl methanethiosulfonate (MMTS), triethylamine, CH$_2$Cl$_2$, rt, two steps 60%; (h) TFA, H$_2$O, rt; (i) Boc$_2$O/TEA, MeOH, rt, two steps 73% yield.

(7R,9S)-tert-butyl-7-(tert-butyldiphenylsilyloxy)-1-(4,5-dimethoxy-2-nitrophenyl)-13,13-dimethyl-3,11-dioxo-2,12-dioxa-4,10-diazatetradecane-9-carboxylate (4-3):
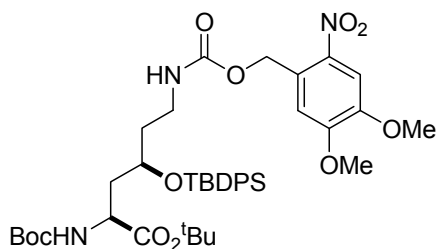
4-3

To a solution of azide **4-2** (0.5 g) in ethyl acetate (4.5 mL, 0.2M) was added 10% Pd/C (0.15 g) at room temperature. The reaction mixture was stirred for 6 h under an H$_2$ atmosphere at room temperature then filtered through a short pad of celite and washed with chloroform\methanol (1:1). The solvent was removed under reduced pressure (quantitative yield). This compound was direct taken for the next step without further purification.

Amine (0.45 g, 0.8 mmol) was dissolved in 2:1 ratio of dioxane\water (10 mL:5 mL) and cooled to 0 °C. Sodium carbonate (0.57 g, 1.76 mmol) was introduced portions wise at the same temperature. Then NVOC-Cl (0.24 g, 0.88 mmol) was added slowly to the reaction mixture. The reaction mixture was allowed to reach room temperature and stirred for 4 h. After being cooled to 0 °C, the reaction mixture was quenched by the addition of 0.5 N HCl (2 mL) and stirred for a further 5 min. The solution was dissolved in ethyl acetate and washed with water, and brine. The organic layer was dried over sodium sulphate and concentrated *in vacuo*, which was purified by column chromatography on silica gel to give desired product **4-3** (0.64 g, 85 %) as a pale yellow solid.

^{1}H NMR (500 MHz, THF-D$_8$): δ in ppm = 7.61-7.55 (m, 5H, H-Ph), 7.29-7.21 (m, 6H, H-Ph), 6.93 (s, 1H, H-Ph), 6.16-6.15 (m, 1H, NH), 6.07 (d, J = 8.1 Hz, 1H, NH), 5.26-5.19 (m,

2H, C$\underline{H}_2$-Ph), 4.09-4.04 (m, 1H, C$\underline{H}$-NH), 3.75 (bs, 1H, C$\underline{H}$-OH), 3.73 (s, 6H, OC$\underline{H}_3$), 2.91-2.86 (m, 2H, C$\underline{H}_2$-NH), 1.86-1.65 (m, 2H, C$\underline{H}_2$), 1.56-1.50 (m, 2H, C$\underline{H}_2$), 1.30 (s, 9H, 3C$\underline{H}_3$), 1.26 (s, 9H, 3C$\underline{H}_3$), 0.93 (s, 9$\underline{H}$, 3C$\underline{H}_3$).

^{13}C NMR (100 MHz, THF-D$_8$): δ in ppm = 171.7 (C=O), 155.7 (C=O), 155.4 (C=O), 153.4 (Ph), 148.1 (Ph), 139.8 (Ph), 135.8 (Ph), 135.8 (Ph), 133.7 (Ph), 133.1 (Ph), 129.9 (Ph), 129.8 (Ph), 128.1 (Ph), 127.7 (Ph), 127.7 (Ph), 110.4 (Ph), 108.1 (Ph), 81.6 ($\underline{C}$Me$_3$), 79.5 ($\underline{C}$Me$_3$), 69.2 ($\underline{C}$H-OTBDPS), 63.3 ($\underline{C}$H$_2$-Ph), 56.3 (O$\underline{C}$H$_3$), 51.9 ($\underline{C}$H-NH), 38.2 ($\underline{C}$H$_2$-NH), 37.1 (CH$_2$), 36.0 (CH$_2$), 28.3 (3$\underline{C}$H$_3$), 27.8 (3$\underline{C}$H$_3$), 26.9 (3$\underline{C}$H$_3$), 19.3 ($\underline{C}$-Si).

IR (CHCl$_3$): ν_{max} = 3018, 1710, 1521, 1215, 669 cm^{-1}.

HRMS (ESI): *m/z*: calcd for C$_{41}$H$_{58}$N$_3$O$_{11}$Si: 796.3841; [*M*]$^+$ found: 796.3835.

(7R,9S)-tert-butyl-1-(4,5-dimethoxy-2-nitrophenyl)-7-hydroxy-13,13-dimethyl-3,11-dioxo-2,12-dioxa-4,10-diazatetradecane-9-carboxylate (4-4):
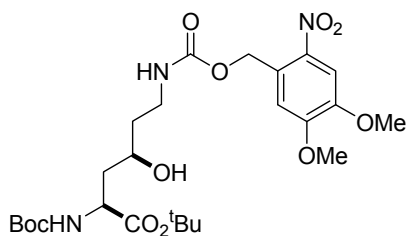
4-4

Compound **4-3** (0.1 g, 0.12 mmol) was dissolved in dry THF (3 mL, 0.04M) and cooled to 0 °C. 1 M solution of TBAF (0.19 ml, 0.19 mmol) was introduced via syringe at the same temperature. The reaction mixture was stirred for 10 h at 0 °C and quenched with saturated ammonium chloride solution. Evaporation of the solvent gave a residue, which was dissolved in ethyl acetate. The solution was washed with water, and brine. The organic layer was dried over sodium sulphate and concentrated *in vacuo*, which was purified by column chromatography on silica gel to give desired product **4-4** (0.053 g, 75 %) as a colorless oil.

¹H NMR (500 MHz, CDCl₃): δ in ppm = 7.70 (s, 1H, Ph), 7.02 (s, 1H, Ph) 5.71 (bs, 1H, NH), 5.50 (s, 2H, CH₂-Ph), 5.42 (d, J = 7.6, 1H, OH), 4.60 (bs, 1H, C<u>H</u>-OH), 4.34 (d, J = 8.8, 1H, C<u>H</u>-NH), 3.98 (s, 3H, OC<u>H</u>₃), 3.94 (s, 3H, OC<u>H</u>₃), 3.71 (bs, 1H, C<u>H</u>₂-NH), 3.55-3.49 (m, 1H, C<u>H</u>₂-NH), 3.25-3.24 (m, 1H, <u>CH</u>₂), 1.63 (d, J = 11.2, 1H, C<u>H</u>₂), 1.67-1.65 (m, 2H, C<u>H</u>₂), 1.46 (s, 9H, 3C<u>H</u>₃), 1.44 (s, 9H, 3C<u>H</u>₃).

¹³C NMR (100 MHz, CDCl₃): δ in ppm = 171.6 (C=O), 157.0 (C=O), 156.1 (C=O), 153.5 (Ph), 147.9 (Ph), 139.6 (Ph), 128.6 (Ph), 109.9 (Ph), 108.0 (Ph), 82.5 (C), 80.6 (C), 66.8 (<u>C</u>H₂-Ph), 63.3 (<u>C</u>H-OH), 56.4 (O<u>C</u>H₃), 56.3 (O<u>C</u>H₃), 51.0 (<u>C</u>H-NH), 41.9 (<u>C</u>H₂), 39.3 (CH₂), 35.7 (3<u>C</u>H₃), 28.2 (3CH₃), 27.9 (3CH₃).

IR (CHCl₃): ν_{max} = 3421, 3018, 1710, 1508, 1215, 669 cm⁻¹.

HRMS (ESI): m/z: calcd for $C_{25}H_{40}N_3O_{11}$: 558.2654; $[M]^+$ found: 558.2663.

(7S,9S)-*tert*-butyl-7-(acetylthio)-1-(4,5-dimethoxy-2-nitrophenyl)-13,13-dimethyl-3,11-dioxo-2,12-dioxa-4,10-diazatetradecane-9-carboxylate (4-5) :
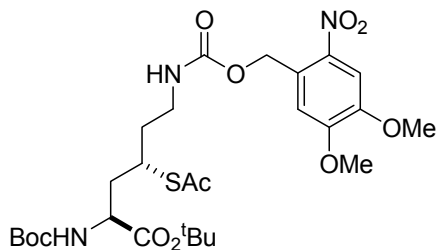
4-5

0.05 mL (0.27 mmol) of diisopropylethylamine (DIPEA) was added to a solution of 0.1 g (0.18 mmol) of alcohol in 3 mL (0.06M) of dichloromethane at room temperature. The solution was cooled to 0 °C, and 0.017 mL (0.21 mmol) of methanesulfonyl chloride was added. The reaction mixture was allowed to reach room temperature and stirred for 2 h. The reaction mixture was quenched with saturated ammonium chloride solution at 0 °C and diluted with dichloromethane. The layers were separated, the aqueous layer was extracted

with dichloromethane two times, and the combined organic extract was washed with water and brine. Drying over anhydrous sodium sulphate and removal of solvent under vacuum resulted in a colorless oily residue which was dissolved in DMF (3 mL, 0.05 M). A total of 0.06 g (0.5 mmol) of potassium thioacetate was added to the solution, which was heated to 40 °C for 12 h. After allowed to reach room temperature, water was added to the solution, which was extracted twice with ethyl acetate. The combined organic layer was washed with water and brine, dried over sodium sulphate and evoporated. The crude product was purified by column chromatography on silica gel to give desired product **4-5** (0.077 g, 70 %) as pale brown oil.

^{1}H NMR (400 MHz, CDCl$_3$): δ in ppm = 7.66 (s, 1H, Ph), 7.01 (s, 1H, Ph), 5.64 (bs, 1H, NH), 5.51-5.40 (m, 2H, CH$_2$-Ph), 5.24-5.22 (m, 1H, NH), 4.23-4.21 (m, 1H, C<u>H</u>-NH), 3.94 (s, 3H, OC<u>H</u>$_3$), 3.90 (s, 3H, OC<u>H</u>$_3$), 3.63-3.61 (m, 1H, C<u>H</u>-S), 3.37-3.34 (m, 1H, C<u>H</u>$_2$-NH), 3.14 (bs, 1H, C<u>H</u>$_2$-NH), 2.28 (s, 3H, CH$_3$), 2.04-2.01 (m, 2H, CH$_2$), 1.84-1.80 (m, 1H, C<u>H</u>$_2$), 1.67-1.57 (m, 1H, CH$_2$) 1.41 (s, 9H, 3C<u>H</u>$_3$), 1.40 (s, 9H, 3C<u>H</u>$_3$).

^{13}C NMR (100 MHz, CDCl$_3$): δ in ppm = 195.8 (COCH$_3$), 171.0 (C=O), 155.9 (C=O), 155.7 (C=O), 153.5 (Ph), 147.9 (Ph), 128.7 (Ph), 109.8 (Ph), 108.0 (Ph), 82.5 (C), 80.1 (C), 63.2 (CH$_2$-Ph), 56.4 (OC<u>H</u>$_3$), 56.3 (OC<u>H</u>$_3$), 52.0 (C<u>H</u>-NH), 39.8 (C<u>H</u>$_2$-NH), 38.3 (C<u>H</u>$_2$), 38.0 (CH$_2$), 33.5 (CH-S), 30.7 (CH$_3$) 28.2 (3C<u>H</u>$_3$), 27.9 (3CH$_3$).

IR (CHCl$_3$): ν_{max} = 3427, 3018, 1689, 1521, 1215, 669 cm^{-1}.

HRMS (ESI): *m/z*: calcd for C$_{27}$H$_{41}$N$_3$O$_{11}$SNa: 638.2360; [*M+Na*]$^+$ found: 638.2350.

(7S,9S)-tert-butyl-1-(4,5-dimethoxy-2-nitrophenyl)-13,13-dimethyl-7-(methyldisulfanyl)-3,11-dioxo-2,12-dioxa-4,10-diazatetradecane-9-carboxylate (4-6) :
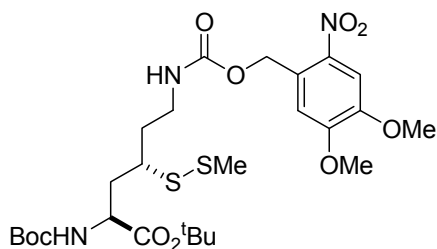
4-6

0.020 g of compound **4-5** was dissolved in 0.6 mL of methanol. 0.175 mL of 1 N NaOH was added. After 30 min at room temperature, the mixture was neutralized carefully with 1 N HCl at 0 °C. The mixture was concentrated and extracted with ethyl acetate three times. The combined organic layer was washed with water and brine, dried over sodium sulphate and evoporated.

The residue was dissolved in 0.4 mL of DCM and added dropwise to a mixture containing 0.4 mL of DCM, 0.006 mL of Et$_3$N and 0.015 mL of S-methyl methanethiosulfonate (MMTS) while stirring at room temperature. The mixture was continuously stirred for 30 min. The solvent was removed and the residue was dissolved in methanol/H$_2$O (3:1) and purified by C18 semi-preparative HPLC followed by lyophilization to give desired product **4-6** (12 mg) as a white solid.

^{1}H NMR (400 MHz, CDCl$_3$): δ in ppm = 7.59 (s, 1H, Ph), 7.03 (s, 1H, Ph), 6.43-6.42 (m, 1H, NH), 6.27 (d, J = 7.6 Hz, 1H, NH), 5.30 (s, 2H, CH$_2$-Ph), 4.09-4.04 (m, 1H, CH-NH), 3.80 (s, 3H, OCH$_3$), 3.76 (s, 3H, OCH$_3$), 3.47 (bs, 1H, CH-S), 3.21-3.05 (m, 1H, CH$_2$-NH), 3.02-2.97 (m, 1H, CH), 2.17 (s, 3H, CH$_3$), 2.00-1.85 (m, 2H, CH$_2$), 1.74-1.62 (m, 2H, CH$_2$) 1.32 (s, 9H, 3CH$_3$), 1.31 (s, 9H, 3CH$_3$).

¹³C NMR (100 MHz, CDCl₃): δ in ppm = 193.9, 171.0 (C=O), 155.5 (C=O), 153.8 (C=O), 148.3 (Ph), 139.7 (Ph), 128.5 (Ph), 110.1 (Ph), 107.9 (Ph), 80.6 (C), 78.2 (C), 62.4 (CH₂-Ph), 55.5 (OC̲H₃), 55.4 (OC̲H₃), 52.1 (C̲H-NH), 33.8 (C̲H₃-S), 29.5 (C̲H₂), 27.6 (3C̲H₃), 27.1 (3C̲H₃), 24.8 (CH-S).

IR (CHCl₃): v_{max} = 3423, 3018, 1215, 758, 669 cm⁻¹. **HRMS (ESI):** *m/z*: calcd for C₂₆H₄₁N₃O₁₀S₂Na: 642.2131; [*M+Na*]⁺ found: 642.2121.

(7S,9S)-1-(4,5-dimethoxy-2-nitrophenyl)-13,13-dimethyl-7-(methyldisulfanyl)-3,11-dioxo-2,12-dioxa-4,10-diazatetradecane-9-carboxylic acid (4-1) :
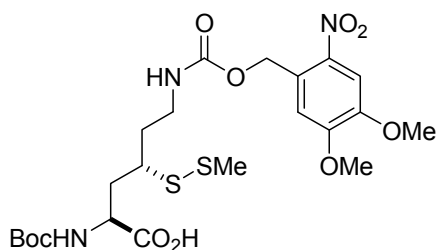
4-1

The white powder of **4-6** was dissolved in 0.5 mL 95% TFA. After 1.5 h at room temperature, the TFA was removed by evaporation. The residue was dissolved in 1 mL methanol/H₂O mixture (3:1). Adjust pH to about 8 with Et₃N. 0.02 mL of Boc₂O was added. After 3 h at room temperature, the sample was subjected to C18 semi-preparative HPLC. After lyophilization 8 mg of **4-1** were isolated.

¹H NMR (400 MHz, CDCl₃): δ in ppm = 10.72 (bs, 1H, COOH), 7.58 (s, 1H, Ph), 7.02 (s, 1H, Ph), 6.54 (bs, 1H, NH), 5.34 (d, *J* = 7.2 Hz, 1H, NH), 5.34-5.25 (m, 2H, CH₂-Ph), 4.25-4.21 (m, 1H, C̲H-NH), 3.79 (s, 3H, OC̲H₃), 3.75 (s, 3H, OC̲H₃), 3.28-3.22 (m, 1H, C̲H₂), 3.15-3.06 (m, 1H, C̲H), 2.78-2.76 (m, 1H), 2.29 (s, 3H, CH₃), 2.29-1.87 (m, 1H, C̲H₂), 1.85-1.83 (m, 2H, C̲H₂), 1.73-1.61 (m, 1H, C̲H₂), 1.32 (s, 9H, 3C̲H₃).

¹³C NMR (100 MHz, CDCl₃): δ in ppm = 172.8 (C=O), 155.6 (C=O), 153.8 (C=O), 148.3

(Ph), 139.7 (Ph), 128.2 (Ph), 128.5 (Ph), 110.1 (Ph), 107.9 (Ph), 78.2 (C), 62.4 (CH$_2$-Ph), 55.5, 55.4, 51.1 (CH-NH), 45.3 (CH$_2$-NH), 38.5 (CH$_3$-S), 32.8 (CH$_2$), 27.6 (3CH$_3$), 23.2 (CH-S).

IR (CHCl$_3$): v_{max} = 3444, 3427, 3018, 1708, 1635, 1215 cm^{-1}.

HRMS (ESI): *m/z*: calcd for C$_{22}$H$_{34}$N$_3$O$_{10}$S$_2$: 564.1686; [*M*]$^+$; found 564.1678.

4.4.2 Solid-phase peptide synthesis:

Synthesis of Ub(L1-K27)-COSCH$_2$CH$_2$CONH$_2$ (**18**) and Ub(Thz28-G47)-COS CH$_2$CH$_2$CONH$_2$ (**19**). **18** and **19** were synthesized manually employing standard *tert*-Butyloxycarbonyl (Boc) chemistry. The synthesis was started with 0.5 g of MBHA resin (0.9 mmol/g). S-trityl mercaptopropanoic acid (4 eq. PyBOP, 4 eq. acid, 12 eq. DIEA, preactivated in DCM for 2 min) was coupled onto the resin and served as the thiol linker. Trityl group with removed by repeated treatment with TFA/β-mercaptoethanol/TIS/DCM (5/2.5/2.5/90). The amino acids were then coupled one by one in a way similar to the loading of S-trityl mercaptopropanoic acid. The coupling was monitored with ninhydrin test. The amino acid derivatives used were Boc-Lys(2Cl-Z)-OH, Boc-Asn(Xan)-OH, Boc-Glu(OcHx)-OH, Boc-Thr(Bzl)-OH, Boc-Asp(OcHx)-OH, Boc-Ser(Bzl)-OH, Boc-Gln(Trt)-OH, Boc-Arg(Tos)-OH, Boc-Thz-OH. After sequence assembly, the Boc group was removed by treatment with 30 % TFA before cleavage. The peptide thioesters were cleaved by HF/*p*-cresol/anisole/ (9:0.5:0.5) for 1 h at 0 °C. The crude was harvested by ether precipitation and centrifugation. The peptides were purified by C18 preparative HPLC. The desired products were characterized with C18 analytical HPLC and ESI-MS.

Synthesis of K48(4-SSMe, NVOC)-G76-OH (**20**). The peptide was synthesized manually employing standard Fmoc chemistry started with 1 g of Wang resin (0.44 mmol/g). The C-terminal Gly was loaded by using 8 eq. DCC, 0.8 eq DMAP, 8 eq. Fmoc-Gly-OH in dry

DCM/DMF overnight. The loading was repeated for another 5 h and the resin was then capped with Ac$_2$O for 1 h. Fmoc group was removed with 20% piperidine in DMF. The following amino acids were coupled using 4 eq. PyBOP, 4. eq. amino acid, 8 eq. DIEA preactivated in DMF. On average, each coupling reaction lasted for 1.5 h. The coupling was monitored with ninhydrin test. The $N\alpha$-Fmoc protected amino acids used were Fmoc-Arg(Pbf)-OH, Fmoc-His(Trt)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Asp(OtBu)-OH. The Gly at the AspGly junction was coupled using Fmoc-(Hmb)Gly-OH (2 eq. PyBOP, 2 eq. amino acid, 4 eq. DIEA). The N-terminal Lys was coupled using Boc-Lys(4-SSMe, NVOC)-OH (**4-1**). For the coupling of **4-1**, 40 mg of **4-1**, 10.6 mg of HOBt and 16 mg of DCC were dissolved in minimum amount of dry DCM/DMF. The mixture was reacted with 300 mg of peptide resin for 3 h. After sequence assembly, the resin was treated with 20 % piperidine in DMF for 20 min to hydrolyze any possible acylation at hydroxyl group of Hmb group of (Hmb)Gly residue. The resin was then cleaved with TFA/TIS/H$_2$O (95/2.5/2.5) for 2.5 h. The crude peptide was harvested by ether precipitation and purified with C18 preparative HPLC. The desired product was analyzed with C18 analytical HPLC and ESI-MS.

4.4.3 Preparation of Ub(1-76)-MES

The construction of ubiquitin-intein fusion protein expression plasmid pTYB1-Ubi and the expression and MESNa thiolysis of ubiquitin-intein fusion were the same as we reported in the previous chapter. The purified product was analyzed with C4 analytical HPLC and ESI-MS.

4.4.4 Free radical mediated desulfurization

The desulfurization was performed under N_2. All the solutions were prepared under N_2 immediately before use. 0.8 mg of sulfur containing diubiquitin was dissolved in 300 μL buffer containing 6 M Gdn•HCl, 0.1 M phosphate, pH 6.5. 100 μL of 0.5 M TCEP solution (neutralized with NaOH) was added. 25 μL of 10 mM glutathione was added. 10 μL of 0.2M VA-044 was added. The solution was stirred at 37 ^{o}C for 6 h. 10 μL of 0.2M VA-044 was added to the mixture and the solution was continuously stirred for another 3 h. The final K48-linked diubiquitin (**24**) was purified by C18 semi-preparative HPLC.

4.4.5 Western blot analysis of K48-linked diubiquitin (24)

24 was dissolved in 8 M urea and analyzed with 18% SDS-PAGE. The samples on the gel were then electrotransferred to PVDF (polyvinylidene difluoride) membrane. The membrane was blocked by 5 % w/v non-fat milk powder in TBS buffer (100 mM Tris-HCl, 150 mM NaCl, pH 7.5) containing 0.1% Tween-20. A 1: 3000 diluted FK2H was used to detect **24** as visualized by chemilluminescence (SuperSignal West Dura Trial Kit, Pierce, USA).

4.4.6 Circular Dichroism (CD) measurement of K48-linked diubiquitin (24)

For the folding of K48-linked diubiquitin (**24**), 0.7 mg of **24** was dissolved in 0.1 mL buffer (6 M Gdn•HCl, 10 mM phosphate, 100 mM NaCl, pH 7.4) and serially dialyzed against 10 mM phosphate, 100 mM NaCl, pH 7.4 buffer containing decreasing concentration of Gdn•HCl. After folding, the CD of **24** was measured with Chirascan spectrometer with the final dialysis solution as baseline. The scan was performed between 180-260 nm and the passlength was 0.1 mm.

References

[1] O. Kerscher, R. Felberbaum, M. Hochstrasser, *Annu. Rev. Cell Dev. Biol.* **2006**, 22, 159.

[2] H. Wang, L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst, R. S. Jones, Y. Zhang, *Nature* **2004**, 431, 873.

[3] W. Zhou, P. Zhu, J. Wang, G. Pascual, K. A. Ohgi, J. Lozach, C. K. Glass, M. G. Rosenfeld, *Mol. Cell* **2008**, 29, 69.

[4] R. Pavri, B. Zhu, G. Li, P. Trojer, S. Mandal, A. Shilatifard, D. Reinberg, *Cell* **2006**, 125, 703.

[5] W. J. Cook, L. C. Jeffrey, M. Carson, Z. Chen, C. M. Pickart, *J. Biol. Chem.* **1992**, 267, 16467.

[6] C. L. Phillips, J. Thrower, C. M. Pickart, C. P. Hill, *Acta Crystallogr. D Biol. Crystallogr.* **2001**, 57, 341.

[7] R. Varadan, O. Walker, C. Pickart, D. Fushman, *J. Mol. Biol.* **2002**, 324, 637.

[8] M. J. Eddins, R. Varadan, D. Fushman, C. M. Pickart, C. Wolberger, *J. Mol. Biol.* **2007**, 367, 204.

[9] T. Tenno, K. Fujiwara, H. Tochio, K. Iwai, E. H. Morita, H. Hayashi, S. Murata, H. Hiroaki, M. Sato, K. Tanaka, M. Shirakawa, *Genes Cells* **2004**, 9, 865.

[10] A. B. Datta, G. L. Hura, C. Wolberger, *J. Mol. Biol.* **2009**, 392, 1117.

[11] D. Komander, F. Reyes-Turcu, J. D. Licchesi, P. Odenwaelder, K. D. Wilkinson, D. Barford, *EMBO Rep.* **2009**, 10, 466.

[12] M. Hochstrasser, *Curr. Opin. Cell Biol.* **1995**, 7, 215.

[13] C. Wang, L. Deng, M. Hong, G. R. Akkaraju, J.-I. Inoue, Z. J. Chen, *Nature* **2001**, 412, 346.

[14] M. L. Matsumoto, K. E. Wickliffe, K. C. Dong, C. Yu, I. Bosanac, D. Bustos, L. Phu, D. S. Kirkpatrick, S. G. Hymowitz, M. Rape, R. F. Kelley, V. M. Dixit, *Mol. Cell* **2010**, 39, 477.

[15] A. Williamson, K. E. Wickliffe, B. G. Mellone, L. Song, G. H. Karpen, M. Rape, *Proc. Natl. Acad. Sci. U.S.A.* **2009**, 106, 18213.

[16] J. Piotrowski, R. Beal, L. Hoffman, K. D. Wilkinson, R. E. Cohen, C. M. Pickart, *J. Biol. Chem.* **1997**, 272, 23712.

[17] R. M. Hofmann, C. M. Pickart, *J. Biol. Chem.* **2001**, 276, 27936.

[18] C. Chatterjee, R. K. McGinty, J.-P. Pellois, T. W. Muir, *Angew. Chem., Int. Ed.,* **2007**, 46, 2814.

[19] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *J. Am. Chem. Soc.* **2009**, 131, 13592.

[20] K. S. Ajish Kumar, M. Haj-Yahya, D. Olschewski, H. A. Lashuel, A. Brik, *Angew. Chem., Int. Ed.* **2009**, 48, 8090.

[21] R. K. McGinty, M. Köhn, C. Chatterjee, K. P. Chiang, M. R. Pratt, T. W. Muir, *ACS Chem. Biol.* **2009**, 4, 958.

[22] X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew. Chem., Int. Ed.* **2009**, 48, 9184.

[23] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, 266, 776.

[24] J. P. Tam, Y. Lu, C.-F. Liu, J. Shao, *Proc. Natl. Acad. Sci. U.S.A.* **1995**, 92, 12485.

[25] C. Chatterjee, R. K. McGinty, B. Fierz, T. W. Muir, *Nat. Chem. Biol.* **2010**, 6, 267.

[26] J. Chen, Y. Ai, J. Wang, L. Haracska, Z. Zhuang, *Nat. Chem. Biol.* **2010**, 6, 270.

[27] K. K. Pasunooti, R. Yang, S. Vedachalam, B. K. Gorityala, C.-F. Liu, X.-W. Liu, *Bioorg. & Med. Chem. Lett.* **2009**, 19, 6268.

[28] D. Alexeev, S. M. Bury, M. A. Turner, O. M. Ogunjobi, T. W. Muir, R. Ramage, L. Sawyer, *Biochem. J.* **1994**, 299, 159.

[29] D. Bang, G. I. Makhatadze, V. Tereshko, A. A. Kossiakoff, S. B. Kent, *Angew. Chem., Int. Ed.* **2005**, 44, 3852.

[30] Q. Wan, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2007**, 46, 9248.

[31] C. Haase, H. Rohde, O. Seitz, *Angew. Chem., Int. Ed.* **2008**, *47*, 6807.

# Chapter 5: N to C Sequential Ligation Using BMEA Peptide Building Blocks

5.1 Introduction

The advent of solid-phase peptide synthesis[1] (SPPS) and chemical ligation methods[for reviews, see 2-8, research articles, see 9-16] has revolutionized the practice of peptide and protein chemical synthesis and dramatically promoted biological structure–function study. For the total synthesis of proteins, two fragments are usually not sufficient to cover the full length of an average-sized protein due to the restriction of the length of the peptides that can be efficiently prepared by SPPS. Therefore, sequential or convergent ligation of multiple fragments is required.[17] Sequential ligation can be done either from C- to N-terminus[18-27] or in the reverse direction[27-34]. For instance, when native chemical ligation (NCL) is involved in the synthesis of a protein from C to N, it is necessary to protect the N-terminal cysteine residue of all the internal thioester fragments to prevent the undesired self-ligation or cyclization reactions.[18-27] The deprotection and purification steps after each round of ligation introduce extra labor and lower the overall yield. While the ligation from C to N is straightforward, the ligation in reverse direction is more challenging. To ensure that the N to C sequential ligation works, the C-terminal ligative moiety of middle Cys-peptide should remain intact during the first ligation and therefore will be available for next ligation step. Based on this principle, many N to C sequential ligation methods were developed by using two or more orthogonal ligation chemistries.[28, 29] When NCL was used as the sole chemistry for N to C sequential ligation, protein assembling was conducted either by using two thioesters with different reactivity as in kinetically controlled ligations[30, 31] or by using thioester precursors.[32-34] The thioester precursor approach may require additional manipulations to convert the precursors

to thioesters after each round of ligation, although the thioesters may also be formed in situ to avoid extra purification steps.[32-34]

In this chapter, we introduce a novel N to C sequential ligation approach based on the differential reactivity between a regular peptidyl thioester and *N,N*-bis(2-mercaptoethyl)-amide (BMEA) which was recently developed by our group[35, 36] and Melnyk's group[37-39]. In the previous studies performed by the two groups, it was found that peptidyl BMEA could be converted to a thioester through N-to-S acyl transfer.[35-39] Under mild acidic conditions (pH 4-6), peptidyl BMEA can (i) be converted to a thioester by exchanging with a thiol and (ii) ligate directly with a Cys-peptide through an in situ formed thioester.[35-39] Under alkaline conditions, it remains in its amide form with relatively low reactivity towards a Cys-peptide. We have also shown that BMEA-mediated ligation tolerates most amino acids (except the much hindered β-branched amino acids) at the C-terminal position.[35, 36] Based on these observations, we realize that, when a peptide thioester reacts with a Cys-peptidyl BMEA at neutral or basic pH, conventional thioester-Cys ligation will be dominant and the BMEA moiety will remain intact during the reaction. Therefore, we propose a novel three-segmental N-to-C sequential ligation strategy. As shown in Scheme 5.1, the first step is the ligation reaction between the N-terminal thioester peptide and the middle Cys-peptidyl BMEA segment. The second step is the BMEA mediated ligation between the product of the first step and the C-terminal Cys-peptide.
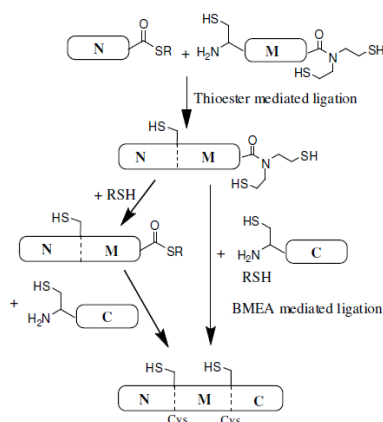
Scheme 5.1 The N to C sequential ligation scheme using the combination of NCL and BMEA mediated ligation.

5.2 Results

5.2.1 Model study for the demonstration of the N to C sequential ligation approach

To test our proposal, we first synthesized a 46-residue model peptide through ligating three short peptides. The sequences of the three peptides are shown in Fig. 5.1. The peptide **25** and **27** were synthesized using standard Boc (with MBHA resin) and Fmoc (with rink amide MBHA resin) chemistry, respectively. Peptide **26**, Cys-peptidyl BMEA was synthesized using the same method as previously reported.[35, 36] For the first ligation step, **25** and **26** were reacted under normal conditions. 15 mM of **25** and 5 mM of **26** were dissolved in ligation buffer containing 6 M guanidine hydrochloride (Gdn•HCl), 0.2 M phosphate, 50 mM TCEP, 2% v/v thiophenol, pH 7.0. The ligation reaction was monitored with HPLC. After 4 h at room temperature, the reaction was completed with only minor side reactions (Fig. 5.1A). The ligation product was purified and subjected to next ligation step with peptide **27**. The 2nd ligation was performed under the optimal conditions for BMEA-mediated ligation. The above ligation product (5 mM) was reacted with peptide **27** (15 mM) in a buffer

containing 6 M Gdn•HCl, 0.2 M NaOAc, 50 mM TCEP, 0.2 M sodium 2-mercaptoethanesulfonate (MESNa), pH 5.0. The ligation was completed under microwave irradiation within 15 h (Fig. 5.1C) with >90% yield based on HPLC analysis. Therefore, this model study demonstrated the feasibility of our N to C sequential ligation strategy.

Peptide **25**: H-**ADKRAHHNALERKRRDHA**-S(CH$_2$)$_2$CONH$_2$
Peptide **26**: H-**CDSFHSLRDSY**-N(CH$_2$CH$_2$SH)$_2$
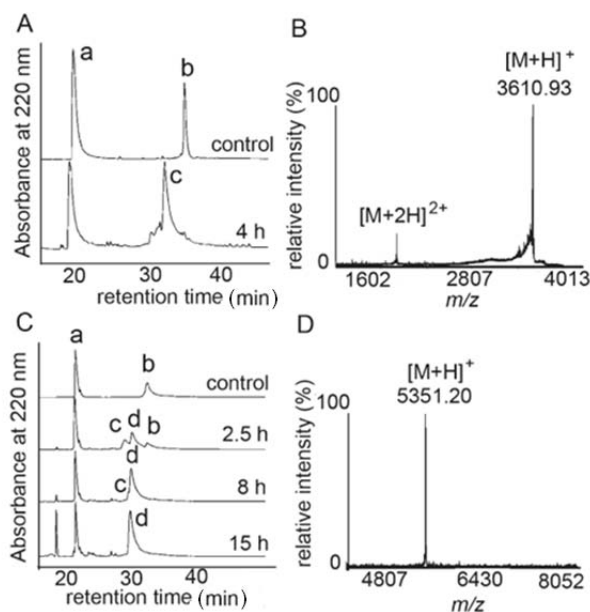Peptide **27**: H-**CLKPLHEKDSES**($_P$)**GGGKD**-NH$_2$
S(p): phosphoserine

Fig. 5.1 The model study of the N to C sequential ligation strategy. A): C18 analytical HPLC analysis of the ligation reaction between peptide **25** and **26**. HPLC condition: 0% to 40% of buffer B (90% acetonitrile, 0.045% trifluoroacetic acid in H$_2$O) in buffer A (0.045% trifluoroacetic acid in H$_2$O) in 40 min. Peak a: peptide **25**; peak b: peptide **26**; peak c: ligation product of **25** and **26**, H-**ADKRAHHNALERKRRDHACDSFHSLRDSY-**N(CH2CH2SH)2 or H-**AY29**-BMEA. B): The MALDI-TOF MS of H-**AY29**-BMEA. [M+H]$^+$ m/z found: 3610.93, calculated mass: 3609.71. C): C18 analytical HPLC analysis of the ligation reaction between H-**AY29**-BMEA and peptide **27**. HPLC condition: 0% to 50% of buffer B in buffer A in 50 min. Peak a: peptide **27**; peak b: H-**AY29**-BMEA; peak c: H-**AY29**-MES; peak d: ligation product, H-**ADKRAHHNALERKRRDHACDSFHSLRDSYCLKPLHEKDSES**(P)**GGGKD**-NH$_2$ or H-**AD46**-NH$_2$. D): The MALDI-TOF MS of H-**AD46**-NH$_2$. [M+H]$^+$ m/z found: 5351.20, calculated mass: 5349.50.

5.2.2 Chemical synthesis of ubiquitin using N to C ligation approach

Next, to test the general utility of our strategy, we applied this approach to the synthesis of a small protein, ubiquitin. Ubiquitin is a highly conserved protein with 76 amino acids and can be linked to the lysine side chain of another ubiquitin or other proteins in a process called ubiquitination. Previously, Kent's group has synthesized ubiquitin through the C to N sequential ligation of three segments with Ala28Cys and Ala46Cys as ligation junctions.[40] Recently, our group has also synthesized ubiquitin from C to N direction with Ala28Cys and Lys48 as ligation junctions.[41] Brik's group also reported the N-to-C synthesis of ubiquitin thioester for the purpose of construction of ubiquitin multimers.[34] Herein, we reported the N to C three-segmental synthesis of ubiquitin monomer using BMEA peptide as building block.

To synthesize ubiquitin with the N to C sequential ligation strategy, Ala28Cys and Ala46Cys were selected as the ligation junctions. After ligation, the two cysteine residues would be converted to Ala. The ubiquitin was divided into three segments which were synthesized by SPPS. Therefore, the middle segment, peptide **29**, was synthesized possessing C-terminal BMEA on a 2-chlorotrityl chloride resin using Fmoc chemistry.[35, 36] The characterization of these three segments by analytical HPLC and MS was shown in Fig. 5.2-5.4.
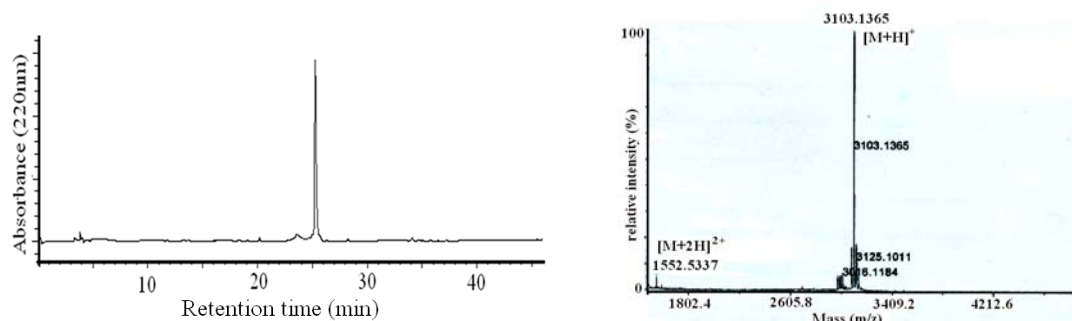
Fig. 5.2 C18 analytical HPLC profile (left) and MALDI-TOF MS spectrum (right) of peptide **28**, Ubi H-**LK$_{27}$**-SCH$_2$CH$_2$CONH$_2$. HPLC condition: 0% to 80% of buffer B in buffer A in 40 min. [M+H]$^+$ m/z found: 3103.1365; isotopic mass calcd: 3102.69 Da.
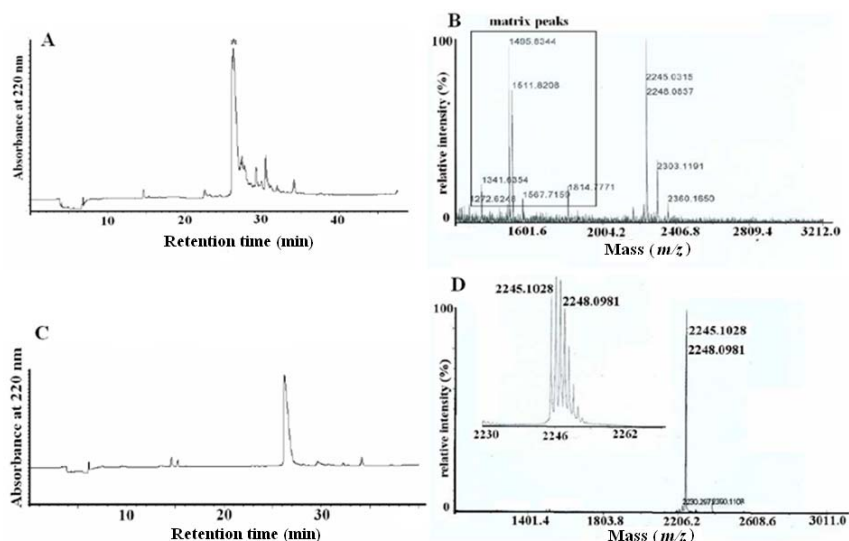

Fig. 5.3 A) and B): C18 analytical HPLC profile (A) and MALDI-TOF MS spectrum (B) of the crude peptide **29**, ubi H-**C$_{28}$-F$_{45}$**-BMEA. The asterisked peak in Fig. A is the desired product containing **29** as well as its BMEA oxidized form (disulfide form). The peaks highlighted with rectangle in Fig B are matrix peaks. The peak with mass 2245.0315 and 2248.0981 is the oxidized and reduced form of **29**, respectively. Peaks 2303.1191 and 2360.1650 are S-alkylated products with addition of one *t*-butyl and two *t*-butyl groups, respectively.

C) and D): C18 analytical HPLC profile (C) and MALDI-TOF MS spectrum (D) of the purified **29**. HPLC gradient: 0% to 80% of buffer B in buffer A in 40 min. [M+H]$^+$ m/z found: 2245.10 (oxidized) and 2248.10 (reduced). The calculated isotopic mass of the oxidized form is 2244.12 Da, reduced form 2246.14 Da.
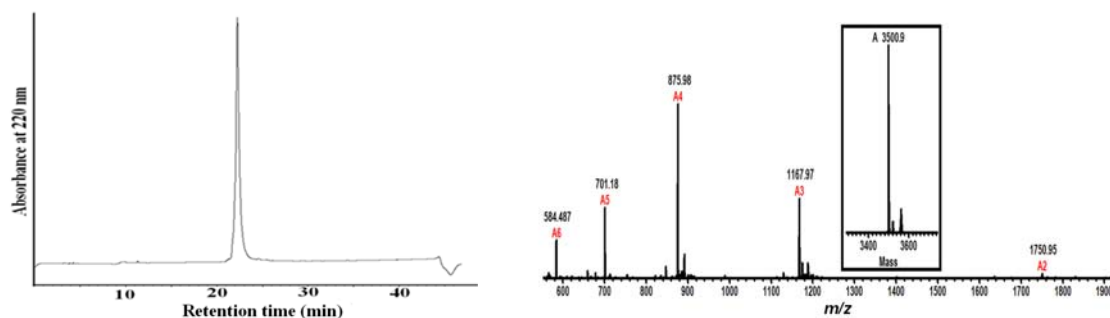
Fig. 5.4 C18 analytical HPLC profile (left) and ESI-MS spectrum (right) of peptide **30**, Ubi H-**C$_{46}$-G$_{76}$**-OH. HPLC gradient: 0% to 80% of buffer B in buffer A in 40 min. Observed mass: 3500.9, calculated mass: 3500.94 Da.

For the first thioester-mediated ligation step, 3.5 mg of the N-terminal peptide thioester **28** (final concentration: 3.8 mM) and 2.3 mg of the middle segment **29** (3.4 mM) were dissolved in 300 µL of ligation buffer containing 6 M Gdn•HCl, 0.2 M phosphate, 20 mM TCEP and 0.2 M MESNa, pH 8.0. As shown in Fig. 5.5A, the ligation was completed within 5 h at room temperature. No BMEA-originated side reactions were observed. Further incubation for 1.5 h did not change the HPLC profile in any significant way, indicating that the BMEA moiety was stable under these conditions. After purification and lyophilization, 2.5 mg of ligation product H-**LF$_{45}$**-BMEA was obtained (isolated yield 47 %).
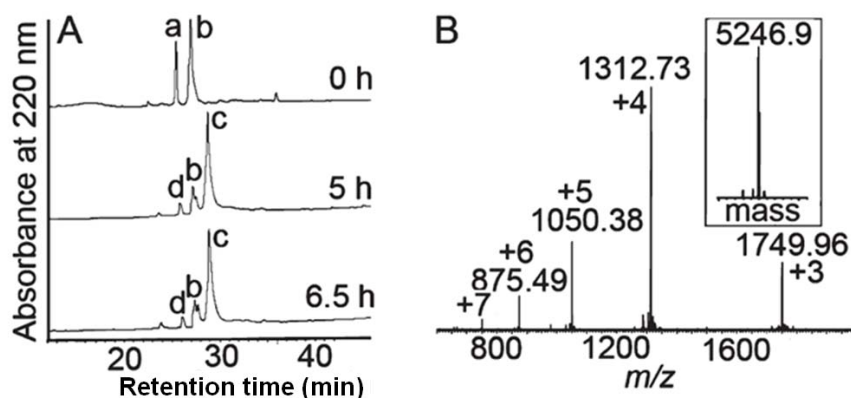

Fig. 5.5 Synthesis of ubiquitin using N to C sequential ligation. A): C18 analytical HPLC analysis of the ligation reaction between peptide **28** and **29**. HPLC condition: 0 % to 80 % of buffer B in buffer A in 40 min. Peak a: peptide **28**; peak b: peptide **29**; peak c: ligation product of **28** and **29**, H-**LF$_{45}$**-BMEA;

peak d: self cyclization (with the C-ter Lys side-chain amine) and hydrolysis product of **28**. B): The raw and deconvoluted ESI-MS of H-**LF₄₅**-BMEA. Observed mass: 5246.9, calculated mass: 5247.16.

For the second step of ligation, we had two options (Scheme 5.1). We could either convert the peptidyl BMEA to thioester and perform ligation between the isolated thioester and the C-terminal Cys-peptide **30** or let the peptidyl BMEA react directly with peptide **30**. We first explored the first option. To convert H-**LF45**-BMEA to thioester, about 3 mg of H-**LF45**-BMEA was dissolved in 300 μL buffer (containing 6 M Gdn•HCl, 0.2 M NaOAc, 0.2 M MESNa, 40 mM TCEP, pH 5.0) and irradiated with low-power microwave. After 10 h, the majority of the BMEA was converted to the thioester form with a small amount of hydrolysis product (Fig. 5.6). 1.5 mg of H-**LF₄₅**-MES was isolated after HPLC purification (isolated yield 50%).
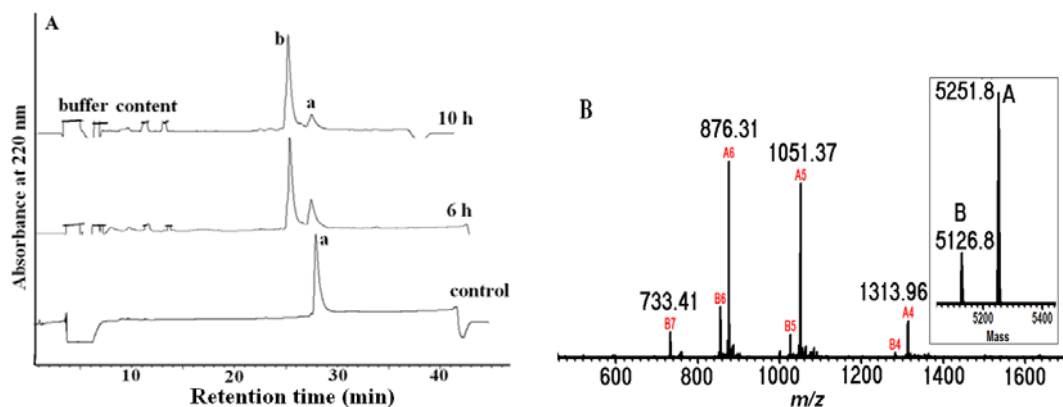
Fig. 5.6 The conversion of ubi H-**LF₄₅**-N(CH₂CH₂SH)₂ to ubi H-**LF₄₅**-MES through the exchange of BMEA moiety with MESNa under acid and microwave condition. A): C18 analytical HPLC monitored reaction at 6 h and 10 h. Peak a: H-**LF₄₅**-N(CH₂CH₂SH)₂; Peak b: H-**LF₄₅**-MES with small amount of hydrolysis product, H-**LF₄₅**-OH. HPLC gradient: 0 % to 80 % of buffer B in buffer A for 40 min. B): The raw and deconvoluted ESI-MS of peak b. Species A: H-**LF₄₅**-MES, Observed mass: 5251.8, calculated mass 5252.09; Species B: Ubi L1-F45-OH. Observed mass: 5126.8, calculated mass: 5127.91.

Next, H-**LF$_{45}$**-MES was reacted with peptide **30** under normal ligation conditions. 1.5 mg of H-**LF$_{45}$**-MES (1.4 mM) and 1.5 mg of **30** (2.1 mM) were dissolved in 200 μL of reaction buffer (6 M Gdn•HCl, 0.2 M phosphate, 20 mM TCEP and 0.2 M MESNa, pH 8.0). The reaction was completed within 6 h (Fig. 5.7). 1.4 mg of purified ligation product was obtained (isolated yield: 57%).
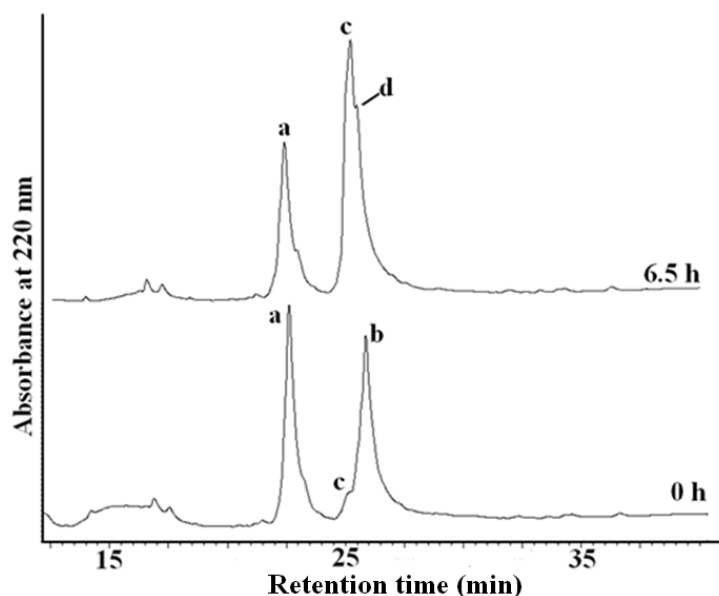

Fig. 5.7 C18 analytical HPLC monitored the ligation between isolated ubi L1-F45-MES with ubi C46-G-76-OH. Peak a: ubi C46-G76-OH; Peak b: Ubi L1-F45-MES (containing small amount of Ubi L1-F45-OH); Peak c: ligation product, Ubi L1-G76-OH A28C, A46C; peak d: Ubi L1-F45-OH. HPLC gradient: 0 % to 80 % of buffer B in buffer A for 40 min.

We then tested the direct ligation between H-**LF$_{45}$**-BMEA and **30** under microwave irradiation at pH 5.0. We first used MESNa as the thiol additive and found that when 1.5 to 2 equivalents of **30** relative to H-**LF$_{45}$**-BMEA were used, the ligation yield was modest. After 5 h irradiation, all H-**LF$_{45}$**-BMEA was transformed to H-**LF$_{45}$**-MES. But only 30-40% of the thioester reacted with **30** to form the ligation product. Continuing irradiation did not increase the yield (data not shown). The ligation reaction was more efficient when a larger excess

(about 3 equivalents) of **30** was used. After microwave irradiation for 10 h, more than 70% of

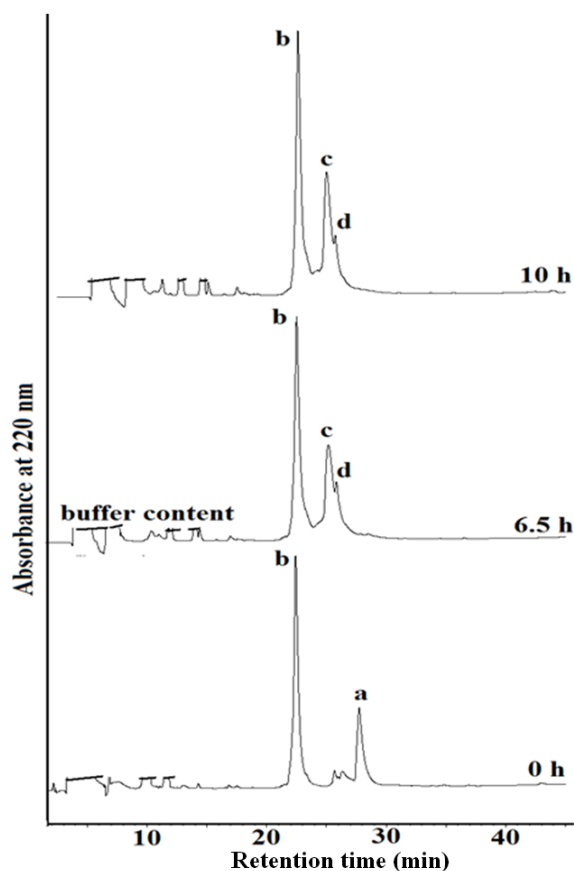H-**LF₄₅**-MES was converted to the ligation product (Fig. 5.8).

Fig. 5.8 C18 analytical HPLC monitored in situ ligation between Ubi H-L_1-F_{45}-BMEA with peptide **30** with MESNa as the thiol additive. The ligation was analyzed at 0 h, 6.5 h and 10 h, respectively. Peak a: H-L_1-F_{45}-BMEA; peak b: **30**; peak c: ligation product; peak d: H-L_1-F_{45}-BMEA and H-L_1-F_{45}-OH. HPLC gradient: 0 % to 80 % of buffer B in buffer A for 40 min.

Based on the above observations, we realized that MESNa might not be a good thiol additive for BMEA mediated ligation when very long peptide segments were involved. It appeared that, although BMEA to thioester conversion was not a problem, the resultant MES thioester would not react efficiently with the cysteinyl peptide at the weakly acidic pH. This is not unexpected considering that, at pH 5, the alkyl thiol of an N-Cys is not appreciably deprotonated for efficient reaction with an MES thioester which is not very reactive itself

because its negatively charged sulfonate moiety reduces the electrophilicity of the thioester carbonyl. Although aromatic thiols have been shown to be effective additives for NCL,[42] they might not be suitable for our case because i) their strong absorption at 220 nm and long retention time may complicate HPLC monitoring and ii) they may have poor solubility in acidic buffers. Taking these considerations into account, we set to look for a simple and more reactive alkyl thiol. We reasoned that methyl mercaptoacetate might be a good thiol additive as the presence of an electron withdrawing ester group near the thiol would make the resultant thioester much more reactive towards a Cys-peptide even at weakly acidic pH. To our delight, this alkyl thiol was indeed an excellent thiol additive in our reaction system. So, when 1 mg of H-**LF$_{45}$**-BMEA (1.9 mM) and 1.4 mg of **30** (4 mM) were mixed in 100 μL buffer (6 M Gdn•HCl, 0.2 M NaOAc, 2 % v/v methyl mercaptoacetate, 40 mM TCEP, pH 5.0), the ligation reaction was completed after microwave irradiation for 7 h (Fig. 5.9). Almost all the H-**LF$_{45}$**-BMEA was consumed and ligated with peptide **30**. Compared to MESNa, the use of methyl mercaptoacetate as the thiol additive gave a much more efficient ligation reaction.
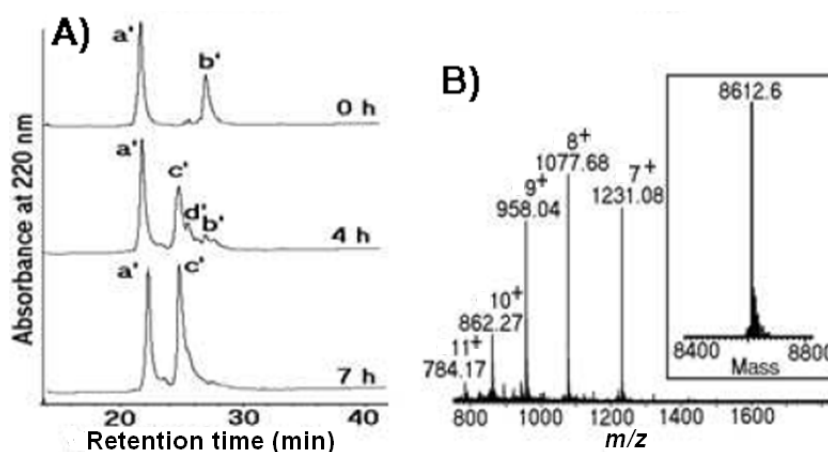

Fig. 5.9 A): C18 analytical HPLC analysis of the ligation reaction between H-**LF$_{45}$**-BMEA and peptide 6 with methyl mercaptoacetate as thiol additive. HPLC condition: 0 % to 80 % of buffer B in buffer A in

40 min. Peak a': peptide **30**; peak b': H-**LF₄₅**-BMEA; peak c': ligation product; peak d': H-**LF₄₅**-SCH₂COOMe. D): The raw and deconvoluted ESI-MS of full length ubiquitin. Observed mass 8612.6, calculated mass 8610.8.

To generate the native sequence of ubiquitin, free radical mediated desulfurization[43] was performed on the full-length ubiquitin synthesized by our N to C sequential ligation approach, which would convert the two cysteine residues at the ligation junctions to Ala residues. Desulfurization was performed with VA-044 as free radical initiator[43] and glutathione as the hydride source[44]. The desulfurization process was monitored with analytical HPLC and ESI-MS and was completed within 8 h (Fig. 5.10). The end product was purified by C18 semi-preparative HPLC. 1.1 mg of the final product was obtained from 1.5 mg starting material (73% isolated yield). Refolding of the lyophilized material was performed through dialysis to give the native folded protein as confirmed by CD analysis (Fig. 5.11).
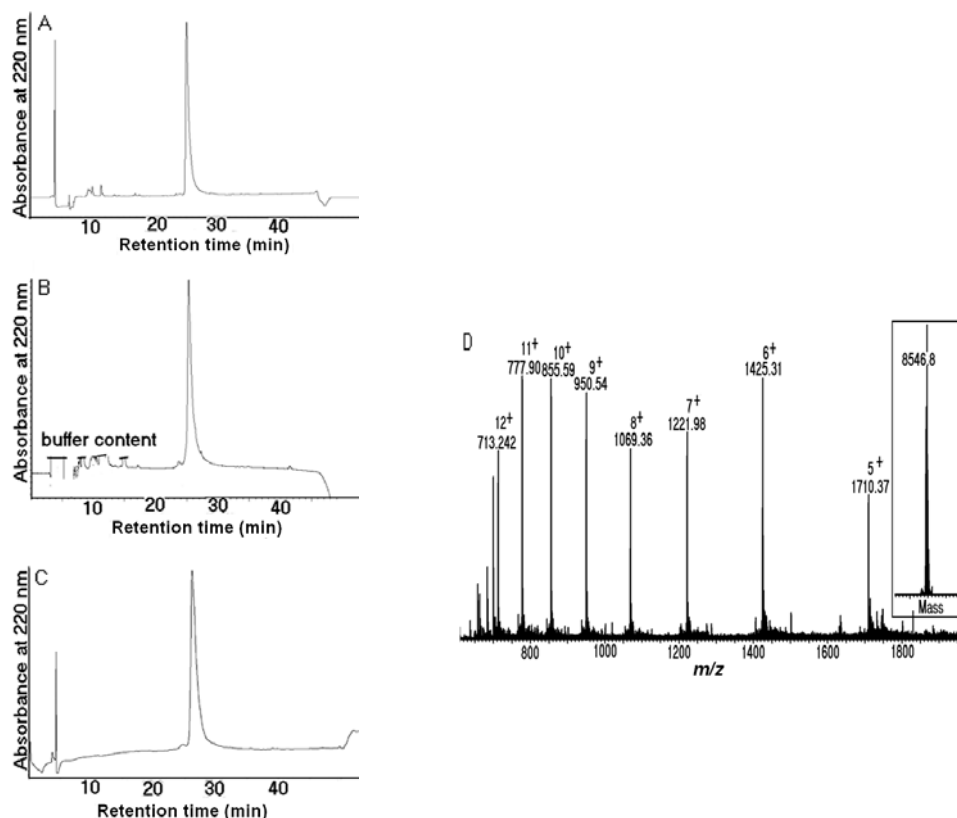

Fig. 5.10 The conversion of Cys28 and Cys46 in synthesized ubiquitin to Ala by free radical mediated desulfurization. A). C18 analytical HPLC of purified ubiquitin Cys28Cys46. B). C18 analytical HPLC

analysis of the desulfurization after 8 hours. C). C18 analytical HPLC of purified desulfurization product. D). The raw and deconvoluted ESI-MS of the desulfurization product. Observed mass 8546.8, calculated mass 8546.70.
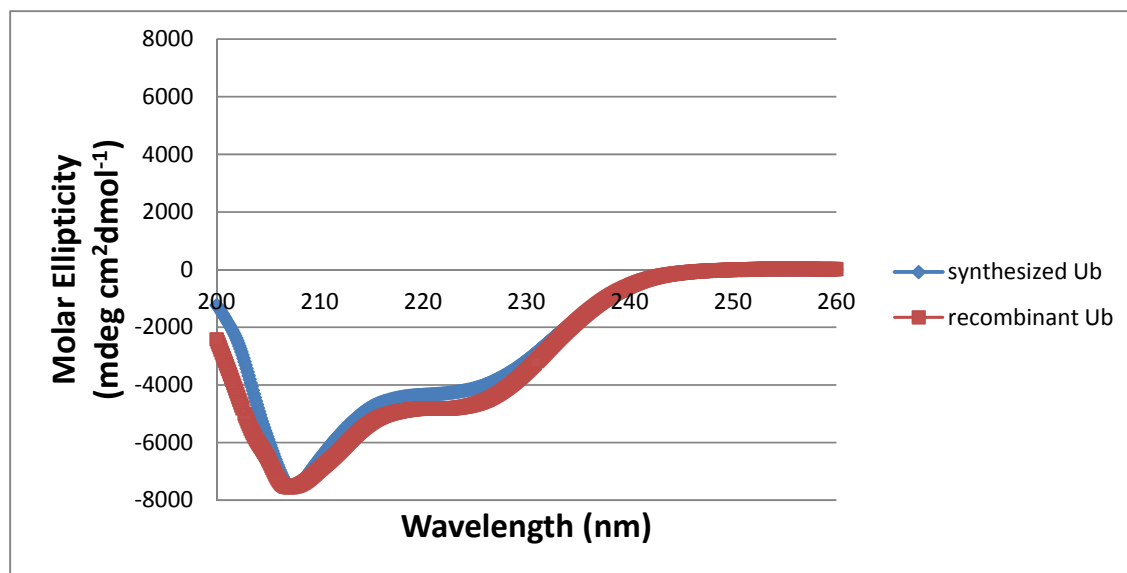
Fig. 5.11 The circular dichroism (CD) spectrum of the refolded synthesized and recombinant ubiquitin. Blue trace: synthesized ubiquitin; Red trace: recombinant ubiquitin.

5.3 Conclusion

In summary, we have demonstrated the utility of a new N- to C-terminus sequential ligation strategy which is based on the combined use of native chemical ligation and BMEA mediated ligation. For the key BMEA mediated ligation step, we have explored ways of ligation with an isolated thioester or direct ligation via an *in situ* generated thioester for which a new alkyl thiol, methyl mercaptoacetate, has been found to be an effective additive. In principle, this N-to-C sequential ligation can be performed for infinite steps if the BMEA is converted to an isolated thioester at every intermediate step. The BMEA ligation methodology works well not only with small peptides but also with large peptide segments.

With this and the fact that peptide BMEA segments can be readily synthesized by Fmoc chemistry, we believe that our newly developed N-to-C sequential ligation strategy has great value in protein synthesis.

5.4 Experimental procedures

5.4.1 General methods

Amino acid derivatives, coupling reagents and resins were purchased from Novabiochem and GL Biochem (Shanghai, China). All the other chemical reagents were purchased from Alfa Aesar, Sigma-Aldrich Chemical Company, Fisher Scientific, and Acros Organics. All the analytical HPLC analyses were performed using a Shimadzu HPLC system equipped with a Jupiter C18 (5 μm, 4.6 x 250 mm) reverse-phase column with a flow rate of 1.0 mL/min. Detection was done with a UV-VIS-detector at 220 nm. The purification was performed using a semi-preparative HPLC column (Jupiter C18, 5 μm, 10 x 250 mm) on a Shimadzu system with a flow rate of 2.5 mL/min. The buffer system for all the analyses was buffer A − H_2O (containing 0.045% TFA) and buffer B − 90% acetonitrile in H_2O (containing 0.045% TFA). Peptide and protein masses were measured using a Thermo FINNIGAN LCQ Deca XP MAX equipped with ESI ion source or a 4800 MALDI TOF/TOF Analyzer with α–cyano-4-hydroxycinnamic acid as the matrix. For the BMEA mediated ligation under microwave condition, the Eppendorf tube containing the reaction mixture (100 ~ 400 μL) was placed in a Panasonic microwave oven (model: NN-MX21WF) set at the lowest power level: the "Low" level which was below the "Defrost" level and with ~20% of the power output (800w). The temperature for microwave irradiated reaction was about 55 ºC.

5.4.2 Solid phase peptide synthesis

Synthesis of thioester peptides. H-**ADKRAHHNALERKRRDHA**-SCH$_2$CH$_2$CONH$_2$ (Peptide **25**) and Ubiquitin **L1-K27-**SCH$_2$CH$_2$CONH$_2$ (peptide **28**) were manually synthesized employing standard *t*-Boc chemistry. First, Trt-SCH$_2$CH$_2$COOH was coupled onto MBHA resin. The trityl group was removed by treatment with a cocktail containing TFA/TIS/β-mercaptoethanol/DCM (5:2.5:2.5:90). For the coupling of amino acids, Boc-amino acid (4 eq.) and benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate (PyBOP) (4 eq.) were dissolved in DCM. DIEA (12 eq.) was added in the solution. After 2 min of activation, the mixture was mixed with resin. The reaction was undertaken for 1.5 h. The coupling efficiency was checked with Kaiser test. The Boc group was removed by treatment with 30% TFA in DCM for 10 min, followed by 15 min. The side chain protected amino acid derivatives used were Boc-Asp(OBzl)-OH, Boc-Lys(2-Cl-Z)-OH, Boc-Arg(Tos)-OH, Boc-His(Tos)-OH, Boc-Asn(Xan)-OH, Boc-Glu(OBzl)-OH, Boc-Gln(Xan)-OH, Boc-Thr(Bzl)-OH, Boc-Ser(Bzl)-OH. After sequence assembly, peptide **25** was cleaved from the resin with a cocktail consisting TFMSA/TFA/*p*-cresol/methyl phenyl sulfide (1:7:1:1) for 1 h. Peptide **28** was cleaved from the resin with HF/*p*-cresol (9:1) at 0 ºC for 1.5 h. The crude peptides were precipitated with cold ether and purified with C18 semi-preparative HPLC.

Synthesis of peptidyl BMEA. Peptide H-**CDSFHSLRDSY**-N(CH$_2$CH$_2$SH)$_2$ (peptide **26**) and ubiquitin H-**C28-F45**-N(CH$_2$CH$_2$SH)$_2$ (peptide **29**) were synthesized as previously described (Hou, W.; Zhang, X.; Li, F.; Liu, C.-F. *Org. Lett.* **2011**, *13*, 386.). The first amino acid was loaded with 4 eq. of DIC, 4 eq. of HOAt and 4 eq. of Fmoc-AA-OH in dry DCM/DMF. The coupling of the remaining amino acids was done using the standard Fmoc chemistry with PyBOP as the coupling reagent. After synthesis, the peptides were cleaved from resin with a mixture containing 95% TFA, 1.5% EDT, 1.5 % TIS and 2% H$_2$O for 1 h or

3 h (if Arg was present). The crude peptide was precipitated with cold ether and purified with C18 semi-preparative HPLC.

Synthesis of the C-terminal Cys-peptide segments. The peptide H-**CLKPLHEKDSES(ₚ)GGGKD**-NH$_2$ (peptide **27**) and the ubiquitin H-**C46-G76**-OH (peptide **30**) were synthesized using the standard Fmoc chemistry. Rink amide MBHA resin (0.65 mmol/g) was used for the synthesis of peptide **27**. First, the Fmoc group was removed with 20% piperidine in DMF for 2 min, followed by 20 min. For the coupling reaction, 4 eq. of Fmoc-AA-OH, 4 eq. of PyBOP and 8 eq. of DIEA were dissolved in DMF. After preactivation for 2 min, the mixture was added to the resin. The coupling reaction was performed for 1.5 to 2 hours. The Fmoc amino acids used were Fmoc-Cys(Trt)-OH, Fmoc-Lys(Boc), Fmoc-His(Trt)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Ser(tBu)-OH. Fmoc-O-(benzylphospho)-L-serine was used for the installation of phosphorylated serine. Peptide **30** was synthesized using Wang resin (0.44 mmol/g). The C-terminal Gly was loaded by using 8 eq. DCC, 0.8 eq DMAP, 8 eq. Fmoc-Gly-OH in dry DCM/DMF overnight. The loading was repeated for another 5 h and the resin was then capped with Ac$_2$O for 1 h. Fmoc group was removed with 20% piperidine in DMF. The remaining amino acids were coupled using 4 eq. PyBOP, 4. eq. amino acid, 8 eq. DIEA preactivated in DMF. On average, each coupling reaction lasted for 1.5 h. The coupling was monitored with ninhydrin test. The *N*α-Fmoc protected amino acids used were Fmoc-Arg(Pbf)-OH, Fmoc-His(Trt)-OH, Fmoc-Thr(tBu)-OH, Fmoc-Ser(tBu)-OH, Fmoc-Glu(OtBu)-OH, Fmoc-Lys(Boc)-OH, Fmoc-Gln(Trt)-OH, Fmoc-Asn(Trt)-OH, Fmoc-Tyr(tBu)-OH, Fmoc-Asp(OtBu)-OH, Fmoc-Cys(Trt)-OH. The Gly at the AspGly junction was coupled using Fmoc-(Hmb)Gly-OH (2 eq. PyBOP, 2 eq. amino acid, 4 eq. DIEA). After sequence assembly, the resin was treated with 20 % piperidine in DMF for 20 min to remove Fmoc group and hydrolyze any possible

acylation at hydroxyl group of Hmb group of (Hmb)Gly residue. The resin was then cleaved with TFA/TIS/H2O/EDT (92.5/2.5/2.5/2.5) for 2.5 h. The crude peptide was precipitated with cold ether and purified with C18 preparative HPLC. The desired product was analyzed with C18 analytical HPLC and ESI-MS.

5.4.3 Free radical mediated desulfurization

The desulfurization was performed under N_2. All the solutions were prepared under N_2 immediately before use. 1.5 mg of sulfur containing ubiquitin was dissolved in 300 μL buffer containing 6 M Gdn•HCl, 0.1 M phosphate, pH 6.5. 50 μL of 1 M TCEP solution (neutralized with 5 M NaOH) was added. 25 μL of 10 mM glutathione was added. 20 μL of 0.2 M VA-044 was added. The solution was stirred at 37 °C for 5 h. 10 μL of 0.2 M VA-044 was added to the mixture and the solution was continuously stirred for another 3 h. The desulfurization reaction was analyzed with C18 analytical HPLC (Fig. S10). The final desulfurized ubiquitin was purified by C18 semi-preparative HPLC and lyophilized. 1.1 mg of final product was obtained.

5.4.4 Circular Dichroism (CD) measurement of refolded ubiquitin

For the refolding of ubiquitin, 1.1 mg of chemically synthesized or recombinant ubiquitin was dissolved in 0.2 mL buffer (6 M Gdn·HCl, 10 mM phosphate, 100 mM NaCl, pH 7.4) and dialyzed against buffer containing 10 mM phosphate, 100 mM NaCl, pH 7.4. After refolding, the CD measurements of ubiquitin were performed with JASCO-810 Spectropolarimeter with the final dialysis solution as baseline. The scan was performed between 190-260 nm and the pathlength was 1 mm. The CD data was converted to molar ellipticity (θ in mdeg*cm_2*dmol-1) according to the equation: $[\theta]_{molar} = \theta_{obs} / (nlc)$, where θ_{obs} is the CD signal in millidegrees, n is the number of peptide bonds, l is the path length in centimeters, and c is the concentration in decimoles per cm^3.

References:

[1] R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, 85, 2149.

[2] S. Aimoto, *Biopolymers* **1999**, 51, 247.

[3] J. P. Tam, Q. Yu, Z. Miao, *Biopolymers* **1999**, 51, 311.

[4] P. E. Dawson, S. B. H. Kent, *Annu. Rev. Biochem.* **2000**, 69, 923.

[5] M. E. Hahn, T. W. Muir, *Trends Biochem. Sci.* **2005**, 30, 26.

[6] C. P. R. Hackenberger, D. Schwarzer, *Angew. Chem., Int. Ed.* **2008**, 47, 10030.

[7] S. B. H. Kent, *Chem. Soc. Rev.* **2009**, 38, 338.

[8] R. J. Payne and C.-H. Wong, *Chem. Commun.* **2010**, 46, 21.

[9] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. H. Kent, *Science* **1994**, 266, 776.

[10] C.-F. Liu, J. P. Tam, *Proc. Natl. Acad. Sci. U.S.A.* **1994**, 91, 6584.

[11] C.-F. Liu, C. Rao, J. P. Tam, *Tetrahedron Lett.* **1996**, 37, 933.

[12] L. Z. Yan, P. E. Dawson, *J. Am. Chem. Soc.* **2001**, 123, 526.

[13] J. W. Bode, R. M. Fox, D. B. Kyle, *Angew. Chem., Int. Ed.* **2006**, 45, 1248.

[14] J. Chen, Q. Wan, Y. Yuan, J. Zhu and S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2008**, 47, 8521.

[15] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *J. Am. Chem. Soc.* **2009**, 131, 13592.

[16] T. Kawakami, S. Aimoto, *Tetrahedron* **2009**, 65, 3871.

[17] For review, see: J. Y. Lee, D. Bang, *Pept. Sci.* **2010**, 94, 441.

[18] C. F. W. Becker, C. L. Hunter, R. Seidel, S. B. H. Kent, R. S. Goody, M. Engelhard, *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 5075.

[19] D. Clayton, G. Shapovalov, J. A. Maurer, D. A. Dougherty, H. A. Lester, G. G. Kochendoerfer, *Proc. Natl. Acad. Sci. U.S.A.* **2004**, 101, 4764.

[20] D. Bang, S. B. H. Kent, *Angew. Chem., Int. Ed.* **2004**, 43, 2534.

[21] D. Bang, S. B. H. Kent, *Proc. Natl. Acad. Sci. U.S.A.* **2005**, 102, 5014.

[22] T. Kawakami, M. Tsuchiya, K. Nakamura, S. Aimoto, *Tetrahedron Lett* **2005**, 46, 5533.

[23] S. Ueda, M. Fujita, H. Tamamura, N. Fujii, A. Otaka, *ChemBiochem* **2005**, 6, 1983.

[24] C. Piontek, D. V. Silva, C. Heinlein, C. Pohner, S. Mezzato, P. ring, A. Martin, F. X. Schmid, C. Unverzagt, *Angew. Chem., Int. Ed.* **2009**, 48, 1941.

[25] J. C. Shimko, J. A. North, A. N. Bruns, M. G. Poirier, J. J. Ottesen, *J. Mol. Biol.* **2011**, 408, 187.

[26] S. Lahiri, M. Brehs, D. Olschewski, C. F. W. Becker, *Angew. Chem., Int. Ed.* **2011**, 50, 3988.

[27] K. Tsuji, A. Shigenaga, Y. Sumikawa, K. Tanegashima, K. Sato, K. Aihara, T. Hara, A. Otaka, *Bioorg. & Med. Chem.* **2011**, 19, 4014.

[28] X.-H. Tan, X. Zhang, R. Yang, C.-F. Liu, *ChemBiochem* **2008**, 9, 1052.

[29] X. Li, H. Y. Lam, Y. Zhang, C. K. Chan, *Org. Lett.* **2010**, 12, 1724.

[30] D. Bang, B. L. Pentelute, S. B. H. Kent, *Angew. Chem., Int. Ed.* **2006**, 45, 3985.

[31] J. S. Zheng, H.-K. Cui, G.-M. Fang, W.-X. Xi, L. Liu, *ChemBiochem* **2010**, 11, 511.

[32] T. Kawakami, S. Aimoto, *Tetrahedron Lett.* **2007**, 48, 1903.

[33] A. Shigenaga, Y. Sumikawa, S. Tsuda, K. Sato, A. Otaka, *Tetrahedron* **2010**, 66, 3290.

[34] L. A. Erlich, K. S. Ajish Kumar, M. Haj-Yahya, P. E. Dawson, A. Brik, *Org. Biomol. Chem.* **2010**, 8, 2392.

[35] W. Hou, R. Yang, F. Li, X. Zhang, C.-F. Liu, *The 11th Chinese International Peptide Symposium (CPS-2010),* July 5-8, 2010. Lanzhou, China.

[36] W. Hou, X. Zhang, F. Li, C.-F. Liu, *Org. Lett.* **2011**, 13, 386.

[37] N. Ollivier, J. Dheur, R. Mhidia, A. Blanpain, O. Melnyk, *Org. Lett.* **2010**, 12, 5238.

[38] J. Dheur, N. Ollivier, O. Melnyk, *Org. Lett.* **2011**, 13, 1560.

[39] J. Dheur, N. Ollivier, A. Vallin, O. Melnyk, *J. Org. Chem.* **2011**, 76, 3194.

[40] D. Bang, G. I. Makhatadze, V. Tereshko, A. A. Kossiakoff, S. B. H. Kent, *Angew. Chem., Int. Ed.* **2005**, 44, 3852.

[41] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *Chem. Commun.* **2010**, 46, 7199.

[42] E. C. B. Johnson, S. B. H. Kent, *J. Am. Chem. Soc.* **2006**, 128, 6640.

[43] Q. Wan, S. J. Danishefsky, *Angew. Chem., Int. Ed.* **2007**, 46, 9248.

[34] C. Haase, H. Rohde, O. Seitz, *Angew. Chem., Int. Ed.* **2008**, 47, 6807.

Final Discussion and Conclusion

During the four years of my Ph.D study, the two major contributions I have made to the peptide chemistry are the development of a new method for the synthesis of N-terminal tail-acetylated H4, especially H4 K16Ac,[1] and the development of a novel method for efficient synthesis of site-specifically ubiquitinated peptides and proteins.[2, 3] To emphasize the originality and timeliness of our work and the fluency of the presentation, I only reviewed those achievements before I initiated my projects or published my work in the relavent research fields. In the past two to three years, more achievements have been made by our group and other researchers in the same areas. Herein, I am going to review these achievements.

The acetylation at H4 K16, as one of the most important histone acetylations, modulates both higher chromatin structure and functional interactions between a nonhistone protein and the chromatin fiber.[4] Before my work, only one chemical ligation method has been reported for the synthesis of H4 K16Ac. The synthesis of H4 K16Ac was done through the ligation between a chemically synthesized thioester containing residues 1 to 22 and acetylated K16 and a recombinant C-terminal domain with R23C mutation.[4] Another method reported for the preparation of H4 K16Ac was the genetic incorporation of Nε-acetyllysine through the suppression of Amber nonsense condon.[5] After these two methods, we then reported the ligation/S-alkylation method for the synthesis of H4 K16Ac.[1] Recently, two new methods are developed for the site-specific installation of acetyllysine analogs. Philip A. Cole's group reported the site-specific installation of acetyllysine mimic by alkylating the cysteine residue in the recombinant H4 with methylthiol carbonylaziridine.[6] The acetyllysine mimic generated this way is still quite different from the native acetyllysine and is resistant to histone deacetylase. Recently, our group developed a new method for the installation of

acetyllysine analog though the thiol-ene coupling reaction between cysteine reside in recombinant H4 and N-vinylacetamide.[7] The cysteine residue at the desired position in the recombinant protein was converted to Nε-acetylaminoethylcysteine, which is a close acetyllysine analog.

In recent years, peptide chemists have made dramatic achievements in chemical ubiquitination. In 2007, Muir's team reported the first method for site-specific peptide ubiquination by using Nα-auxiliary mediated ligation.[8] Using this approach, an ubiquitinated H2B C-terminal peptide with native isopeptidic linkage was synthesized.[8] After that, the approach was applied to synthesis of full-length ubiquinated H2B through a two-step ligation approach.[9] As we mentioned earlier, the Nα-auxiliary mediated ubiquitination is slow due to the steric hindrance of the amine at ligation site. Based on Muir's pioneer work, we set to develop more efficient ways for chemical ubiquitination. In 2009, we reported the dual native chemical ligation at lysine for efficient site-specific ubiquitination.[2] In this approach, the thiol group of 4-thiolysine was used to mediate the ligation with thioester both at the α- and ε-amines of the thiolysine. The ligation at ε-amine can be used to install the ubiquitin tag by reacting with ubiquitin thioester. At the same time we published our results, Brik A. et. al. reported a similar approach for site-specific peptide ubiquitination by using 5-thiolysine to mediate the ligation of the ε-amine with ubiquitin thioester.[10]

All the three methods mentioned above generate native ubiquitin linkage. Meanwhile, many methods generating non-native ubiquitin linkages were also developed. Methods for the synthesis of proteins ubiquitinated with Gly76Ala ubiquitin were reported by two research groups independently.[11, 12] The ubiquitination was done by ligating ub(1-75)-thioester with either genetically incorporated[11] or chemically installed[12] Nε-cysteinyl-lysine. After ligation, the thiol of the cysteine was removed through desulfurization to convert to alanine.

Ubiquitination through disulfide bond linkage was also reported independently by two groups.[13, 14] In their approach, the ubiquitin was linked to the protein through mixed disulfide bond between ubiquitin-$CONHCH_2CH_2SH$ (generated by lysis of ubiquitin-intein fusion with cysteamine) and the cysteine residue in the recombinant proteins. These two forms of analogous ubiquitination were shown to excert similar functions as the native form as long as the functions are relied on the overall conformation of ubiquitin and ubiquitinated protein rather than the recognition of the isopeptidic linkage.

The chemical synthesis of diubiquitins or polyubiquitins is also an important research topic. As polyubiquitin chains are more difficult to synthesize, many research groups start with the synthesis of diubiquitins. Indeed, diubiquitin is the repeative unit of polyubiquitin. The study of diubiquitin can reveal many structural and functional effects of polyubiquitin. In 2009, Przybylski M. et. al. reported the first chemical synthesis of K63-linked diubiquitin through the thioether ligation of cyteinyl-ubiquitin peptide building blocks.[15] In 2000, our group and the other two groups independently reported three different ways of diubiquitin synthesis.[3, 16, 17] We applied the dual native chemical ligation to the synthesis of K48-diubiquitin.[3] The detailed synthesis is presented in chapter 4 of this thesis. Brik A. et. al. synthesized all seven kinds of diubiquitins through 5-thiolysine mediated ubiquitination.[16] In their approach, the 5-thiolysine was incorporated to the desired position of the monoubiquitin through a one-step ligation. Chin J. W. et. al. reported the synthesis of K6- and K29-linked diubiquitin through a global protection strategy.[17] The lysine involved in ubiquitination was incorporated as Nε-Boc-lysine to be distinguished from other lysines. All the other amines are protected with benzyloxylcarbonyl (Cbz). The Boc group was then deprotected to expose the amine which was then condensed with partially protected ubuiqitin thioester to form the isopeptide bond. Ovaa H et al. reported an approach for the synthesis of diubiquitin which is similar to Brik's

approach.[18] The main difference is that the 5-thiolysine was incorporated through total solid-phase synthesis rather than chemical ligation.

The synthesis of diubiquitin with non-native linkage was also reported. Scheffner M. et. al. reported the synthesis of diubiquitins through a triazole linkage generated through click reaction between genetically incorporated azide at the ubiquitin C-terminus and genetically incorporated alkyne at the side chain of the lysine involved in ubiquitination.[19]

Recently, researchers have made further achievements in chemical ubiquitination. Cropp T. A.'s group reported the controlled enzymatic synthesis of naturally-linkage, defined length polyubiquitin chains using lysines with removable protecting groups.[20] In their approach, the enzymatic ubiquitination was performed between a proximal ubiquitin with C-terminal diglycine deletion and a distal ubiquitin donor containing protected lysine which involved in ubiquitin linkage. After the formation the diubiquitin, the protected group was removed and the enzymatic ubiquitination reaction was repeated one more cycle to form a triubiquitin. Brik's group reported the first total chemical synthesis of K48-linked tetraubiquitin based on 5-thiolysine mediated ubiquitination.[21] The tetraubiquitin synthesized is probably the largest-sized protein ever synthesized by total chemical synthesis. Another achievement is the genetic incorporation of 5-thiolysine derivative for the ubiquitination reported by Chin's group.[22] As mentioned, 4- or 5-thiolysines can mediate efficient site-specific peptide and protein ubiquitination. However, the introduction of these thiolysines through total chemical synthesis or ligations is laborious work. The direct genetic incorporation of 5-thiolysine provides a much easier way to do so.

Despite the excellent achievements in chemical ubiquitination, the chemical synthesis of ubiquitinated proteins and polyubiquitin chains remain a challenging and labor-intensive task. In future, we set to develop more straightforward methods for chemical synthesis of ubiquitinated peptides and proteins.

References

[1] A. Allahverdi, R. Yang, N. Korolev, Y. Fan, C. A. Davey, C.-F. Liu, L. Nordenskiöld, *Nucleic Acids Res.* **2011**, 39, 1680.

[2] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *J. Am. Chem. Soc.* **2009**, 131, 13592.

[3] R. Yang, K. K. Pasunooti, F. Li, X.-W. Liu, C.-F. Liu, *Chem. Commun.* **2010**, 46, 7199.

[4] M. Shogren-Knaak, H. Ishii, J. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson, *Science* **2006**, 311, 844.

[5] H. Neumann, S. Y. Peak-Chewand, J. W. Chin, *Nat. Chem. Biol.* **2008**, 4, 232.

[6] R. Huang, M. A. Holbert, M. K. Tarrant, S. Curtet, D. R. Colquhoun, B. M. Dancy, B. C. Dancy, Y. Hwang, Y. Tang, K. Meeth, R. Marmorstein, R. N Cole, S. Khochbin, P. A. Cole, *J. Am. Chem. Soc.* **2010**, 132, 9986.

[7] F. Li, A. Allahverdi, R. Yang, G. B. J. Lua, X. Zhang, Y. Cao, N. Korolev, L. Nordenskiöld, C.-F. Liu, A direct method for site-specific protein acetylation. *Angew. Chem., Int. Ed.* Accepted manuscript.

[8] C. Chatterjee, R. K. McGinty, J.-P. Pellois, T. W. Muir, *Angew. Chem., Int. Ed.* **2007**, 46, 2814.

[9] R. K. McGinty, J. Kim, C. Chatterjee, R. G. Roeder, T. W. Muir, *Nature* **2008**, 453, 812.

[10] K. S. Ajish Kumar, M. Haj-Yahya, D. Olschewski, H. A. Lashuel, A. Brik, *Angew. Chem., Int. Ed.* **2009**, 48, 8090.

[11] X. Li, T. Fekner, J. J. Ottesen, M. K. Chan, *Angew. Chem., Int. Ed.* **2009**, 48, 9184.

[12] R. K. McGinty, M. Köhn, C. Chatterjee, K. P. Chiang, M. R. Pratt, T. W. Muir, *ACS Chem. Biol.* **2009**, 4, 958.

[13] C. Chatterjee, R. K. McGinty, B. Fierz, T. W. Muir, *Nat. Chem. Biol.* **2010**, 6, 267.

[14] J. Chen, Y. Ai, J. Wang, L. Haracska, Z. Zhuang, *Nat. Chem. Biol.* **2010**, 6, 270.

[15] E. J. Ji, H.-P. Wollscheid, A. Marquadt, M. Manea, M. Scheffner, M. Przybylski, *Bioconugate Chem.* **2009**, 20, 1152.

[16] K. S. A. Kumar, L. Spasser, L. A. Erlich, S. N. Bavikar, A. Brik. *Angew. Chem., Int. Ed.* **2010**, 49, 9126.

[17] S. Virdee, Y, Ye, D. P. Nguyen, D. Komander, J. W. Chin, *Nat. Chem. Biol.* **2010**, 6, 750.

[18] F. E. Oualid, R. Merkx, R. Ekkebus, D. S. Hameed, J. J. Smit, A. de Jong, H. Hilkmann, T. K. Sixma, H. Ovaa, *Angew. Chem., Int. Ed.* **2010**, 49, 10149.

[19] S. Eger, M. Scheffner, A. Marx, M. Rubini, *J. Am. Chem. Soc.* **2010**, 132, 16337.

[20] C. A. Castañeda, J. Liu, T. R. Kashyap, R. K. Singh, D. Fushman, T. A. Cropp, *Chem. Commun.* **2011**, 47, 2026.

[21] K. S. A. Kumar, S. N. Bavikar, L. Spasser, T. Moyal, S. Ohayon, A. Brik, *Angew. Chem., Int. Ed.* **2011**, 50, 6137.

[22] S. Virdee, P. B. Kapadnis, T. Elliott, K. Lang, J. Madrzak, D. P. Nguyen, L. Riechmann, J. W. Chin, *J. Am. Chem. Soc.*, **2011**, 133, 10708.