SITE-SPECIFIC PROTEIN FUNCTIONALIZATION USING ACTIVATED CYSTEINE BASED CHEMICAL BIOLOGY METHODS

A Dissertation

by

YUCHEN QIAO

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe R. Liu
Committee Members,	Marcetta Y. Darensbourg
	Jonathan T. Sczepanski
	Junjie Zhang
Head of Department,	Simon W. North

August 2021

Major Subject: Chemistry

ABSTRACT

Proteins, as one of the most essential biomacromolecules, involve in various of cellular processes such as enzymatic metabolism, small molecule transportation, immune response, DNA replication, etc. within living organisms. To obtain enough biological tools that assist scientists to study the cellular functions of human proteins, protein functionalization, in the other word, the incorporation of novel functional groups onto protein sequences has become more and more important and well-developed.

Proteins containing a *C*-terminal modification are critical to the protein synthesis via expressed protein ligation. They are usually made by recombinant fusion to intein. Although powerful, the intein fusion approach suffers from premature hydrolysis and low compatibility with denatured conditions. To totally bypass the involvement of an enzyme for expressed protein ligation, we developed an activated-cysteine directed protein ligation (ACPL) technique using 2-nitro-5-thiocyanatobenzoic acid (NTCB) as cysteine cyanylating reagent for undergoing nucleophilic acyl substitution with amines including a number of L- and D-amino acids and hydrazine. The afforded protein hydrazides could be used further for expressed protein ligation. We demonstrated the versatility of this approach with the successful synthesis of ubiquitin conjugates, ubiquitin-like protein conjugates, histone H2A with a *C*-terminal posttranslational modification, RNAse H that actively hydrolyzed RNA, and exenatide which is a commercial therapeutic peptide. The technique, which is exceedingly simple but highly

useful, expands to a great extent the synthetic capacity of protein chemistry and will therefore make a large avenue of new research possible.

Dehydroalanine (Dha) exists natively in certain proteins and can also be chemically made from a protein cysteine. As a strong Michael acceptor, dehydroalanine in proteins has been explored to undergo reactions with different thiolate reagents for making close analogues of posttranslational modifications (PTMs) including a variety of lysine PTMs. We explored an NTCB-triggered dehydroalanine formation which is highly efficient when cysteine is at the flexible *C*-terminal end of the protein. We produced ubiquitin and ubiquitin-like proteins containing a *C*-terminal dehydroalanine residue with high yields. Although this method was found not effective when cysteine is at an internal region of a protein, we believe this method will find broad applications in studying ubiquitin and ubiquitin like protein pathways and functional annotation of many PTMs in proteins such as histones.

ACKNOWLEDGEMENTS

My deepest gratitude is first to my advisor and committee chair, Dr. Wenshe R. Liu, who recruited me as a research assistant, provided me research guidance and supported me a lot for my academic career plan.

Secondly, I would like to express my heartfelt thanks to all my committee members, Dr. Marcetta Y. Darensbourg, Dr. Jonathan T. Sczepanski and Dr. Junjie Zhang for their great suggestions throughout my research projects and oral presentations. I would also like thank all previous and current lab members from the Liu research group for the unforgettable experience and useful suggestions.

Thirdly, I am also very grateful to my dear wife, Dr. Xuemei Yang, who graduated as a Doctor of Philosophy from Department of Chemistry, Texas A&M University in 2020 for her love and scientific support to me. I would like to thank my parents and parents-in-law for their patience and encouragement from the other side of the Earth.

Finally, I would like to thank my close friends, Yuting Luo and Xizhen Lian, as well as their young daughter, Weilin Lian who make my life more colorful during the holiday and weekends. I also want to thank Tingyuan Yang from Dr. Xin Yan's research group for our collaborative project and taking care of my cat, Pebbles. Special thanks to my high school friends, Zecheng He, Xiaoran Zhang, Yinzhuo Ding and Zhengyang Wang who organized and participated in our online gathering and entertainments to fulfill my Ph.D. life during the COVID-19 pandemic.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Dr. Wenshe R. Liu [advisor], Dr. Marcetta Y. Darensbourg and Dr. Jonathan T. Sczepanski of the Department of Chemistry and Dr. Junjie Zhang of the Department of Biochemistry & Biophysics.

The synthesis of dipeptides and the following ACPL reactions with propargylamine shown in Chapter II, 2.1 were done by Kaci Kratch in Liu's research group. The RNAseH expression, hydrazide ligation reaction and catalytic hydrolysis assay depicted in Chapter II, 2.4 were conducted by Dr. Ge Yu in Liu's research group. The cloning and expression of H2A-K129C-6H protein, as well as the H2A-H2B dimer folding assay described in Chapter II, 2.5 were performed by Dr. Wei Wang in Liu's research group. The expression vector for SUMO protease and pET28a-SUMO cloning vector were kindly provided by Dr. Pingwei Li from Department of Biochemistry & Biophysics at Texas A&M University. The cloning of Ub-G76C-6H and Ub-C-6H, and the cell stock preparation of Ub-KxC mutants mentioned in Chapter IV were conducted by Dr. Xiaoyan Wang in Liu's research group. The expression and purification of Ub-K48C and Ub-K63C in Chapter IV were performed by Dr. Ge Yu.

All other work conducted for this dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a teaching assistantship from Department of Chemistry at Texas A&M University and research assistantship from Dr. Wenshe R. Liu.

This work was also made possible in part by National Institutes of Health under Grant Number R01GM127575 and R01GM121584, and Welch Foundation under Grant Number A-1715.

NOMENCLATURE

Aa	Allylamine
ABC	Ammonium bicarbonate
ACPL	Activated Cysteine-Directed Protein Ligation
AMC	7-amido-4-methylcoumarin
Boc	tert-Butyloxycarbonyl
DCM	Dichloromethane
Dha	Dehydroalanine
DMSO	Dimethyl sulfoxide
DUB	Deubiquitinase
ESI-MS	Electrospray ionization mass spectrometry
ESI-LC-MS	Electrospray ionization liquid chromatography mass spectrometry
Fmoc	Fluorenylmethyloxycarbonyl
FPLC	Fast protein liquid chromatography
GndCl	Guanidinium chloride / Guanidine hydrochloride
^{1}H NMR	Proton nuclear magnetic resonance
Ha	Hydrazine
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High performance liquid chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
mmPylRS	*methanosarcina mazei* pyrrolysyl-tRNA synthetase

MWCO	Molecular weight cut-off
ncAA	Non-canonical amino acid
NTCB	2-nitro-5-thiocyanatobenzoic acid
Pa	Propargylamine
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PTM	Post-translational modification
SDS	Sodium dodecyl sulfate
SPPS	Solid phase peptide synthesis
TCEP	Tris(2-carboxyethyl)phosphine
TNB	Thionitrobenzoic acid
Tris	Tris(hydroxymethyl)aminomethane
uAA	Unnatural amino acid
Ub	Ubiquitin
Ubl	Ubiquitin-like protein
ULP	Ubiquitin-like protease

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

1.1. Overview of protein functionalization

Proteins as the most fundamental biomacromolecules in living organisms, are like gears providing basic but essential supports for our cellular machineries. Proteins are typically composed by twenty widely existing natural amino acids called canonical amino acids that encoded by universal genetic codons. Each amino acid contains a unique side chain such as hydroxyl group and thiol group where chemical reactions are able to occur. In order to expand protein diversity to ensure enough capability for massive cellular duties, post-translational modifications (PTM) are introduced to proteins after their biosynthesis using ribosome.[1-3] In mammalian cells, over 80% proteins are found post-translationally modified on most amino acids in more than 200 different formats.[4-5] Those delicate *in vivo* processes, including lysine alkylation[6-8], acylation[9-10] and ubiquitination[11-12], arginine methylation[13] and citrullination[14], serine and threonine phosphorylation[15-16] and glycosylation[17-18], cysteine oxidation[19] and nitrosylation[20], glutamic acid carboxylation[21] as well as *N*- and *C*-terminal modifications[22], *et al* (Figure 1), offer the possibility for modified substrates to regulate target protein activity, modulate DNA transcription, stabilize and destabilize protein structures, as well as control cellular signal transduction. This also makes the investigation of protein PTMs a popular research hotspot for scientists to interrogate biological processes in human bodies.

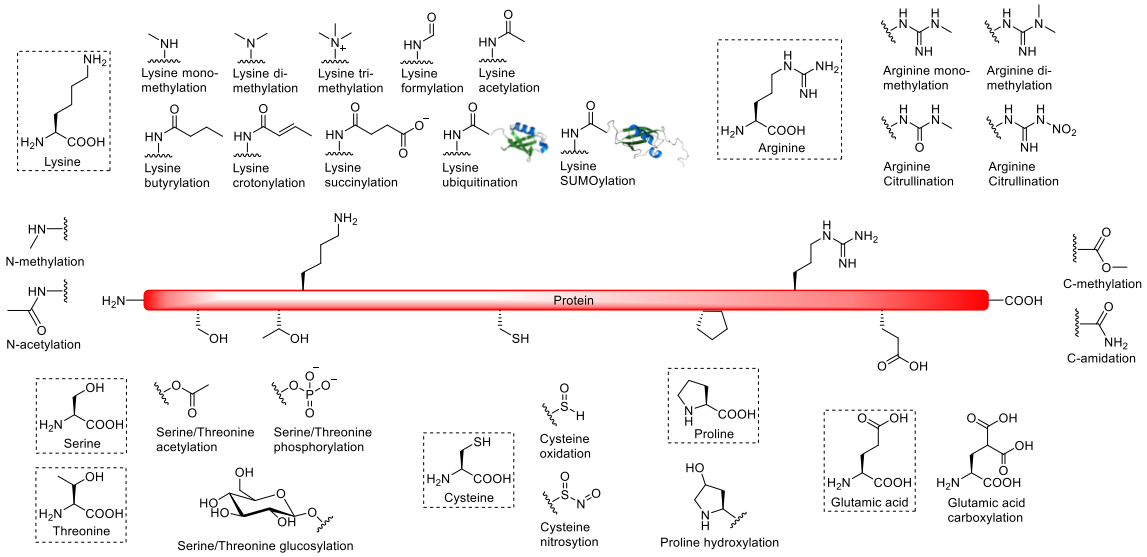

Figure 1 Common PTMs in nature.

Besides the studies of protein PTMs, many other research aspects such as structural biology, activity-based protein profiling and epigenetics also require vast scale of functionalized proteins. Inspired by native PTMs, selective installation of unique functional groups at specific amino acid positions has become an emerging direction to significantly expanded the protein toolbox for chemical biologists. Such concept, also known as protein functionalization, can be used to synthesized various modified proteins for specific research purposes.[23] Traditionally, like native PTMs which are usually obtained via enzymatic reactions, protein functionalization was also achieved with the assistance of enzymes. Although their high efficiency and selectivity benefit a lot to some specific protein modifications, the involvement of enzymes still leads to several shortages. The requirement of unique proteins sequences for each enzyme restricts its range of application.[24] Most of PTMs have low stoichiometry which means only several

molecules of a particular protein are modified at a given time.[25] In addition, currently discovered enzymes cannot generate of all types of functional groups. The enzymatic synthetic processes of many commonly existing PTMs also remain unknow. As the demand of novel protein-based biological tools building up remarkably, biorthogonal chemical reactions that can site-specifically applied on different proteins have been pushed towards the front stage instead of enzymatic approaches and quickly proved beneficial to the synthesis of functionalized proteins.

1.2. Chemical reactions on protein cysteine

Currently available chemical protein functionalization methods typically involve the modification reaction on a single amino acid, either a canonical amino acid such as lysine, or an unnatural amino acid (uAA). Comparing to uAAs that usually incorporated using genetic code expansion technique, canonical amino acids have become the better choices owing to their ease of expression and purification using bacterial protein over-expression system.[26-27] Chemoselective protein modifications strategies have been widely and successfully used to target lysine, cysteine, aromatic amino acids, and protein *N*-terminus amino group.[3,23,28-30] Unlike living organisms which only generate limited kinds of cysteine PTMs, cysteine has turned out to be one of the most popular residues for chemical reactions due to its relatively low natural occurrence.[31]

1.2.1. Nucleophilic substitution reactions

Cysteine is an amino acid that contains a strong nucleophilic side chain, thiol

group, which is significantly suitable for nucleophilic substitution reactions.[32-33] By carefully controlling the pH value of the reaction system, cysteine can be selectively deprotonated without the activation of other nucleophilic amino acids such as lysine and serine.[34] Cysteine alkylation reactions via nucleophilic substitution pathway can be traced to nearly a century ago. In 1935, Goddard and Michalis firstly reported that iodoacetic acid (IAA), iodoacetamide (IAM), iodoethyl alcohol and α-bromopropionic acid can be used for the α-halocarbonylation reaction with sulfhydryl proteins in order to achieve better solubility (Figure 2).[35] In later reports, IAM which was discovered as the

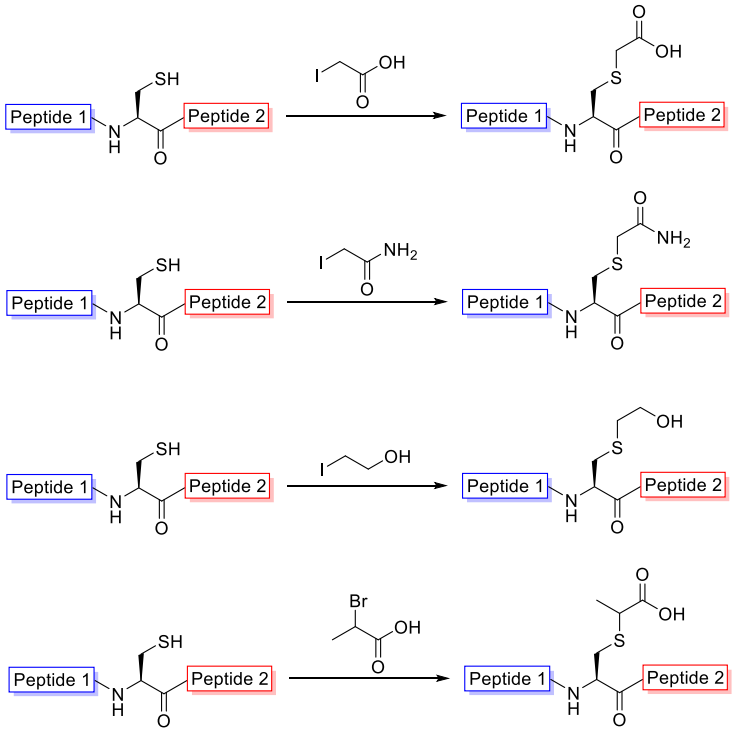

Figure 2 Cysteine alkylation via nucleophilic substitution using α-halocarbonyl compounds as electrophiles.

most powerful alkylation reagent among all its derivatives, were widely used in sample preparation of peptide mapping for cysteine containing therapeutic proteins.[36-37] Furthermore, IAM conjugated with *N*-acetyl-glucosamine was successfully used as a novel glycosylation reagent to selectively modify cysteines on proteins or peptides.[38]

Besides α-halocarbonylation, cysteine alkylation via halogen substituted ethylamine is another big aspect of halogen-based cysteine nucleophilic substitution reaction. Benefitting from the close mimic to lysine side chain with the only methyl to sulfur change using this specific construct, this type of reaction has been widely applied to generate proteins containing post-translationally modified lysine residues such as mono-, di-, or tri-methyl-lysine (Figure 3).[39] Moreover, in 2019, Liu's group reported a Cysteine-Aminoethylation-Assisted Chemical Ubiquitination (CAACU) method that combined the alkylation of cysteine using an *N*-alkylated 2-bromoethylamine and protein hydrazide ligation. The auxiliary group that used to achieve the requirement of native chemical ligation can be chemically removed. By using this method,

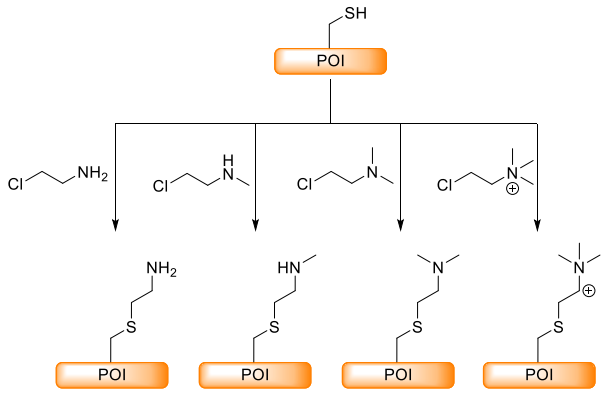

Figure 3 The incorporation concept of methylated lysine mimics on proteins using cysteine aminoethylation reactions.

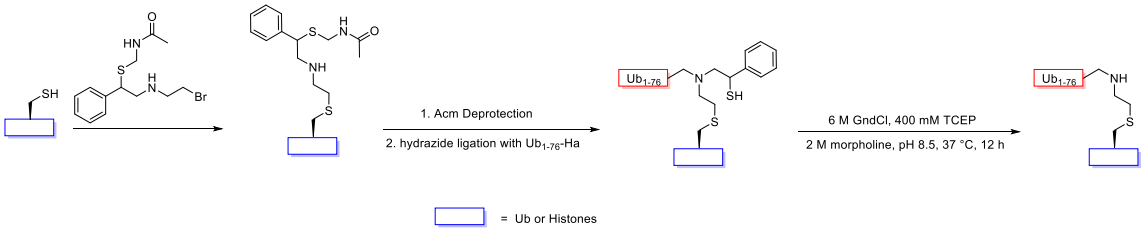

Figure 4 The scheme showing the aminoethylation step and auxiliary group deprotection step of CAACU method.

ubiquitination, one of the most common post-translational modification can be successfully installed onto recombinant histone proteins or ubiquitin (Figure 4).[40-41] The cysteine aminoethylation can also be achieved using ethylene imine made from 2-bromoethylamine in basic condition. In 2018, this alternative way was developed by Reilly and co-workers to identify the digested peptides of rituximab, an antibody.[42]

Inspired by cysteine mediated nucleophilic substitution reactions aforementioned, bifunctional linkers were then existed and widely developed. 1,3-dichloroacetone that generates a special acetone-based linkage was described by Wilkinson, *et al* and employed to synthesis ubiquitinated proteins (Figure 5).[43-44] Although, the manipulated cross-linking bonding contains one more atom than the native

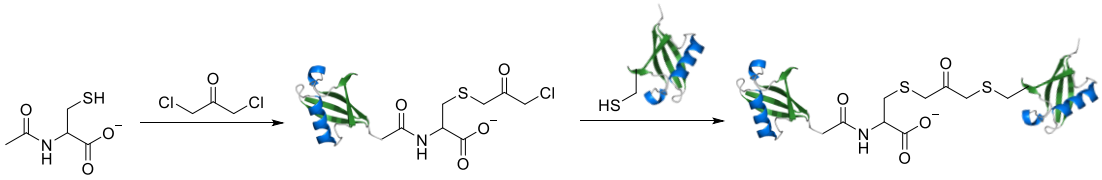

Figure 5 The scheme showing cysteine-based protein ubiquitination using 1,3-dichloroacetone as a bifunctional linker.

isopeptide linkage of protein ubiquitination, however, the polarity and spatial positioning are majorly preserved. In addition, this covalent bonding is exempted from the hydrolysis catalyzed by deubiquitinases which further allowed these mimics successfully retained stable ubiquitination linkages during biological assays.[44-45]

There were also other kinds of cysteine alkylation reactions reported in literature. For instance, an IR active bromo-benzyl nitrile was used to site-specifically modify cysteine residues in proteins for protein environment analysis (Figure 6).[46] Moreover, cysteine alkylation also occurs with aryl halides. A facile perfluoroarylation reaction that reacts with a couple of cysteine thiolates via nucleophilic aromatic substitution pathway to form rigid perfluoroaromatic staples was published by Pentelute and co-workers in 2013 (Figure 7). These stapled analogues have revealed enhanced abilities in protein-

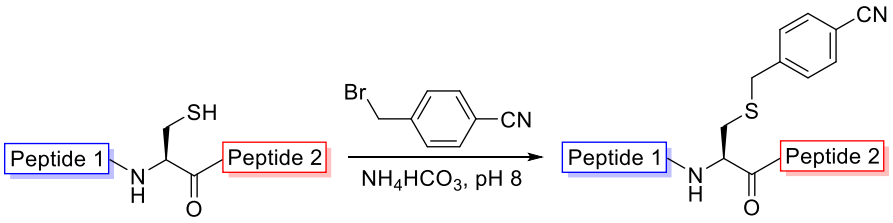

Figure 6 The formation of benzyl nitrile modified cysteine containing protein.

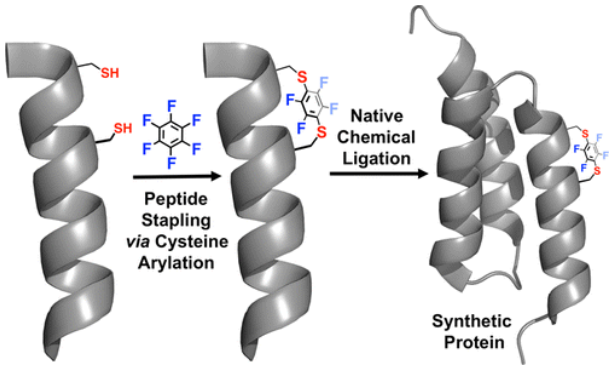

Figure 7 The idea of bi-cysteines perfluoroarylation reaction.[47]

protein interaction, peptide stabilization, and cell permeation compared to unmodified proteins.[47]

1.2.2. Michael addition reactions

Another major type of cysteine alkylation is Michael addition reaction. Maleimide derivatives were the first category that used for cysteine modification via Michael addition. In 1956, the first protein-based cysteine maleimidation was reported by Moore and Ward using *N*,*N*'-bis-maleimides as a bifunctional linker to covalently cross-link wool cysteines (Figure 8a).[48] *N*-ethyl maleimide (NEM) was also described as a cell membrane permeable agent by Bailey's research group that can undergo Michael addition reactions with cysteine side chain (Figure 8b).[49-50] *N*-(1-phenylethyl) maleimide (NPEM) which is more hydrophobic can provide five times better ionization ability of the cysteine thiol group (Figure 8b). However, because of in more uncontrollable side products as well as less-selectivity, NPEM was determined as a worse choice rather than NEM for cysteine maleimidation.[51] Recent discovery on maleimide-based cysteine alkylation reactions performed by Zhou's group has shown a novel cysteine multi-functionalization using 3-bromo-5-methylene pyrrolones (3Br-5MPs). According to a proposed addition-elimination mechanism, two cysteine residues are able to be linked together either intramolecularly or intermolecularly (Figure 9).[52]

Other than maleimides, vinyl sulfones were also broadly used to react with cysteine via Michael addition. The first example was given by Masri and Friedman in 1988 using aliphatic vinyl sulfone to react with cysteine.[53] It has been proved useful

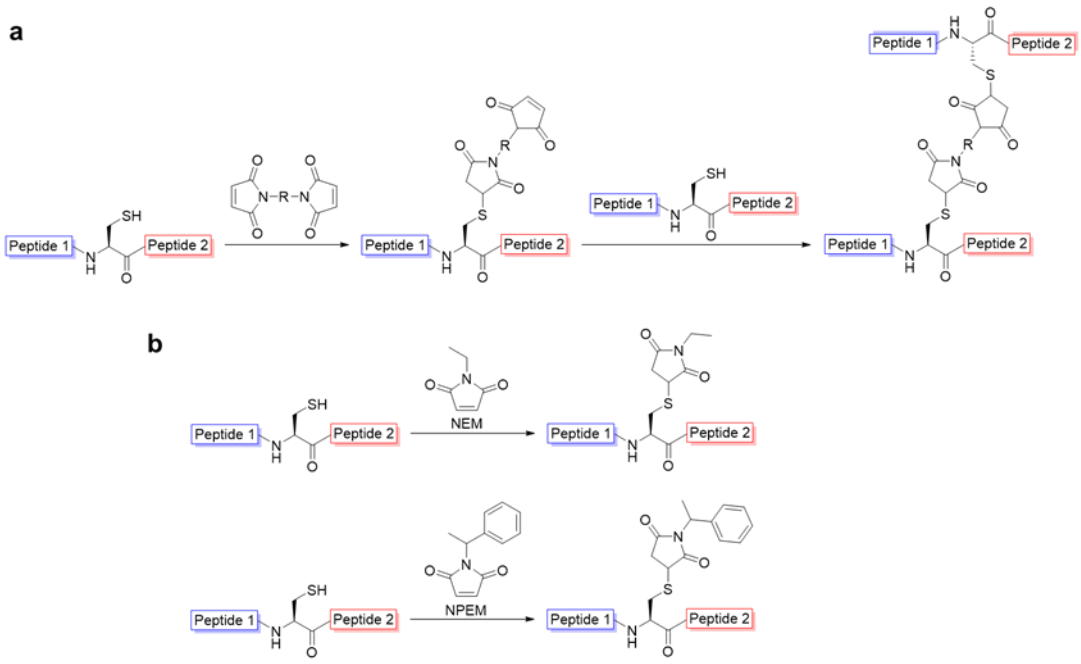

Figure 8 Protein modification on cysteine with maleimide reagents. (a) The scheme showing *N*,*N*'-bis-maleimides as a bifunctional linker; (b) The schemes showing cysteine modified with NEM/NPEM.

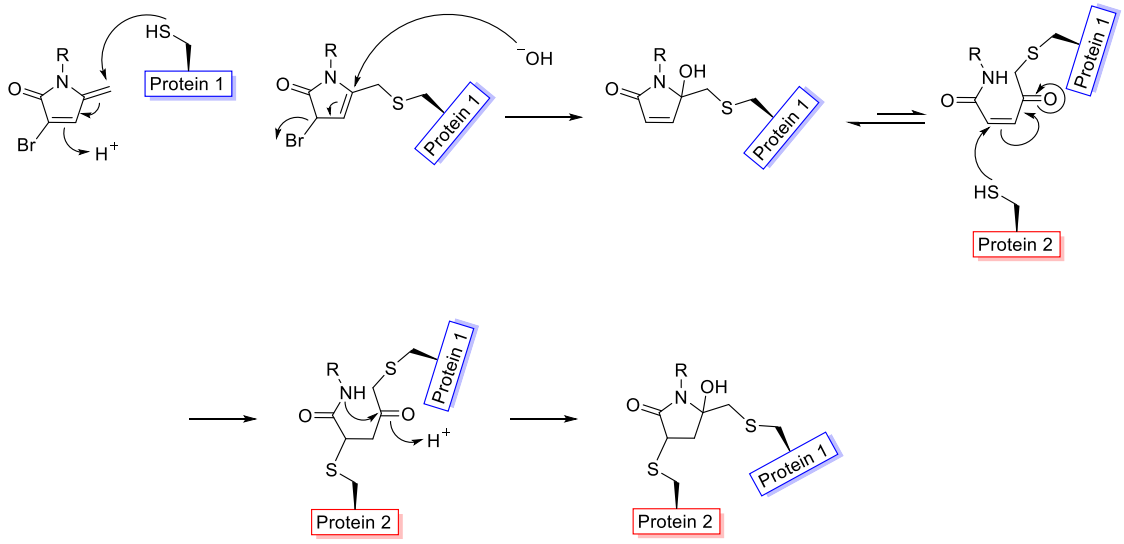

Figure 9 Proposed mechanism of protein multi-functionalization via Michael addition reaction using 3Br-5MPs as linkers.

while making ubiquitin-like protein based probes that specifically target their conjugating and deconjugating enzymes (Figure 10).[54] This reaction is efficient enough which does not require large equivalent of reagent. This benefits the cysteine specificity of the reaction rather than too much side-reactions on histidine or lysine.[55]

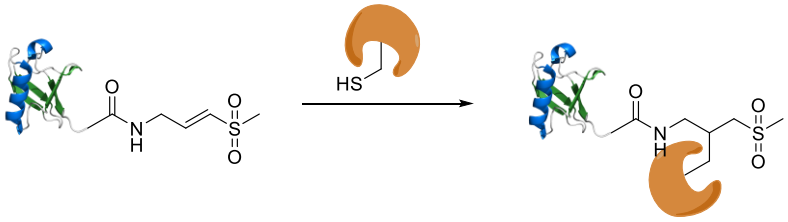

Figure 10 Covalent conjugation between a vinyl sulfone based NEDD8 probe and its conjugating enzyme.

1.2.3. Oxidation reactions

Cysteine oxidation has been found widely existing in the form of PTMs which play essential roles in cellular redox regulations.[19,34] This has encouraged scientists develop mild oxidation reactions using chemical methods to functionalize cysteine. However, not like inside cells where enzymes assist numerous complicated oxidation processes, facile oxidation *in vitro* is always much more difficult. Among all the possibilities, cysteine disulfide formation, in the other word, cysteine being oxidized by atmospheric oxygen during exposure to air turns to be the easiest and most significant choice for chemical reaction. Many protein refolding processes, and cysteine-based protein adducts have been reported using this automatic method.[56]

Besides air oxidation, the generation of disulfide bond on cysteine residues can also be achieved using disulfide exchange reaction. The first well-known disulfide exchange compound that was used to detect protein sulfhydryl content was Ellman's reagent (5,5-dithiobis(2-nitrobenzoate)) (Figure 11a).[57] At pH 8, the color of released 5-thio-2-nitrobenzonate (NTB) can be captured using UV-vis spectrometry analysis. Facile radical oxidation using iodine as the activator was also reported as one the initiation of cysteine disulfide usage.[58] As the bioorganic chemistry kept developing, mild chemical oxidation reagents such as sulfenic acid and its derivatives such as sulfenyl halides and were reported to form *S*-sulfenylation at protein cysteine residues (Figure 11b).[59-62]

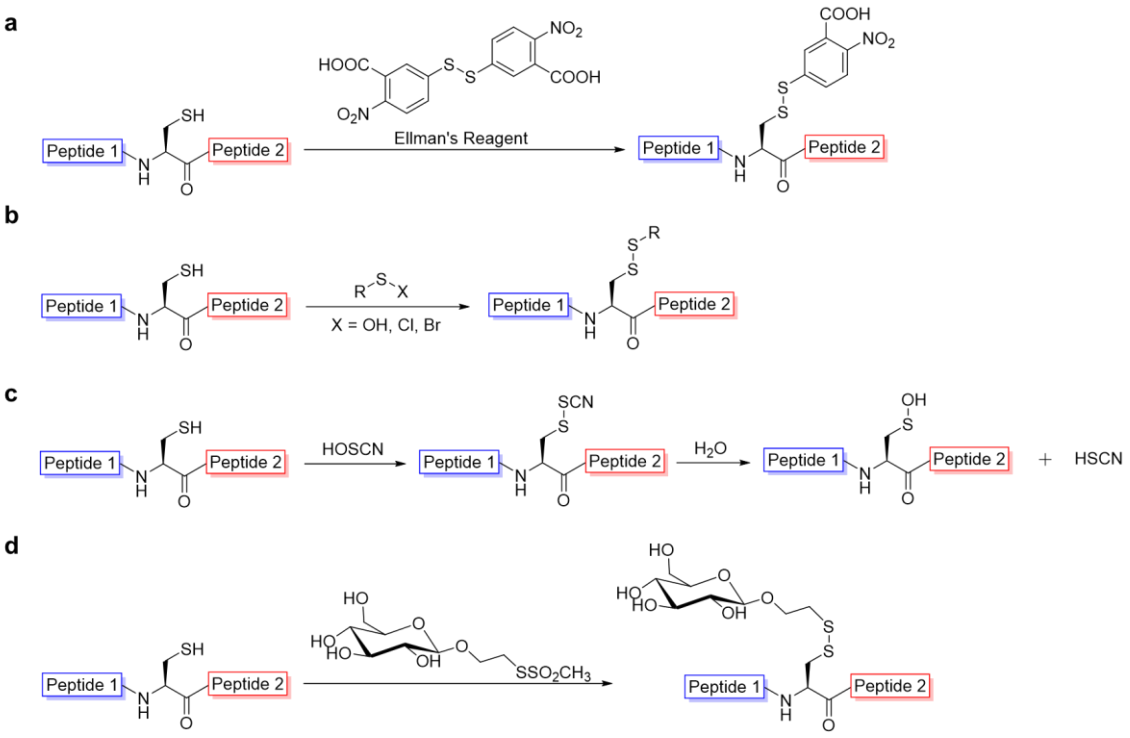

Figure 11 The schemes of cysteine disulfide bond formation reaction using (a) Ellman reagent, (b) sulfenic acid and its derivatives, (c) hypothiocyanous acid, and (d) glycosylated methanethiosulfonate.

Later, a similar concept that using hypothiocyanous acid (HOSCN) instead of acyl or alkylated sulfenic acid was reported by Hawkins's research group. They have cleared observed both reversible and irreversible inactivation of enzymes which proved the toxicity of this reagent via the generation of sulfenyl thiocyanate at protein cysteine position (Figure 11c).[63] In addition, benefitting from the great leaving tendency of methyl sulfonate ion, methanethiosulfonates have also been discovered as a powerful reagent to oxidize the cysteine thiol group and successfully installed protein glycosylation via cysteine disulfide bond (Figure 11d).[64]

Disulfide linkages are a series of redox sensitive covalent bonds. At the presence of reducing agents such as tris(2-carboxyethyl)phosphine (TCEP), such linkages are easily to be break apart. This special chemical property does provide cysteine disulfide bond advantages to serve as an ideal drug carrier for spontaneously release if the targeting cell environment is highly reducing.[65] In the other hand, such labile linkages also restricts its biological usages when reducing agent is inevitable. Therefore, other stable covalent linkages are still highly desired.

1.2.4. Cysteine to dehydroalanine conversion

Dehydroalanine (Dha), as an α,β-unsaturated carbonyl structure containing amino acid, reveals strong reactivity for nucleophilic addition reaction. This has provided it great potential to become an ideal precursor for further site-specific protein modifications.[66-68] Moreover, Dha which was installed in certain protein sequences can provide an ideal covalent binding site for their downstream targets. Various mono-/poly-

ubiquitin (Ub) or ubiquitin like protein (Ubl) based Dha probes have been successfully used to capture and characterize their enzymatic intermediates.[69-72] Those Ub/Ubl-Dha conjugates have also been used for the synthesis of ubiquitinated or Ubl tagged proteins to discover the enzymatic selectivity and linkage specificity of deubiquitinases (DUBs) and ubiquitin-like proteases (ULPs).[73-75] Therefore, the demand for easily and efficiently synthesizing Dha-containing proteins keeps building up.

In order to obtain enough Dha containing proteins, cysteine has become the most popular precursor among all amino acids. There have been several methods reported in literature that can site-specifically convert cysteine to Dha. The possibility of thiol based elimination using cysteine disulfide was strongly benefitted from the highly efficient cysteine disulfide formation using the water soluble Ellman's reagent (*vide supra*) that generates a stable electron deficient leaving group. Under basic condition using 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), cysteine can be directly convert to Dha in a simple and scalable manner (Figure 12a).[76] However, to prevent the formation of sulfinic

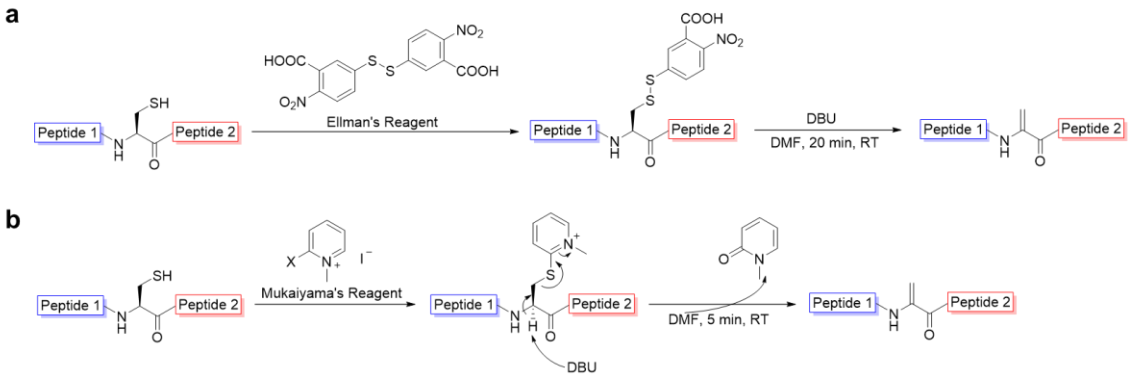

Figure 12 The DBU directed basic elimination of (a) cysteine disulfide formed by Ellman's reagent and (b) alkylated cysteine using Mukaiyama's reagent.

acid from cysteine disulfide at the occurrence of hydroxide, the reaction was only capable for pure organic solvent which is not favorable to most protein substrates.[77] To overcome that issue, instead of Ellman's reagent, Davis's group used Mukaiyama's reagent[78] to activate cysteine thiol via an aromatic nucleophilic substitution pathway (Figure 12b).[76] Although this methods successful evaded the sulfinic acid formation problem, the requirement of intolerably high pH value still strongly restricted its scope of usage.

Besides directly basic elimination, the conversion from cysteine disulfide to Dha can also be driven by reductive elimination using electron rich phosphine reductants. Davis and co-workers reported a synthetic method of glycoproteins that involved a disulfide to thiol ester linkage conversion. In this method, an electron rich phosphine, tris(dimethylamino)phosphine (HMPT), was used to reduce the disulfide bond formed by Ellman's reagent followed by the elimination to obtain Dha intermediate (Figure 13a).[79] Similar idea was taken by Xian and co-workers one year later where *S*-nitrosocysteine was used instead of cysteine disulfide (Figure 13b).[80] Both methods are applicable under neutral pH value and room temperature which led out this approach

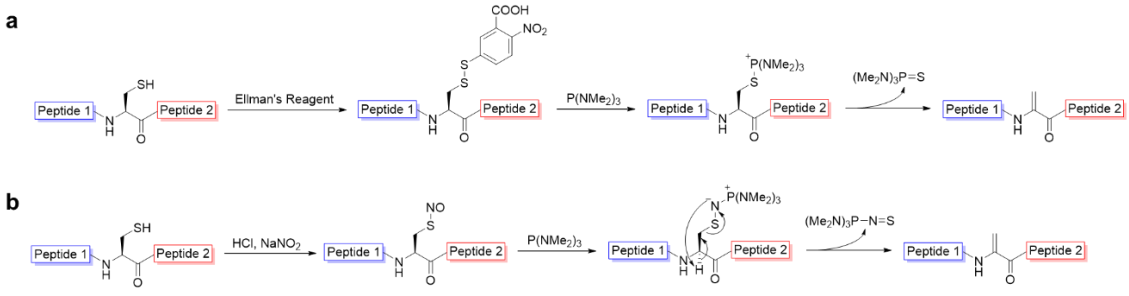

Figure 13 The reductive elimination using HMPT to generate Dha from (a) cysteine disulfide formed by Ellman's reagent and (b) *S*-nitrosocysteine.

good capability for protein based reactions. However, due to the usage of strong reducing agent, the disulfide reduction to cysteine free thiol is always inevitable.[76]

Another category of reactions that convert cysteine to Dha is oxidative elimination. The first oxidative elimination reaction was reported by Holmes and Lawton in 1977 in which 2-methyl-N^1-benzenesulfonyl-N^4-bromoacetylquinonediimide was used as the reagent (Figure 14a).[81] To generate milder and more biorthogonal methods for cysteine to Dha conversion, Davis's group develop an improved oxidative elimination method using *O*-mesitylenesulfonylhydroxylamine (MSH) to effectively remove cysteine thiol under robust pH and temperature (Figure 14b).[82] Unfortunately, this approach has only been determined efficiently convert cysteine to Dha on a single serine to cysteine mutant of subtilisin SBL-S156C where the cysteine residue located on a exposed loop region.[82] For further trials on other proteins, especially the ones containing super buried cysteines, randomized amination was observed on histidine, methionine, aspartic acid and glutamic acid which demonstrated the its poor selectivity.[83]

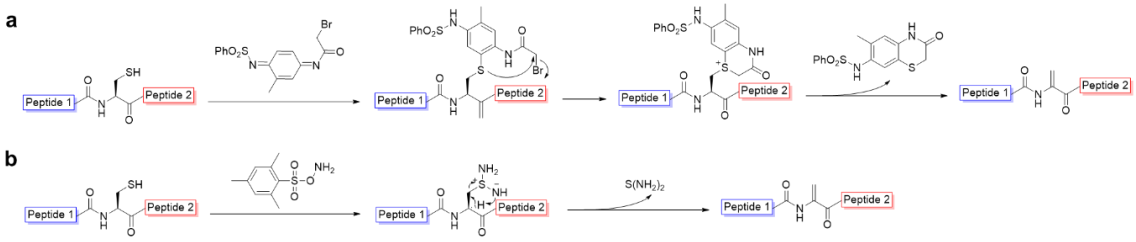

Figure 14 The oxidative elimination reaction to site-specifically convert cysteine to dehydroalanine using (a) 2-methyl-N^1-benzenesulfonyl-N^4-bromoacetylquinonediimide, (b) *O*-mesitylenesulfonylhydroxylamine.

More recently, a group of bis-alkylation-elimination reactions were assessed by Davis's lab to surmount all limitations we described above. The first compound they tested was a di-iodide species. After 1 hour incubation with SBL-S156C at pH 8, 37 °C, the mass spectrometry analysis confirmed the formation of Dha at cysteine position (Figure 15a).[76] Furthermore, a di-bromide compound, 2,5-dibromohexanediamide (DBHDA), was evaluated using the same protein substrate. After 30 min incubation at room temperature, single alkylation was detected by mass spectrometry. Cyclized bis-alkylation intermediate as well as final elimination product were both able to be observed after 1 hour prolongation of incubation (Figure 15b).[76] The feasibility of this di-bromide reagent was further confirmed by expanding the usage onto single cysteine mutant of Np276 and camel single-domain antibody cAb-Lys3.[76] In 2017, Davis's group reported an improve format of di-bromide bis-alkylation reagent, methyl 2,5-

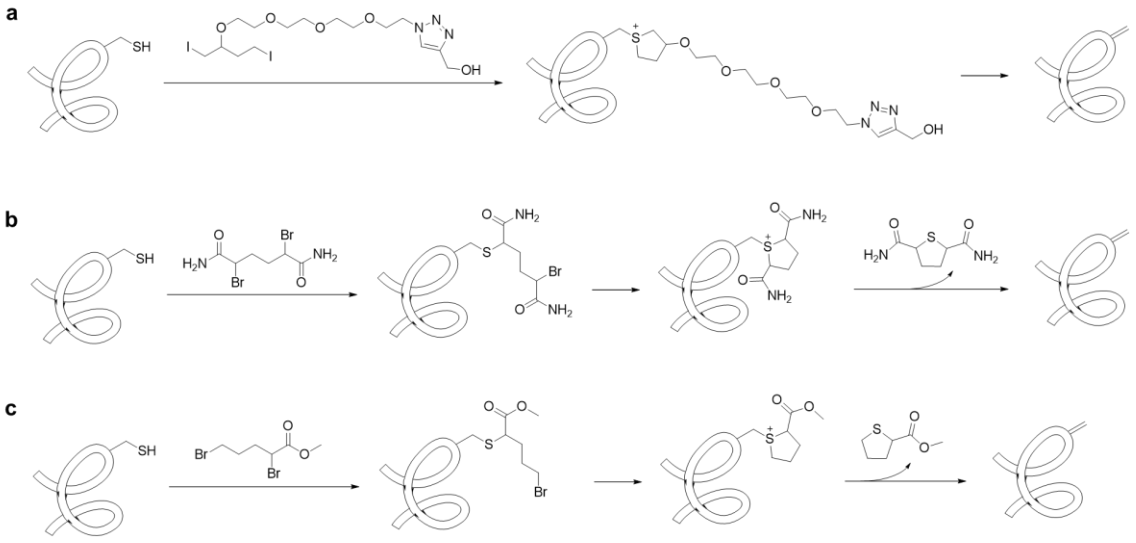

Figure 15 Bis-alkylation-elimination reaction to site-specifically convert cysteine to dehydroalanine using (a) di-iodine reagent, (b) DBHDA, and (c) MDBP.

dibromopentanoate (MDBP), which was used to pursue bis-alkylation-elimination on a super sensitive enzyme, *M. tuberculosis* pantothenate synthetase, where the conditions required by DBHDA were too harsh to introduce Dha directly (Figure 15c).[84] The synthesis of an antibody-drug conjugate that used the same compound was also reported by Bernardes *et al*. in the same year.[85] Based on those thorough assessments, bis-alkylation-elimination was considered as the most facile and versatile method to convert cysteine to Dha.

1.2.5. Vinyl sulfide linkage: cysteine-alkyne reaction

Inspired by cysteine involved Michael addition reaction, scientist started to investigate the possibility of the reaction between cysteine and alkyne. Unlike α,β-unsaturated carbonyl species, alkynes are much weaker electrophiles which results in its relatively stable chemical property. To initiate the alkyne involved reaction on protein cysteine, radical initiators are usually essential for the assistance (*vide infra*). While getting rid of the participation of radicals, there must be some other factors to drive alkyne's reactivity.

The first group of reaction between cysteine and alkynes were directly derived from Michael addition by Che's lab. Herein, electron deficient acyl alkynes were efficiently added to peptide cysteine side chain thiol via a Michael addition pathway (Figure 16a). The reaction was then expanded to site-specifically label a fluorescent dye into BSA which further proved its capability to work on proteins (Figure 16b).[86] It is notable that when R_2 referred to a hydrogen atom, due to the presence of a labile

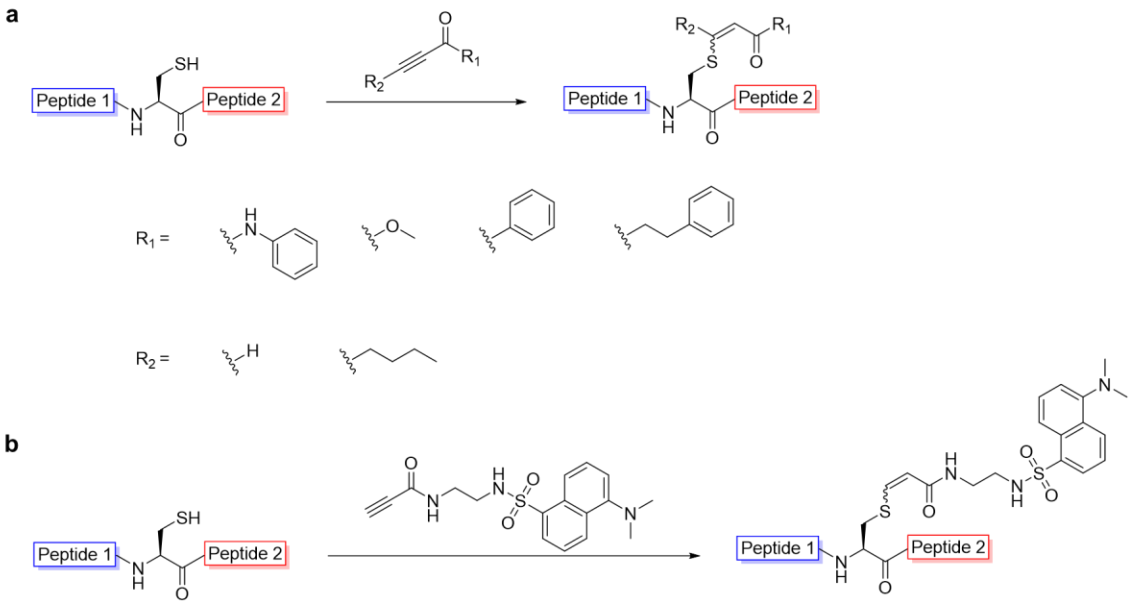

Figure 16 (a) Chemoselective cysteine modification using electron deficient alkynes. (b) Site-specific incorporation of a fluorescent dye via cysteine alkenylation reaction.

unsaturated vinyl acyl linkage as well as no steric hindrance at carbon-carbon double bond, by providing excess of free thiol, the reaction would continue to form a cleavable vinyl sulfide linkage and revive the original free cysteine residue.[86] This benefit was successfully taken by Che's lab to obtain fluorescence resonance energy transfer (FRET) based probe[87] and luminescent cyclometalated iridium(iii) complexes conjugated peptides and proteins.[88]

Except chemical driving force, enzymatic catalysis can also facilitate the reaction between cysteine and alkynes. In 2013, Ovaa's group observed a enzymatic recognition of deubiquitinase (DUB) driven Ub-propargylamine (Ub-Pa) conjugation reaction with the active site cysteine thiol of DUBs (Figure 17).[89] This reaction only occurred when

certain DUB could recognize full-length ubiquitin sequence and would not resulted in any non-specific modification. This reaction was also determined as a universal approach to obtain a covalent linkage between an enzyme and its protein substrate.

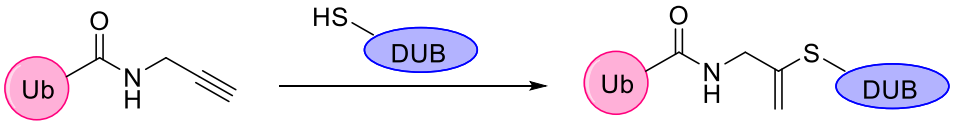

Figure 17 Enzyme promoted covalent bond formation between the active site cysteine of a deubiquitinase and ubiquitin terminal alkyne.

1.2.6. Radical thiol-ene & thiol-yne coupling

The nucleophilic addition reactions occur between thiol and unsaturated carbon-carbon bonds have been described in-depth above. In addition to traditional ionic addition, thiyl radicals that are produced at cysteine residue can also trigger the effective addition to alkenes or alkynes. Such reactions are typically initiated via UV irritation or radical initiator and has been widely applied in various research aspects (Figure 18). Applying thiol-ene coupling on protein substrates first initiated in 2009 by Dondoni *et al*. to synthesize glycosylated bovine serum albumin (BSA) using a glycosylated alkene and BSA bearing a single cysteine mutant (Figure 19a).[90] In the same year, Davis *et al*. used an opposite design where thiyl glycan and olefinic protein were used to undergo the coupling reaction (Figure 19b).[91] A few years later, similar idea was expanded to site specifically install fluorescent dye into an *E. coli* acid-chaperone HdeA by Chen's research lab (Figure 19c).[92] Furthermore, the thiol-ene coupling reaction was also used by Liu *et al*. to incorporate lysine acetylation (Figure 19d).[93]

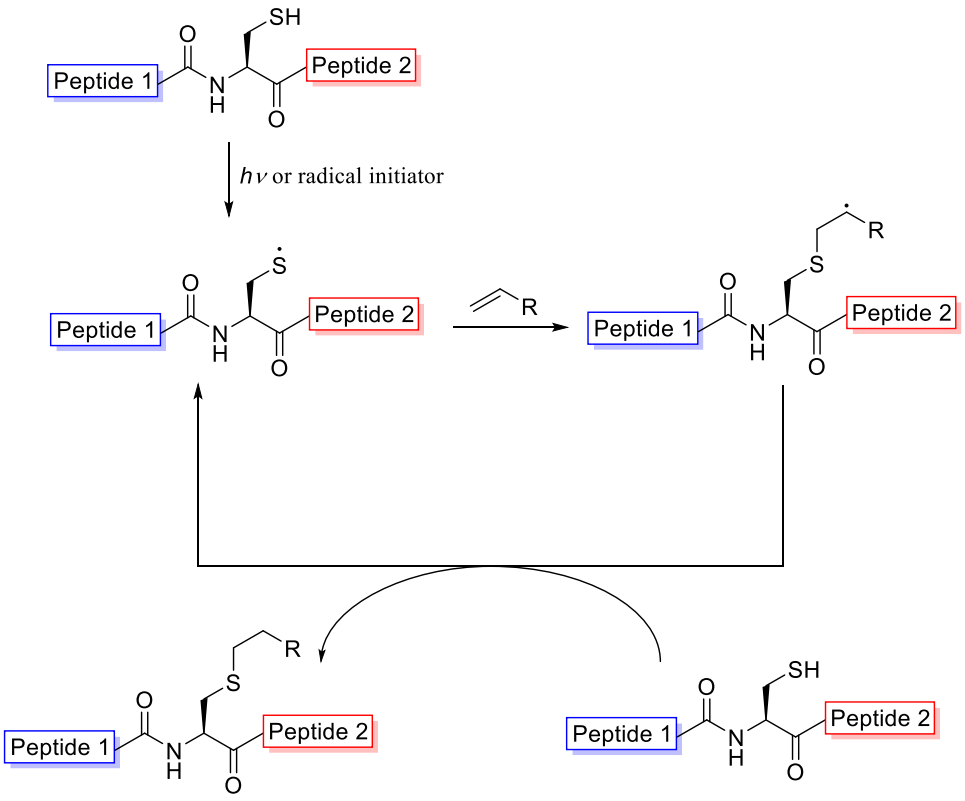

Figure 18 General concept of thiol-ene coupling reaction.

In 2012, Strieter and co-workers developed a thiol-ene coupling involved approach to chemically obtain thioether bond between the *C*-terminal of one ubiquitin and the side chain of the other. The ubiquitin carrying a *C*-terminal allylamine (Ub-Aa) was rapidly obtained via a UCHs promoted hydrolysis and amidation reaction.[24] Lithium acyl phosphinate (LAP), as an ideal radical initiator, was subsequentially used to start the reaction between a Ub K to C mutant and Ub-Aa in order to acquire various Ub dimer linkages (Figure 20a).[94] In addition to the synthesis of diUb, branched triUb can

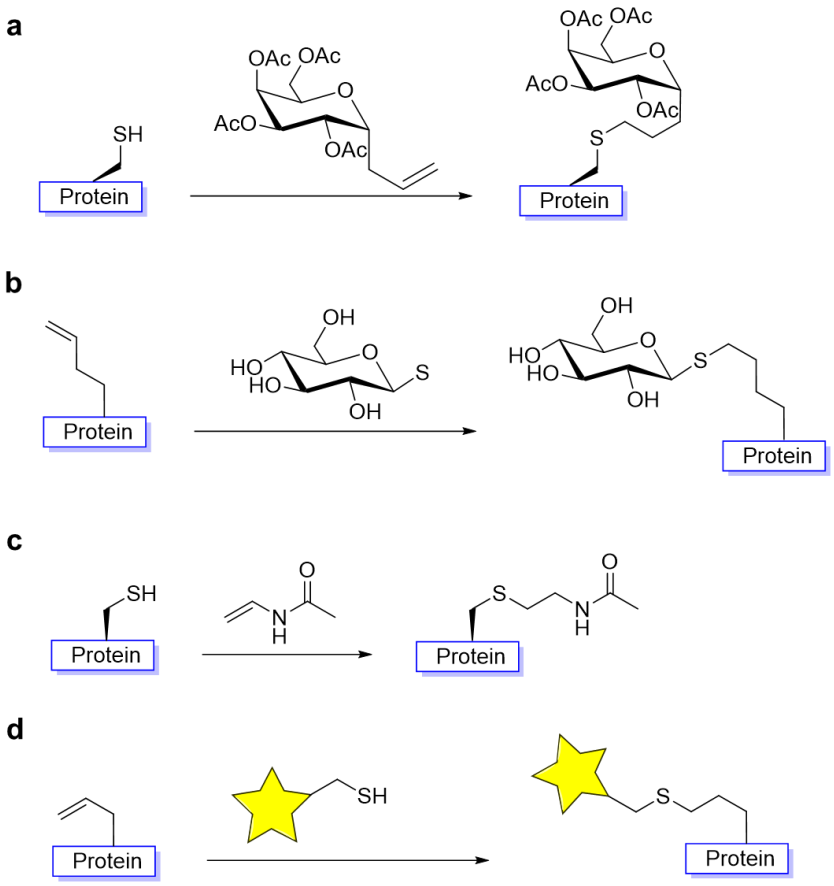

Figure 19 Application of thiol-ene coupling in (a) protein glycosylation using alkene containing glycan; (b) protein glycosylation using alkene containing protein; (c) synthesis of lysine acetylation mimic; and (d) fluorophore labeled protein synthesis.

be generated using Ub bearing two K to C mutation sites using the same method mentioned above (Figure 20b). Additionally, by using bifunctionalized Ub that bearing both *C*-terminal Aa and K6C mutation, homogeneous Lys6-linked ubiquitin oligomers which containing two to seven Ubs can be attained (Figure 20c).[95]

More recently, Anseth and co-workers reported a reversible bioconjugation reaction using thiol-ene reaction to synthesize protein hydrogel using poly(ethylene

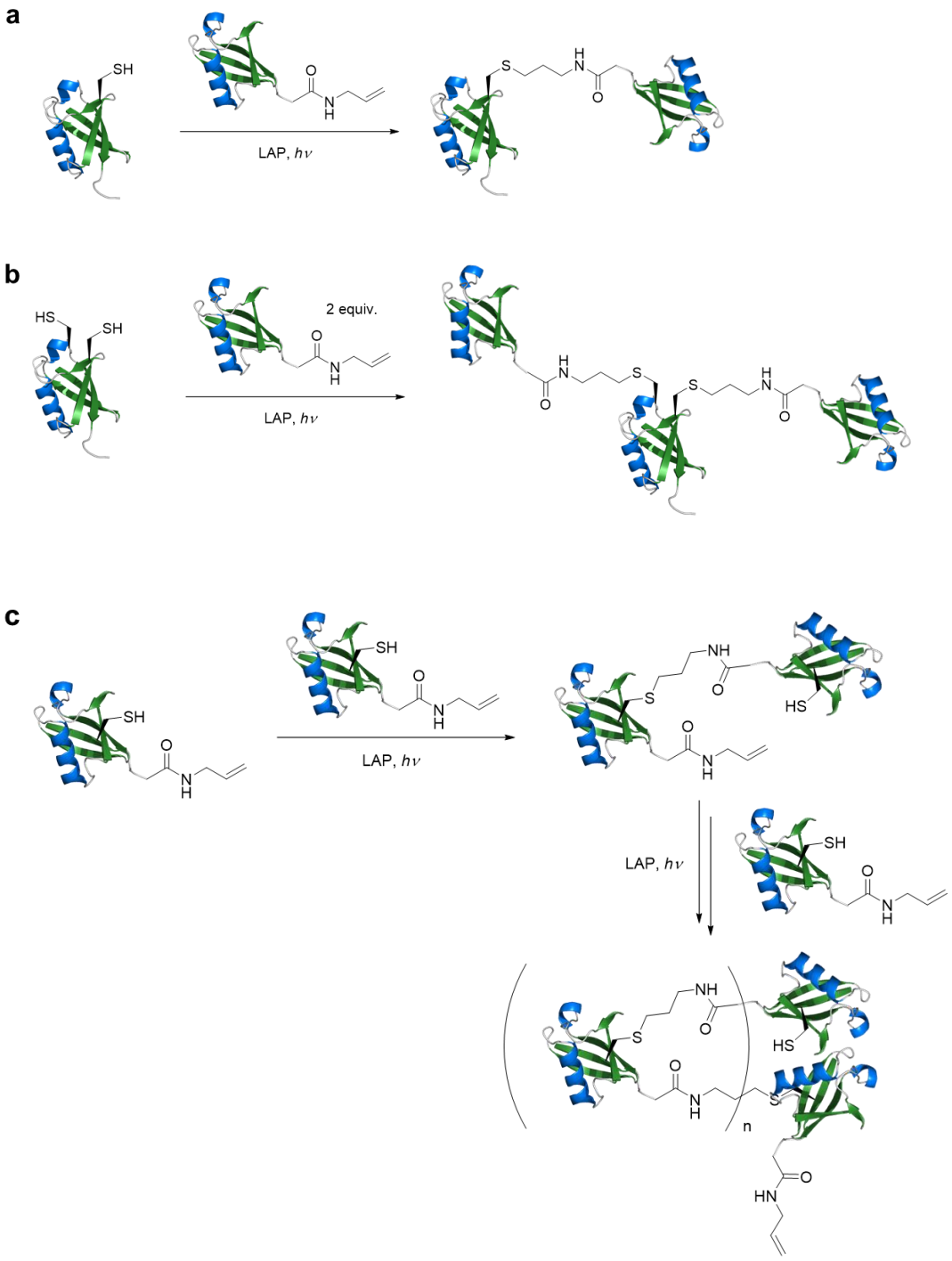

Figure 20 The synthesis of (a) di-Ub (b) tri-Ub, and (c) poly-Ub chain using thiol-ene coupling method.

glycol) (PEG) (Figure 21). They have proved that the whole signaling proteins can be patterned and released in the presence of cells, and that cells respond to their presentation with spatial fidelity.[96]

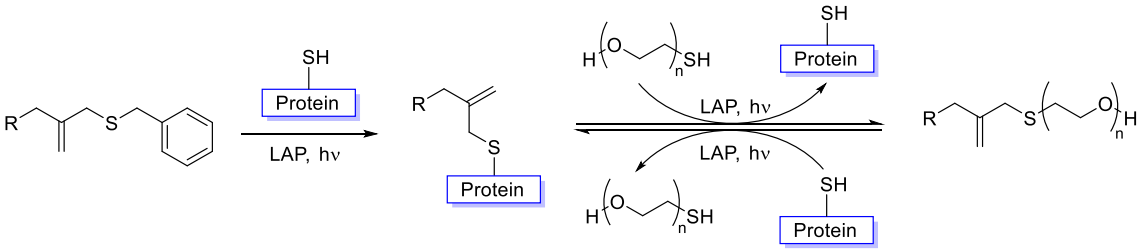

Figure 21 Proposed strategy to reversibly tether signaling proteins to hydrogels through the allyl sulfide moiety.

Although thiol-yne reactions were much less frequently used than thiol-ene coupling for protein chemistry, we still anticipate it as an intelligent direction for protein multi-functionalization. Davis and co-workers reported the usage of thiol-yne coupling to site-specifically conjugate an alkyne containing glycan molecule as well as a thiol containing fluorescent dye in a sequential manner onto a BSA mutant.[97] This result strongly demonstrated our assumption and provided fundamental basis of the synthetic techniques for protein bi-modification.

1.2.7. Thiazolidine linkage

Protein thiazolidine can be formed by conjugating *N*-cysteine peptide with a *C*-terminated aldehyde. This is an alternative method to build up covalent linkage between two peptide fragments. This method was successfully used to form antibody-drug

conjugates that can be specifically delivered to the target cells and release the cytotoxic drug inside the cell due to the reversibility of such reaction.[98-99]

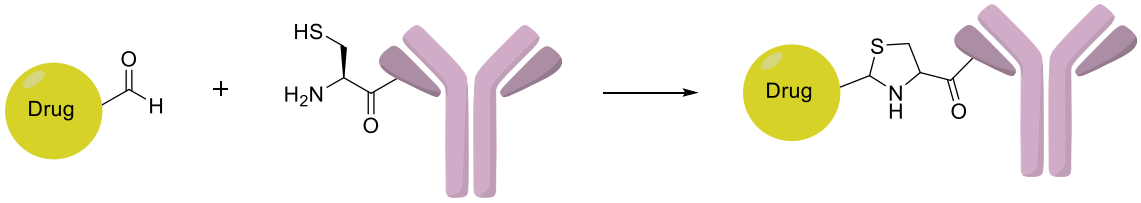

Figure 22 Thiazolidine heterocycle formation reaction.

1.3. Current methods to functionalize protein *C*-termini

Protein *C*-terminus, as a distal end, is an ideal site for protein labelling as well as making active intermediates for protein synthesis. Unfortunately, although some enzyme mediated protein *C*-terminal activation reactions have been reported, due to the thermodynamic stability of carboxylic acid, it is extremely hard to directly functionalize protein *C*-termini using chemical methods without the help of enzymes. Therefore, cleavage of amide backbones in proteins at a specific position followed by a new amide bond formation becomes an alternative way to solve that problem.

We have already described a lot of protein chemical reactions happened on protein cysteine. Most of them were driven by the strong nucleophilicity of cysteine side chain thiol group. Besides traditional nucleophilic alkylation reactions, the sulfhydryl group also involves in nucleophilic acyl substitution reactions to form a thioester linkage. If the substrate protein or peptide contain an *N*-terminal cysteine, with the assistance of adjacent β-amino group, a brand-new amide bond can be easily obtained

via *S*-to-*N* acyl transfer reaction. This provided protein chemists the theoretical basis of protein ligation reactions.

1.3.1. Protein total synthesis and native chemical ligation

In 1963, inspired by the ribosomal biosynthesis of proteins, Merrifield developed solid phase peptide synthesis (SPPS) that was one of the major milestones in protein total synthesis.[100] Solid phase, referring to the insoluble resin that the first amino acid is covalently attached to, while peptide synthesis proceeds by a repetitive *N*-terminus de-protection and coupling with a following selected amino acid, until the desired peptide is totally synthesized and then ultimately cleaved from the solid phase resin. Instead of the *N*- to *C*- sequential biosynthesis by ribosome, Merrifield's technique proceeds in a *C*- to *N*- direction.

Solid phase peptide synthesis has allowed the protein total synthesis to generate up to 166 amino acid residues.[101-102] The automatic stepwise elongation of SPPS has greatly reduced the time that required for traditional organic peptide synthesis. It also strongly facilitates the synthesis of small proteins that contain uAAs or involve site-specific modifications.[103]

The ligation between cysteine and a thioester was first observed and reported in 1953 by Wieland.[104] In 1994, the Kent group firstly reported this whole process and termed it as Native Chemical Ligation (NCL).[105] Herein, two peptide fragments, one containing a *C*-terminal thioester while the other containing an *N*-terminal cysteine residue, are participating. During the reaction, thioester exchange reaction firstly

happens between the two peptide segments mentioned above to generate a cysteine-based thioester. The reaction then undergoes the intramolecular *S*-to-*N* acyl transfer reaction to obtain an amide bond linkage (Figure 23). Notably, NCL is a robust method with wild accessibility and high biocompatibility, which can undergo in pure aqueous solution for fully unprotected peptide segments without the concern of amino acid

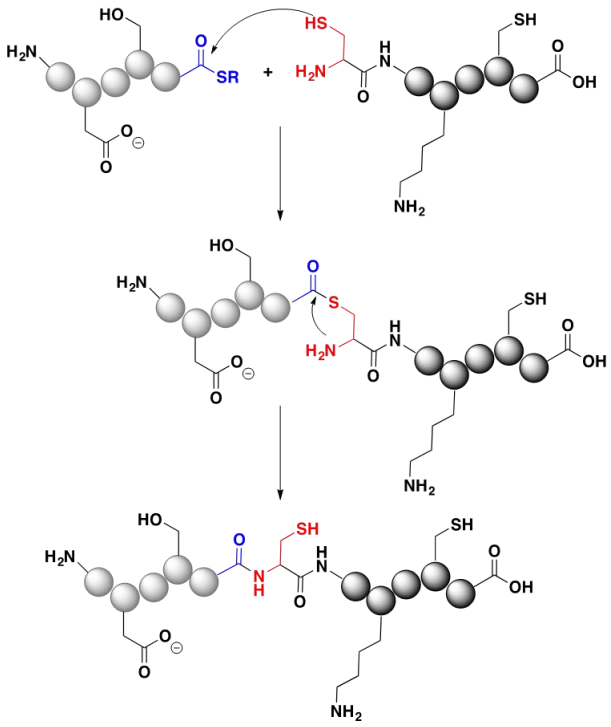

Figure 23 The mechanism of native chemical ligation including the cysteine thioester formation (step 1) and the *S*-to-*N* acyl transfer (step 2).

racemization.[106] With the development of NCL, many important protein targets have been synthesized in their native or non-native structures, which was previously unattainable by other ligation approaches. The idea of NCL has also been applied to the lysine side chain for the generation of ubiquitinated proteins[107-108].

Benefitting from the development of SPPS and NCL, numerous *C*-terminal functionalized proteins have been successful synthesized. The synthetic process typically involves two steps: the synthesis of *N*-peptide fragments containing either thioester C-terminus and *C*-peptide fragments including *N*-terminal cysteine and a desired *C*-terminal modification using SPPS, and the ligation reaction between those two peptide segments (Figure 24). This systematic approach has been widely applied to synthesize protein activity-based probes and fluorogenic probes involving *C*-terminal functionalization.[109]

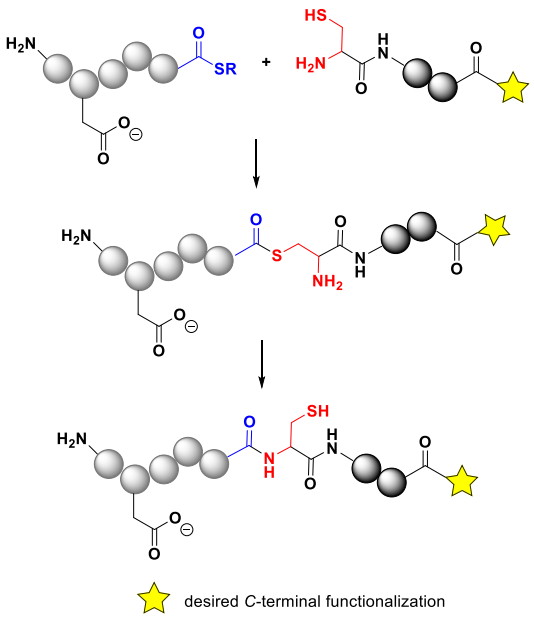

Figure 24 Synthetic cascade of C-terminal functionalized protein using SPPS and NCL.

1.3.2. Protein semi-synthesis and expressed protein ligation

Intein was first discovered in 1988 by Bowman.[110] Since it does not perform any known function in biological system but instead, it undergoes an intramolecular

chemical process called protein splitting to separate itself into fragments and combine two protein fragments together, it is also called protein intron.[111] Protein splitting process is comprised of the removal of an intein fragment followed by the combination of *N*- and *C*-extein to form a novel protein (Figure 25). The reactive thiol side chain at the *N*-terminus of intein firstly attacks the carbonyl group on the amide bond at the *C*-terminus of *N*-extein in order to form a reactive thioester intermediate. Then, the thioester undergoes the second nucleophilic attack by the thiol group on the Cysteine residue at *N*- terminus of *C*-extein to form the new thioester bond. Subsequently, the asparagine residue right at the *N*-terminal of the thioester cleaves the amide backbone and releases the free β-amino thiol at the cysteine residue as well as removes the intein fragment. Eventually, an *S*-to-*N* acyl transfer reaction like NCL gives out a mature protein for cellular functions.

Native chemical ligation relies on organic peptide synthesis, usually solid-phase peptide synthesis (SPPS) using resins to obtain the *C*-terminal thioester fragment. The length limitation of SPPS always significantly limits the protein size that can be synthesis via NCL. Such multi-step SPPS also results in huge amount of waste as well as relatively low yield. To remedy those shortcomings, in 1998, a new concept, expressed protein ligation (EPL) was came discovered inspired by intein-based protein splitting reaction.[112] Instead of organic synthesized thioesters, EPL uses intein as a mediate to generate protein thioester.

As shown in Figure 26, unlike natural protein splitting that using asparagine cleavage to assist the removal of intein, EPL uses a strategy similar to NCL to break the

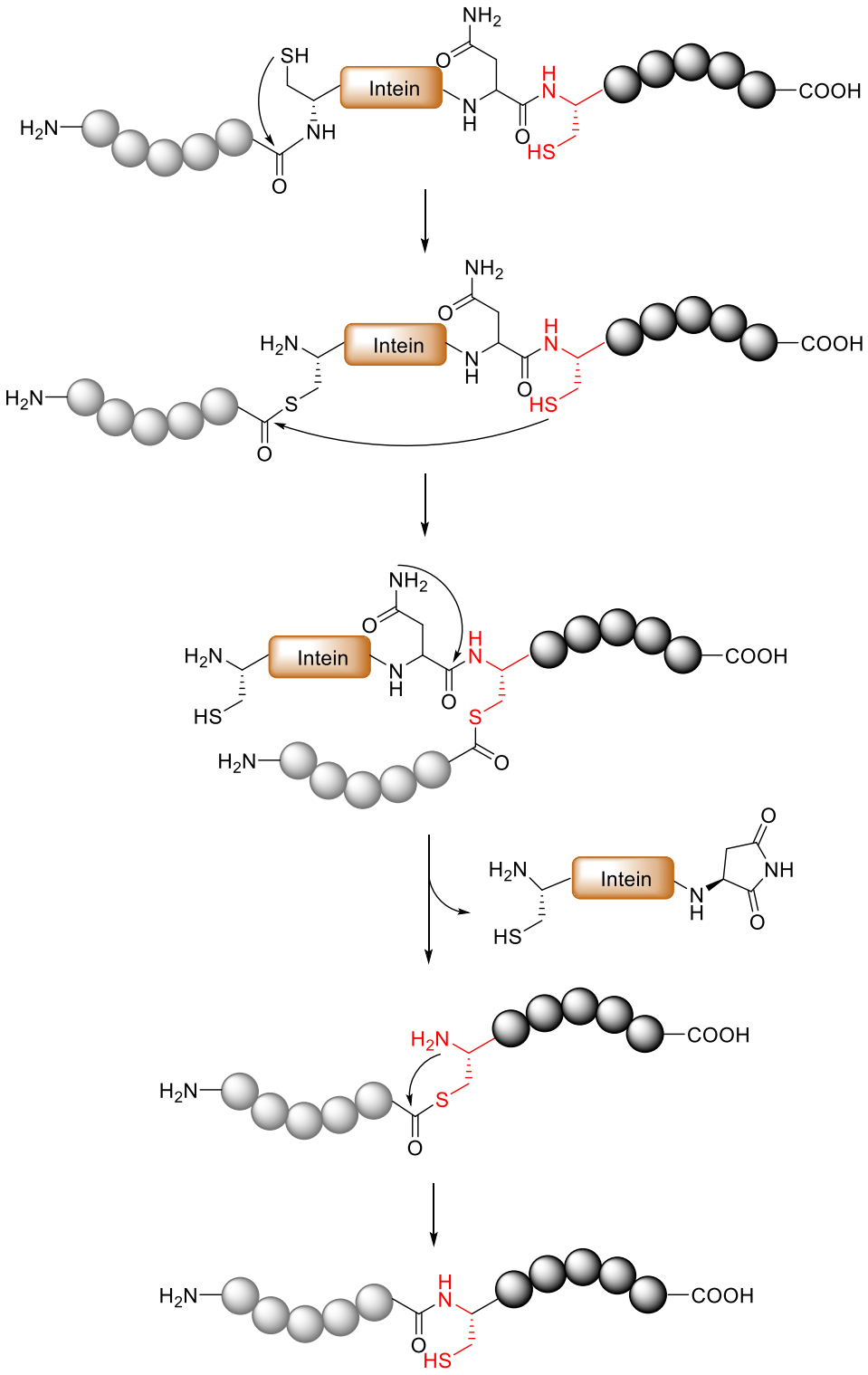

Figure 25 The reaction mechanism of intein-based protein splitting reaction.

whole protein sequence into two peptide fragments. In the first step, an expressed recombinant protein which contains both the first part of the target protein and an intein sequence binds to chitin affinity resin through a chitin binding domain (CBD). After that, such recombinant protein undergoes an intramolecular acyl shift reaction to generate the thioester. Finally, another part of the target protein which contain an *N*-terminal cysteine residue undergoes a NCL process to conjugate with the first part and

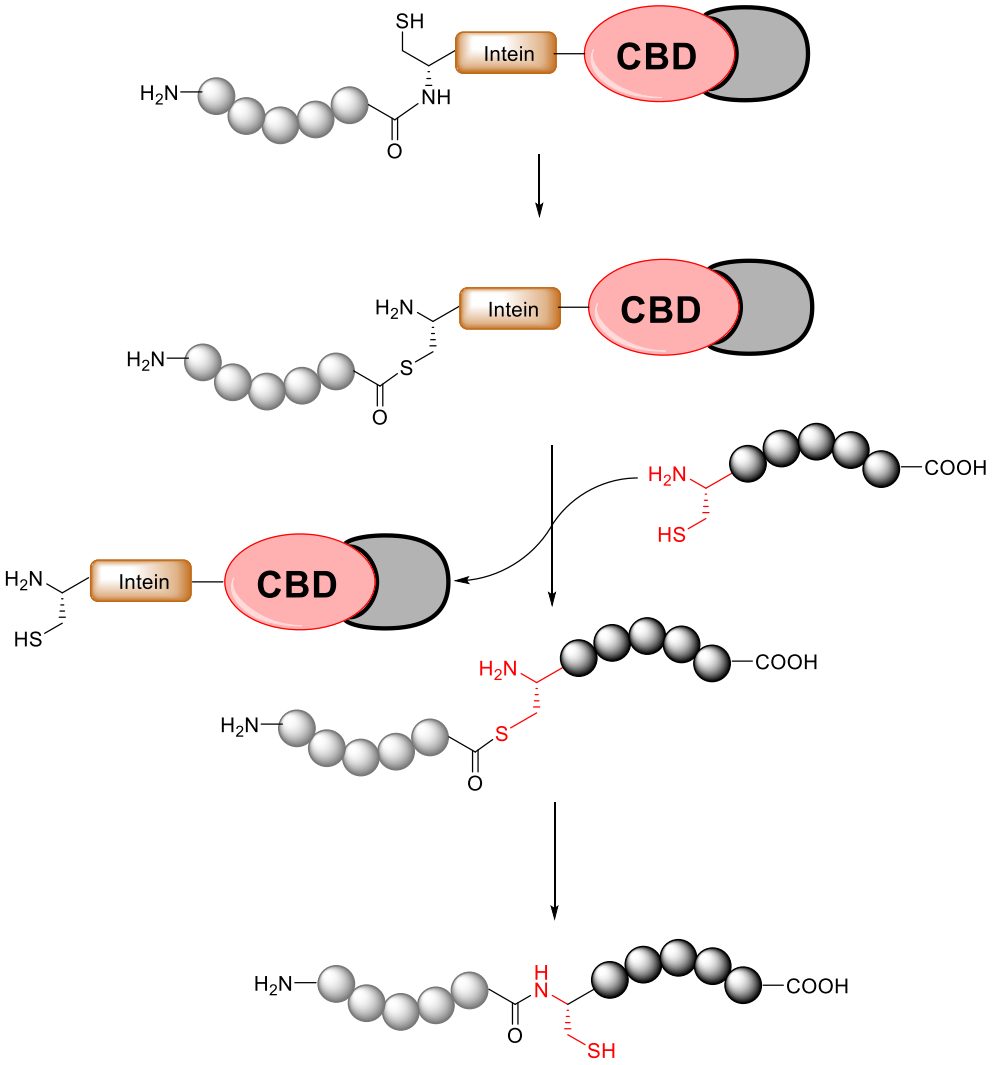

Figure 26 The reaction scheme of expressed protein ligation.

generate the entire target protein. At the same time, The CBD contain intein still binds with affinity resin which has been separated with the expected protein.[113] By using EPL approach, the majority of a protein sequence can be obtained via *E coli.* Protein expression system. Only the small *C*-terminal cysteine containing fragment requires SPPS which extremely simplified the synthetic process.

Standard EPL has already been widely used as a method to synthesize *C*-terminal functionalized recombinant proteins. However, it still has some drawbacks especially the low efficiency on generating the thioester.[114] Right at the upstream of ligation thioester site, the premature intein cleavage can occur and this leads to the decreasing of active thioester produced.[115] For these reasons, EPL needs to be updated.

In biological systems, there is another similar intein splitting pathway called protein trans-splitting (PTS). It does not use the whole intein to undergo chemical ligation, instead, it separates the intein into two fragments, one containing an *N*-extein at the *N*-terminus of the *N*-part intein while the other containing the *C*-part intein followed by the *C*-extein. Only if *N*-intein and *C*-intein non-covalently bind together, the *N*-to-*S* acyl shift to generate thioester can be initiated.[116-117] Like normal EPL, PTS also use a chitin affinity resin as a purification method. However, just the *C*-intein part binds with the resin. This highly decreases the chance of premature intein cleavage since none of the whole intein sequence is present in both expressed cell lysate and on the affinity resin. Once the *N*-intein part is attached to the resine that has already been treated by *C*-intein fragment, it will undergo a thioester formation reaction right away and then followed by the conjugation with another part of target protein with *N*-terminal cysteine

(Figure 27). In this method, the *C*-intein part which binds with chitin resin can be recycled in the whole synthesis process.

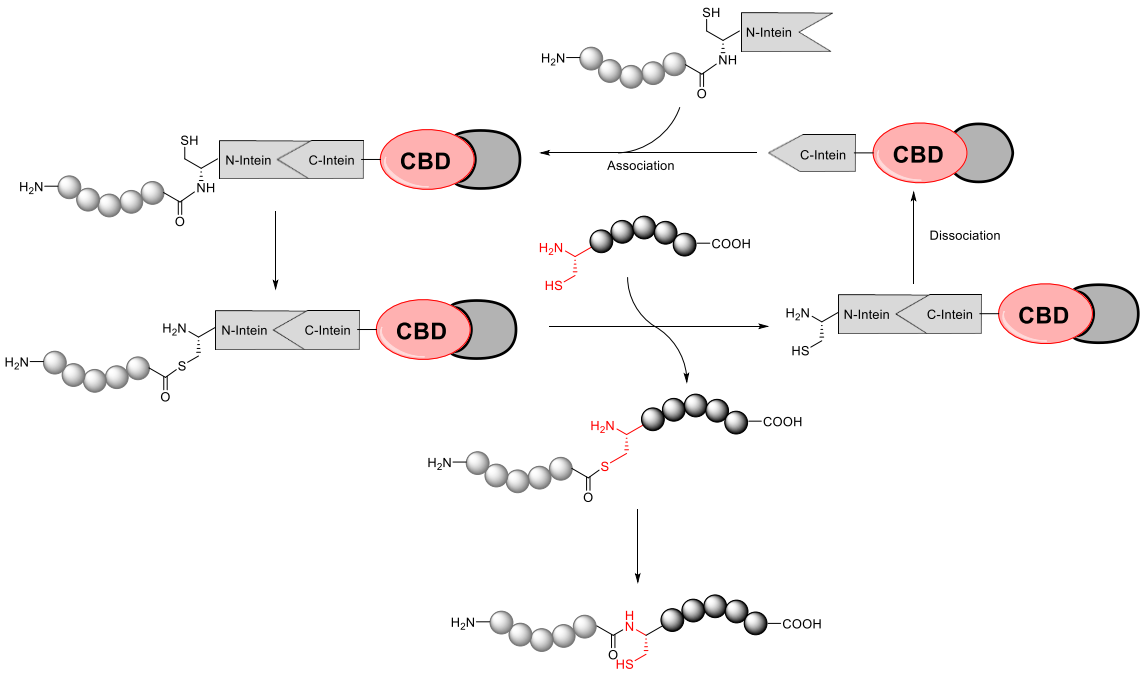

Figure 27 The steps of protein trans-splitting used for expressed protein ligation.

1.3.3. Improved variants of native chemical ligation

Although native chemical ligation has handled massive amount of protein synthesis, the Fmoc protected peptide synthesis of *N*-terminal fragment that contains thioester *C*-terminus still remain challenging.[118] In 2011, a NCL complementary method called peptide hydrazide ligation was reported by Liu etc. where a synthetic peptide hydrazide was used instead of protein thioester (Figure 28). The *C*-terminal hydrazide containing peptide can be obtained by either Boc-/Fmoc-based SPPS or protein semisynthesis.[119-120] Similarly, in 2018, Dawson's group reported another method that uses acetyl acetone to convert the chemically stable peptide hydrazide to a peptide acyl

pyrazole and then a peptide thioester for further native chemical or expressed protein ligation (Figure 16).[121] Both techniques which fully avert the protein side chain protections, have greatly expanded the synthetic scope of NCL and EPL and become an advance method for protein synthesis. Those two methods, benefitted a lot from the simpler synthesis of peptide hydrazide rather than peptide thioester, has become the most popular method to obtain *C*-terminal modified Ub and Ubl probes.[122]

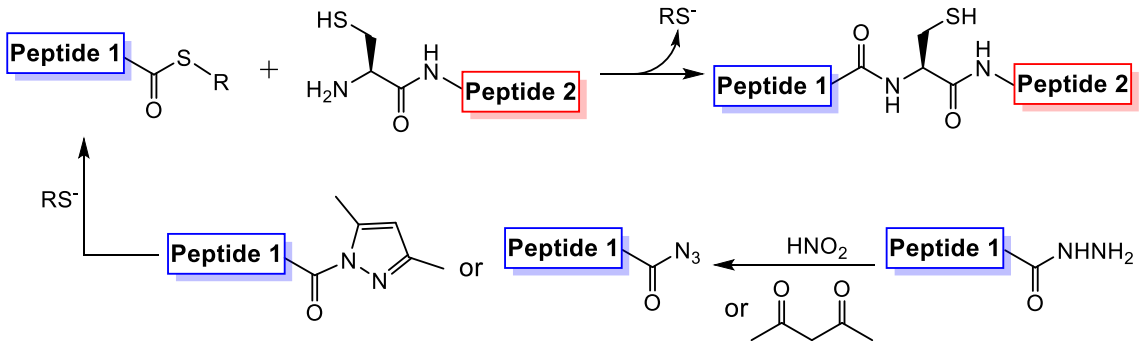

Figure 28 The reaction scheme of peptide hydrazide ligation.

Besides peptide hydrazide ligation, another method that used a phenolic ester equipped with an unsymmetrical disulfide was invented by Danishefsky and co-workers in 2004. Under mild aqueous reductive condition, the disulfide moiety which located at the *ortho*- position of phenolic ester was quickly reduced to a thiol group and undergo an orientation preferred intramolecular *O*-to-*S in situ* acyl transfer reaction to obtain a thioester. Although ester is less thermodynamically favorable, at the presence of incoming *N*-terminal cysteine containing peptide, the reaction equilibrium kept moving forward that eventually gave out ligation product (Figure 29). This approach has

manifested strong reactivity and versality to synthesis a lot of protein such as the bifunctional glycopeptides.[123]

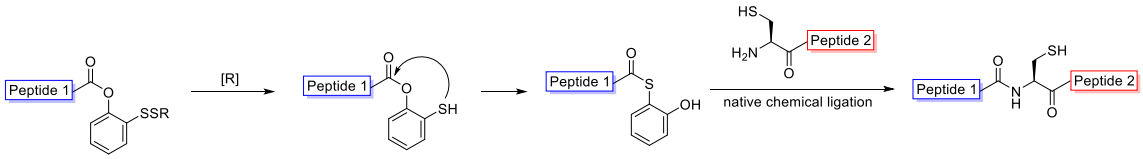

Figure 29 Proposed mechanism of phenolic ester mediated protein ligation reaction.

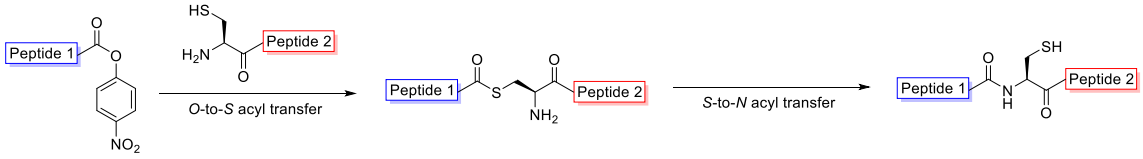

Figure 30 The reaction scheme of *para*-nitrophenyl ester mediated protein ligation reaction.

Although the thioester mediated NCL approaches, including the *ortho*-thiophenolic ester variant, are very powerful, they still suffer many limitations. Among them, the most significant one is the ligation rate dependency on the steric hindrance of *C*-terminal thioester amino acid. For proteins containing more branched side chains, such as valine, isoleucine and proline, ligation step sometimes is extremely slow or completely inhibited.[124] To circumvent this problem, a direct oxo-ester peptide ligation method that using an activated *C*-terminal *para*-nitrophenyl ester was developed by Danishefsky's group in 2008.[125] Comparing to *ortho*-thiophenolic ester method which still requires the intramolecular *O*-to-*S* acyl transfer to activate the ester, this approach that benefitts from the strong acyl transfer ability given by *para*-nitrophenyl group

allows *N*-terminal cysteine containing peptide quickly undergoing intermolecular *O*-to-*S* acyl transfer followed by *S*-to-*N* acyl transfer to generate ligation product (Figure 30).

1.3.4. Cysteine desulfurization reactions

All protein ligation methods that we described above requires a cysteine residue for the ligation process. However, the low natural existence of cysteine limits the number of sites to undergo ligation reactions. To overcome this major shortage, cysteine desulfurization to alanine was firstly invented by Dawson's lab in 2001 where reductive hydrogenation using a palladium/Al_2O_3 catalyst was employed to convert cysteine to alanine.[126] Similar hydrogen reductive desulfurization was also reported using Ni based catalysts by Kent's group in both 2005 and 2007 (Figure 31a).[127-128] Soon later, an improved method that using VA-044 (a water-soluble radical initiator), TCEP, and thiopropionic acid to undergo a free radical based cysteine desulfurization reaction (Figure 31b) was used by Danishefsky's lab to synthesize homogeneous glycopolypeptides such as Erythropoietin (EPO) via "alanine" based native chemical ligation.[129-130]

With reliable cysteine desulfurization methods in hand, scientists started to put effort in developing system that can handle protein ligation at non-cysteine positions. For native chemical ligation, because the reactive motif of cysteine residue is the β-aminothiol group, any peptide fragments containing an *N*-terminal β-aminothiol type structure are all eligible for NCL or EPL. Therefore, NCL followed by desulfurization has been quickly expanded onto the protein synthesis at "phenylalanine",[131]

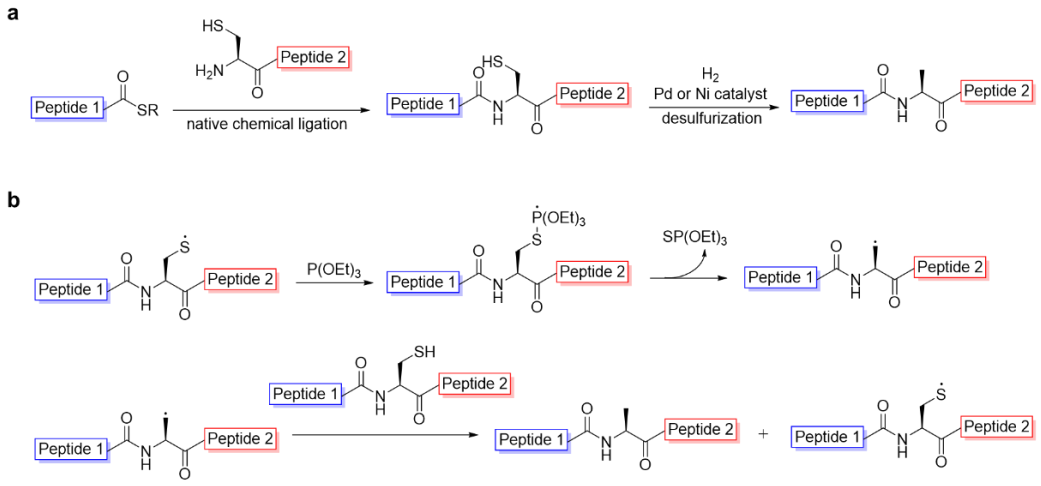

Figure 31 (a) The chemical reaction of metal catalyzed reductive desulfurization; (b) The proposed mechanism of free radical based cysteine desulfurization.

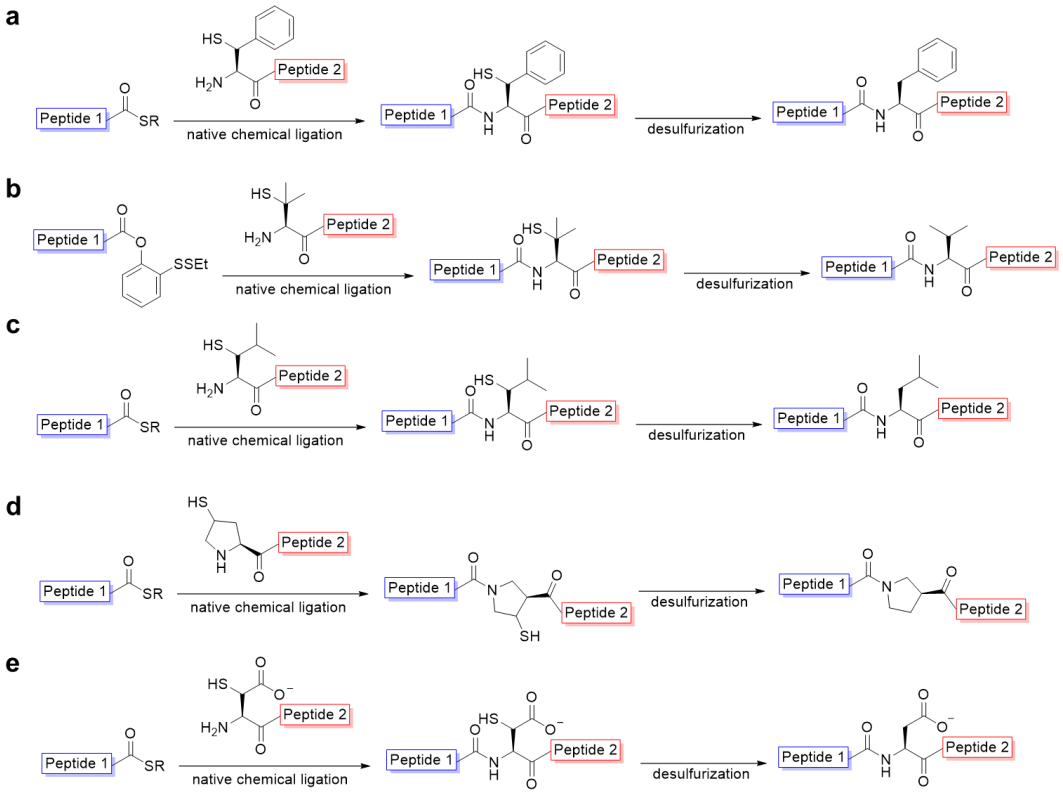

Figure 32 The scheme of native chemical ligation at (a) phenylalanine, (b) valine, (c) leucine, (d) proline, and (e) aspartic acid position using β-aminothiol.

"valine",[130,132-133] leucine,[134] proline,[135] and aspartic acid[136] using synthetic β-thiolized amino acid containing peptides (Figure 32).

Analogizing from β-aminothiol triggered protein ligation, γ-aminothiol should also play the same role during the ligation process. This idea has been successful taken by Danishefsky *et al*. to process the ligation reaction and resulted in an alternative way to ligate proteins at valine position (Figure 33a).[133] It also enable the NCL at threonine (Figure 33b)[137] and glutamine (Figure 33c)[138] position to further expand protein ligation synthetic scope. Moreover, by installing a γ-thiol group on lysine side chain, both

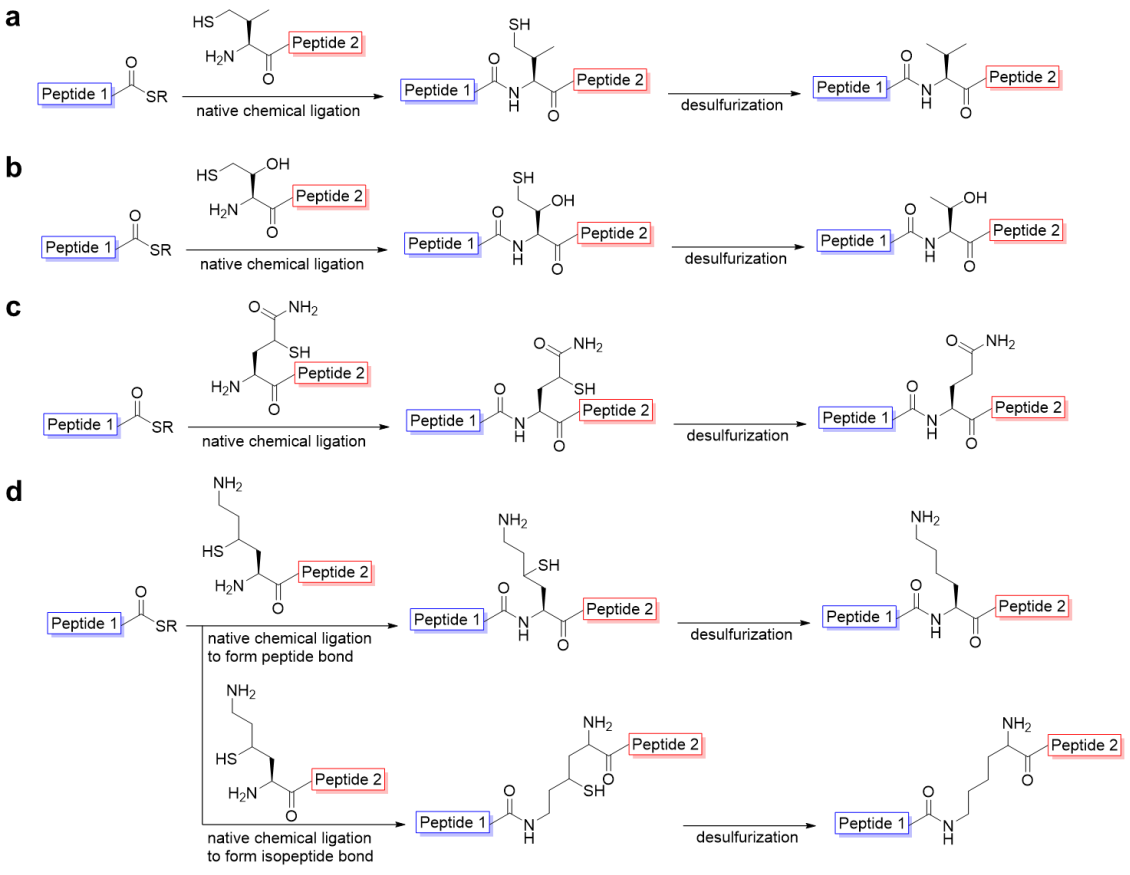

Figure 33 The scheme of native chemical ligation at (a) valine, (b) threonine, (c) glutamine, and (d) lysine position using γ-aminothiol.

peptide and isopeptide bond were able to be synthesized (Figure 33d).[139] This approach has been effectively used for the preparation of ubiquitinated protein synthesis.[140-141]

Besides regular desulfurization reaction that converting thiol to hydrogen, in 2009, Kajihara *et al*. reported a brilliant reaction which transform cysteine to serine based on the reaction scheme shown in Figure 34. Using this acidic CNBr condition, the native structure of a glycopeptide was successfully generated.[142]

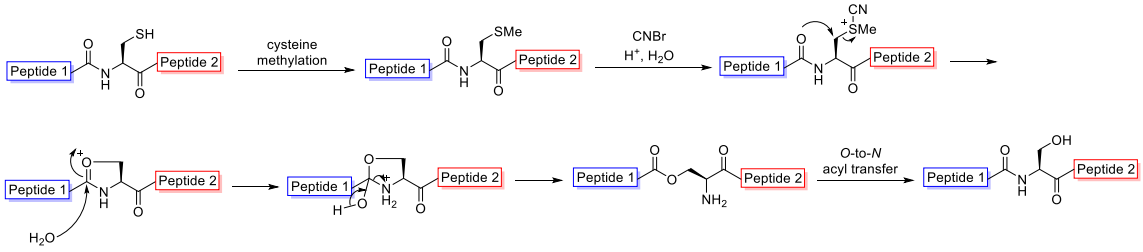

Figure 34 The proposed mechanism of CNBr induced cysteine to serine conversion.

1.3.5. Enzymatic approaches for protein C-terminal functionalization

Besides synthetic chemistry and protein semi-synthesis, protein C-terminal functionalization can also be achieved via enzyme mediated reaction. The first notable method is called sortagging in which an enzyme called sortase catalyzes the amide bond cleavage after threonine residue to afford a protein *C*-terminal thioester by recognizing a special pentapeptide sequence LPXTG. The obtained thioester containing peptide is able to rebuild the amide bond with another *N*-terminal oligoglycine containing peptide.[143] By conjugating a small functionalized motif right after the oligoglycine, by incubating with sortase, the *C*-terminal modifications are easily accomplished (Figure 35).[144] However, the sortase recognition require the unique LPXTG motif which is not

commonly existing at the *C*-termini of proteins. This would lead to the mandatory introduction of this extra sequence into protein *C*-termini. For some biological assay that are extremely sensitive to the protein sequence, this method would not be a perfect choice.

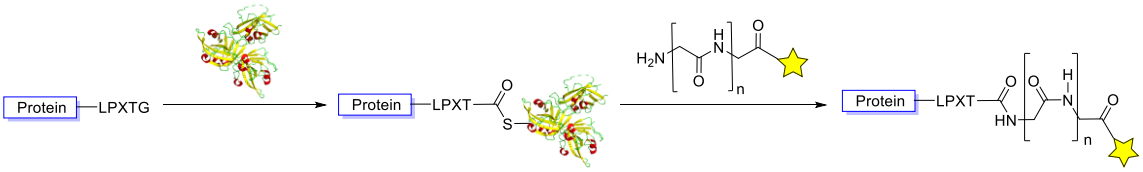

Figure 35 Protein *C*-terminal functionalization using sortase mediated ligation. Sortase A structure PDB: 1T2P.[145]

Our group previously showed that a recombinantly produced full-length Ub was directly activated by ubiquitin activating enzyme (E1) in the presence of ATP to undergo amidation reaction at its *C*-terminus with small molecule amines (Figure 36).[24] In principle, the same concept can be applied to the synthesis of Ubl-based probes. However, this E1-based activation approach does not work for Ub or Ubl variants without an intact *C*-terminus.

In 2016, Janssen and co-workers explored the peptide amidase from *Stenotrophomonas maltophilia* as a versatile catalyst for diverse carboxy-terminal peptide modification reactions. Using advanced computational protein engineering, a mutant containing 12-mutatation was determined successfully functionalized protein *C*-terminal carboxyl to amide, hydroxyl amine and ester (Figure 37).[146] For large functional group, they

anticipated further mutations characterized via computational chemistry would solve the problem.

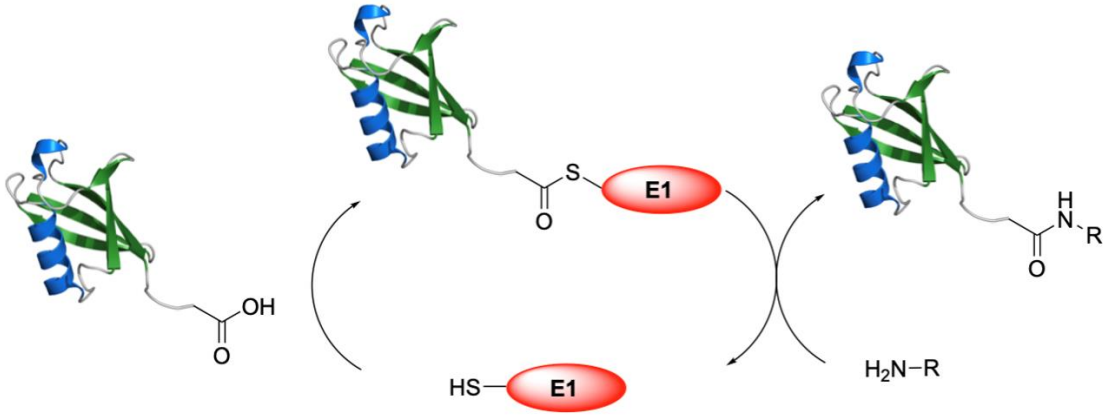

Figure 36 Facile *C*-terminal functionalization on full length ubiquitin using E1-mediated amidation reaction.

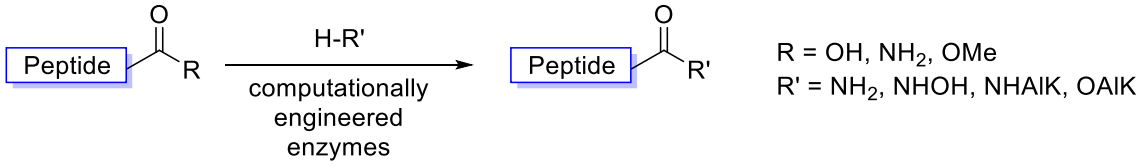

Figure 37 Peptide amidase catalyzed regioselective peptide *C*-terminal functionalization.

CHAPTER II

SITE-SPECIFIC FUNCTIONALIZATION AT PROTEIN C-TERMINI: EXPRESSED PROTEIN LIGATION WITHOUT INTEIN

2.1. Introduction

The native chemical ligation concept was first developed in 1994, several improved methods such as expressed protein ligation and peptide hydrazide ligation have extremely expand the synthetic scope of protein chemistry (Figure 38A).[105,112,118,121] Evident by the original publication garnering more than 2700 citations so far, the advent of native chemical and expressed protein ligation techniques has revolutionized the protein and peptide chemistry field. Groundbreaking applications include the synthesis of a large variety of proteins such as histones, kinases, and RAS proteins with posttranslational modifications for driving basic research advances and the production of many proteins or enzymes for therapeutic and biotechnological purposes.[115,147-153] Although developed extensively, further technological improvement in protein ligation is still necessary. The production of a protein thioester using the intein fusion approach is not guaranteed for a lot of proteins. The stringent requirement for the

---

*This chapter is reprinted with permission from Yuchen Qiao, Ge Yu, Kaci C. Kratch, Xiaoyan Aria Wang, Wesley Wei Wang, Sunshine Z. Leeuwon, Shiqing Xu, Jared S. Morse, and Wenshe Ray Liu *Journal of the American Chemical Society* **2020** *142* (15), 7047-7054 DOI: 10.1021/jacs.0c00252.

intein catalysis to generate a protein thioester prevents the processing of many fusion proteins that are expressed insolubly and hard to fold[154]. The *C*-terminal residue of a targeted protein that is immediate to the intein *N*-terminus also significantly impacts the protein splicing efficiency, which leads to low splicing efficiency for residues such as proline at this site.[155-156] The purification of an intein fusion also requires significant caution for avoiding premature hydrolysis.[112,157] A split intein may be used to prevent premature hydrolysis but adds more procedural complexity.[114-115] Using a protein ligase for expressed protein ligation resolves some issues related to the intein fusion approach but requires a specific amino acid sequence context at the ligation site.[158] Therefore, a simple method to functionalize a recombinant protein at its *C*-terminus for expressed protein ligation that requires no enzymatic catalysis, can be broadly applied, and maintains high efficiency in different protein *C*-terminal sequence contexts is highly desired. In this work, we report such a method and its application in the synthesis of a number of proteins or peptides that can be used in both basic research and therapy.

Among 20 proteinaceous amino acids, cysteine is the most nucleophilic. Its direct activation for generating a peptide thioester was previously explored by Kajihara, Otaka, and their coworkers.[159-160] To develop a more general approach that could be applied to proteins, we followed a century old industrial chemical process, leather tanning by cyanides. Cyanide salts that reduce disulfide bonds in proteins were used in the early 20th century to treat animal hides and wools.[161-162] During the process, a cyanide covalently attached to a protein cysteine to form a thiocyanate that underwent reversible intramolecular addition with the cysteine *N*-amide to generate a 1-acyl-2-

iminothiazolidine intermediate. The amide bond in this intermediate was significantly weakened in comparison to a regular protein amide and therefore slowly hydrolyzed to split the protein (Figure 38B).[163-164] Early protein chemists exploited this reaction for mapping protein sequences and replaced cyanide salts with other cyanylating reagents

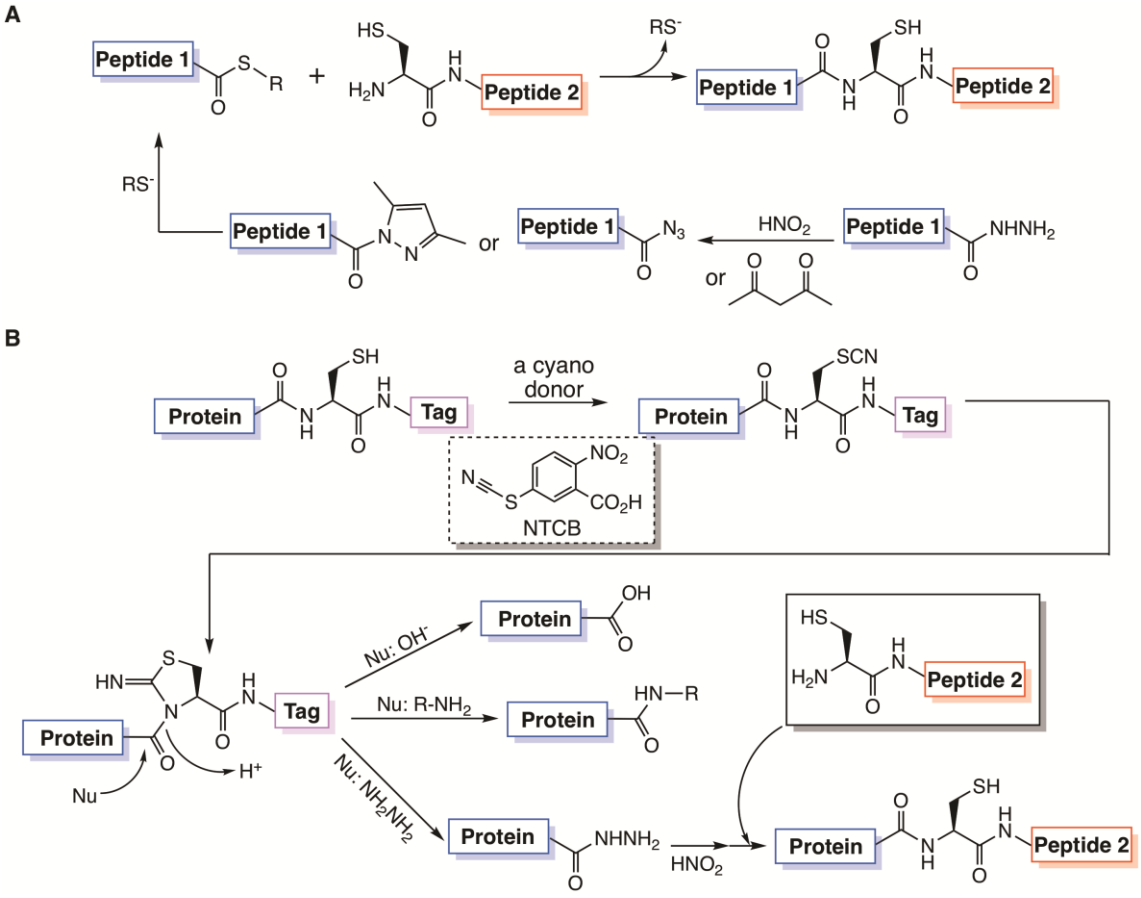

Figure 38 Protein synthesis by ligation techniques. (A) Native chemical ligation and a derivative technique, peptide hydrazide ligation; (B) A proposed protein ligation technique based on nucleophilic acyl substitution of an activated cysteine residue in a recombinant protein with a nucleophilic amine. Without a nucleophilic amine, the protein undergoes hydrolysis. When the nucleophile is hydrazine, the afforded protein hydrazide can then undergo peptide hydrazide ligation to form a larger protein.

such as 2-nitro-5-thiocyanatobenzoic acid (NTCB) that transfers the cyano group directly to a reduced protein cysteine for avoiding the formation of highly toxic cyanide wastes.[165] According to this reaction mechanism, providing a strongly nucleophilic amine in the reaction will trigger nucleophilic acyl substitution with the 1-acyl-2-iminothiazolidine intermediate to replace 2-iminothiazolidine and potentially circumvent the hydrolysis process. The afforded small molecule amine-ligated product can be used for a further protein ligation process. Since this proposed expressed protein ligation that we termed as activated cysteine-based protein ligation (ACPL) does not involve an enzyme and is purely chemically based, it can be highly controllable, selective, and versatile such as functioning for proteins both soluble and insoluble.

2.2. Results and discussion

2.2.1. Feasibility of ACPL

To demonstrate the feasibility of ACPL, we synthesized Boc-Xxx-Cys-OMe dipeptides in which the Xxx identity varied between seven native amino acids including proline and carried out their reactions with an equivalent amount of NTCB and then ligation with propargylamine (Pa) in dichloromethane (DCM). Our result showed that all dipeptides reacted with Pa to form desired products with varied yields (Table 1). Since the reactions were performed in DCM, amino acids with a hydrophobic and large side chain tended to have high yields, while amino acids with a relatively small and hydrophilic side chain resulted in relatively low yields. To further characterize the effect of the reaction on the chirality of the amino acid at the *N*-terminal side of cysteine, we

Table 1 Yields of dipeptide ligation with propargylamine.

Peptide	Amount (g)	Ligation product	Amount (g)	Yield (%)
Boc-Gly-Cys-OMe	0.6	Boc-Gly-Pa	0.12	56.6
Boc-Ala-Cys-OMe	0.6	Boc-Ala-Pa	0.11	48.6
Boc-Leu-Cys-OMe	2.0	Boc-Leu-Pa	0.7	87.5
Boc-Phe-Cys-OMe	1.8	Boc-Phe-Pa	0.65	88.3
Boc-Trp-Cys-OMe	2.5	Boc-Trp-Pa	0.8	78.4
Boc-Asp-Cys-OMe	0.9	Boc-Asp-Pa	0.05	13.8
Boc-Pro-Cys-OMe	2.0	Boc-Pro-Pa	0.16	25.4

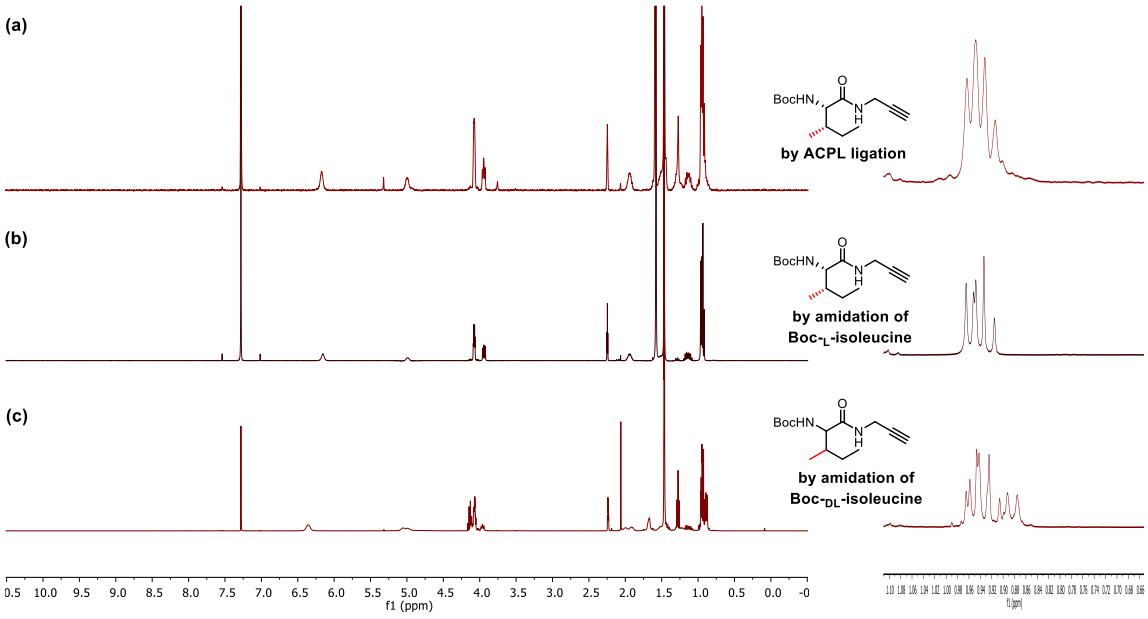

Figure 39 ^{1}H NMR spectra of tert-butyl (3-methyl-1-oxo-1-(prop-2-yn-1-ylamino)pentan-2-yl)carbamate from ACPL ligation (a), amidation of Boc-L-isoleucine (b), and amidation of Boc-DL-isoleucine (c) ($CDCl_3$, 400 MHz).

synthesized the Boc-L-isoleucine-Pa using ACPL and compared it with both Boc-$_{L}$-isoleucine-Pa and Boc-$_{DL}$isoleucine-Pa that were synthesized using the standard amidation approach as controls. As shown in the ^{1}H NMR spectra (Figure 39), no diastereomeric CH_3 peak was observed for the ligated product indicating that this reaction did not lead to the racemization of the amino acid before cysteine.

2.2.2. Versality of ACPL

Encouraged by our small molecule results, we tested ACPL further with recombinant proteins. Ubiquitin (Ub) is natively devoid of cysteine.[166-167] We chose it as a model protein for our demonstration. We produced recombinant native Ub and Ub with both a G76C mutation and a *C*-terminal 6×His tag (Ub-G76C-6H) in *E. coli* and purified them to homogeneity. We then ligated Ub-G76C-6H with 12 small molecule amines including Pa, allylamine (Aa), hydrazine (Ha), and $_{L}$- and $_{D}$-amino acids (Figure 40A) by adding 5 mM NTCB and a 50-1000 mM amine simultaneously to a 2 mg/mL Ub-G76C-6H solution at pH 9 for an overnight incubation at 37 °C. We carefully selected seven $_{L}$-amino acids for our reactions to represent amino acids in different chemical categories and also different sizes. For all tested compounds including proline that has a secondary amine and two $_{D}$-amino acids, we obtained ligation products with 50-90% yields that were estimated by the SDS-PAGE analysis of reaction mixtures (Figure 41A, Table 2). After using Ni^{2+} charged resins to simply remove unreacted intermediates, we analyzed all 12 ligation products and the two original Ub and Ub-G76C-6H proteins by electrospray ionization mass spectrometry (ESI-MS) analysis. For all analyzed proteins,

their deconvoluted ESI-MS spectra displayed clearly observable monoisotopic peaks. Since there is no commercial software for calculating protein monoisotopic peaks, we wrote a Python script to calculate all theoretical monoisotopic masses and their relative intensities for all proteins and compared them to the determined ESI-MS spectra. Our results showed that determined monoisotopic masses for all proteins agreed very well with their theoretic values in terms of both molecular weight and intensity (Figures 58-71). Hydrolysis products were either non-detectable or at very low levels. To simplify the comparison, we wrote another Python script to integrate deconvoluted monoisotopic peaks and then calculate the average molecular weights and intensities for all detected protein species in a particular spectrum. The final results are presented in Figures 40B and 40C. For all determined average molecular weights, they matched their theoretical values with a deviation of ± 0.3 Da (Table 2). For all 12 ligation products, we detected very few minor peaks in their ESI-MS spectra indicating that all reactions were very selective. One ligation product Ub-G76G is native Ub itself. Its ESI-MS spectrum in Figure 40C matched that of recombinantly expressed native Ub in Figure 40B. So far, our data demonstrated that ACPL works exactly according to what we proposed on a recombinant protein and this reaction is effective for amines that are primary, secondary, Ha, and amino acids with different configurations, characteristics, and sizes. The ligation with Ha was done in both native and denatured conditions. The results from two conditions showed minimal differences (Figures 71 and 72). Since conditions used were at 37 °C with a long incubation time that might be a concern for some proteins, we also carried out reactions under lower temperatures and shorter times. We performed the

reaction between Ub-G76C-6H and Pa under three different temperatures (8, 23, and 37 °C) as well as four different reaction times (4, 8, 12, and 16 h). The SDS-PAGE analysis of the final reaction mixtures exhibited no clear differences among all conditions. All these conditions yielded the desired product as 50 % (Figure 41C). Therefore, a reaction at 8 °C for 4 h is well enough for ACPL.

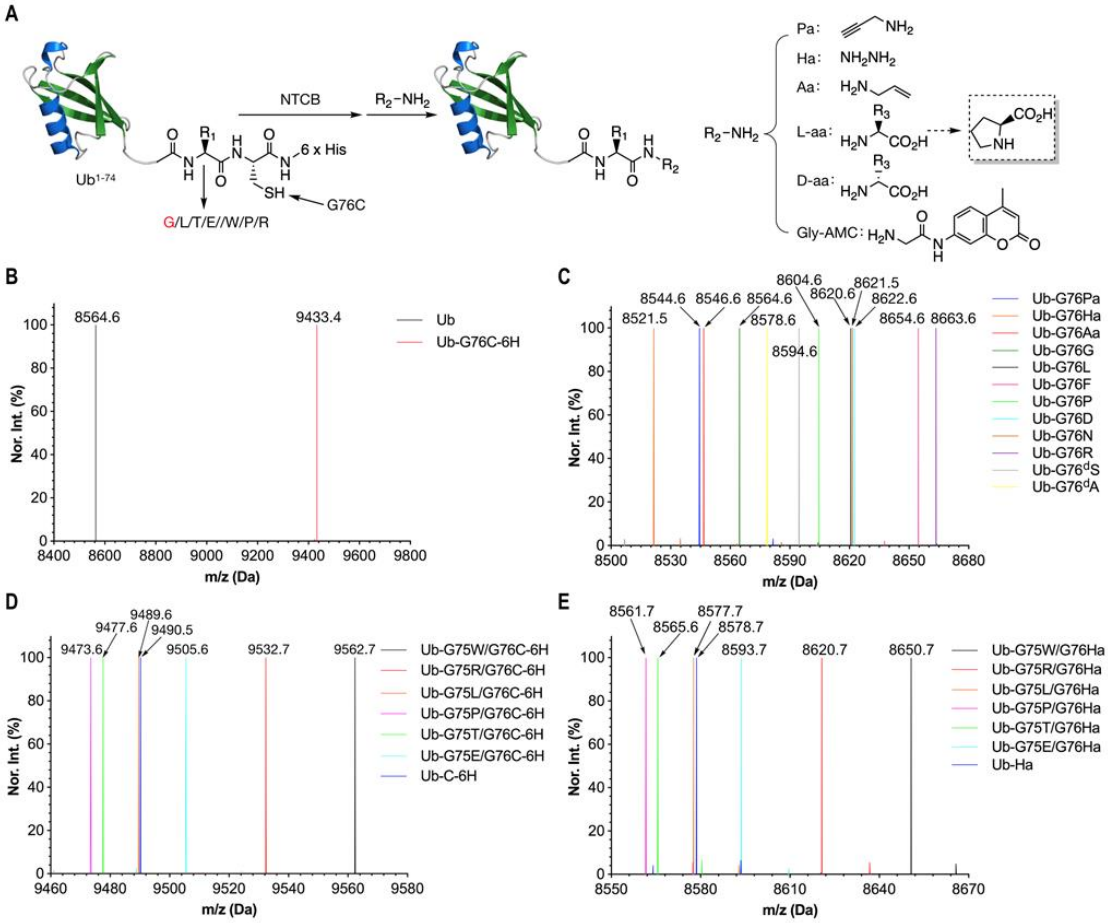

Figure 40 The synthesis of Ub conjugates by activated cysteine-directed protein ligation. (A) A schematic diagram to show the activation of recombinant Ub proteins containing a cysteine by NTCB followed by nucleophilic acyl substitution with amines, both primary and secondary, to generate different Ub conjugates. The native Ub has 76

residues and glycine at the 75th and 76th positions. (B) The deconvoluted and integrated ESI-MS of wild type Ub and Ub-G76C-6H. 6H represents a 6×His tag. (C) The deconvoluted and integrated ESI-MS of Ub conjugates that were converted from Ub-G76C-6H and had different ligated molecules at the G76 position. Pa, Ha, and Aa are three small molecule amines shown in A. All other ligated molecules are amino acids whose one letter codes are used for labeling. All amino acids are in the L-configuration except two D-amino acids with a footnote d. (D-E) The deconvoluted and integrated ESI-MS of 7 recombinant Ub proteins and products of their reactions with NTCB and Ha. C in Ub-C-6H represents cysteine. All detected molecular weights agreed well with theoretic values in a deviation range of ± 0.3 Da.

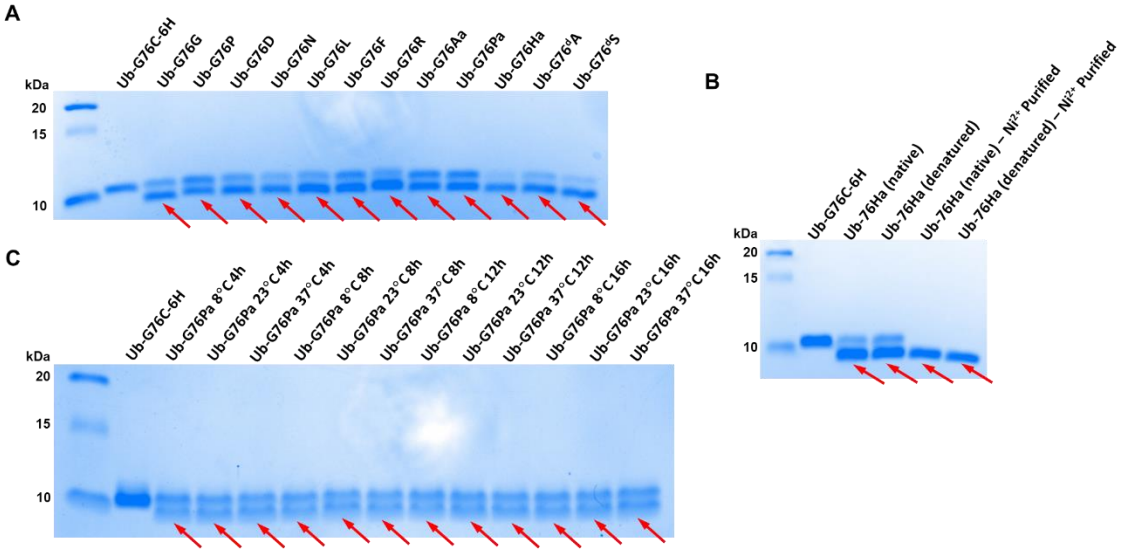

Figure 41 (A) SDS-PAGE of Ub-G76C-6H reactions with different amine-containing compounds. (B) SDS-PAGE of Ub-G76Ha synthesized under both native and denatured conditions. (C) Time and temperature depending assay between Ub-G76C-6H and Pa.

Table 2 Theoretical and determined average molecular weights of Ub, Ub-G76C-6H, and the ligation products of Ub-G76C-6H with 12 amine-containing compounds.

Protein	Observed Mass (Da)	Theoretical Mass (Da)	Yield (%)
Ub	8564.6	8564.8	N/A
Ub-G76C-6H	9433.4	9433.7	N/A
Ub-G76G	8564.6	8564.8	70
Ub-G76P	8604.6	8604.8	50
Ub-G76D	8622.6	8622.8	70
Ub-G76N	8621.5	8621.8	80
Ub-G76L	8620.6	8620.9	75
Ub-G76R	8663.6	8663.9	85
Ub-G76F	8654.6	8654.9	65
Ub-G76^{d}A	8578.6	8578.8	85
Ub-G76^{d}S	8594.6	8594.8	90
Ub-G76Aa	8546.6	8546.8	60
Ub-G76Pa	8544.6	8544.8	55
Ub-G76Ha	8521.5	8521.7	90
Ub-G76Ha (native)	8521.6	8521.7	85

Ubiquitin natively has a G75 residue that has the lowest steric hindrance among all amino acids. In Ub-G76C-6H, the glycine immediately *N*-terminal to G76C might have permitted easy processing of the ligation. Other residues that have different chemical properties and/or are sterically hindered might impede the ligation. To resolve this concern, we mutated G75 in Ub-G76C-6H to six other residues that are large in size, charged, and/or having a secondary amine, recombinantly expressed them, analyzed them with ESI-MS (Figures 40D and 73-78), and then reacted them in a one-pot fashion with NTCB and Ha. We chose Ha in our demonstration since its ligation products are protein hydrazides that can undergo further peptide hydrazide ligation for making even larger proteins. All reactions progressed exceedingly well and their reaction products displayed average molecular weights matching well to their theoretic values (Figures 40E and 79-84; Table 3), demonstrating that the residue immediately

Table 3 Theoretical and determined molecular weights of six Ub-G75X/G76C-6H proteins, Ub-C-6H, and their ligation products with hydrazine (data for the ligation products of Ub-C-6H with glycine and for the ligation products of Ub-K48C and Ub-K63C with hydrazine are also included). X denotes one of six amino acids: W, R, L, P, T, and E.

Protein	Observed Mass (Da)	Theoretical Mass (Da)	Yield (%)
Ub-G75P-G76C-6H	9473.6	9473.8	N/A
Ub-G75T-G76C-6H	9477.6	9477.7	N/A
Ub-G75L-G76C-6H	9489.6	9489.8	N/A
Ub-G75E-G76C-6H	9505.6	9505.8	N/A
Ub-G75R-G76C-6H	9532.7	9532.8	N/A
Ub-G75W-G76C-6H	9562.7	9562.8	N/A
Ub-G75P-G76Ha	8561.7	8561.8	30
Ub-G75T-G76Ha	8565.6	8565.8	70
Ub-G75L-G76Ha	8577.7	8577.8	20
Ub-G75E-G76Ha	8593.7	8593.8	45
Ub-G75R-G76Ha	8620.7	8620.9	55
Ub-G75W-G76Ha	8650.7	8650.9	20
Ub-C-6H	9490.5	9490.7	N/A
Ub-Ha	8578.7	8578.8	50
Ub-Aa	8603.8	8603.8	60
Ub-Pa	8601.7	8601.8	50
Ub-G	8621.7	8621.8	70
Ub-K48C	8539.5	8539.7	N/A
Ub-K63C	8539.5	8539.7	N/A
Ub_{1-47}-Ha	5255.8	5256.1	N/A
Ub_{1-62}-Ha	7017.8	7017.9	N/A

N-terminal to the targeted cysteine has little detrimental effect on the ligation process. However, large amino acids such as tryptophan did lead to lower yields in comparison to small amino acids such as threonine (Figure 42A, Table 3). Putting a cysteine residue right after Ub G76 led to similar ligation results with Ha, Aa, Pa and glycine with an at least 50 % conversion rate (Figures 40D, 40E, 42B and 85-89 and Table 3). Ub has a flexible *C*-terminus that may facilitate the ligation. To show that the ligation may work in a more structurally constrained environment, we introduced a cysteine mutation at

K48 and K63, two residues in the globular region of Ub and used the two afforded Ub mutants (Figures 90-91 and Table 3) to undergo ACPL with Ha. ESI-MS of reaction mixtures showed successful formation of two desired protein hydrazides (Figures 92-93 and Table 3) indicating that ACPL works well in a structurally constrained protein region. Ligation both in a structurally constrained protein region and under a denatured condition is something that the traditional intein and ligase-based methods cannot perform well. Collectively our data strongly demonstrates the versatility of the ACPL technique.

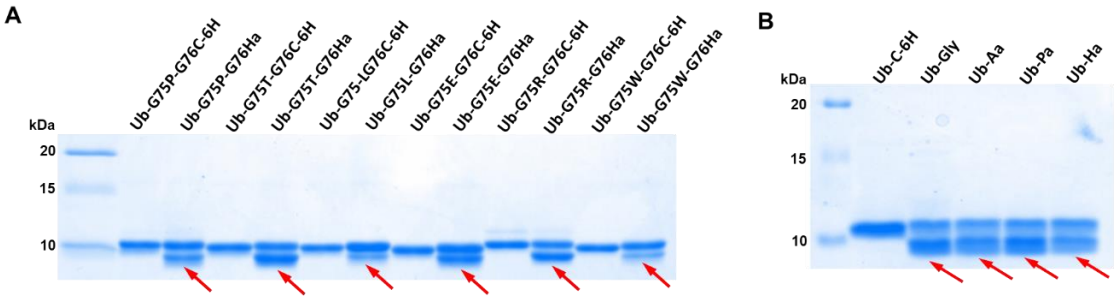

Figure 42 (A) SDS-PAGE of Ub-C-6H reactions with 4 amine-containing compounds. (B) SDS-PAGE of Ub-G75X/G76C-6H with hydrazine (X denotes P, T, L, E, R, or W).

2.2.3. Use of ACPL to synthesize Ub and Ub-like protein probes

In eukaryotic cells, Ub and Ubls can be posttranslationally attached to proteins for their functional regulation.[168-170] It has been shown that replacing the *C*-terminal glycine in Ub and Ubl proteins including SUMO1-3, NEED8, and ISG15 with Pa using either the intein based approach or total synthesis afforded excellent probes to conjugate covalently to DUBs or ULPs that catalytically remove Ub or Ubls from their conjugated proteins in cells.[89,171-174] To recapitulate these results and demonstrate the broad

application scope of our ACPL technique in the probe synthesis, we recombinantly expressed Ub and a number of Ubl proteins including SUMO1-4, NEDD8, ISG15, GABARAP, GABARAPL2, UFM1, URM1, and MNSFβ (FLAG-Ub/Ubl-GxC-6H: x denotes the terminal glycine position) that all contained a *C*-terminal Gly-to-Cys mutation and were also fused with a *N*-terminal FLAG tag and a *C*-terminal 6×His tag, purified them to homogeneity, and then carried out their reactions with Pa in the presence of NTCB to afford their Pa-conjugated products. ISG15, SUMO1-4, and MNSFβ natively contain a cysteine residue. This cysteine was mutated to alanine or serine to avoid non-targeted reaction at its location. The yields of all 12 reactions varied between 25% and 80% as indicated by their SDS-PAGE analysis (Figure 43, Table 4). ESI-MS analysis of all 12 products showed their successful and efficient synthesis (Figures 44A and 94-117; Table 4). The final results are presented in Figures 3A that displayed very little side products for all 12 Pa-conjugated products. In comparison to both intein based and total synthesis approach, our method for the synthesis of these Pa conjugates is much simpler and easier to control. To reproduce some literature results, we used our synthesized Pa-conjugated FLAG-Ub (FLAG-Ub-G76Pa) to react with seven DUBs and observed efficient covalent adduct formation for all tested enzymes in both SDS-PAGE analysis and Western blotting (Figures 44B and 45). We also performed similar tests for seven Pa-conjugated FLAG-Ubl probes and observed their covalent binding to a number of ULPs as shown in Figure 44C. Some ULPs such as SENP1 have only been vaguely confirmed in previous work to deconjugate corresponding Ubls such as SUMO4.[175] All synthesized FLAG-Ub/Ubl conjugates, of

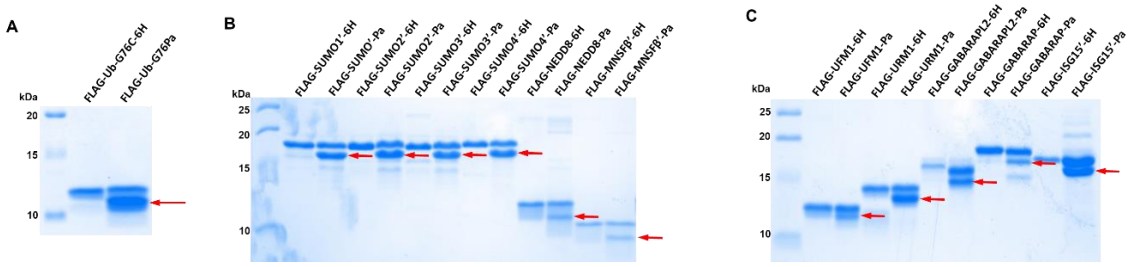

Figure 43 (A) SDS-PAGE of FLAG-Ub-G76C-6H reactions with Pa under different temperatures and durations. (B) and (C) SDS-PAGE of 11 different FLAG-Ub/Ubl-GxC-6H reactions with propargylamine (x denotes the *C*-terminal glycine position).

Table 4 Theoretical and determined molecular weights of FLAG-Ub/Ubl-GxC-6H, their ligation products with propargylamine and AMC (Ub-AMC was generated from Ub-G76C-6H instead of FLAG-Ub-G76C-6H). x denotes the last glycine positions in different Ub and Ubls.

Protein	Observed Mass (Da)	Theoretical Mass (Da)	Yield (%)
FLAG-Ub-G76C-6H	10701.8	10702.0	N/A
FLAG-Ub-G76Pa	9812.9	9813.1	80
FLAG-SUMO1(C52A)-G97C-6H	13237.4	13237.6	N/A
FLAG-SUMO1(C52A)-G97Pa	12348.4	12348.7	60
FLAG-SUMO2(C48A)-G93C-6H	12713.8	12714.0	N/A
FLAG-SUMO2(C48A)-G93Pa	11825.6	11825.1	60
FLAG-SUMO3(C47A)-G92C-6H	12629.2	12629.9	N/A
FLAG-SUMO3(C47A)-G92Pa	11741.3	11741.0	60
FLAG-SUMO4(C48A)-G93C-6H	12527.6	12527.9	N/A
FLAG-SUMO4(C48A)-G93Pa	11639.5	11639.0	60
FLAG-NEDD8-G76C-6H	10696.8	10697.1	N/A
FLAG-NEDD8-G76Pa	9808.0	9808.2	40
FLAG-ISG15(C89S)-G157C-6H	19266.6	19266.8	N/A
FLAG-ISG15(C89S)-G157Pa	18377.8	18377.8	50
FLAG-GABARAP-G116C-6H	15941.7	15942.0	N/A
FLAG-GABARAP-G116Pa	15052.9	15053.1	25
FLAG-GABARAPL2-G116C-6H	15640.5	15640.7	N/A
FLAG-GABARAPL2-G116Pa	14751.6	14751.8	40
FLAG-UFM1-G83C-6H	11064.3	11064.5	N/A
FLAG-UFM1-G83Pa	10175.4	10175.6	30
FLAG-URM1-G101C-6H	13517.0	13517.2	N/A

which six are synthesized for the first time, are activity-based probes that can be potentially used to profile DUB and ULP proteomes in different tissues or cells. As a demonstration, we incubated the HEK293T cell lysate with FLAG-Ub-G76Pa and then probed the FLAG-Ub-conjugated proteins by an anti-FLAG antibody in Western blotting. The result showed the formation of a number of higher molecular weight species compared to the original FLAG-Ub-G76Pa, indicating conjugation with other proteins in the cell lysate (Figure 46). However, control reaction using FLAG-Ub-G76C-6H showed no covalent conjugation with any other proteins in the HEK293T cell lysate.

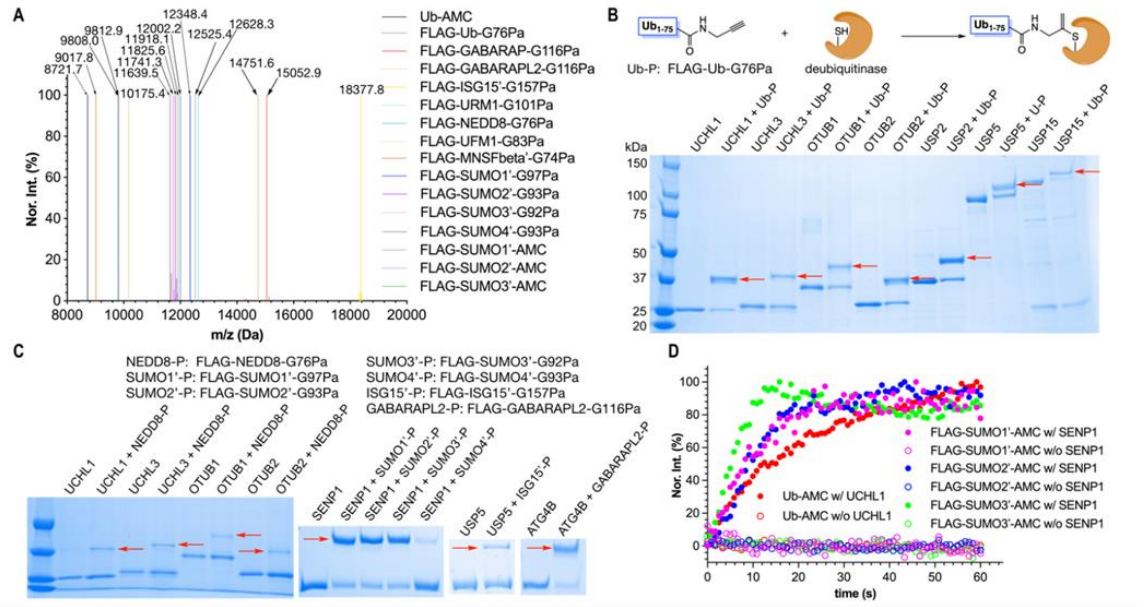

Figure 44 The synthesis of FLAG-Ub/Ubl-Pa and Ub/FLAG-SUMO1-3-AMC probes and their applications in covalent conjugation or activity assays of DUB/ULPs. (A) The deconvoluted and integrated ESI-MS of FLAG-Ub/Ubl-Pa and Ub/FLAG-SUMO1-3-AMC probes. Ub-AMC was synthesized from Ub-G76C-6H. All other Pa- and AMC-

conjugated Ub/Ubls were generated from FLAG-tagged proteins. Ub/Ubls with their *C*-terminal glycine mutated to cysteine were expressed and purified as a protein fused with a *N*-terminal FLAG tag and a *C*-terminal 6×His tag. ISG15, SUMO1-4, and MNSF have a native cysteine residue. This cysteine was mutated to alanine or serine in all six expressed proteins for avoiding side reactions. The label "′" indicates this mutation. All detected molecular weights agreed well with their theoretic values with a deviation range of 0.5 Da. (B) The formation of covalent adducts between FLAG-Ub-G76Pa and a number of DUBs. Red arrows point to the generated adducts. (C) The formation of covalent adducts, indicated by red arrows, between different FLAG-Ubl-GxPa probes and DUB/ULPs. (D) The DUB/ULP-catalyzed AMC release from Ub-AMC and three FLAG-SUMO-AMC conjugates.

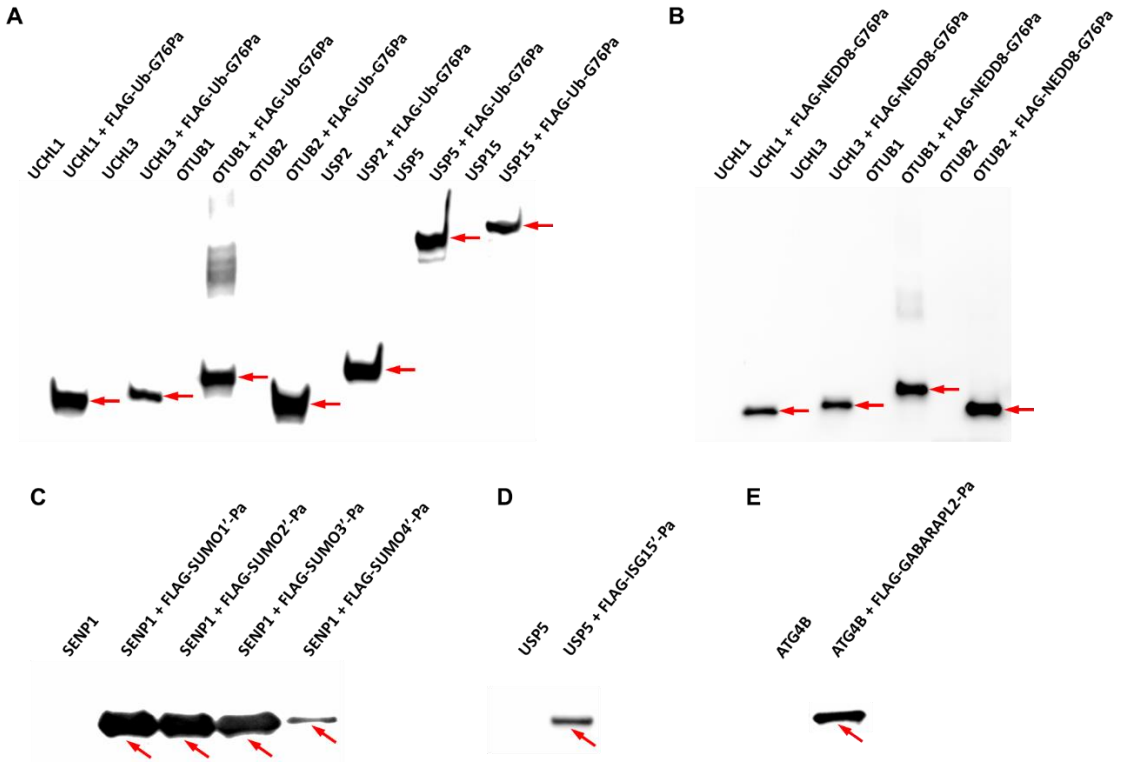

Figure 45 Western blotting. (A) The indication of the formation of covalent adducts between FLAG-Ub-G76Pa and UCHL1, UCHL3, OTUB1, OTUB2, USP2, USP5, USP15 by using anti-DDDDK antibody. (B) The formation of covalent adducts between FLAG-NEDD8-G76Pa and UCHL1, UCHL3, OTUB1, OTUB2 by using anti-DDDDK antibody. (C) The formation of covalent adducts between FLAG-SUMO1'/2'/3'/4'-GxPa (x denotes the *C*-terminal glycine position) and SENP1 by using anti-DDDDK antibody. (D) The formation of covalent adducts between FLAG-ISG15'-G157Pa and USP5 by using anti-DDDDK antibody. (E) The formation of covalent adducts between FLAG-GABARAPL2-G116Pa and ATG4B by using anti-DDDDK antibody. Arrows in red point to DUB-Ub/Ubl covalent adducts.

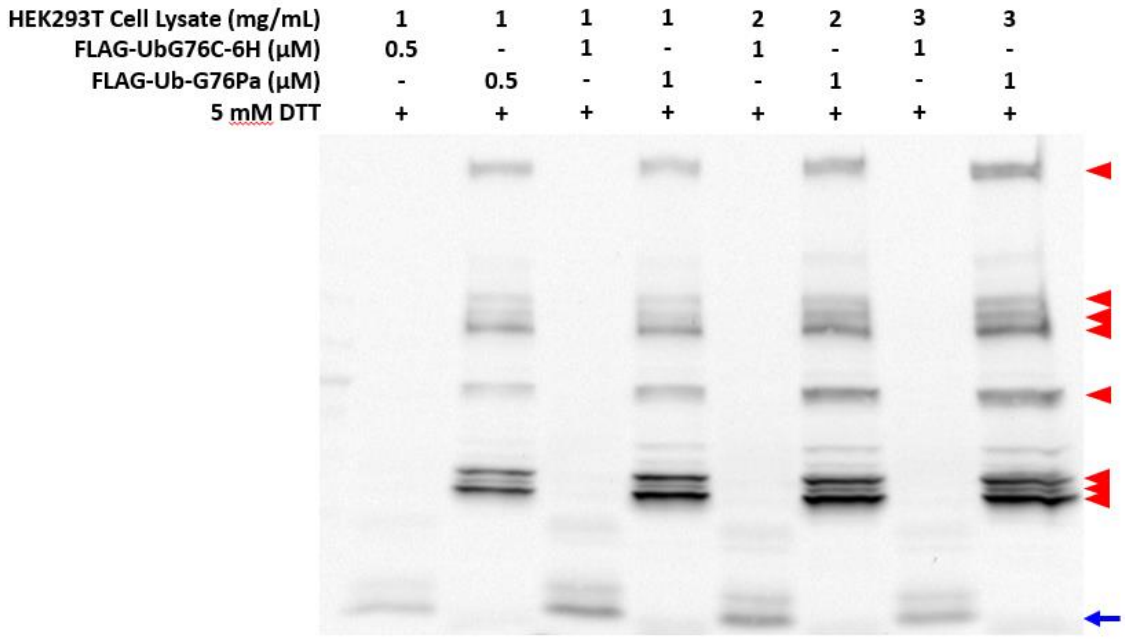

Figure 46 Profiling DUBs in the HEK293T cell lysate using FLAG-Ub-G76Pa. The gel was blotted by the anti-FLAG antibody. The cell lysate concentration was varied from 1

to 3 mg/mL. The reaction with FLAG-Ub-G76C-6H was used as a negative control. High molecular weight bands indicated by red triangles revealed multiple proteins presumably DUBs that covalently bound to FLAG-Ub-G76Pa. Blue arrow points to the FLAG-UbG76C-6H dimer.

Ub and Ubls conjugated directly to 7-amido-4-methylcoumarin (AMC) at their *C*-terminus are useful fluorogenic substrates of DUBs and ULPs.[176] To demonstrate the synthesis of Ub/Ubl-AMC conjugates using our ACPL technique, we made Ub-AMC and FLAG-SUMO1-3-AMC by reacting recombinantly produced Ub-G76C-6H and FLAG-SUMO1-3-GxC-6H proteins with Gly-AMC in the presence of NTCB. The ESI-MS analysis of all four products confirmed their successful formation (Figures 44A and 118-121) and the following activity assays showed that they served as active substrates for cysteine proteases UCHL1 and SENP1, respectively (Figure 44D). Overall, our combined data of Ub/Ubl probe synthesis establish the broad application scope of the ACPL technique and this technique can make Ub/Ubl probes readily available in a manner that can be performed in almost any biology lab for advancing Ub and Ubl biology studies.

2.2.4. Use of ACPL to synthesize H2AK129ac

To further demonstrate its broad application scope, we also carried out ACPL to make non-ubiquitin or non-ubiquitin-like proteins. In human cells, histone H2A can undergo posttranslational acetylation at its terminal lysine K129.[177] The functional

investigation of this acetylation such as how it influences the structure and dynamics of the nucleosome will require the synthesis of the corresponding acetyl-histone, H2AK129ac. We chose to synthesize H2AK129ac to demonstrate that our method can be applied to the synthesis of histones with *C*-terminal modifications. As a matter of fact, except H3 that natively has a cysteine, all other core histones are lack of cysteine residues. We first recombinantly produced H2A-K129C-6H, an H2A protein with a K129C mutation and a *C*-terminal 6×His tag and then ligated it to N^{ε}-acetyl-lysine with the assistance of NTCB. The ESI-MS spectrum of the reaction product showed the formation of H2AK129ac (Figures 47A and 122-123; Table 5) with an at least 60% yield. We folded successfully H2AK129ac into a dimer with H2B and subsequently into a nucleosome (Figure 47B) making it possible to study effects of H2AK129ac on the nucleosome structure and function. In the ESI-MS spectrum of H2AK129ac in Figure 47A, we noticed a minor peak at 13813.6 Da. H2A has two lysine residues, K125 and K127 that are adjacent to K129. Both crystal and cryo-EM structures have indicated all three lysines are located at a flexible *C*-terminal region of H2A.[178-179] Potentially either K125 or K127 can undergo intramolecular cyclization with an activated K129C. These intramolecular cyclization products will have a theoretical molecular weight (13813.9 Da) that matches the minor peak in the ESI-MS spectrum. Based on the determined intensities of peaks in the ESI-MS spectrum, this minor peak was lower than 15% of the desired product demonstrating the selectivity of ACPL. By tuning the reaction conditions, this minor peak might be further reduced.

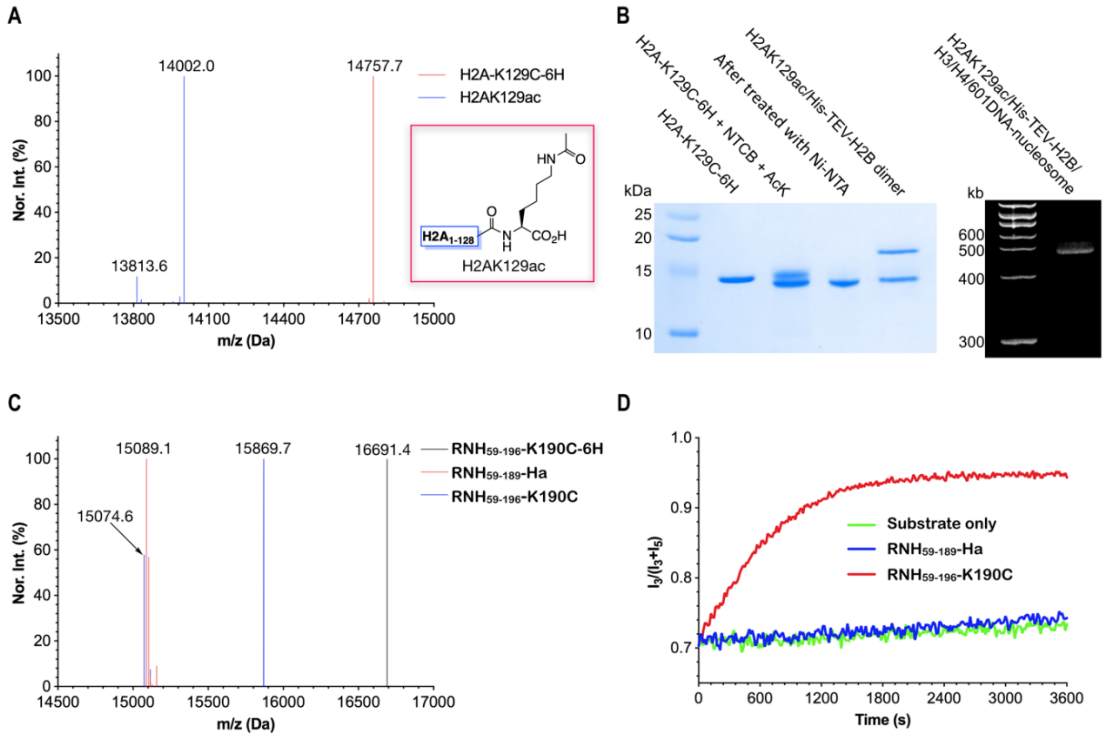

Figure 47 The synthesis of H2AK129ac and RNase H by activated cysteine-directed protein ligation. (A) The deconvoluted and integrated ESI-MS of H2A-K129C-6H and H2AK129ac. H2A-K129C-6H was recombinantly expressed and then reacted with NTCB and N^{ε}-acetyl-lysine to afford H2AK129ac. (B) The synthesis of H2AK129ac, its isolation, and folding into an H2AK129ac/H2B dimer and then a nucleosome. The purification of H2AK129ac was achieved by extracting the unreacted intermediate using Ni-NTA resins. (C) The deconvoluted and integrated ESI-MS of $RNH_{59\text{-}196}$-K190C-6H, $RNH_{59\text{-}189}$-Ha, and $RNH_{59\text{-}196}$-K190C. $RNH_{59\text{-}196}$-K190C-6H was recombinantly expressed in *E. coli*. It was reacted with NTCB and Ha to afford $RNH_{59\text{-}189}$-Ha that then underwent peptide hydrazide ligation with a 7-mer NH_2-CADYGRK-OH peptide to form a catalytic active $RNH_{59\text{-}196}$-K190C. (D) The catalytic hydrolysis of an RNA

substrate by RNH_{59-196}-K190C. The RNA substrate had a sequence 5'-Cy3-GACACCUGAUUC-Cy5-3'. A DNA fragment 5'-GAATCAGGTGTC-3' was used to form a double strand with the RNA substrate for binding to RNH_{59-196}-K190C. The hydrolysis led to improved Cy3 (I_3) and decrease Cy5 (I_5) emission.

Table 5 Theoretical and determined molecular weights of two H2A, three RNase H proteins and two exenatide peptides.

Protein	Observed Mass (Da)	Theoretical Mass (Da)
H2A-K129C-6H	14757.7	14757.9
H2AK129ac	14002.0	14002.2
$RNAseH_{59-196}$-K190C-6H	16691.4	16691.7
$RNAseH_{59-189}$-Ha	15089.1	15088.9
$RNAseH_{59-196}$-K190C	15869.7	15868.8
exenatide-S39C-SA-Strep	5402.0	5401.9
exenatide	4186.4	4186.6

2.2.5. Use of ACPL in combination with peptide hydrazide ligation to synthesize an active RNase H

For all ligation reactions that we performed thus far, they involved small molecules with only one amino group for avoiding side product formation. For ligation with larger molecules or peptides that have more than one amino group, one can couple ACPL with peptide hydrazide ligation to resolve non-specificity issues. To demonstrate this prospect, we recombinantly produced a *B. halodurans* RNase H region with a *C*-terminal Cys-6×His tag (RNH_{59-196}-K190C-6H). Its ligation with Ha in the presence of NTCB led to the synthesis of RNH_{59-189}-Ha, a protein hydrazide that we proceeded further to undergo peptide hydrazide ligation with a 7-mer peptide, NH_2-CADYGRK-OH to afford a ligated product RNH_{59-196}-K190C.[180] ESI-MS analysis showed the

successful synthesis of both $RNH_{59\text{-}189}$-Ha and $RNH_{59\text{-}196}$-K190C (Figures 47C and 124-126; Table 5). Similar to what has been found in previous peptide hydrazide ligation reactions, we also detected a minor hydrolysis product at 15074.6 Da.[181] The ligated product $RNH_{59\text{-}196}$-K190C was catalytically active to hydrolyze an RNA substrate as shown in Figure 47D. In the contrary, $RNH_{59\text{-}189}$-Ha was completely inactive toward this substrate. Our data related to the synthesis of RNase H demonstrated that ACPL can couple to peptide hydrazide ligation for conjugation with large peptides or even protein fragments.

2.2.6. Use of ACPL to synthesize exenatide, an anti-diabetic medication from a recombinant precursor

Exenatide is a 39-mer commercial anti-diabetic peptide drug that has a *C*-terminal amide. The presence of this *C*-terminal amide makes it difficult to produce exenatide using the recombinant expression technique.[182] This bottleneck can be presumably resolved using our ACPL technique by recombinant expression of a precursor protein and then an ACPL reaction with serinamide. To demonstrate this potential, we expressed a 6×His-SUMO-exenatide-S39C-SA-Strep fusion protein that can be largely produced in *E. coli* followed by the treatment with SUMO protease to obtain the exenatide-S39C-SA-Strep peptide. We then proceeded to carry out its reaction with L-serinamide in the presence of NTCB, purified the final product with a Strep-Tactin column, and analyzed it by ESI-MS. Our results showed that exenatide can be easily procured by this approach (Figures 127-128).

2.3. Summary

In summary, we have developed a novel expressed protein ligation technique that uses a cyanylating reagent to directly activate a cysteine in a recombinant protein for ligation with small molecule amines and large peptide or protein fragments when coupling with peptide hydrazide ligation. These small molecule amines are both primary and secondary and include a number of L- and D-amino acids and functional amines such as Ha, Aa, Pa, and Gly-AMC. The technique that is termed ACPL requires no enzymatic catalysis and is controllable, versatile, specific, and very simple to process. Therefore, it can be broadly applied to synthesize a large variety of proteins with unique functionalities for advanced applications in both basic and applied research. One potential industrial application of the technique is to synthesize therapeutic peptides or proteins like exenatide. ACPL requires the activation of cysteine, one of the two lowest occurring amino acids in proteins. Non-targeted cysteines need to be mutated. For proteins with essential cysteines, one solution for using ACPL is to couple it with the noncanonical amino acid mutagenesis technique. Photocaged cysteines have been genetically incorporated into proteins by amber suppression.[183-184] The incorporation of a photocaged cysteine to essential cysteine sites in a protein followed by ACPL and then decaging to release protected essential cysteines will allow the processing of proteins with non-targeted cysteines. Overall, our ACPL technique expands to a large extent the synthetic capacity of protein chemistry and will energize the whole field. We anticipate its broad applications in a large variety of research fields and industrial processing of proteins and peptides.

CHAPTER III

SITE-SPECIFIC CONVERSION OF CYSTEINE IN A PROTEIN TO DEHYDROALANINE USING 2-NITRO-5-THIOCYANATOBENZOIC ACID

3.1. Introduction

The currently known genetic code specifies 22 proteinogenic amino acids that are incorporated biosynthetically into proteins during translation. Built upon them, organisms, especially eukaryotes, undergo a plethora of posttranslational modifications (PTMs) to impart proteins with a large variety of unique functions. Functional annotation of PTMs requires the synthesis of proteins that contain them. In cells, PTMs are mostly installed in proteins via enzymatic reactions. Due to the facts that a lot of PTMs have no known enzymes and many PTM-installing enzymes are promiscuous, chemical biologists have been focusing long on the development of synthetic and semi-synthetic methods for the synthesis of proteins with PTMs for their functional investigations.[109,185-187] Notable methods include native chemical ligation,[188] expressed protein ligation,[112] and the genetic code expansion technique that relies on the

---

*This chapter is reprinted with permission from Yuchen Qiao, Ge Yu, Sunshine Z. Leeuwon, and Wenshe R. Liu. Site-Specific Conversion of Cysteine in a Protein to Dehydroalanine Using 2-Nitro-5-Thiocyanatobenzoic Acid. *Molecules* **2021**, *26* (9), 2619. DOI: 10.3390/molecules26092619. Published by MDPI.

suppression of amber codon for the incorporation of noncanonical amino acids that are post-translationally modified proteinogenic amino acids themselves or can be chemically modified to form amino acids with PTMs.[189-193] Built upon the unique reactivity of cysteine, methods have also been developed for its conversion to a variety of PTM analogues. As the most nucleophilic amino acid, cysteine can be directly conjugated to many electrophiles to form PTM analogues[194] or chemically converted to dehydroalanine (Dha).[82] Dha contains an α, β-unsaturated carbonyl moiety that can undergo both Michael addition and cross coupling reactions to generate PTM analogues or genuine PTMs respectively.[68,76] Related applications include functional annotation of histone PTMs.[195] Moreover, Dha in a protein might form a reactive chemical probe for covalent conjugation with associating partners for their identification. Various mono-/poly-ubiquitin (Ub) or ubiquitin like protein (Ubl) based Dha probes have been successfully used to capture enzymes functioning in Ub and Ubl pathways.[69-72] Those Ub/Ubl-Dha molecules have also been used for the synthesis of ubiquitinated or Ubl tagged proteins to discover the selectivity and linkage specificity of deubiquitinases (DUBs) and ubiquitin-like proteases (ULPs).[31-32,74]

To chemically convert cysteine to Dha in a protein in relatively mild conditions, Davis *et al.* developed several reagents.[76] Two of them, 2,5-dibromohexanediamide (DBHDA)[76] and methyl 2,5-dibromopentanoate (MDBP),[84] are the most used that selectively convert cysteine to Dha by bis-alkylation and then elimination in the aqueous system. Both reagents have been successfully practiced in the synthesis of antibody-drug conjugates and proteins with PTM analogues.[76,84-85] For efficient conversion of cysteine

to Dha when using DBHDA and MDBP, a high pH is typically used to achieve a high yield. Sometimes a high temperature might also be required.[84-85] There are other methods available, but they generally require harsh conditions that are not well compatible with proteins. Dha can also be generated in a protein through the genetic incorporation of an alkylated selenocysteine followed by oxidation and then elimination or a phosphoserine followed by elimination as demonstrated by Schultz, Liu, Park, Chen, and their coworkers.[196-199] Another technique called genetically encoded chemical conversion (GECCO) in which serine/threonine is selectively converted to Dha or dehydrobutyrine (Dhb) by the assistance of a proximal, genetically encoded fluorosulfo-tyrosine (FSY) residue was also reported.[200] In comparison to cysteine modification methods, these later methods are more complicated. In the current work, we wish to report a simple method that uses the readily available, highly water soluble 2-nitro-5-thiocyanatobenzoic acid (NTCB) to react with cysteine in a protein for its conversion to Dha at pH 7 and its use in the successful synthesis of several Ub- and Ubl-based, Dha-containing probes.

3.2. Results

3.2.1. Inspiration of NTCB induced dehydroalanine formation

NTCB is a highly reactive reagent that transfers rapidly its cyano group to a nucleophilic thiolate. When provided to a protein, it will quickly cyanylate its cysteine to form *S*-cyano-cysteine that undergoes reversible intramolecular addition with the cysteine *N*-amide to generate 1-acyl-2-iminothiazolidine, an intermediate that can

undergo nucleophilic acyl substitution. By taking advantage of this nucleophilic acyl substitution, we have previously developed a technique termed activated cysteine-based protein ligation (ACLP).[201] During the development of ACPL, we noticed that a protein was usually not fully converted to a ligation product. The SDS-PAGE analysis of the reaction showed a protein band that ran at the same place as the original protein (Figure 41). We previously reasoned that this was the unreacted original protein. HPLC analysis of products after the ACPL reaction between Ub-G76C-6H and propargylamine (Pa) showed two protein peaks (Figure 48a). Electrospray ionization mass spectrometer (ESI-MS) analysis of Peak 2 indicated the desired product from the ACPL reaction (Figure 48b). However, the deconvoluted ESI-MS spectrum of Peak 1 revealed a molecular weight of 9454.4 Da (Figure 48c) that did not match the theoretical mass of Ub-G76C-6H (9433.7 Da), disapproving our assumption of incomplete conversion of Ub-G76C-6H. This detected molecular weight matches the replacement of the cysteine thiol group in Ub-G76C-6H with propargylamine. This replacement can potentially happen in two possible pathways: the cyanylation of the cysteine thiolate makes it an easy leaving group for undergoing a SN2 reaction with propargylamine or the formed thiocyanate-protein adduct undergoes a beta elimination reaction to generate a Dha residue that then reacts with propargylamine via aza-Michael addition (Figure48d). The Dha formation via beta elimination of *S*-cyano-cysteine was previously predicted and observed at a very low level when NTCB was used to hydrolyze proteins at a high pH.[202-204] We deemed that this beta elimination to generate Dha and then addition with propargylamine is the most probable mechanism of the formation of the Peak 1 protein. Based on the HPLC

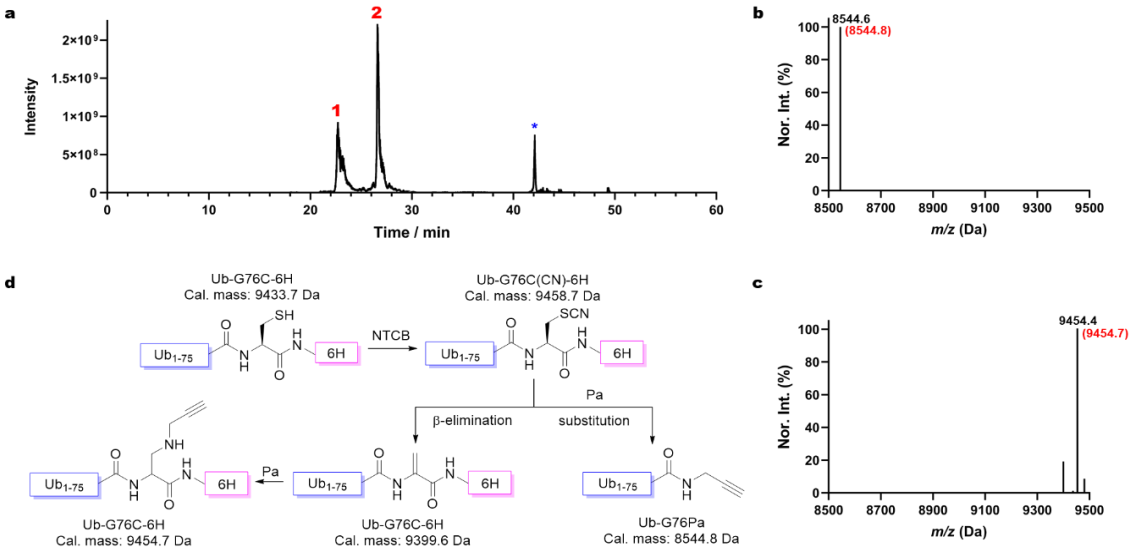

Figure 48 The observation of an aza-Michael addition product. (a) The HPLC chromatogram of products after the ACPL reaction between Ub-G76C-6H and Pa. * indicates the solvent peak. (b) The deconvoluted and integrated ESI-MS spectrum of Peak 2 shown in A. The parenthesized number colored in red is the theoretical molecular weight of the conjugation production of Ub-G76Pa (Pa replaces G76C-6H in Ub-G76C-6H). (c) The deconvoluted and integrated ESI-MS spectrum of Peak 1 shown in A. The parenthesized number colored in red is the theoretical molecular weight of the product in which Pa replaces the thiol group in Ub-G76C-6H. (d) A diagram showing two reaction pathways of a cyanylated Ub-G76C-6H product, one undergoing conjugation with Pa and the other undergoing beta elimination to form Dha and then aza-Michael addition with Pa.

chromatogram in Figure 1a, the Peak 1 protein is about 40% of the overall final products. This high level of Dha formation is likely due to a relatively milder pH we

used to do the ACPL reaction in comparison to that for the traditional hydrolysis reaction.

3.2.2. Optimizations on NTCB triggered Dha formation

According to the scheme shown in Figure 48d, the substitution with Pa and the beta elimination are two competing reactions. Withdrawing Pa from the reaction will potentially improve the yield toward Dha and also keep Dha intact (Figure 49a). We used Ub-G76C-6H to test this prospect. After incubation in a buffer containing 0.5 mM TCEP and 5 mM NTCB at pH 9 and 37 °C overnight, about 80% of Ub-G76C-6H were successfully converted to Ub-G76Dha-6H (Figure 49b and 129, Table 6), confirming an improved Dha formation yield in the absence of a strong amine nucleophile. The major byproduct was determined to be $Ub_{(1-75)}$, a hydrolysis product in which a water molecule substituted G76C-6H (Figure 49c). We then further optimized the reaction condition by lowering down the pH value to reduce the undesirable hydrolysis. After overnight incubation, LC-MS analysis of reaction products indicated that as pH decreased, the percentage of Ub-G76Dha-6H kept increasing. The yield ranged from 79.9% to 89.5% (Figure 130-132, Table 1). Since the yield was not improved further when pH was decreased from 7 to 6.5, we selected pH 7 as the standard pH value for further optimization. We then tested the temperature effect on this reaction. Despite that the hydrolysis product diminished significantly when the temperature went down, the Dha formation was also dramatically decreased (Table 6, Figure 130, 133 and 134). At 5 °C, the predominant product was the cyanylated intermediate of Ub-G76C-6H (Table 6,

Figure 49d). Another minor side product that we observed when reactions were carried out at room or low temperature was a thionitrobenoate (TNB) adduct of Ub-G76C-6H (Figure 49e). The addition of TCEP was to reduce the disulfide bond such as in the TNB adduct. Its observation indicated that TCEP was exhausted during the overnight incubation process. TNB is a product of the cyanylation step and will form an oxidized dimer when TCEP is exhausted. This dimer will then react with Ub-G76C-6H to form

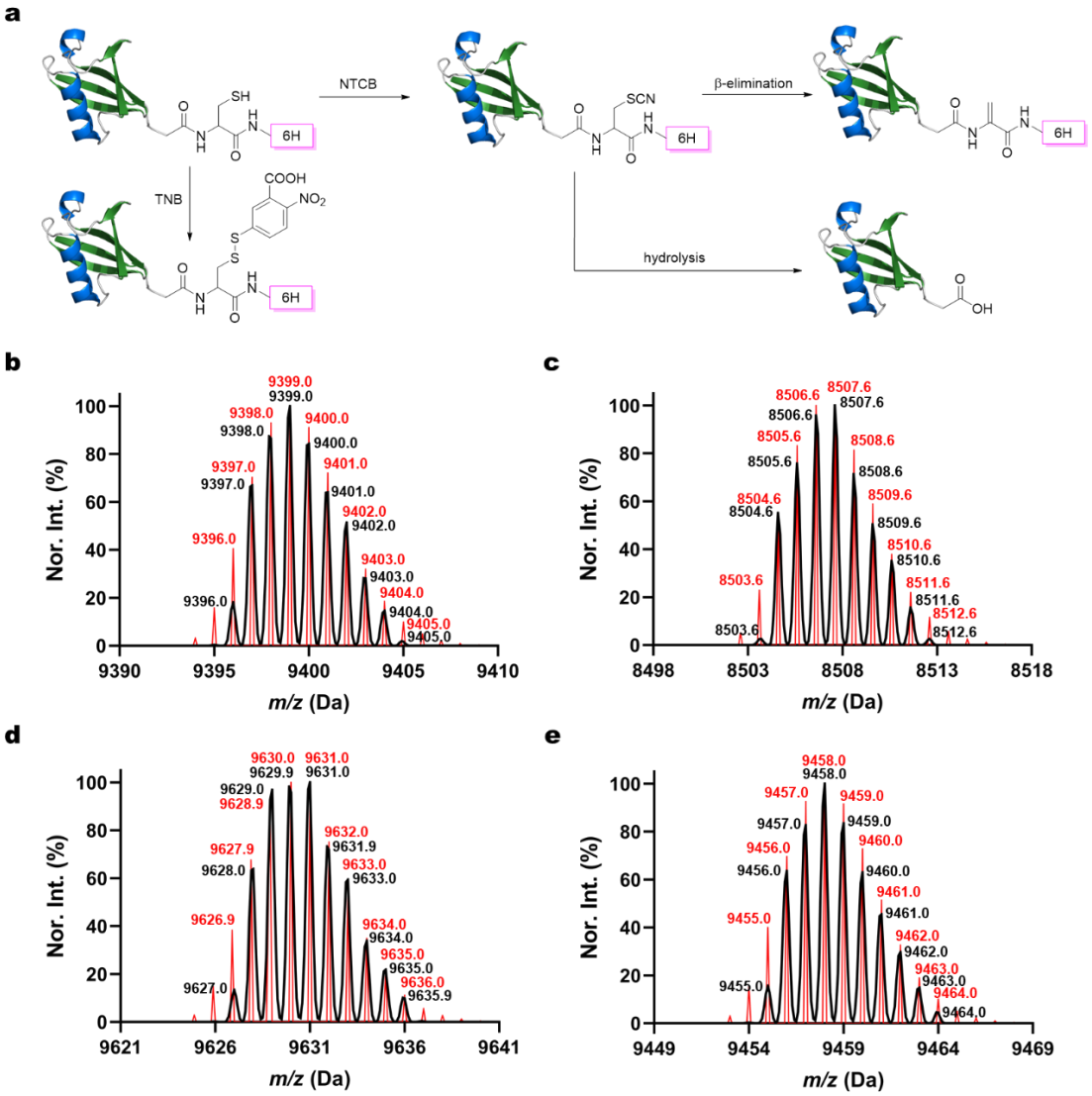

Figure 49 NTCB induced Dha formation using Ub-G76C-6H as substrate. (a) Pathways for the formation of four different possible products when Ub-G76C-6H reacts NTCB. Deconvoluted ESI-MS spectra of (b) the Dha product Ub-G76Dha-6H, (c) the hydrolysis byproduct $Ub_{(1-75)}$, (d) the cyanylated intermediate, Ub-G76C(CN)-6H and (e) the NTB conjugate Ub-G76C(NTB)-6H. Black lines show the detected monoisotopic mass peaks and their relative intensities. Red lines refer to the calculated theoretical monoisotopic peaks and their relative intensities. (b), (c) were from the LC-ESI-MS analysis of reaction products between Ub-G76C-6H and NTCB in the presence of 1× PBS, pH 9 under 37 °C; (d) (e) were from the LC-ESI-MS analysis of reaction products between Ub-G76C-6H and NTCB in the presence of 1× PBS, pH 7 under 5 °C.

Table 6 Quantitation of NTCB induced Dha formation of Ub-G76C-6H.

Reaction condition			**% yield**		
Buffer	**pH**	**Temp/**°C	**Dha**	**Hydro**	**+TNB**
1× PBS	9	37	79.9	20.1	0
1× PBS	8	37	84.4	15.6	0
1× PBS	7	37	89.3	10.7	0.9
1× PBS	6.5	37	89.5	10.5	0
1× PBS	7	23	61.6	4.8	1.5
1× PBS	7	5	12.2	0.6	10.2
20 mM HEPES, 10 mM Py, 6 M GndCl	7	37	97.8	2.2	0
20 mM HEPES, 10 mM Py	7	37	90.2	9.8	0
20 mM HEPES, 10 mM Py, 20% DMSO	7	37	82.9	7.1	2.3
20 mM HEPES, 10 mM Py, 40% DMSO	7	37	78.3	4.7	3.5
20 mM HEPES, 10 mM Py, 60% DMSO	7	37	76.9	1.8	5.0
20 mM HEPES, 10 mM Py, 80% DMSO	7	37	73.1	0.4	3.5
20 mM HEPES, 10 mM Py, 100% DMSO	7	37	70.5	0	1.9

All reactions contained 200 μM Ub-G76C-6H, 0.5 mM TCEP and 5 mM NTCB. For the conditions without a specifying DMSO percentage, the solvent that was used was pure water.

the TNB adduct. Observations we made so far indicated that the decreased hydrolysis did not result from better selectivity but was due to incomplete reactions. Furthermore, reactions were compared under both denatured and native conditions at 37 °C. To improve the beta elimination process, 10 mM pyridine (Py) was also included in the reaction as a non-nucleophilic base. The results showed a better yield in the presence of 10 mM Py under both denatured (6 M guanidine hydrochloride (GndCl)) and native conditions compared to the same reaction conditions without the addition of 10 mM Py (Table 6), confirming that Py improves the beta elimination process. We also observed slightly more Dha product formed under denaturing conditions than under native conditions possibly due to that cysteine was less restrained under denatured conditions (Table 6, Figure 135 and 136). It is known that the use of organic solvents will curb hydrolysis. For this reason, we explored how solvent composition influences Dha formation. We made solvent mixtures with different ratios of DMSO to water and performed the Dha formation in the presence of 5 mM NTCB, 0.5 mM TCEP and 10 mM Py in these solvent mixtures at pH 7 and 37 °C overnight. ESI-MS analysis of reaction products showed that the DMSO percentage increase correlated with the hydrolysis product decrease (Figure 135, 137-141) as we expected. However, the addition of DMSO led to a significant amount of the cyanylated Ub-G76C-6H intermediate not converted to the Dha-containing product (Figure 135, 137-141). This was to the contrary of the observation under pure aqueous conditions that typically led to complete conversion of the intermediate. The actual yield of the Dha production formation in the presence of DMSO kept decreasing when the DMSO percentage

increased (Table 6). Moreover, a high DMSO percentage also resulted in low protein solubility that made the reaction difficult to perform. Collectively, our results confirmed that the addition of an organic solvent does not improve NTCB-triggered conversion from cysteine to Dha in a protein. Therefore, we concluded that using denatured conditions at pH 7, as well as providing 10 mM pyridine, 0.5 mM TCEP, 5 mM NTCB for a reaction under 37 °C overnight will provide an effective approach to convert a protein cysteine to Dha.

3.2.3. Use of NTCB triggered Dha formation to synthesize Ubl-Dha probes

In eukaryotic cells, proteins can be post-translationally modified by Ub or Ubl proteins to undergo proteasome degradation or serve as function regulators for cellular processes.[169,205-208] To study pathways involving Ub and Ubls, diverse Ub and Ubl probes have been synthesized and used for various research purposes.[109,209-211] Dha probes which contain an α, β-unsaturated carbonyl structure at the *C*-terminus of Ub and Ubl proteins are ideal in the formation of covalent adducts with a catalytic cysteine in many enzymes in Ub and Ubl pathways. Mono-Ub, di-Ub, SUMO2, LC3 and NEDD8 based Dha probes have all been successfully synthesized by either intein-based protein semi-synthesis or total synthesis.[69,72-73,75,212-214] To simplify the synthesis of these Ub/Ubl-Dha probes, we explored the use of NTCB-triggered Dha formation from cysteine for their generation. We expressed a series of recombinant *N*-terminal FLAG tagged and *C*-terminal 6×His tagged Ubl proteins (FLAG-Ubl-GxC-6H) including NEDD8, MNSFβ, GABARAP, GABARAPL2, UFM1, URM1, ISG15 and

SUMO1/2/3/4 (Figure 50a). Therein, SUMO1/2/3/4, ISG15 and MNSFβ natively contain a cysteine residue in their sequences. To avoid unexpected modification at these cysteines, they were mutated to alanine or serine. ESI-MS analysis of purified proteins showed their expected molecular weights (Figure 50b). These proteins were then used to undergo the conversion of the installed cysteine at the *C*-terminal glycine position to Dha using the aforementioned optimized conditions. All synthesized Dha-containing products (FLAG-Ubl-GxDha-6H) displayed expected molecular weights (Figure 50c). Since the only type of byproducts that we detected was a hydrolysis species (Figure 142-152) in which the *C*-terminal 6×His tag was cleaved, all Dha products were easily accumulated and purified by Ni-NTA resin after reactions. By comparing to the intein-based approaches and protein total synthesis, our method is exceedingly simple.

3.2.4. Position specificity of NTCB triggered Dha formation

So far, all synthesized Dha-containing proteins have Dha at their flexible *C*-terminal ends. To explore whether our NTCB-triggered Dha formation from cysteine would also work with a cysteine in an internal region of a protein, we mutated seven lysine residues (K6, K11, K27, K29, K33, K48, K63) separately in Ub, expressed the afforded seven Ub mutants in *E. coli*, and then used them to undergo NTCB-triggered Dha formation. All seven expressed proteins had molecular weights matching perfectly with their theoretic values (Figure 51a, 153-159). However, reaction products varied according to different cysteine mutants. Surprisingly, ESI-LC-MS analysis showed almost complete hydrolysis with only a trace amount of Dha formation for the reaction

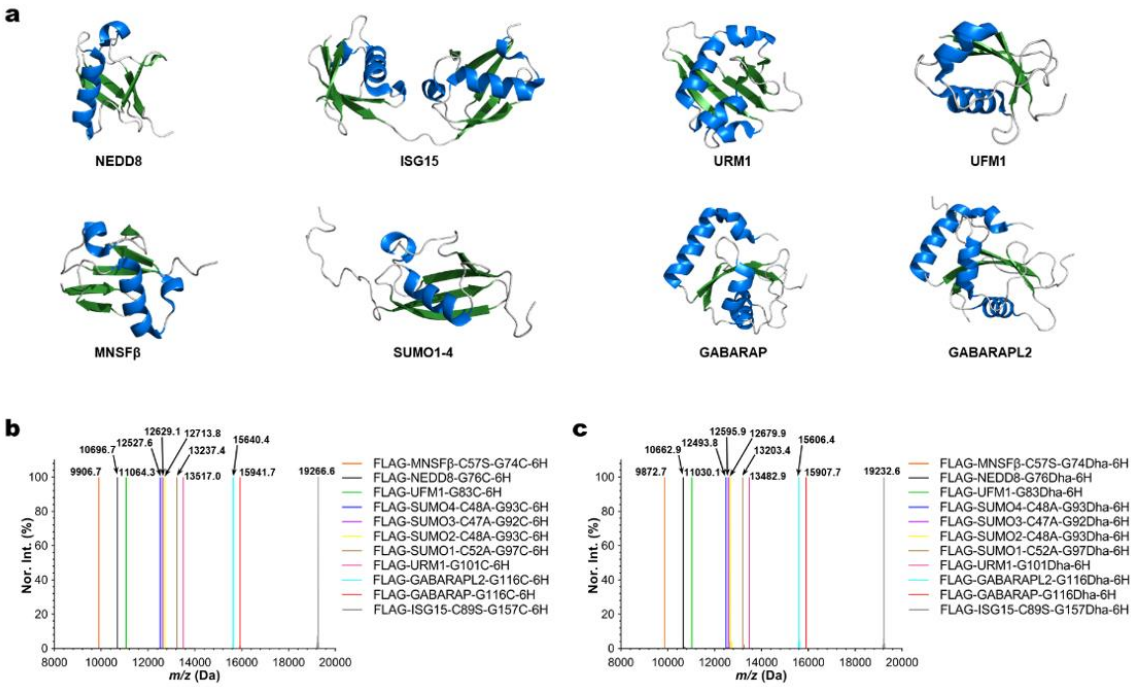

Figure 50 The synthesis of FLAG-Ubl-GxDha-6H probes by NTCB-triggered Dha formation from cysteine. (a) The structures of different Ubls used. (b) The deconvoluted and integrated ESI-MS spectra of 11 recombinant FLAG-Ubl-GxC-6H proteins. (c) The deconvoluted and integrated ESI-MS spectra of 11 synthesized FLAG-Ubl-Dha-6H probes. All detected molecular weights agreed well with theoretical values in a deviation range of ± 0.4 Da.

between Ub-K63C and NTCB (Figure 51b, 160 and 161). For all other six mutants, the detected predominant species had a mass of 8564.6 Da that matched their corresponding cyanylated intermediates (Figure 51c, 162, 164, 166, 168, 170, and 172). The reaction seemed to be stalled after the cyanylation process. When we switched to use a native condition, the Dha product was still not visible. The only significantly increased peak

was as the TNB adduct (Figure 55d, 163, 165, 167, 169, 171 and 173), which is likely due to more oxidized TNB dimer formed in a native condition. The reaction was also not obviously improved when we increased the incubating time or temperature (data not shown). Since Ub is a highly stable protein with a melting temperature higher than 100 °C, it is likely that all seven mutants do not fully denature in 6 M GndCl to certain extents. Local structural constraints of the introduced cysteine mutations might have

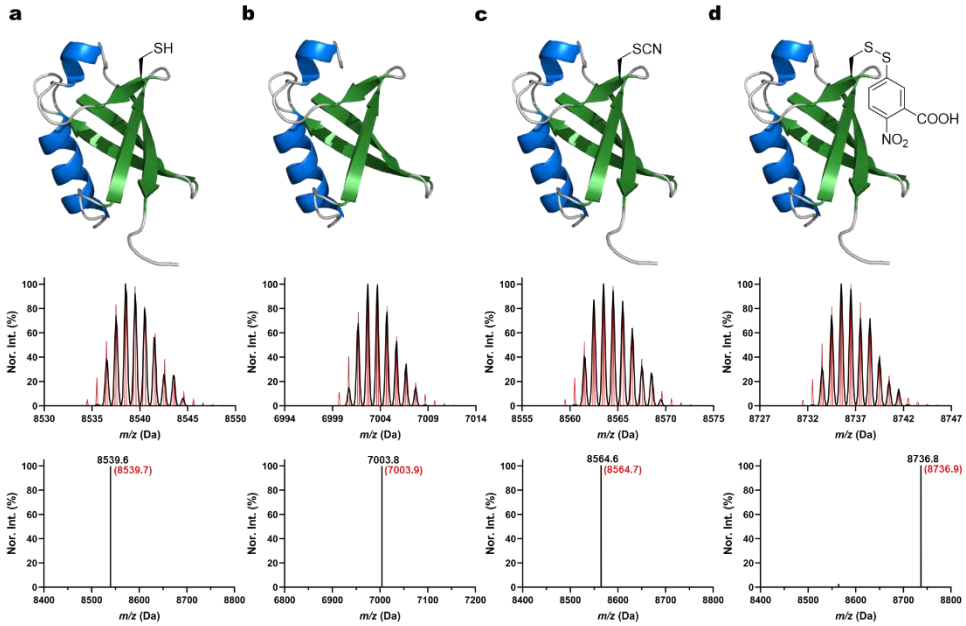

Figure 51 Examples for seven lysine to cysteine mutants of Ub, structures and their corresponding deconvoluted and integrated ESI-MS spectra after NTCB-triggered Dha Formation reaction. (a) Ub-K33C, (b) the hydrolysis product of Ub-K63C, $Ub_{(1-62)}$, (c) the cyanylation intermediate of Ub-K33C and (d) the TNB adduct of Ub-K33C. In deconvoluted MS spectra, black lines show the detected monoisotopic mass peaks and their relative intensities. Red lines refer to the calculated theoretical monoisotopic peaks

and their relative intensity. In integrated MS, experimental and theoretical values are labeled as black and red respectively.

prevented the beta elimination process. More investigations are needed to conclude that NTCB-triggered Dha formation does not work for an internal cysteine in a protein. Nevertheless, all our current data support that NTCB-triggered Dha formation from cysteine is highly efficient when the cysteine is placed at the *C*-terminus of a protein.

3.3. Discussion

Although the NTCB induced Dha formation approach is only suitable for a type of specific protein sequence, the phenomenon of our results based on Ub K to C mutants can still lead us to find more evidence to further optimize the ACPL approach. In principle, the transcyanylation step should be able to undergo either nucleophilic substitution or β-elimination reaction quickly. However, even after we purified the +CN products formed between Ub-K6/11/27/29/33/48C mutants and NTCB, and followed by incubating with zinc chloride solution which is able to form a strong ligand with HSCN to force the reaction equilibrium towards the β-elimination, such cyanylated species still maintained stable. We also thought of the effect caused by 6×His tag right on the C-terminal of reactive cysteine residue, so up to 200 mM imidazole was incubated with +CN product overnight as a substitution to provide an environment which mimic the 6×His tag for the protein. Unfortunately, there was still no obvious production of Ub-KxDha. One possible explanation could be the formation of a chemically stable and mass-neutral rearrangement formed after the transcyanylating step. Our previous

reported mechanism for ACPL approach contains an intramolecular cyclization step to forms the key 5-membered ring intermediate.[201] In a similar fashion, an analogical cyclization reaction may also occur between cysteine *C*-amide and protein thiocyanate (Figure 52a). The acquired product is so stable due to its 6-membered ring structure and lacking stable leaving group like 2-imonothiazolidine-4-carbonyl peptide formed after acyl substitution reaction that no following reactions are able to undergo. Despite the successful hydrazine conjugation reaction on Ub-K48C had proved the reactivity of ACPL approach when small amine applied,[201] at the absence of strong nucleophile which drive the consumption of *N*-side cyclized intermediate and facilitate the thermodynamic equilibrium towards *N*-amide rearrangement, the unreactive *C*-side cyclization species kept accumulating and eventually terminated the reaction.

There are several possibilities that may cause the variation of cyclization preference on either *N*-side or *C*-side. In the beginning, we supposed that protein secondary structure around the cysteine residue might regulate the cysteine accessibility. However, it has been overturned due to the fact that the denatured condition in which the folding of Ub-KxC proteins was fully destroyed could only promote the formation of +CN product.

The second possible enhancer for the reactions happened after cyanylation is 6×His tag. Although the high concentration imidazole assay had been proved to be not helpful at all (Figure 174), we believed that simulated reaction environment could not completely restore the same effect resulting from the 6×His tag right attached to the *C*-terminal of cysteine residue. Under basic conditions, the adjacent histidine residue could be able to serve as a base to subtract the α-hydrogen on cysteine residue due to its proximity

(Figure 52b). This could explain the preference to obtain Dha product while using protein containing *C*-terminal cysteine and 6×His tag.

Another parameter which may lead to the different reactivity we discovered is the sequence construction. It has been reported that the difference of the amino acid side chain right at the *N*-terminus of cysteine residue did play an important role in the yield of ACPL reaction. In general, most amino acids containing large side chains prevent the access of nucleophiles approaching the acyl group. In addition to that, comparing the yield difference of the ACPL reaction performed on Ub-G75X-G76C-6H models, we figured out that amino acids including a leucine-type side chain structure (D/N/W/Y/F) revealed strongest steric hindrance which exceedingly hampered the deprotonation of cysteine *N*-amide and resulted in poor reactivity for ACPL approach. Such steric effect impeded the 5-membered ring intermediate formation and extraordinarily retarded the nucleophilic substitution pathway to achieve ligation products. Since the proposed formation of *C*-side cyclization product also involves amide deprotonation, it might experience a similar manner. For protein like Ub-G76C-6H which has glycine as the cysteine *N*-terminal amino acid, the *C*-terminal histidine side chain is large enough to avoid *C*-amide rearrangement. In contrast, for proteins containing large *N*- amino acid and small *C*- amino acid such as Ub-K27C which has valine and alanine respectively, without providing strong nucleophiles, the *C*-amide rearrangement becomes inevitable.

Even so, except Ub-K27C, all other five Ub K to C mutants which revealed similar product selectivity contain a larger amino acid at *N*-terminus of cysteine rather than *C*-terminus. To explain their disability to form either Dha or hydrolysis product,

chemical property of amino acids right at the *N*-terminus of cysteine residue should be considered as a key factor. We assumed that amino acids which produce larger electron

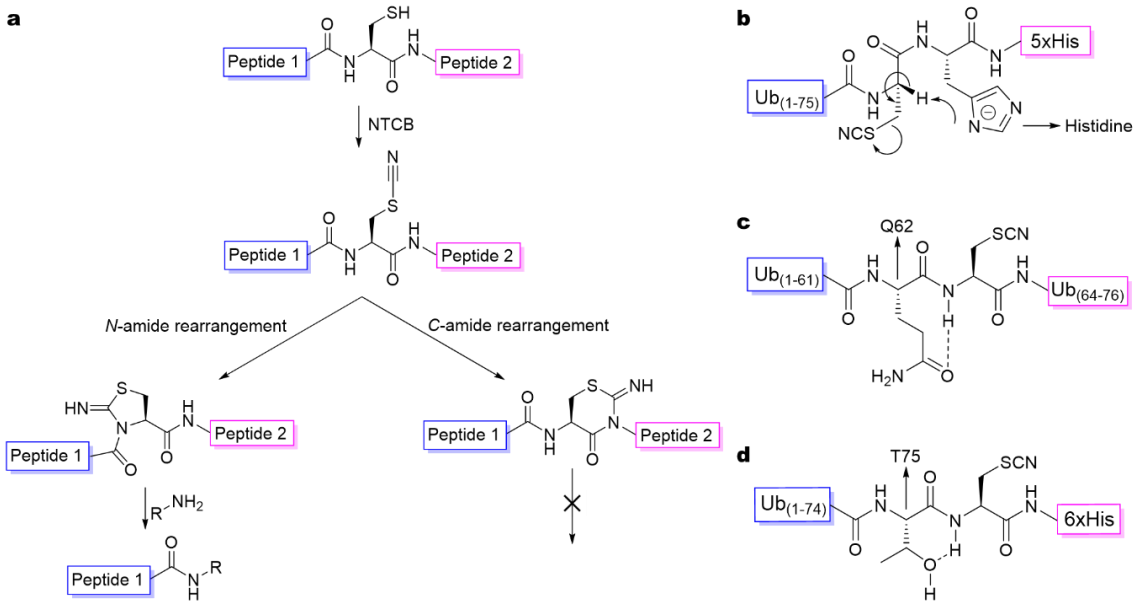

Figure 52 Proposed explanations for reactivity difference at each cysteine position. (a) The scheme showing the formation of 1-acyl-2-iminothiazolidine intermediate that activates the cysteine *N*-amide for ACPL ligation via *N*-amide rearrangement and a proposed analogical *C*-amide rearrangement which is inert to undergo following reactions; (b) Mechanism showing histidine assisted NTCB induced Dha product formation reaction; (a) The structure showing the hydrogen bonding that assists the *N*-amide rearrangement on Ub-K63C; (d) The structure showing the hydrogen bonding that assists the *N*-amide rearrangement on Ub-G75T-G76C-6H.

donating effects obstruct the *C*-side cyclization reaction and vice versa. By comparing the sequences of all Ub mutants we tested before, most of them have an electron pulling *C*- amino acid (E/T/Q) to weaken the cysteine α-hydrogen. For Ub-K29C, although the direct *C*- amino acid is an aliphatic amino acid, isoleucine, the amino acids right after that isoleucine (QDK) all contain an electron negative side chain which still led to the easiness of *C*-amide rearrangement. For Ub-K63C, its robust hydrolysis reactivity can be explained by the hydrogen bond formed between its cysteine *N*-glutamine and *N*-amide hydrogen (Figure 52c). The promotion effect on the subtraction of *N*-amide hydrogen resulting from such hydrogen bond formation beat the obstacle caused by *C*-side amino acid and eventually made Ub-K63C super reactive for hydrolysis. A similar result was also observed as the increase of ACPL ligation yield while using Ub-G76T-G76C-6H as the reaction substrate compared to other Ub G75 mutants (Figure 52d). On the other hand, if the cysteine *N*-amino acid has the ability to stabilize the *N*-amide hydrogen, it could further prevent the product formation after the cysteine cyanylation.

3.4. Conclusion

NTCB is a protein cyanylating reagent we used previously for the development of ACPL, a protein ligation technique. By reprograming NTCB for the generation of Dha from cysteine, we optimized the reaction condition in the presence of a denaturant and a non-nucleophilic base at pH 7. Using a Ub with a cysteine at its flexible *C*-terminal side, we were able to obtain a Dha-containing product with a yield of close to completion. Applying this same condition to a number of Ubl proteins with a cysteine

installed at their flexible *C*-terminal sides also resulted in their corresponding Dha-containing derivatives. Ub/Ubl-Dha probes are useful in proteomic analysis of cysteine-containing enzymes functioning in Ub and Ubl pathways. Our method greatly simplifies their synthesis. Its broad adoption for the study of Ub and Ubl pathways is expected.

CHAPTER IV

GENERAL EXPERIMENTAL DETAILS

4.1. General materials

Isopropyl β-D-1-thiogalactopyranoside (IPTG) was bought from INDOFINE Chemical Company Inc. (Hillsborough, NJ, USA). Sodium phosphate monobasic, sodium chloride, imidazole, pyridine, Tris and HEPES were all provided by VWR (Radnor, PA, USA). NTCB was purchased from TCI America (Portland, OR, USA). TCEP was obtained from Alfa Aesar (Tewksbury, MA, USA). LC/MS grade water and acetonitrile for liquid chromatography were both purchased from Fisher Scientific (Waltham, MA, USA) and supplied with 0.1% Optima™ LC/MS grade formic acid (Fisher Scientific, Waltham, MA, USA). All oligonucleotide primers and DNA sequences used for DNA mutagenesis were synthesized by Integrated DNA Technologies (Coralville, IA, USA). Sanger sequencing was performed by Eton Bioscience Inc. (San Diego, CA, USA).

4.2. Compound synthesis

4.2.1. Synthesis of dipeptides

Boc-Cystine-OMe(1a)

To a suspension of Boc-Cystine-OH (1 eq.) in dry methanol (20 mL per 0.8 g) thionyl chloride (2 eq.) was added slowly at room temperature. Temperature was increased to 80 °C over 15 min. Reaction was stirred for 3 h then left to cool down and

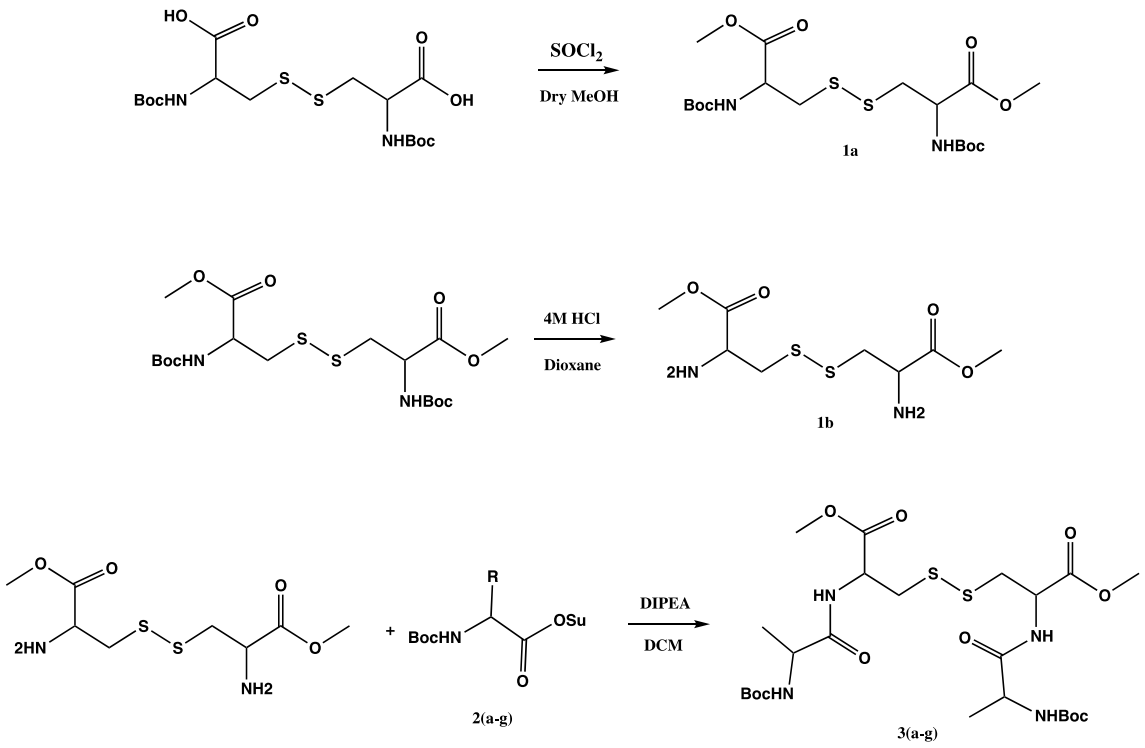

Figure 53 Synthetic route for dipeptides.

stirred overnight at room temperature. Volatiles were evaporated in vacuo and the residue was dissolved in minimal amount of methanol. Crystals of crude product crashed out from solution by addition of diethyl ether. Product was filtered, washed twice with diethyl ether and dried to obtain **1a** as white crystals. Final yield was 90%. ^{1}H NMR ($CDCl_3$) δ =1.45 (s, 18 H), 3.16 (d, 4 H, J = 5.0 Hz), 3.76 (s, 6 H), 4.60 (dt, 2 H, J = 6.3, 5.0 Hz), 5.38 (d, 2H, J = 6.3 Hz).

Cystine-OMe(1b)

A solution of **1a** (1 eq.) in DCM was cooled to 0 °C. 4 M HCl in Dioxane (1.1 eq.) was added dropwise. Solution was stirred for 2 h and the reaction progress was

monitored by TLC. Solvent was removed in vacuo and product was obtained as white solid without further purification. Quantitative yield. ^{1}H NMR ($CDCl_3$) δ = 3.16 (d, 4 H, J = 5.0 Hz), 3.76 (s, 6 H), 4.60 (dt, 2 H, J = 6.3, 5.0 Hz), 5.38 (d, 2H, J = 6.3 Hz) 7.80 (s, 1H).

Gly-Cystine-OMe(3a)

A solution of Boc-Gly-Osu **2a** (2 eq.) in dry DCM was added into a mixture of **1b** (1 eq.) and DIPEA (2.2 eq.) at 0 °C and stirred for 15-30 min. The solution was then warmed to room temperature and left to stir overnight. The solution was washed with 3 M HCl twice then washed with brine twice. Organic layer was separated, dried over sodium sulfate, and concentrated under reduced pressure. The residue was purified by column chromatography (30% EtOAc/hexane). ^{1}H NMR ($CDCl_3$) δ = 1.42 (s, 18H), 3.16 (d, 4 H, J = 5.0 Hz), 3.76 (s, 6 H), 3.88 (s, 2H) 4.60 (t, 2 H) 7.80 (s, 1H).

Ala-Cystine-OMe (3b)

A solution of Boc-Ala-Osu **2b** (2 eq.) in dry DCM was added dropwise into a mixture of **1b** (1 eq.) and DIPEA (2.2 eq.) at 0 °C and stirred for 15-30 min. The solution was then warmed to room temperature and left to stir overnight. The solution was washed with 3 M HCl twice then washed with brine twice. Organic layer was separated, dried over sodium sulfate, and concentrated under reduced pressure. The residue was purified by column chromatography (30% EtOAc/hexane). ^{1}H NMR

($CDCl_3$) δ = 0.98 (d, 6H, J = 6.8 Hz), 1.42 (s, 18H) 3.16 (d, 4 H, J = 5.0 Hz), 3.76 (s, 6 H), 4.60 (t, 2 H, J = 6.3, 5.0 Hz) 7.80 (s, 1H).

Leu-Cystine-OMe (3c)

A solution of Boc-Leu-Osu **2c** (2 eq.) in dry DCM was added dropwise into a mixture of **1b** (1 eq.) and DIPEA (2.2 eq.) at 0 °C and stirred for 15-30 min. The solution was then warmed to room temperature and left to stir overnight. The solution was washed with 3 M HCl twice then washed with brine twice. Organic layer was separated, dried over sodium sulfate, and concentrated under reduced pressure. The residue was purified by column chromatography (30% EtOAc/hexane). ^{1}H NMR ($CDCl_3$) δ = 0.90 (d, 12H, J = 6.8 Hz), 1.42 (s, 18H), 1.49 (q, 2H), 1.76 (t, 4H), 3.16 (d, 4H, J = 5.0 Hz), 3.76 (s, 6 H), 4.40 (t, 2H), 4.60 (t, 2H, J = 6.3, 5.0 Hz), 7.80 (s, 1H).

Phe-Cystine-OMe (3d)

A solution of Boc-Phe-Osu **2d** (2 eq.) in dry DCM was dropped into a mixture of **1b** (1 eq.) and DIPEA (2.2 eq.) at 0 °C and stirred for 15-30 min. The solution was then warmed to room temperature and left to stir overnight. The solution was washed with 3 M HCl twice then washed with brine twice. Organic layer was separated, dried over sodium sulfate, and concentrated under reduced pressure. The residue was purified by column chromatography (30% EtOAc/hexane). ^{1}H NMR ($CDCl_3$) δ = 1.42 (s, 18H), 3.16 (d, 2H, J = 5.0 Hz), 3.40 (q, 2H, J = 7.0 Hz) 3.76 (s, 6 H), 4.40 (t, 2H), 4.60 (t, 2H, J = 6.3, 5.0 Hz), 7.14 (m, 10H), 7.80 (s, 1H).

Trp-Cystine-OMe (3e)

A solution of Boc-Trp-Osu **2e** (2 eq.) in dry DCM was added dropwise into a mixture of **1b** (1 eq.) and DIPEA (2.2 eq.) at 0 °C and stirred for 15-30 min. The solution was then warmed to room temperature and left to stir overnight. The solution was washed with 3 M HCl twice then washed with brine twice. Organic layer was separated, dried over sodium sulfat,e and concentrated under reduced pressure. The residue was purified by column chromatography (30% EtOAc/hexane). ^{1}H NMR ($CDCl_3$) δ = 1.42 (s, 18H), 3.16 (d, 2H, J = 5.0 Hz), 3.40 (q, 2H, J = 7.0 Hz) 3.76 (s, 6 H), 4.40 (t, 2H), 4.60 (t, 2H, J = 6.3, 5.0 Hz), 6.9 (t, 1H), 7.14 (t, 1H), 7.3 (t, 1H), 7.60 (t, 1H), 7.90 (s, 1H).

Asp-Cystine-OMe (3f)

A solution of Boc-Asp-Osu **2f** (2 eq.) in dry DCM was added dropwise into a mixture of **1b** (1 eq.) and DIPEA (2.2 eq.) at 0 °C and stirred for 15-30 min. The solution was then warmed to room temperature and left to stir overnight. The solution was washed with 3 M HCl twice then washed with brine twice. Organic layer was separated, dried over sodium sulfate, and concentrated under reduced pressure. The residue was purified by column chromatography (30% EtOAc/hexane). ^{1}H NMR ($CDCl_3$) δ = 1.42 (s, 18H), 2.79 (q, 4H, J = 7.0 Hz) 3.16 (d, 2H, J = 5.0 Hz), 3.40 (q, 2H, J = 7.0 Hz) 3.76 (s, 6 H), 4.40 (t, 2H), 4.60 (t, 2H, J = 6.3, 5.0 Hz), 4.90 (t, 1H), 7.90 (s, 1H), 11.20 (s, 1H).

Pro-Cystine-OMe (3g)

A solution of Boc-Asp-Osu **2g** (2 eq.) in dry DCM was added into a mixture of **1b** (1 eq.) and DIPEA (2.2 eq.) at 0 °C and stirred for 15-30 min. The solution was then warmed to room temperature and left to stir overnight. The solution was washed with 3 M HCl twice then washed with brine twice. Organic layer was separated, dried over sodium sulfate, and concentrated under reduced pressure. The residue was purified by column chromatography (30% EtOAc/hexane). ^{1}H NMR ($CDCl_3$) δ = 1.42 (s, 18H), 2.13 (m. 8H) 3.16 (m, 4H), 3.40 (m, 4H) 3.76 (s, 6 H), 4.26 (br, 2H), 4.78 (br, 2H), 6.90 (br, 1H), 7.90 (s, 1H).

4.2.2. Dipeptide reactions with NTCB and propargylamine

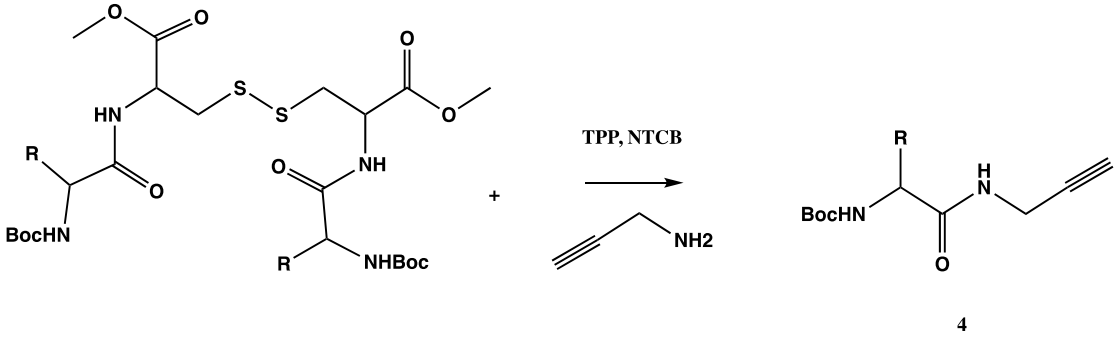

Figure 54 Conjugation reaction between dipeptides and propargylamine using NTCB.

Boc-Gly-Propargylamine (4a)

Boc-Gly-Cys-OMe **3a** (1 eq.) and TPP (1.1 eq.) were mixed in DCM at room temperature for 40 min or until reaction completion by monitoring with TLC. After

completion NTCB (2 eq.) was added at room temperature and left to stir for 1 h. After the formation of the thiazolidine intermediate, confirmed by TLC, propargylamine was added (1.1 eq.) and left to stir overnight. The reaction was washed in 1 M HCl thrice and brine thrice, the organic layer was separated, dried over sodium sulfate, and concentrated in vacuo. Crude residue was purified by column chromatography with a 0-20% EtOAc/hexane gradient. ^{1}H NMR ($CDCl_3$) δ = 7.29 (s, 1H), 5.41 (s, 1H), 4.08 (dd, 2H, *J* = 5.4 Hz, 2.7 Hz), 3.84 (d, 2H, *J* = 5.7 Hz), 2.25 (t, 1H, *J* = 2.7 Hz), 1.47 (s, 9H).

Boc-Ala-Propargylamine(4b)

Boc-Ala-Cys-OMe **3b** (1 eq.) and TPP (1.1 eq.) were mixed in DCM at room temperature for 40 min or until reaction completion by monitoring with TLC. After completion NTCB (2 eq.) was added at room temperature and left to stir for 1 h. After the formation of the thiazolidine intermediate, confirmed by TLC, propargylamine was added (1.1 eq.) and left to stir overnight. The reaction was washed in 1 M HCl thrice and brine thrice, the organic layer was separated, dried over sodium sulfate, and concentrated in vacuo. Crude residue was purified by column chromatography with a 0-20% EtOAc/hexane gradient. ^{1}H NMR ($CDCl_3$) δ= 1.37 (3H, d, *J* = 6.8 Hz). 1.46 (s, 9H), 2.23 (t, 1H *J* = 2.4 Hz), 4.05 (m, 2H), 4.16 (m, 1H), 4.92 (s,1H), 6.43 (s,1H).

Boc-Leu-Propargylamine(4c)

Boc-Leu-Cys-OMe **3c** (1 eq.) and TPP (1.1 eq.) were mixed in DCM at room temperature for 40 min or until reaction completion by monitoring with TLC. After

completion NTCB (2 eq.) was added at room temperature and left to stir for 1 h. After the formation of the thiazolidine intermediate, confirmed by TLC, propargylamine was added (1.1 eq.) and left to stir overnight. The reaction was washed in 1 M HCl thrice and brine thrice, the organic layer was separated, dried over sodium sulfate, and concentrated in vacuo. Crude residue was purified by column chromatography with a 0-20% EtOAc/hexane gradient. ^{1}H NMR (300 MHz, CDCl3): δ= 0.90 (br d, 6H), 1.41 (s, 9H), 1.59 (m, 3H) 2.18 (s, 1H), 3.99 (s, 2H), 4.18 (br s, 1H), 5.25 (s, 1H), 7.13 (br s, 1H).

Boc-Phe-Propargylamine(4d)

Boc-Phe-Cys-OMe **3d** (1 eq.) and TPP (1.1 eq.) were mixed in DCM at room temperature for 40 min or until reaction completion by monitoring with TLC. After completion NTCB (2 eq.) was added at room temperature and left to stir for 1 h. After the formation of the thiazolidine intermediate, confirmed by TLC, propargylamine was added (1.1 eq.) and left to stir overnight. The reaction was washed in 1 M HCl thrice and brine thrice, the organic layer was separated, dried over sodium sulfate, and concentrated in vacuo. Crude residue was purified by column chromatography with a 0-20% EtOAc/hexane gradient. ^{1}H NMR ($CDCl_3$) δ = 1.37 (s, 9H), 2.18 (s, 1H), 2.98 (m, 1H), 3.06 (m, 1H), 3.95 (s, 2H), 4.47 (m, 1H), 5.49 (d, J = 9.0 Hz, 1H), 7.04 (br s, 1H), 7.17-7.27 (m, 5H).

Boc-Trp-Propargylamine(4e)

Boc-Trp-Cys-OMe **3e** (1 eq.) and TPP (1.1 eq.) were mixed in DCM at room temperature for 40 min or until reaction completion by monitoring with TLC. After completion NTCB (2 eq.) was added at room temperature and left to stir for 1 h. After the formation of the thiazolidine intermediate, confirmed by TLC, propargylamine was added (1.1 eq.) and left to stir overnight. The reaction was washed in 1 M HCl thrice and brine thrice, the organic layer was separated, dried over sodium sulfate, and concentrated in vacuo. Crude residue was purified by column chromatography with a 0-20% EtOAc/hexane gradient. ^{1}H NMR ($CDCl_3$) δ =1.41 (s, 9H), 1.86 (m, 2H), 2.19 (m, 2H), 3.28-3.41 (m, 2H), 3.98-4.22 (m, 3H), 6.38 (br s, 1H), 7.23 (br s, 1H).

Boc-Asp-Propargylamine(4f)

Boc-Asp-Cys-OMe **3f** (1 eq.) and TPP (1.1 eq.) were mixed in DCM at room temperature for 40 min or until reaction completion by monitoring with TLC. After completion NTCB (2 eq.) was added at room temperature and left to stir for 1 h. After the formation of the thiazolidine intermediate, confirmed by TLC, propargylamine was added (1.1 eq.) and left to stir overnight. The reaction was washed in 1 M HCl thrice and brine thrice, the organic layer was separated, dried over sodium sulfate, and concentrated in vacuo. Crude residue was purified by column chromatography with a 0-20% EtOAc/hexane gradient. ^{1}H NMR ($CDCl_3$) δ = 1.37 (s, 9H), 2.23 (d, 2H, J = 2.4 Hz), 3.06 (m, 1H), 4.05 (m, 2H), 4.92 (t, 1H), 5.7 (br, 1H), 8.1 (br, 1H), 12.30 (br, 1H).

Pro-Propargylamine(4g)

Pro-Cys-OMe **3g** (1 eq.) and TPP (1.1 eq.) were mixed in DCM at room temperature for 40 min or until reaction completion by monitoring with TLC. After completion NTCB (2 eq.) was added at room temperature and left to stir for 1 h. After the formation of the thiazolidine intermediate, confirmed by TLC, propargylamine was added (1.1 eq.) and left to stir overnight. The reaction was washed in 1 M HCl thrice and brine thrice, the organic layer was separated, dried over sodium sulfate, and concentrated in vacuo. Crude residue was purified by column chromatography with a 0-20% EtOAc/hexane gradient. ^{1}H NMR ($CDCl_3$) δ =1.41 (s, 9H), 1.86 (m, 2H), 2.19 (m, 2H), 3.28-3.41 (m, 2H), 3.98-4.22 (m, 3H), 6.38 (br s, 1H), 7.23 (br s, 1H).

4.3. Plasmid construction

The gene coding Ub-G76C-6H and Ub-C-6H were inserted into the pETDuet-1-Ub vector[24] to afford the expression vector pETDuet-1-Ub-G76C-6H and pETDuet-1-Ub-C-6H. Constructs for Ub-K6/11/27/29/33/48/63C were obtained from lab cell stocks. Constructs for Ub G75P/T/L/E/R/W mutants were made by site-directed mutagenesis from the pETDuet-1-Ub-G76C-6H vector. The expression vector for FLAG-Ub-G76C-6H was generated by inserting a FLAG coding DNA at the *N*-terminal side of the Ub-G76C-6H coding sequence in the pETDuet-1-Ub-G76C-6H vector. The DNA fragments for all Ubls that contained both an *N*-terminal FLAG tag and a *C*-terminal 6×His tag were ordered from IDT Inc. and were cloned into the pETDuet-1 vector to afford expression vectors for them. The H2A-K129C-6H expression vector was constructed from the pETDuet-1-6H-TEV-H2A vector. For the $RNH_{59\text{-}196}$-K190C-6H expression

vector, the DNA fragment for *B. halodurans* RNase H (residues 59-196 and a K190C mutation) was cloned into the pET28a vector right before the sequence of a *C*-terminal 6×His tag using the MEGAWHOP technique. The gene fragment for exenatide-40C-GGGSA-Strep was cloned into the pET28a-SUMO vector right after the SUMO cleavage site using the MEGAWHOP technique. The pET28a-6H-SUMO-exenatide-S39C-SA-Strep was constructed by site-directed mutagenesis of the pET28a-6H-SUMO-exenatide-40C-GGGSA-Strep vector. pET28a-SUMO protease was kindly provided by Prof. Pingwei Li.

4.4. Recombinant protein expression and purification

4.4.1. Expression and purification of Ub proteins

The expression of all Ub mutants and FLAG-Ub were performed according to the protocol described before.[24] Briefly, overnight culture of each protein was inoculated into 1 L 2xYT medium and let grow under 37 °C until OD_{600} reached 0.6-1.0. 1 mM IPTG was added to induce the expression at 18 °C overnight. Upon saturation, cells were collected by centrifugation (6000 rpm, 20 min, 4 °C) and stored at -80 °C if not lysed directly.

For purification of each all seven Ub K to C mutant, the bacterial pellet was resuspended in Ub lysis buffer (50 mM Tris, 1 mM TCEP, pH 7.8) supplied with 0.2 mg/mL lysozyme (Sigma-Aldrich, St. Louis, MO, USA) and lysed by sonication on ice. The total cell lysate was clarified by centrifugation (10000 rpm, 30 min, 4 °C) and supernatant was collected. After that, 6 M HCl solution was gradually added into the

supernatant with constant stirring to adjust pH value to 1.5-2. The white precipitate was then removed by centrifugation (10000 rpm, 30 min, 4 °C) and the pH of the acid purification supernatant was adjusted back to 7.8 using 6 M NaOH and then concentrated using Amicon stirring filtration system with a 5k MWCO membrane (EMD Millipore, Burlington; MA, USA). Subsequently, each protein was desalted into 50 mM ammonium bicarbonate (ABC) buffer by HiPrep 26/10 Desalting column (GE Healthcare, Chicago, IL, USA) using NGC™ chromatography system (Bio-Rad Laboratories, Hercules, CA, USA) and analyzed by 15% SDS-PAGE. The concentration of protein solution was measured by Pierce™ 660nm Protein Assay Reagent (Thermo Fisher Scientific, Waltham, MA, USA) and dispensed into 100 nmol aliquots for lyophilization. Eventually, protein pellets were either used for chemical reactions or kept at -80 °C for long-term storage.

For purification of all his-tagged Ub mutants, Ni binding buffer (50 mM NaH_2PO_4, 500 mM NaCl, 5 mM imidazole, 1 mM TCEP, pH 7.8) was used instead of Ub lysis buffer to resuspend and lyse the cell pellet. The subsequent steps were identical to the purification of Ub-KxC until finishing the acid purification. Acid supernatant was directly passed through high affinity Ni charged resin (Genescript, Piscataway, NJ, USA) without concentration. Resin was then washed by Ni washing buffer (50 mM NaH_2PO_4, 500 mM NaCl, 25 mM imidazole, 1 mM TCEP, pH 7.8) and eluted in 7 mL elution buffer (50 mM NaH_2PO_4, 500 mM NaCl, 300 mM imidazole, 1 mM TCEP, pH 7.8). The following desalting, SDS-PAGE, concentration measurement, lyophilization and storage as described above.

4.4.2. Expression and purification of FLAG-Ubl proteins

An overnight culture of *E. coli* BL21(DE3) cells harboring a Ubl expression vector was inoculated (1:100 dilution) into LB medium containing 100 µg/mL ampicillin. Cells were let grow at 37 ºC until OD_{600} reached 0.6-0.9 and then 1 mM IPTG was added to induce protein expression. Induced cells were let grow in a refrigerated incubating shaker at 18 ºC overnight and harvested by centrifugation (4000 rpm, 20 min, 4 ºC). The cell pellet was then lysed, purified by Ni^{2+}-NTA resins, and desalted as same as mentioned previously for his-tagged Ub proteins. Yield for all Ubl proteins were around 10-40 mg/L. Eventually, the proteins were aliquoted, lyophilized, and stored at -80 ºC.

4.4.3. Expression and purification of H2A-K129-6H

The expression vector pETDuet-1-H2A-K129C-6H was used to transform *E. coli* BL21(DE3) strain, and a single colony was inoculated into 2×YT medium containing 100 µg/mL ampicillin. After cell grew to OD_{600} about 0.6 at 37 ºC, 0.5 mM IPTG was added to induce protein expression. Four hours later, cells were harvested and purified according to a previous protocol.[215]

4.4.4. Expression and purification of RNH_{59-196}-K190C-6H

An overnight culture of *E. coli* BL21(DE3) cells which contained the pET28a-RNH_{59-196}-K190C-6H vector was inoculated (1:100 dilution) into 2×YT medium containing 50 µg/mL kanamycin. The protein was expressed by inducing with 0.5 mM

IPTG when OD_{600} reached 0.6 at 37 ºC for 3 h. The harvested cells were lysed in a buffer (20 mM Tris, 200 mM NaCl, pH 8.0) by sonication at 4 ºC, the lysate was clarified by centrifugation (10,000 rpm, 30 min, 4 ºC). Supernatant was collected and incubated with Ni^{2+}-NTA resins at rt for 45 min before washed with a 20× resin volume of lysis buffer supplemented with 35 mM imidazole. The expressed 6×His-tagged protein was eluted by another buffer (20 mM Tris, 200 mM NaCl, 300 mM imidazole, pH 8.0). Collected elution was concentrated and desalting to 50 mM ABC buffer by HiTrap Desalting columns (GE Healthcare). Proteins were aliquoted and lyophilized for further assays.

4.4.5. Expression and purification of SUMO protease

An overnight culture of *E. coli* BL21(DE3) cells harboring a pET28a-SUMO vector was inoculated (1:100 dilution) into LB medium containing 50 µg/mL kanamycin. Cells were let grown at 37 ºC until OD_{600} reached 1.3 and 1 mM IPTG was added to induce protein expression. Inducted cells were let grow in a refrigerated incubating shaker at 16 ºC overnight and harvested by centrifugation (4000 rpm, 20 min, 4 ºC). The cell pellet was resuspended, lysed, purified by Ni^{2+}-NTA resins, and desalted according to the same protocol for Ub proteins except the use of DTT instead of TCEP. The expression yield was about 4.1 mg/L. The finally purified protein was aliquoted into 0.2 mg aliquots and stored at -80 ºC for the future usage.

4.4.6. Expression and purification of exenatide

An overnight culture of *E. coli* BL21(DE3) cells harboring the 6H-SUMO-exenatide-S39C-SA-Strep expression vector was inoculated (1:100 dilution) into LB medium containing 50 µg/mL kanamycin. Cells were let grow at 37 ºC until OD_{600} reached 0.6-0.9 and then 1 mM IPTG was added to induce protein expression. Induced cells were let grow in a refrigerated incubating shaker at 18 ºC overnight and harvested by centrifugation (4000 rpm, 20 min, 4 ºC). The sonication with addition of one protease inhibitor cocktail tablet (Sigma-Aldrich), Ni^{2+}-NTA purification, and desalting were performed as same as what were described previously. After desalting, one aliquot of SUMO protease stock solution (0.2 mg) was added into the eluted protein solution along with 1 mM DTT and stirred at 4 ºC overnight. On the next day, the digested sample was loaded onto a HisTrap HP column (GE Healthcare) and the flow-through was collected. The exenatide-S39C-SA-Strep yield was determined as 11 mg/L. The final exenatide-S39C-SA-Strep solution was aliquoted and lyophilized into protein pellets and then stored at -80 ºC.

4.5. Chemical reactions on protein

4.5.1. ACPL of Ub and Ubl proteins

A 500 mM TCEP stock and NTCB stock were prepared in water and DMSO, respectively. For reactions with all Ub proteins, solutions of small amine-containing compounds (glycine, proline, aspartic acid, arginine, D-alanine, D-Serine, allylamine, and propargylamine) were prepared in water to a concentration of 1 M along with 6 M guanidinium chloride (GndCl). For asparagine, leucine, and phenylalanine, their

concentrations were 500, 150, and 100 mM respectively due to their poor solubility. Hydrazine solution was prepared as 50 mM in 1× PBS buffer along with 6 M GndCl. The pH of all solutions was pre-adjusted to 9 before setting up reactions. 40 mM Gly-AMC solution was prepared in 40% DMSO without pH adjustment. To set up reactions, Ub protein pellets were dissolved by the previously described amine-containing compound solutions and 0.5 mM TCEP and 5 mM NTCB were sequentially introduced into the reaction mixture. For the reaction with Gly-AMC, pH was adjusted to 8 after adding TCEP and NTCB prior to the incubation. The final concentration of proteins was 200 μM (2 mg/mL). All reaction mixtures were incubated at 37 ºC for 16 h and followed by desalting using HiTrap Desalting column (GE Healthcare, Chicago, IL, USA). Fractions with UV absorbance was collected and analyzed by 15% SDS-PAGE. Then, Ni^{2+}-NTA resins were incubated with the desalted products at rt for 30 min. After that, Ni^{2+}-NTA resins were removed by micro syringe filters and flowthroughs were collected for ESI-MS analysis. For reactions between UbK48/63C or Ub-G76C-6H and hydrazine without GndCl, the reaction setup was identical except that no GndCl was provided.

For all Ubl proteins, the reaction setup was identical as aforementioned except in the amine buffer preparation. In this case, no GndCl was added and the amine used was only 1 M propargylamine. The reaction between Gly-AMC and SUMO-1/2/3 was also initiated by mixing all reagents first and then adjusting pH to 8.

4.5.2. Synthesis of H2AK129ac

H2A-K129C-6H pellets were solubilized in reaction buffer (6 M GndCl, 1 M N^{ε}-acetyl-L-lysine). To the solution, 0.5 mM TCEP and 5 mM NTCB were sequentially added. After incubating the reaction at 37 ºC overnight, the reaction mixture was dialyzed to water for twice and then dialyzed to a Ni^{2+}-NTA pull-down buffer (6 M GndCl, 20 mM imidazole, pH 8.0). After that, Ni^{2+}-NTA resins were added and incubated at 4 ºC for 30 min on a rotator. Then Ni^{2+}-NTA resins were removed by a micro syringe filter and the supernatant was collected, dialyzed to water and analyzed by ESI-MS. The remaining product was lyophilized to pellets for further assays.

4.5.3. Synthesis of $RNH_{59\text{-}196}$-K190C

Lyophilized original protein was dissolved in an aqueous solution containing 6 M GndCl and 50 mM hydrazine (pH 9). The following setups and treatments were identically performed as mentioned for the synthesis of Ub ligation products. The flow-through was analyzed by ESI-MS and then lyophilized for further use.

50 μM $RNH_{59\text{-}189}$-Ha was dissolved into a buffer containing 6 M GndCl and 0.2 M NaH_2PO_4 (pH 3). Protein solution was pre-cooled to -20 ºC before adding 50 mM freshly made $NaNO_2$. The first step of the reaction occurred at -20 ºC for 30 min, followed by the addition of 20 mM 4-mercaptophenylacetic acid and 250 μM of the 7-mer NH_2-CADYGRK-OH peptide. pH was adjusted to 6.8 and the reaction was let undergo at 37 ºC for 16 h. The ligation product finally desalted and lyophilized for ESI-MS analysis and further use.

4.5.4. Synthesis of exenatide

Lyophilized exenatide-S39C-SA-Strep pellets were dissolved in an aqueous pH 9 buffer containing 1 M serinamide hydrochloride, 0.5 mM TCEP, and 5 mM NTCB. The reaction mixture was incubated under rt overnight. After the post-reaction desalting, a StrepTrap HP column (GE Healthcare) was used to capture the unreacted peptide. Flow-through was collected as the final product for ESI-MS analysis.

4.5.5. NTCB induced Dha formation reaction on Ub/Ubl proteins

The stock solution of 500 mM TCEP and 500 mM NTCB was prepared in water and DMSO respectively. For all the optimization reactions, each aliquot of 100 nmol Ub-G76C-6H was dissolved in a buffer mentioned in Table 1 and followed by adding 0.5 mM TCEP, 5 mM NTCB sequentially. The reactions were then incubated at 5/18/37 °C for 18 h based on the desired testing requirements. Later, each reaction mixture was desalted to 50 mM ABC buffer by HiTrap Desalting column using NGC™ chromatography system to quench the reaction. Fractions corresponding to the peak of UV signal without conductivity change were collected for ESI-LC-MS analysis.

For reactions on eleven FLAG-Ubl-G76C-6H, instead of changing buffer and temperature for each condition, a buffer containing 20 mM HEPES, 10 mM pyridine at pH 7 as well as 37 °C incubation were used constantly for all Ubl proteins. All other conditions were identical as mentioned above.

For reactions on seven Ub K to C mutants, besides one set of reaction set up using the same conditions for FLAG-Ubl proteins, another series of reaction using

denatured condition was performed. For that set, the reaction buffer was further supplied with 6 M GndCl (VWR, Radnor, PA, USA) as a denaturing agent. All other conditions were identical as mentioned above.

4.6. ESI-MS analysis and data processing

Samples for MS were placed in tube inserts inside HPLC sample vials for auto injection. The positive ESI-LC-MS was carried out using Q Exactive Orbitrap mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) connected to a liquid chromatography instrument. The settings for all ESI-MS data acquisition were listed in Table 7. For the liquid chromatography, 100% water and 100% acetonitrile that both supplied with 0.1% formic acid were used as mobile phase A and B respectively. All protein samples were separated by Accucore™ 150-C4 analytical HPLC column (150 mm × 2.1 mm, 2.6 μm particle size) (Thermo Fisher Scientific, Waltham, MA, USA). The scan range of mass spectrometer and the gradient of liquid chromatography applied for each analysis varied according to Table 8.

ESI-MS raw data was exported as .txt files using Xcalibur 4.1.31.9 Qual Browser (Thermo Fisher Scientific, Waltham, MA, USA). Deconvolution of each raw file was processed by Bayesian Protein Reconstruction using Analyst™ QS v1.1 software (Applied Biosystems, Foster City, CA, USA). The deconvolution parameters are set as following: Adduct: hydrogen; Step mass: 0.1 Da; S/N threshold: 20; Minimum intensity (%): 5; Iteration: 20. Start and stop mass were inputted according to the observed mass of each protein sample. Deconvoluted results were then replotted using GraphPad Prism

Table 7 General Settings of Q Exactive Orbitrap ESI-MS.

Mass Spectrometer Parameter	Value
Spray Voltage (+)	3.75 kV
Capillary Temperature	320 °C
Sheath Gas Flow Rate	45
Auxiliary Gas Flow Rate	30
Sweep Gas Flow Rate	0
S-Lens RF level	70
Auxiliary Gas Heat Temperature	30 °C
Ion Source	HESI
Resolution	70000
AGC target	3e6
Maximum Injection Time	150 ms

Table 8 LC-MS Method Settings.

Protein	Method Duration (min)	Mass Spec Scan Range (m/z)	Gradient (%B)
Ub mutants	55	700-1600	5.1-40 min, 15-35%; 40.1 min, 100%
Ub mutants	40	700-1600	3-20 min, 20-30%; 25.1 min, 100%
Ub mutants	30	700-1600	3.1-20 min, 25-32%; 20.1 min, 100%
SUMO1-4	30	700-1600	3.1-20 min, 25-32%; 20.1 min, 100%
NEDD8	30	700-1600	3.1-20 min, 25-35%; 20.1 min, 100%
MNSFβ	30	700-1600	3.1-20 min, 30-40%; 20.1 min, 100%
UFM1	55	700-1600	5.1-40 min, 25-45%; 40.1 min, 100%
URM1	55	700-1800	7.1-32 min, 30-45%; 37.1 min, 60%; 40 min, 98%
ISG15	30	700-2000	3.1-20 min, 32-42%; 20.1 min, 100%
GABARAP	30	700-1800	3.1-20 min, 27-37%; 20.1 min, 100%
GABARAPL2	30	700-1800	3.1-20 min, 27-37%; 20.1 min, 100%
Others	55	700-1800	5.1-40 min, 15-75%; 40.1 min, 100%

8 (GraphPad, San Diego, CA, USA) to obtain normalized spectra. After that, normalized deconvoluted results were integrated by a Python script to achieve their experimental averaged mass values. The theoretic monoisotopic peaks distribution and values were calculated using another Python script. All integrated and theoretic data were plotted using GraphPad Prism 8.

4.7. Biological assays

4.7.1. Covalent capture of DUBs and ULPs by FLAG-Ub/Ubl-Pa probes

0.1 μg/μL DUB (UCHL1, UCHL3, OTUB1, OTUB2, USP5 (Sino Biological), USP2, or USP15 (ENZO Life Sciences)) and 0.1 μg/μL FLAG-Ub-G76Pa was mixed and adjusted into a 20 μL 1× PBS buffer. All reactions were incubated at 37 ºC for 45 min and quenched by adding a 6× SDS loading dye containing 30% β-mercaptoethanol. The mixtures were then analyzed by SDS gel electrophoresis (8-12%) and western blotting using a peroxidase labeled anti-FLAG antibody (Abcam). For FLAG-NEDD8-G76Pa, only the first four enzymes mentioned above were tested. For FLAG-SUMO1'/2'/3'/4'-GxPa, FLAG-ISG15'-G157Pa and FLAG-GABARAP-G116Pa, SENP1 (Abcam), USP5, and ATG4B (Abcam) were used respectively. Reactions were carried out and analyzed similarly.

4.7.2. Profiling of DUBs in the HEK293T cell lysate using FLAG-Ub-G76Pa

HEK293T cells (ATCC CRL-11268) were cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (Gibco). Cell was grown on a 10 cm culture plate inside a 37 ºC, 5% CO_2 incubator until about 90% confluency and collected in 1× PBS buffer into a 1.5 mL tube. 1% Triton X-100 along with 0.1 mM phenylmethylsulfonyl fluoride (Sigma) was then introduced into the cell suspension and shaken under refrigeration for one hour in order to lyse the cells. After the lysis process, the cell lysate was clarified by centrifugation (14000 rpm, 10 min, 4 ºC), the supernatant

was collected, and the concentration was measured by the BCA assay. To perform the binding reaction, FLAG-Ub-G76C-6H and Ub-G76Pa protein pellet were dissolved in 1× PBS buffer and quantitate their concentration by the BCA assay. Then, all components were mixed together to afford the reaction mixture which has 1 μM FLAG-Ub-G76Pa, 5 mM DTT and 1-3 mg/mL HEK293T cell lysate proteins. The reaction mixtures were incubated at rt for 30 min and quenched by adding 6× SDS loading dye containing 30% β-mercaptoethanol. All reactions were finally analyzed by SDS gel electrophoresis (8-12%) and western blotting using a peroxidase labeled anti-FLAG antibody to visualize the binding results.

4.7.3. DUB/ULP catalytic activity assays using Ub/Ubl-AMC

Each reaction mixture contained 50 nM DUB, 200 nM Ub-AMR or FLAG-SUMO1-3-AMC, and 1 mM DTT in a 100 μL 1× PBS buffer. Meanwhile, control reactions which did not have the 50 nM DUB inside were also set up. Enzymes used for Ub-AMC and FLAG-SUMO1-3-AMC were UCHL1 and SENP1, respectively. All mixtures were prepared in wells of a 96-well black plate and the fluorescent signals were recorded by the plate reader (BioTek Synergy H1) per 50 s in a 1 h time frame. The excitation and emission wavelengths were set as 380 and 460 nm, respectively.

4.7.4. Assembly of the H2AK129ac-containing nucleosome

His-TEV-H2B, His-TEV-H3 and His-SUMO-TEV-H4 were expressed and purified according to procedures in a previous publication[215]. All histone pellets were

dissolved in a 6 M GndCl buffer (20 mM Tris, 500 mM NaCl, pH 7.5), and concentration was measured by UV absorption at 280 nm. To prepare H2AK129ac/H2B dimer, H2AK129ac and His-TEV-H2B were mixed in the molar ratio of 1:1, and 6 M GndCl buffer was added to adjust total protein concentration to 4 mg/mL. The denatured H2AK129ac/H2B solution was dialyzed sequentially at 4 ºC against 2/1/0.5 M TE buffers (2/1/0.5 M NaCl, 20 mM Tris, 1 mM EDTA, pH 7.8). The resulting dimer solution was centrifuged (14000 rpm, 5 min, 4 ºC) to remove precipitates and its concentration was determined by UV absorption at 280 nm. Steps for His-TEV-H3/His-SUMO-TEV-H4 tetramer refolding were generally similar to the refolding of the H2AK129ac/H2B dimer except that the total protein concentration was adjusted to 2 mg/mL and there was no stirring in the sequential dialysis. To generate a histone octamer, the refolded H2AK129ac/H2B dimer was mixed with the refolded His-TEV-H3/His-SUMO-TEV-H4 tetramer in a molar ratio of 1:1, and NaCl solid was added to adjust NaCl concentration to 2 M. A 601 nucleosome-positioning DNA (147 bp) was prepared by PCR and purified with the PCR cleanup kit (Epoch Life Science). Purified 601 DNA was re-dissolved in a 2 M TE buffer and added to histone octamer solution in the molar ratio of 0.9:1. Additional 2 M TE buffer was added to adjust the final DNA concentration to about 2-3 μM. The DNA-histone solution was then transferred to a dialysis bag and placed inside a 200 mL 2 M TE buffer. While it was stirred at rt, 20 mM Tris buffer was slowly added into the 2 M salt buffer through a liquid transfer pump (VWR). When salt concentration was reduced to about 200 mM (measured by EX170 salinity meter), the dialysis bag would be transferred to nucleosome storage buffer (20

mM Tris, 20 mM NaCl, 1 mM EDTA, pH 7.8) for another 4 h, and precipitation was removed through centrifugation (14000 rpm, room temperature, 10 min). TEV protease was then added to the nucleosome solution (TEV: nucleosome 1:30, w:w) for about 1 h at 37 ºC to remove all the histone tags. Native 0.2× TBE 5% PAGE gel electrophoresis was run to check nucleosome assembly efficiency

4.7.5. Activity assays of $RNH_{59\text{-}189}$-Ha and $RNH_{59\text{-}196}$-K190C

The activity of $RNH_{59\text{-}189}$-Ha and $RNH_{59\text{-}196}$-K190C were analyzed using a 5'-Cy3 and 3'-Cy5 labeled 12-mer RNA/DNA hybrid double strand in 50 mM Tris (pH 7.9), 50 mM NaCl, 5 mM $MgCl_2$,1 mM DTT, and 20 μg/mL BSA. The FRET-labeled 12-mer RNA and its reverse complement DNA were purchased separately from IDT Inc. and annealed by gradient cooling process. Protein and substrate were incubated at 5:1 molar ratio, emission of both Cy3(I_3) and Cy5(I_5) under single 550 nm excitation wavelength were recorded in a 60 min time frame.

4.8. Product yield quantitation

Quantitation of the reaction yield was calculated by the peak areas on the mass intensity chromatogram using the Xcalibur 3.1 Qual Browser. For example, the yield of Dha product ($Y_{Dha}\%$) was calculated using equation (1) where *A* referred to each peak area. Considered possible tailing effect and background noise, each peak area only included the space above the line between its beginning and ending. In addition, for peaks that showed up on the previous peak's tailing area, we assumed such contributions

were all caused by the detection of new species. Further correction applied when a single peak contained both Dha product and +CN product (the intensity of the lower signal species higher than 5 % the intensity of higher signal species). In this case, the ratio of Dha product (R_{Dha}) was quantified by the proportion of its signal intensity (I_{Dha}) on the integrated deconvoluted mass spectra using equation (2). That the A_{Dha} on the numerator in equation (1) was substituted by A_{Dha}xR_{Dha} led to equation (3) to calculate yield of Dha product.

$$Y_{Dha}\% = \frac{A_{Dha}}{A_{Dha} + A_{hydro} + A_{TNB}} \times 100\ \% \quad (1)$$

$$R_{Dha} = \frac{I_{Dha}}{I_{Dha} + I_{CN}} \quad (2)$$

$$Y_{Dha}\% = \frac{A_{Dha} \times R_{Dha}}{A_{Dha} + A_{hydro} + A_{TNB}} \times 100\ \% \quad (3)$$

CHAPTER V

SUMMARY AND OUTLOOKS

In summary, we have developed a novel expressed protein ligation technique called activated cysteine-directed protein ligation that uses a cyanylating reagent, 2-nitro-5-thiocyanatobenzoic acid, to directly activate a cysteine in a recombinant protein for ligation with small molecule amines and large peptides or protein fragments when coupling with peptide hydrazide ligation. Although cysteine is required for processing ACPL reactions, the eventual cleavage of cysteine residue and all other *C*-terminal amino acids allows the flexible design for substrate protein. This technique requires no enzymatic catalysis and is very easy to perform. The reaction itself can be carried out in both native and denatured aqueous conditions that are weakly basic and is expected to work with a large variety of more nucleophilic species. We also anticipate quick and wide adoption of the technique and its broad application in both biological research and industry processes.

Encouraged by our successful HEK293T cell lysate pulling-down assay using Ub-G76Pa as well as the generation of Ubl-based probes, we anticipate more DUBs and ULPs that specifically recognize those Ubl proteins would be discovered in the similar way. We have also successful conjugated glycyl aminoacetonitrile (Gly-AACN) at the *C*-terminus of FLAG-$Ub_{(1-74)}$-G75C-6H (Figure 55a) that may be able to serve as an activity-based probe to substitute propargylamine-based probes. This FLAG-tagged probe should also have the ability to conjugate human PARKIN, a 465 amino acids RBR

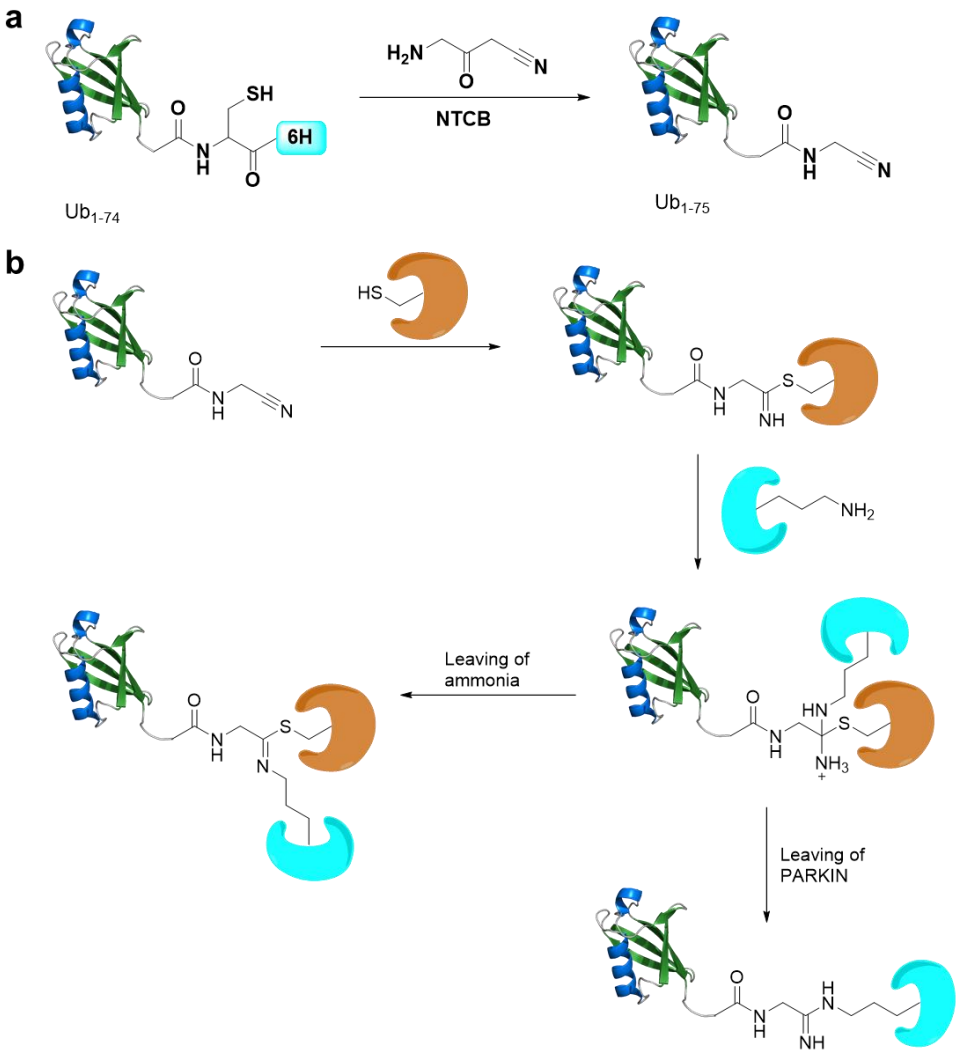

Figure 55 ACPL synthesized Ub probes for PARKIN capturing analysis. (a) The reaction to synthesize FLAG-Ub-G76AACN; (b) The proposed recognition pathway when incubating FLAG-Ub-G76AACN and human cell lysates.

type E3 ubiquitin ligase,[216] *in vitro*. This formation mimics the natural construction of PARKIN linked ubiquitin during the ubiquitin cascade and would recognize by downstream target proteins. By incubating this ubiquitinated PARKIN with human cell

lysate, we anticipate to pulling down several targeted proteins via either amidine formation (Figure 55b). Because the mutation of PARKIN is the major cause of the autosomal recessive form of Parkinson's disease,[217-218] those proteomics information generates by our proposed design would lead the pathology and drug discovery for Parkinson's disease.

In addition, ACPL can also be used to synthesized ubiquitinated proteins. It has been proved that ACPL can site-specifically install 2-chloroethylamine (ClEa) at the *C*-terminus of Ub (Figure 56a). By incubating with Ub-KxC mutants, di-Ub with various linkages at different positions can be generated (Figure 56b). Furthermore, Ub-Ha that achieved via ACPL (Figure 56c) can undergo a hydrazide ligation with 2-mercaptoethylamine (MEa) to obtain Ub-MEa (Figure 56d). This conjugate should be able to react with Ub-KxDha via Michael addition to obtain same di-Ub linkages (Figure 56e). This same concept can be further expanded onto the synthesis of ubiquitinated histone proteins as well.

For proteins natively contain other cysteine residues, as we mentioned above, using a genetic encoded photocaged cysteine (PhC) at the original cysteine position followed by ACPL reaction, the reaction at original cysteine position can be fully inhibited which results in only *C*-terminal conjugation. A mutagenic mmPylRS, photocaged cysteinyl t-RNA synthetase (PhCRS) that used to incorporate PhC at amber suppression codon has been successfully generated and confirmed by sequencing (Figure 57a). Test experiment using co-expressed PhCRS and sfGFP-N134TAG determined the successful of PhC incorporation (Figure 57b, c). The co-expression of Ub-K48TAG-

G76C-6H and PhCRS also generated PhC incorporated ubiquitin in a good yield (Figure 57d). The following ACPL reaction using Pa resulted in only C-terminal Pa conjugated Ub that further confirmed the inhibition of cleavage at K48 position (Figure 57e). We are currently working on optimization of the photolysis step. In principle, this method can fully overcome the major shortcoming of ACPL approach and allow the reaction to be processed on all proteins that native containing one cysteine residue such as SUMOs (Figure 56f). Therefore, Ubl-based probes that contain the closest construction as their native structure are able to be obtained using our ACPL technique.

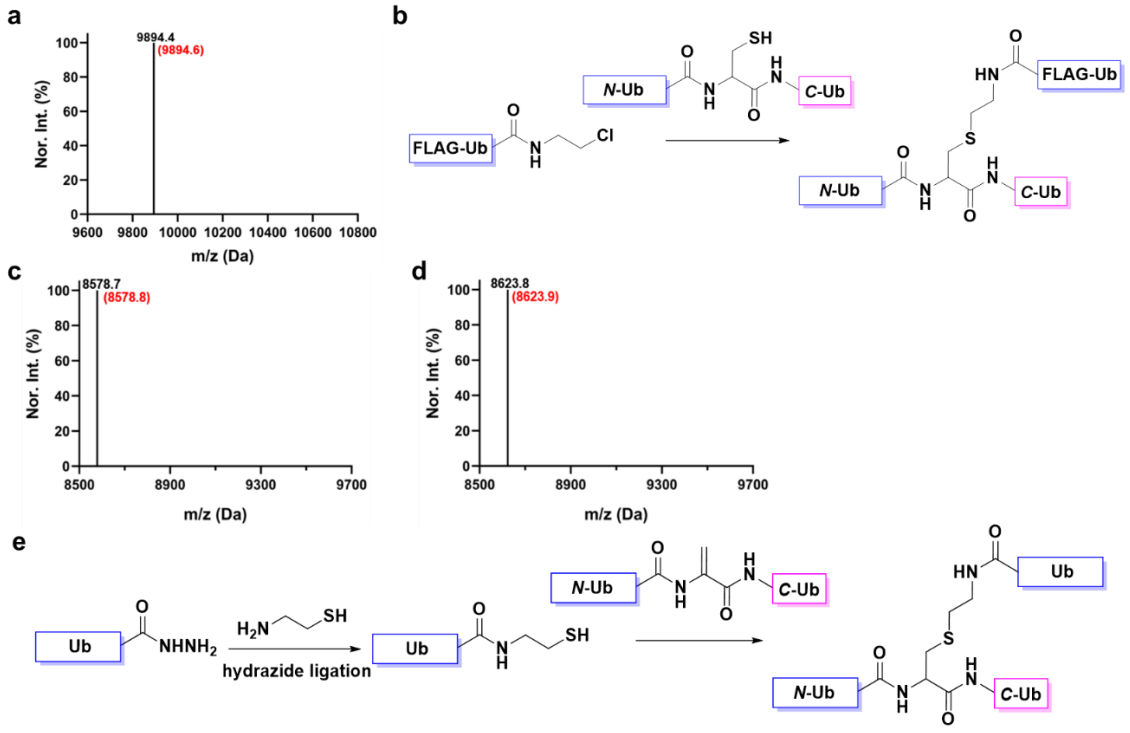

Figure 56 Two approaches to chemically synthesize bi-Ub. (a) The deconvoluted and integrated MS of FLAG-Ub-ClEa; (b) Synthesis of di-Ub using FLAG-Ub-ClEa and Ub-KxC; (c) The deconvoluted and integrated MS of Ub-Ha; (d) The deconvoluted and integrated MS of Ub-MEa; (e) Synthesis of di-Ub using Ub-MEa and Ub-KxDha.

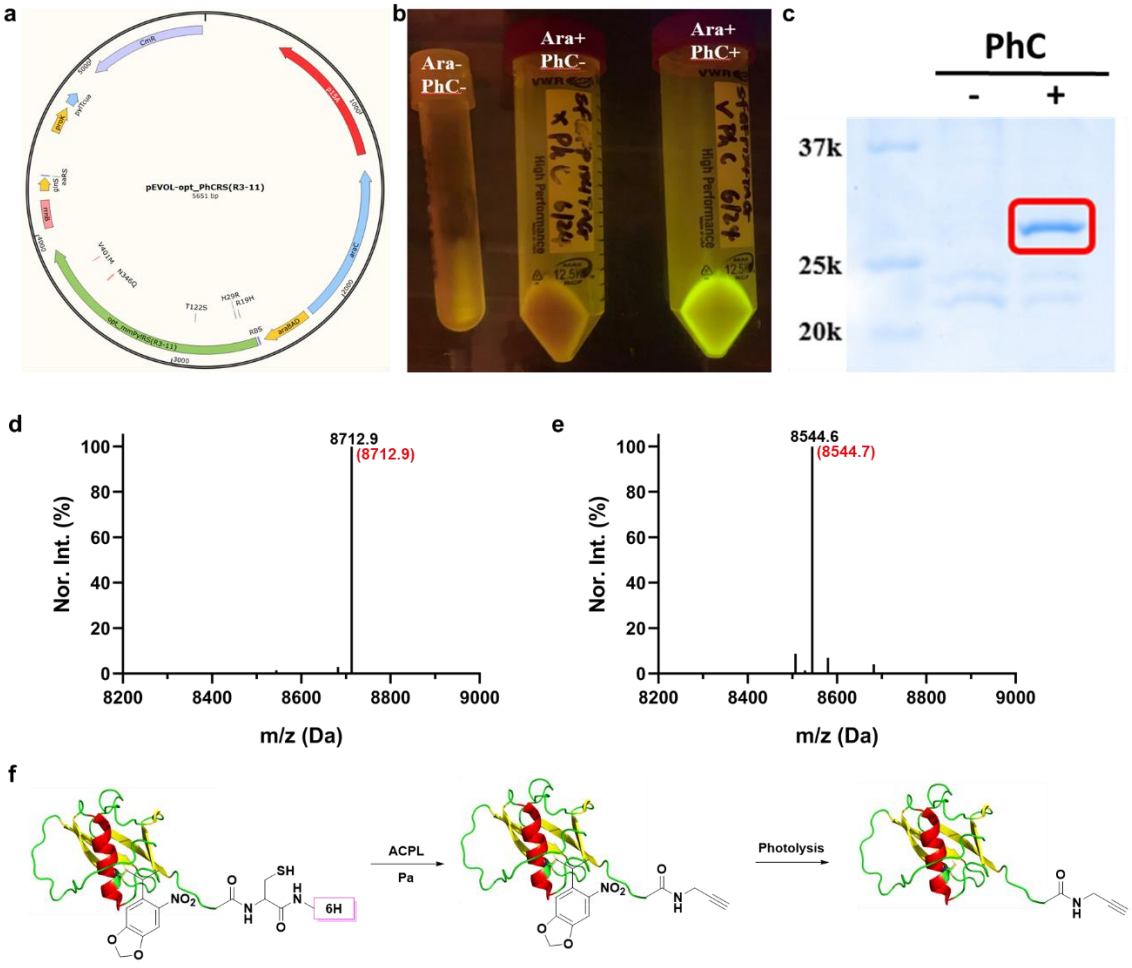

Figure 57 The concept and initial tests of photocaged cysteine assisted ACPL reaction. (a) Plasmid map for PhCRS; (b) Test co-expression of PhCRS and sfGFP-N134TAG, Ara refers to 0.2% arabinose for PhCRS protein induction; (c) The SDS-PAGE showing the incorporation of PhC when PhC was added during expression; (d) The deconvoluted and integrated MS of co-expressed Ub-K48PhC-6H; (e) The deconvoluted and integrated MS of Ub-K48PhC-G76Pa, red number indicated the theoretical average mass; (f) The proposed synthetic pathway for FLAG-SUMOs-GxPa that contain one native cysteine.

Furthermore, we have successfully used the super simple NTCB triggered Dha formation method to synthesize a bunch of super useful FLAG-Ub/Ubl-Dha-6H probes. We anticipate using all of them for proteomics discovery of cysteine-containing enzymes as while as DUBs and ULPs. The majority of histone PTMs are found at the flexible *N*-terminal tail of all four histones. Since the flexible nature of the histone tails potentiates NTCB-triggered Dha formation from cysteine, we foresee potential applications of our method in the generation of histones with PTM analogues as well for their functional annotation. This method was later found not reactive enough for cysteines at internal positions of a protein, however, by careful editing *C*-terminal sequence of the cysteine residue, such as introducing additional linkers or electron donating amino acids, we expected to see improved reactivity on both nucleophilic substitution and beta-elimination pathway.

REFERENCES

1. Walsh, C. T.; Garneau-Tsodikova, S.; Gatto, G. J., Jr., Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem., Int. Ed.* **2005,** *44* (45), 7342-72.

2. Spicer, C. D.; Davis, B. G., Selective chemical protein modification. *Nat. Commun.* **2014,** *5* (1), 4740.

3. Xu, L.; Kuan, S. L.; Weil, T., Contemporary Approaches for Site-Selective Dual Functionalization of Proteins. *Angew. Chem., Int. Ed.* **2020,** *60*, 2-23.

4. Walsh, C., *Posttranslational modification of proteins: expanding nature's inventory*. Roberts and Company Publishers: Englewood, CO, 2006.

5. Santos, A. L.; Lindner, A. B., Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease. *Oxid. Med. Cell. Longev.* **2017,** *2017*, 5716409.

6. Martin, C.; Zhang, Y., The diverse functions of histone lysine methylation. *Nat. Rev. Mol. Cell Biol.* **2005,** *6* (11), 838-849.

7. Han, D.; Huang, M.; Wang, T.; Li, Z.; Chen, Y.; Liu, C.; Lei, Z.; Chu, X., Lysine methylation of transcription factors in cancer. *Cell Death Dis.* **2019,** *10* (4), 290.

8. Levy, D., Lysine methylation signaling of non-histone proteins in the nucleus. *Cell. Mol. Life Sci.* **2019,** *76* (15), 2873-2883.

9. Lin, H.; Su, X.; He, B., Protein Lysine Acylation and Cysteine Succination by Intermediates of Energy Metabolism. *ACS Chem. Bio.* **2012,** *7* (6), 947-960.

10. Ali, I.; Conrad, R. J.; Verdin, E.; Ott, M., Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics. *Chem. Rev.* **2018,** *118* (3), 1216-1252.

11. Xu, G.; Jaffrey, S. R., The new landscape of protein ubiquitination. *Nat. Biotechnol.* **2011,** *29* (12), 1098-1100.

12. Mattiroli, F.; Sixma, T. K., Lysine-targeting specificity in ubiquitin and ubiquitin-like modification pathways. *Nat. Struct. Biol.* **2014,** *21* (4), 308-316.

13. Guccione, E.; Richard, S., The regulation, functions and clinical relevance of arginine methylation. *Nat. Rev. Mol. Cell Biol.* **2019,** *20* (10), 642-657.

14. van Venrooij, W. J.; Pruijn, G. J., Citrullination: a small change for a protein with great consequences for rheumatoid arthritis. *Arthritis Res.* **2000,** *2* (4), 249-251.

15. McCubrey, J. A.; May, W. S.; Duronio, V.; Mufson, A., Serine/threonine phosphorylation in cytokine signal transduction. *Leukemia* **2000,** *14* (1), 9-21.

16. Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L., The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy (Review). *Int. J. Mol. Med.* **2017,** *40* (2), 271-280.

17. Van den Steen, P.; Rudd, P. M.; Dwek, R. A.; Opdenakker, G., Concepts and principles of O-linked glycosylation. *Crit. Rev. Biochem. Mol. Biol.* **1998,** *33* (3), 151-208.

18. Reily, C.; Stewart, T. J.; Renfrow, M. B.; Novak, J., Glycosylation in health and disease. *Nat. Rev. Nephrol.* **2019,** *15* (6), 346-366.

19. Turell, L.; Zeida, A.; Trujillo, M., Mechanisms and consequences of protein cysteine oxidation: the role of the initial short-lived intermediates. *Essays Biochem.* **2020,** *64* (1), 55-66.

20. Hess, D. T.; Matsumoto, A.; Kim, S.-O.; Marshall, H. E.; Stamler, J. S., Protein S-nitrosylation: purview and parameters. *Nat. Rev. Mol. Cell Biol.* **2005,** *6* (2), 150-166.

21. Shah, A. A.; Khan, Y. D., Identification of 4-carboxyglutamate residue sites based on position based statistical feature and multiple classification. *Sci. Rep.* **2020,** *10* (1), 16913.

22. Marino, G.; Eckhard, U.; Overall, C. M., Protein Termini and Their Modifications Revealed by Positional Proteomics. *ACS Chem. Bio.* **2015,** *10* (8), 1754-1764.

23. Shadish, J. A.; DeForest, C. A., Site-Selective Protein Modification: From Functionalized Proteins to Functional Biomaterials. *Matter* **2020,** *2* (1), 50-77.

24. Wang, X. A.; Kurra, Y.; Huang, Y.; Lee, Y. J.; Liu, W. R., E1‐Catalyzed Ubiquitin C‐Terminal Amidation for the Facile Synthesis of Deubiquitinase Substrates. *ChemBioChem* **2014,** *15* (1), 37-41.

25. Barber, K. W.; Rinehart, J., The ABCs of PTMs. *Nat. Chem. Bio.* **2018,** *14* (3), 188-192.

26. Noren, C.; Anthony-Cahill, S.; Griffith, M.; Schultz, P., A general method for site-specific incorporation of unnatural amino acids into proteins. *Science* **1989,** *244* (4901), 182-188.

27. Ovaa, H.; wals, k., Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2014,** *2* (15).

28. Hoyt, E. A.; Cal, P. M. S. D.; Oliveira, B. L.; Bernardes, G. J. L., Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **2019,** *3* (3), 147-171.

29. Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H.-C., Selective N-terminal functionalization of native peptides and proteins. *Chem. Sci.* **2017,** *8* (4), 2717-2722.

30. van Vught, R.; Pieters, R. J.; Breukink, E., Site-specific functionalization of proteins and their applications to therapeutic antibodies. *Comput. Struct. Biotechnol. J.* **2014,** *9* (14), e201402001.

31. Miseta, A.; Csutora, P., Relationship between the occurrence of cysteine in proteins and the complexity of organisms. *Mol. Biol. Evol.* **2000,** *17* (8), 1232-1239.

32. Krall, N.; Da Cruz, F. P.; Boutureira, O.; Bernardes, G. J., Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016,** *8* (2), 103-113.

33. Gunnoo, S. B.; Madder, A., Chemical protein modification through cysteine. *ChemBioChem* **2016,** *17* (7), 529-553.

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

35. Goddard, D. R.; Michaelis, L., Derivatives of keratin. *J. Biol. Chem.* **1935,** *112* (1), 361-371.

36. Dowling, L.; Sparrow, L., Sequences of wool keratin proteins: the CSIRO connection. *Trends. Biochem. Sci.* **1991,** *16*, 115-118.

37. Lundell, N.; Schreitmüller, T., Sample preparation for peptide mapping—a pharmaceutical quality-control perspective. *Anal. Biochem.* **1999,** *266* (1), 31-47.

38. Davis, N. J.; Flitsch, S. L., A novel method for the specific glycosylation of proteins. *Tetrahedron Lett.* **1991,** *32* (46), 6793-6796.

39. Simon, M. D.; Chu, F.; Racki, L. R.; Cecile, C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M., The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007,** *128* (5), 1003-1012.

40. Chu, G.-C.; Pan, M.; Li, J.; Liu, S.; Zuo, C.; Tong, Z.-B.; Bai, J.-S.; Gong, Q.; Ai, H.; Fan, J., Cysteine-aminoethylation-assisted chemical ubiquitination of recombinant histones. *J. Am. Chem. Soc.* **2019,** *141* (8), 3654-3663.

41. Chu, G. C.; Hua, X.; Zuo, C.; Chen, C. C.; Meng, X. B.; Zhang, Z.; Fu, Y.; Shi, J.; Li, Y. M., Efficient Semi‐Synthesis of Atypical Ubiquitin Chains and Ubiquitin‐Based Probes Forged by Thioether Isopeptide Bonds. *Chem. Eur. J.* **2019,** *25* (72), 16668-16675.

42. DeGraan-Weber, N.; Reilly, J. P., Use of Cysteine Aminoethylation To Identify the Hypervariable Peptides of an Antibody. *Anal. Chem.* **2018,** *90* (3), 1608-1612.

43. Yin, L.; Krantz, B.; Russell, N. S.; Deshpande, S.; Wilkinson, K. D., Nonhydrolyzable diubiquitin analogues are inhibitors of ubiquitin conjugation and deconjugation. *Biochemistry* **2000,** *39* (32), 10001-10010.

44. Morgan, M. T.; Haj-Yahya, M.; Ringel, A. E.; Bandi, P.; Brik, A.; Wolberger, C., Structural basis for histone H2B deubiquitination by the SAGA DUB module. *Science* **2016,** *351* (6274), 725-728.

45. Morgan, M.; Jbara, M.; Brik, A.; Wolberger, C., Chapter One - Semisynthesis of ubiquitinated histone H2B with a native or nonhydrolyzable linkage. In *Methods in Enzymology*, Hochstrasser, M., Ed. Academic Press: 2019; Vol. 618, pp 1-27.

46. Jo, H.; Culik, R. M.; Korendovych, I. V.; DeGrado, W. F.; Gai, F., Selective incorporation of nitrile-based infrared probes into proteins via cysteine alkylation. *Biochemistry* **2010,** *49* (49), 10354-10356.

47. Spokoyny, A. M.; Zou, Y.; Ling, J. J.; Yu, H.; Lin, Y.-S.; Pentelute, B. L., A Perfluoroaryl-Cysteine SNAr Chemistry Approach to Unprotected Peptide Stapling. *J. Am. Chem. Soc.* **2013,** *135* (16), 5946-5949.

48. Moore, J. E.; Ward, W. H., Cross-linking of Bovine Plasma Albumin and Wool Keratin. *J. Am. Chem. Soc.* **1956,** *78* (11), 2414-2418.

49. Tsao, T. C.; Bailey, K., The extraction, purification and some chemical properties of actin. *Biochim. Biophys. Acta* **1953,** *11*, 102-113.

50. Crankshaw, M. W.; Grant, G. A., Modification of cysteine. *Curr. Protoc. Protein Sci.* **1996,** *3* (1), 15.1. 1-15.1. 18.

51. Russo, M. S.; Napylov, A.; Paquet, A.; Vuckovic, D., Comparison of N-ethyl maleimide and N-(1-phenylethyl) maleimide for derivatization of biological thiols using liquid chromatography-mass spectrometry. *Anal. Bioanal. Chem.* **2020,** *412* (7), 1639-1652.

52. Zhang, Y.; Zang, C.; An, G.; Shang, M.; Cui, Z.; Chen, G.; Xi, Z.; Zhou, C., Cysteine-specific protein multi-functionalization and disulfide bridging using 3-bromo-5-methylene pyrrolones. *Nat. Commun.* **2020,** *11* (1), 1015.

53. Masri, M. S.; Friedman, M., Protein reactions with methyl and ethyl vinyl sulfones. *J. Protein Chem.* **1988,** *7* (1), 49-54.

54. Hemelaar, J.; Borodovsky, A.; Kessler, B. M.; Reverter, D.; Cook, J.; Kolli, N.; Gan-Erdene, T.; Wilkinson, K. D.; Gill, G.; Lima, C. D.; Ploegh, H. L.; Ovaa, H., Specific and Covalent Targeting of Conjugating and Deconjugating Enzymes of Ubiquitin-Like Proteins. *Mol. Cell. Biol.* **2004,** *24* (1), 84-95.

55. Wu, Z.; Li, L.; Liu, S.; Yakushijin, F.; Yakushijin, K.; Horne, D.; Conti, P. S.; Li, Z.; Kandeel, F.; Shively, J. E., Facile preparation of a thiol-reactive 18F-labeling agent and synthesis of 18F-DEG-VS-NT for PET imaging of a neurotensin receptor–positive tumor. *J. Nucl. Med.* **2014,** *55* (7), 1178-1184.

56. Crankshaw, M. W.; Grant, G. A., Modification of Cysteine. *Curr. Protoc. Protein Sci.* **1996,** *3* (1), 15.1.1-15.1.18.

57. Ellman, G. L., Tissue sulfhydryl groups. *Arch. Biochem. Biophys.* **1959,** *82* (1), 70-77.

58. Anson , M. L., THE REACTIONS OF IODINE AND IODOACETAMIDE WITH NATIVE EGG ALBUMIN. *J. Gen. Physiol.* **1940,** *23* (3), 321-331.

59. Fontana, A.; Scoffone, E.; Benassi, C. A., Sulfenyl halides as modifying reagents for polypeptides and proteins. II. Modification of cysteinyl residues. *Biochemistry* **1968,** *7* (3), 980-986.

60. Kenyon, G. L.; Bruice, T. W., [40] Novel sulfhydryl reagents. In *Methods in Enzymology,* Academic Press: 1977; Vol. 47, pp 407-430.

61. Devarie-Baez, N. O.; Silva Lopez, E. I.; Furdui, C. M., Biological chemistry and functionality of protein sulfenic acids and related thiol modifications. *Free Radic. Res.* **2016,** *50* (2), 172-194.

62. Hawkins, C. L.; Davies, M. J., Detection, identification, and quantification of oxidative protein modifications. *J. Biol. Chem.* **2019,** *294* (51), 19683-19708.

63. Barrett, T. J.; Pattison, D. I.; Leonard, S. E.; Carroll, K. S.; Davies, M. J.; Hawkins, C. L., Inactivation of thiol-dependent enzymes by hypothiocyanous acid: role of sulfenyl thiocyanate and sulfenic acid intermediates. *Free Radic. Biol. Med.* **2012,** *52* (6), 1075-1085.

64. Davis, B. G.; Maughan, M. A. T.; Green, M. P.; Ullman, A.; Jones, J. B., Glycomethanethiosulfonates: powerful reagents for protein glycosylation. *Tetrahedron Asymmetry* **2000,** *11* (1), 245-262.

65. Wender, P. A.; Goun, E. A.; Jones, L. R.; Pillow, T. H.; Rothbard, J. B.; Shinde, R.; Contag, C. H., Real-time analysis of uptake and bioactivatable cleavage of luciferin-transporter conjugates in transgenic reporter mice. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (25), 10340-10345.

66. Yang, A.; Ha, S.; Ahn, J.; Kim, R.; Kim, S.; Lee, Y.; Kim, J.; Söll, D.; Lee, H.-Y.; Park, H.-S., A chemical biology route to site-specific authentic protein modifications. *Science* **2016,** *354* (6312), 623-626.

67. Galan, S. R.; Wickens, J. R.; Dadova, J.; Ng, W.-L.; Zhang, X.; Simion, R. A.; Quinlan, R.; Pires, E.; Paton, R. S.; Caddick, S., Post-translational site-selective protein backbone α-deuteration. *Nat. Chem. Bio.* **2018,** *14* (10), 955-963.

68. Dadová, J.; Galan, S. R.; Davis, B. G., Synthesis of modified proteins via functionalization of dehydroalanine. *Curr. Opin. Chem. Biol.* **2018,** *46*, 71-81.

69. Haj-Yahya, N.; Hemantha, H. P.; Meledin, R.; Bondalapati, S.; Seenaiah, M.; Brik, A., Dehydroalanine-based diubiquitin activity probes. *Org. Lett.* **2014,** *16* (2), 540-543.

70. Weber, A.; Elliott, P. R.; Pinto-Fernandez, A.; Bonham, S.; Kessler, B. M.; Komander, D.; El Oualid, F.; Krappmann, D., A linear diubiquitin-based probe for efficient and selective detection of the deubiquitinating enzyme OTULIN. *Cell Chem. Biol.* **2017,** *24* (10), 1299-1313. e7.

71. Xu, L.; Fan, J.; Wang, Y.; Zhang, Z.; Fu, Y.; Li, Y.-M.; Shi, J., An activity-based probe developed by a sequential dehydroalanine formation strategy targets HECT E3 ubiquitin ligases. *Chem. Commun.* **2019,** *55* (49), 7109-7112.

72. An, H.; Statsyuk, A. V., Facile synthesis of covalent probes to capture enzymatic intermediates during E1 enzyme catalysis. *Chem. Commun.* **2016,** *52* (12), 2477-2480.

73. Jiang, H.-K.; Kurkute, P.; Li, C.-L.; Wang, Y.-H.; Chen, P.-J.; Lin, S.-Y.; Wang, Y.-S., Revealing USP7 Deubiquitinase Substrate Specificity by Unbiased Synthesis of Ubiquitin Tagged SUMO2. *Biochemistry* **2020,** *59* (40), 3796-3801.

74. Meledin, R.; Mali, S. M.; Singh, S. K.; Brik, A., Protein ubiquitination via dehydroalanine: development and insights into the diastereoselective 1, 4-addition step. *Org. Biomol. Chem.* **2016,** *14* (21), 4817-4823.

75. Meledin, R.; Mali, S. M.; Kleifeld, O.; Brik, A., Activity-Based Probes Developed by Applying a Sequential Dehydroalanine Formation Strategy to Expressed Proteins Reveal a Potential α-Globin-Modulating Deubiquitinase. *Angew. Chem., Int. Ed.* **2018,** *57* (20), 5645-5649.

76. Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G., Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2011,** *2* (9), 1666-1676.

77. Donovan, J. W.; White, T. M., Alkaline hydrolysis of the disulfide bonds of ovomucoid and of low molecular weight aliphatic and aromatic disulfides. *Biochemistry* **1971,** *10* (1), 32-8.

78. Teruaki, M.; Masahiro, U.; Eiichiro, S.; Kazuhiko, S., A convenient method for the synthesis of carboxylic esters. *Chem. Lett.* **1975,** *4* (10), 1045-1048.

79. Bernardes, G. J. L.; Grayson, E. J.; Thompson, S.; Chalker, J. M.; Errey, J. C.; El Oualid, F.; Claridge, T. D. W.; Davis, B. G., From Disulfide- to Thioether-Linked Glycoproteins. *Angew. Chem., Int. Ed.* **2008,** *47* (12), 2244-2247.

80. Wang, H.; Zhang, J.; Xian, M., Facile Formation of Dehydroalanine From S-Nitrosocysteines. *J. Am. Chem. Soc.* **2009,** *131* (37), 13238-13239.

81. Holmes, T. J.; Lawton, R. G., Cysteine modification and cleavage of proteins with 2-methyl-N1-benzenesulfonyl-N4-bromoacetylquinonediimide. *J. Am. Chem. Soc.* **1977,** *99* (6), 1984-1986.

82. Bernardes, G. J. L.; Chalker, J. M.; Errey, J. C.; Davis, B. G., Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **2008,** *130* (15), 5052-5053.

83. Fernández-González, M.; Boutureira, O.; Bernardes, G. J. L.; Chalker, J. M.; Young, M. A.; Errey, J. C.; Davis, B. G., Site-selective chemoenzymatic construction of synthetic glycoproteins using endoglycosidases. *Chem. Sci.* **2010,** *1* (6), 709-715.

84. Dadová, J.; Wu, K.-J.; Isenegger, P. G.; Errey, J. C.; Bernardes, G. a. J.; Chalker, J. M.; Raich, L.; Rovira, C.; Davis, B. G., Precise probing of residue roles by post-translational β, γ-C, N aza-Michael mutagenesis in enzyme active sites. *ACS Cent. Sci.* **2017,** *3* (11), 1168-1173.

85. Freedy, A. M.; Matos, M. J.; Boutureira, O.; Corzana, F.; Guerreiro, A.; Akkapeddi, P.; Somovilla, V. J.; Rodrigues, T.; Nicholls, K.; Xie, B., Chemoselective installation of amine bonds on proteins through aza-Michael ligation. *J. Am. Chem. Soc.* **2017,** *139* (50), 18365-18375.

86. Shiu, H.-Y.; Chan, T.-C.; Ho, C.-M.; Liu, Y.; Wong, M.-K.; Che, C.-M., Electron-Deficient Alkynes as Cleavable Reagents for the Modification of Cysteine-Containing Peptides in Aqueous Medium. *Chem. Eur. J.* **2009,** *15* (15), 3839-3850.

87. Shiu, H.-Y.; Chong, H.-C.; Leung, Y.-C.; Wong, M.-K.; Che, C.-M., A Highly Selective FRET-Based Fluorescent Probe for Detection of Cysteine and Homocysteine. *Chem. Eur. J.* **2010,** *16* (11), 3308-3313.

88. Shiu, H.-Y.; Chong, H.-C.; Leung, Y.-C.; Zou, T.; Che, C.-M., Phosphorescent proteins for bio-imaging and site selective bio-conjugation of peptides and proteins with luminescent cyclometalated iridium(iii) complexes. *Chem. Commun.* **2014,** *50* (33), 4375-4378.

89. Ekkebus, R.; van Kasteren, S. I.; Kulathu, Y.; Scholten, A.; Berlin, I.; Geurink, P. P.; de Jong, A.; Goerdayal, S.; Neefjes, J.; Heck, A. J., On terminal alkynes that can react with active-site cysteine nucleophiles in proteases. *J. Am. Chem. Soc.* **2013,** *135* (8), 2867-2870.

90. Dondoni, A.; Massi, A.; Nanni, P.; Roda, A., A New Ligation Strategy for Peptide and Protein Glycosylation: Photoinduced Thiol–Ene Coupling. *Chem. Eur. J.* **2009,** *15* (43), 11444-11449.

91. Floyd, N.; Vijayakrishnan, B.; Koeppe, J. R.; Davis, B. G., Thiyl Glycosylation of Olefinic Proteins: S-Linked Glycoconjugate Synthesis. *Angew. Chem., Int. Ed.* **2009,** *48* (42), 7798-7802.

92. Li, Y.; Yang, M.; Huang, Y.; Song, X.; Liu, L.; Chen, P. R., Genetically encoded alkenyl–pyrrolysine analogues for thiol–ene reaction mediated site-specific protein labeling. *Chem. Sci.* **2012,** *3* (9), 2766-2770.

93. Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C.-F., A Direct Method for Site-Specific Protein Acetylation. *Angew. Chem., Int. Ed.* **2011,** *50* (41), 9611-9614.

94. Valkevich, E. M.; Guenette, R. G.; Sanchez, N. A.; Chen, Y.-c.; Ge, Y.; Strieter, E. R., Forging isopeptide bonds using thiol–ene chemistry: site-specific coupling of ubiquitin molecules for studying the activity of isopeptidases. *J. Am. Chem. Soc.* **2012,** *134* (16), 6916-6919.

95. Trang, V. H.; Valkevich, E. M.; Minami, S.; Chen, Y. C.; Ge, Y.; Strieter, E. R., Nonenzymatic polymerization of ubiquitin: single‐step synthesis and isolation of discrete ubiquitin oligomers. *Angew. Chem., Int. Ed.* **2012,** *51* (52), 13085-13088.

96. Grim, J. C.; Brown, T. E.; Aguado, B. A.; Chapnick, D. A.; Viert, A. L.; Liu, X.; Anseth, K. S., A Reversible and Repeatable Thiol–Ene Bioconjugation for Dynamic Patterning of Signaling Proteins in Hydrogels. *ACS Cent. Sci.* **2018,** *4* (7), 909-916.

97. Conte, M. L.; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.; Dondoni, A., Multi-molecule reaction of serum albumin can occur through thiol-yne coupling. *Chem. Commun.* **2011,** *47* (39), 11086-11088.

98. Casi, G.; Huguenin-Dezot, N.; Zuberbühler, K.; Scheuermann, J. r.; Neri, D., Site-specific traceless coupling of potent cytotoxic drugs to recombinant antibodies for pharmacodelivery. *J. Am. Chem. Soc.* **2012,** *134* (13), 5887-5892.

99. Bernardes, G. J.; Steiner, M.; Hartmann, I.; Neri, D.; Casi, G., Site-specific chemical modification of antibody fragments using traceless cleavable linkers. *Nat. Protoc.* **2013,** *8* (11), 2079.

100. Merrifield, R. B., Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J. Am. Chem. Soc.* **1963,** *85* (14), 2149-2154.

101. Merrifield, B., Solid phase synthesis. *Science* **1986,** *232* (4748), 341-7.

102. Nilsson, B. L.; Soellner, M. B.; Raines, R. T., Chemical synthesis of proteins. *Annu. Rev. Biophys. Biomol. Struct.* **2005,** *34*, 91-118.

103. Palomo, J. M., Solid-phase peptide synthesis: an overview focused on the preparation of biologically relevant peptides. *RSC Adv.* **2014,** *4* (62), 32658-32672.

104. McGrath, N. A.; Raines, R. T., Chemoselectivity in Chemical Biology: Acyl Transfer Reactions with Sulfur and Selenium. *Acc. Chem. Res.* **2011,** *44* (9), 752-761.

105. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-779.

106. Kent, S. B. H., Total chemical synthesis of proteins. *Chem. Soc. Rev.* **2009,** *38* (2), 338-351.

107. Virdee, S.; Ye, Y.; Nguyen, D. P.; Komander, D.; Chin, J. W., Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. *Nat. Chem. Bio.* **2010,** *6* (10), 750-757.

108. Virdee, S.; Kapadnis, P. B.; Elliott, T.; Lang, K.; Madrzak, J.; Nguyen, D. P.; Riechmann, L.; Chin, J. W., Traceless and site-specific ubiquitination of recombinant proteins. *J. Am. Chem. Soc.* **2011,** *133* (28), 10708-10711.

109. Sui, X.; Wang, Y.; Du, Y.-X.; Liang, L.-J.; Zheng, Q.; Li, Y.-M.; Liu, L., Development and application of ubiquitin-based chemical probes. *Chem. Sci.* **2020,** *11* (47), 12633-12646.

110. Bowman, B. J.; Allen, R.; Wechser, M. A.; Bowman, E. J., Isolation of genes encoding the Neurospora vacuolar ATPase. Analysis of vma-2 encoding the 57-kDa polypeptide and comparison to vma-1. *J. Biol. Chem.* **1988,** *263* (28), 14002-14007.

111. Anraku, Y.; Mizutani, R.; Satow, Y., Protein Splicing: Its Discovery and Structural Insight into Novel Chemical Mechanisms. *IUBMB Life* **2005,** *57* (8), 563-574.

112. Muir, T. W.; Sondhi, D.; Cole, P. A., Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (12), 6705-6710.

113. Berrade, L.; Camarero, J. A., Expressed protein ligation: a resourceful tool to study protein structure and function. *Cell. Mol. Life Sci.* **2009,** *66* (24), 3909-3922.

114. Vila-Perelló, M.; Liu, Z.; Shah, N. H.; Willis, J. A.; Idoyaga, J.; Muir, T. W., Streamlined expressed protein ligation using split inteins. *J. Am. Chem. Soc.* **2012,** *135* (1), 286-292.

115. Muralidharan, V.; Muir, T. W., Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat. Methods* **2006,** *3* (6), 429.

116. Wu, H.; Hu, Z.; Liu, X. Q., Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (16), 9226-31.

117. Southworth, M. W.; Adam, E.; Panne, D.; Byer, R.; Kautz, R.; Perler, F. B., Control of protein splicing by intein fragment reassembly. *EMBO J.* **1998,** *17* (4), 918-926.

118. Fang, G. M.; Li, Y. M.; Shen, F.; Huang, Y. C.; Li, J. B.; Lin, Y.; Cui, H. K.; Liu, L., Protein chemical synthesis by ligation of peptide hydrazides. *Angew. Chem., Int. Ed.* **2011,** *50* (33), 7645-7649.

119. Wang, S. S., p-alkoxybenzyl alcohol resin and p-alkoxybenzyloxycarbonylhydrazide resin for solid phase synthesis of protected peptide fragments. *J. Am. Chem. Soc.* **1973,** *95* (4), 1328-33.

120. Chang, J. K.; Shimizu, M.; Wang, S. S., Fully automated solid phase synthesis of protected peptide hydrazides on recycling hydroxymethyl resin. *J. Org. Chem.* **1976,** *41* (20), 3255-8.

121. Flood, D. T.; Hintzen, J. C. J.; Bird, M. J.; Cistrone, P. A.; Chen, J. S.; Dawson, P. E., Leveraging the Knorr Pyrazole Synthesis for the Facile Generation of Thioester Surrogates for use in Native Chemical Ligation. *Angew. Chem., Int. Ed.* **2018,** *57* (36), 11634-11639.

122. Rut, W.; Zmudzinski, M.; Snipas, S. J.; Bekes, M.; Huang, T. T.; Drag, M., Engineered unnatural ubiquitin for optimal detection of deubiquitinating enzymes. *Chem. Sci.* **2020,** *11* (23), 6058-6069.

123. Warren, J. D.; Miller, J. S.; Keding, S. J.; Danishefsky, S. J., Toward Fully Synthetic Glycoproteins by Ultimately Convergent Routes: A Solution to a Long-Standing Problem. *J. Am. Chem. Soc.* **2004,** *126* (21), 6576-6578.

124. Hackeng, T. M.; Griffin, J. H.; Dawson, P. E., Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (18), 10068-73.

125. Wan, Q.; Chen, J.; Yuan, Y.; Danishefsky, S. J., Oxo-ester Mediated Native Chemical Ligation: Concept and Applications. *J. Am. Chem. Soc.* **2008,** *130* (47), 15814-15816.

126. Yan, L. Z.; Dawson, P. E., Synthesis of Peptides and Proteins without Cysteine Residues by Native Chemical Ligation Combined with Desulfurization. *J. Am. Chem. Soc.* **2001,** *123* (4), 526-533.

127. Bang, D.; Makhatadze, G. I.; Tereshko, V.; Kossiakoff, A. A.; Kent, S. B., Total Chemical Synthesis and X-ray Crystal Structure of a Protein Diastereomer: [D-Gln 35]Ubiquitin. *Angew. Chem., Int. Ed.* **2005,** *44* (25), 3852-3856.

128. Pentelute, B. L.; Kent, S. B. H., Selective Desulfurization of Cysteine in the Presence of Cys(Acm) in Polypeptides Obtained by Native Chemical Ligation. *Org. Lett.* **2007,** *9* (4), 687-690.

129. Wan, Q.; Danishefsky, S. J., Free-Radical-Based, Specific Desulfurization of Cysteine: A Powerful Advance in the Synthesis of Polypeptides and Glycopolypeptides. *Angew. Chem., Int. Ed.* **2007,** *46* (48), 9248-9252.

130. Kan, C.; Trzupek, J. D.; Wu, B.; Wan, Q.; Chen, G.; Tan, Z.; Yuan, Y.; Danishefsky, S. J., Toward homogeneous erythropoietin: chemical synthesis of the Ala1-Gly28 glycopeptide domain by "alanine" ligation. *J. Am. Chem. Soc.* **2009,** *131* (15), 5438-43.

131. Crich, D.; Banerjee, A., Native Chemical Ligation at Phenylalanine. *J. Am. Chem. Soc.* **2007,** *129* (33), 10064-10065.

132. Haase, C.; Rohde, H.; Seitz, O., Native Chemical Ligation at Valine. *Angew. Chem., Int. Ed.* **2008,** *47* (36), 6807-6810.

133. Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J., Native Chemical Ligation at Valine: A Contribution to Peptide and Glycopeptide Synthesis. *Angew. Chem., Int. Ed.* **2008,** *47* (44), 8521-8524.

134. Harpaz, Z.; Siman, P.; Kumar, K. S. A.; Brik, A., Protein Synthesis Assisted by Native Chemical Ligation at Leucine. *ChemBioChem* **2010,** *11* (9), 1232-1235.

135. Shang, S.; Tan, Z.; Dong, S.; Danishefsky, S. J., An Advance in Proline Ligation. *J. Am. Chem. Soc.* **2011,** *133* (28), 10784-10786.

136. Thompson, R. E.; Chan, B.; Radom, L.; Jolliffe, K. A.; Payne, R. J., Chemoselective Peptide Ligation–Desulfurization at Aspartate. *Angew. Chem., Int. Ed.* **2013,** *52* (37), 9723-9727.

137. Chen, J.; Wang, P.; Zhu, J.; Wan, Q.; Danishefsky, S. J., A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. *Tetrahedron* **2010,** *66* (13), 2277-2283.

138. Siman, P.; Karthikeyan, S. V.; Brik, A., Native Chemical Ligation at Glutamine. *Org. Lett.* **2012,** *14* (6), 1520-1523.

139. Pasunooti, K. K.; Yang, R.; Vedachalam, S.; Gorityala, B. K.; Liu, C.-F.; Liu, X.-W., Synthesis of 4-mercapto-l-lysine derivatives: Potential building blocks for sequential native chemical ligation. *Bioorg. Med. Chem. Lett.* **2009,** *19* (22), 6268-6271.

140. Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X.-W.; Liu, C.-F., Dual native chemical ligation at lysine. *J. Am. Chem. Soc.* **2009,** *131* (38), 13592-13593.

141. Ajish Kumar, K. S.; Haj-Yahya, M.; Olschewski, D.; Lashuel, H. A.; Brik, A., Highly Efficient and Chemoselective Peptide Ubiquitylation. *Angew. Chem., Int. Ed.* **2009,** *48* (43), 8090-8094.

142. Okamoto, R.; Souma, S.; Kajihara, Y., Efficient Substitution Reaction from Cysteine to the Serine Residue of Glycosylated Polypeptide: Repetitive Peptide Segment Ligation Strategy and the Synthesis of Glycosylated Tetracontapeptide Having Acid Labile Sialyl-TN Antigens. *J. Org. Chem.* **2009,** *74* (6), 2494-2501.

143. Tsukiji, S.; Nagamune, T., Sortase-mediated ligation: a gift from Gram-positive bacteria to protein engineering. *Chembiochem* **2009,** *10* (5), 787-98.

144. Popp, M. W.; Antos, J. M.; Grotenbreg, G. M.; Spooner, E.; Ploegh, H. L., Sortagging: a versatile method for protein labeling. *Nat. Chem. Biol.* **2007,** *3* (11), 707-8.

145. Zong, Y.; Bice, T. W.; Ton-That, H.; Schneewind, O.; Narayana, S. V. L., Crystal Structures of Staphylococcus aureus Sortase A and Its Substrate Complex*. *J. Biol. Chem.* **2004,** *279* (30), 31383-31389.

146. Wu, B.; Wijma, H. J.; Song, L.; Rozeboom, H. J.; Poloni, C.; Tian, Y.; Arif, M. I.; Nuijens, T.; Quaedflieg, P. J. L. M.; Szymanski, W.; Feringa, B. L.; Janssen, D. B., Versatile Peptide C-Terminal Functionalization via a Computationally Engineered Peptide Amidase. *ACS Catal.* **2016,** *6* (8), 5405-5414.

147. Shogren-Knaak, M.; Ishii, H.; Sun, J. M.; Pazin, M. J.; Davie, J. R.; Peterson, C. L., Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **2006,** *311* (5762), 844-7.

148. McGinty, R. K.; Kim, J.; Chatterjee, C.; Roeder, R. G.; Muir, T. W., Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. *Nature* **2008,** *453* (7196), 812.

149. Chu, N.; Salguero, A. L.; Liu, A. Z.; Chen, Z.; Dempsey, D. R.; Ficarro, S. B.; Alexander, W. M.; Marto, J. A.; Li, Y.; Amzel, L. M., Akt kinase activation mechanisms revealed using protein semisynthesis. *Cell* **2018,** *174* (4), 897-907. e14.

150. Schwarzer, D.; Cole, P. A., Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr. Opin. Chem. Biol.* **2005,** *9* (6), 561-569.

151. Pickin, K. A.; Chaudhury, S.; Dancy, B. C.; Gray, J. J.; Cole, P. A., Analysis of protein kinase autophosphorylation using expressed protein ligation and computational modeling. *J. Am. Chem. Soc.* **2008,** *130* (17), 5667-5669.

152. Dann, G. P.; Liszczak, G. P.; Bagert, J. D.; Müller, M. M.; Nguyen, U. T.; Wojcik, F.; Brown, Z. Z.; Bos, J.; Panchenko, T.; Pihl, R., ISWI chromatin remodellers sense nucleosome modifications to determine substrate preference. *Nature* **2017,** *548* (7669), 607.

153. Rak, A.; Pylypenko, O.; Durek, T.; Watzke, A.; Kushnir, S.; Brunsveld, L.; Waldmann, H.; Goody, R. S.; Alexandrov, K., Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase. *Science* **2003,** *302* (5645), 646-650.

154. Stevens, A. J.; Sekar, G.; Shah, N. H.; Mostafavi, A. Z.; Cowburn, D.; Muir, T. W., A promiscuous split intein with expanded protein engineering applications. *Proc. Natl. Acad. Sci. U. S. A.* **2017,** *114* (32), 8538-8543.

155. Amitai, G.; Callahan, B. P.; Stanger, M. J.; Belfort, G.; Belfort, M., Modulation of intein activity by its neighboring extein substrates. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (27), 11005-11010.

156. Oeemig, J. S.; Zhou, D.; Kajander, T.; Wlodawer, A.; Iwaï, H., NMR and crystal structures of the Pyrococcus horikoshii RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *J. Mol. Biol.* **2012,** *421* (1), 85-99.

157. Evans, T. C.; Benner, J.; Xu, M.-Q., The in vitro ligation of bacterially expressed proteins using an intein from Methanobacterium thermoautotrophicum. *J. Biol. Chem.* **1999,** *274* (7), 3923-3926.

158. Henager, S. H.; Chu, N.; Chen, Z.; Bolduc, D.; Dempsey, D. R.; Hwang, Y.; Wells, J.; Cole, P. A., Enzyme-catalyzed expressed protein ligation. *Nat. Methods* **2016,** *13* (11), 925.

159. Okamoto, R.; Morooka, K.; Kajihara, Y., A synthetic approach to a peptide α-thioester from an unprotected peptide through cleavage and activation of a specific peptide bond by N-acetylguanidine. *Angew. Chem., Int. Ed.* **2012,** *51* (1), 191-6.

160. Miyajima, R.; Tsuda, Y.; Inokuma, T.; Shigenaga, A.; Imanishi, M.; Futaki, S.; Otaka, A., Preparation of peptide thioesters from naturally occurring sequences using reaction sequence consisting of regioselective S-cyanylation and hydrazinolysis. *Biopolymers* **2016,** *106* (4), 531-46.

161. Rimington, C., The relation between cystine yield and total sulphur in wool. *Biochem. J.* **1929,** *23* (1), 41.

162. Farnworth, A.; Speakman, J., Reactivity of the Sulphur Linkage in Wool. *Nature* **1949,** *163* (4151), 798.

163. Wood, J. L.; Catsimpoolas, N., Cleavage of the peptide bond at the cystine amino group by the action of cyanide. *J. Biol. Chem.* **1963,** *238* (8), PC2887-PC2888.

164. Catsimpoolas, N.; Wood, J. L., Specific cleavage of cystine peptides by cyanide. *J. Biol. Chem.* **1966,** *241* (8), 1790-1796.

165. Patchornik, A.; Degani, Y.; Neumann, H., Selective cyanylation of sulfhydryl groups. *J. Am. Chem. Soc.* **1970,** *92* (23), 6969-6971.

166. Komander, D.; Rape, M., The ubiquitin code. *Annu. Rev. Biochem.* **2012,** *81*, 203-229.

167. Shafer, D. E.; Toll, B.; Schuman, R. F.; Nelson, B. L.; Mond, J. J.; Lees, A., Activation of soluble polysaccharides with 1-cyano-4-dimethylaminopyridinium tetrafluoroborate (CDAP) for use in protein-polysaccharide conjugate vaccines and immunological reagents. II. Selective crosslinking of proteins to CDAP-activated polysaccharides. *Vaccine* **2000,** *18* (13), 1273-81.

168. Swatek, K. N.; Komander, D., Ubiquitin modifications. *Cell Res.* **2016,** *26* (4), 399.

169. Cappadocia, L.; Lima, C. D., Ubiquitin-like protein conjugation: structures, chemistry, and mechanism. *Chem. Rev.* **2018,** *118* (3), 889-918.

170. van der Veen, A. G.; Ploegh, H. L., Ubiquitin-like proteins. *Annu. Rev. Biochem.* **2012,** *81*, 323-357.

171. Pruneda, J. N.; Durkin, C. H.; Geurink, P. P.; Ovaa, H.; Santhanam, B.; Holden, D. W.; Komander, D., The molecular basis for ubiquitin and ubiquitin-like specificities in bacterial effector proteases. *Mol. Cell* **2016,** *63* (2), 261-276.

172. Sommer, S.; Weikart, N. D.; Linne, U.; Mootz, H. D., Covalent inhibition of SUMO and ubiquitin-specific cysteine proteases by an in situ thiol–alkyne addition. *Bioorg. Med. Chem.* **2013,** *21* (9), 2511-2517.

173. Paudel, P.; Zhang, Q.; Leung, C.; Greenberg, H. C.; Guo, Y.; Chern, Y.-H.; Dong, A.; Li, Y.; Vedadi, M.; Zhuang, Z., Crystal structure and activity-based labeling reveal the mechanisms for linkage-specific substrate recognition by deubiquitinase USP9X. *Proc. Natl. Acad. Sci. U. S. A.* **2019,** *116* (15), 7288-7297.

174. Basters, A.; Knobeloch, K. P.; Fritz, G., How USP18 deals with ISG15-modified proteins: structural basis for the specificity of the protease. *FEBS J.* **2018,** *285* (6), 1024-1029.

175. Catic, A.; Fiebiger, E.; Korbel, G. A.; Blom, D.; Galardy, P. J.; Ploegh, H. L., Screen for ISG15-crossreactive deubiquitinases. *PloS one* **2007,** *2* (7), e679.

176. Dang, L. C.; Melandri, F. D.; Stein, R. L., Kinetic and mechanistic studies on the hydrolysis of ubiquitin C-terminal 7-amido-4-methylcoumarin by deubiquitinating enzymes. *Biochemistry* **1998,** *37* (7), 1868-1879.

177. Basu, A.; Rose, K. L.; Zhang, J.; Beavis, R. C.; Ueberheide, B.; Garcia, B. A.; Chait, B.; Zhao, Y.; Hunt, D. F.; Segal, E., Proteome-wide prediction of acetylation substrates. *Proc. Natl. Acad. Sci. U. S. A.* **2009,** *106* (33), 13785-13790.

178. Frouws, T. D.; Barth, P. D.; Richmond, T. J., Site-Specific disulfide crosslinked nucleosomes with enhanced stability. *J. Mol. Biol.* **2018,** *430* (1), 45-57.

179. Matsumoto, S.; Cavadini, S.; Bunker, R. D.; Grand, R. S.; Potenza, A.; Rabl, J.; Yamamoto, J.; Schenk, A. D.; Schübeler, D.; Iwai, S., DNA damage detection in nucleosomes involves DNA register shifting. *Nature* **2019,** *571* (7763), 79-84.

180. Nowotny, M.; Gaidamakov, S. A.; Crouch, R. J.; Yang, W., Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis. *Cell* **2005,** *121* (7), 1005-1016.

181. Zheng, J.-S.; Tang, S.; Qi, Y.-K.; Wang, Z.-P.; Liu, L., Chemical synthesis of proteins using peptide hydrazides as thioester surrogates. *Nat. Protoc.* **2013,** *8* (12), 2483.

182. Raufman, J.-P., Bioactive peptides from lizard venoms. *Regul. Pept.* **1996,** *61* (1), 1-18.

183. Wu, N.; Deiters, A.; Cropp, T. A.; King, D.; Schultz, P. G., A genetically encoded photocaged amino acid. *J. Am. Chem. Soc.* **2004,** *126* (44), 14306-14307.

184. Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W., Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* **2014,** *136* (6), 2240-2243.

185. Klemm, T.; Ebert, G.; Calleja, D. J.; Allison, C. C.; Richardson, L. W.; Bernardini, J. P.; Lu, B. G.; Kuchel, N. W.; Grohmann, C.; Shibata, Y.; Gan, Z. Y.; Cooney, J. P.; Doerflinger, M.; Au, A. E.; Blackmore, T. R.; van der Heden van Noort, G. J.; Geurink, P. P.; Ovaa, H.; Newman, J.; Riboldi-Tunnicliffe, A.; Czabotar, P. E.;

Mitchell, J. P.; Feltham, R.; Lechtenberg, B. C.; Lowes, K. N.; Dewson, G.; Pellegrini, M.; Lessene, G.; Komander, D., Mechanism and inhibition of the papain-like protease, PLpro, of SARS-CoV-2. *EMBO J.* **2020,** *39* (18), e106275.

186. Ward, J. A.; McLellan, L.; Stockley, M.; Gibson, K. R.; Whitlock, G. A.; Knights, C.; Harrigan, J. A.; Jacq, X.; Tate, E. W., Quantitative chemical proteomic profiling of ubiquitin specific proteases in intact cancer cells. *ACS Chem. Bio.* **2016,** *11* (12), 3268-3272.

187. Chuh, K. N.; Batt, A. R.; Pratt, M. R., Chemical methods for encoding and decoding of posttranslational modifications. *Cell Chem. Biol.* **2016,** *23* (1), 86-107.

188. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776-9.

189. Neumann, H.; Peak-Chew, S. Y.; Chin, J. W., Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. *Nat. Chem. Biol.* **2008,** *4* (4), 232-4.

190. Nguyen, D. P.; Mahesh, M.; Elsässer, S. J.; Hancock, S. M.; Uttamapinant, C.; Chin, J. W., Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* **2014,** *136* (6), 2240-3.

191. Lee, Y. J.; Wu, B.; Raymond, J. E.; Zeng, Y.; Fang, X.; Wooley, K. L.; Liu, W. R., A genetically encoded acrylamide functionality. *ACS Chem Biol* **2013,** *8* (8), 1664-70.

192. Wang, Z. A.; Kurra, Y.; Wang, X.; Zeng, Y.; Lee, Y. J.; Sharma, V.; Lin, H.; Dai, S. Y.; Liu, W. R., A Versatile Approach for Site-Specific Lysine Acylation in Proteins. *Angew. Chem., Int. Ed.* **2017,** *56* (6), 1643-1647.

193. Wang, Z. A.; Zeng, Y.; Kurra, Y.; Wang, X.; Tharp, J. M.; Vatansever, E. C.; Hsu, W. W.; Dai, S.; Fang, X.; Liu, W. R., A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation. *Angew. Chem., Int. Ed.* **2017,** *56* (1), 212-216.

194. Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M., The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007,** *128* (5), 1003-12.

195. Wright, T. H.; Bower, B. J.; Chalker, J. M.; Bernardes, G. J.; Wiewiora, R.; Ng, W. L.; Raj, R.; Faulkner, S.; Vallee, M. R.; Phanumartwiwath, A.; Coleman, O. D.; Thezenas, M. L.; Khan, M.; Galan, S. R.; Lercher, L.; Schombs, M. W.; Gerstberger, S.; Palm-Espling, M. E.; Baldwin, A. J.; Kessler, B. M.; Claridge, T. D.; Mohammed, S.; Davis, B. G., Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **2016,** *354* (6312).

196. Wang, J.; Schiller, S. M.; Schultz, P. G., A Biosynthetic Route to Dehydroalanine-Containing Proteins. *Angew. Chem., Int. Ed.* **2007,** *46* (36), 6849-6851.

197. Wang, Z. U.; Wang, Y. S.; Pai, P. J.; Russell, W. K.; Russell, D. H.; Liu, W. R., A facile method to synthesize histones with posttranslational modification mimics. *Biochemistry* **2012,** *51* (26), 5232-4.

198. Yang, A.; Ha, S.; Ahn, J.; Kim, R.; Kim, S.; Lee, Y.; Kim, J.; Soll, D.; Lee, H. Y.; Park, H. S., A chemical biology route to site-specific authentic protein modifications. *Science* **2016,** *354* (6312), 623-626.

199. Lin, S.; He, D.; Long, T.; Zhang, S.; Meng, R.; Chen, P. R., Genetically encoded cleavable protein photo-cross-linker. *J. Am. Chem. Soc.* **2014,** *136* (34), 11860-3.

200. Yang, B.; Wang, N.; Schnier, P. D.; Zheng, F.; Zhu, H.; Polizzi, N. F.; Ittuveetil, A.; Saikam, V.; DeGrado, W. F.; Wang, Q.; Wang, P. G.; Wang, L., Genetically Introducing Biochemically Reactive Amino Acids Dehydroalanine and Dehydrobutyrine in Proteins. *J. Am. Chem. Soc.* **2019,** *141* (19), 7698-7703.

201. Qiao, Y.; Yu, G.; Kratch, K. C.; Wang, X. A.; Wang, W. W.; Leeuwon, S. Z.; Xu, S.; Morse, J. S.; Liu, W. R., Expressed Protein Ligation Without Intein. *J. Am. Chem. Soc.* **2020,** *142* (15), 7047-7054.

202. Jacobson, G. R.; Schaffer, M. H.; Stark, G. R.; Vanaman, T. C., Specific chemical cleavage in high yield at the amino peptide bonds of cysteine and cystine residues. *J. Biol. Chem.* **1973,** *248* (19), 6583-6591.

203. Degani, Y.; Patchornik, A., Cyanylation of sulfhydryl groups by 2-nitro-5-thiocyanobenzoic acid. High-yield modification and cleavage of peptides at cysteine residues. *Biochemistry* **1974,** *13* (1), 1-11.

204. Tang, H.-Y.; Speicher, D. W., Identification of alternative products and optimization of 2-nitro-5-thiocyanatobenzoic acid cyanylation and cleavage at cysteine residues. *Anal. Biochem.* **2004,** *334* (1), 48-61.

205. Pickart, C. M.; Eddins, M. J., Ubiquitin: structures, functions, mechanisms. *Biochim. Biophys. Acta Mol. Cell Res.* **2004,** *1695* (1-3), 55-72.

206. Hu, H.; Sun, S.-C., Ubiquitin signaling in immune responses. *Cell Res.* **2016,** *26* (4), 457-483.

207. Deng, L.; Meng, T.; Chen, L.; Wei, W.; Wang, P., The role of ubiquitination in tumorigenesis and targeted drug discovery. *Signal Transduct. Target. Ther.* **2020,** *5* (1), 1-28.

208. Hochstrasser, M., Origin and function of ubiquitin-like proteins. *Nature* **2009,** *458* (7237), 422-429.

209. Mali, S. M.; Singh, S. K.; Eid, E.; Brik, A., Ubiquitin signaling: chemistry comes to the rescue. *J. Am. Chem. Soc.* **2017,** *139* (14), 4971-4986.

210. Love, K. R.; Catic, A.; Schlieker, C.; Ploegh, H. L., Mechanisms, biology and inhibitors of deubiquitinating enzymes. *Nat. Chem. Bio.* **2007,** *3* (11), 697.

211. Hewings, D. S.; Flygare, J. A.; Bogyo, M.; Wertz, I. E., Activity‐based probes for the ubiquitin conjugation–deconjugation machinery: new chemistries, new tools, and new insights. *FEBS J.* **2017,** *284* (10), 1555-1576.

212. Witting, K. F.; Mulder, M. P.; Ovaa, H., Advancing our understanding of ubiquitination using the Ub-toolkit. *J. Mol. Biol.* **2017,** *429* (22), 3388-3394.

213. van Wijk, S. J.; Fulda, S.; Dikic, I.; Heilemann, M., Visualizing ubiquitination in mammalian cells. *EMBO Rep.* **2019,** *20* (2), e46520.

214. Mulder, M. P.; Witting, K.; Berlin, I.; Pruneda, J. N.; Wu, K.-P.; Chang, J.-G.; Merkx, R.; Bialas, J.; Groettrup, M.; Vertegaal, A. C., A cascading activity-based probe sequentially targets E1–E2–E3 ubiquitin enzymes. *Nat. Chem. Bio.* **2016,** *12* (7), 523.

215. Wang, W. W.; Zeng, Y.; Wu, B.; Deiters, A.; Liu, W. R., A chemical biology approach to reveal Sirt6-targeted histone H3 sites in nucleosomes. *ACS Chem. Bio.* **2016,** *11* (7), 1973-1981.

216. Rankin, C. A.; Joazeiro, C. A.; Floor, E.; Hunter, T., E3 ubiquitin-protein ligase activity of Parkin is dependent on cooperative interaction of RING finger (TRIAD) elements. *J. Biomed. Sci.* **2001,** *8* (5), 421-9.

217. Seirafi, M.; Kozlov, G.; Gehring, K., Parkin structure and function. *FEBS J.* **2015,** *282* (11), 2076-2088.

218. Arkinson, C.; Walden, H., Parkin function in Parkinson's disease. *Science* **2018,** *360* (6386), 267-268.

APPENDIX A

DETAILED MASS SPECTROMETRY ANALYSIS RESULTS

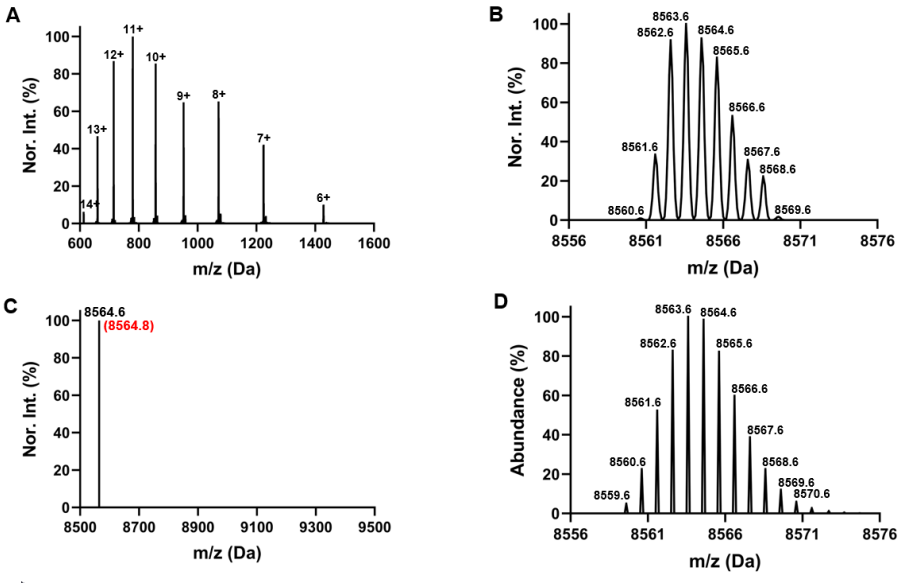

Figure 58 ESI-MS spectra of wild type Ub. (A) raw data, (B) deconvoluted spectra. (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

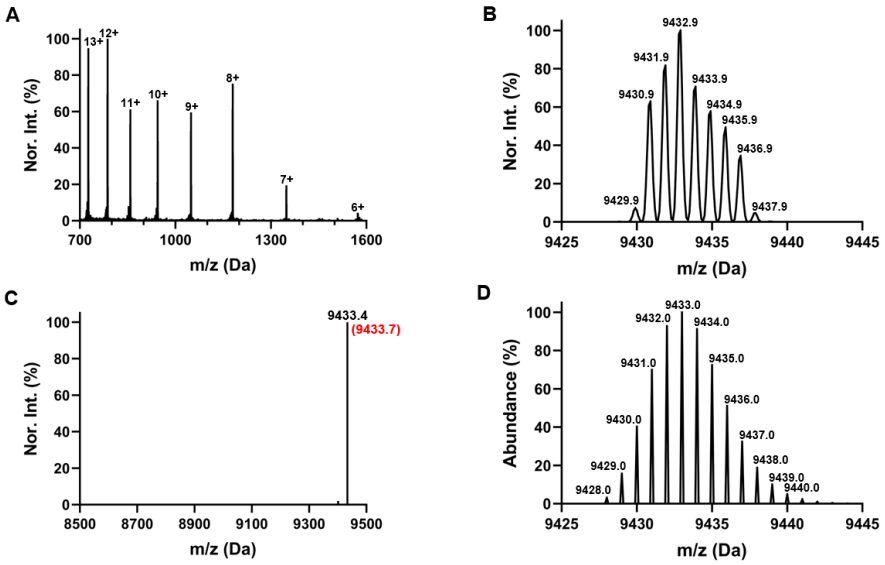

Figure 59 ESI-MS spectra of Ub-G76C-6H. (A) raw data, (B) deconvoluted spectra. (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

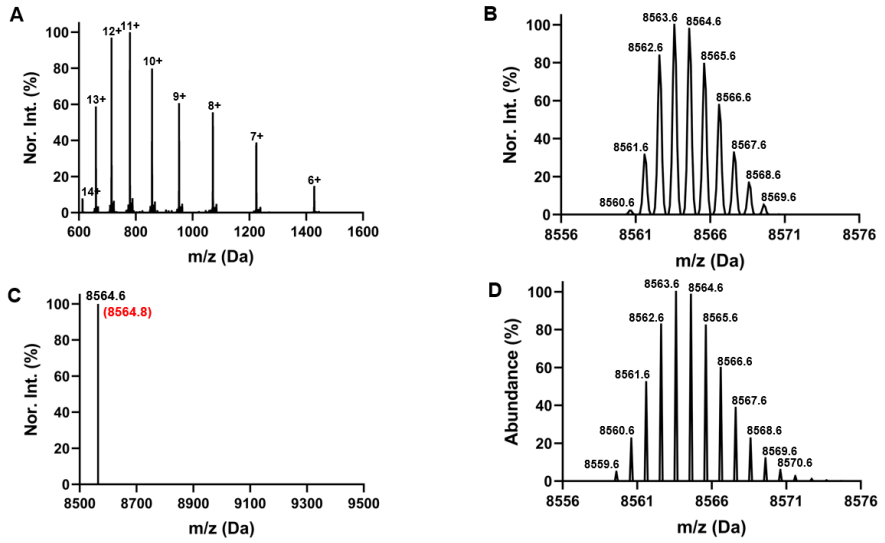

Figure 60 ESI-MS spectra of Ub-G76G. (A) raw data, (B) deconvoluted spectra. (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

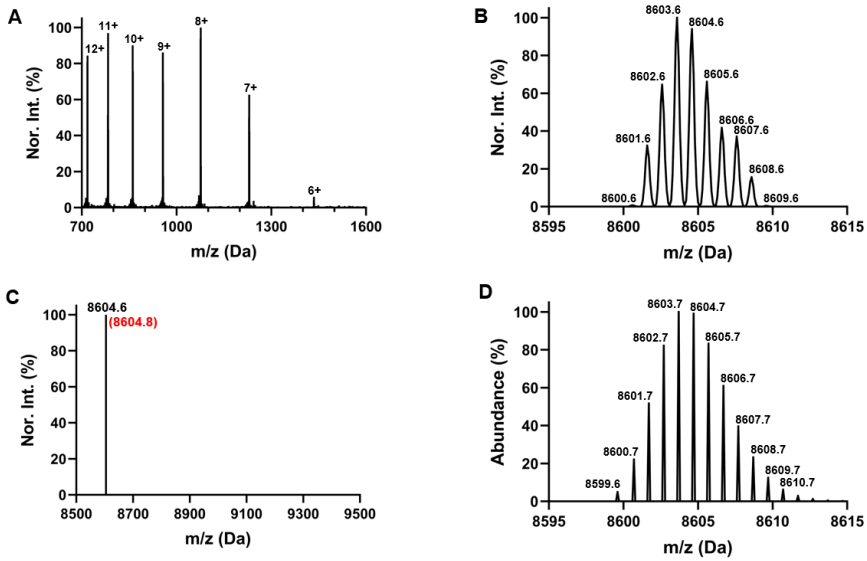

Figure 61 ESI-MS spectra of Ub-G76P. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

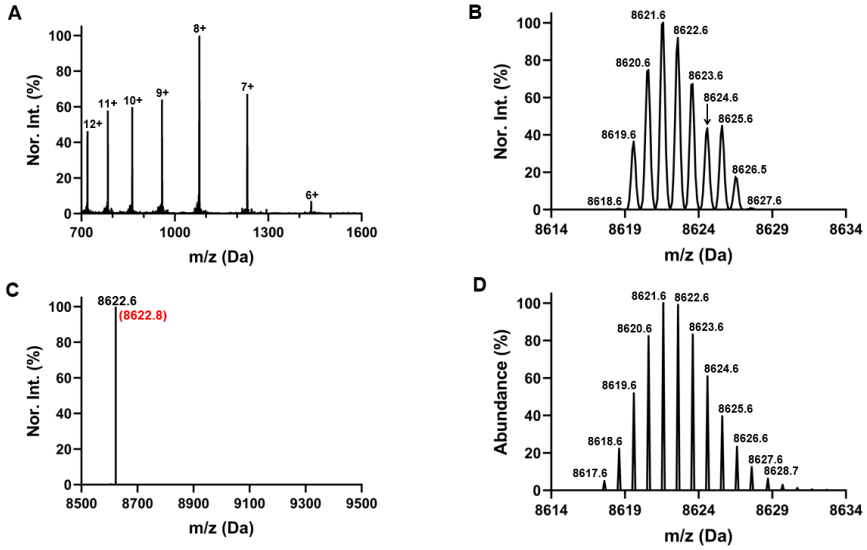

Figure 62 ESI-MS spectra of Ub-G76D. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

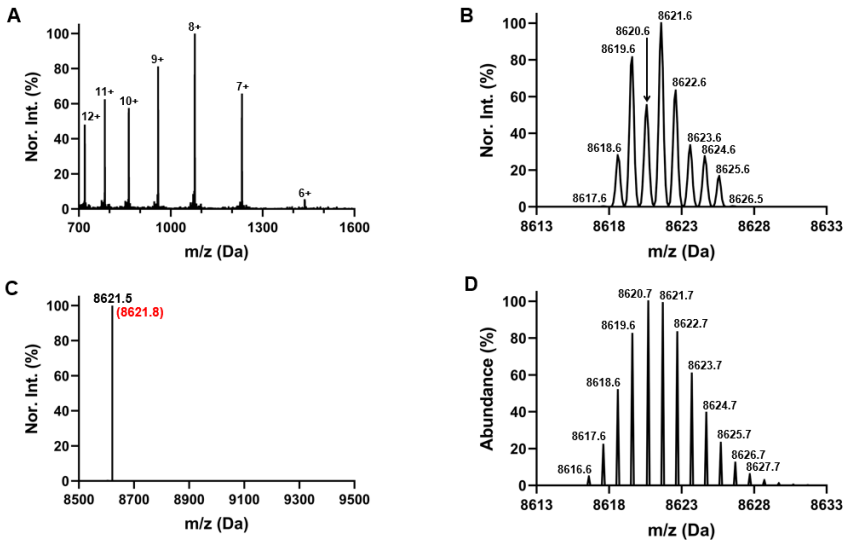

Figure 63 ESI-MS spectra of Ub-G76N. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

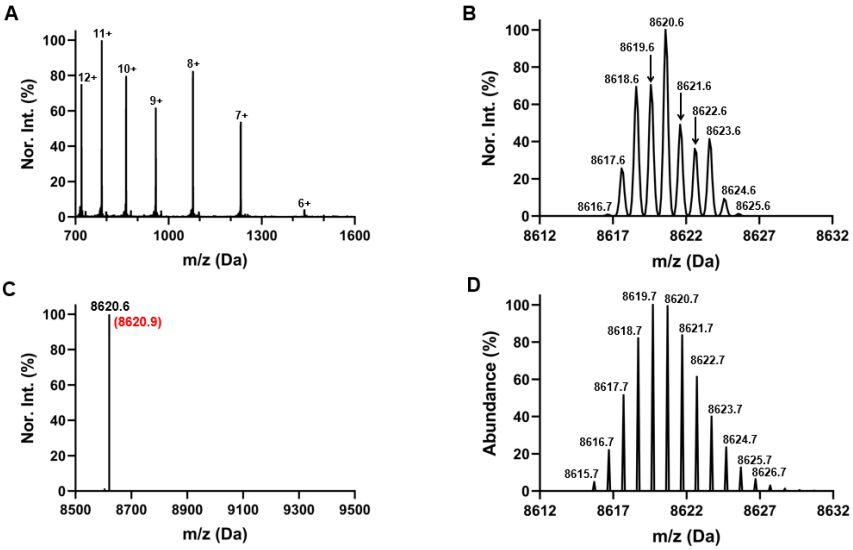

Figure 64 ESI-MS spectra of Ub-G76L. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

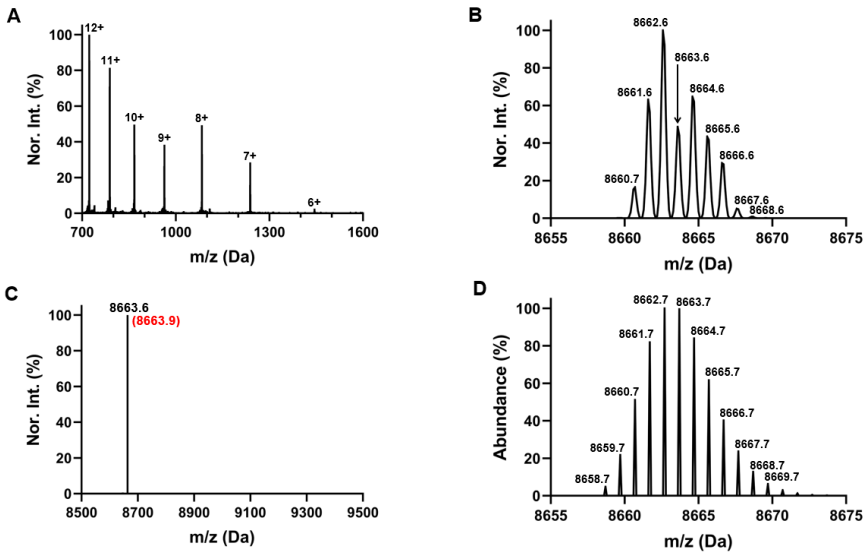

Figure 65 ESI-MS spectra of Ub-G76R. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

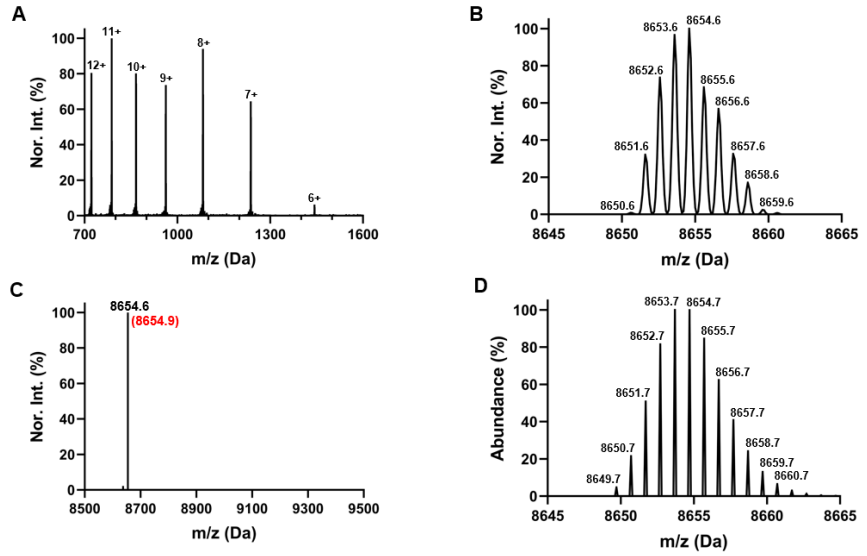

Figure 66 ESI-MS spectra of Ub-G76F. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

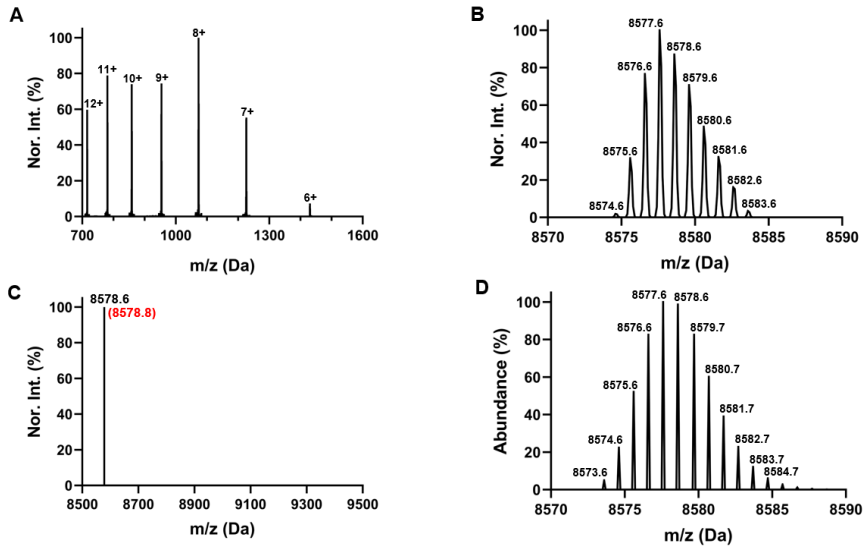

Figure 67 ESI-MS spectra of Ub-G76A$_D$. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

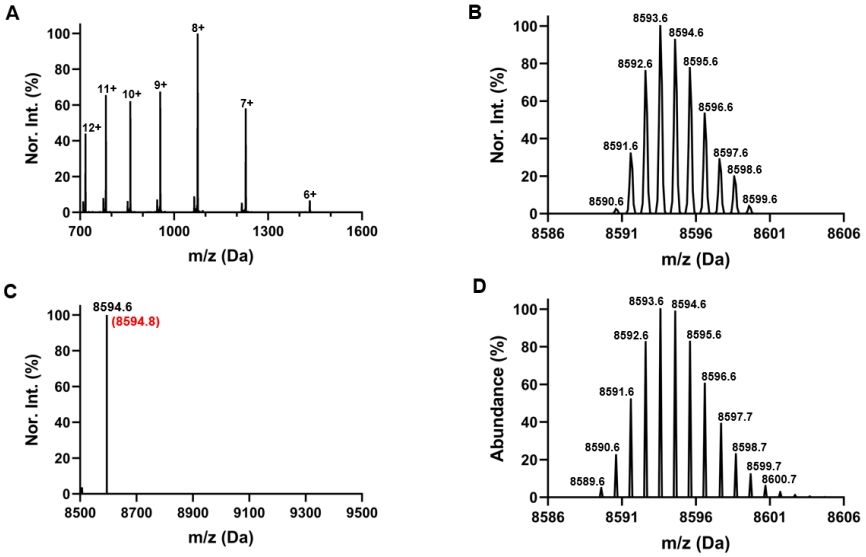

Figure 68 ESI-MS spectra of Ub-G76S$_D$. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

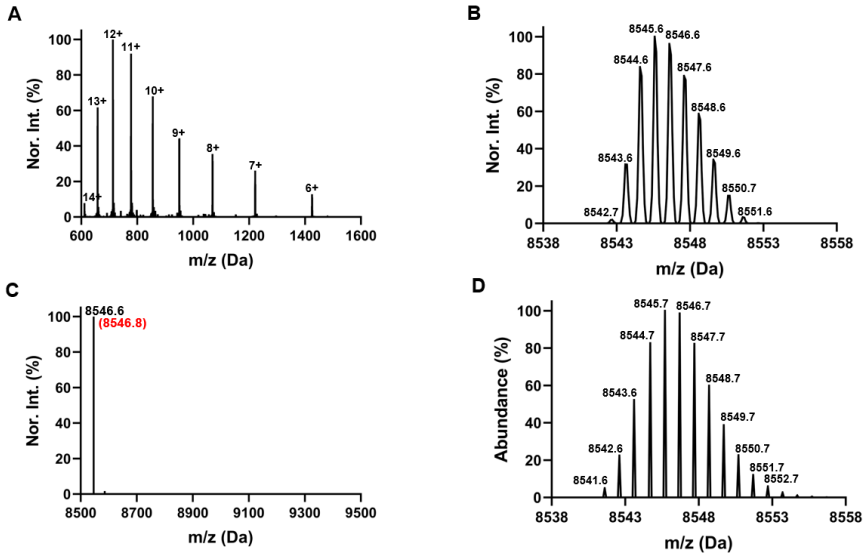

Figure 69 ESI-MS spectra of Ub-G76Aa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

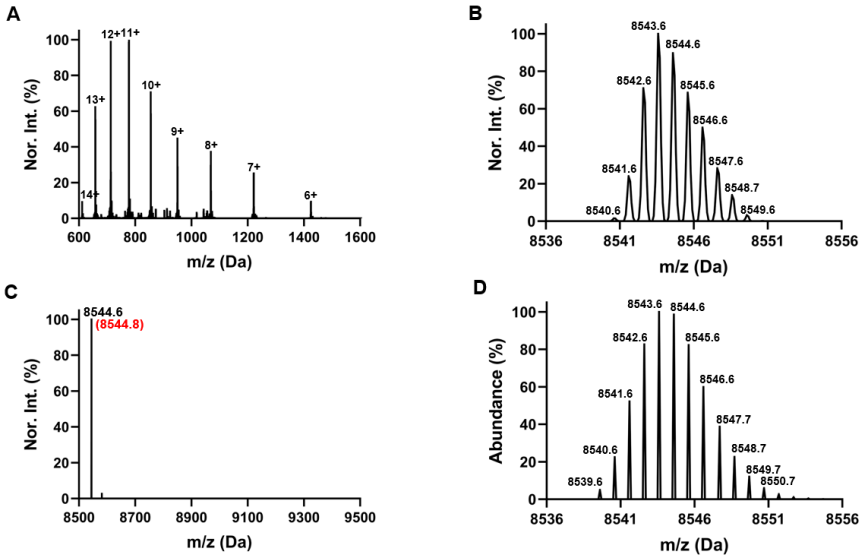

Figure 70 ESI-MS spectra of Ub-G76Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

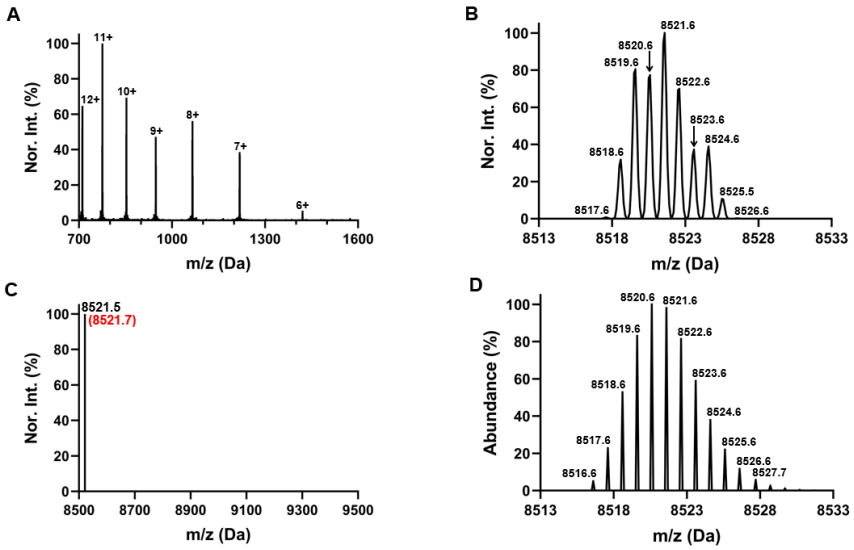

Figure 71 ESI-MS spectra of Ub-G76Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

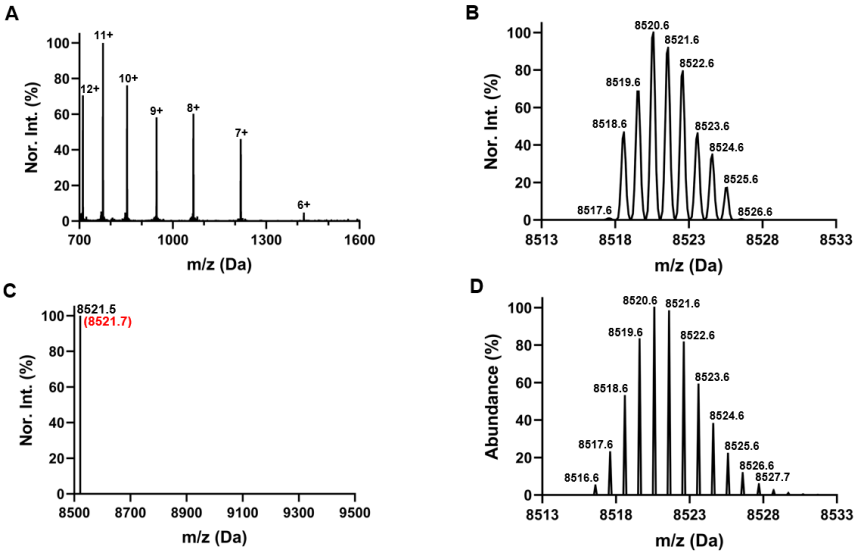

Figure 72 ESI-MS spectra of Ub-G76Ha (native condition). (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

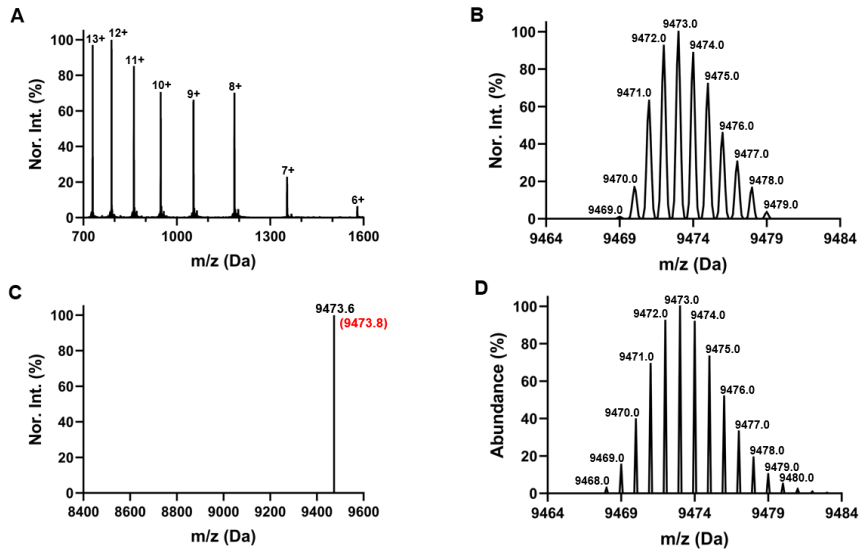

Figure 73 ESI-MS spectra of Ub-G75P/G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

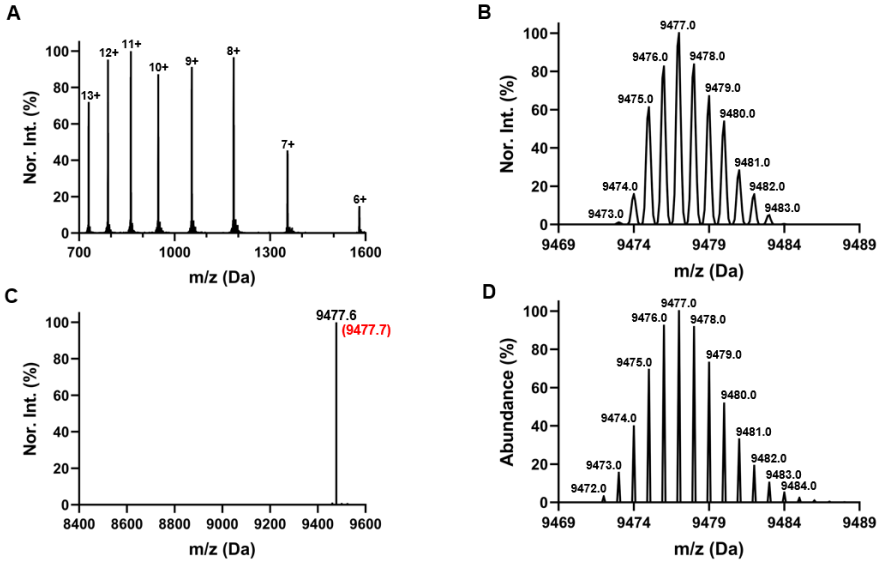

Figure 74 ESI-MS spectra of Ub-G75T/G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

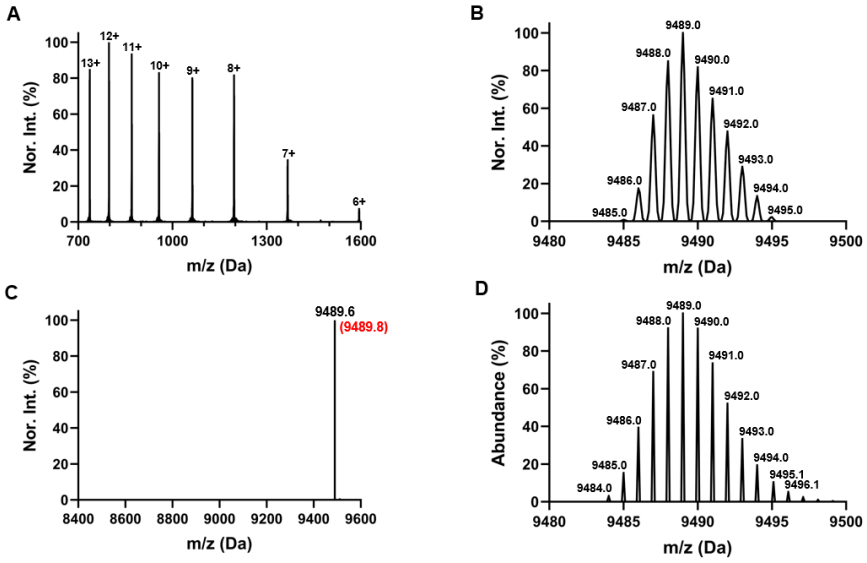

Figure 75 ESI-MS spectra of Ub-G75L/G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

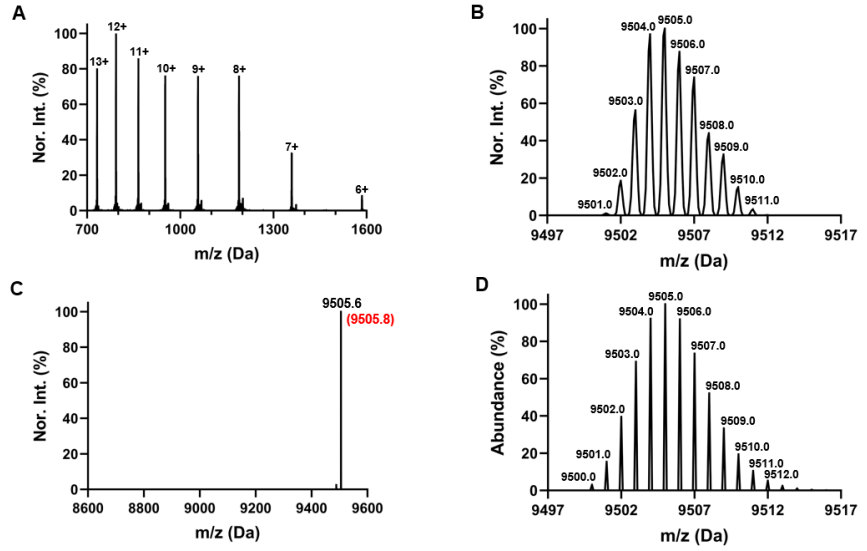

Figure 76 ESI-MS spectra of Ub-G75E/G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

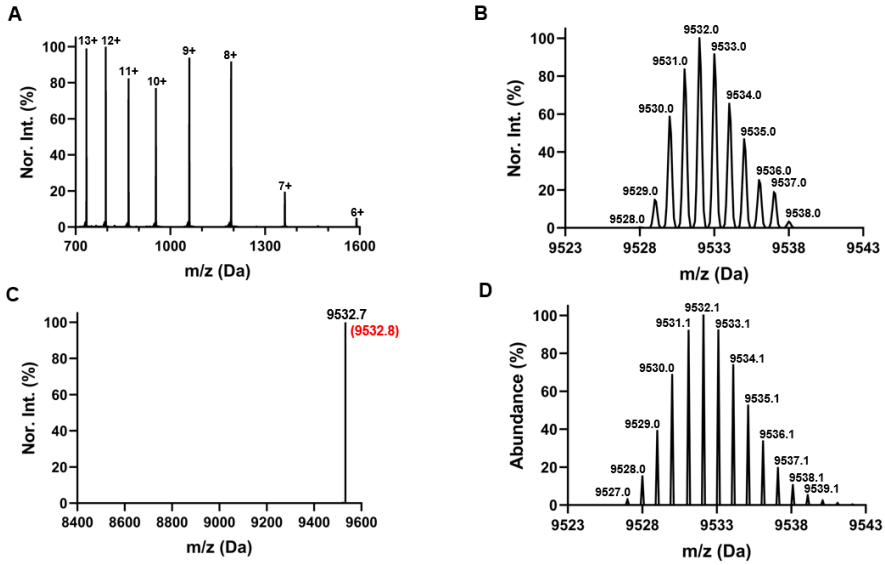

Figure 77 ESI-MS spectra of Ub-G75R/G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

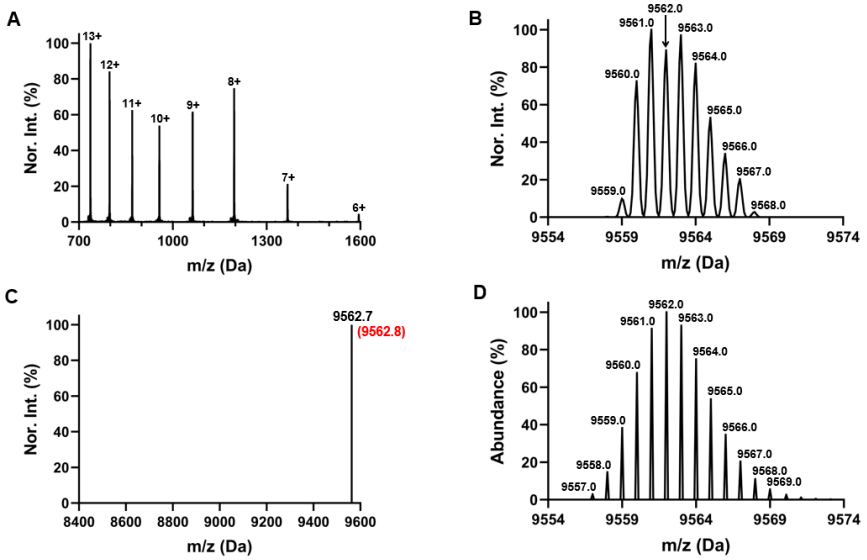

Figure 78 ESI-MS spectra of Ub-G75W/G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

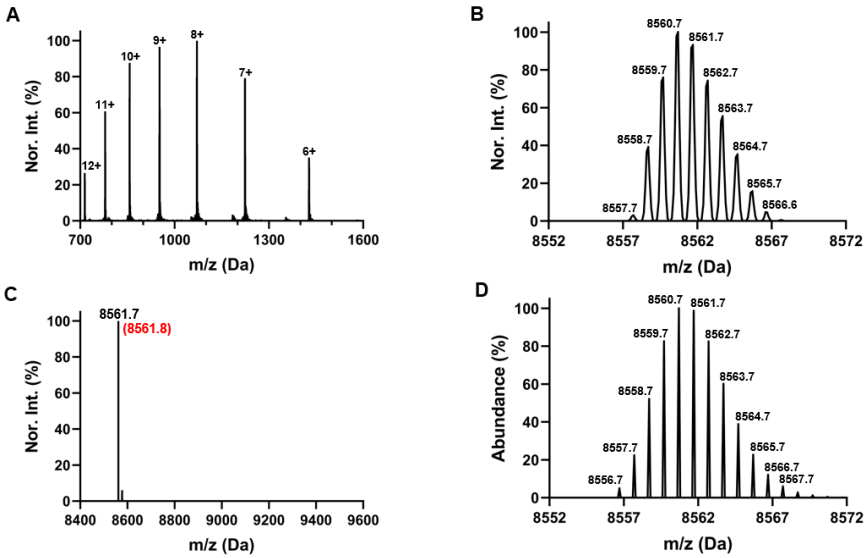

Figure 79 ESI-MS spectra of Ub-G75P/G76Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

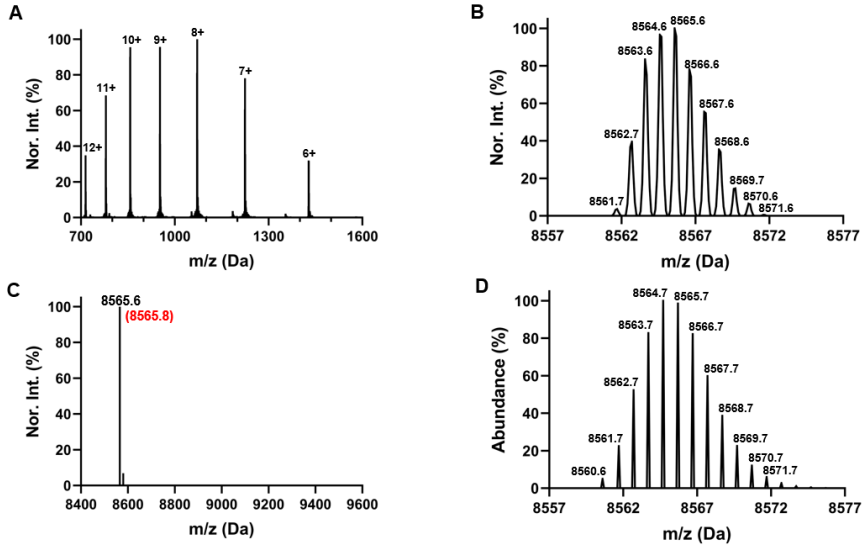

Figure 80 ESI-MS spectra of Ub-G75T/G76Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

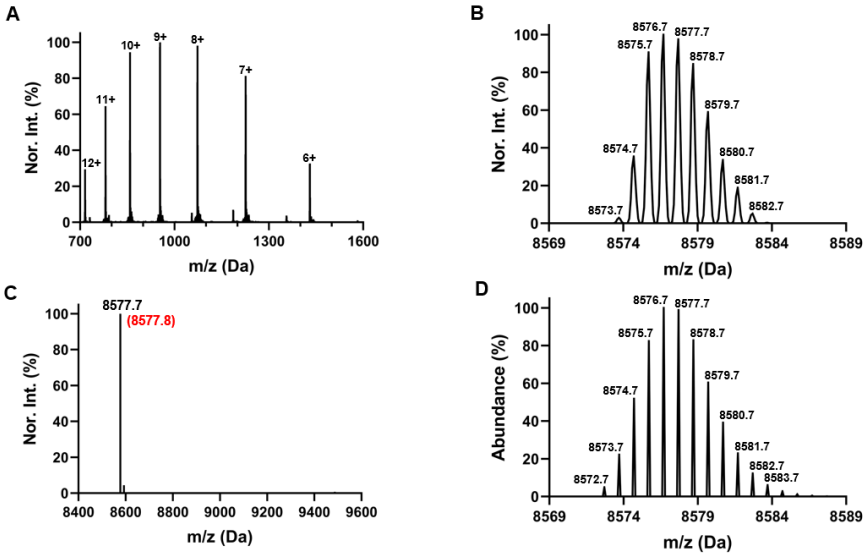

Figure 81 ESI-MS spectra of Ub-G75L/G76Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

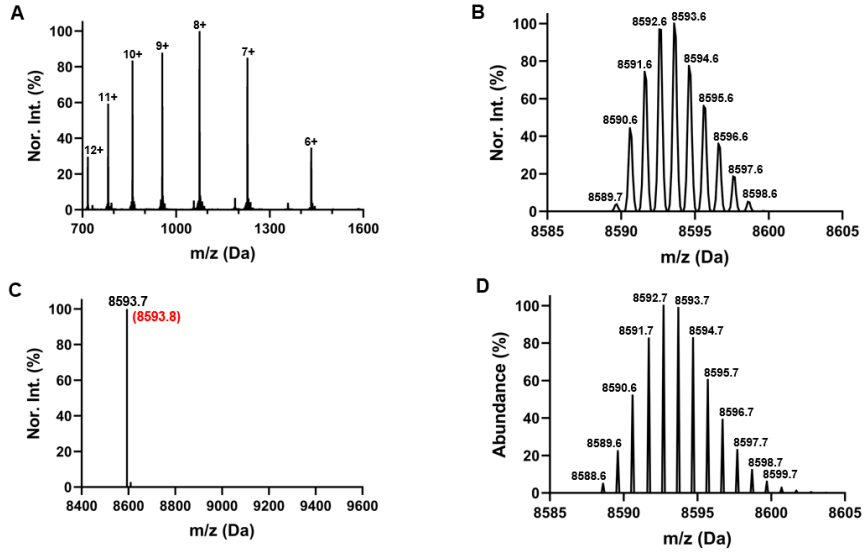

Figure 82 ESI-MS spectra of Ub-G75E/G76Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

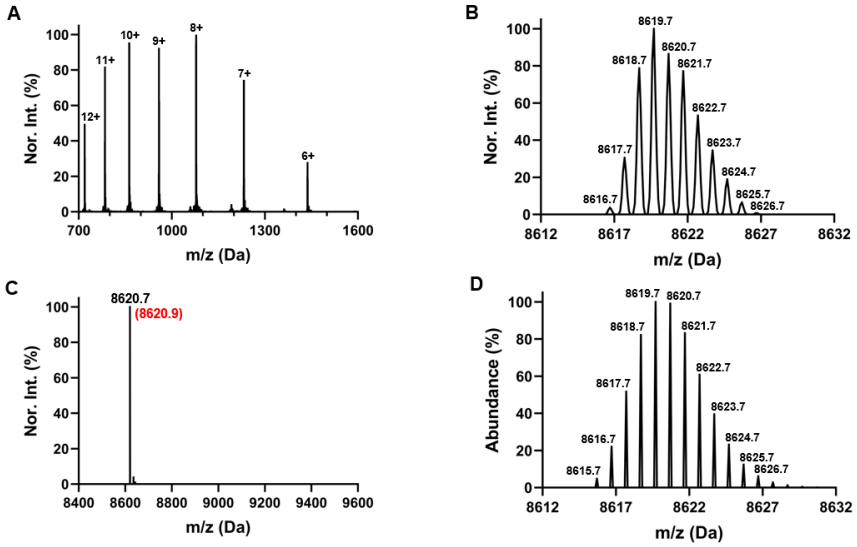

Figure 83 ESI-MS spectra of Ub-G75R/G76Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

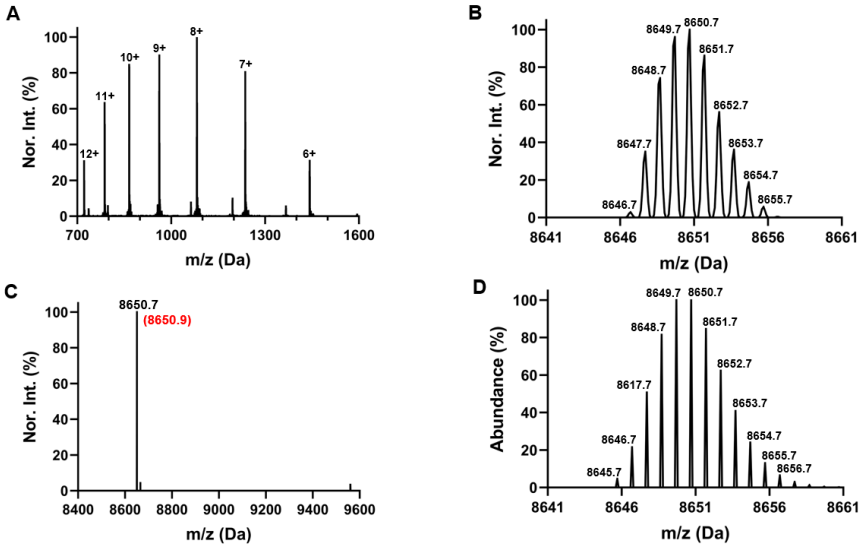

Figure 84 ESI-MS spectra of Ub-G75W/G76Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

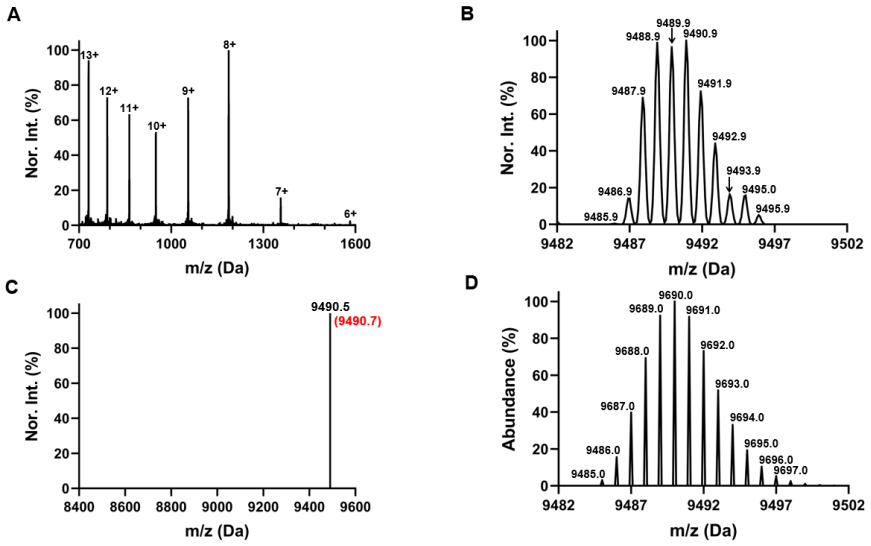

Figure 85 ESI-MS spectra of Ub-C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

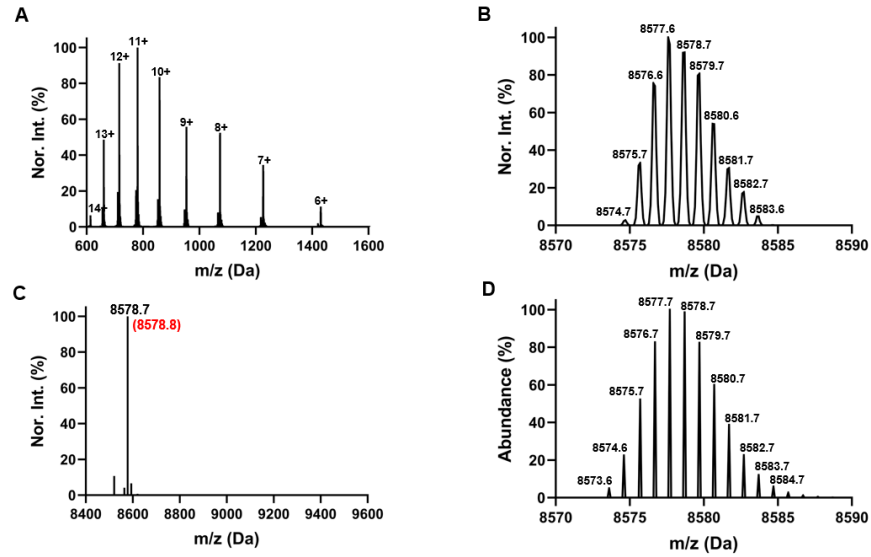

Figure 86 ESI-MS spectra of Ub-Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

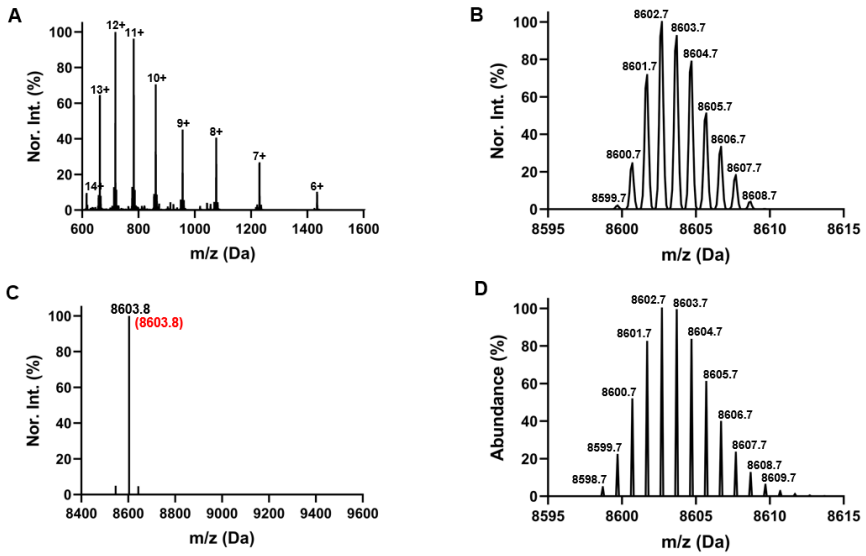

Figure 87 ESI-MS spectra of Ub-Aa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

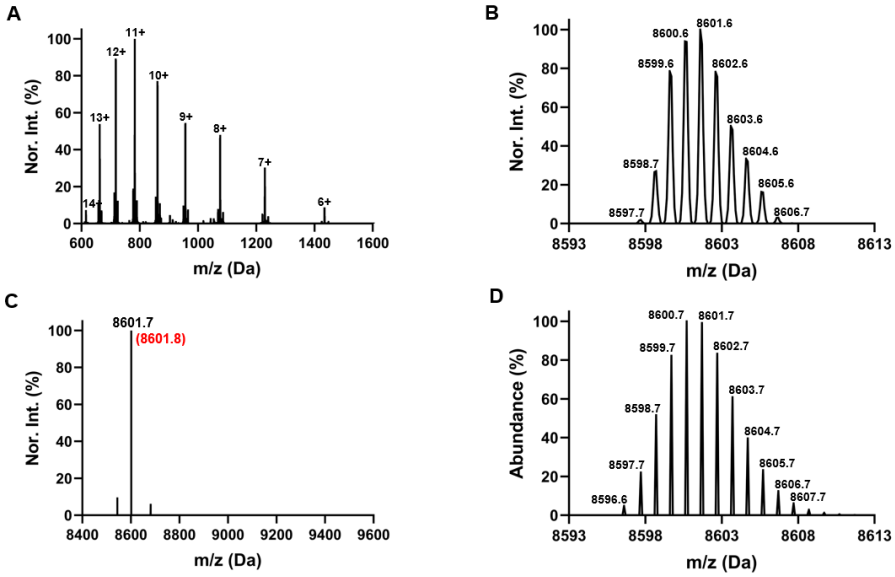

Figure 88 ESI-MS spectra of Ub-Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

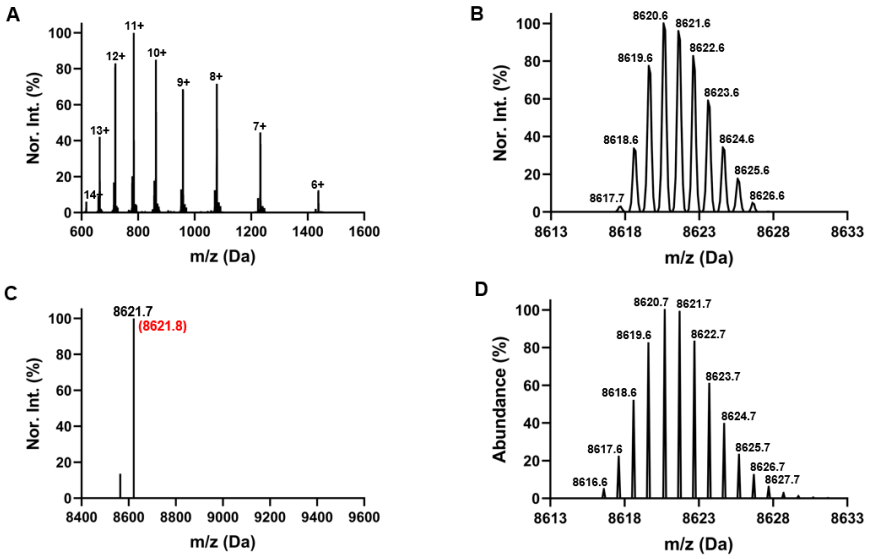

Figure 89 ESI-MS spectra of Ub-G. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

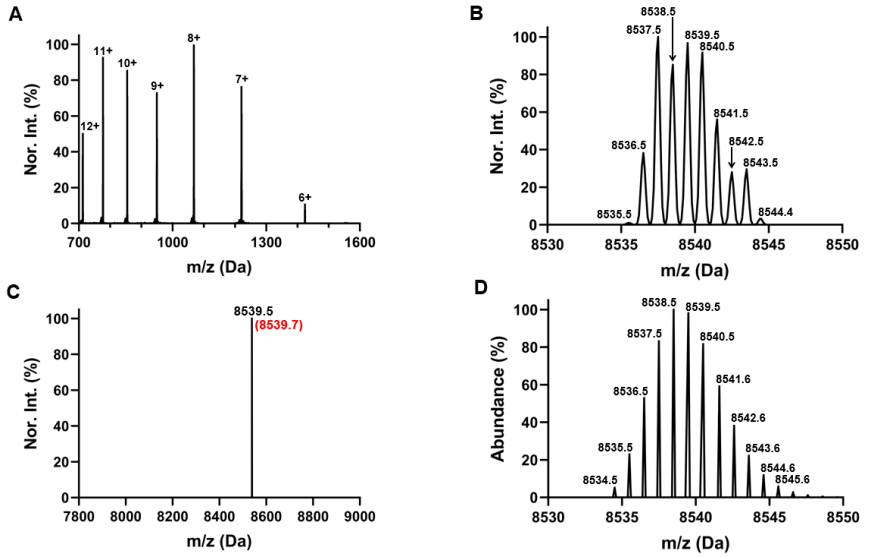

Figure 90 ESI-MS spectra of Ub-K48C. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

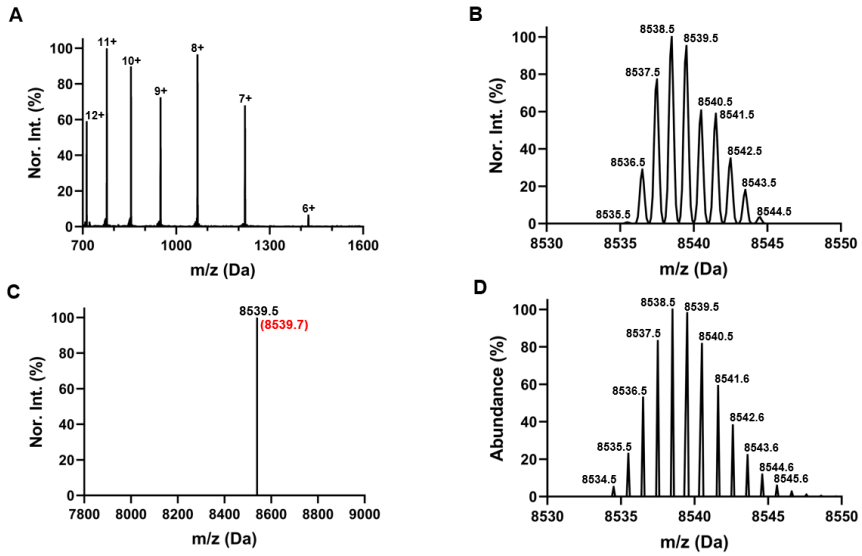

Figure 91 ESI-MS spectra of Ub-K63C. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

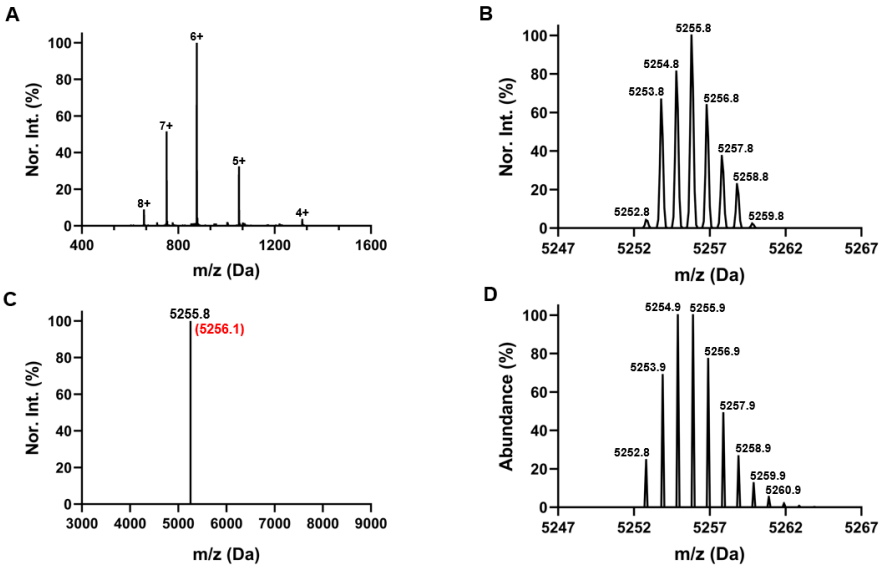

Figure 92 ESI-MS spectra of Ub_{1-47}-Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

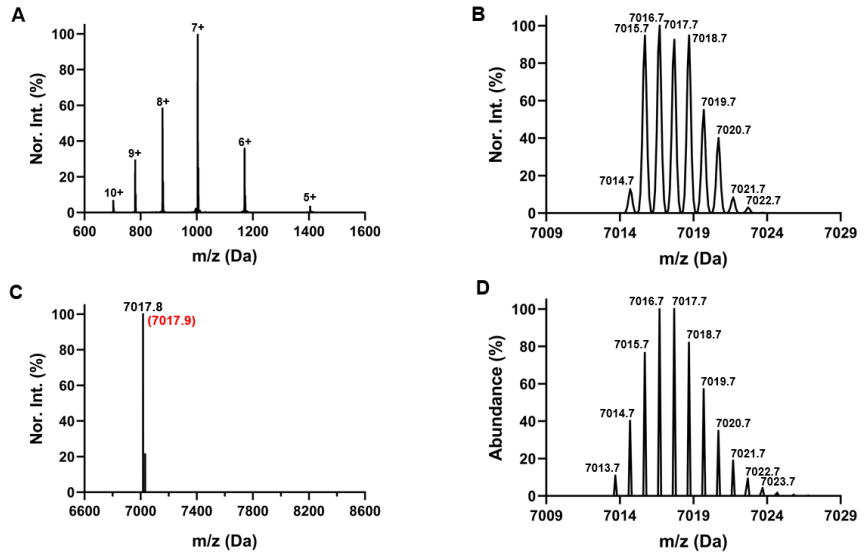

Figure 93 ESI-MS spectra of Ub_{1-62}-Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

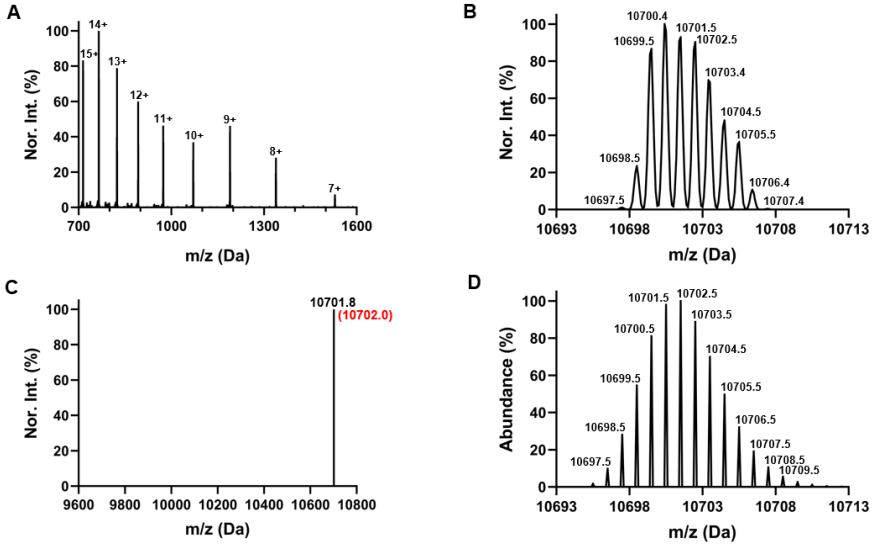

Figure 94 ESI-MS spectra of FLAG-Ub-G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

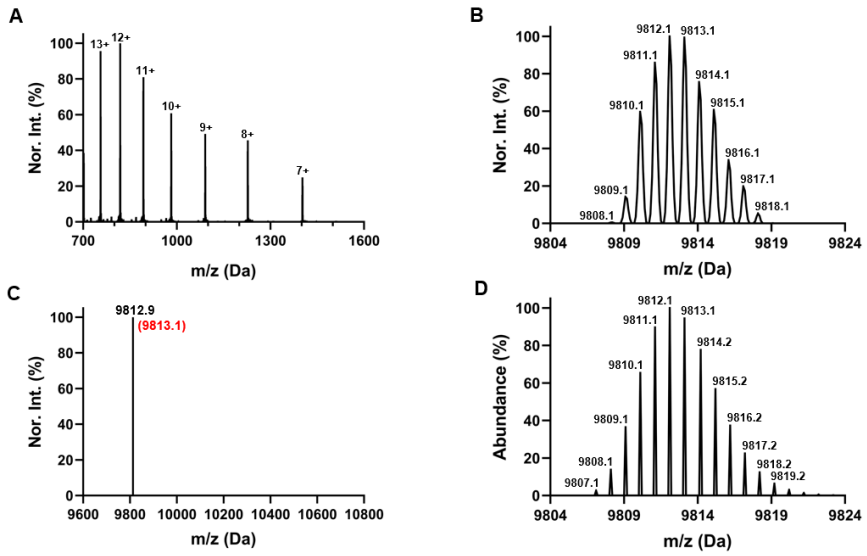

Figure 95 ESI-MS spectra of FLAG-Ub-G76Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

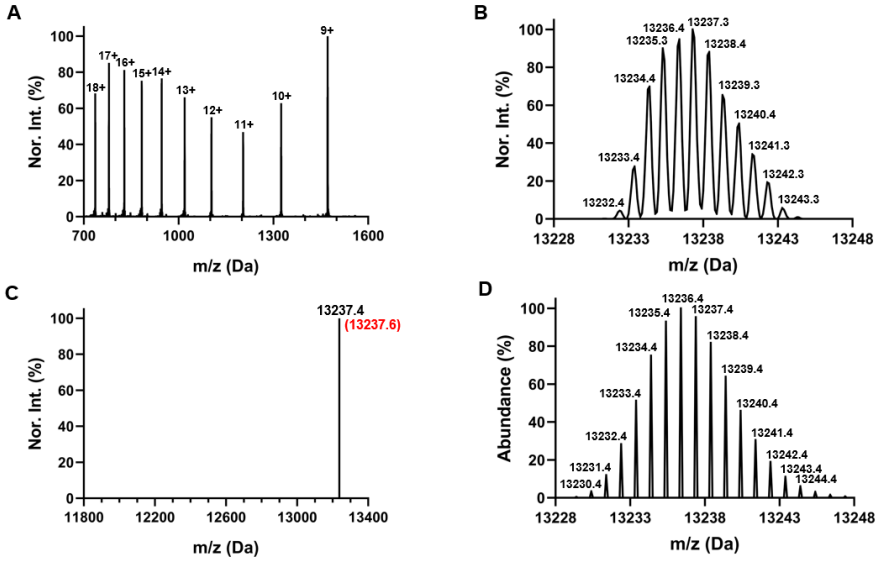

Figure 96 ESI-MS spectra of FLAG-SUMO1-C52A/G97C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

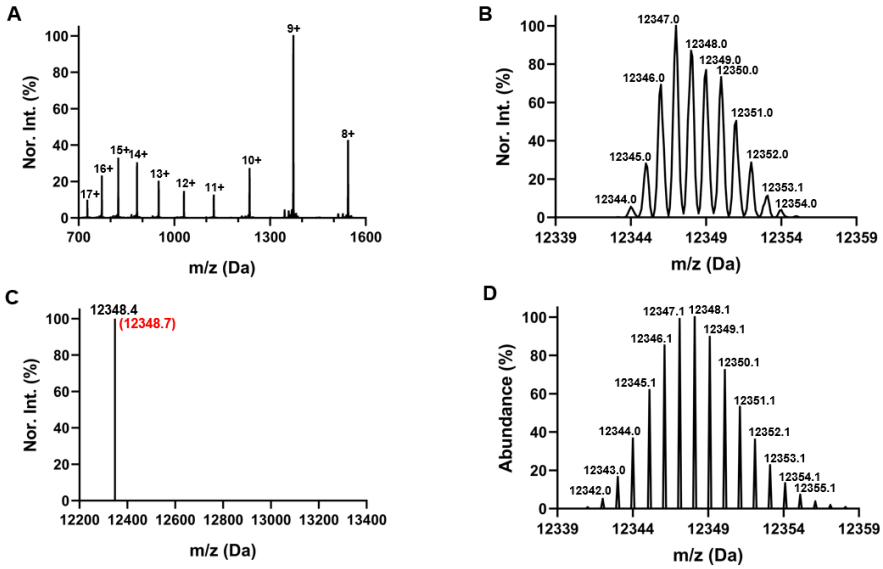

Figure 97 ESI-MS spectra of FLAG-SUMO1-C52A/G97Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

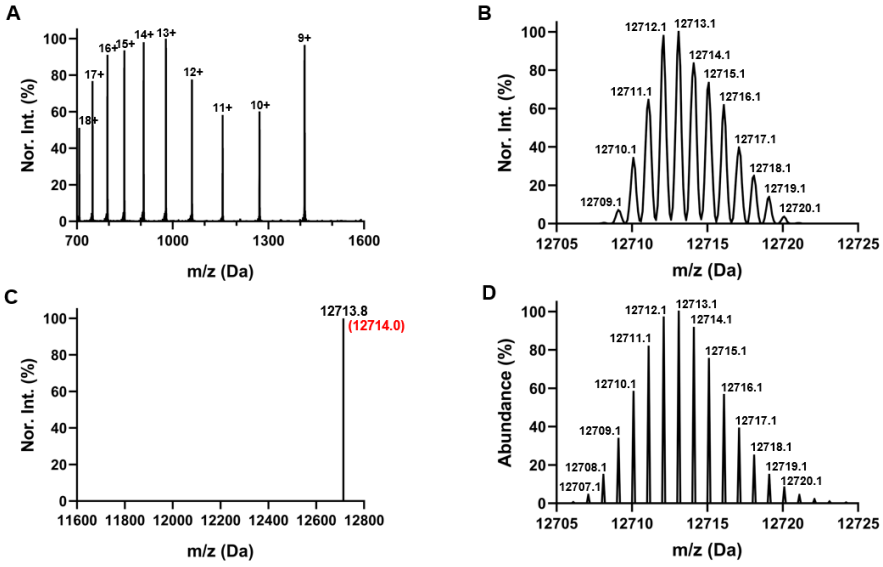

Figure 98 ESI-MS spectra of FLAG-SUMO2-C48A/G93C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

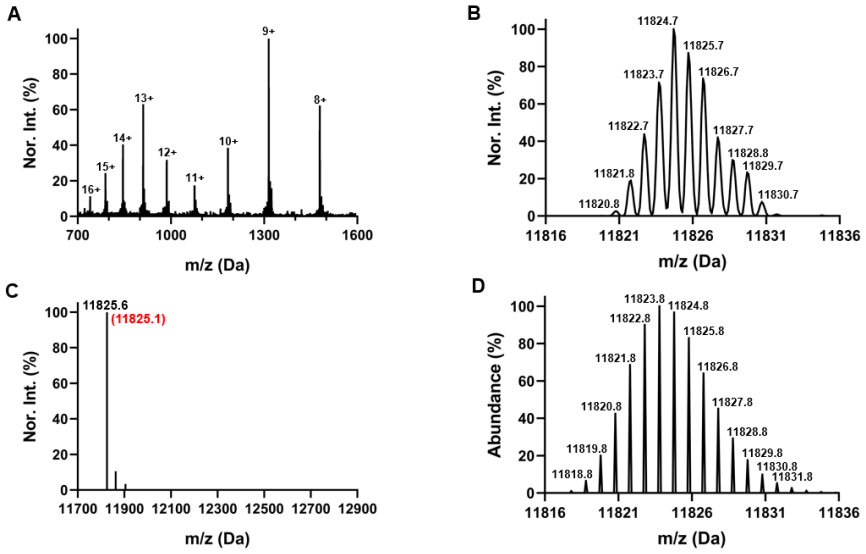

Figure 99 ESI-MS spectra of FLAG-SUMO2-C48A/G93Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

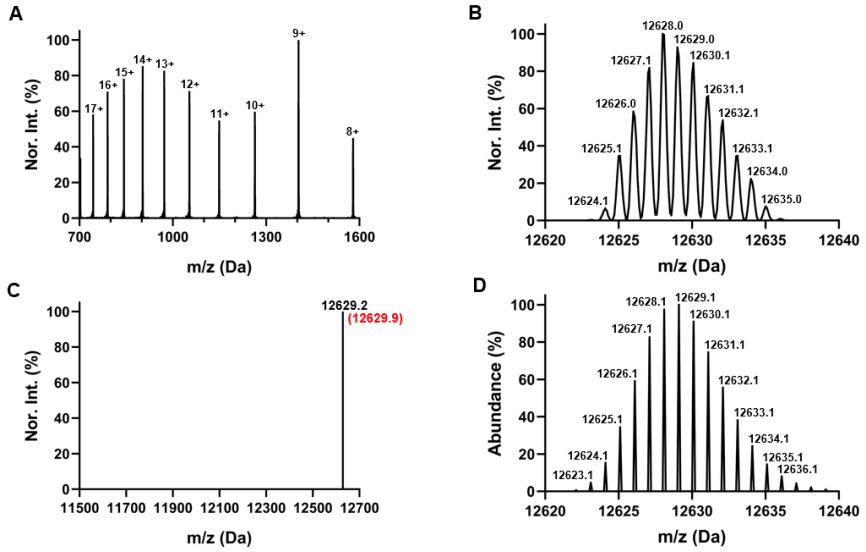

Figure 100 ESI-MS spectra of FLAG-SUMO3-C47A/G92C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

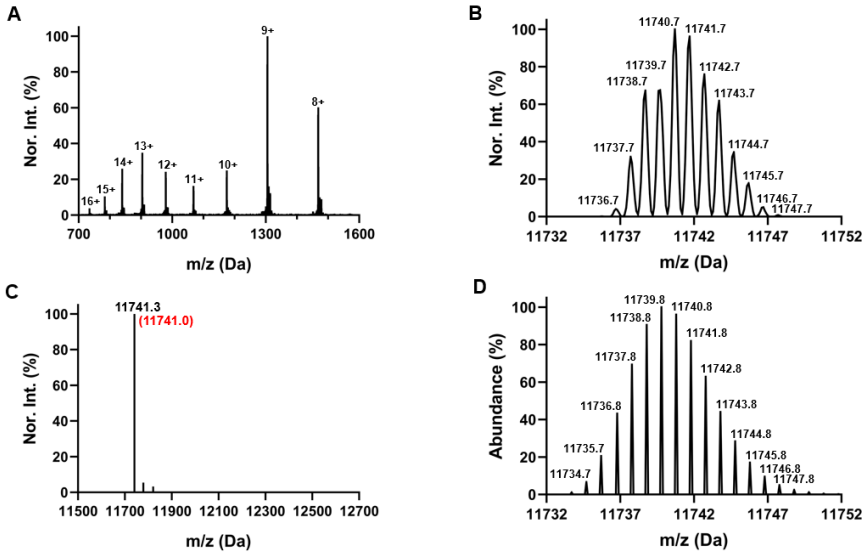

Figure 101 ESI-MS spectra of FLAG-SUMO3-C47A/G92Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

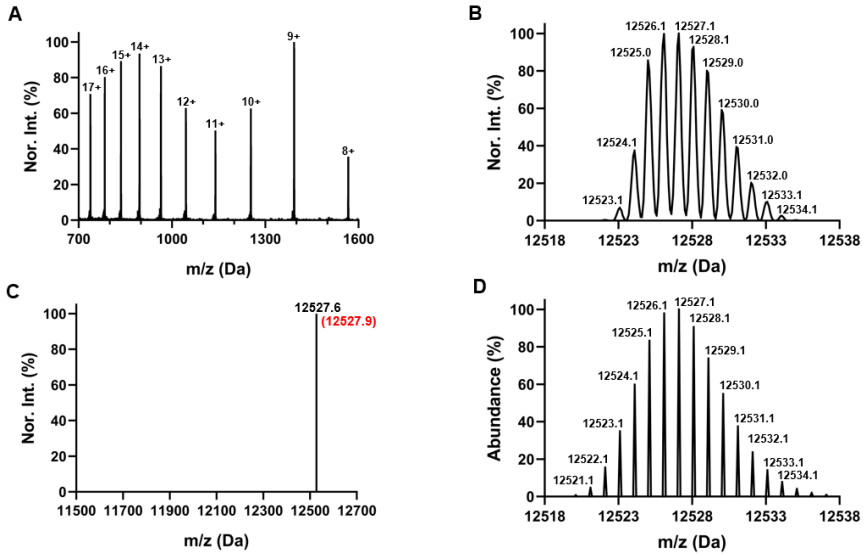

Figure 102 ESI-MS spectra of FLAG-SUMO4-C48A/G93C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

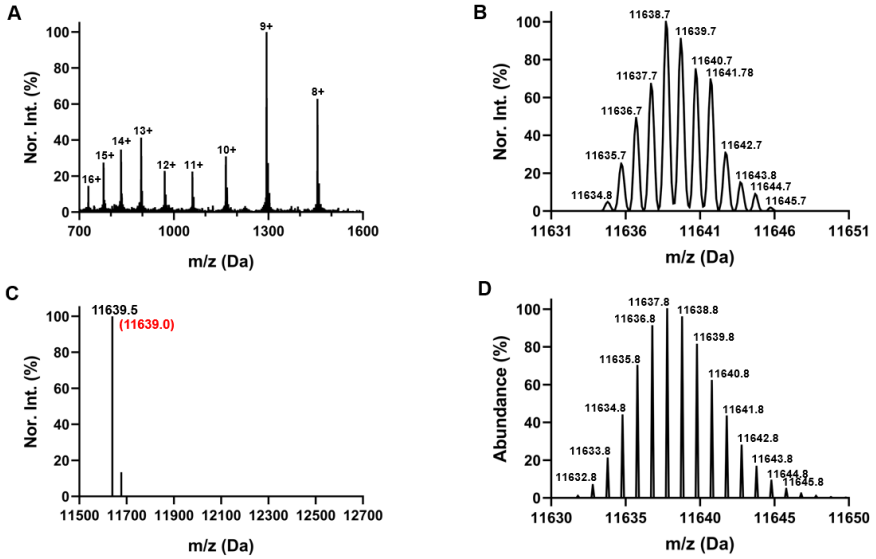

Figure 103 ESI-MS spectra of FLAG-SUMO4-C48A/G93Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

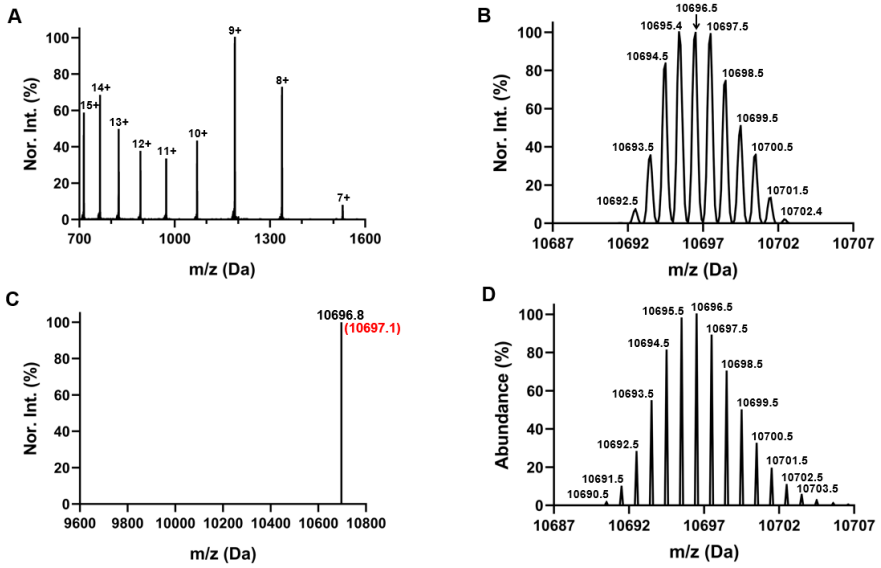

Figure 104 ESI-MS spectra of FLAG-NEDD8-G76C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

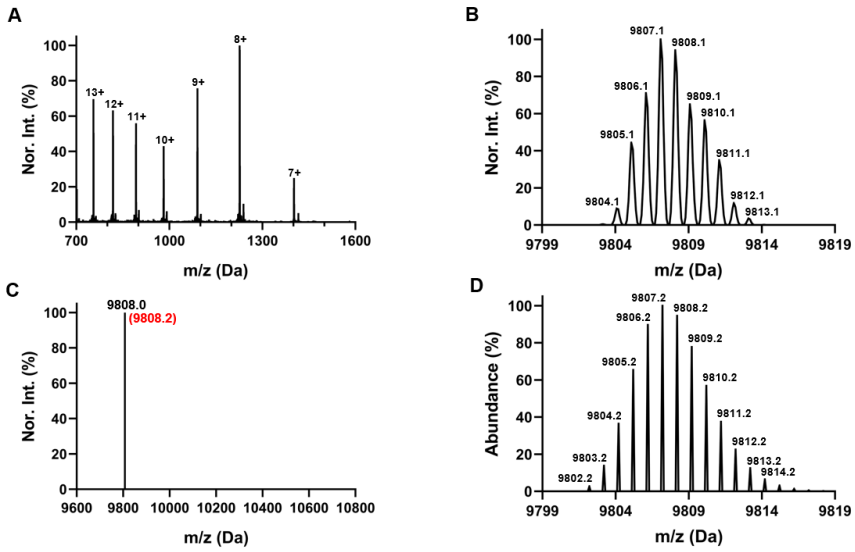

Figure 105 ESI-MS spectra of FLAG-NEDD8-G76Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

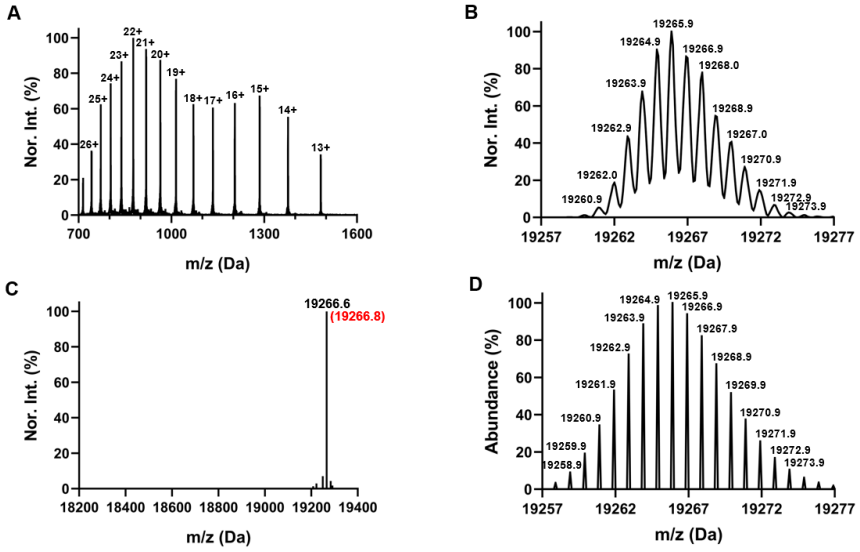

Figure 106 ESI-MS spectra of FLAG-ISG15-C89S/G157C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

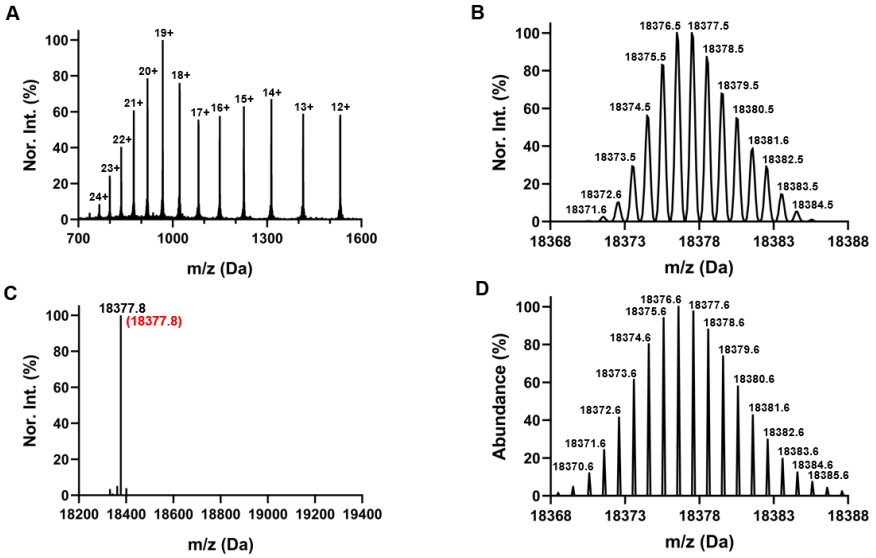

Figure 107 ESI-MS spectra of FLAG-ISG15-C89S/G157Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

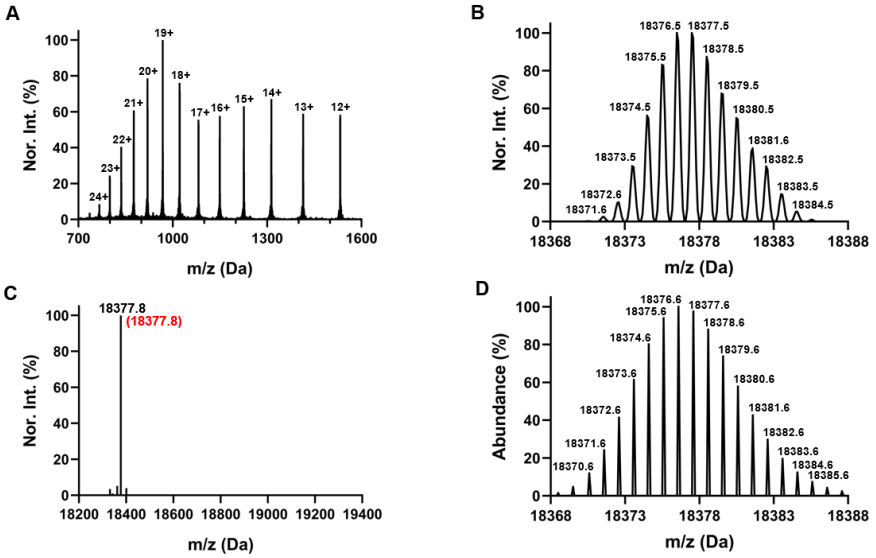

Figure 108 ESI-MS spectra of FLAG-GABARAP-G116C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

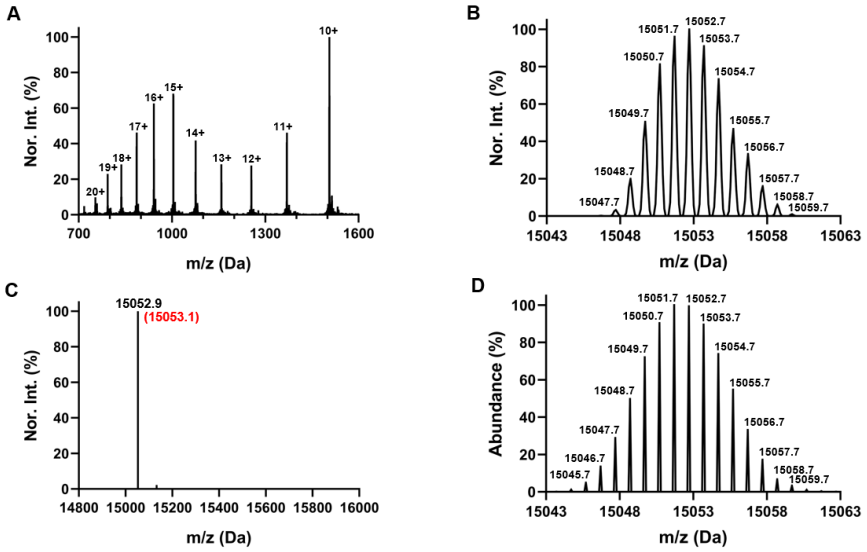

Figure 109 ESI-MS spectra of FLAG-GABARAP-G116Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

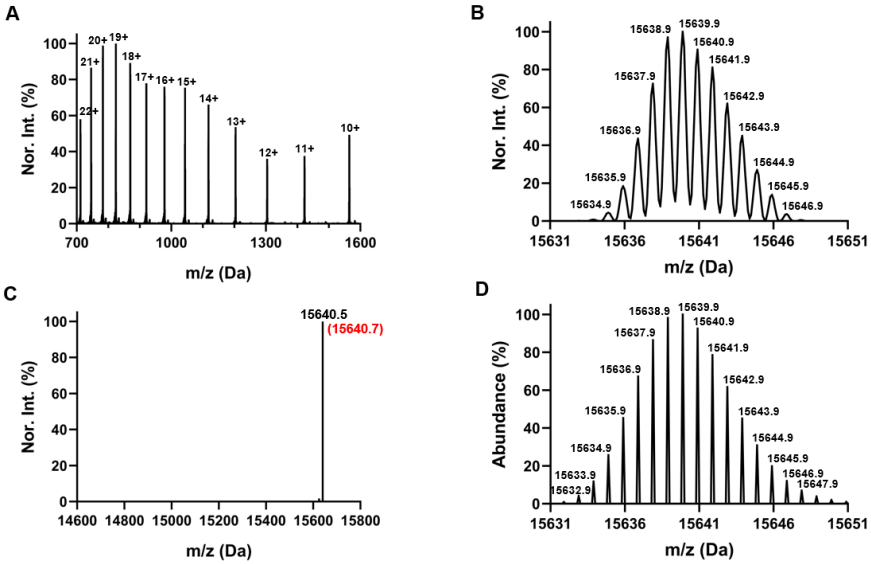

Figure 110 ESI-MS spectra of FLAG-GABARAPL2-G116C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

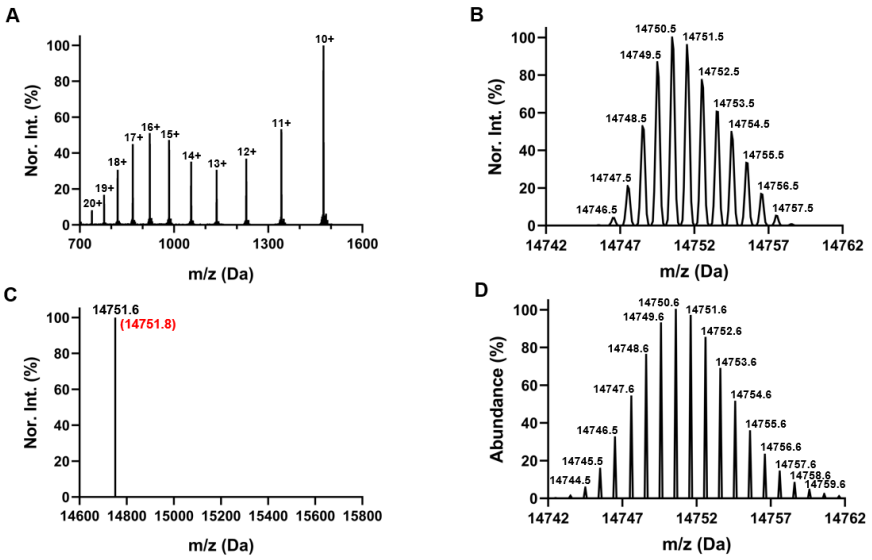

Figure 111 ESI-MS spectra of FLAG-GABARAPL2-G116Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

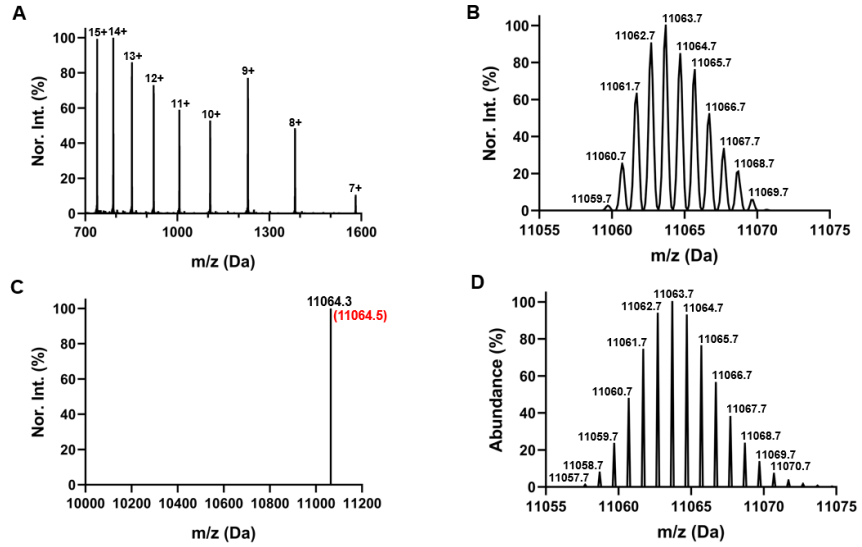

Figure 112 ESI-MS spectra of FLAG-UFM1-G83C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

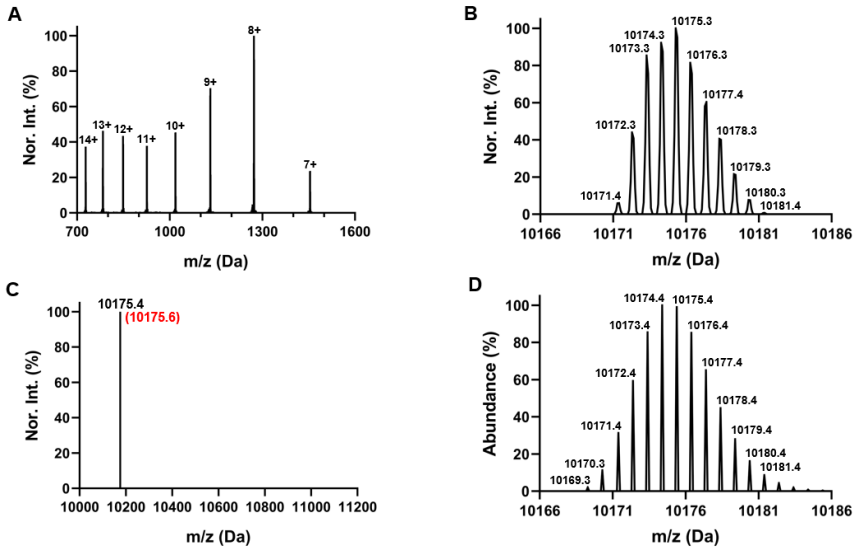

Figure 113 ESI-MS spectra of FLAG-UFM1-G83Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

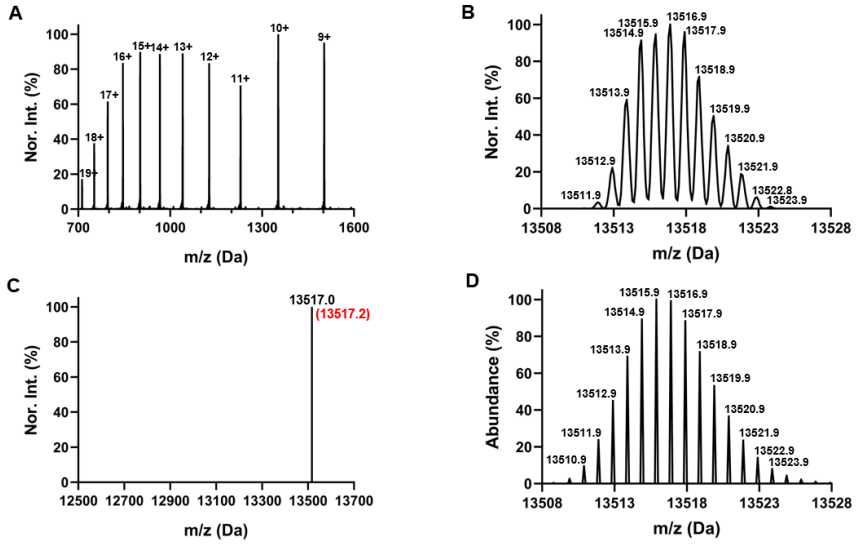

Figure 114 ESI-MS spectra of FLAG-URM1-G101C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

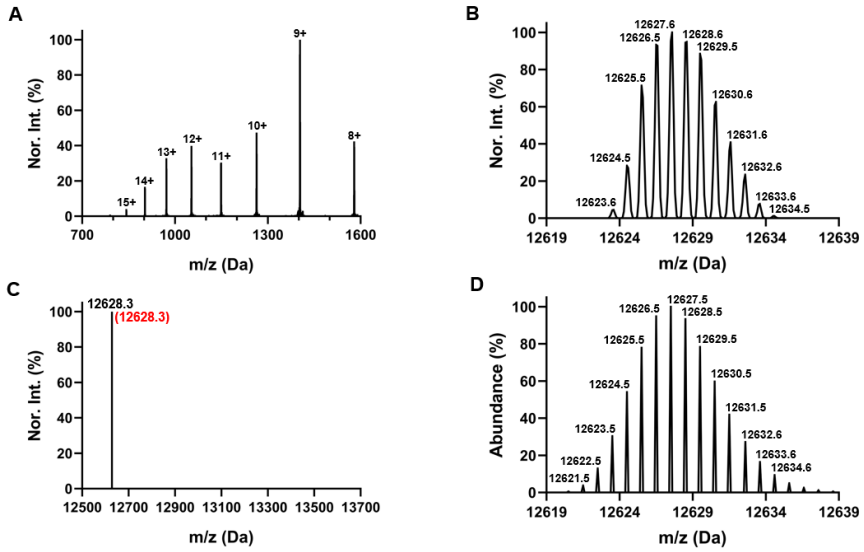

Figure 115 ESI-MS spectra of FLAG-URM1-G101Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

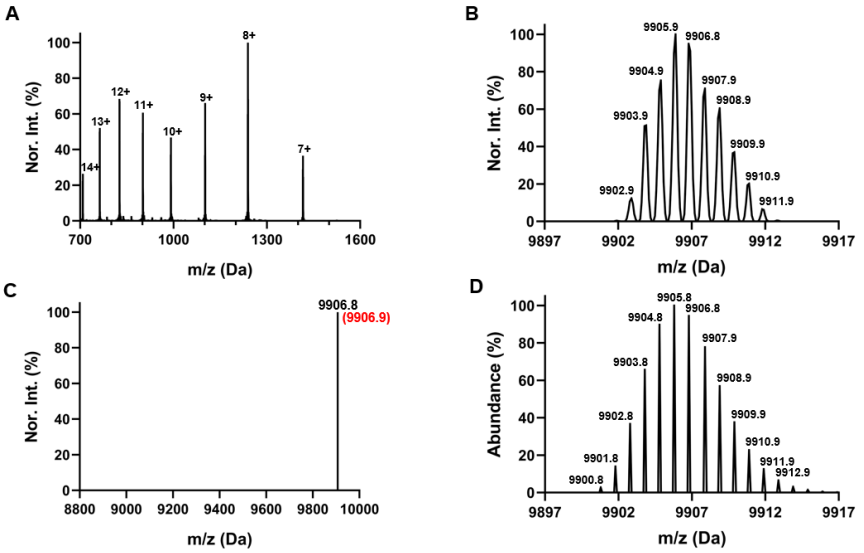

Figure 116 ESI-MS spectra of FLAG-MNSFβ-C57S/G74C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

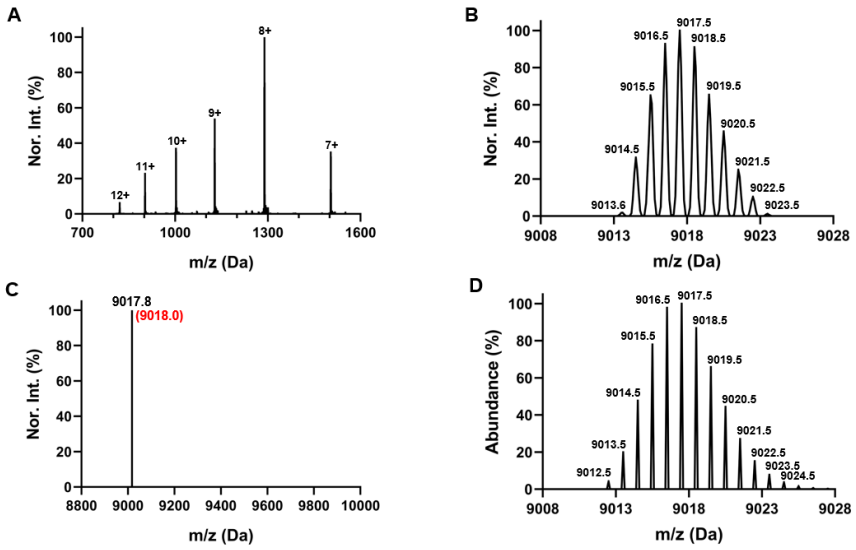

Figure 117 ESI-MS spectra of FLAG-MNSFβ-C57S/G74Pa. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

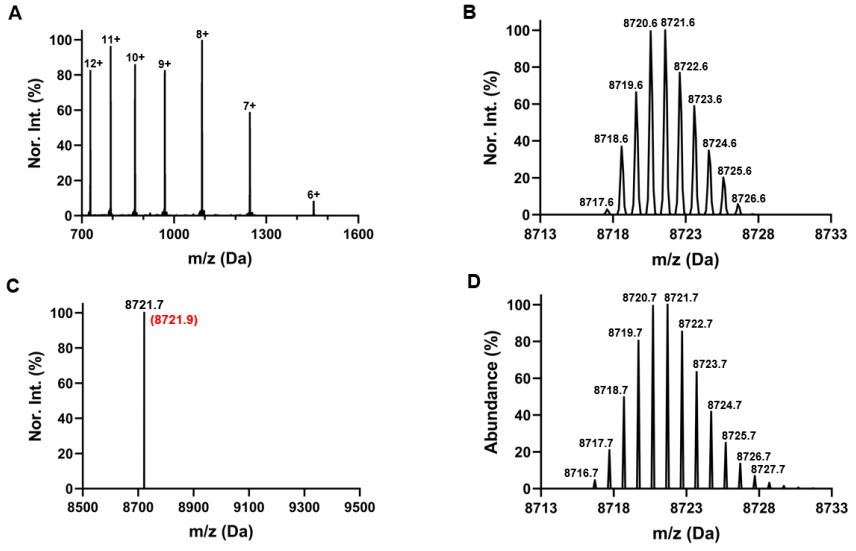

Figure 118 ESI-MS spectra of Ub-AMC. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

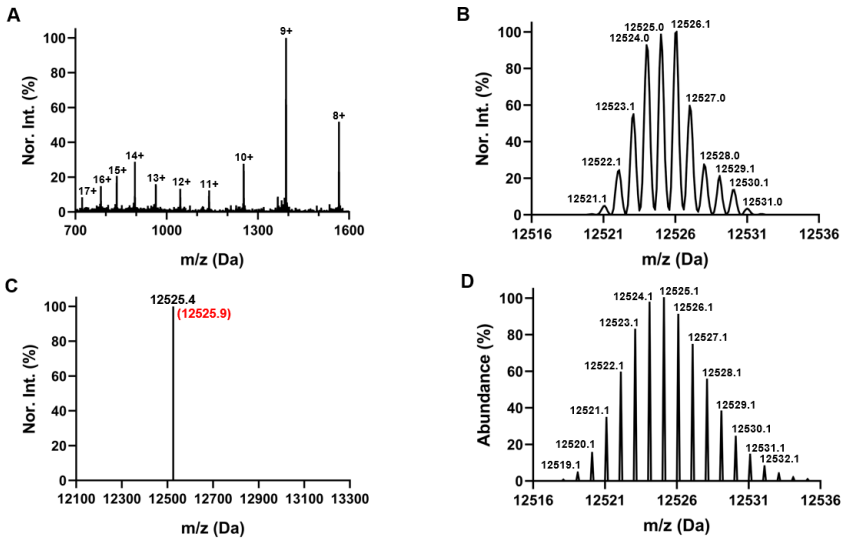

Figure 119 ESI-MS spectra of FLAG-SUMO1-C52A-AMC. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

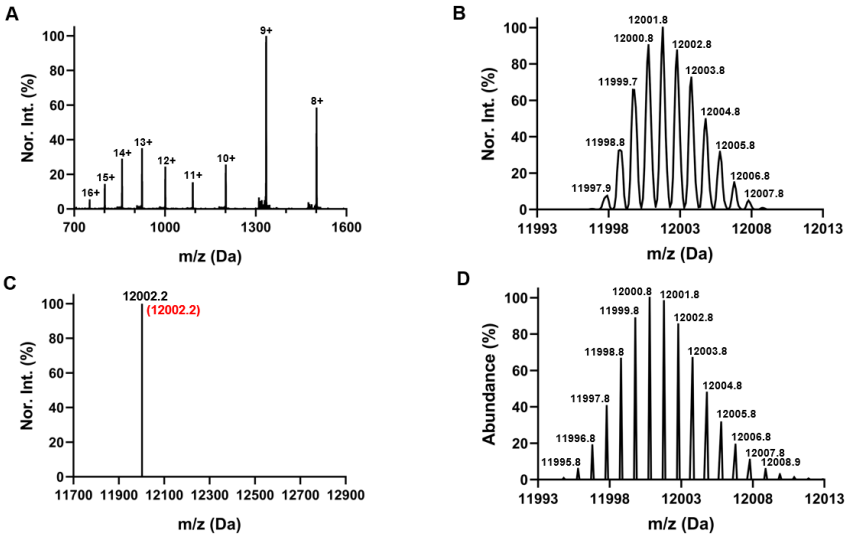

Figure 120 ESI-MS spectra of FLAG-SUMO2-C48A-AMC. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

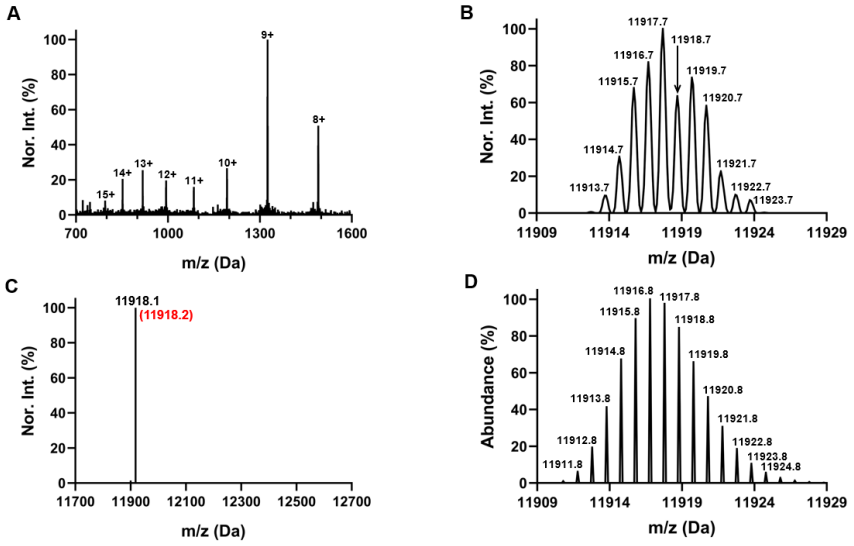

Figure 121 ESI-MS spectra of FLAG-SUMO3-C47A-AMC. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

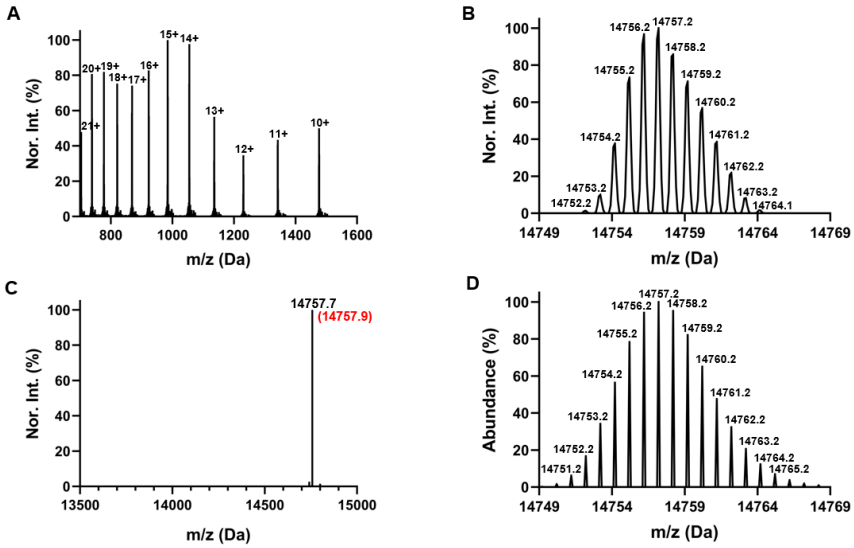

Figure 122 ESI-MS spectra of H2A-K129C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

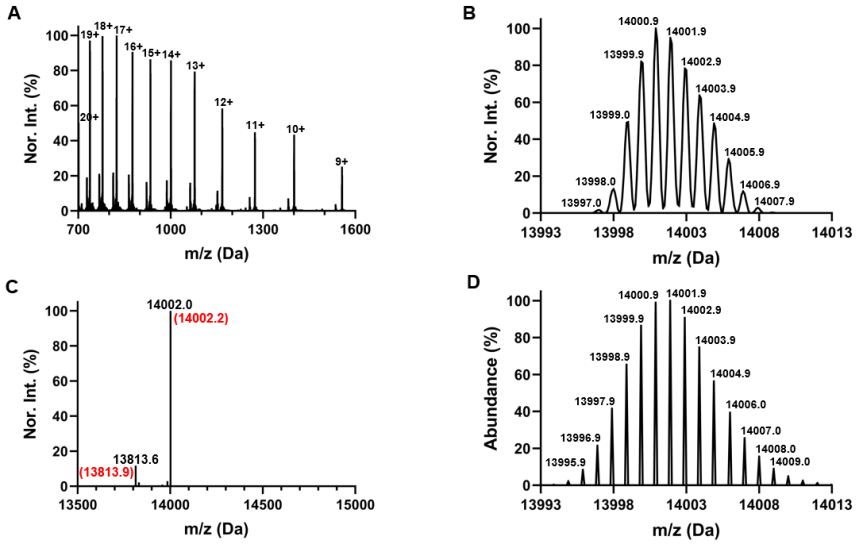

Figure 123 ESI-MS spectra of H2AK129ac. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

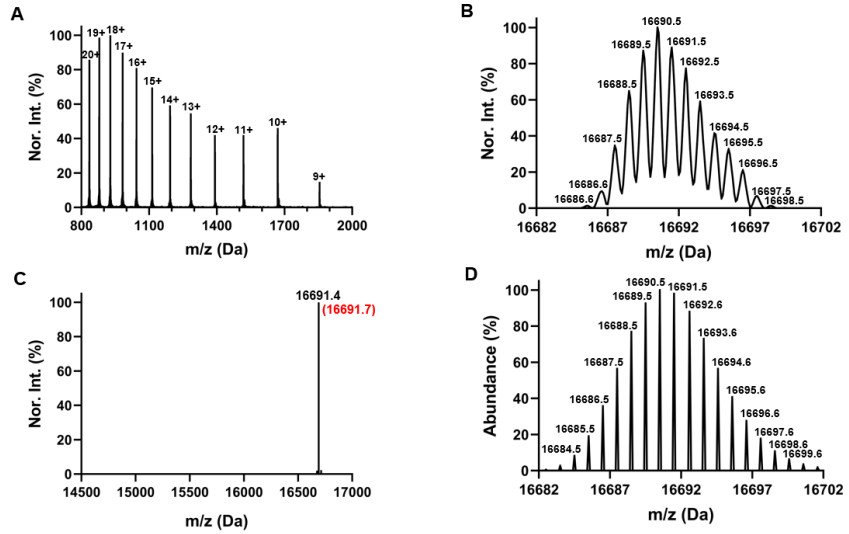

Figure 124 ESI-MS spectra of $RNH_{59\text{-}196}$-K190C-6H. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

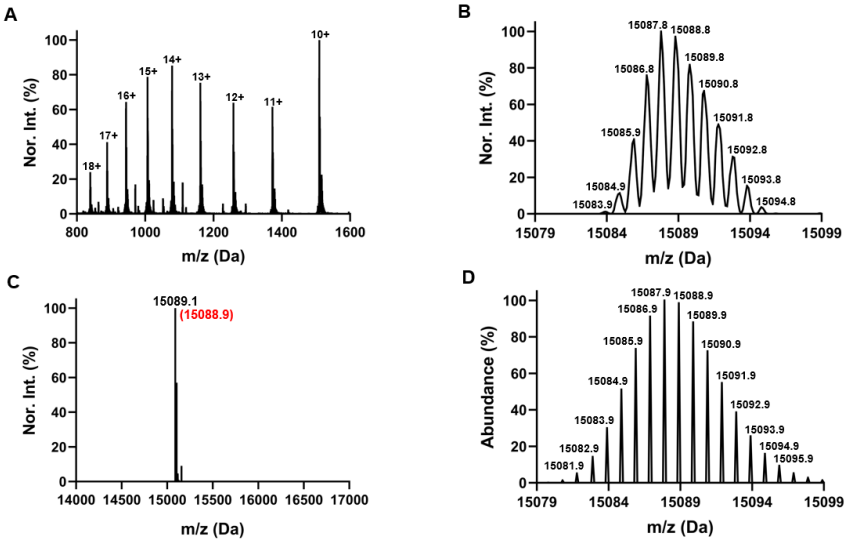

Figure 125 ESI-MS spectra of $RNH_{59\text{-}189}$-Ha. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

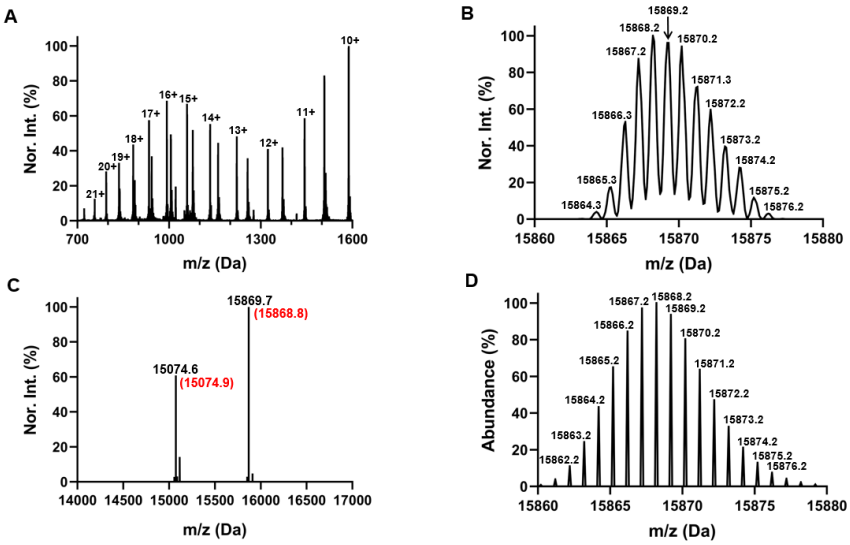

Figure 126 ESI-MS spectra of $RNH_{59\text{-}196}$-K190C. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

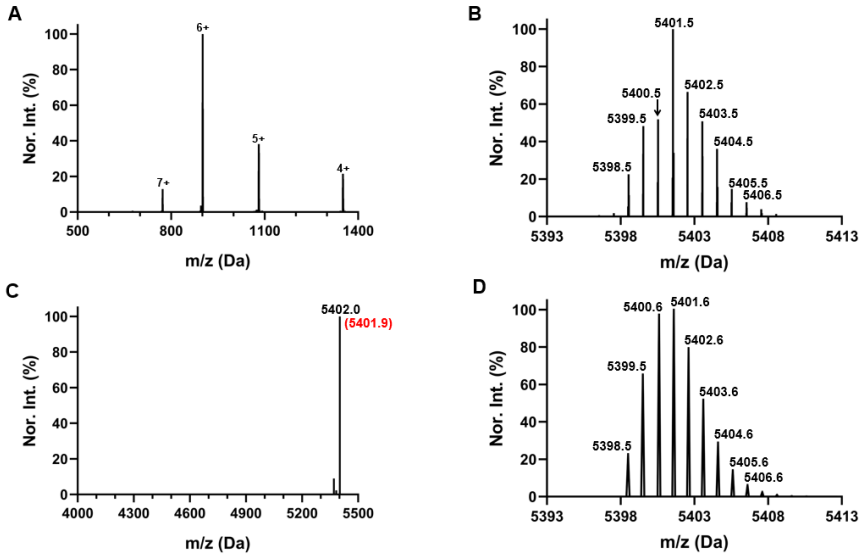

Figure 127 ESI-MS spectra of exenatide-S39C-SA-Strep. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

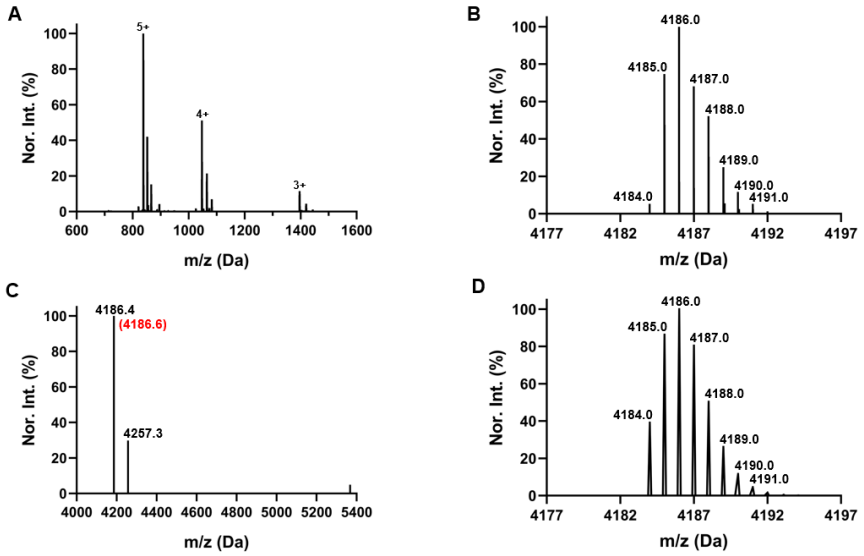

Figure 128 ESI-MS spectra of exenatide. (A) raw data, (B) deconvoluted spectra, (C) experimental and theoretical average mass (shown in red), (D) theoretical monoisotopic peaks and relative natural abundance.

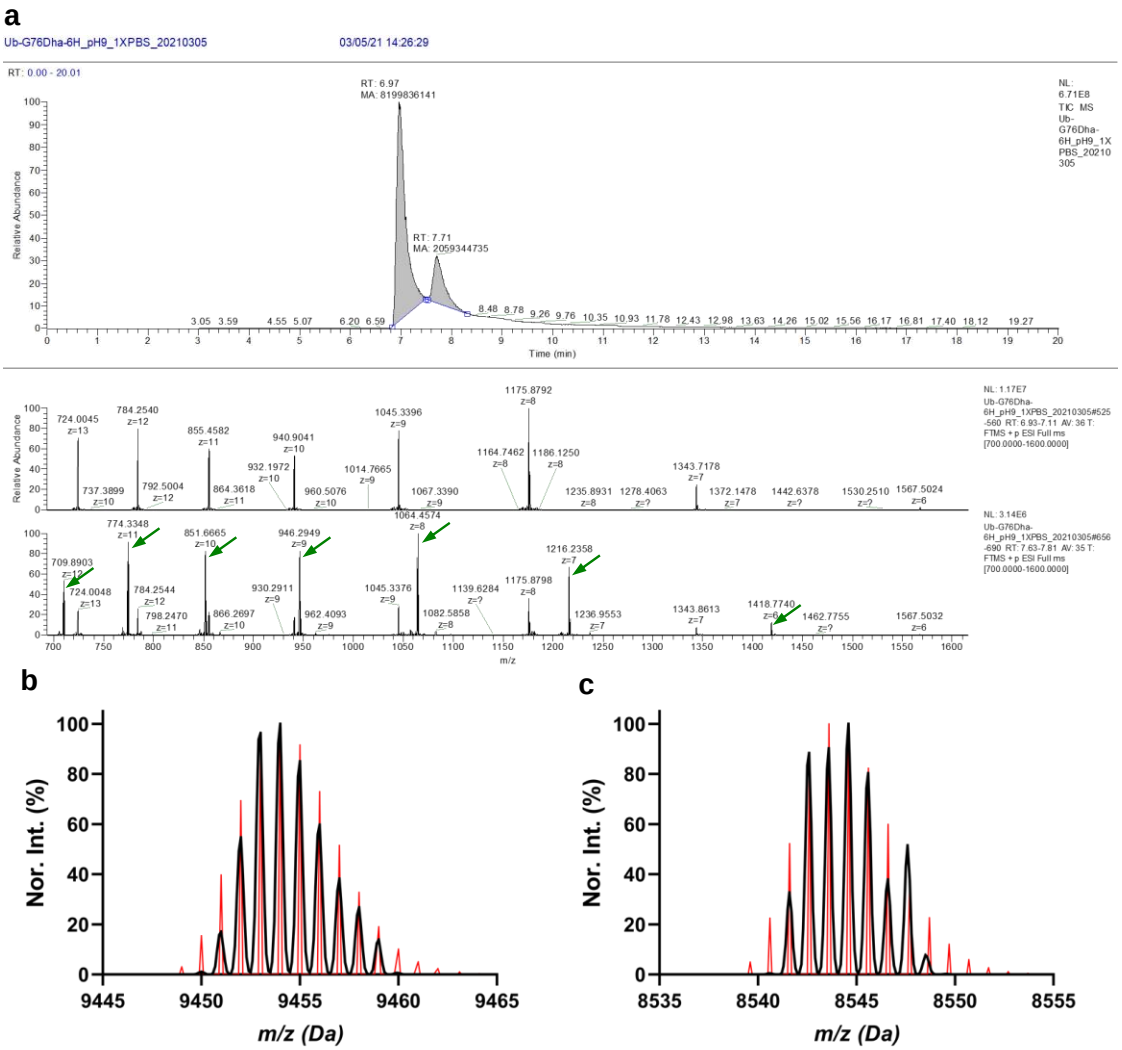

Figure 129 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate at pH 9. Reaction condition: 1× PBS pH 9, 37 °C, 18 h incubation. (a) The top mass diagram that came from the 6.97 min peak represented Ub-G76Dha-6H. The major mass signal pattern indicated in green in the bottom mass diagram that came from the 7.71 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$. (b) Deconvoluted mass spectrum (black line) of the Dha derivative product formed via Michael addition reaction between Pa and Ub-G76Dha-

6H, and the theoretic natural abundance of its monoisotopic peaks (red line). (c) Deconvoluted mass spectrum (black line) of the Ub-G76Pa, and the theoretic natural abundance of its monoisotopic peaks (red line).

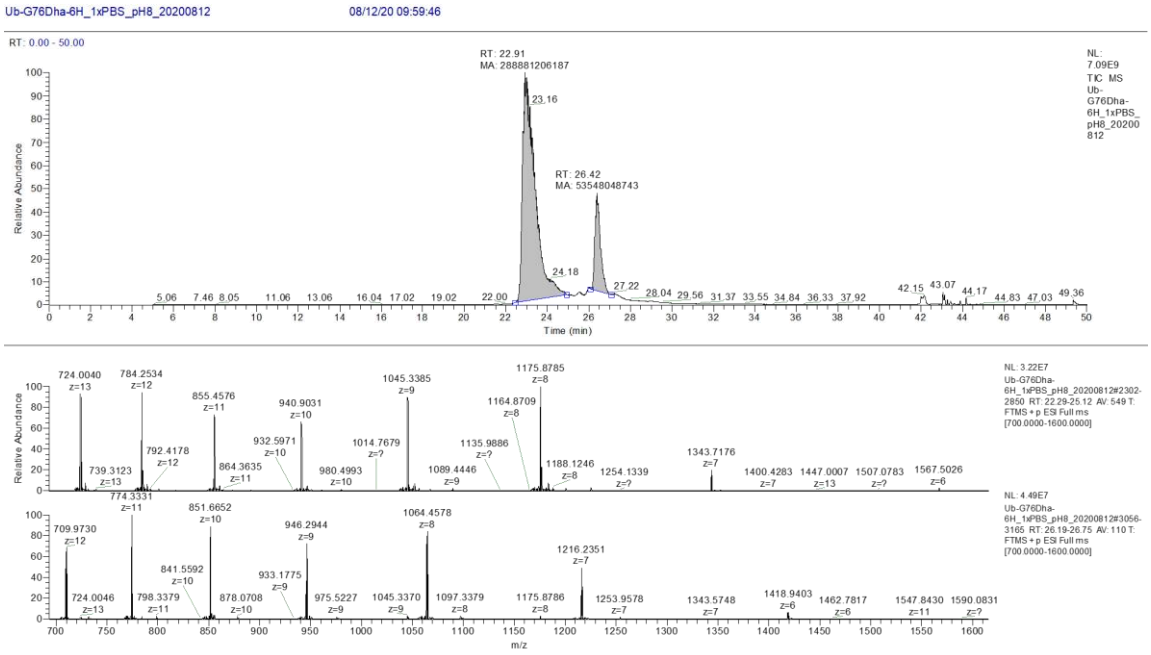

Figure 130 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate at pH 8. Reaction condition: 1× PBS pH 8, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 6.97 min peak represented Ub-G76Dha-6H. The major mass signal pattern indicated in green in the bottom mass diagram that came from the 7.71 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

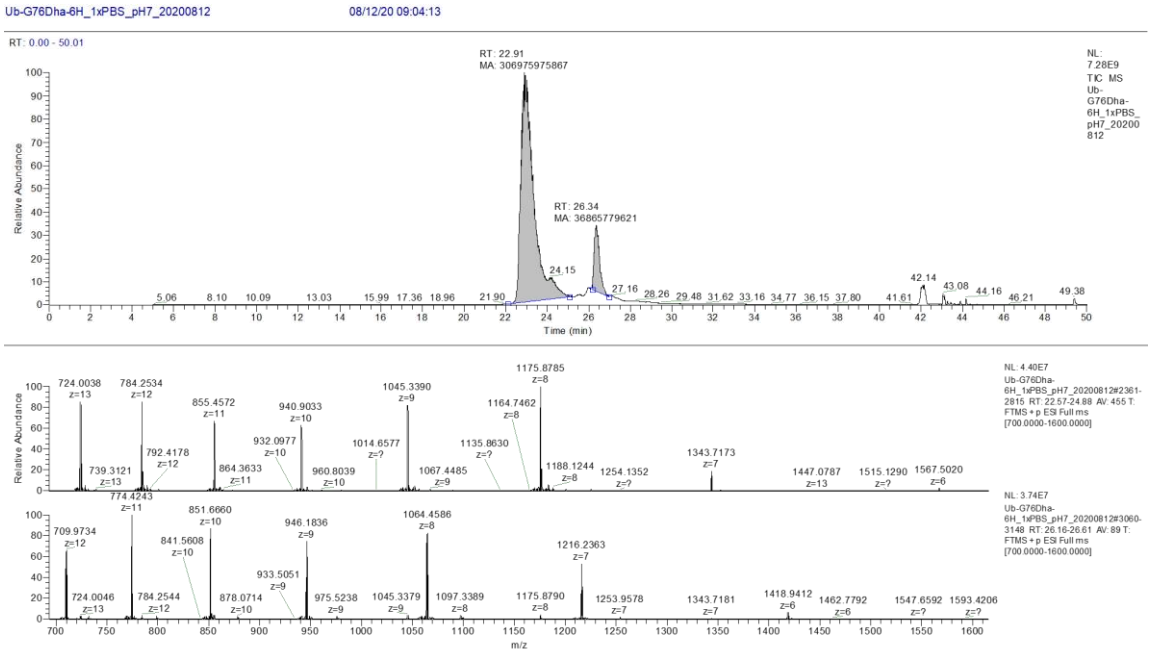

Figure 131 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate at pH 7. Reaction condition: 1× PBS pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 22.91 min peak represented Ub-G76Dha-6H. The bottom mass diagram that came from the 26.34 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

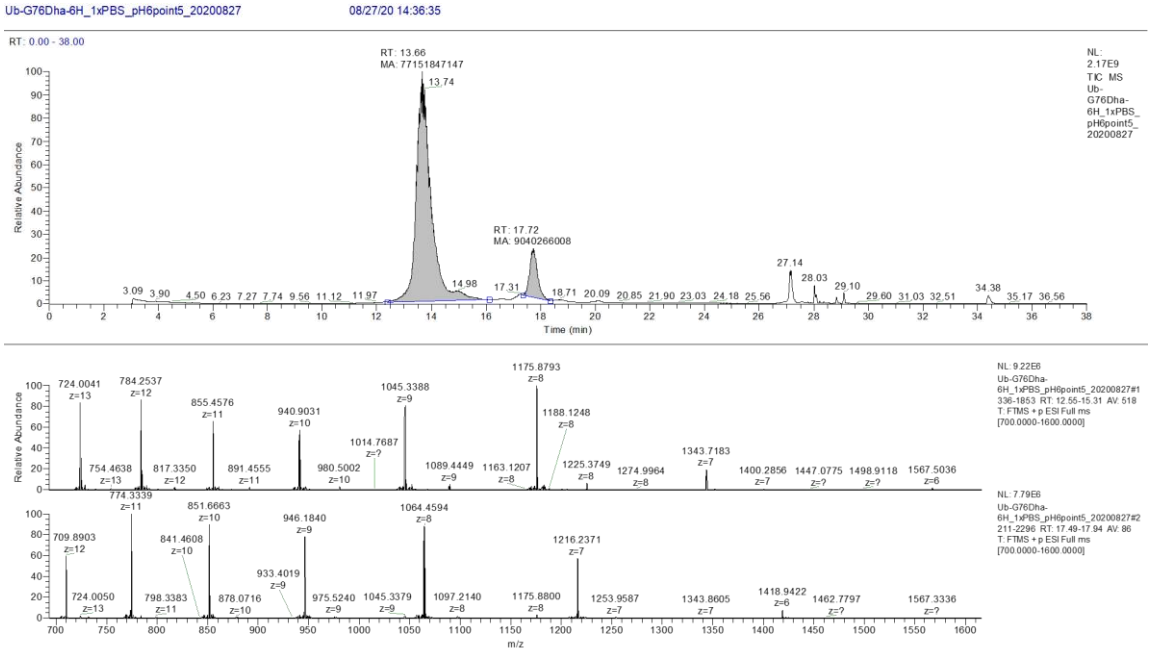

Figure 132 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate at pH 6.5. Reaction condition: 1× PBS pH 6.5, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.66 min peak represented Ub-G76Dha-6H. The bottom mass diagram that came from the 17.72 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

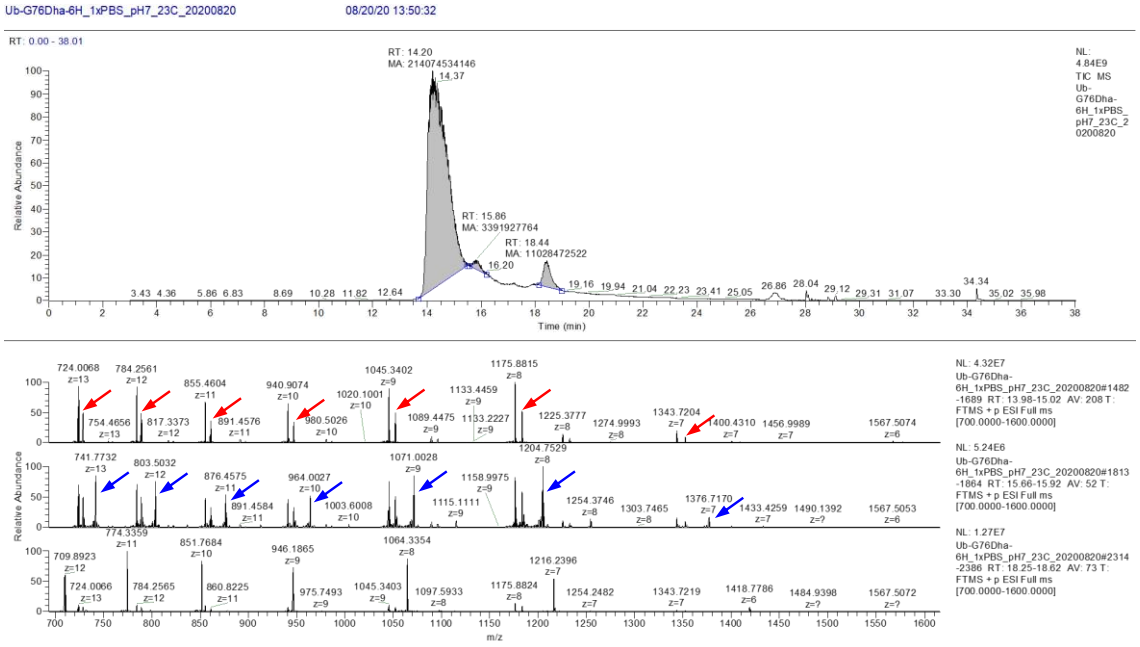

Figure 133 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate at 23 °C. Reaction condition: 1× PBS pH 7, 23 °C, 18 h incubation. The major mass signal pattern in the top mass diagram that came from the 14.20 min peak represented Ub-G76Dha-6H. The minor mass signal pattern indicated in red on the right side of each Dha product signal showed the formation of the +CN product, Ub-G76C(CN)-6H. The mass signal pattern indicated in blue in the middle mass diagram that came from the 15.86 min peak determined the formation of +TNB product, Ub-G76C(TNB)-6H. The bottom mass diagram that came from the 18.44 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

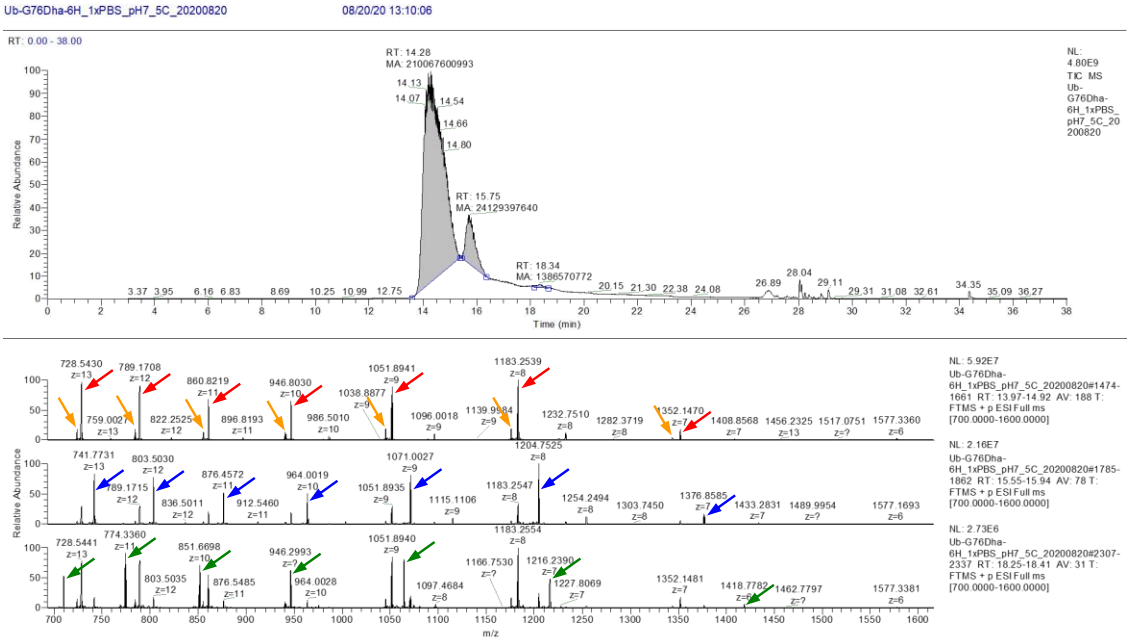

Figure 134 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate at 5 °C. Reaction condition: 1× PBS pH 7, 5 °C, 18 h incubation. The major mass signal pattern indicated in red in the top mass diagram that came from the 14.28 min peak showed the formation of +CN product, Ub-G76C(CN)-6H. The minor mass signal pattern indicated in orange on the left side of each +CN product signal represented Ub-G76Dha-6H. The mass signal pattern indicated in blue in the middle mass diagram that came from the 15.75 min peak determined the formation of +TNB product, Ub-G76C(TNB)-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the 18.34 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

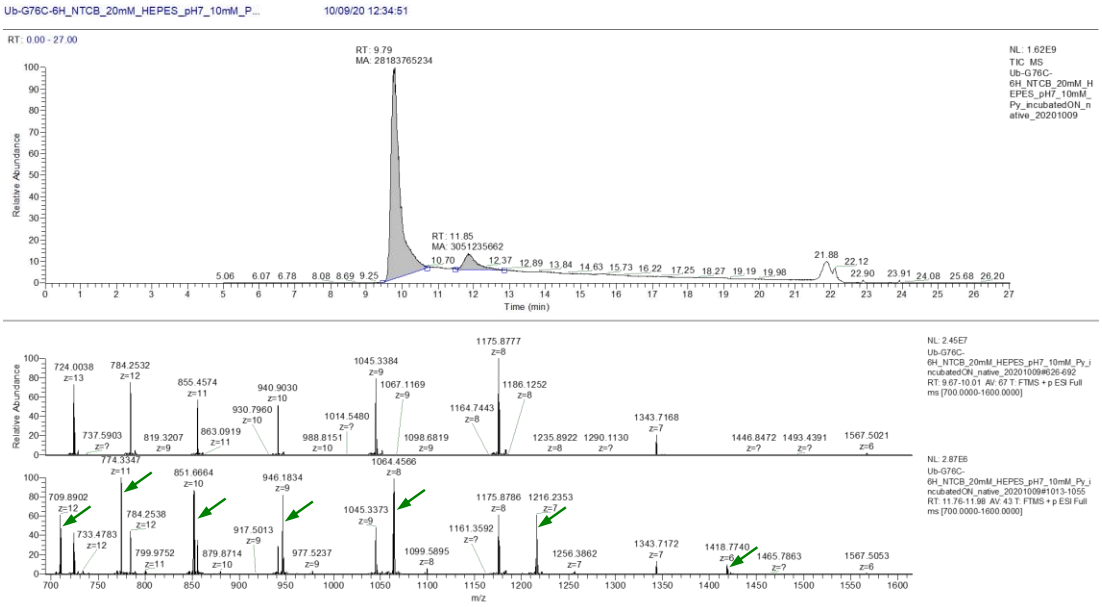

Figure 135 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The top mass diagram that came from the 9.79 min peak represented Ub-G76Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the 11.85 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

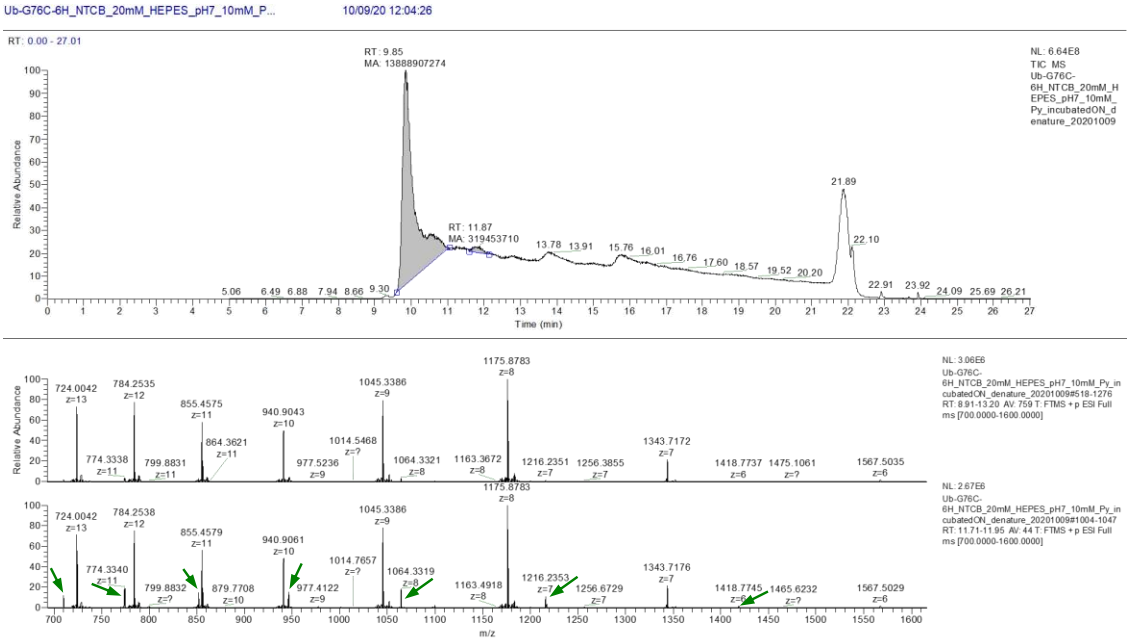

Figure 136 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate in pure water under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The top mass diagram that came from the 9.85 min peak represented Ub-G76Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the 11.87 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

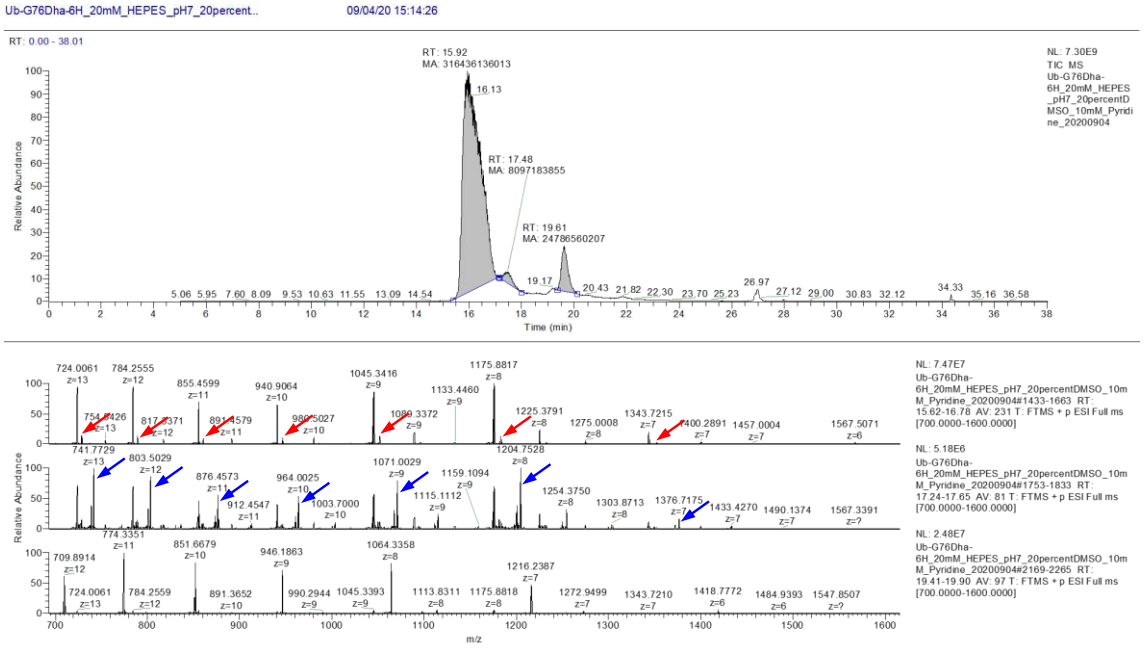

Figure 137 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate in 20% DMSO. Reaction condition: 20 mM HEPES pH 7, 20% DMSO, 37 °C, 18 h incubation. The major mass signal pattern in the top mass diagram that came from the 15.92 min peak represented Ub-G76Dha-6H. The minor mass signal pattern indicated in red on the right side of each Dha product signal showed the formation of +CN product, Ub-G76C(CN)-6H. The mass signal pattern indicated in blue in the middle mass diagram that came from the 17.48 min peak determined the formation of +TNB product, Ub-G76C(TNB)-6H. The mass signal pattern in the bottom mass diagram that came from the 19.61 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

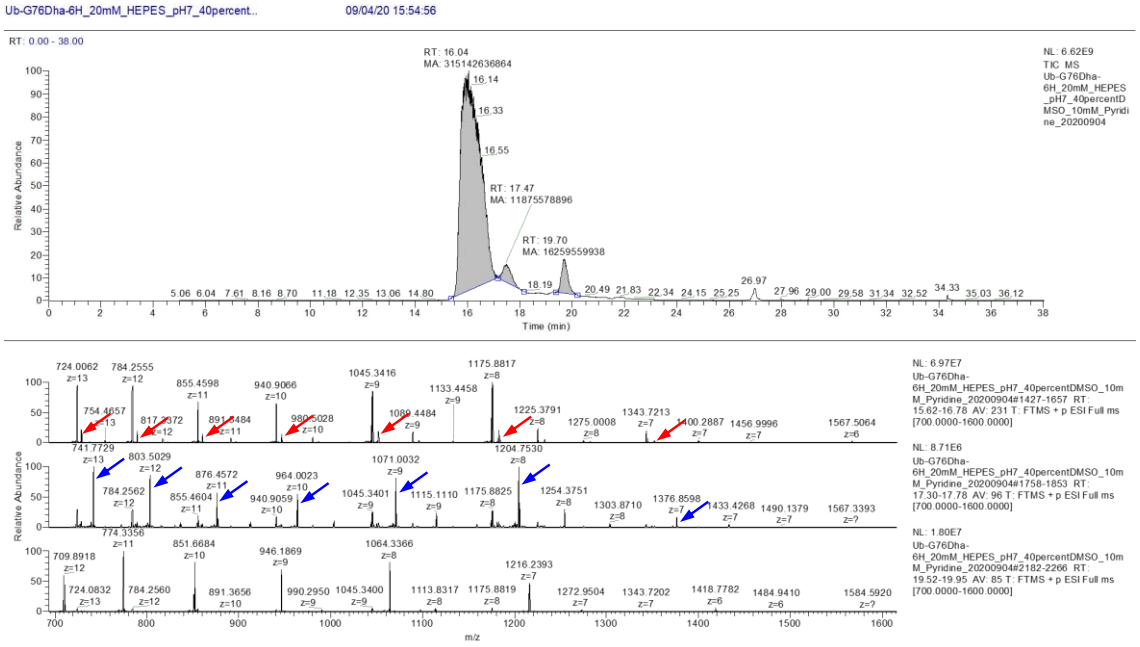

Figure 138 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate in 40% DMSO. Reaction condition: 20 mM HEPES pH 7, 40% DMSO, 37 °C, 18 h incubation. The major mass signal pattern in the top mass diagram that came from the 16.04 min peak represented Ub-G76Dha-6H. The minor mass signal pattern indicated in red on the right side of each Dha product signal showed the formation of +CN product, Ub-G76C(CN)-6H. The mass signal pattern indicated in blue in the middle mass diagram that came from the 17.47 min peak determined the formation of +TNB product, Ub-G76C(TNB)-6H. The mass signal pattern in the bottom mass diagram that came from the 19.70 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

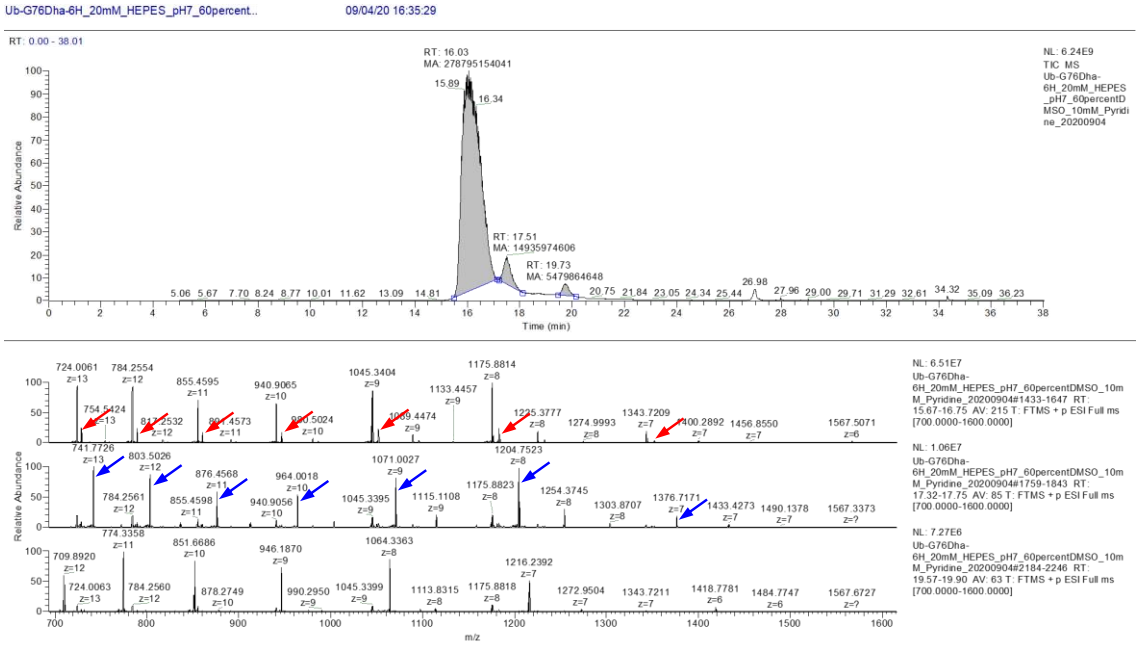

Figure 139 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate in 60% DMSO. Reaction condition: 20 mM HEPES pH 7, 60% DMSO, 37 °C, 18 h incubation. The major mass signal pattern in the top mass diagram that came from the 16.03 min peak represented Ub-G76Dha-6H. The minor mass signal pattern indicated in red on the right side of each Dha product signal showed the formation of +CN product, Ub-G76C(CN)-6H. The mass signal pattern indicated in blue in the middle mass diagram that came from the 17.51 min peak determined the formation of +TNB product, Ub-G76C(TNB)-6H. The mass signal pattern in the bottom mass diagram that came from the 19.73 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

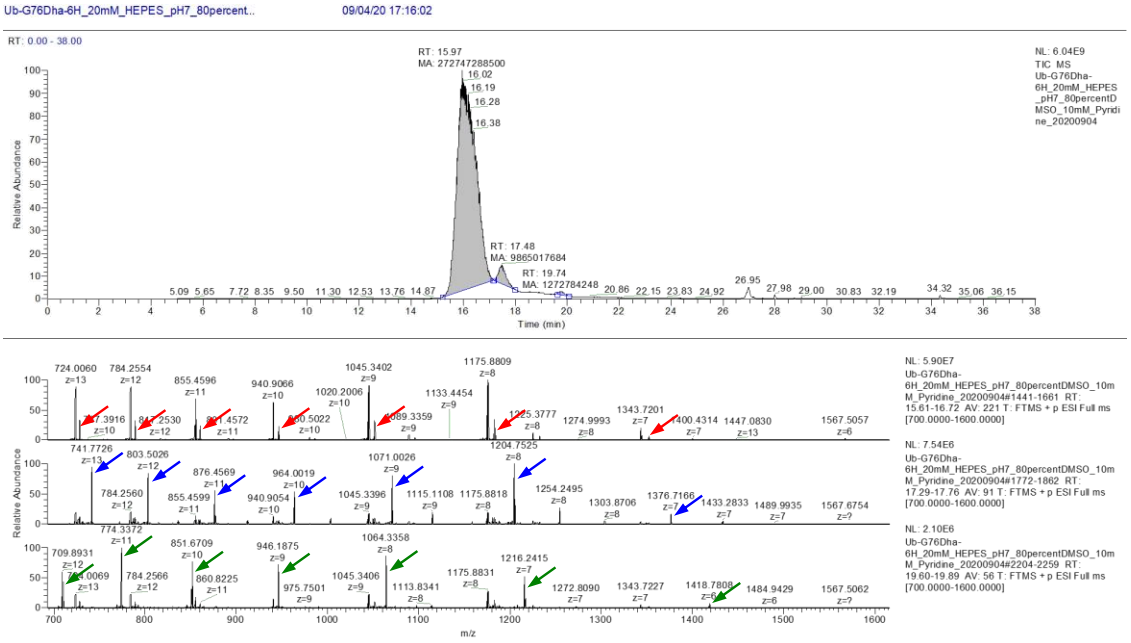

Figure 140 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate in 80% DMSO. Reaction condition: 20 mM HEPES pH 7, 80% DMSO, 37 °C, 18 h incubation. The major mass signal pattern in the top mass diagram that came from the 15.97 min peak represented Ub-G76Dha-6H. The minor mass signal pattern indicated in red on the right side of each Dha product signal showed the formation of +CN product, Ub-G76C(CN)-6H. The mass signal pattern indicated in blue in the middle mass diagram that came from the 17.48 min peak determined the formation of +TNB product, Ub-G76C(TNB)-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the 19.74 min peak designated the formation of hydrolysis product, $Ub_{(1-75)}$.

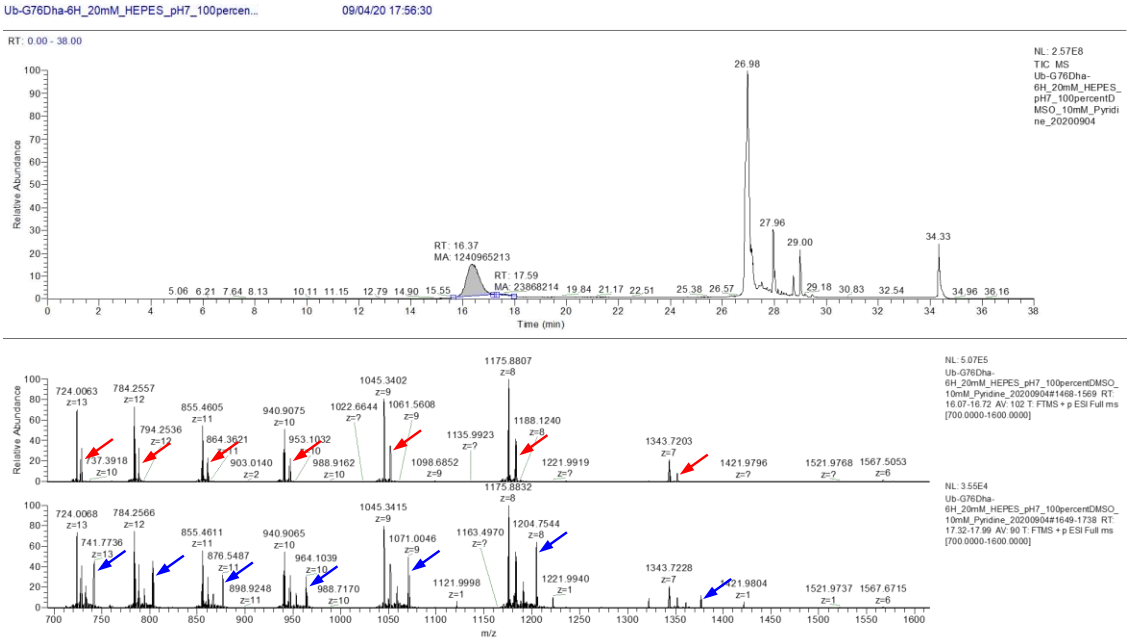

Figure 141 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-G76C-6H as the substrate in 100% DMSO. Reaction condition: 20 mM HEPES pH 7, 100% DMSO, 37 °C, 18 h incubation. The major mass signal pattern in the top mass diagram that came from the 16.37 min peak represented Ub-G76Dha-6H. The minor mass signal pattern indicated in red on the right side of each Dha product signal showed the formation of +CN product, Ub-G76C(CN)-6H. The mass signal pattern indicated in blue in the bottom mass diagram that came from the tailing area of the Dha peak determined the formation of +TNB product, Ub-G76C(TNB)-6H. No hydrolysis product was detected due to the lack of water as solvent.

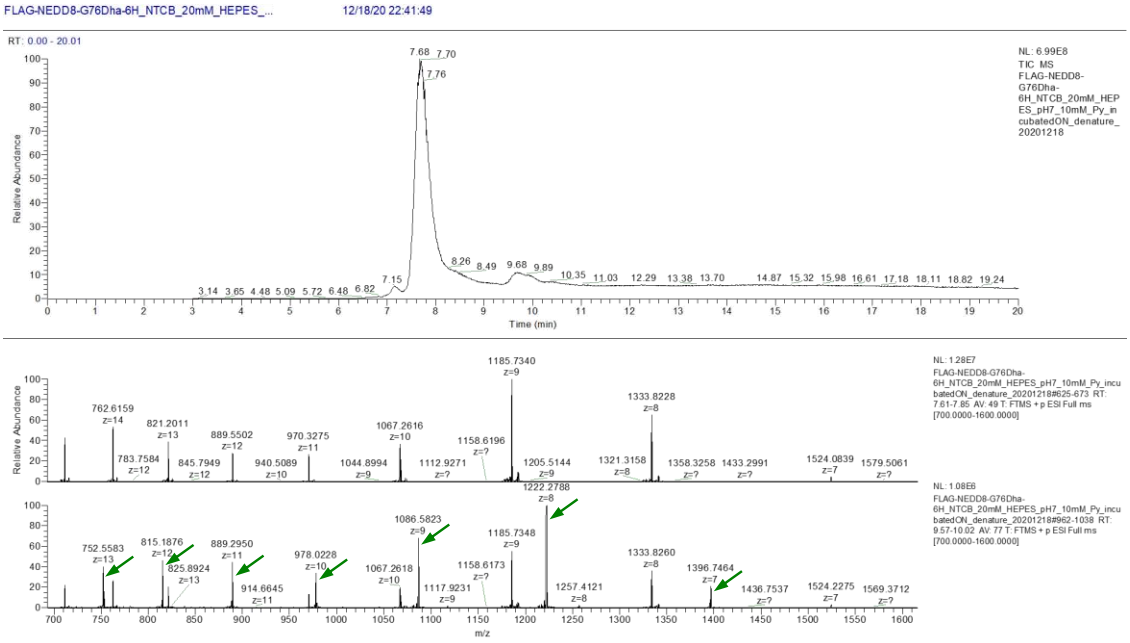

Figure 142 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-NEDD8-G76C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 7.68 min peak represented FLAG-NEDD8-G76Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the 9.68 min peak showed the formation of hydrolysis product, FLAG-NEDD8$_{(1-75)}$.

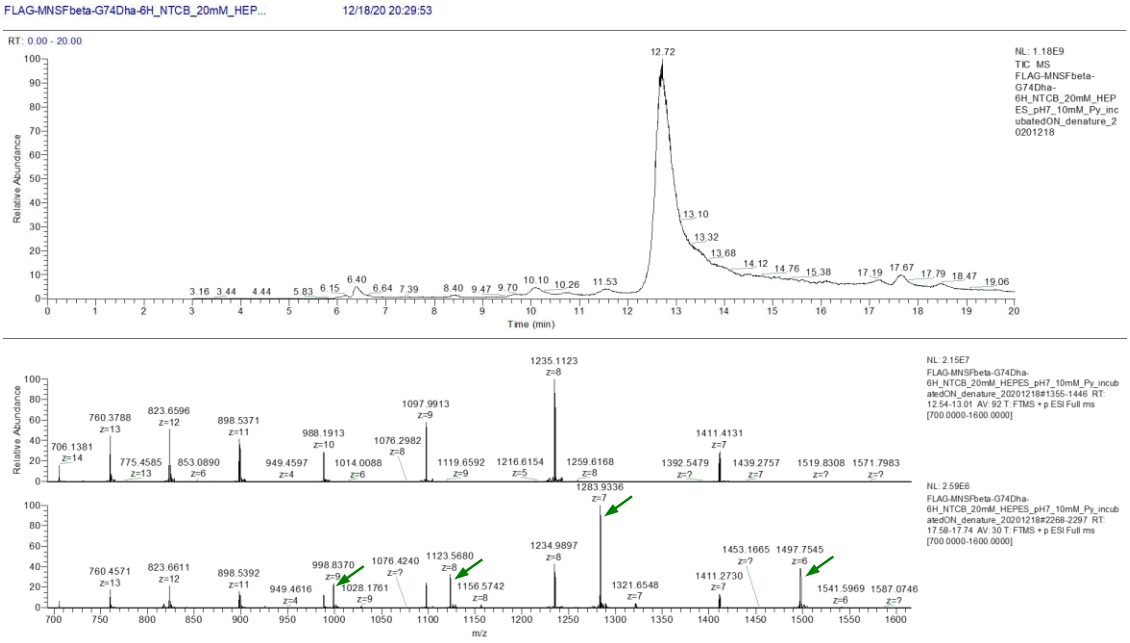

Figure 143 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-MNSFβ-C57S-G74C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 12.72 min peak represented FLAG-MNSFβ-C57S-G74Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the 17.67 min peak showed the formation of hydrolysis product, FLAG-MNSF$\beta_{(1\text{-}73)}$-C57S.

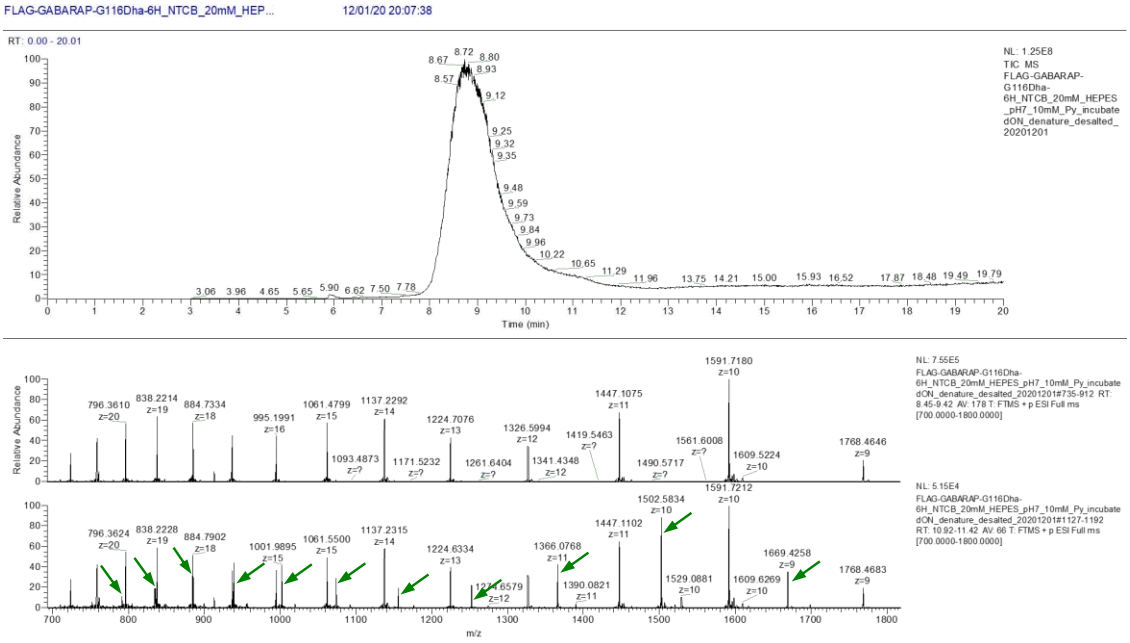

Figure 144 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-GABARAP-G116C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 8.72 min peak represented FLAG-GABARAP-G116Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the tailing area of Dha product peak showed the formation of hydrolysis product, FLAG-GABARAP$_{(1\text{-}115)}$.

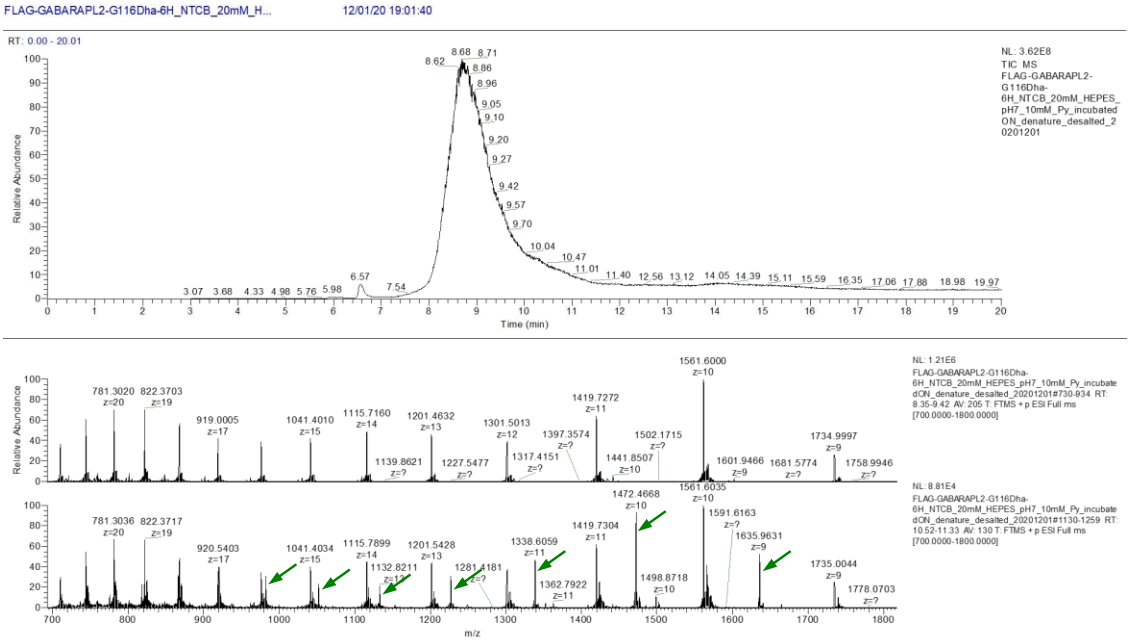

Figure 145 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-GABARAPL2-G116C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 8.68 min peak represented FLAG-GABARAPL2-G116Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the tailing area of Dha product peak showed the formation of hydrolysis product, FLAG-GABARAPL2$_{(1\text{-}115)}$.

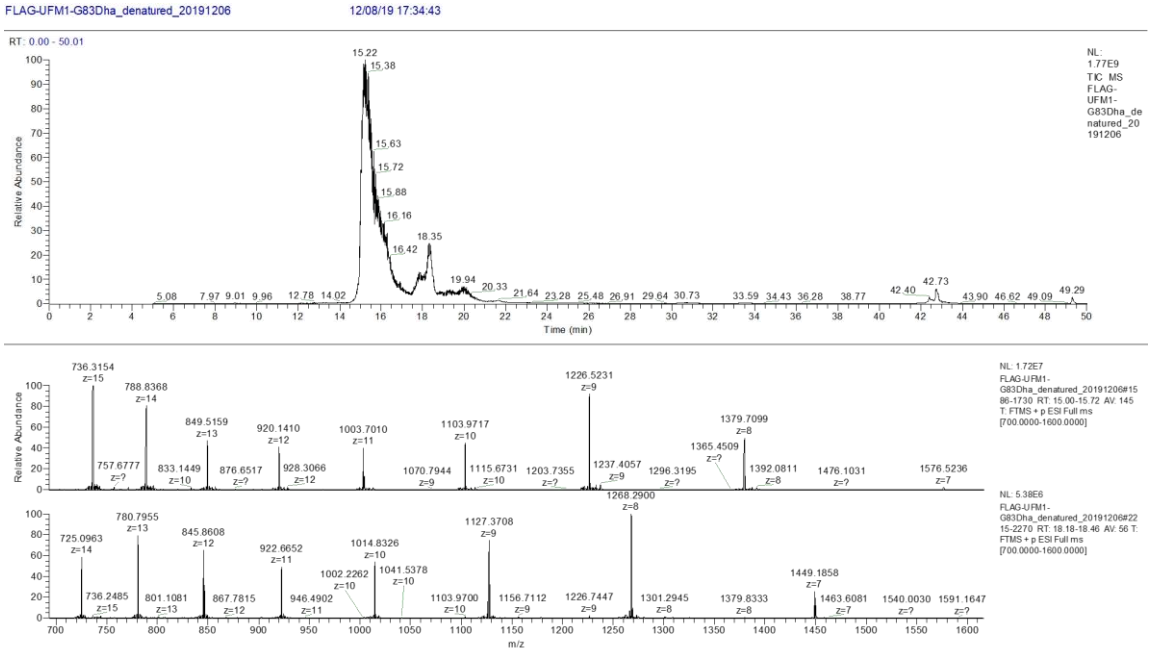

Figure 146 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-UFM1-G83C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 15.22 min peak represented FLAG-UFM1-G83Dha-6H. The mass signal pattern in the bottom mass diagram that came from the 18.35 min peak showed the formation of hydrolysis product, FLAG-UFM1$_{(1\text{-}82)}$.

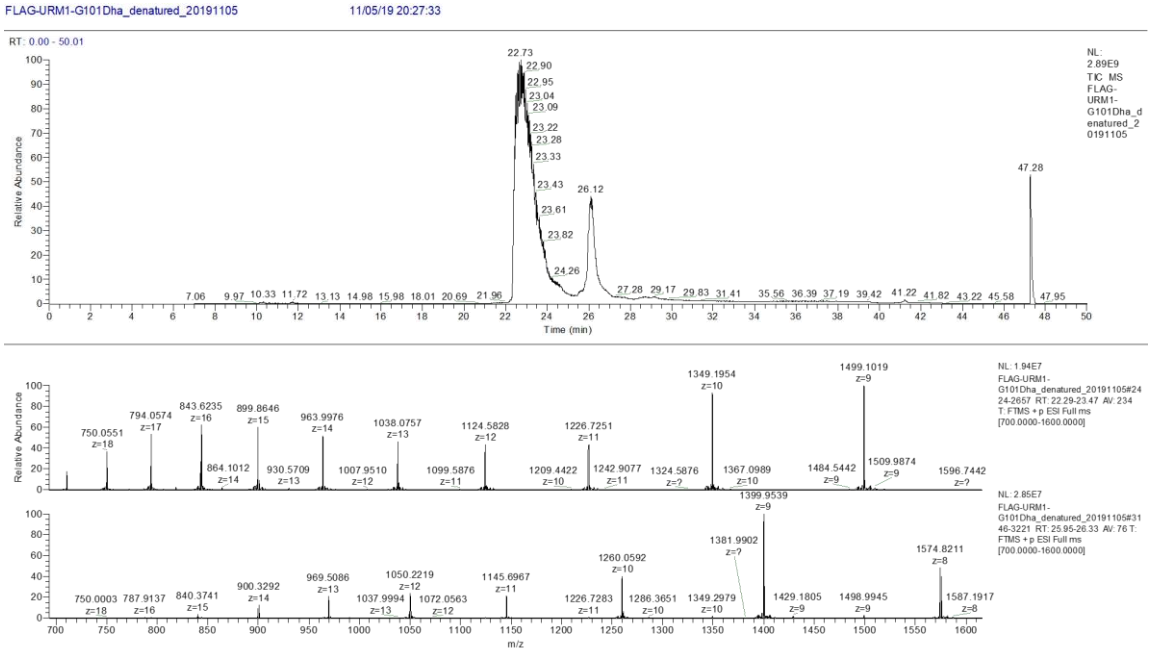

Figure 147 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-URM1-G101C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 22.73 min peak represented FLAG-URM1-G101Dha-6H. The mass signal pattern in the bottom mass diagram that came from the 26.12 min peak showed the formation of hydrolysis product, FLAG-URM1$_{(1\text{-}100)}$.

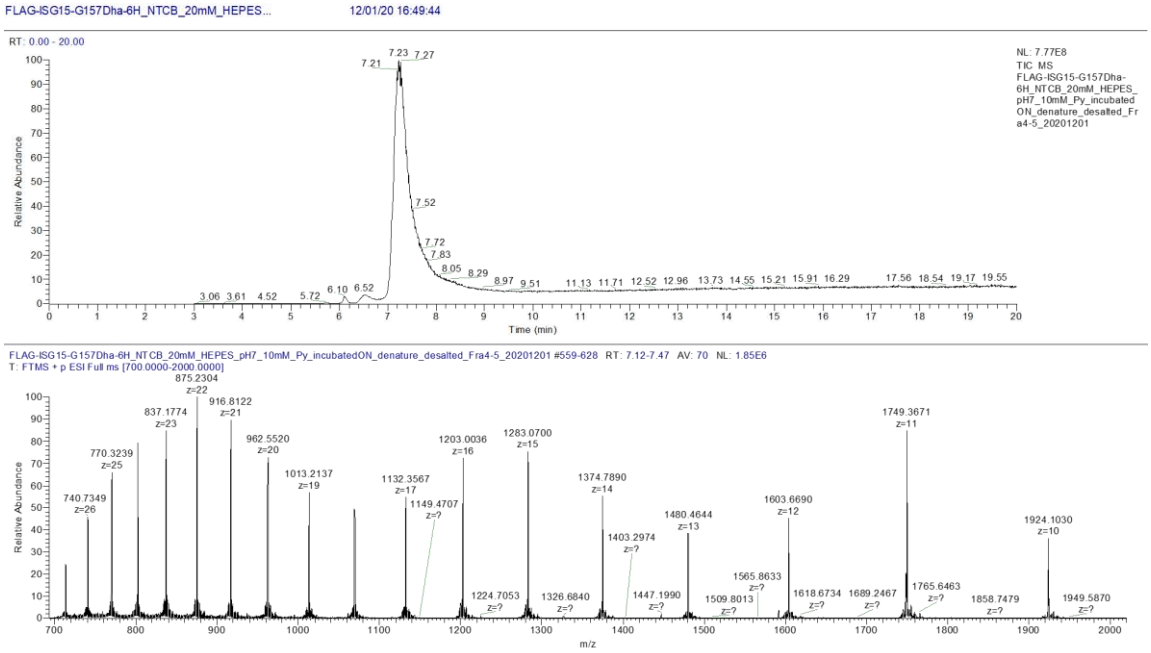

Figure 148 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-ISG15-C89S-G157C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern that came from the 7.23 min peak represented FLAG- ISG15-C89S-G157Dha-6H. The formation of hydrolysis product was too minor to be detected.

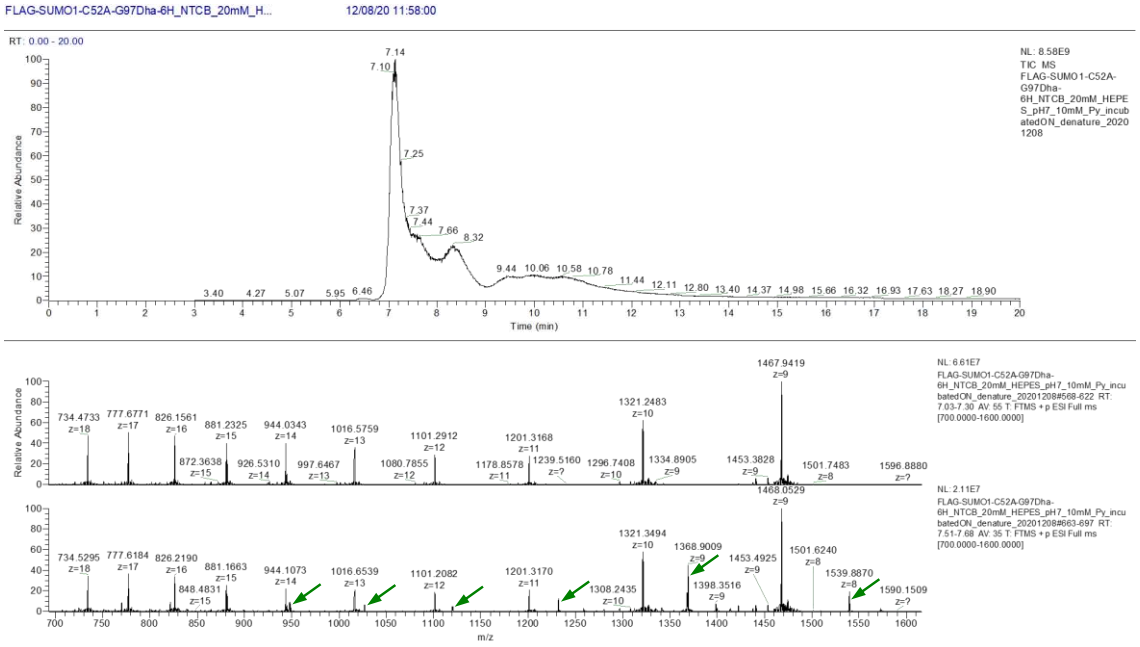

Figure 149 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-SUMO1-C52A-G97C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 7.14 min peak represented FLAG-SUMO1-C52A-G97Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the tailing area (~7.66 min) of Dha product peak showed the formation of hydrolysis product, FLAG-SUMO1$_{(1\text{-}96)}$-C52A.

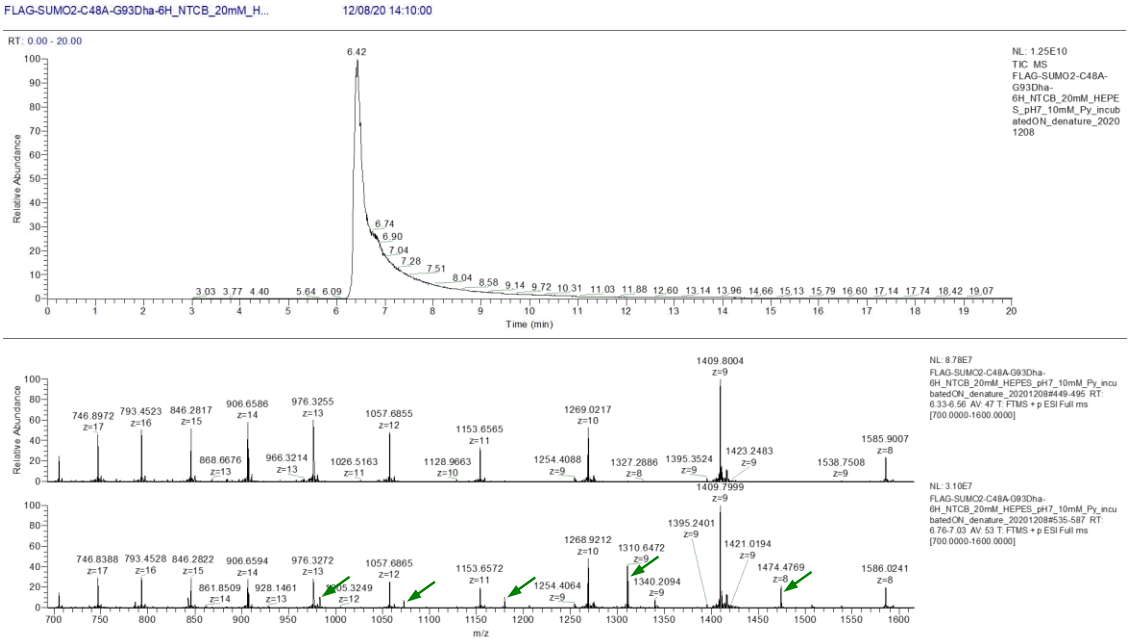

Figure 150 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-SUMO2-C48A-G93C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 6.42 min peak represented FLAG- SUMO2-C48A-G93Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the tailing area of Dha product peak showed the formation of hydrolysis product, FLAG-SUMO2$_{(1\text{-}92)}$-C48A.

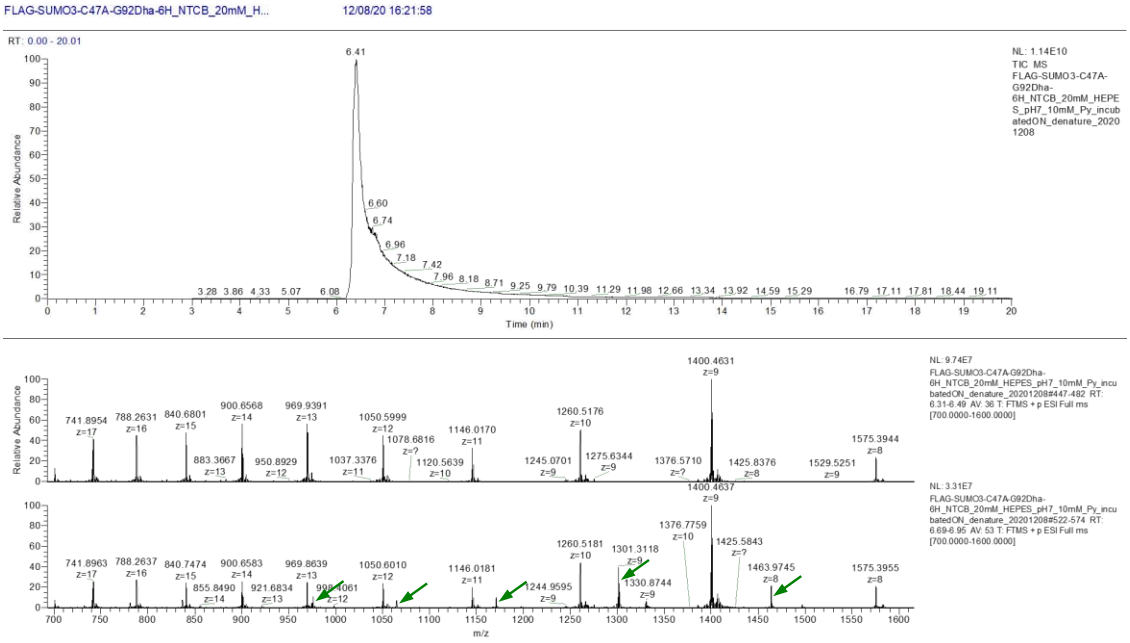

Figure 151 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-SUMO3-C47A-G92C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 6.41 min peak represented FLAG- SUMO3-C47A-G92Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the tailing area of Dha product peak showed the formation of hydrolysis product, FLAG-SUMO3$_{(1\text{-}91)}$-C47A.

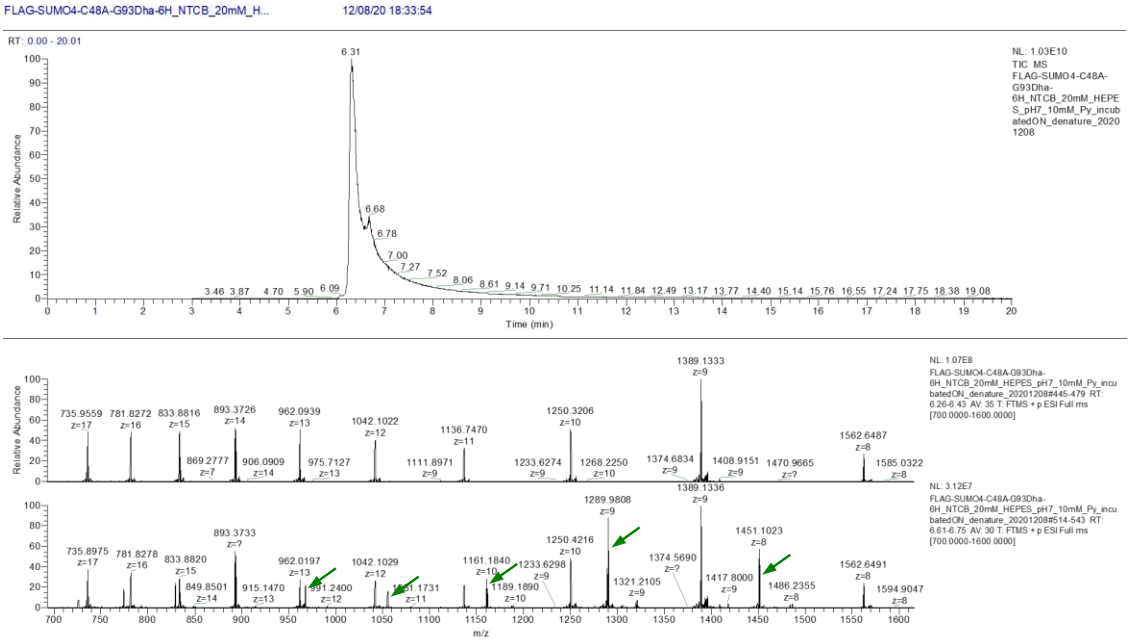

Figure 152 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using FLAG-SUMO4-C48A-G93C-6H as the substrate. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 6.31 min peak represented FLAG- SUMO4-C48A-G93Dha-6H. The mass signal pattern indicated in green in the bottom mass diagram that came from the tailing area (~6.68 min) of Dha product peak showed the formation of hydrolysis product, FLAG-SUMO4$_{(1\text{-}92)}$-C48A.

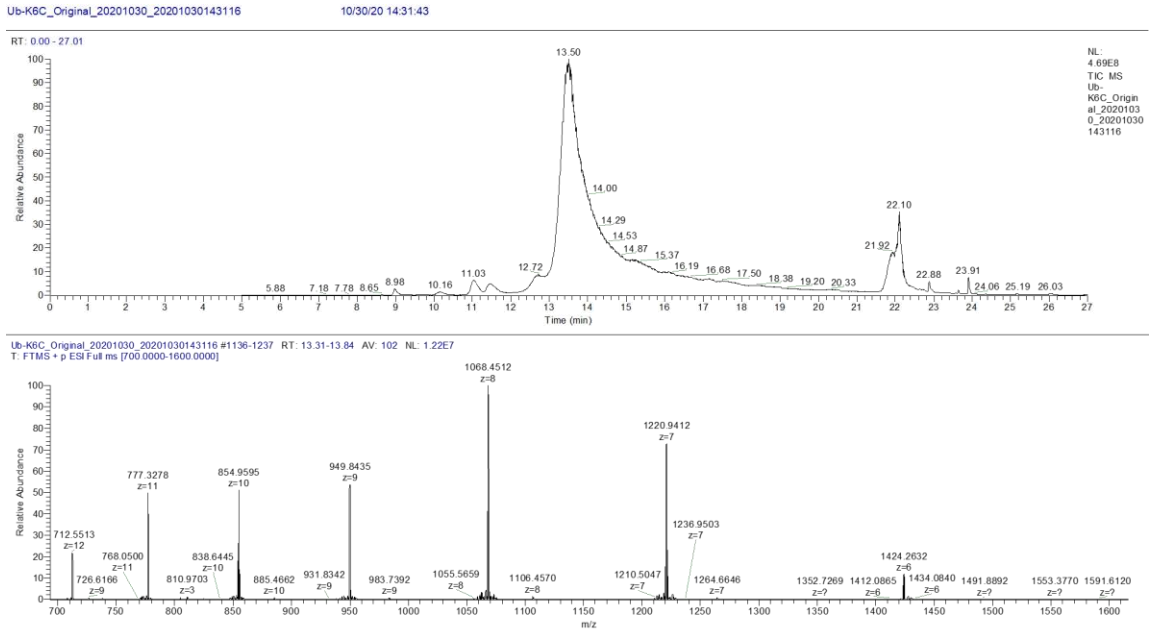

Figure 153 ESI-LC-MS analysis and deconvoluted spectra of purified Ub-K6C.

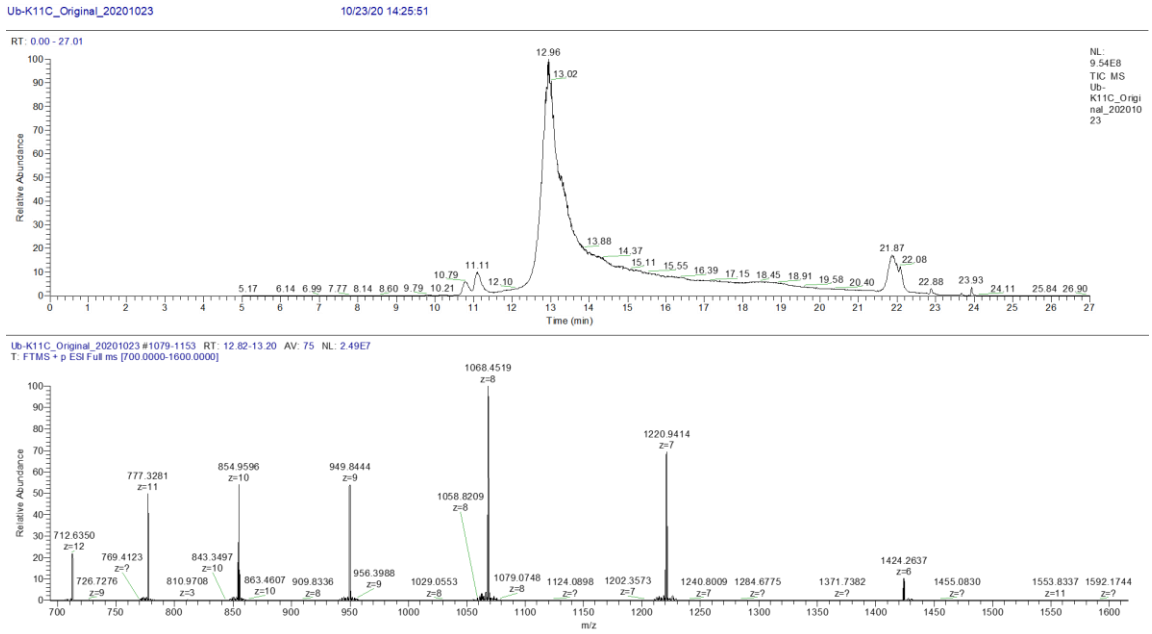

Figure 154 ESI-LC-MS analysis and deconvoluted spectra of purified Ub-K11C.

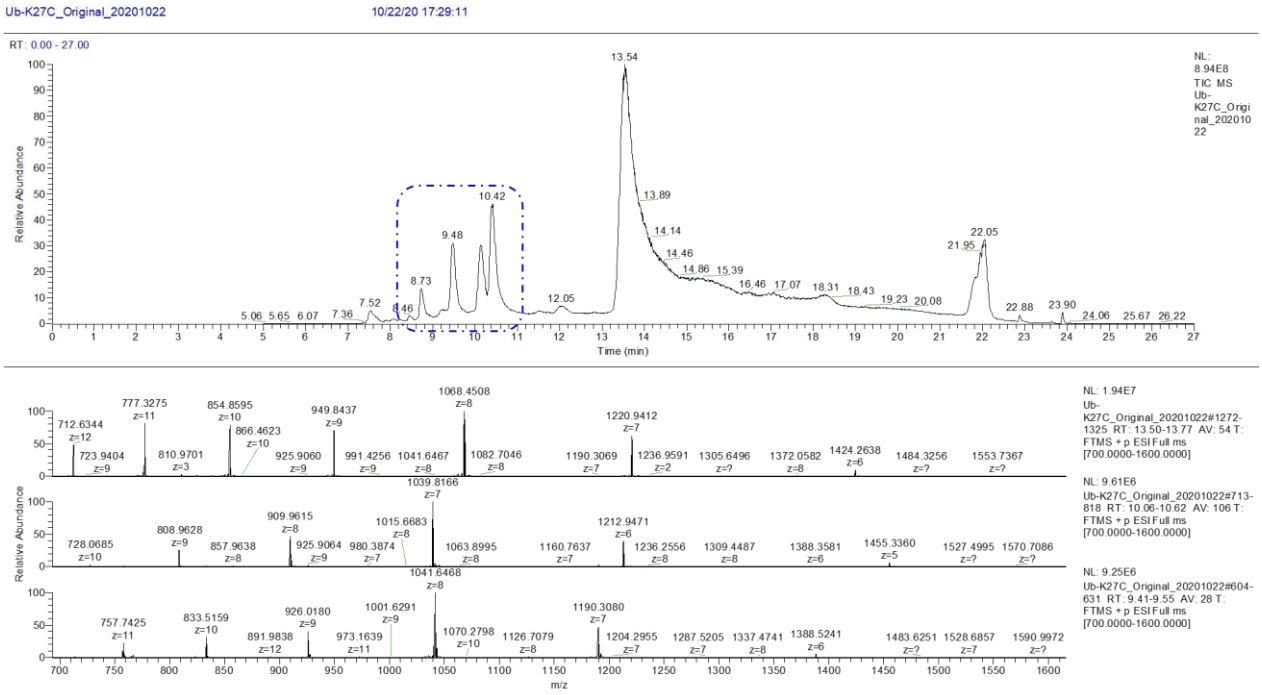

Figure 155 ESI-LC-MS analysis and deconvoluted spectra of purified Ub-K27C.

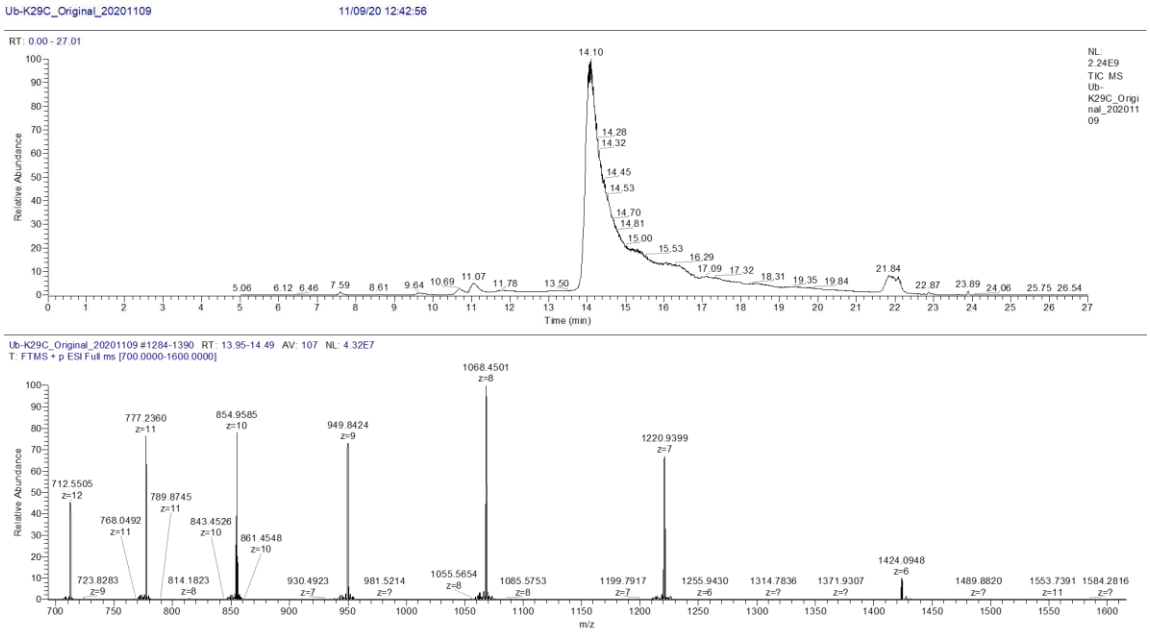

Figure 156 ESI-LC-MS analysis and deconvoluted spectra of purified Ub-K29C.

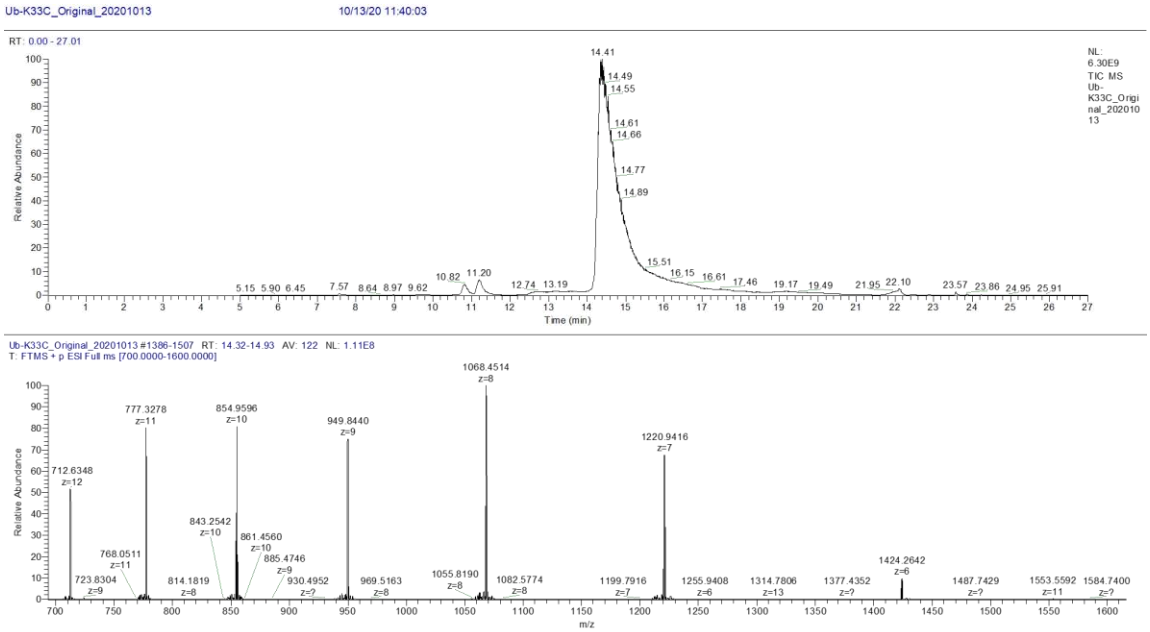

Figure 157 ESI-LC-MS analysis and deconvoluted spectra of purified Ub-K33C.

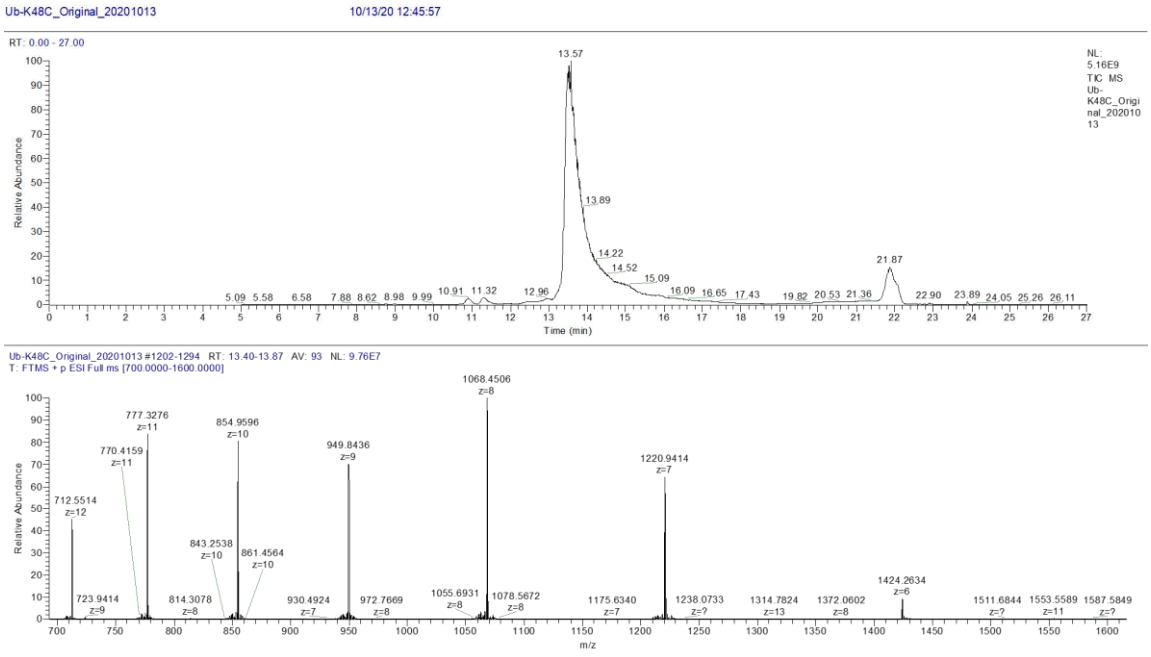

Figure 158 ESI-LC-MS analysis and deconvoluted spectra of purified Ub-K48C.

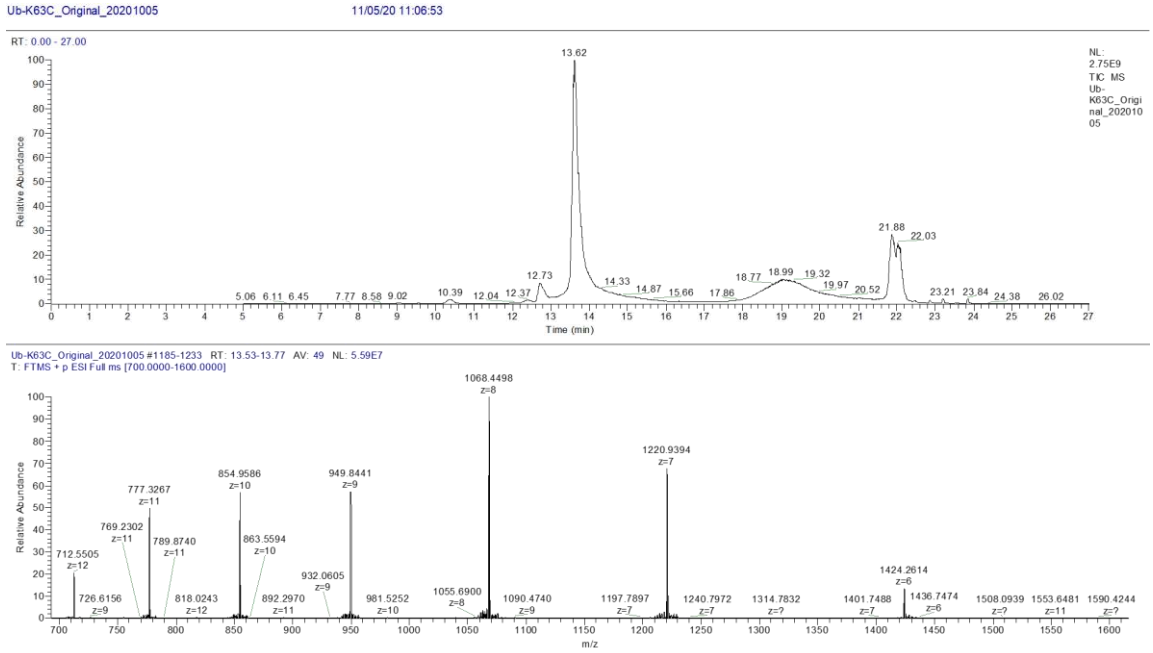

Figure 159 ESI-LC-MS analysis and deconvoluted spectra of purified Ub-K63C.

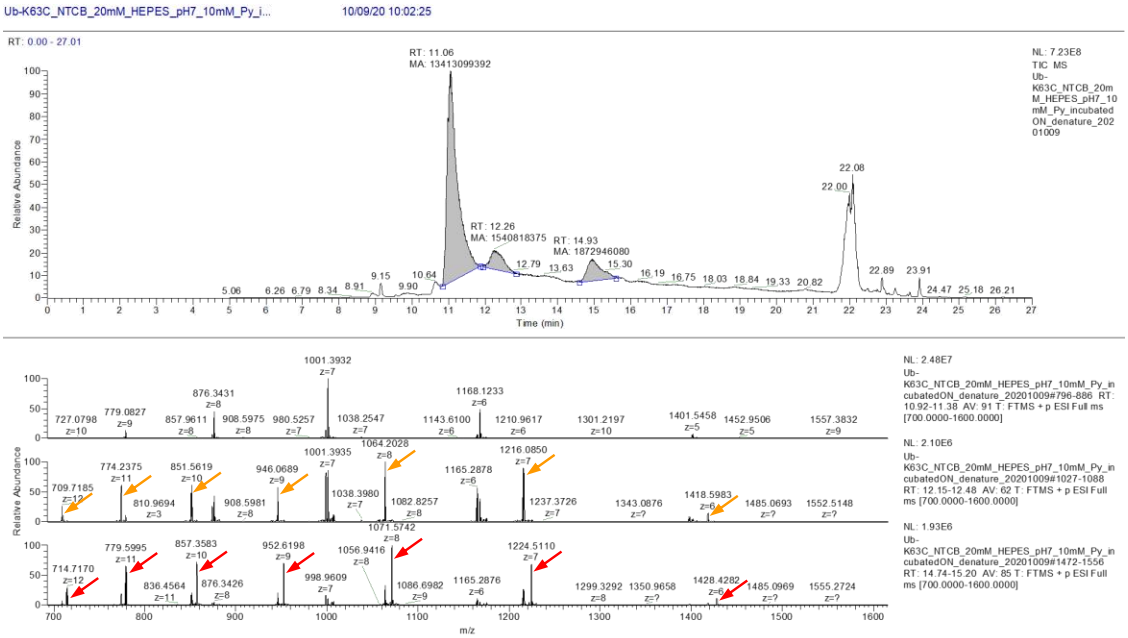

Figure 160 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K63C as the substrate under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 11.06 min peak determined the formation of hydrolysis product, $Ub_{(1-62)}$. The mass signal pattern indicated in orange in the middle mass diagram that came from the 12.26 min peak represented Ub-K63Dha. The mass signal pattern indicated in red in the bottom mass diagram that came from the 14.93 min peak showed the formation of +CN product, Ub-K63C(CN).

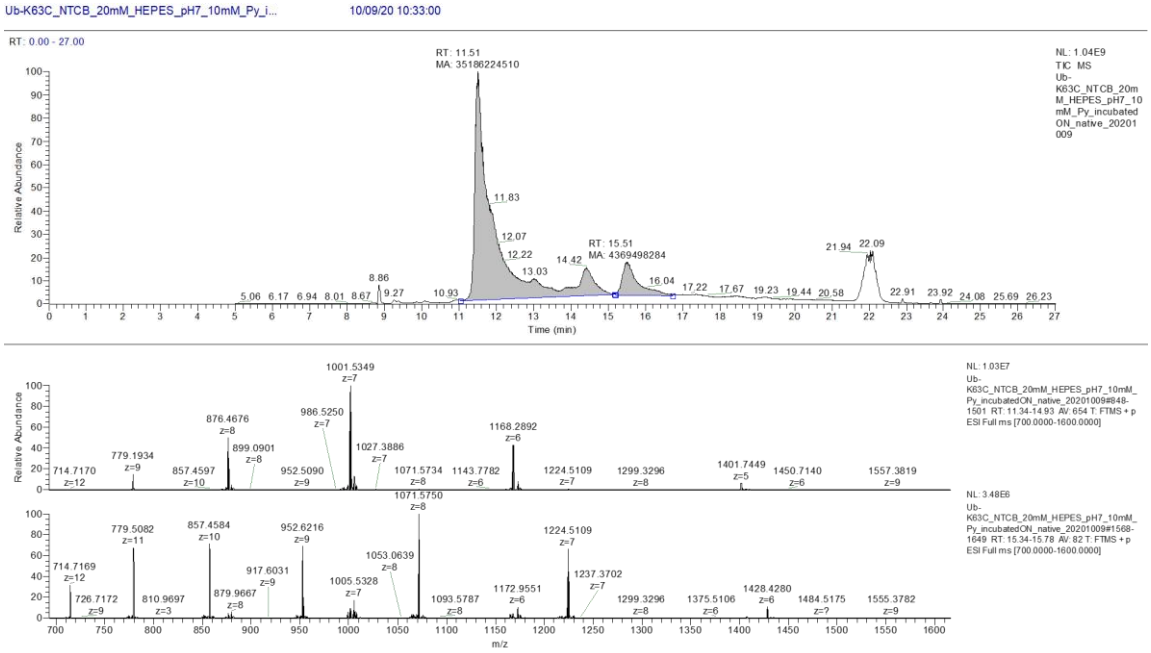

Figure 161 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K63C as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 11.51 min peak determined the formation of hydrolysis product, $Ub_{(1-62)}$. The mass signal pattern in the bottom mass diagram that came from the 15.51 min peak showed the formation of +CN product, Ub-K63C(CN).

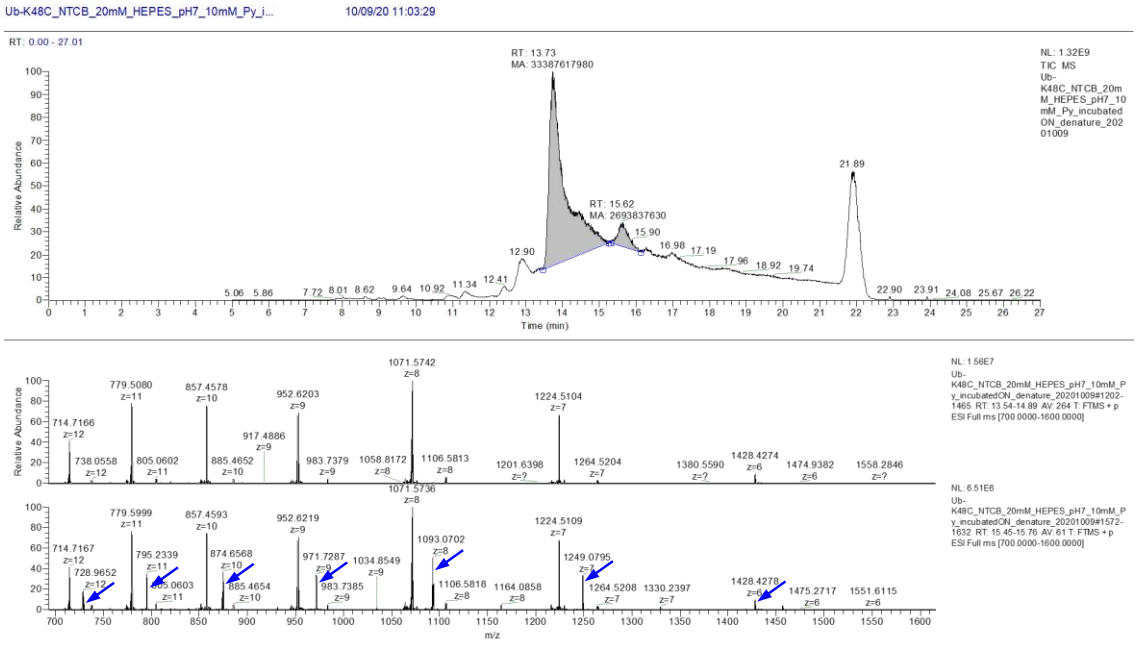

Figure 162 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K48C as the substrate under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.73 min peak represented +CN product, Ub-K48C(CN). The mass signal pattern in the bottom mass diagram indicated in blue that came from the 15.62 min peak showed the formation of +TNB product, Ub-K48C(TNB).

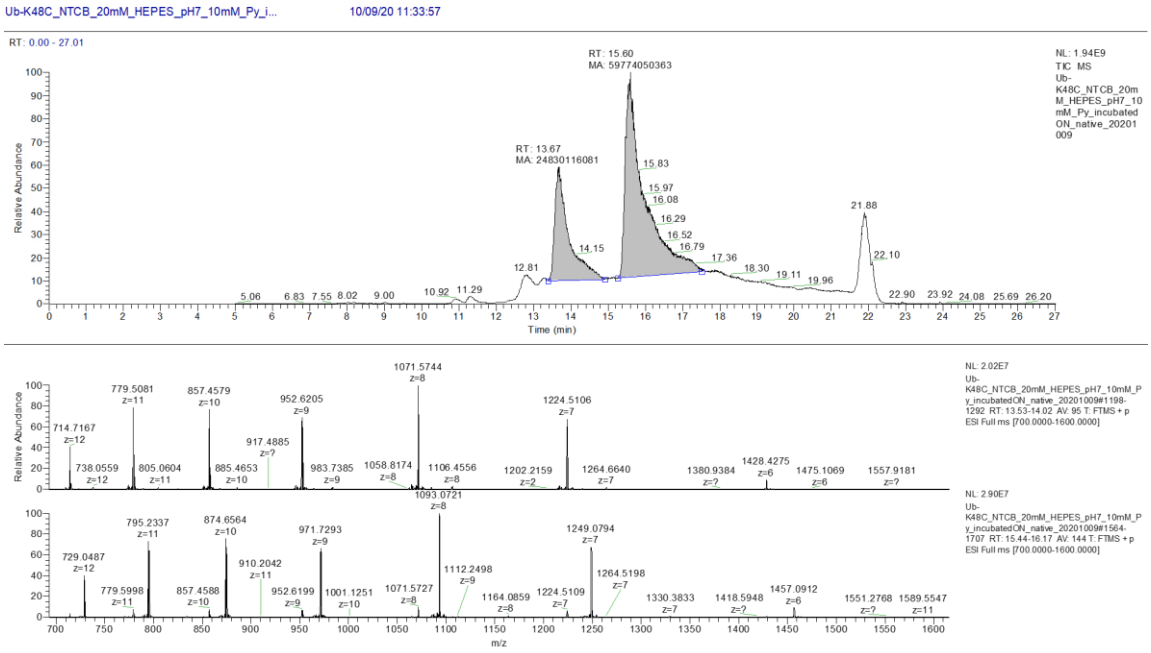

Figure 163 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K48C as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.67 min peak represented +CN product, Ub-K48C(CN). The mass signal pattern in the bottom mass diagram that came from the 15.60 min peak showed the formation of +TNB product, Ub-K48C(TNB).

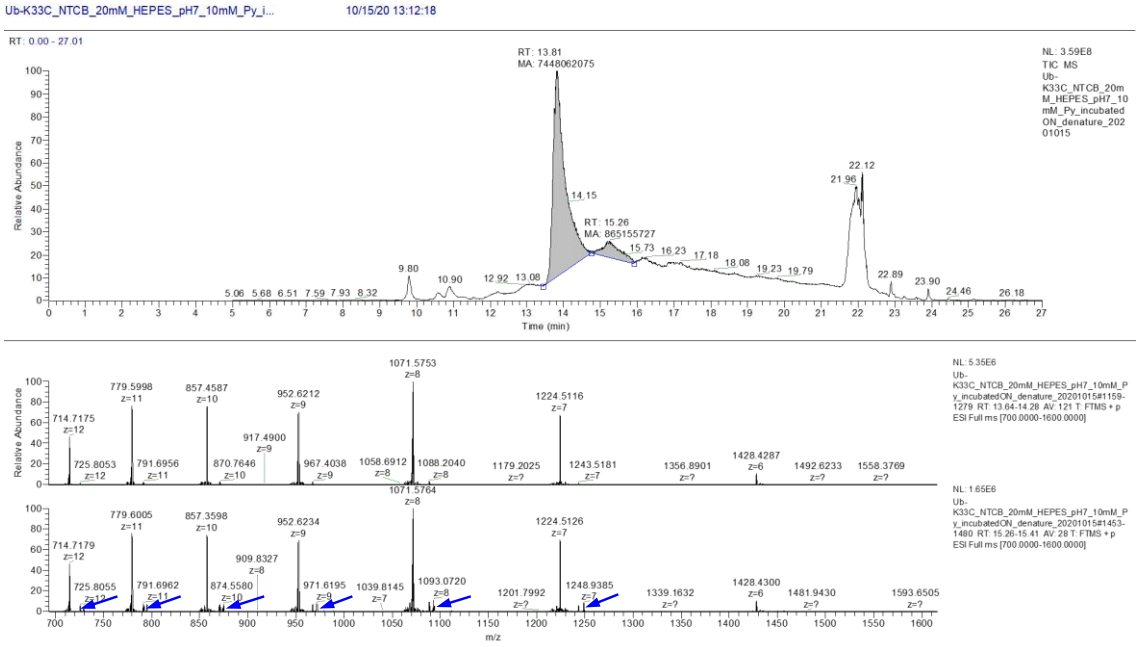

Figure 164 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K33C as the substrate under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.81 min peak represented +CN product, Ub-K33C(CN). The mass signal pattern in the bottom mass diagram indicated in blue that came from the 15.26 min peak showed the formation of +TNB product, Ub-K33C(TNB).

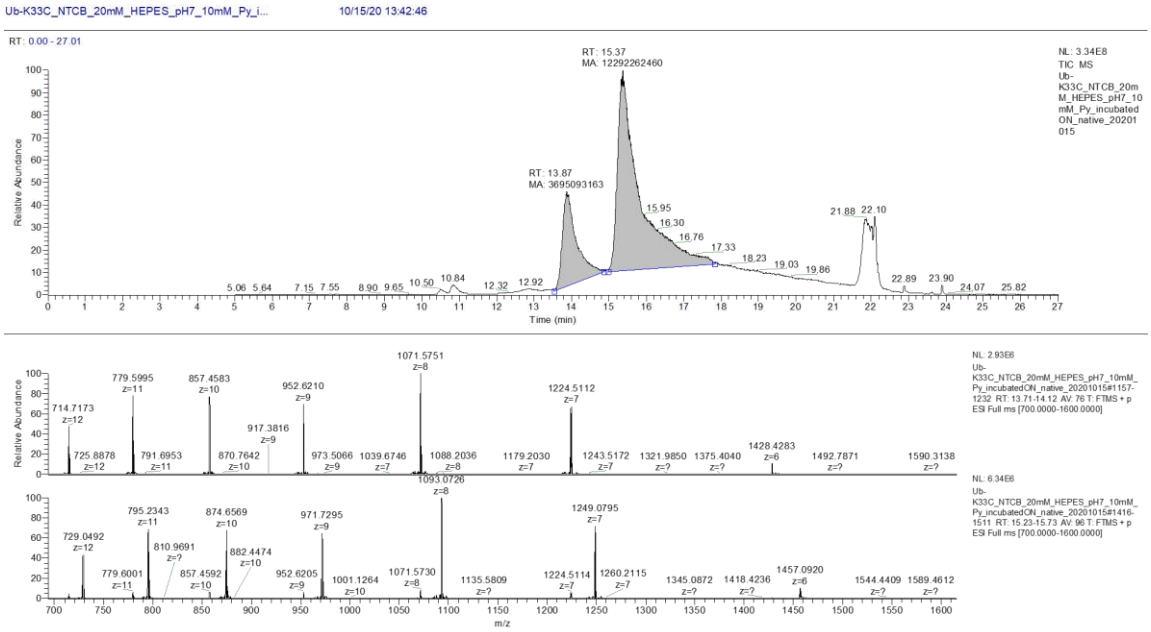

Figure 165 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K33C as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.87 min peak represented +CN product, Ub-K33C(CN). The mass signal pattern in the bottom mass diagram that came from the 15.37 min peak showed the formation of +TNB product, Ub-K33C(TNB).

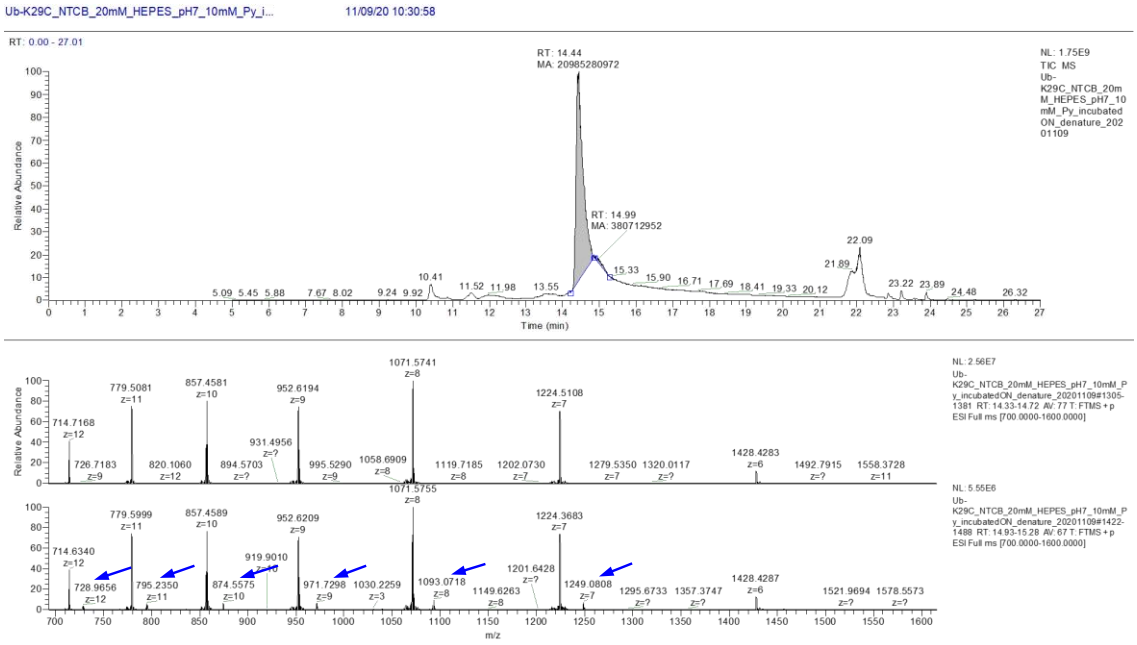

Figure 166 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K29C as the substrate under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 14.44 min peak represented +CN product, Ub-K29C(CN). The mass signal pattern in the bottom mass diagram indicated in blue that came from the 14.99 min peak showed the formation of +TNB product, Ub-K29C(TNB).

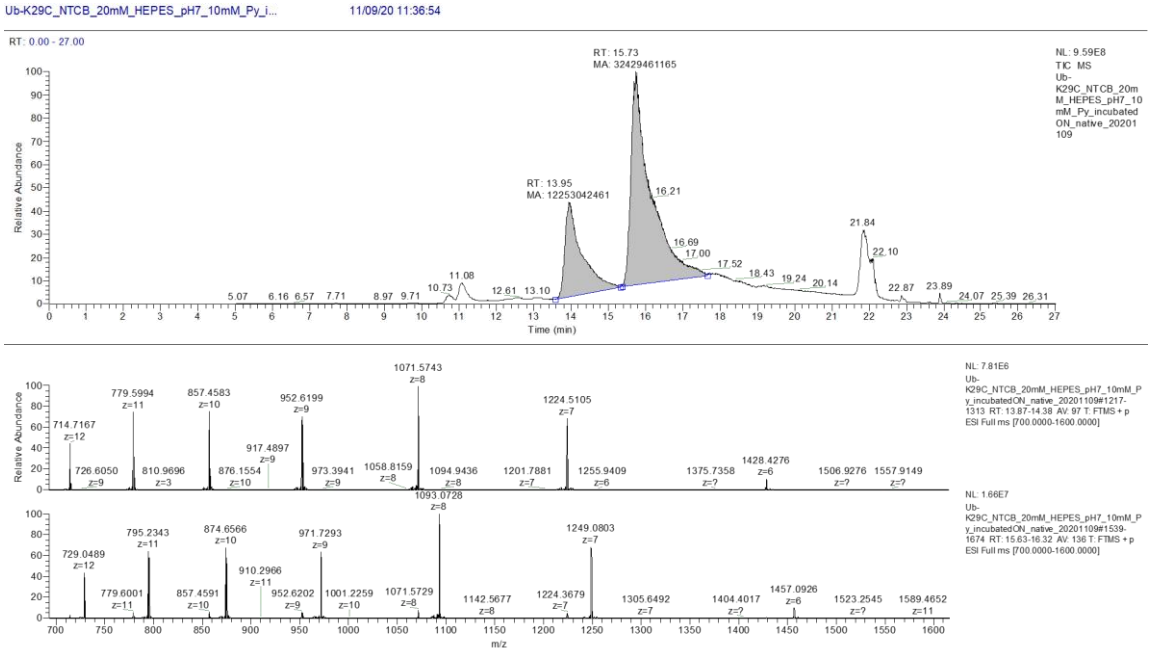

Figure 167 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K29C as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.95 min peak represented +CN product, Ub-K29C(CN). The mass signal pattern in the bottom mass diagram that came from the 15.73 min peak showed the formation of +TNB product, Ub-K29C(TNB).

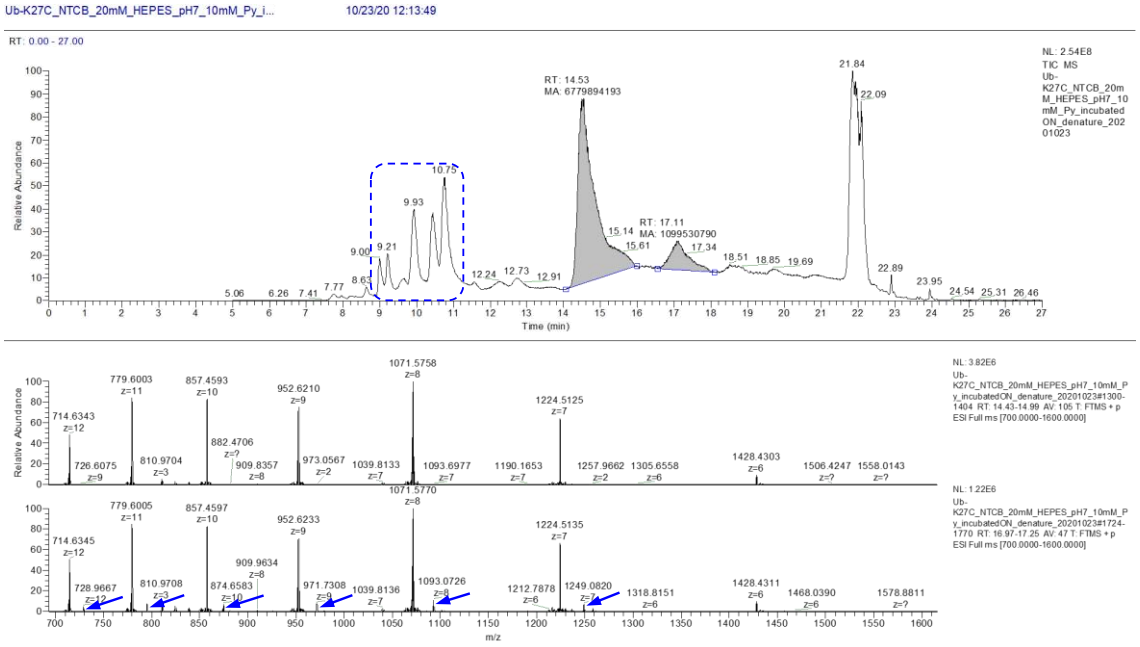

Figure 168 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K27C as the substrate under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 14.53 min peak represented +CN product, Ub-K27C(CN). The mass signal pattern in the bottom mass diagram indicated in blue that came from the 17.11 min peak showed the formation of +TNB product, Ub-K27C(TNB). The dash line circled peaks were unreactive impurities from original Ub-K27C sample.

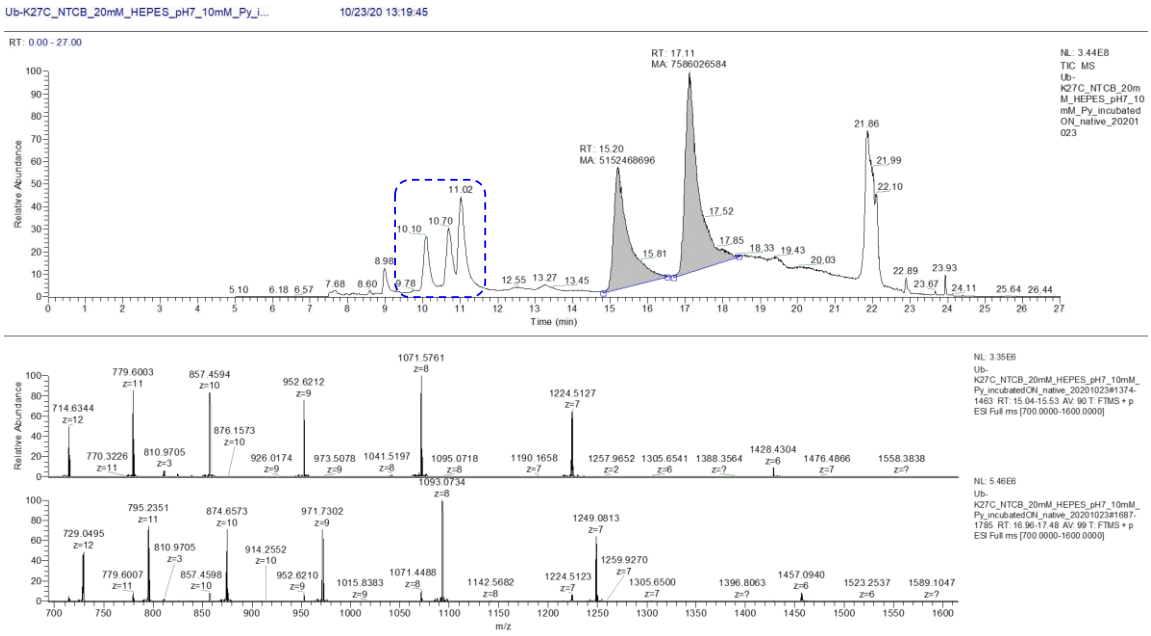

Figure 169 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K27C as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 15.20 min peak represented +CN product, Ub-K27C(CN). The mass signal pattern in the bottom mass diagram that came from the 17.11 min peak showed the formation of +TNB product, Ub-K27C(TNB). The dash line circled peaks were unreactive impurities from original Ub-K27C sample.

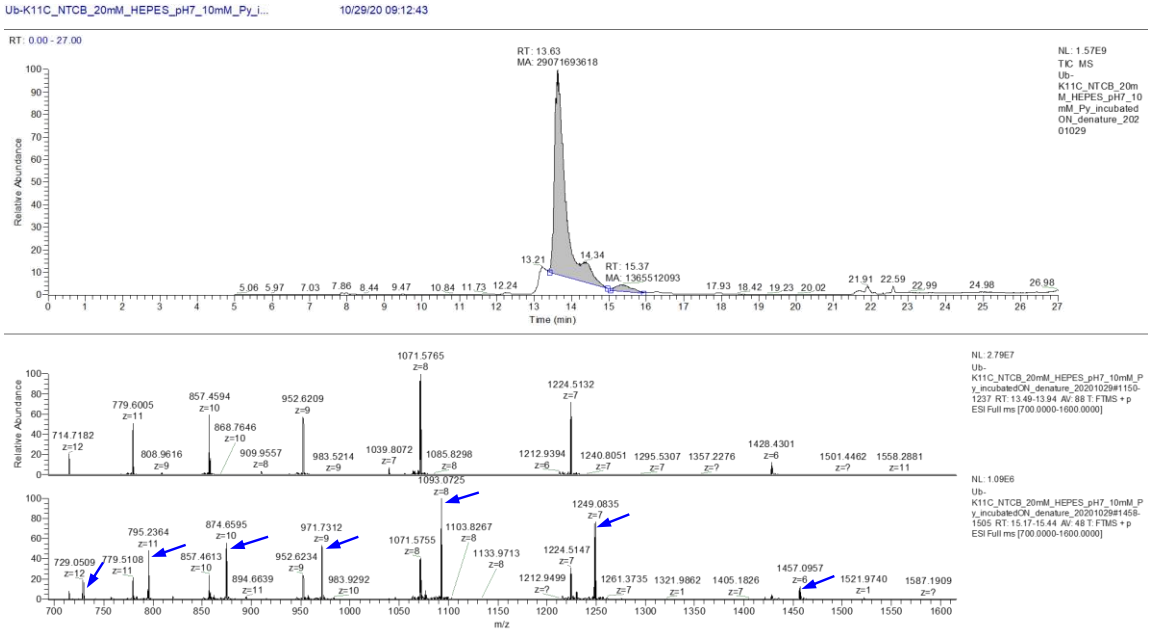

Figure 170 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K11C as the substrate under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.63 min peak represented +CN product, Ub-K11C(CN). The mass signal pattern in the bottom mass diagram indicated in blue that came from the 15.37 min peak showed the formation of +TNB product, Ub-K11C(TNB).

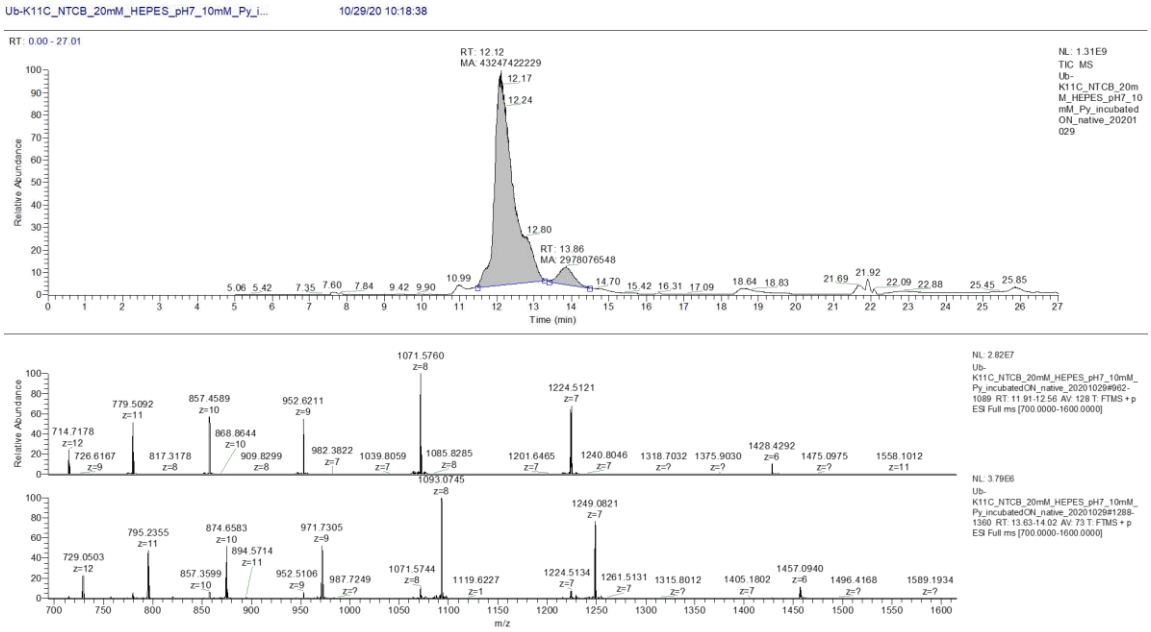

Figure 171 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K11C as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 12.12 min peak represented +CN product, Ub-K11C(CN). The mass signal pattern in the bottom mass diagram that came from the 13.86 min peak showed the formation of +TNB product, Ub-K11C(TNB).

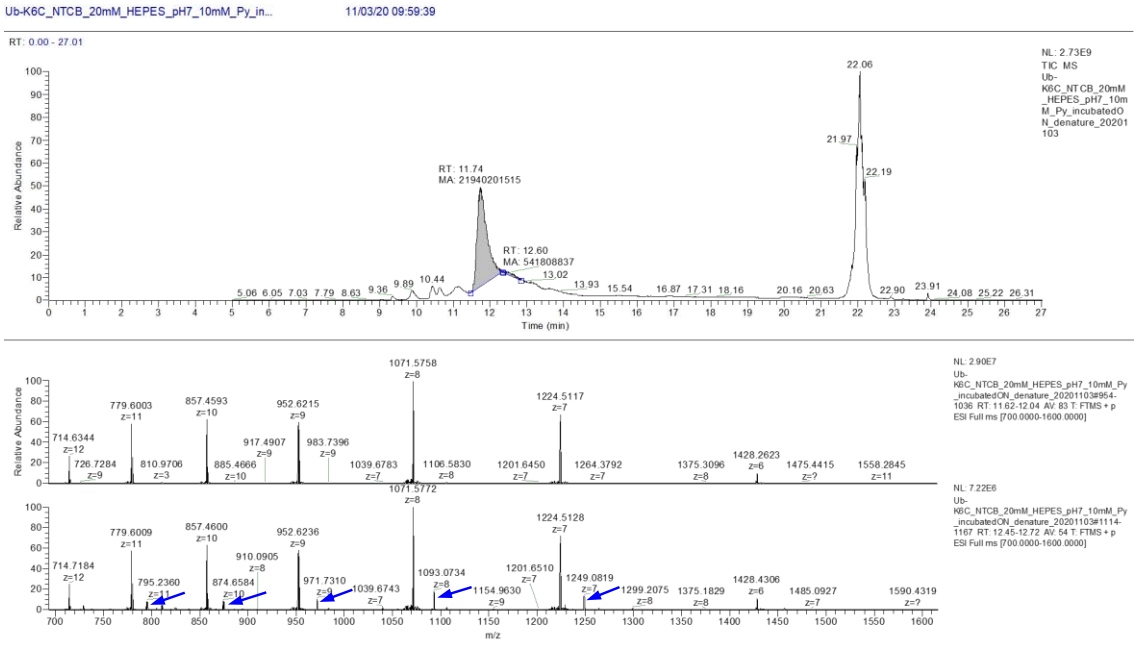

Figure 172 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K6C as the substrate under denatured condition. Reaction condition: 20 mM HEPES pH 7, 6 M GndCl, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 11.74 min peak represented +CN product, Ub-K6C(CN). The mass signal pattern in the bottom mass diagram indicated in blue that came from the 12.60 min peak showed the formation of +TNB product, Ub-K6C(TNB).

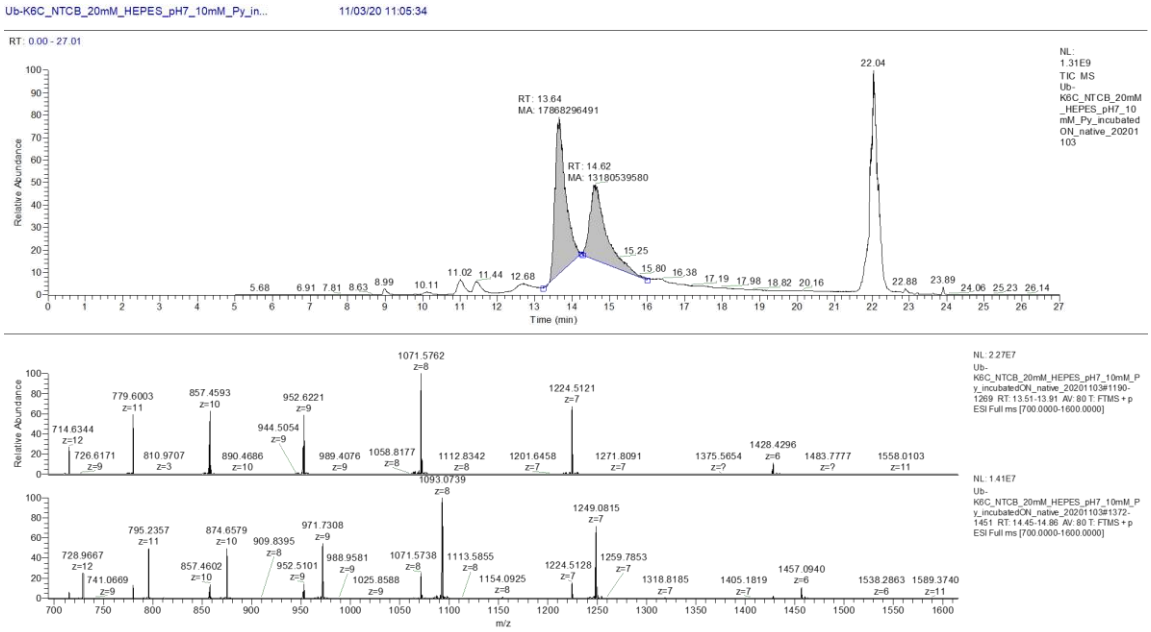

Figure 173 ESI-LC-MS analysis of product mixture of NTCB induced Dha formation reaction using Ub-K6C as the substrate under native condition. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 13.64 min peak represented +CN product, Ub-K6C(CN). The mass signal pattern in the bottom mass diagram that came from the 14.62 min peak showed the formation of +TNB product, Ub-K6C(TNB).

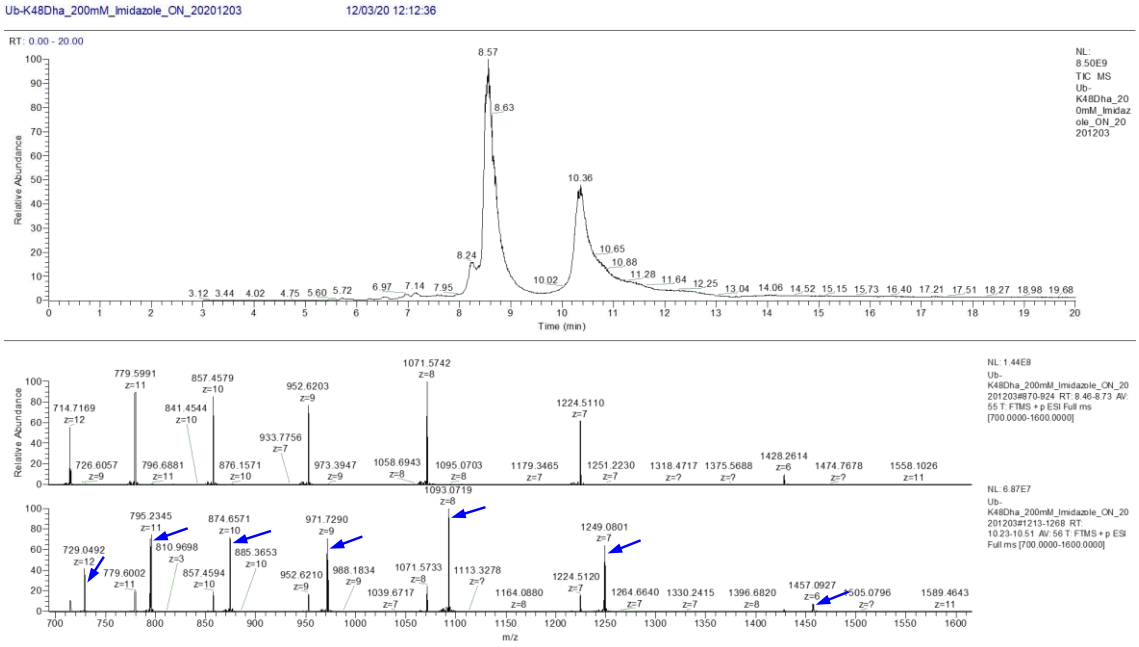

Figure 174 ESI-LC-MS analysis of Ub-K48C(CN) after incubating with 200 mM imidazole overnight. Reaction condition: 20 mM HEPES pH 7, 37 °C, 18 h incubation. The mass signal pattern in the top mass diagram that came from the 8.57 min peak represented +CN product, Ub-K6C(CN). The mass signal pattern in the bottom mass diagram indicated in blue that came from the 10.36 min peak referred to +TNB product, Ub-K6C(TNB).

APPENDIX B

SEQUENCES OF PCR PRIMERS AND RECOMBINANTLY EXPRESSED PROTEIN

Table 9 Sequences of PCR primers used for cloning.

Ub-G76C-6H / Ub-C-6H forward primer	CACCATCACTGAGGTACCCTCGAGTCTGGT
Ub-G76C-6H reverse primer	ATGGTGATGGCAACCTCTGAGACGGAGTAC
Ub-C-6H reverse primer	ATGGTGATGGCAACCACCTCTGAGACGGAG
Ub-G75P-G76C-6H mutagenesis forward primer	ACCTGGTACTCCGTCTCAGACCGTGCCATCACC
Ub-G75P-G76C-6H mutagenesis reverse primer	GGTGATGGCACGGTCTGAGACGGAGTACCAGGT
Ub-G75T-G76C-6H mutagenesis forward primer	ACCTGGTACTCCGTCTCAGAACCTGCCATCACC
Ub-G75T-G76C-6H mutagenesis reverse primer	GGTGATGGCACAGTCTGAGACGGAGTACCAGGT
Ub-G75L-G76C-6H mutagenesis forward primer	ACCTGGTACTCCGTCTCAGACTGTGCCATCACC
Ub-G75L-G76C-6H mutagenesis reverse primer	GGTGATGGCACAGTCTGAGACGGAGTACCAGGT
Ub-G75E-G76C-6H mutagenesis forward primer	ACCTGGTACTCCGTCTCAGATGGTGCCATCACC
Ub-G75E-G76C-6H mutagenesis reverse primer	GGTGATGGCATTCTCTGAGACGGAGTACCAGGT
Ub-G75R-G76C-6H mutagenesis forward primer	ACCTGGTACTCCGTCTCAGACGTTGCCATCACC

Table 9 Continued

Ub-G75R-G76C-6H mutagenesis reverse primer	GGTGATGGCAACGTCTGAGACGGAGT ACCAGGT
Ub-G75W-G76C-6H mutagenesis forward primer	ACCTGGTACTCCGTCTCAGATGGTGCC ATCACC
Ub-G75W-G76C-6H mutagenesis reverse primer	GGTGATGGCACCATCTGAGACGGAGT ACCAGGT
FLAG-Ub-G76C-6H N'-FLAG tag insertion forward primer	ATGATGACAAAGCTGCAATGCAGATC TTCGTGAAGACTCTGAC
FLAG-Ub-G76C-6H N'-FLAG tag insertion reverse primer	CGTCTTTGTAGTCCATATGTATATCTC CTTCTTATACTTAACTAATATACTAAG ATGGGG
FLAG-SUMO1-G97C-6H C52A mutagenesis forward primer	AGCTGAAAGAAAGCTACGCGCAACGT CAGGGCGTGCC
FLAG-SUMO1-G97C-6H C52A mutagenesis reverse primer	TCTTCAGATGGGTAGTCATCTTCACTT TGAAGTGG
FLAG-SUMO2-G93C-6H C48A mutagenesis forward primer	AGCTGATGAAGGCCTACGCGGAACGC CAAGGTCTGAGC
FLAG-SUMO2-G93C-6H C48A mutagenesis reverse primer	TGCTCAGCGGAGTATGGCGTTTG
FLAG-SUMO3-G92C-6H C47A mutagenesis forward primer	AGGCCTACGCGGAACGCCAAGGTCTG AGCATGCG
FLAG-SUMO3-G92C-6H C47A mutagenesis reverse primer	TCATCAGCTTGCTCAGCGGAGTATGG
FLAG-SUMO4-G93C-6H C48A mutagenesis forward primer	ACTGATGAAAGCCTATGCGGAACCAC GCGGTCTGAGCATG
FLAG-SUMO4-G93C-6H C48A mutagenesis reverse primer	TTGCTCAGCGGAGTCTGGCG

Table 9 Continued

H2A *N*-His tag deletion forward primer	TCTGGTCGTGGTAAACAAGGTGGTAAAGC
H2A *N*-His tag deletion reverse primer	CATGGTATATCTCCTTCTTAAAGTTAAACAAAATTATTTCTAGAGG
H2A *C*-His tag insertion forward primer	TGCCATCACCATCATCACCACTGAAAGCTTGCGGCCGCATAATG
H2A *C*-His tag insertion reverse primer	ACCTTTCGCTTTGTGGTGGGATTC
pET28a-RNH59-196-6H MEGAWHOP forward primer	CAATTCCCCTCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATACCATGGCGAAGGAAGAGATT
pET28a-RNH59-196-6H MEGAWHOP reverse primer	TTAGCAGCCGGATCTCAGTGATGATGGTGATGGTGTTTGCGCCCGTAGTC
RNH_{59-196}-6H K190C mutagenesis forward primer	AAGTGGGGTGAGATCTGTGCGGACTACGGGC
RNH_{59-196}-6H K190C mutagenesis reverse primer	GCCCGTAGTCCGCACAGATCTCACCCCACTT
pET28a-exenatide-40C-GGGSA-Strep MEGAWHOP forward primer	GATAACGATATTATTGAGGCTCACAGAGAACAGATTGGTGGACACGGTGAAGGTACGTTC
pET28a-exenatide-40C-GGGSA-Strep MEGAWHOP reverse primer	GCTTGTCGACGGAGCTCGAATTCGGATCACTTCTCAAACTGAGGATGTG
exenatide-S39C-SA-Strep mutagenesis forward primer	TGCTCCTCCTCCGTGTTCAGCGTGGTCACATC
exenatide-S39C-SA-Strep mutagenesis reverse primer	GCACCTGAGGACGGTCCGCCGTTCTTCAAC

Table 10 Sequences of all expressed recombinant proteins.

Ub	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGG
Ub-K6C	MQIFVCTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGG
Ub-K11C	MQIFVKTLTGCTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGG
Ub-K27C	MQIFVKTLTGKTITLEVEPSDTIENVCAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGG
Ub-K29C	MQIFVKTLTGKTITLEVEPSDTIENVKACIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGG
Ub-K33C	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDCEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGG
Ub-K48C	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGCQLEDGRTLSDYNIQKESTLHLVLR LRGG

Table 10 Continued

Ub-K63C	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQCESTLHLVLR LRGG
Ub-G76C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGCHHHHHH
Ub-G75P-G76C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRPCHHHHHH
Ub-G75T-G76C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRTCHHHHHH
Ub-G75L-G76C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRLCHHHHHH
Ub-G75E-G76C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRECHHHHHH
Ub-G75R-G76C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRRCHHHHHH
Ub-G75W-G76C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRWCHHHHHH
Ub-C-6H	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIP PDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLR LRGGCHHHHHH

Table 10 Continued

FLAG-Ub-G76C-6H	MDYKDDDDKAAMQIFVKTLTGKTITLEVEPSDTIE NVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDY NIQKESTLHLVLRLRGCHHHHHH
FLAG-SUMO1(C52A)-G97C-6H	MDYKDDDDKAAMSDQEAKPSTEDLGDKKEGEYI KLKVIGQDSSEIHFKVKMTTHLKKLKESYAQRQG VPMNSLRFLFEGQRIADNHTPKELGMEEEDVIEVY QEQTGCHHHHHH
FLAG-SUMO2(C48A)-G93C-6H	MDYKDDDDKAAMADEKPKEGVKTENNDHINLKV AGQDGSVVQFKIKRHTPLSKLMKAYAERQGLSMR QIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTG CHHHHHH
FLAG-SUMO3(C47A)-G92C-6H	MDYKDDDDKAAMSEEKPKEGVKTENDHINLKVA GQDGSVVQFKIKRHTPLSKLMKAYAERQGLSMRQ IRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGC HHHHHH
FLAG-SUMO4(C48A)-G93C-6H	MDYKDDDDKAAMANEKPTEEVKTENNNHINLKV AGQDGSVVQFKIKRQTPLSKLMKAYAEPRGLSMK QIRFRFGGQPISGTDKPAQLEMEDEDTIDVFQQPTG CHHHHHH
FLAG-NEDD8-G76C-6H:	MDYKDDDDKAAMLIKVKTLTGKEIEIDIEPTDKVE RIKERVEEKEGIPPQQQRLIYSGKQMNDEKTAADY KILGGSVLHLVLALRGCHHHHHH
FLAG-ISG15(C89S)-G157C-6H	MDYKDDDDKAAMGWDLTVKMLAGNEFQVSLSS SMSVSELKAQITQKIGVHAFQQRLAVHPSGVALQD RVPLASQGLGPGSTVLLVVDKSDEPLSILVRNNKG RSSTYEVRLTQTVAHLKQQVSGLEGVQDDLFWLT FEGKPLEDQLPLGEYGLKPLSTVFMNLRLRGCHHH HHH

Table 10 Continued

FLAG-GABARAP-G116C-6H	MDYKDDDDKAAMKFVYKEEHPFEKRRSEGEKIR KKYPDRVPVIVEKAPKARIGDLDKKKYLVPSDLTV GQFYFLIRKRIHLRAEDALFFFVNNVIPPTSATMGQ LYQEHHEEDFFLYIAYSDESVYCHHHHHH
FLAG-GABARAPL2-G116C-6H	MDYKDDDDKAAMKWMFKEDHSLEHRSVESAKIR AKYPDRVPVIVEKVSGSQIVDIDKRKYLVPSDITVA QFMWIIRKRIQLPSEKAIFLFVDKTVPQSSLTMGQL YEKEKDEDGFLYVAYSGENTFCHHHHHH
FLAG-UFM1-G83C-6H	MDYKDDDDKAAMSKVSFKITLTSDPRLPYKVLSV PESTPFTAVLKFAAEEFKVPAATSAIITNDGIGINPA QTAGNVFLKHGSELRIIPRDRVCHHHHHH
FLAG-URM1-G101C-6H	MDYKDDDDKAAMAAPLSVEVEFGGGAELLFDGI KKHRVTLPGQEEPWDIRNLLIWIKKNLLKERPELFI QGDSVRPGILVLINDADWELLGELDYQLQDQDSV LFISTLHGCHHHHHH
FLAG-MNSFβ(C57S)-G74C-6H	MDYKDDDDKAAMQLFVRAQELHTLEVTGQETVA QIKDHVASLEGIAPEDQVVLLAGSPLEDEATLGQS GVEALTTLEVAGRMLGCHHHHHH
H2A-K129C-6H	MSGRGKQGGKARAKAKTRSSRAGLQFPVGRVHR LLRKGNYAERVGAGAPVYLAAVLEYLTAEILELA GNAARDNKKTRIIPRHLQLAIRNDEELNKLLGKVTI AQGGVLPNIQAVLLPKKTESHHKAKGCHHHHHH
RNH$_{59-196}$-K190C-6H	MAKEEIIWESLSVDVGSQGNPGIVEYKGVDTKTGE VLFEREPIPIGTNNMGEFLAIVHGLRYLKERNSRKP IYSDSQTAIKWVKDKKAKSTLVRNEETALIWKLV DEAEEWLNTHTYETPILKWQTDKWGEICADYGRK HHHHHH

Table 10 Continued

SUMO Protease	MGSSHHHHHHSSGLVPRGSHMLVPELNEKDDDQ VQKALASRENTQLMNRDNIEITVRDFKTLAPRRW LNDTIIEFFMKYIEKSTPNTVAFNSFFYTNLSERGY QGVRRWMKRKKTQIDKLDKIFTPINLNQSHWALG IIDLKKKTIGYVDSLSNGPNAMSFAILTDLQKYVM EESKHTIGEDFDLIHLDCPQQPNGYDCGIYVCMNT LYGSADAPLDFDYKDAIRMRRFIAHLILTDALK
6H-SUMO-exenatide-S39C-SA-Strep	MGSSHHHHHHSSGLVPRGSHMSDSEVNQEAKPEV KPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLME AFAKRQGKEMDSLRFLYDGISIQADQTPEDLDME DNDIIEAHREQIGGHGEGTFTSDLSKQMEEEAVRL FIEWLKNGGPSSGAPPPCSAWSHPQFEK