

THE CHEMICAL SYNTHETIC INVESTIGATION OF PROTEINS WITH SITE-  
SPECIFIC LYSINE POST-TRANSLATIONAL MODIFICATIONS

A Dissertation

by

ZHIPENG WANG

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,	Tadhg P. Begley
	Kevin Burgess
	David Russell
Head of Department,	Simon North

May 2018

Major Subject: Chemistry

Copyright 2018 Zhipeng Wang

## ABSTRACT

In the recent decade, an increasing amount of protein post-translational modifications (PTMs) have been discovered, which are important epigenetic markers widespread on nucleic and cytoplasmic proteins. Lysine (Lys), as the only  $\epsilon$ -amino group containing amino acid, carries many different PTMs, which play important functions in the regulation of chromatin structure and gene transcription. Methylation and acylation as the two main groups of Lys modifications are of great significance to the biochemistry studies. However, the detailed molecular mechanisms and signaling pathways are remaining elusive, due to the difficulty of site-specific incorporation into target proteins for molecular biologists.

Therefore, it becomes a challenging but fruitful task for chemical biologists to prepare sufficient amount of protein of interests (POIs) with PTMs in their homogeneous states. Based on the noncanonical amino acid (ncAA) incorporation technique, the Liu group designed a combinatory methodology for the semi-synthesis of protein with site-specific Lys PTMs. Mainly two systems have been set up based on the introduction of different noncanonical amino acids via amber codon suppression followed by certain bioorthogonal reactions.

This dissertation will begin with a summary of recent PTM functional studies, followed by a detailed analysis of two ncAA systems for the preparation of POIs with site-specific Lys PTMs, *N*<sup>ε</sup>-(4-azidobenzoxycarbonyl)- $\delta,\epsilon$ -dehydrolysine (AcdK) for Lys methylation, and azidonorleucine (AznL) for Lys acylation. The further biological investigations are highlighted to demonstrate their applications. Another AcdK based

design for the installation of Lys acylation via KAHA ligation strategy will be discussed, followed by a brief discussion of some other extensions. Last but not the least, we will talk about a novel visualization strategy for protein fatty acylation modifications by HaloTag recognition. Optimizations are still needed for this on-going project.

With the combination of ncAA incorporation and bioorthogonal reactions, we can theoretically incorporate any Lys PTMs with their native structures into target proteins at specific sites. The protein products we generate can be directly applied for enzymatic assays *in vitro* performing similarly as native proteins. Notably, we also provide two useful functional groups, aldehyde and azido, as very important chemical biology handles for other protein functionalization purposes.

## DEDICATION

To my parents

Without your love, encouragement, and support, this work would not have been completed.

## ACKNOWLEDGEMENTS

First of all, I would like to thank my advisor Professor Dr. Wenshe R. Liu for his guidance, help, and support throughout my graduate study, research, and life. Thanks to Dr. Liu's broad knowledge and deep understanding of both biochemistry and chemical biology, I had the chance to work on many sophisticated designs and fantastic projects.

I would also like to thank my committee members, Professors Tadgh P. Begley, Professor Kevin Burgess, and Professor David Russell for their support and helpful discussions.

Second, I want to thank my colleagues in Liu group for their suggestions and help during my working there. I would like to first thank Erol Can Vatansever for his tolerance of my endless questions and asking for favors. Without your advisor-like scientific discussions, bother-like physical support and father-like mental support, I would have been having many more troubles. Then, I want to thank Jeffery Tharp and Willie Hsu for scientific discussion, kindly joking, lunch eating, and, more importantly, instructions of US culture, language, behavior, and altitude, which is playing a significant role in adjusting myself to be more Americanized. Thirdly, I want to thank three masters, Dr. Yadagiri Kurra, Dr. Yu Zeng, and Dr Yanjun Lee, for been both teachers and collaborators with me, one for organic synthesis, one for molecular biology and one for both. Without standing on your giants' shoulders, I would never achieve my research goals.

For people out of group, I want to first thank Dr. Xin Wang as my collaborator. Besides helping me with hundreds of MS/MS samples, Xin's personal charming still

encourages me to chase hard to be excellent in all aspects. Thanks to Dr. Joshua Yuan, Dr. Susie Dai, Man Li, Brian Young, Yuanyou Wang, Yi Zheng for their scientific help, suggestions, and supports. I want to thank Amy Liu, Sandy Horton, Carrie Nichols, Bo Wang in Chemistry department for all the help necessary to me.

I am also feeling thankful to my friends Xinyu Ma, Yuchen Qiao, Zhenxiong Jiang, Bing Xu, Xuemei Yang, Jie Tian, Xiaoshan Wang, Zhiye Wang, Zhihua Liu and many others for precipitation in the Card game “Sanguosha” parties I organized, from which I can train my skill of organizing and communicating.

For people out of Texas, I want to thank Dr. Mingxuan Wu in Harvard Medical School for all the suggestions, help, and jokes. I also thank Prof. Dr. Lei Liu in Tsinghua University and Prof. Dr. Jishen Zheng in USTC for continuously academic advices and scientific encourages.

Finally, thanks to my mother, father, and entire family members for their care, encouragement and love.

## CONTRIBUTORS AND FUNDING SOURCES

All of the projects in this thesis were conducted under the guidance of Professor Dr. Wenshe R. Liu of the Department of Chemistry and supervised by a thesis committee consisting of Professors Tadgh P. Begely of the Department of Chemistry, Professor Kevin Burgess of the Department of Chemistry, and Professor David Russell of the Department of Chemistry and Department of Biochemistry and Biophysics. All work for the dissertation was completed independently by the student.

The organic synthesis of AcdK in Chapter 3, dPPMT-Ac, dPPMT-Su, and dPPMT-NB-Su in Chapter 4 were originally conducted by Dr. Yadagiri Kurra of the Department of Chemistry. The selection of AcdK in Chapter 3, AznL in Chapter 4 were carried out by Dr. Yu Zeng of the Department of Chemistry. The MS/MS in Chapter 3, 4, and part of 5 were carried out by Dr. Xin Wang of the Department of Plant Pathology and Microbiology, while the rest in Chapter 5 was conducted by Man Li. The LSD1 expression in Chapter 3, the Tip60 expression, and the hp53 study were carried out by Erol Vatansever, Willie Hsu, and Jeffery Tharp, respectively. All other work conducted for the dissertation was completed by the student independently.

The financial support for this work was supported by the National Institutes of Health (Grants R01CA161158 and R01GM121584) and the Welch Foundation (Grant A-1715). My personal graduate study was partially supported by a teaching fellowship from Texas A&M University, Organic lab.

## NOMENCLATURE

AA	Amino acid
Ac	Acetylation
ABC	Ammonium bicarbonate
AcdK	<i>N</i> <sup>ε</sup> -(4-azidobenzoxycarbonyl)- $\delta$ , $\epsilon$ -dehydrolysine
Alk	Allysine
Amp	Ampicillin
AznL	Azidonorlucine
Bu	Butylation
C	Cysteine
CloK	<i>N</i> <sup>ε</sup> -(8-chlorooctanyl)lysine
Cm	Chloramphenicol
CoA	Coenzyme A
Cro	Crotonylation
DMSO	Dimethyl sulfoxide
DUB	Deubiquitinase
E. coli	Escherichia coli
EPL	Expressed Protein Ligation
EtOAc	Ethanyl acetate
ESI-MS	Electrospray ionization mass spectrometry
FAD	Flavin adenine dinucleotide



FPLC	Fast protein liquid chromatography
HMTs	Histone methyltransferases
HRP	Horseradish peroxidase
K	Lysine
Kac	Lys acetylation
KAHA	$\alpha$ -Ketoacid–hydroxylamine peptide ligation
KAT	Potassium acyltrifluoroborate
KAT	Lysine acetyltransferase
Kan	Kanamycin
Kcr	Lys crotonylation or crotonyllysine
KDAC	Lysine deacetylase
Kmal	Lysine malenylation
Kgl	Lysine glutarylation
Kme1	Lysine monomethylation or mono-methyllysine
Kme2	Lysine dimethylation or di-methyllysine
Kme3	Lysine trimethylation or tri-methyllysine
Kpr	Lysine propionylation or propionyllysine
Ksu	Lysine succinylation or succinyllysine
Lys	Lysine
MeOH	Methanol
MS/MS	Tandem mass spectrum technique
MW	Molecular weight

ncAA	Noncanonical Amino Acid
NCL	Native Chemical Ligation
o/n	Overnight
POI	Protein of Interests
PTM	Post-Translational Modification
Pr	Propionylation
PylRS	Pyrrolysyl-tRNA synthetase
r.t.	Room temperature
sfGFP	Superfolder green fluorescent protein
Sirt	Sirtuin deacetylase
TCEP	Tris-(2-carboxyethyl)phosphine
Tet	Tetracycline
Ub	Ubiquitin or Lys ubiquitination
ULP	Ubiquitin-like proteins
WT	Wild-type

## TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
DEDICATION .....	iv
ACKNOWLEDGEMENTS .....	v
CONTRIBUTORS AND FUNDING SOURCES.....	vii
NOMENCLATURE .....	viii
TABLE OF CONTENTS.....	xi
LIST OF FIGURES.....	xvi
LIST OF TABLES.....	xxii
CHAPTER I INTRODUCTION .....	1
1.1 Extensions of proteinogenic amino acids .....	1
1.2 PTM and Epigenetics.....	2
1.3 Traditional Lys PTMs.....	3
1.3.1 Lys methylation .....	3
1.3.2 Lys Acetylation.....	4
1.4 Novel Lys PTMs.....	6
1.4.1 Lys malonylation, succinylation, and glutarylation.....	7
1.4.2 Lys ubiquitination and Ub-like protein modifications.....	9
1.4.3 Other Acylation Modifications.....	10
1.5 Case studies.....	10
1.5.1 Histone Lys PTM and cross-talk .....	10
1.5.2 Protein Lys ubiquitination.....	12
1.5.3 Protein Lys lipidation.....	13
1.6 Biological Trials for the Preparation of Proteins with Site-specific Lys PTMs	15
1.6.1 The enzymochemical synthesis .....	15
1.6.2 The direct incorporation.....	16
CHAPTER II PROTEIN CHEMICAL BIOLOGY TECHNIQUES FOR LYSINE PTMS.....	18
2.1 Protein Total/semi-synthesis.....	18
2.1.1 Peptide Synthesis and PTM Analogs.....	18
2.1.2 Peptide ligation strategies and the application on Lys PTMs .....	19

2.1.3 The shortage of protein total and semi-synthesis .....	21
2.2 Bioorthogonal Reactions and Protein Labeling .....	22
2.2.1 The definition and application of bioorthogonal reactions .....	22
2.2.2 To install Lys PTM mimics via bioorthogonal reactions.....	23
2.3 NcAA Incorporation Technique .....	26
2.3.1 Introduction to ncAA technique .....	26
2.3.2 The direct incorporation of Lys PTM into proteins.....	28
2.3.3 The combination of ncAA with bioorthogonal reaction approach .....	29

## CHAPTER III THE SYNTHETIC INVESTIGATION OF LYSINE

DIMETHYLATION.....	31
3.1 Research background .....	31
3.1.1 The chemical installation of Kme1 .....	31
3.1.2 The chemical installation of Kme2.....	32
3.2 Methodology Design of AcdK.....	33
3.3 The Synthesis of AcdK.....	35
3.3.1 General description .....	35
3.3.2 Detailed procedure.....	36
3.4 The Construction of a PylRS mutant library .....	40
3.5 The Selection, Screening and Incorporation of AcdK .....	41
3.5.1 The positive selection of AcdK .....	42
3.5.2 The negative selection of AcdK .....	42
3.6 The expression, purification of sfGFP-D134AcdK and its conversion to AIK .....	44
3.6.1 Sequence information .....	44
3.6.2 The expression of sfGFP-D134AcdK.....	45
3.6.3 The characterization and labeling of sfGFP-D134AcdK .....	46
3.6.4 The conversion to sfGFP-D134AIK .....	48
3.7 Methodology investigation on sfGFP .....	48
3.8 The preparation and application of histone with site-specific Kme2.....	51
3.8.1 Background and general design.....	51
3.8.2 The expression of H3 with AcdK incorporated.....	51
3.8.3 The conversion of AcdK to Kme2 in H3 .....	53
3.8.4 Substrate preparation for LSD1 assay.....	54
3.9 The LSD1 assay over H3K4me2, H3K9me2 and H3K36me2 .....	55
3.9.1 The expression of LSD1.....	55
3.9.2 The setting up of LSD1 enzymatic assay .....	56
3.10 The expression of p53 and preparation of methylated p53 .....	57
3.10.1 Background and general design.....	57
3.10.2 The expression of hp53-K372AcdK .....	58
3.10.3 The conversion of hp53-K372AcdK to Kme2 .....	59
3.11 The expression of Tip60 and enzymatic assay .....	60
3.11.1 The expression of Tip60.....	60
3.11.2 The setting up of Tip60 enzymatic assay .....	61

3.12 Summary.....	62
CHAPTER IV THE SYNTHETIC INVESTIGATION OF LYSINE ACYLATION ....	64
4.1 Research background .....	64
4.2 The design and synthesis of AznL .....	65
4.2.1 The design of AznL .....	65
4.2.2 The synthesis of AznL .....	67
4.3 The selection and screening of AznL.....	70
4.3.1 The selection of AznL.....	70
4.3.2 The selection result .....	71
4.4 The expression and purification of sfGFP-D134AznL .....	72
4.4.1 The incorporation of AznL.....	72
4.4.2 The fluorescence labeling of sfGFP-D134AznL .....	74
4.5 The expression and purification of Ub-K48AznL, Ub-K48ac, and WT-Ub....	75
4.5.1 The DNA and protein sequence of Ub.....	75
4.5.2 The expression of Ub-K48AznL .....	75
4.5.3 The expression of WT-Ub and Ub-K48AcK.....	77
4.6 The synthesis of (diphenylphosphino)methanethiol thioesters.....	77
4.6.1 The synthesis of (diphenylphosphino)methanethiol acetate (dPPMT-Ac) .....	77
4.6.2 The synthesis of (diphenylphosphino)methanethiol succinate (dPPMT-Su).....	80
4.7 Methodology investigation on Ub-K48AznL.....	85
4.7.1 Reaction setting up of dPPMT-Ac.....	85
4.7.2 Reaction characterization .....	86
4.7.3 Result analysis .....	88
4.8 Methodology examination and optimization for Lys succinylation .....	90
4.8.1 Reaction test of dPPMT-Su .....	90
4.8.2 Result analysis .....	91
4.8.3 The synthesis of photocaged (diphenylphosphino)thiol succinate .....	92
4.8.4 Reaction test of photocaged succinyl diphenylphosphinothioester .....	94
4.9 The expression of histone H3-K4AznL, WT-H3, and H4-His6-SUMO .....	96
4.9.1 The DNA and protein sequence of histone H3.....	96
4.9.2 The purification of H3K4AznL .....	97
4.10 General process of time-dependent traceless-Staudinger ligation reaction of H3K4AznL .....	99
4.10.1 The setting up of time-dependent Succinylation assay .....	99
4.11 The assembly of H3K4su-H4 tetramer.....	101
4.11.1 The preparation of H3K4su .....	101
4.11.2 The assembly of H3K4su-H4 tetramer .....	102
4.12 Sirt5 enzymatic assay .....	103
4.12.1 The expression of Sirt1 and Sirt5 .....	103
4.12.2 The setting up of Sirt1 and Sirt5 enzymatic assay.....	104

4.13 Summary.....	105
<b>CHAPTER V THE SYNTHETIC INVESTIGATION OF LYSINE ACYLATION VIA KAHA REACTION.....</b>	
	110
5.1 Background.....	110
5.2 The Design of acylation from AcdK.....	111
5.3 The optimization of AcdKRS plasmid.....	113
5.3.1 The cloning of opt-AcdKRS plasmid.....	113
5.3.2 The expression of UbK48AcdK.....	114
5.4 Methodology investigation of the reductive amination with hydroxylamine.....	115
5.5 Methodology investigation of the KAHA ligation reaction with $\alpha$ - ketoglutarate.....	116
5.5.1 The reaction setting up for Lys succinylation.....	116
5.5.2 The reaction working up and characterization.....	117
5.5.3 Analysis.....	118
5.6 Methodology investigation of the KAHA ligation reaction with oxoadipic acid.....	119
5.6.1 The reaction setting up for Lys glutarylation.....	119
5.6.2 The reaction working up and characterization.....	119
5.7 KAHA ligation reaction on H3K9.....	120
5.8 Sirt5 assay over H3K9Ksu and H3K9Kgl.....	123
5.9 Other Lys acylation strategies investigation based on AcdK.....	124
5.9.1 The succinylation analogs.....	124
5.9.2 The KAT reaction strategy.....	127
5.9.3 The N-O band cleavage strategy.....	130
5.9.4 The auxiliary group-assisting strategy.....	132
5.9.5 The Ub cleavage with TCEP.....	133
5.10 Summary.....	135
<b>CHAPTER VI THE CHEMICAL LABELING AND INVESTIGATION OF LYSINE LIPIDATION.....</b>	
	137
6.1 Background.....	137
6.2 Previous research and disadvantages.....	140
6.2.1 Surrogates for acetylation.....	141
6.2.2 Surrogates for medium and long chain fatty acylation.....	142
6.3 Our design and potential advantages.....	144
6.3.1 Project initiation.....	144
6.3.2 Advantage analysis.....	144
6.3.3 Introduction to Halo-tag and our strategy.....	146
6.4 The synthesis of chlorofatty <sub>n</sub> -acids (n = 4, 7, 10, 12) and FITC- Lys(chloroacyl)-OMe.....	147
6.4.1 The synthesis of chlorofatty acids.....	147

6.4.2	The synthesis of Boc-Lys-OMe intermediate.....	154
6.4.3	The synthesis of FITC-Lys(chloroacyl)-OMe.....	155
6.5	Kinetics study of Halo-tag reaction.....	157
6.5.1	The expression of Halo protein.....	157
6.5.2	Gel-based kinetics study.....	158
6.6	Test on proteins.....	160
6.6.1	The synthesis of H-Lys(8-chlorooctanyl)-OH (42).....	160
6.6.2	The incorporation of H-Lys(8-chlorooctanyl)-OH (13).....	163
6.6.3	The binding test of HaloTag protein.....	164
6.7	HaloTag protein labeling.....	165
6.7.1	The labeling of HaloTag protein.....	166
6.7.2	The blotting examination of HaloTag protein on membrane.....	167
6.8	Metabolic labeling on HEK293T cells.....	168
6.8.1	Cell culture of HEK293T cell.....	168
6.8.2	Cell culture with chlorofatty acids.....	169
6.8.3	Cell lysis and sample preparation.....	170
6.8.4	Gel analysis.....	171
6.9	Metabolic labeling on HEK293T cells.....	172
6.9.1	Background design.....	172
6.9.2	Cys myristoylation.....	173
6.9.3	Lys myristoylation.....	174
6.10	Summary.....	175
CHAPTER VII CONCLUSION AND OUTLOOK.....		178
7.1	Conclusion.....	178
7.2	Outlook.....	179
REFERENCES.....		181

## LIST OF FIGURES

	Page
Figure 1 Lys methylation and chemical biology approaches for the site-selective Lys methylation in proteins .....	4
Figure 2 Various Lys PTMs structures including acylations and methylations.....	7
Figure 3 Scheme for the convergent synthesis of H2B K34Ub .....	21
Figure 4 Cys selective modification for Lys alkylation or acylation analog installation (A) Cys alkylation for Lys alkylation analog installation; (B) Cys thiol-ene reaction for Lys acylation analog installation .....	24
Figure 5 The generation of Dha for Lys PTM analog installation. (A) the Dha generation from Cys; (B) the indirect installation of Dha with selenium containing ncAA; (C) the conversion of Dha to Lys alkylation mimics via thiol-ene reaction; (D) the conversion of Dha to Lys PTM and analogs.....	25
Figure 6 The selection, screening and expression of ncAA. (A) a “double sieved” selection process; (B) the screening process; (C) the expression of GFP with ncAA .....	28
Figure 7 Examples of directly incorporation of PTMs or bioorthogonal reactions on ncAA-containing proteins (A) the direct incorporation of AcK; (B) the installation of Kme1 with the incorporation of Boc protected precursor followed by the deprotection process. ....	29
Figure 8 The combination of ncAA and bioorthogonal reactions for the site-specific incorporation of Kme2 with a global protection-deprotection based approach.....	32
Figure 9 Chemical biology approaches for the site-selective Lysine dimethylation in proteins. (A) the general design of AcdK; (B) the structure of AcdK; (C) the conversion from AcdK to Kme2. ....	35
Figure 10 Scheme for the synthesis of AcdK.....	36
Figure 11 <sup>1</sup> HNMR and <sup>13</sup> CNMR of AcdK.....	40
Figure 12 Screening of the clones after the third positive selection on LB plates. ....	44
Figure 13 The expression of sfGFP-D134AcdK and its conversion to sfGFP- D134Aik (A) SDS-PAGE of sfGFP-D134AcdK; (B) ESI-MS of sfGFP-	



D134AcdK; (C) fluorescence labeling of sfGFP-D134AcdK showing the partial deprotection in cell; (D) the conversion of sfGFP-D134AcdK.....	47
Figure 14 Methodology investigation of reductive amination on sfGFP-D134AIK (A) SDS-PAGE and Western blot of sfGFP-D134AIK after the reductive amination reaction with dimethylamine; (B) ESI-MS of the same product after reductive amination reaction; (C) MS/MS analysis the same product. ...	50
Figure 15 The conversion of AcdK to Kme2 in histone H3 (A) SDS-PAGE and Western blot for the expression of three H3 variants with AcdK installed; (B) Western blot for the AcdK containing H3 after reductive amination with dimethylamine; (C) Western blot of H3K4me2.....	54
Figure 16 The LSD1 enzymatic assay over histone with Kme2 installed. (A) SDS-PAGE of LSD1 expression; (B) LSD1 assay for all three H3 variants.....	56
Figure 17 The expression of hp53 with AcdK incorporated. (A) the plasmid for the expression of hp53; (B) SDS-PAGE for the GST-p53-sfGFP expression. ....	59
Figure 18 The expression of Tip60 and its enzymatic assay. (A) SDS-PAGE for the expression of GST-Tip60; (B) Tip60 assay for the cross-talk between p53-K372 methylation and acetylation; (C) quantification of cross-talk between p53-K372 methylation and acetylation.....	61
Figure 19 The conversion of AznL to Lys acylation via traceless Staudinger ligation...	66
Figure 20 Scheme for the synthesis of AznL.....	67
Figure 21 <sup>1</sup> HNMR spectrum of AznL .....	70
Figure 22 Screening of the clones after the third positive selection on LB plates. The expression of GFP can be seen on the plate with either 1 mM or 5 mM AznL but not the one without.....	71
Figure 23 The expression of sfGFP-D134AznL (A) Coomassie staining of SDS-PAGE analysis for sfGFP-D134AznL; (B) ESI-MS of sfGFP-D134AznL; (C) fluorescence labeling of sfGFP-D134AznL with azido-dye.....	74
Figure 24 Scheme for the synthesis of dPPMT-Ac.....	77
Figure 25 Scheme for the synthesis of dPPMT-Su .....	80
Figure 26 Methodology examination on UbK48AznL (A) the structure of dPPMT-Ac; (B) the gel-based time dependent assay for Lys acetylation; (C) MALDI-TOF of UbK48AznL after traceless Staudinger acetylation reaction for 48	

h; (D) MS/MS analysis of UbK48AznL after traceless Staudinger acetylation reaction for 48 h.....	88
Figure 27 Mechanism of traceless Staudinger ligation reaction on proteins.....	89
Figure 28 Methodology examination and optimization of Lys succinylation (A) the structure of dPPMT-Su; (B) the gel-based time dependent assay for Lys succinylation; (C) the potential side reaction of non-specific Lys succinylation; (D) the structure of dPPMT-NB-Su.....	91
Figure 29 Scheme for the synthesis of dPPMT-NB-Su.....	92
Figure 30 Methodology examination of Lys succinylation via dPPMT-NB-Su. (A) the structure of dPPMT-NB-Su; (B) the gel-based time dependent assay for Lys succinylation; (C) MALDI-TOF of UbK48AznL after traceless Staudinger succinylation reaction for 48 h; (D) MS/MS analysis of UbK48AznL after traceless Staudinger succinylation reaction for 48 h. ....	96
Figure 31 Preparation of histone with Lys succinylation via dPPMT-NB-Su (A) the gel-based time dependent assay for succinylation of H3; (B) MS/MS analysis of H3 succinylation product.....	101
Figure 32 Preparation of H3K4su-H4 tetramer (A) scheme of the assembly of H4-H3K4su tetramer; (B) SDS-PAGE analysis of H4-H3K4su tetramer. ....	103
Figure 33 The setting up of Sirt1 and Sirt5 enzymatic assay of H3K4su-H4 tetramer. ....	105
Figure 34 <sup>1</sup> HNMR, <sup>13</sup> CNMR and <sup>31</sup> PNMR of dPPMT-Ac.....	107
Figure 35 <sup>1</sup> HNMR, <sup>13</sup> CNMR and <sup>31</sup> PNMR of dPPMT-Su.....	108
Figure 36 <sup>1</sup> HNMR, <sup>13</sup> CNMR and <sup>31</sup> PNMR of dPPMT-NB-Su.....	109
Figure 37 The combination of AcdK incorporation and KAHA ligation reaction for the installation of Lys acylation.....	113
Figure 38 The optimization of AcdKRS by the comparison of sfGFP expression with different optimized AcdKRS mutants (A) the GFP expression with optimized AcdKRS; (B) SDS-PAGE of the expression of Ub-K48AcdK. ....	114
Figure 39 The reductive amination of hydroxylamine on UbK48AIK (A) MALDI-TOF-MS spectrum of product Ub-K48K <sub>OH</sub> ; (B) MS/MS analysis of 48K <sub>OH</sub> -containing trypsinized peptide fragments. ....	116
Figure 40 Methodology examination of Lys succinylation (A) after reacting with $\alpha$ -ketoglutarate, the protein product Ub-K48Ksu was analyzed by SDS-PAGE	

and the Western blot probed by a pan-anti-succinyllysine antibody; (B) MALDI-TOF-MS spectrum of the KAHA reaction product as Ub-K48Ksu; (C) MS/MS analysis of the trypsinized 48Ksu-containing peptide fragments. ....	118
Figure 41 Methodology examination of Lys glutarylation (A) After reaction with oxoadipic acid, the protein product Ub-K48Kgl was analyzed by SDS-PAGE and the Western blot probed by a pan-anti-glutaryllysine antibody; (B) MALDI-TOF-MS spectrum of the KAHA reaction product as Ub-K48Kgl; (C) MS/MS analysis of the trypsinized 48Kgl-containing trypsinized peptide fragments. ....	121
Figure 42 After reaction with oxoadipic acid, the protein product Ub-K48Kgl was analyzed by SDS-PAGE and the Western blot probed by a pan-anti-glutaryllysine antibody (A) After reaction with $\alpha$ -ketoglutarate, the protein product H3K9Ksu was analyzed by the Western blot probed by a pan-anti-succinyllysine antibody and an anti-H3 antibody; (B) After reaction with oxoadipic acid, the protein product H3K9Kgl was analyzed by the Western blot probed by a pan-anti-glutaryllysine antibody and an anti-H3 antibody. ....	123
Figure 43 Sirt5 enzymatic assay over H3-H4 tetramer (A) Desuccinylation of the H3K9Ksu by incubating H3K9Ksu with or without 0.5 $\mu$ M Sirt5 in the presence of 1 mM NAD <sup>+</sup> and 1 mM dithiothreitol (DTT) for 4 h. After reaction, proteins were analyzed with western blot analysis by a pan-anti-succinyllysine antibody and anti-H3 antibody; (B) Desuccinylation of the H3K9Kgl under the same condition. After reaction, proteins were analyzed with western blot analysis by a pan-anti-Kgl antibody and anti-H3 antibody.....	124
Figure 44 Scheme for the installation of acidic PTM analogs.....	125
Figure 45 <sup>1</sup> HNMR and <sup>13</sup> CNMR of 2-(aminoxy)acetate .....	126
Figure 46 Reductive amination investigation of acidic PTM analog (A) the MALDI-TOF of the product after reductive amination with 2-(aminoxy)acetate; (B) Western blot analysis of the product after reductive amination with 2-(aminoxy)acetate. ....	127
Figure 47 Scheme for the Lys acylation installation with KAT reaction.....	128
Figure 48 <sup>1</sup> HNMR and <sup>13</sup> CNMR of N-((aminoxy)carbonyl)-N-ethylethanamine .....	129
Figure 49 Scheme of the Lys acylation installation with N-O cleavage strategy .....	130
Figure 50 <sup>1</sup> HNMR and <sup>13</sup> CNMR of 2-(aminoxy)ethanethiol.....	131

Figure 51 Scheme for the Lys acylation installation with the auxiliary group-assisting strategy.....	132
Figure 52 Methodology investigation of the Lys acylation installation with the auxiliary group-assisting strategy (A) MALDI-TOF of reductive amination; (B) Western blot for the acetylation and propionylation; (C) MS/MS of acetylation product. ....	133
Figure 53 Scheme for Ub with Cys mutations cleaved by TCEP .....	134
Figure 54 MALDI-TOF of Ub with Cys mutations cleaved by TCEP (A) UbK48C with His <sub>6</sub> tag as MW 10044 Da before reaction (...FAG-C-QLE...); (B) UbK48C with His <sub>6</sub> tag was cleaved by 20 mM TCEP to generate fragment 1 4716 Da (QLE...), and fragment 2 5244 Da (...FAG), and the Cys was eliminated; (C) UbK11C without His <sub>6</sub> tag as MW 8542 Da before reaction (...LTG-C-TIT...); (D) UbK11C without His <sub>6</sub> tag was partially cleaved by 20 mM TCEP to generate fragment 3 MW 7319 Da (TIT...) and a cleavage at T site to generate fragment 4 MW 7218 Da (IT...), while the C-terminal fragment was too small to be detected.....	135
Figure 55 Protein fatty acylation and fatty acid metabolism (A) examples of protein lipidation; (B) $\beta$ -oxidation pathway for fatty acid metabolism.....	139
Figure 56 Surrogates of fatty acids and their corresponding detection strategies (A) terminal olefin group labeled by tetrazine; (B) terminal azido group labeled by Click reaction or Staudinger ligation; (C) the basic mechanism of dehalogenase enzyme; (D) design of terminal chloride labeled by HaloTag protein. ....	143
Figure 57 Scheme for the synthesis of 8-bromooctanoic acid .....	147
Figure 58 <sup>1</sup> HNMR and <sup>13</sup> CNMR of 8-bromooctanoic acid .....	148
Figure 59 Scheme for the synthesis of 11-chloroundecanoic acid.....	149
Figure 60 <sup>1</sup> HNMR and <sup>13</sup> CNMR of 11-chloroundecanoic acid.....	150
Figure 61 Scheme for the synthesis of 13-chlorotridecanoic acid .....	151
Figure 62 <sup>1</sup> HNMR and <sup>13</sup> CNMR of 13-chlorotridecanoic acid .....	153
Figure 63 Scheme for the synthesis of Boc-Lys-OMe intermediate.....	154
Figure 64 Scheme for the synthesis of Boc-Lys-OMe intermediate.....	155
Figure 65 The purification of GST-HaloTag protein .....	158

Figure 66 Gel-based kinetics study of Lys conjugated dye reacting with HaloTag protein .....	159
Figure 67 Scheme for the synthesis of H-Lys(8-chlorooctanyl)-OH .....	160
Figure 68 <sup>1</sup> HNMR and <sup>13</sup> CNMR of H-Lys(8-chlorooctanyl)-OH .....	162
Figure 69 The expression of sfGFP-D134CloK.....	164
Figure 70 The binding test of HaloTag protein with sfGFP-D134CloK.....	165
Figure 71 Screen for the metabolic labeling with biotin conjugated HaloTag protein .	166
Figure 72 The biotin labeling of HaloTag protein and its gel characterization (A) the insertion of TEV protease site at the C-terminal of HaloTag; (B) the TEV cleavage followed by the CBT-dye labeling of HaloTag; (C) the SDS-PAGE and Western blot of labeled HaloTag by Streptavidin-HRP.....	167
Figure 73 The membrane blotting with Biotin conjugated HaloTag protein.....	168
Figure 74 The metabolic labeling of HEK293T cell lysate via biotin labeled HaloTag protein (A) SDS-PAGE and Western blot of cells fed with $\omega$ -chlorofatty-acid n = 7; (B) SDS-PAGE and Western blot of cells fed with $\omega$ -chlorofatty-acid n = 10; (C) SDS-PAGE and Western blot of cells fed with $\omega$ -chlorofatty-acid n = 12.....	172
Figure 75 The metabolic labeling of HEK293T cell lysate under different conditions via biotin labeled HaloTag protein (A) SDS-PAGE and Western blot of cells with hydroxylamine treatment to remove Cys myristoylation; (B) SDS-PAGE and Western blot of cells fed with nicotinamide to inhibit Sirtuin enzymes.....	174
Figure 76 <sup>1</sup> HNMR of FITC-Lys(5-chloropentanyl)-OMe.....	176
Figure 77 <sup>1</sup> HNMR of FITC-Lys(8-chlorooctanyl)-OMe.....	176
Figure 78 <sup>1</sup> HNMR of FITC-Lys(11-chloroundecanyl)-OMe .....	177
Figure 79 <sup>1</sup> HNMR of FITC-Lys(13-chlorotridecanyl)-OMe.....	177

## LIST OF TABLES

	Page
Table 1 The mutation sites for the selected PyIRS candidates.....	72
Table 2 The comparison of fatty acids with different terminal functional groups.....	145

# CHAPTER I

## INTRODUCTION

### **1.1 Extensions of proteinogenic amino acids**

The Mother Nature chooses 20 proteinogenic amino acids (AAs) for almost the entire life systems, encoded by the combination of three base pair as the information transduction of “Centre dogma”<sup>[1]</sup>. Although large amount of protein sequences and structures can be made with the combination of these 20 AAs, challenges still exist due to the limited chemical diversity when facing the various earth environments<sup>[2]</sup>. The selenocysteine and pyrrolysine as the 21<sup>st</sup> and 22<sup>nd</sup> amino acids are also installed in native proteins. But due to its limited application in rare species, extensions of AA side chains are still needed. Coenzymes and cofactors have been one of the decent solutions to provide functional groups not exist among the 22 proteinogenic AAs to facilitate many enzymatic processes. However, they are mainly linked to enzyme active sites which can hardly regulate the entire protein.

Similar to the mRNA maturation process, some of the proteins are also undergoing editing process after been expressed by the ribosome. A typical example is the maturation of insulin, where the original proinsulin was an entire peptide that will be cleaved into three fragments. The fragment A and B was conjugated via disulfide bonds after the removal of C-peptide<sup>[3]</sup>. Another example of protein editing is known as “Protein splicing”, in which the removal of intein followed by the conjugating of C-terminus to another component, mainly peptide but also other small molecules<sup>[4]</sup>. However, these approaches

can only change the substrate protein between two states and are generally irreversible, which makes it not suitable in response to the frequently changing environments.

To meet the need of more robust and widely applicable protein covalent regulations, post-translational modifications (PTMs) on the protein amino acid residues have been gradually discovered. These modifications with various structures cannot only regulate the physical and chemical properties of the original proteins<sup>[5]</sup>, but also provide diverse chemical handles for protein-protein interactions (PPI) or protein-small molecule interactions<sup>[6]</sup>. PTM on proteins, together with DNA modifications and RNA interference, is becoming three key subfields of epigenetics. Although the study of PTMs has become a cutting edge of chemical biology field since its infancy, the chemical biology techniques are still one of the bottlenecks for further functional studies.

## **1.2 PTM and Epigenetics**

Cellular proteins are undergoing post-translational modifications (PTMs) including phosphorylation on Ser, Thr, Tyr and His, methylation on Lys and Arg, glycosylation on Ser and Asn, lipidation on Lys and Cys, etc. Although these PTMs are generally related to many important cellular activities, detailed molecular mechanisms are remaining unclear<sup>[7]</sup>. Histone proteins, for instance, have various types of PTMs at their N-terminal tails, which is closely regulating gene expression and chromatin structure<sup>[8]</sup>. Especially with the existence of sophisticated cross-talk regulation between different PTMs, it is a prerequisite to get sufficient amount of proteins with site-specific PTMs for further biological studies<sup>[9]</sup>.



In the post-genomic era, a growing number of different protein PTMs are gradually being discovered on both cytoplasmic and nucleic proteins with the help of tandem mass spectrum technique (MS/MS) and proteomic techniques<sup>[10]</sup>. Three groups of proteins or enzymes are closely related to them, namely “Writer”, “Reader”, and “Eraser”, which is catalyzing the reactions of installing, recognition and removing of different PTMs. Known as “epigenetic markers”, these PTMs are generally known as critical to the gene transcription and metabolic regulation<sup>[11]</sup>. Especially on histones, these so called “Histone code” are attracting more and more research focus, but the detailed molecular mechanisms as well as their signaling pathways remain elusive<sup>[12]</sup>.

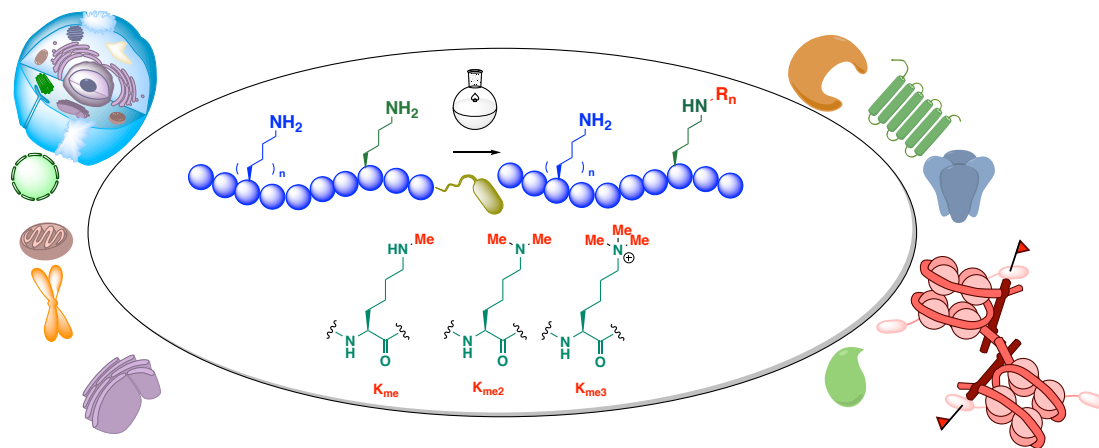
### **1.3 Traditional Lys PTMs**

Lysine (Lys, K), as the only free  $\epsilon$ -amino group containing amino acid among 20 proteinogenic amino acids, can undergo many different types of acylations and alkylations. Some traditional ones are listed and summarized here.

#### **1.3.1 Lys methylation**

Lys methylations widely exist on histone and non-histone proteins, which can be divided into three levels, namely mono-methylation (Kme), di-methylation (Kme2), and trimethylation (Kme3)<sup>[13]</sup>. As one of the main groups of Lys modifications, histone methylation influences the structure and functions of eukaryotic chromatin, and recruits various binding proteins for transcription regulation<sup>[14]</sup>. In addition, Lys methylation can also regulate the functions of important transcription factors such as p53 and NF- $\kappa$ B<sup>[15]</sup>.

Therefore, a primary prerequisite for the further biological investigations is to get protein of interests (POIs) with sufficient quantity in their homogeneous PTM states (Figure 1) [16].



**Figure 1 Lys methylation and chemical biology approaches for the site-selective Lys methylation in proteins**

### 1.3.2 Lys Acetylation

As one of the main groups of protein PTMs, Lys acetylation (Kac) have been known and studied for decades, which are closely related to many important cellular activities. Especially for the PTMs on the histone N-terminal tails, also known as “histone code”, can regulate chromatin structures and gene transcription functions. As a conserved protein PTM mode, Lys acetylation links acetyl-coenzyme A from metabolism with genetic information delivering, as well as the cellular signaling pathway.

From a biochemistry point of view, Lys acetylation can mask the positive charge on Lys under physiological condition. This charge removal fact can change the basic physical properties of protein substrates, and significantly impact their interactions with

other micro-molecular binders. Especially for the nucleosome, the charge-charge interaction between highly positively charged histone proteins with DNA is of great importance, which can be interrupted by Lys acetylation.

For PTM studies, one of the key points for our investigation is their regulation, including the installation, the recognition or binding, and the removal. All these three processes are enzymatic procedure, which are catalyzed by three groups of very specialized enzymes. Marked as “Writer, reader, and eraser”, the three enzyme families corresponding to acetylation are also known as Lys acetyltransferases (KATs) to add acetyl groups to proteins, acetyllysine binders to selectively interact with acetylated proteins, and Lys deacetylases (KDACs) to remove acetyl groups from proteins<sup>[17]</sup>. Till now, KATs identified in human and mouse cells can be divided into three major groups as GCN5, CBP/p300 and MYST11, which need acetyl-coenzyme A (acetyl-CoA) as the only essential cofactor to install an Ac group to POIs. Bromodomain has been reported as a AcK binder existing in about 46 proteins in the human genome. Over 50 % have been shown to bind with AcK, especially with the histone N-terminal tail.

On the other hand, 18 KDACs have been identified in the human genomes. Since they mainly catalyze the deacetylation over histones, they are also called as histone deacetylases (HDACs)<sup>[18]</sup>. All the enzymes belonging to HDAC family can be divided into two distinct families. The first group enzymes relay on an  $Zn^{2+}$  at the catalytic site (HDAC1–11), while the second group depends on  $NAD^+$ . Specially named as Sirtuin deacetylases (Sirt), there are in total 7 members (Sirt1–7)<sup>[19]</sup>, existing in the nucleus, cytoplasm and mitochondria. Although sharing similar structures and mechanisms,

different SIRT enzymes may catalyze the removal of some other types of Lys acylations, which will be discussed in Part 1.4.1.

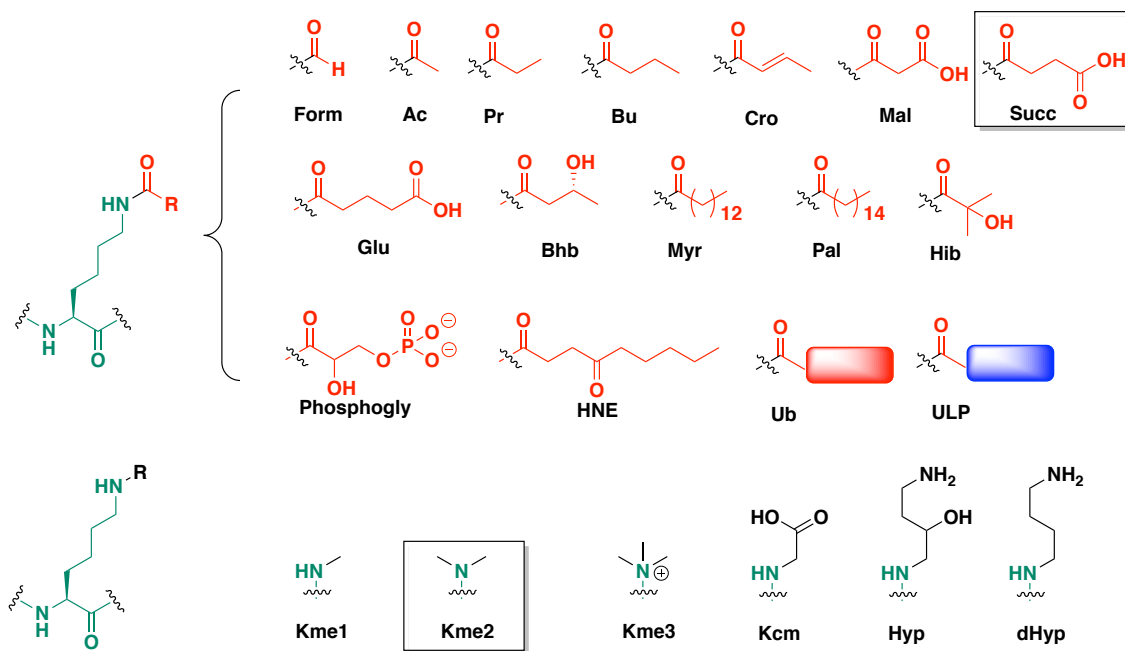
Recently MS and MS/MS advance have showed the wide existence of Lys acetylation in the human proteome. Noted as acetylome, the identification and quantification of Lys acetylation have dramatically increased our understanding of the function, regulation and localization of AcK.

#### **1.4 Novel Lys PTMs**

Besides the widely existing acylation types as Kac, Lys propionylation (Kpr), Lys butylation (Kbu), and Lys lipoylation (Klip), modern tandem mass spectrometry (MS/MS) and proteomic techniques have identified an increasing amount of other acylation types with more complicated structures and functions (Figure 2). These novel post-translational acylations also widely exist in human proteome, including histone and non-histone proteins. Two typical examples are lysine succinylation (Ksu) and glutarylation (Kgl) discovered recently. The introduction of Ksu or Kgl can reverse the positively charged state of lysine side chain into a negatively charged one. Although these modifications are believed to play important roles in transcription and metabolic regulation, detailed molecular mechanisms are still remaining unclear.

Among the novel Lys PTMs, three anion-carrying modification models are becoming more and more interesting, malonylation, succinylation, and glutarylation, which can reverse the positive charge state of the lysine  $\epsilon$ -amine and thus regulate protein structures and functions<sup>[2]</sup>. Besides, over ten different Lys acylation types have been

discovered, the study of which will advance our understanding of protein regulations and their related cell activities. Although cell biologists have showed the important roles of PTMs, the detailed molecular mechanisms are remaining unclear.



**Figure 2 Various Lys PTMs structures including acylations and methylations**

#### 1.4.1 Lys malonylation, succinylation, and glutarylation

Protein Lys acetylation is a well-studied PTM with a fine regulatory mechanism for several important cellular processes. As one of the simplest acylation structures, its chemical essence is to mask the positive charge on Lys under physiological condition. Additionally, several novel acylation PTMs cannot just mask the positive charge, but even introduce a new negative charge. This charge reversing process can obviously perform even stronger interference to the charge interaction and bring in new binding possibilities.

Three of these new Lys PTMs as malonylation, succinylation, and glutarylation have greatly expanded our imagination of the natural types of modifications on proteins. With an excess free carboxyl group, these acidic Lys modifications belong to the same family with similar structures. Thus, discovered in the last decades by the Zhao group, these modifications are drawing more and more interests recently<sup>[20]</sup>.

Although discovered recently, their existences are not low at all<sup>[21]</sup>. MS/MS based proteomic studies have shown their wide existence over metabolic proteins and histone proteins. Especially in mitochondria, Ksu is believed to be closely related to the tricarboxylic acid (TCA) cycle, which builds up the bridge between metabolism and genetic regulation.

As for the PTM enzymes of “Writer, reader, and eraser”, very little has been known for these three acidic modifications. Although they share very similar structures, their regulation and signaling pathways may be completely distinct. With a high level of succinyl-CoA in mitochondria and malonyl-CoA, for instance, it is even unclear whether the succinylation is an enzymatic process or not. Recently, GCN5 as the lysine acetyltransferase 2A together with its complex has been reported as the first potential histone H3 succinyltransferase<sup>[22]</sup>. But nothing has been reported for Kmal or Kgl.

The removal of the three acidic PTMs has been reported as the Sirtuin 5 (SIRT5) as we mentioned above. The hindrance of investigation mainly comes from a difficulty in the site-specific introduction of these novel PTMs into POIs. Therefore, we will provide a new synthetic technique for the study of Lys succinylation in Part 4. By applying our technique, we will further study their regulation by SIRT5.

#### 1.4.2 Lys ubiquitination and Ub-like protein modifications

There is a very specialized class of Lys acylation modifications as peptidonylation and proteinylation, in which a peptide or a small protein is covalently linked to the  $\epsilon$ -amino group of Lys. Protein ubiquitination, for instance, is one of the most important and widely existing PTM in eukaryotes. Discovered only in last decades, numerous studies have been carried out to investigate its critical roles in a vast number of cellular activities and diseases. Due to a large variety of different linking modes including mono-, di-, and poly-ubiquitination, the biological functions and regulations of ubiquitination are generally elusive as a special “Ubiquitin code”<sup>[23]</sup>.

Apart from ubiquitination, several other small proteins are also widely added to protein substrates. Named as “ubiquitin-like protein (UBL)”, this ubiquitin family contains about 20 eukaryotic proteins sharing  $\beta$ -grasp fold structures<sup>[24]</sup>. Typical examples are SUMO and NEDD, corresponding to SUMOylation and Neddylation. UBLs have related structures, however, they participate in very different cellular processes and are controlled by various mechanisms. The protein substrates, mechanisms, and functions of Ub and UBLs are largely unclear, which brings many research opportunities to investigate their molecular mechanisms. The synthetic study of proteins with Ub or UBL modifications is one of the main challenges for chemical biologists and bioorganic chemists<sup>[25]</sup>. More discussion about Lys ubiquitination can be found in Part 1.5.2.

### 1.4.3 Other Acylation Modifications

Beside all the Lys modifications discussed above, many other minor modification modes have also been discovered recently. Although not common, they can play significant roles in certain cellular processes. Typical examples are Lys phosphoglylation, Lys 2-hydroxyisobutylation<sup>[26]</sup>, Lys  $\beta$ -hydroxybutylation<sup>[27]</sup>, Lys 4-oxononanoylation<sup>[28]</sup>, etc.

## 1.5 Case studies

Since Lys PTMs are playing important roles in many cellular activities, several fields have drawn special efforts. In this part, we will discuss several special Lys modification cases as PTMs on the histone tails, protein Lys ubiquitination, and protein Lys lipidation.

### 1.5.1 Histone Lys PTM and cross-talk

One of the main battlefields for PTM study is the Lys modification on histone proteins, which is of primary interest for epigenetics. Here a brief introduction of histone and nucleosome is provided. Human chromatin is made up with heavily condensed DNA over many protein complex units. Noted as nucleosome, each of these units has a histone octamer wrapped by 146 bp of DNA. The octamer is composed with two copies of each histone protein as H2A, H2B, H3 and H4. These single nucleosomes are linked together by DNA in between to form a nucleosomal array, which is stabilized by histone H1. The



nucleosomal arrays are further packed into higher-order chromatin structures with the help of other non-histone proteins interacting.

Histones have a conserved sequence and a special histone fold structure with its flexible N-terminal tail. When folding as octamer, the globular domains of different histone proteins are packed together, while the N-terminal tails are all protruding outside of the nucleosome. Since the binding with other regulatory proteins such as transcription factors are mainly with its N-terminal tails, the modifications on these residues can significantly influence the protein-protein interactions<sup>[29]</sup>.

Many PTMs have been discovered heavily localized at all four histone N-terminal tails such as traditional modes as Lys acetylation, Lys methylation, Ser phosphorylation; new modes as Lys ubiquitination, Lys sumoylation, Lys succinylation, etc. As we mentioned above, histone modifications at its N-terminal tail can interrupt the histone-DNA interactions, which will further affect the higher-order chromatin structures and alter the transcription activities by recruiting binding proteins. These epigenetic effects are involved in many cellular activities such as DNA replication, DNA repair, cell cycle, cell differentiation etc. Therefore, the misregulation of histone N-terminal PTMs can lead to the generation of many human diseases. To investigate these binding proteins are still one of the main tasks for current chemical biology and enzymology fields<sup>[30]</sup>.

Another important point for the histone PTM is about the sophisticated network of cross-regulation<sup>[9]</sup>. Also named as cross-talks, One PTM can play a regulatory role in another PTM, which really makes the histone PTM as another level of genetic code. Noted as “Histone code”, one PTM is able to influence the installation of one or multiple PTMs

subsequent on the same histone or between histone molecules or even across different nucleosomes<sup>[31]</sup>. These two ways are marked as “*in cis*” or “*in trans*”, respectively. For instance, we will study the cross-talk between methylation and acetylation on human p53 in the Chapter 4. These special “histone code” are recognized and read by special binders to facilitate the installation of other PTMs or to initiate the downstream signaling pathway.

### 1.5.2 Protein Lys ubiquitination

Ubiquitination is a special modification discovered not long ago. As a small protein with 76 AAs, ubiquitination is quite distinct from other small modifications due to a large area it can provide. Till now, more than 57000 Lys sites have been found with ubiquitination, making it as the second most widely existing PTMs. Besides modifications over other protein substrates, Ub can undergo ubiquitination itself as well. With seven different Lys sites (K6, K11, K27, K29, K33, K48, and K63) and a N-terminal amino group, there are eight possible di-Ub linkages. Ubiquitination can have a large variety of different modes as mono- ubiquitination, di-ubiquitination, oligo-ubiquitination, and poly-ubiquitination. As for poly-ubiquitination, they can form long chain or branched chain structures with one Ub linked with two Ubs.

Due to many different linking modes, ubiquitination plays various roles in cells, including many physiological activities. The malfunction and misregulation of protein Lys ubiquitination can result in many disease-related processes, such as DNA damage and proteasomal degradation<sup>[32]</sup>. More importantly, poly-ubiquitination is generally believed for the target protein degradation process. Marked as “Ubiquitin code”, it has been a case-

by-case issue, and the detailed molecular mechanism is still remaining unclear so far<sup>[33]</sup>. Another important direction is to investigate the PTM enzymes related as “writer, reader, and eraser”. Many efforts have been focused on the enzymatic ubiquitination process. Also known as ubiquitination cascade, a sequential action of three enzymes is needed to add a Ub unit to the protein target. Noted as E1, E2, and E3, different numbers of them can produce an even larger selectivity combination over protein substrates<sup>[34]</sup>. The E3 ligases will finally attach an Ub unit to the target protein. All these three enzymes are tightly regulated by other binding proteins to avoid cellular pathology.

Similar to other Lys PTMs, Lys ubiquitination is also a reversible process that can be removed by a group of enzymes deubiquitinases (DUBs). Belonging to Cys hydrolases, DUBs carefully control the ubiquitination level and sites for protein substrates to maintain cells in the physiological state. Extended studies have been carried out to study the installation and removal of Ub in the past two decades<sup>[35]</sup>, however, many important questions are still remaining unclear.

### 1.5.3 Protein Lys lipidation

Beside the acidic modifications discussed above, another special modification mode will be brief talked about in this paragraph. Proteins can undergo lipidation on Lys, Cys or N-terminal amino group in general<sup>[36]</sup>. As a widely existing phenomenon, lipidation together with the lipid trafficking can control many important cellular functions. With a large amount of different lipids and their derivatives including fatty acids, phospholipids, cholesterol, etc, there is an even larger library for the choices of protein lipidation<sup>[37]</sup>.

The discovery of lipidation as one of the widely existing covalent modification to many protein substrates has revealed a complex network of cellular membrane system. Proteins with lipidated residues can play vital roles under physiology conditions<sup>[38]</sup>, and the malfunction or misregulation can also cause many human diseases<sup>[39]</sup>. Since cell membranes are made of lipids in all kingdoms of life, the lipid linked on proteins can play roles as an interaction handle for membrane or other binding proteins for the purpose of discretion out of cells. This evolves in the both metabolism and epigenetic controls, because even transcription factors can also undergo Lys lipidation, such as myristilation. Additionally, the potential changes in the lipid composition can have influence over membrane architecture, and further control the recruitment of protein binders to membranes. This process can play important roles on organismal physiology.

However, the lack of biochemical methods and chemical biology tools is the main hindrance for the study of protein lipidation. Besides Sirt6 reported as an eraser for fatty acylation<sup>[40]</sup>, little information has been reported for lipidation regulation. With a high hydrophobicity, a non-reactive and non-polar hydrocarbon chain, fatty acylation has a very poor handling capacity which is hard to develop antibody against. Due to the bad physical performance, chemically synthesizing membrane proteins become equally difficult<sup>[41]</sup>. For the purpose of lipidomic study, precise and sensitive labeling molecules and methods are of great significance especially with the existence of a complicated lipid trafficking and metabolic systems<sup>[42]</sup>. Projects related to Lys lipidation will be discussed in Chapter 5 and 6.

## 1.6 Biological Trials for the Preparation of Proteins with Site-specific Lys PTMs

Although the study of protein PTMs have been highly evaluated for long, limited achievement has been made in the last decades, because traditional biological techniques fail to achieve the goal to install PTMs into specific sites of POIs. It is almost impossible to purify target proteins with PTMs from natural source due to the wide distributions, heterogeneous states and dynamic properties, and additionally these proteins usually contain more than one modification. Recombinant protein expression is a routinely used technique for the preparation of proteins from certain heterogenetic cells. But traditional biological technique can hardly control the PTM states from the genetic level.

### 1.6.1 The enzymochemical synthesis

Enzymatic preparation is another strategy for the installation of certain PTMs, in which a specific “Writer” enzyme can catalyze the installation of a corresponding modifications into proteins substrates. Histone methyltransferases (HMTs) are a group of enzymes to catalyze the methylation of histones in biological systems, which can potentially be applied to methylate protein substrates at few identified sites in vivo or in vitro. However, HMT-catalyzed enzymatic assay may result in heterogeneous protein methylation due to its substrate promiscuity and inability to control the three-methylation levels, which is very difficult to be separated from each other.

Researchers started with the enzymatic synthesis approach, nonetheless, the enzymes for most of the novel modifications are unexplored<sup>[5]</sup>. Protein total and semi-synthesis can theoretically provide native proteins with natural or non-natural structures.

However, both NCL and EPL are time-consuming comparing with the normal protein expression with a general requirement for Cys. In addition, Cys alkylation and Dha based bioorthogonal reactions can generate proteins Lys acylation mimics. With the introduction of metal-catalyzed reactions into classic bioorthogonal reaction field, coupling reaction on Dha can generate carbon-carbon linking modifications<sup>[6]</sup>. However, the products are in general not stereochemically pure.

### 1.6.2 The direct incorporation

Another approach is to apply ncAA for the incorporation of proteins with Lys acylations. Initiated by the Chin group, many simple PTMs have been successfully incorporated mainly based on the pyrrolysine system, such as Kac, Kpr, Kbu, Kcr, and Khib<sup>[8]</sup>. Although it is a straightforward and convenient method, a PylRS mutant is necessary to recognize the particular modified Lys, which can be tedious and does not guarantee to be successful. For instance, Lys with charge-containing side chain such as succinylation, or large acylation modifications such as lipidation or Ubiquitination is not accessible for the direct incorporation strategy. For instance, all our trials for the incorporation of malonyllysine, succinyllysine, and glutaryllysine have turned out to be failure in the selection of PylRS mutants. And therefore, a more robust and versatile strategy is needed for the installation of Lys PTMs.

Meanwhile, peptides with Kme2 have been prepared for enzymatic assays<sup>[43]</sup>. Due to the lack of a robust synthetic method for the preparation of proteins with Kme2, many important biological questions remain unclear. Therefore, to develop universal and

reliable methodologies for the protein methylation is a fruitful but challenging task in front of protein chemists. So far, several protein chemistry approaches have been developed, and we will summarize and compare their advantages and disadvantages below.

## CHAPTER II

### PROTEIN CHEMICAL BIOLOGY TECHNIQUES FOR LYSINE PTMS

#### **2.1 Protein Total/semi-synthesis**

In this chapter, I would like to give a brief introduction to several key chemical biology techniques for the synthetic study of Lys PTMs that will be directly or indirectly applied to the later chapters, protein synthesis, bioorthogonal reactions, and ncAA<sup>[44]</sup>.

In the past several decades, protein total synthesis and semi-synthesis set up a new avenue for the study of macromolecules and laid the foundation of protein chemical biology. When discussing about current protein chemical biology techniques, the first key word is synthesis without any doubt. Thus, a brief discussion of protein synthesis will initiate this chapter.

##### 2.1.1 Peptide Synthesis and PTM Analogs

In the recent decades, the development of bioorganic chemistry has enabled chemical biologist a detailed control of biomacromolecules with defined structures. The development of novel protecting group, new generations of coupling reagents, and solid phase peptide synthesis (SPPS) are greatly facilitating the formation of peptides<sup>[45]</sup>. With excess amount of another amino acid moiety and coupling reagents, the amide bond formation between amino acids can be very efficient with a yield of around 99 %. Besides natural amino acids applicable for SPPS, any molecules with both a carboxyl group and an amino group can be coupled into a peptide chain. Therefore, amino acids carrying PTMs or their structures can be directly installed into target peptides, which can be



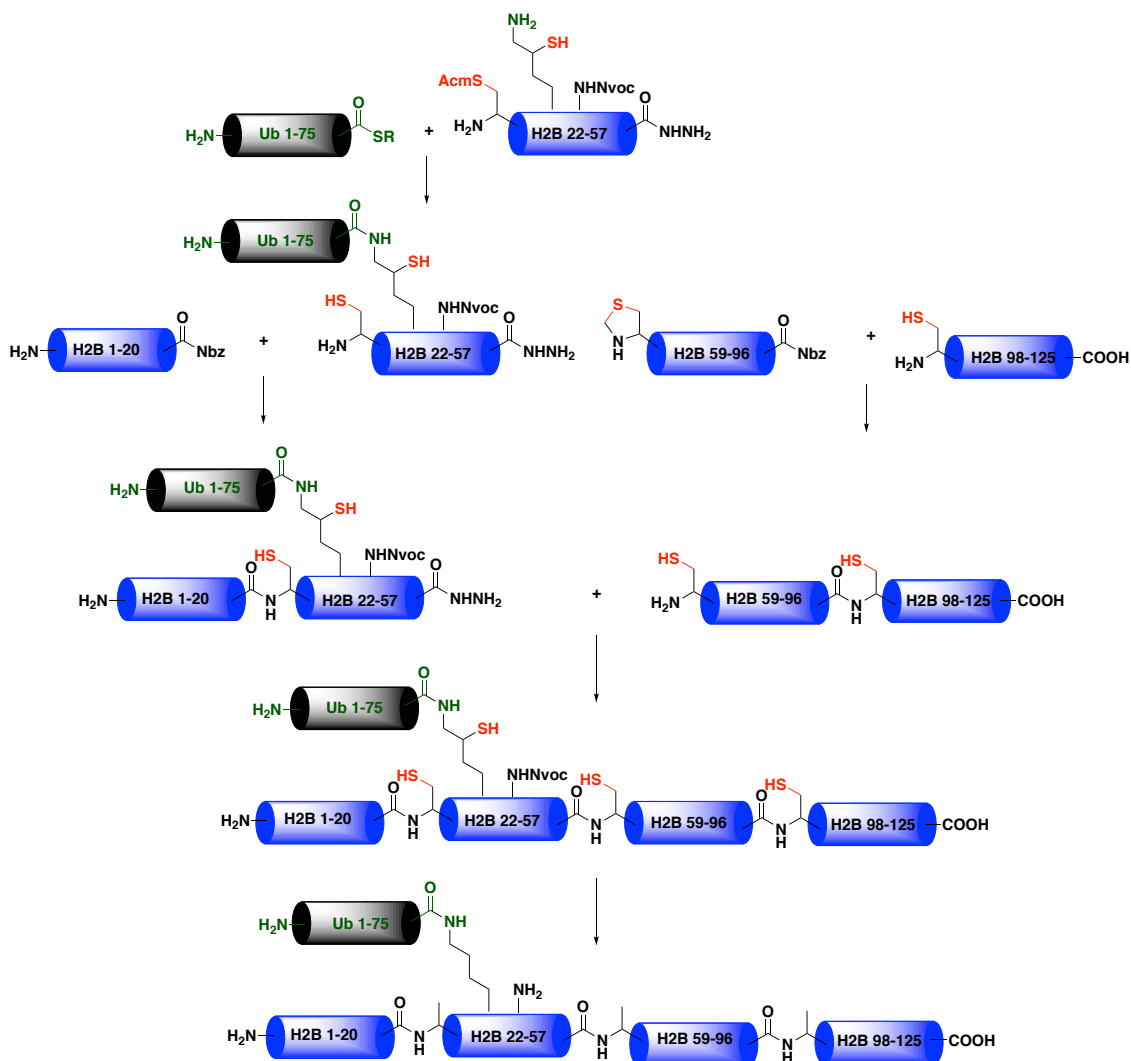
otherwise hard to synthesize or easily degraded. Phosphohistidine (pHis), for instance, is one of the naturally existing phosphorylated amino acids, which is generally an intermediate for protein phosphorylation mechanism<sup>[46]</sup>. Phosphorhistidinetriazole designed and synthesized by the Muir lab can be an easily synthesized analog for it<sup>[47]</sup>. Additionally, the same lab has designed and prepared a *N*-substituted aminooxy amino acids for the conjugating to ADP-ribose as analog for ADP-ribosylation. Although easily accessible, concerns exist about whether these peptide based PTM mimics can be able to reflect the real biological functions or regulatory mechanism accurately enough comparing with the full protein with three-dimensional folded structures<sup>[48]</sup>.

### 2.1.2 Peptide ligation strategies and the application on Lys PTMs

Although SPPS largely increased the peptide synthesis efficiency, the overall yield still drops to a very low level after a sequential ligation of 50 AAs. Thus, ligation strategies are necessary to join multiple peptide chains into an entire protein after cleavage from the resin. To carry out the peptide ligation under mild aqueous condition without any side chain protecting groups are always a dream for peptide chemists. And in 1994 the Kent group<sup>[49]</sup> first introduced the native chemical ligation (NCL) reaction to lay the foundation of protein total synthesis. And by applying the discovery of protein splicing process, intein based protein semi-synthesis, also known as expressed protein ligation (EPL) has been developed in 1998<sup>[50]</sup>. In the later decades, many modifications have been made to extend the limitation of the necessary *N*-terminal Cys and important improvements have been achieving to build up masked thioesters. Examples are the ligation of peptidylhydrozine.

Protein total synthesis<sup>[51]</sup> and semi-synthesis<sup>[52]</sup> can generate tailor-made proteins with theoretically any natural or non-natural structures<sup>[53]</sup> to solve many biological questions that are otherwise hard to achieve<sup>[54]</sup>. For instance, protein total-synthesis procedures have been applied for the synthesis of many target proteins with site-specific Kme3 so far, such as the histone H3K9me3 by the McCafferty group in 2003<sup>[55]</sup>, histone H3K9me3 by the Aimoto group in 2013<sup>[56]</sup>, histone H3K4me3 by the Liu group in 2014<sup>[57]</sup>, and H3Kme3 library by the Muir group in 2014<sup>[58]</sup>, etc. However, to the best of our knowledge, no other strategies have been known to be able to site-specifically install a Kme3 in proteins<sup>[59]</sup>.

Apart from the ligation to form the backbone of peptide chain, NCL has also been applied for the formation of isopeptide bonds, especially for the ligation of Ub and UBLs mentioned above. So far, Ub chains with different unit numbers have been achieved with total-synthesis<sup>[25]</sup>. Here one typical example from the Brik group of chemical ubiquitination of H2B at Lys34 will be discussed<sup>[60]</sup>. In the process, Ub was first ligated to the second segment (22-57) by isopeptide bond at thiol-functionalized Lys side chain. The product was then conjugated to the first segment (1-20). With a convergent peptide ligation strategy, a Thz protected segment 3 (59-96) and a segment 4 (98-125) was ligated, and the activation of the peptide 1+2 hydrazine facilitated its ligation with peptide 3+4 (Figure 3).



**Figure 3 Scheme for the convergent synthesis of H2B K34Ub**

### 2.1.3 The shortage of protein total and semi-synthesis

Although NCL and EPL have been widely used to prepare proteins with natural or non-natural structures, several shortages exist to cause significant concerns about the protein products. Artificial proteins ligated from synthetic peptides need HPLC purification followed by the refolding process before biological studies.

Another key point hinders the real extension of protein synthesis to a wider biological interest is the hardcore organic synthesis behind it. Till now, few labs over the world have the expertise to perform a total synthesis of a protein over 200 AAs, although it had been claimed not a challenge anymore for over ten years. Therefore, to discover simpler and more convenient strategies of protein synthesis especially suitable for biologists are one of the main directions for the further chemical biology development.

## **2.2 Bioorthogonal Reactions and Protein Labeling**

The foundation of bioorganic chemistry is the development and application of different reactions suitable to life-related systems for the functionalization of POIs<sup>[61]</sup>. In this part, a brief introduction of bioorthogonal reaction will be provided followed by a discussion of applying it to generate Lys PTM mimetics.

### **2.2.1 The definition and application of bioorthogonal reactions**

The term “Bioorthogonal reaction”, first proposed by the Bertozzi lab, was used to describe those chemical reactions with the capability of undergoing in living cells<sup>[62]</sup>. Due to the lack of reactions powerful enough, not too many of them can meet this requirement. Therefore, the term was extended to reactions on biological molecules with selectivity. Some of them can undergo in cell lysate with the selectivity over various proteins, while the others can undergo with purified proteins over different amino acid residues. Although being distinct from the original dream, bioorthogonal reactions have still been an important weapon for further biological studies<sup>[63]</sup>.

Till now, a large variety of bioorthogonal reactions have been developed for various purposes such as bioconjugation, labeling, localization, stabilization, etc. Typical reactions are Staudinger reaction, traceless-Staudinger ligation, thiol-ene reaction, Click reaction, reductive amination reaction, etc, with special functional groups. These warheads can be native amino acid side chains or been installed via other approaches<sup>[64]</sup>.

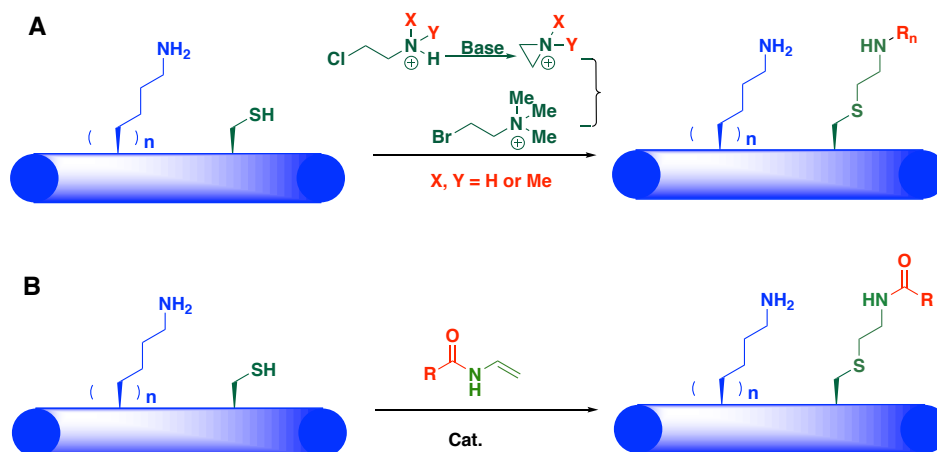
With the development of organic chemistry especially metal-catalyzed carbon-carbon coupling reactions, an increasing number of reactions have been transferred from glass flasks to macromolecular systems<sup>[65]</sup>. To develop and discover novel bioorthogonal reaction with good reliability, high efficient, no toxicity is still one of the main tasks for bioorganic chemistry.

### 2.2.2 To install Lys PTM mimics via bioorthogonal reactions

The early trials of applying bioorthogonal reactions into the PTM studies started with the design and synthesis of POIs with methyllysine analogs (MLAs). Based on the alkylation of cysteine (Cys) on the entire expressed protein level, Shokat and coworkers<sup>[66]</sup> installed *S*-(2-aminoethyl)-*L*-Cys containing various methylation levels as mimics of Kme1, Kme2, or Kme3, respectively (Figure 4A). With the incorporation of MLAs into recombinant histones, they were able to further investigate the regulatory functions of methylation to the chromatin structure and functions.

Besides MLAs, by using thiol-ene reaction, we can convert Cys to Lys analogs with acylation modifications such as small molecular acetylation or ubiquitination<sup>[67]</sup> (Figure 4B). However, due to the lack of residue selectivity over different Cys, this

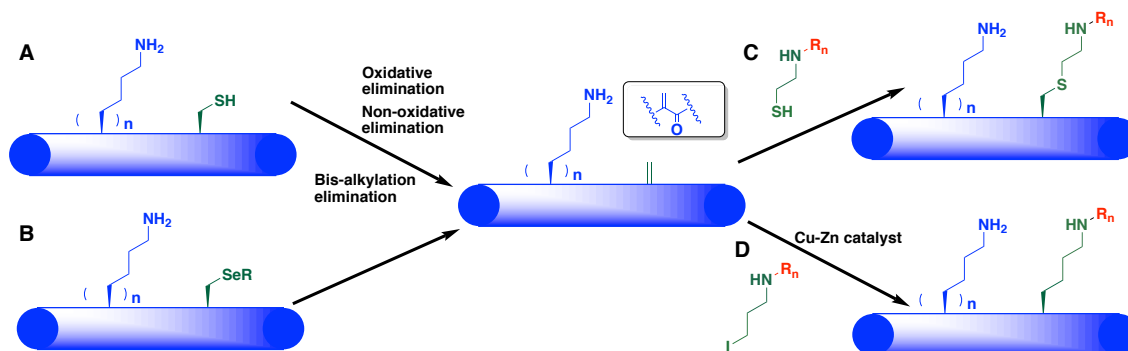
strategy requires a single Cys among the POI sequence, which can be generated from site-directed mutagenesis.



**Figure 4 Cys selective modification for Lys alkylation or acylation analog installation**  
 (A) Cys alkylation for Lys alkylation analog installation; (B) Cys thiol-ene reaction for Lys acylation analog installation

A similar but slightly advanced approach is to carry out bioorthogonal reactions on dehydroalanine (Dha) at a specific site to generate modified proteins. Michael addition reactions on Dha, as an  $\alpha,\beta$ -unsaturated carbonyl group, will result in MLAs with similar structures as the alkylation of Cys. In these methods, the Dha is first installed into POIs in a site-specific manner via different strategies, such as the oxidative elimination mechanism<sup>[68]</sup>, non-oxidative elimination mechanism<sup>[69]</sup>, *bis*-alkylation-elimination mechanism<sup>[70]</sup>, etc (Figure 5A&B). Then, the Dha can undergo thiol-ene reactions with various kinds of nucleophiles to generate corresponding PTM analogs including methylation, phosphorylation, glycosylation, etc. For instance, by using this approach, the Davis group<sup>[69]</sup> prepared modal proteins with site-specific Kme2 analog incorporated with over 95 % yield (Figure 5C). Although the site specificity is guaranteed, an extra sulfur

atom is usually remaining in the final products, introduced from the Michael addition reactions. To avoid the appearance of a sulfur atom, Park et al.<sup>[71]</sup> developed Zn-Cu bimetal-catalyzed radical addition reactions over Dha to introduce pure carbon-carbon bond linked modified Lys structures in 2016. Methylated H3K79 was generated for the further investigation of the biological functions and PTM cross-talks (Figure 5D). Additionally, Davis and the coworkers<sup>[72]</sup> reported a similar approach for the generation of theoretically any PTM structures. Termed as “Posttranslational mutagenesis”, alkyl halides are used to react with Dha with radical mechanism to form native carbon-carbon bond linkages in this method. Although an improvement has been made, this strategy still provides a D/L enantiomer mixture, which is unable to be separated. Additionally, radical based reactions can also be applied to other amino acid modifications<sup>[73]</sup>.



**Figure 5 The generation of Dha for Lys PTM analog installation.** (A) the Dha generation from Cys; (B) the indirect installation of Dha with selenium containing ncAA; (C) the conversion of Dha to Lys alkylation mimics via thiol-ene reaction; (D) the conversion of Dha to Lys PTM and analogs

## 2.3 NcAA Incorporation Technique

The last main chemical biology technique is the ncAA approach for the installation of non-natural structures, which will be discussed in this part. Comparing with the residue-replacement approach<sup>[74]</sup>, the amber-suppression approach can guarantee the site-specificity, which makes it more suitable for the synthetic purpose. People's effort to introduce methyllysines into proteins with their native structures started from the direct incorporation, in which the codon suppression technique was applied to install them during translation for the generation of posttranslationally modified proteins.

### 2.3.1 Introduction to ncAA technique

The trials of non-natural amino acid incorporation via biosynthesis approach started with the chemical misacylation of tRNAs with the corresponding amino acids in vitro<sup>[75]</sup>. The Schultz group developed a strategy in which a modified amino acid can be installed into a peptide chain at specific site by an aminoacyl tRNA synthetase (aaRS) and tRNA pair in response to the amber stop codon (TAG)<sup>[76]</sup>. Also known as non-canonical amino acid incorporation (ncAA), This aaRS-tRNA pair is specially generated to charge the unnatural amino acid from a selection of a mutant library to guarantee its orthogonality over other native aaRS for natural ones.

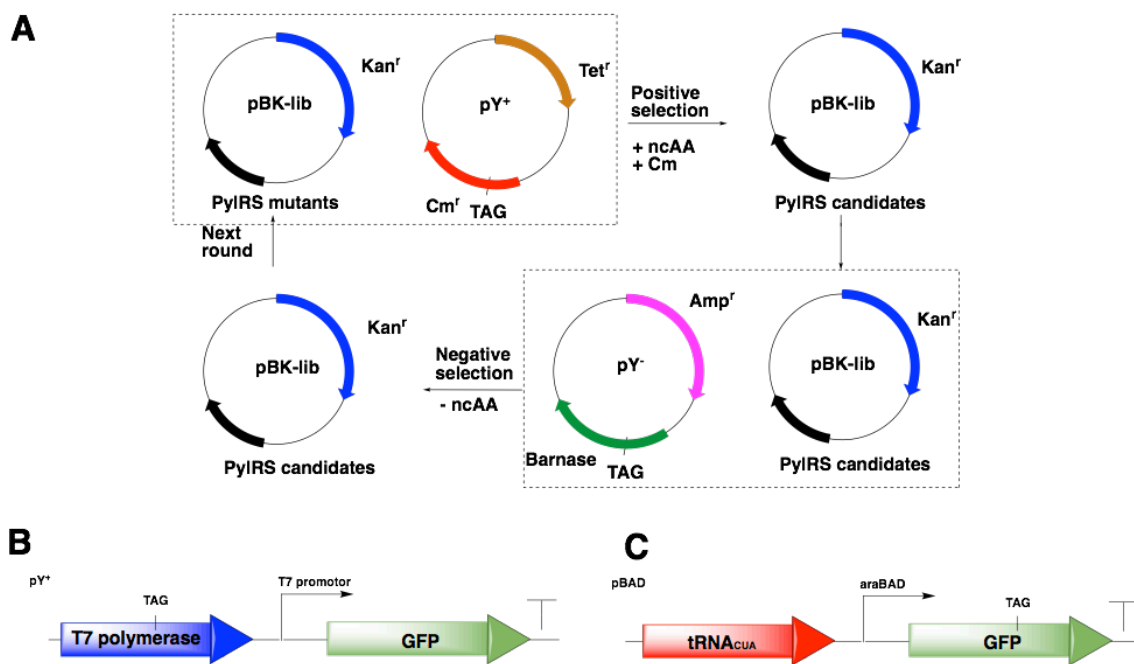
For the selection of PylRS mutant, a “double sieved” selection process followed by a screening process is generally used. A pBK plasmid with the PylRS library was used to transform a pY<sup>+</sup> cell line containing a pY<sup>+</sup> plasmid. The Cm resistance gene on pY<sup>+</sup> plasmid has a TAG codon. The cells are let to grow on a media with the existence of both



ncAA and Cm. Only if the PylRS mutant can encode the amber codon with either natural or non-natural amino acid, the cell can still be alive. Those cells will be harvested and lysed to purify plasmids to transform pY<sup>-</sup> cells. The pY<sup>-</sup> cells contain a pY<sup>-</sup> plasmid with TAG in bacterial RNA Ribonuclease (Barnase). When growing on plate without ncAA, all the mutants which charge TAG with natural AAs will not survive. The cells are harvested and lysed to purify their pBK plasmids. The positive-negative selection process will be repeated for over three times. The A detailed process will be discussed in Part 3.5, which generally provides several mutant candidates (Figure 6A).

To further optimize and find out the best one, a screening step will be carried out through comparing the expression of GFP. The pY<sup>+</sup> plasmid with certain PylRS mutant candidate under the control of T7 promoter with an amber codon TAG incorporated. Under the existence of ncAA, the T7 polymerase can be expressed and sequentially facilitate the expression of downstream GFP. The fluorescence visible by naked eyes is corresponding to the efficiency of ncAA incorporation (Figure 6B). Generally, after the selection and screen to get the best mutant, the gene will be installed to pBAD plasmid. An expression of model proteins with the ncAA installed will be carried out for ESI-MS analysis to confirm the molecular weight (Figure 6C).

Till now, about 150 ncAAs have been installed into proteins with various structures<sup>[77]</sup>. With these new functional groups incorporated, many novel bioorthogonal reactions can be applied to protein bioorganic field, which also lays the foundation of this thesis.



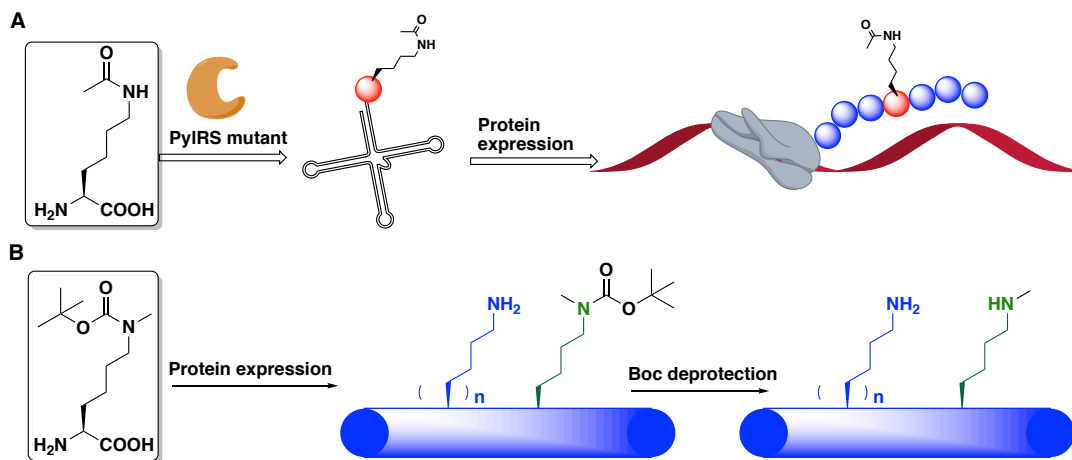
**Figure 6** The selection, screening and expression of ncAA. (A) a “double sieved” selection process; (B) the screening process; (C) the expression of GFP with ncAA

### 2.3.2 The direct incorporation of Lys PTM into proteins

With the help of ncAA technique, many naturally existing modified amino acids have been installed into POIs with their naturally modified structures, including Kac<sup>[78]</sup>, Kpr<sup>[79]</sup>, *N*<sup>ε</sup>-crotonyllysine (Kcr)<sup>[80]</sup>, *N*<sup>ε</sup>-2-hydroxyisobutyryllysine (Khib)<sup>[26]</sup>, etc (Figure 7A). A routine selection process was carried out for each of these PyIRS mutants.

Although convenient and direct, the direct incorporation may not be perfect to all the Lys PTMs for several reasons. First, the selection of PyIRS mutant is routine but requiring special expertise that not many biochemistry labs can have. Second, the searching for optimized PyIRS mutant may not always work. Due to the non-polarity of the enzyme active site, all the trials of direct incorporation for acidic PTM as Kmal, Ksu,

Kgl have turned out to be failure. Last but not the least, some of the Lys modifications with large volumes such as ubiquitination and lipidations are not applicable to this approach.



**Figure 7** Examples of directly incorporation of PTMs or bioorthogonal reactions on ncAA-containing proteins (A) the direct incorporation of AcK; (B) the installation of Kme1 with the incorporation of Boc protected precursor followed by the deprotection process.

### 2.3.3 The combination of ncAA with bioorthogonal reaction approach

Besides the direct incorporation of posttranslationally modified Lys into proteins, a combination of ncAA with bioorthogonal reaction is opening another avenue. As mentioned above, the main advantage for the ncAA incorporation is to install a non-natural functional group (FG) into a desired position in protein, which can further undergo certain reactions to be converted to Lys PTMs. At the same time, the bioorthogonal reactions can largely extend the scope for ncAA.

Lys mono-methylation, for example, cannot be directly incorporated because aaRS fails to differentiate it from Lys. However, several precursors with protecting groups on the  $\epsilon$ -amino group, also known as “Masked Lys”, such as the *tert*-Butyloxycarbonyl (Boc) group<sup>[81]</sup>, allylcarbamoyl group<sup>[82]</sup>, and photocaging groups<sup>[83]</sup>, were successfully incorporated via the ncAA technique. The deprotection of these groups under various conditions could release proteins with site-specific Kme1 *in situ* (Figure 7B).

However, the approach for the Kme1 incorporation is not applicable for the Kme2, because a protected Kme2 side chain will generate a quaternary amino group with a positive charge, whose recognition will be hard for aaRS with a hydrophobic active site<sup>[84]</sup>. A novel approach for the installation of Kme2 will be discussed in the next Chapter.

## CHAPTER III

### THE SYNTHETIC INVESTIGATION OF LYSINE DIMETHYLATION

#### 3.1 Research background

Because of a limited understanding of protein Lys methylation, chemical biologists have been trying to achieve the site-selective installation of different level of methyllysine into proteins for decades, Kme, Kme2 and Kme3. The Cys alkylation strategy mentioned in Part 2.2 can only provide methyllysine mimics in general. With a much higher abundance of Lys comparing with Cys, it's almost not possible to directly distinguish one desired Lys from a pool of others, especially for histone proteins. Therefore, we cannot eraser all the other Lys residues to guarantee the site-specificity. Additionally, the selection of a special PylRS mutant for Kme has turned out to be a failure. The small structural difference after methylation cannot be easily distinguished with the original Lys, and the negative selection can hardly pass. Therefore, protein chemistry has been drawing extensive efforts so far.

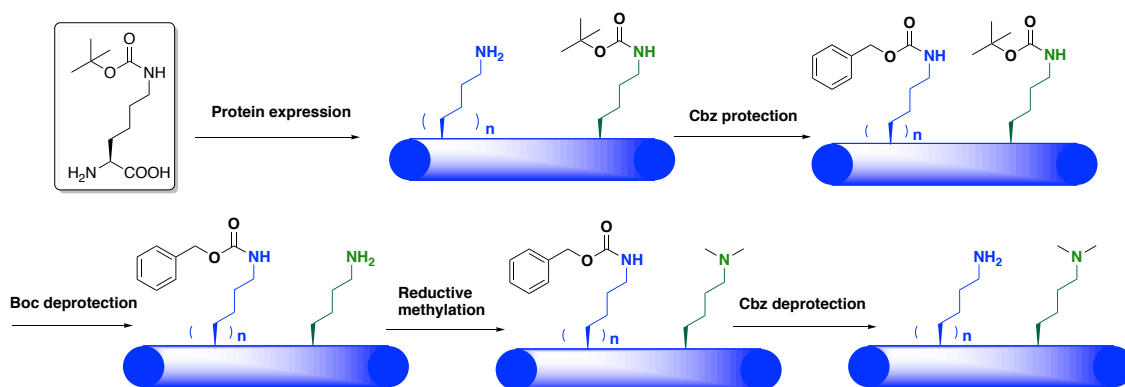
##### 3.1.1 The chemical installation of Kme1

One widely used approach is to install a protected Kme into POIs. Also known as “masked Lys”, these protected Lys derivatives can undergo a deprotection step to release Kme as expected. The Chin group<sup>[81]</sup> introduced a first example of installing Boc protected Kme into histone H3K9, whose acid-deprotection could release Kme with native strature. With the development of these “bond-cleavage” bioorthogonal reactions, many other ncAA with protecting groups have been applied, such as an allylcarbamoyl-protected Kme

into H2BK27 by the Schultz group, photocaged Kme derivatives separately by the Liu group into GFP and the Chin group.

### 3.1.2 The chemical installation of Kme2

However, these approaches abovementioned can only apply for Kme, which will generate a positively charged state for Kme2 and not suitable for the enzymatic recognition anymore. It was not until 2010 that the Chin group<sup>[85]</sup> developed a novel method for the generation of Kme2-containing histone H3K9me2. First, a Boc-protected Lys was installed at the H3K9 site via the ncAA technique, followed by the global protection of all the other existing Lys side chains with the Benzyloxycarbonyl (Cbz) protecting group. Then, the Boc group deprotection will selectively release the only Lys at the specific site, which can further undergo reductive methylation reaction with formaldehyde to form Kme2 only at H3K9 site (Figure 8).



**Figure 8** The combination of ncAA and bioorthogonal reactions for the site-specific incorporation of Kme2 with a global protection-deprotection based approach.

Finally, all the Cbz groups are removed to finally generate the Kme2-containing protein in its fully native structure. However, concerns exist about proteins sensitive to the global protection/deprotection conditions of the other Lys side chains.

### 3.2 Methodology Design of AcdK

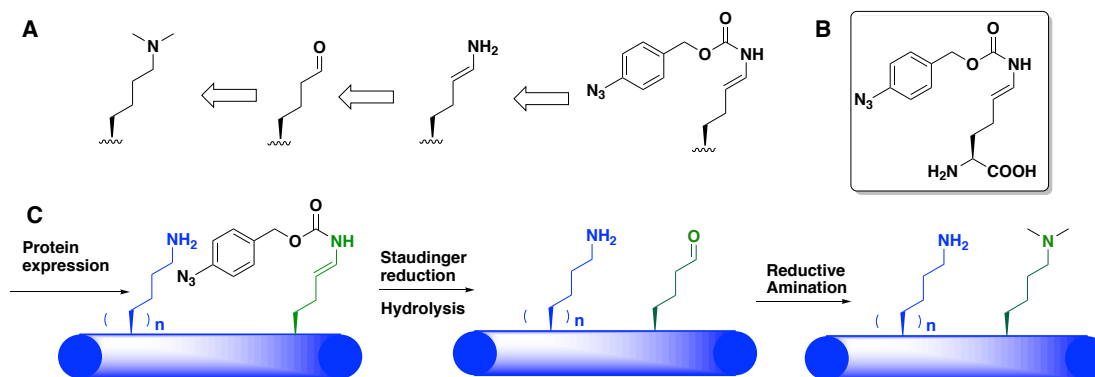
Although elegant, the global protection and deprotection procedures, as well as the acid deprotecting process for Boc group, are not under mild condition, which raises concern about protein unfolding. Thus, we plan to develop novel methods for the robust and convenient introduction of Lys dimethylation. In order to install lysine dimethylation into POIs at specific sites, we envision a reversed approach that allysine (AlK) can undergo the reductive amination with dimethylamine to generate Kme2 in proteins, which may be genetically encoded via the amber suppression mutagenesis approach. Comparing with the method mentioned above to react formaldehyde with a Lys on POI, the aldehyde group is directly installed onto POIs, which excessive amount of dimethylamine will be provided outside (Figure 9A).

With the concern of the cellular toxicity from the aliphatic aldehyde group on its side chain, a precursor of AlK is needed which can further undergo deprotection to generate AlK in vivo. Therefore, *N*<sup>ε</sup>-(4-azidobenzoxycarbonyl)- $\delta,\epsilon$ -dehydrolysine (AcdK) was then designed out as an amino acid precursor of AlK, with a shielded side chain aldehyde as a similar design an *o*-azidobenzoyloxycarbonyl lysine from the Deiters group<sup>[86]</sup> (Figure 9B). In the structure, the azidobenzoxycarbonyl moiety reduced by a phosphine reagents can further trigger the self-cleavage process to release a  $\delta,\epsilon$ -dehydrolysine<sup>[65]</sup>.

The automatic hydrolysis of  $\delta$ ,  $\epsilon$ -dehydrolysine in water will instantaneously generate AIK. In that case, the AcdK genetically incorporated into proteins can undergo Staudinger reduction with tris-(2-carboxyethyl)phosphine (TCEP) to release AIK, which will further undergo reductive amination with dimethyllysine with the existence of NaCNBH<sub>3</sub> to finally generate proteins with site-specific lysine dimethylation (Figure 9C). As discussed above, both Staudinger reduction and reductive amination are typical bioorthogonal reactions which can be carried out under mild physiological conditions, our approach can be theoretically applied for the preparation of any native proteins with site-specific Kme<sub>2</sub> as well as the preparation of proteins with site-specific Kme or even other alkylations by simply changing dimethylamine to methylamine or any corresponding alkylamines during the reductive amination step<sup>[87]</sup>.

The idea of AIK comes from the human elastin and collagen, in which AIK is a naturally existing derivative of lysine<sup>[88]</sup>. Although the free aldehyde group may form intramolecular imine with other free amino groups on either Lys side chains or the N-terminal amine, this process is reversible and the incubation with high concentration of external small molecular amine for hours can guarantee the formation of desire imine product.



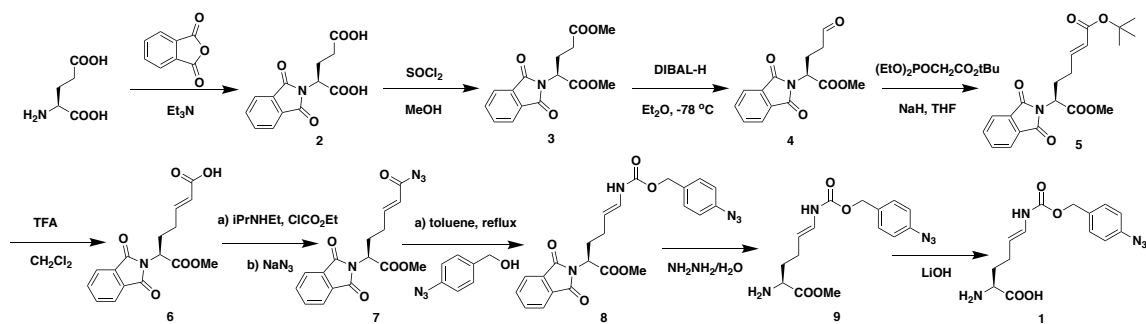


**Figure 9 Chemical biology approaches for the site-selective Lysine dimethylation in proteins.** (A) the general design of AcDK; (B) the structure of AcDK; (C) the conversion from AcDK to Kme2.

### 3.3 The Synthesis of AcDK

#### 3.3.1 General description

A synthetic route of AcDK is a complicated process. It starts with the selective DIBAL reduction of the side chain of natural *L*-glutamate to form an aldehyde group. The yield reaction followed by the Curtius rearrangement to form the acylazide as an activated acyl group. After the coupling reaction to form an ester bond, the deprotection will finally generate the product as dilithium salt of AcDK as designed. Although the overall synthesis involves 9 steps in total, gram quantities of AcDK have been routinely produced. The synthesis was further optimized to improve the overall yield. The finally optimized yields for each step was presented in Figure 10. This part is originally designed and achieved by Dr. Yadagiri Kurra.



**Figure 10** Scheme for the synthesis of AcDK

### 3.3.2 Detailed procedure

#### **The synthesis of compound (S)-2-(1,3-dioxoisindolin-2-yl)pentanedioic acid (2):**

To a 500 mL round bottom flask was added L-glutamic acid (30 g, 0.2 mol), phthalic anhydride (29.6 g, 0.2 mol), triethylamine (2.79 mL, 0.02 mol) and toluene (250 mL). The mixture was heated to reflux, and the water formed in the reaction was collected using a Dean-Stark apparatus. After 24 h, the reaction was cooled to room temperature and solvent was removed under reduced pressure. EtOAc was added to dilute the reaction and pH was adjusted to 2 with HCl (3.0 M). Then the mixture was extracted with EtOAc (300 mL × 3) and the organic layers were dried and concentrated. The resulting solid was washed with EtOAc and then filtered to collect the solid. Drying the solid under high vacuum afforded the *N*-protected *L*-Glutamic acid (2) (22.5 g, 41% yield).

#### **The synthesis of compound dimethyl (S)-2-(1,3-dioxoisindolin-2-yl)pentanedioate (3):**

(2) (22.5 g, 0.081 mol) was dissolved in MeOH (200 mL) and cooled to 0 °C. SOCl<sub>2</sub> (23.5 mL, 4 e.q.) was added dropwise. The mixture was warmed to room temperature naturally and stirred overnight. All the volatiles were removed under reduced pressure and

then diluted with EtOAc. The organic layer was washed with aqueous NaHCO<sub>3</sub>, dried, and concentrated to give methyl ester 3 (25.4 g, 100%).

**The synthesis of compound methyl (S)-2-(1,3-dioxoisindolin-2-yl)-5-oxopentanoate (4):** 3 (25.4 g, 0.083 mol) was dissolved in anhydrous diethyl ether (300 mL), cooled to -78 °C. DIBAL-H (99.8 mL, 1.0 M in hexane, 1.2 eq.) was added dropwise under argon protection. Stirring continued for another 1 h. Water (10 mL) was added. The reaction was warmed to room temperature and continually stirred for 1 h. The precipitate was removed with Celite and the filtrate was concentrated to produce the crude aldehyde (4), which was used directly to the next step without further purification.

**The synthesis of compound 1-(tert-butyl) 7-methyl (S,E)-6-(1,3-dioxoisindolin-2-yl)hept-2-enedioate (5):** Anhydrous THF (80 mL) was added to a round bottom flask and cooled to 0 °C. NaH (1.9 g, 60% dispersion in mineral oil, 1.2 eq.) was added carefully. Then tert-butyl 2-(diethoxyphosphoryl)acetate (11.99 g, 1.2 eq.) was added dropwise. After bubbles ceased to release, 4 (~ 0.0396 mol) in THF (20 mL) was added and the mixture was warmed to room temperature. After 2 h, NH<sub>4</sub>Cl aqueous was added to quench the reaction. After diethyl ether extraction, the organic phase was dried and concentrated. The residue was purified via column chromatography with hexane/EtOAc (3:1 v/v) as eluent to give product 5 (6.9 g, 22% yield, two steps).

**The synthesis of compound (S,E)-6-(1,3-dioxoisindolin-2-yl)-7-methoxy-7-oxohept-2-enoic acid (6):** (5) (6.9 g, 0.0184788 mol) was dissolved in CH<sub>2</sub>Cl<sub>2</sub> (70 mL), cooled to 0 °C. Trifluoroacetic acid (14.15 mL, 10 eq.) was added dropwise. Then the solution was warmed to room temperature and stirred for 1.5 h. The reaction was diluted with CH<sub>2</sub>Cl<sub>2</sub>,

washed with H<sub>2</sub>O, then dried and concentrated. The residue was purified via column chromatography with hexane/EtOAc (1:1 v/v) as eluent to afford pure acid 6 (6.1 g, 85% yield).

**The synthesis of compound methyl (*S,E*)-7-azido-2-(1,3-dioxoisindolin-2-yl)-7-oxohept-5-enoate (7):** 6 (5.0 g, 0.015758 mol) was dissolved in acetone (150 mL), cooled to 0 °C and then added diisopropylethylamine (6.587 mL, 2.4 eq.). Ethyl chloroformate (3.3 mL, 2.2 eq.) was added dropwise and the solution was stirred at 0 °C for 1 h. Sodium azide (5.12 g, 5 eq.) in water (25 mL) was added and the reaction mixture was warmed to room temperature. After 1 h with stirring, brine was added and extracted with EtOAc. The organic layer was dried and concentrated. The residue was purified via column chromatography with hexane/EtOAc (3:1 v/v) as eluent to afford pure methyl ester 7 (4.6 g, 85% yield).

**The synthesis of compound methyl (*S,E*)-6-(((4-azidobenzyl)oxy)carbonyl)amino)-2-(1,3-dioxoisindolin-2-yl)hex-5-enoate (8):** 7 (4.6 g, 0.01344 mol) was dissolved in toluene, and heated to reflux for 2 h under argon protection. Then (2-azidophenyl)methanol (2 g, 1.0 eq.) was added and the mixture was continuously refluxed for another 2 h under argon protection. It was then cooled to room temperature. The solvent was removed under reduced pressure and the residue was purified via column chromatography with hexane/EtOAc (3:1 to 1:1 v/v) as eluent to afford product 8 (5.1 g, 92% yield).

**The synthesis of compound methyl (*S,E*)-2-amino-6-(((4-azidobenzyl)oxy)carbonyl)amino)hex-5-enoate (9):** 8 (5.1 g, 0.011 mol) was dissolved

in ethanol/dichloromethane (16 mL/25 mL) and hydrazine monohydrate (0.85 mL, 1.6 eq.) was added. The mixture was stirred at room temperature overnight. The white solid formed in the reaction was filtered away and the filtrate was concentrated. The residue was purified via column chromatography with methanol/ammonium hydroxide/dichloromethane (15:50:450 v/v/v) as eluent to afford amine 9 (2.2 g, 60% yield).

**The synthesis of compound AcdK (1):** 9 (2.2 g, 0.0066 mol) was dissolved in THF/H<sub>2</sub>O (25 mL/12.5 mL) and LiOH (0.3168 g, 2 eq.) was added. After stirring for 1 h, the undissolved impurities were filtered away and the filtrate was concentrated under vacuum to afford amino acid 1 (2.1 g, 96% yield). <sup>1</sup>H NMR (CD<sub>3</sub>OD, 300 MHz) 7.40 (d, 2H, J = 4.5 Hz), 7.07 (d, 2H, J = 4.5 Hz), 6.43 (dd, 1H, J = 0.9, 8.7 Hz), 5.12 (dd, 1H, J = 4.5, 8.7 Hz), 5.08 (s, 2H), 3.20 (dd, 1H, J = 0.9, 4.2 Hz), 2.12-2.07 (m, 2H), 1.76-1.73 (m, 1H), 1.60-1.56 (m, 1H). <sup>13</sup>C NMR (CD<sub>3</sub>OD, 75 MHz) 181.4, 154.6, 139.8, 133.6, 129.3, 123.9, 118.6, 110.4, 65.5, 55.7, 36.0, 26.1 (Figure 11).

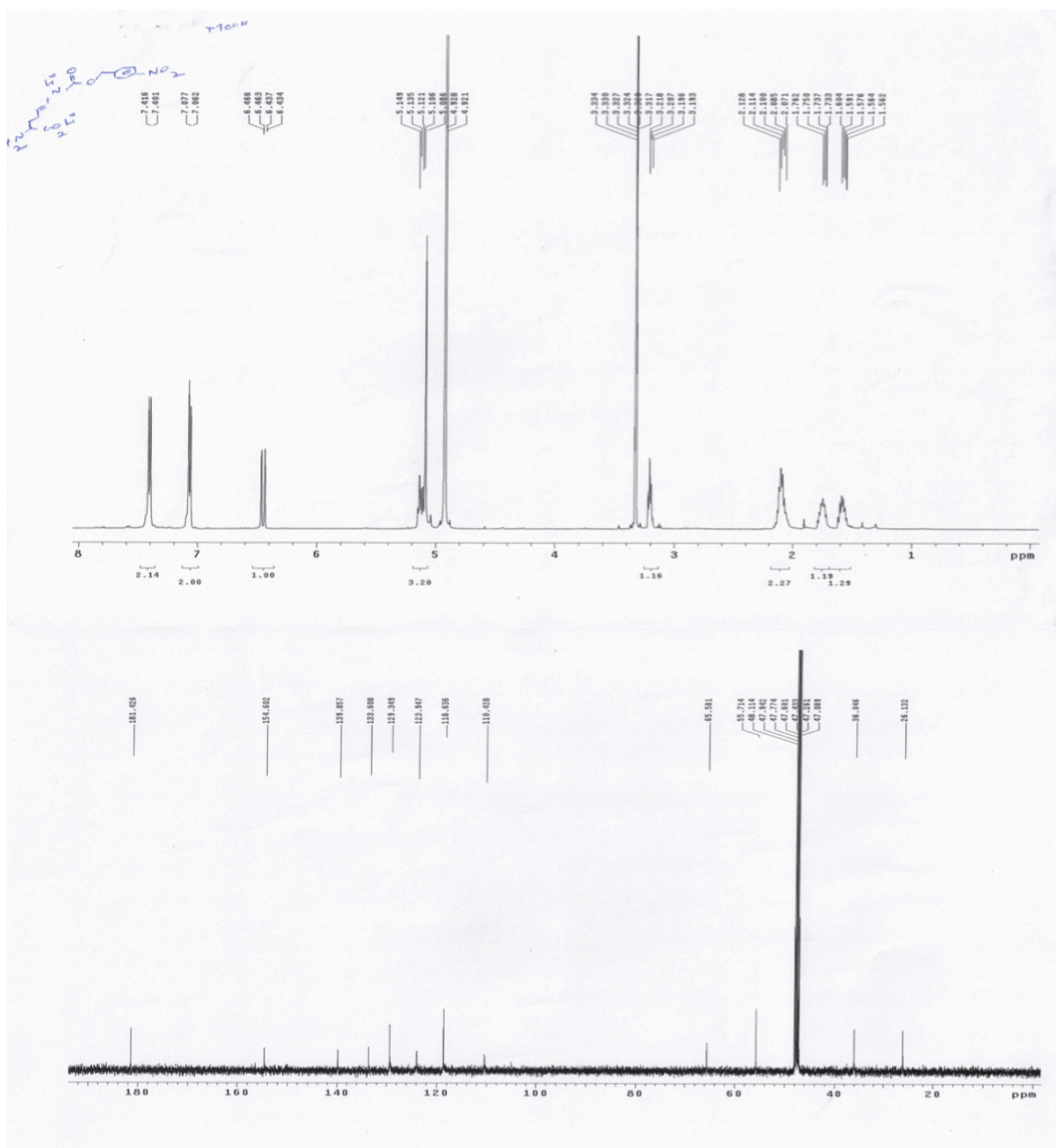


Figure 11  $^1\text{H}$ NMR and  $^{13}\text{C}$ NMR of AcdK

### 3.4 The Construction of a PylRS mutant library

To start with, we built up a *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (MmPylRS or short for PylRS) gene library with four sites random mutations. The codon

for one active site residue Y384 was randomized as codons for residues Y, F, and W, and then codons for three active site residues Y306, L309, and C348 were randomized as NNK nucleotides (N=A or C or G or T, K=G or T). For the site-direct mutagenesis as described, several pairs of primers were used for PCR with Phusion DNA polymerase, as pBK-mmPylRS-348NNK-F (5'-ACC ATG CTG AAC TTC NNK CAG ATG GGA TCG GGA TGC ACA CGG-3'), pBK-mmPylRS-348NNK-R (5'-AAA CTC TTC GAG GTG TTC TTT GCC GTC GGA CTC-3'), pBK-mmPylRS-306-309-NNK-F (5'-CTT GCT CCA AAC CTT NNK AAC TAC NNK CGC AAG CTT GAC AGG GCC CTG CCT-3'), pBK-mmPylRS-306-309-NNK-R (5'-CAT GGG TCT CAG GCA GAA GTT CTT GTC AAC CCT-3'). The gene was finally cloned into pBK plasmid as reported by the Schultz group. The products from PCR were phosphorylated by T4 polynucleotide kinase and ligated by T4 ligase. The final mutant pBK-MmPylRS library plasmid mix was used to transform Top10 cells with a kanamycin (Kan) selection marker.

### **3.5 The Selection, Screening and Incorporation of AcdK**

WT PylRS-tRNA<sup>Pyl</sup> pair and its mutants have been widely applied for the genetic incorporation of many ncAAs into naturally expressed proteins at amber mutation sites in different cell strains<sup>[84]</sup>. Mainly lysine derivatives, the Liu group also extended it for phenylalanine derivatives<sup>[89]</sup>. After getting sufficient amount of AcdK in its dilithium salt form, we started to carry out the selection and screening of PylRS-tRNA<sup>Pyl</sup> mutant library for the incorporation of AcdK into *E. coli* via the amber codon suppression strategy<sup>[90]</sup>. A double sieved selection procedure<sup>[78]</sup> generally reported for the identification of mutant

aminoacyl-tRNA synthetases for ncAAs in *E. coli* has been employed, which contains a three positive selections coupled with a two negative selections in between.

### 3.5.1 The positive selection of AcdK

To start with, we carried out the first positive selection. The mutant library pBK-MmPylRS plasmid mix was used to transform Top10 electrocompetent cells that contains the positive selection plasmid to yield a cell library greater than  $1 \times 10^9$  cfu. For this positive selection plasmid, also known as pY<sup>+</sup>, was previously constructed containing several parts including a gene coding tRNA<sup>Pyl</sup>, a tetracycline (Tet) selection marker, a chloramphenicol (Cm) acetyltransferase gene for Cm resistance, an amber mutation at its D112 position, a T7 RNA polymerase gene with two amber mutations at positions 1 and 107, and a green fluorescent protein GFP<sub>UV</sub> gene under control of a T7 promoter. After the transform, cells were then plated on the GMMML agar plates (1×M9 Salts, 1 % glycerol, 300 μM Leucine, 2 mM MgSO<sub>4</sub>, 0.1 mM CaCl<sub>2</sub>, and 0.2 % NaCl) with 12 μg/mL Tet, 25 μg/mL Kan, 102 μg/mL Cm, and 1 mM AcdK, and let to grow at 37 °C for 72 h. Colonies on the plates were collected and selected pBK-MmPylRS mutant plasmids were extracted.

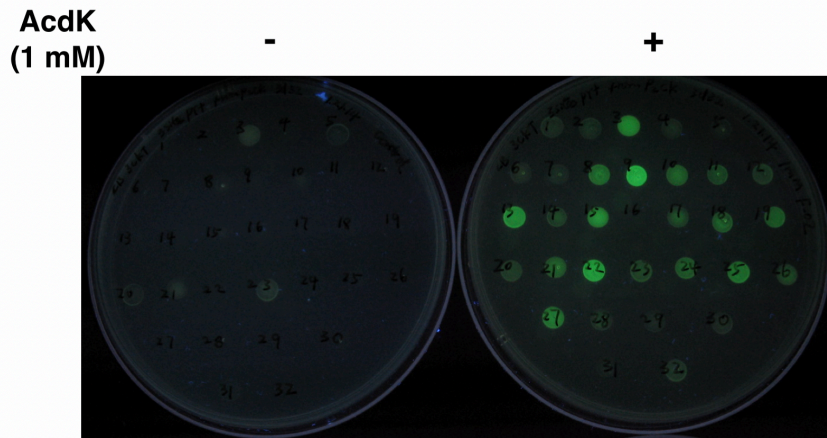
### 3.5.2 The negative selection of AcdK

Then we can carry out the first negative selection by using the extracted pBK-MmPylRS plasmids from the positive selection to transform Top10 electrocompetent cells with the negative selection plasmid as pY<sup>-</sup>. Similar to the positive selection plasmid, the negative selection plasmid is also a specialized plasmid for the selection process, with an



Amp selection marker, genes coding tRNA<sup>Pyl</sup>, and barnase with two amber mutations at Q2 and D44. After been plated on LB agar plates with 50 µg/mL Kan, 200 µg/mL ampicillin (Amp), and 0.2 % arabinose, but without the existence of AcdK to make sure that any non-specific installation of natural amino acids will cause cell death. After incubating the plate at 37 °C for 16 h, survived cells were collected and their plasmids were extracted for the second positive selection.

After three rounds of positive selections and two rounds of negative selections, we carried out the screening by the visualization of GFP<sub>UV</sub> expression. The clones picked after the third positive selection was let to grow on LB plate with 102 µg/mL Cm, 25 µg/mL Kan, and 12 µg/mL Tet but with or without 1 mM AcdK. Besides a pBK plasmid contained a mutant PylRS gene, the pY<sup>+</sup> plasmid in these cells had a tRNA<sup>Pyl</sup> gene, a Cm resistance gene with an amber mutation at D112, as well as a GFP<sub>UV</sub> gene under the control of a T7 promoter. Since two amber mutations were installed in T7 RNA polymerase gene at positions 1 and 107, the expression of GFP<sub>UV</sub> would be promoted by the successful suppression of two amber mutations in the T7 RNA polymerase. By measuring the fluorescent intensity of GFP<sub>UV</sub> expression, we can select a best colony from a large number of candidates after the third positive selection. The mutant with the best amber suppression efficiency in *E. coli* was identified as L309T/C348G/Y384F (Figure 12). The gene mutant corresponding to this PylRS mutant noted as AcdKRS was then cloned into pEvol plasmid backbone together with the tRNA<sup>Pyl</sup>.



**Figure 12** Screening of the clones after the third positive selection on LB plates.

### 3.6 The expression, purification of sfGFP-D134AcdK and its conversion to AIK

#### 3.6.1 Sequence information

*DNA sequence: sfGFP-D134TAG-His6:*

atggttagcaaaggtgaagaactgtttaccggcgttgccgattctggtggaactggatggtgatgtgaatggcataaatttag  
 cgttcgtggcgaaggcgaaggtgatgcgaccaacggtaaactgacctgaaattattgcaccaccggtaaactgccggtcc  
 gtggccgacctggtgaccaccctgacctatggcggtcagtgcttagccgctatccggatcatatgaaacgccatgattcttta  
 aaagcgcgatgccggaaggctatgtgcaggaacgtaccattagcttcaaagatgatggcacctataaaaccgctgcggaagt  
 aaattgaaggcgataccctggtgaaccgattgaactgaaaggtattgatttaagaatagggaacattctgggtcataaact  
 ggaatataattcaacagccataatgtgtatattaccgccgataaacagaaaaatggcatcaaagcgaacttaaaatccgtcaca  
 acgtggaagatggtagcgtgcagctggcggatcattatcagcagaataccccgattggtgatggcccggctgctgctgccgat  
 aatcattatctgagcaccagagcgttctgagcaaagatccgaatgaaaaacgtgatcatatggtgctgctggaatttaccgc  
 cgcgggcattaccacgggatggatgaactgtataaaggcagccaccatcatcaccattga

*Protein sequence: sfGFP-D134AznL-His6 (MW: 27866.35):*

(M)VSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLLKFICTTGK  
LPVPWPTLVTTTLTYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGT  
YKTRAEVKFEGDTLVNRIELKGIDFKE(AznI)GNILGHKLEYNFNShNVYITADK  
QKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLS  
KDPNEKRDHMLLEFVTAAGITHGMDELYKGSHHHHHH

### 3.6.2 The expression of sfGFP-D134AcK

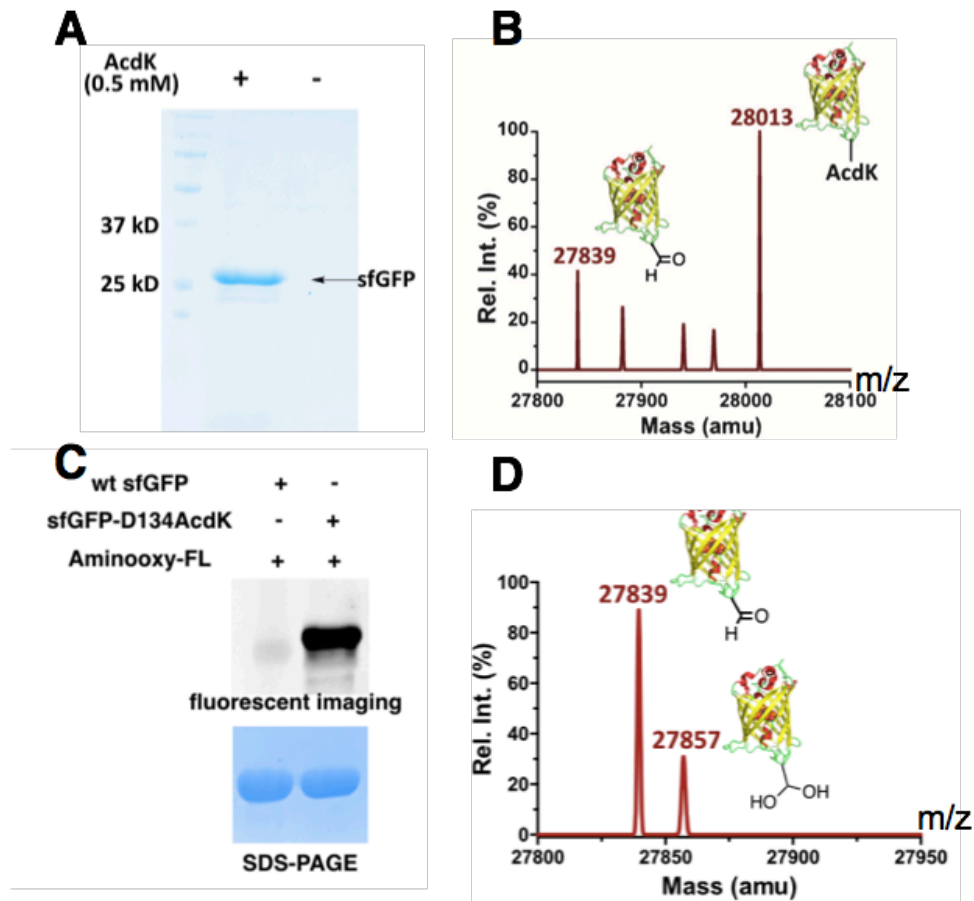
With the pEVOL-AcdKRS-PylT at hand, we also got another plasmid pBAD-sfGFP-D134TAG containing a gene for the expression of the superfolder green fluorescent protein (sfGFP) with an amber mutation at the D134 position, as well as a His<sub>6</sub> tag at the C-terminal as gift from Ryan Mehl at Oregon State University. The newly constructed plasmid pEVOL-AcdKRS-tRNA<sup>Pyl</sup> and the expression plasmid pBAD-sfGFP-D134TAG were used to co-transform *E. coli* BL21(DE3) cells. A single colony was picked and let to grow in 5 mL LB medium with 100 mg/mL Amp and 34 mg/mL Cm at 37 °C for o/n. Then, the transformed cells were used to inoculate 50 mL 2YT medium with 100 mg/mL Amp and 34 mg/mL Cm, which can express full-length sfGFP with AcK at its D134 position (sfGFP-D134AcK) after been induced by 0.2% arabinose and 0.5 mM IPTG. After growing at 37 °C for 6 h, the cells were collected by centrifugation (6,000 rpm for 15 min). After been resuspended in a 25 mL lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0) and sonicated in an ice water bath three times (4 min each, with a 5 min interval between each run), the cell lysate was precipitated by centrifugation (10000 rpm for 30 min). The supernatant was added to 1 mL Ni

Sepharose™ Fast Flow column (GE Healthcare) and the flow-through was then removed. The resin was washed with a wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 8.0) and finally eluted with an elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 8.0). After been concentrated with Amicon Ultra-15 Centrifugal Filter Devices (10000 MWCO cut, Millipore), the purified sfGFP-D134AcdK was analyzed by SDS-PAGE. A clear band can be clearly seen only with 0.5 mM AcdK afforded in the media, while none for the one without AcdK added (Figure 13A). The yield of sfGFP-D134AcdK is 7 mg/L with the presence of 0.5 mM AcdK, determined by the BCA method.

### 3.6.3 The characterization and labeling of sfGFP-D134AcdK

Additionally, we took electrospray ionization mass spectrometry (ESI-MS) to analyze the purified sfGFP-D134AcdK, in which two major peaks showed up (Figure 13B). One of the major peaks showing as 28013 Da suggests sfGFP-D134AcdK with a theoretical MW of 28013Da, while the other major peak showing as 27839 Da agrees with the reduction product of AcdK after incorporation into sfGFP-D134AcdK to sfGFP-D134AIK with a theoretical MW of 27839 Da. This implies part of the AcdK is already reduced and rearranged to AIK without the addition of phosphine reagents during the protein purification process with no reducing reagent added, possibly due to reactions with endogenous thiol- containing reagents such as glutathione and H<sub>2</sub>S in *E. coli* cells. Since AcdK will finally been converted to AIK in POIs, this partial conversion of AcdK to AIK is not a problem. To further confirm the key aldehyde group in sfGFP-D134AIK, a simple

dye labeling test was also carried out using a hydroxylamine-fluorescein conjugated dye. (Figure 13C). Because the boiling process can cause the denature of sfGFP and losing its original fluorescence, no fluorescence could be detected for the negative control WT-sfGFP sample.



**Figure 13** The expression of sfGFP-D134AcidK and its conversion to sfGFP-D134AIK (A) SDS-PAGE of sfGFP-D134AcidK; (B) ESI-MS of sfGFP-D134AcidK; (C) fluorescence labeling of sfGFP-D134AcidK showing the partial deprotection in cell; (D) the conversion of sfGFP-D134AcidK. Adapted with permission from [87].

#### 3.6.4 The conversion to sfGFP-D134AIK

After confirming the expression of sfGFP-D134AcDK, Tris(2-carboxyethyl)phosphine (TCEP) was added to the purified protein sample to a final concentration of 5 mM, and the reaction mixture was let to incubate at 37 °C for 1 h, followed by the dialysis to 10 mM pH=7.8 ammonium bicarbonate (ABC) buffer and sent for the characterization by ESI-MS. The main peak showing as 27839 Da similar to the peak in the previous figure implies sfGFP-D134AIK. Since AIK contains an aliphatic aldehyde group, which will undergo addition reaction to form gem-diol in aqueous solution, another unexpected peak was also detected as 27857 Da suggesting its corresponding hydrated form with water addition (Figure 13D). Since the gem-diol formation is a reversible step in equilibrium, it is not supposed to interfere with the following bioorthogonal reactions. This suggests that sfGFP-D134AcDK can be completely converted to sfGFP-D134AIK by simply reacting with TCEP.

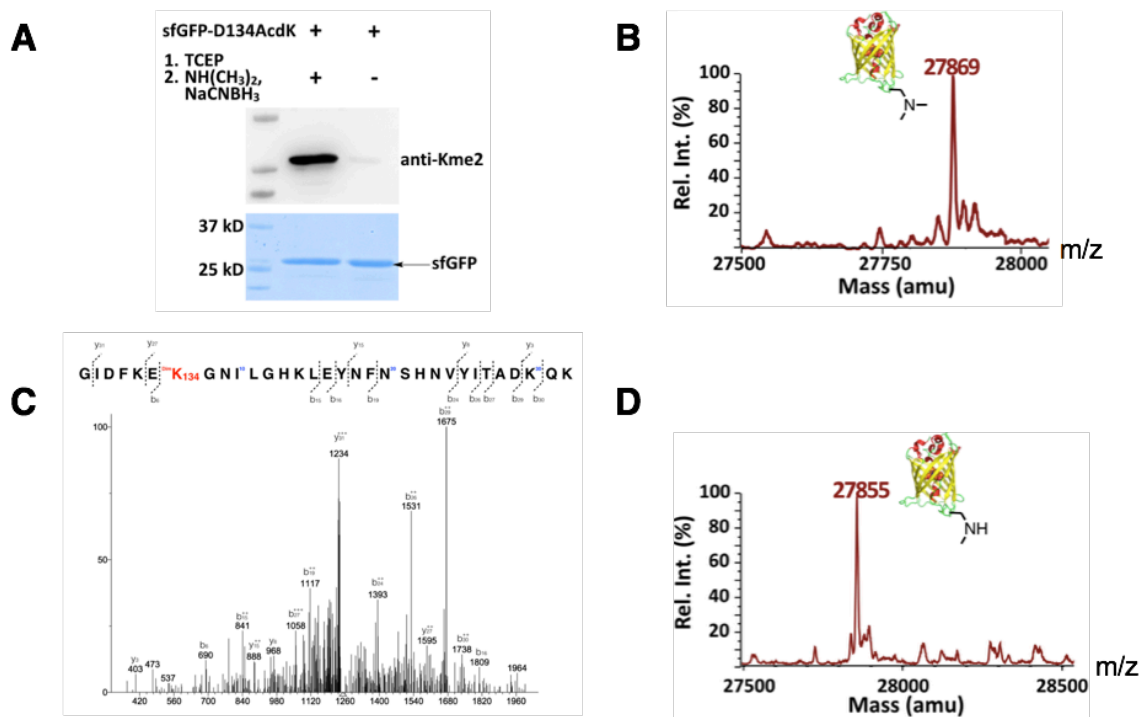
#### 3.7 Methodology investigation on sfGFP

With the protein product sfGFP-D134AIK at hand, we wanted to test our design for the generation of Kme2 at the specific site. After removing all the small molecule reactants by dialysis to 200 mM phosphate buffer pH=7, the reductive amination reaction was then carried out to convert sfGFP-D134AIK to sfGFP with Kme2 at D134 site (sfGFP-D134Kme2) by incubating with excess dimethylamine followed by the reduction of NaCNBH<sub>3</sub>.

The protein with 10-100  $\mu\text{M}$  final concentration was first incubated with 100 mM of dimethylamine for 6 h in order to convert the allysine into imine. The potential cross-linking between allysine and the native lysine residues can also be removed because the imine formation is a reversible process under aqueous condition especially with the existence of excess external small molecular amine. After the incubation with alkylamine, 100 mM final concentration of  $\text{NaCNBH}_3$  was then added and let to incubate at 37  $^\circ\text{C}$  for o/n. Different conditions have been tried to optimize the reaction yield. After finishing the reaction, the reactant mixture was dialyzed back to phosphate buffer to remove all the small molecular reagents. The reaction product was first analyzed by the Western blot with pan-anti-Kme2 antibody, which showed a clear band after reaction but not for the control without reaction (Figure 14A). An ESI-MS analysis was further applied to confirm the conversion from sfGFP-D134AIK to sfGFP-D134Kme2 as 27869 Da (Figure 14B). Besides, tandem MS (MS/MS) analysis was also carried out in order to guarantee the site-specificity of our reactions. The trypsinized sfGFP-D134Km2 fragments were detected to show that the Kme2 at the desire position (Figure 14C).

Additionally, by simply changing dimethylamine to monomethylamine, we also made sfGFP with methyllysine (Kme) at its D134 site (sfGFP-D134Kme) via a similar reaction procedure. Although we fail to detect sfGFP-D134Kme via Western blot analysis due to the lack of a robust pan-anti-Kme antibody, the finally synthesized sfGFP-D134Kme product successfully gave out a peak at 27855 Da in ESI-MS, which agreed with the theoretical mass as 27854 Da (Figure 14D). Notably, some side peaks in Figure

14B or Figure 14D are representing some unknown impurities or potential side reactions not fully studied.



**Figure 14 Methodology investigation of reductive amination on sfGFP-D134AIK (A)** SDS-PAGE and Western blot of sfGFP-D134AIK after the reductive amination reaction with dimethylamine; (B) ESI-MS of the same product after reductive amination reaction; (C) MS/MS analysis the same product. Adapted with permission from [87].

To summarize, both Staudinger reduction and reductive amination can be carried out in the phosphate buffer pH 7 at room temperature. Under this physiological condition, no protein precipitation or aggregation has been observed, while sfGFP's fluorescence remains without being quenched. These facts again proof that the tandem reactions are compatible with natural proteins with native folding states.



## **3.8 The preparation and application of histone with site-specific Kme2**

### **3.8.1 Background and general design**

After the test of methodology on model protein sfGFP as a water-soluble protein with medium size, we moved to the synthesis of histone H3 proteins with site-specifically installed Kme2 as a water non-soluble protein purified from the inclusion body. Histone dimethylation is a very important epigenetic marker widely existing over different lysine sites, which plays important roles in the regulation of gene expression and transcription factor interaction. Lysine-Specific Histone Demethylase 1 (LSD1) belongs to an important LSD family enzyme catalyzing the removal of methylation, which is a Flavin adenine dinucleotide (FAD)-dependent histone demethylase that selectively targets H3 at K4 and K9 positions but inert toward other lysine sites. Therefore, we would like to use different methylated H3 generated from our method as probes to study the substrate specificity of it. To further confirm and extend the previous reported results, we chose three different lysine sites on histone H3 as H3K4, H3K9, and H3K36 to install dimethylation modifications.

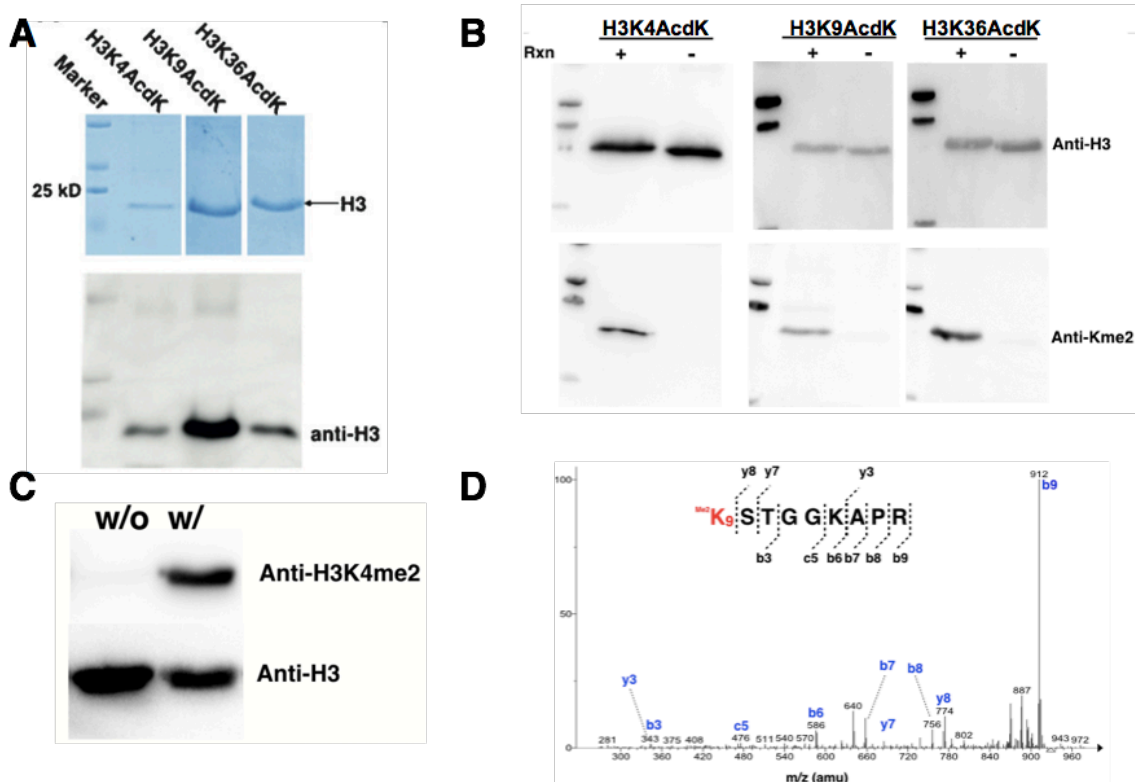
### **3.8.2 The expression of H3 with AcdK incorporated**

To start with, AcdK was installed into all the three histone protein precursors at specific positions, and the proteins were expressed and purified according to a typical histone purification protocol. Plasmids pETduet-H3K4TAG, pETduet-H3K9TAG, and pETduet-H3K36TAG constructed by our group was used separately to co-transform *E. coli* BL21(DE3) cells together with the plasmid pEVOL-AcdKRS. The transformed cells

were placed on LB agar plates with 100  $\mu\text{g}/\text{mL}$  Amp and 34  $\mu\text{g}/\text{mL}$  Cm. A single colony was picked, and 5 mL LB medium seed culture was set up with 100  $\mu\text{g}/\text{mL}$  Amp and 34  $\mu\text{g}/\text{mL}$  Cm. After been inoculated to 250 mL 2YT medium, cells were let to grow at 37 °C for 2.5 h until  $\text{OD}_{600}$  reached 0.6, followed by the induction of protein expression with the addition of 1 mM IPTG, 0.2 % arabinose and 0.5 mM AcDK. After induction for 8 h, cells were collected (4000 rpm for 20 min), washed with the lysis buffer (20 mM Tris, 500 mM NaCl, 0.1% Triton X-100, and pH 7.5; 20 mL) and resuspended in the lysis buffer. The cell pellets were then sonicated under ice bath for 3 times (2 min each time with an interval of 10 min) and centrifuged under 6000 rpm for 20 min to remove the supernatant. The precipitate was resuspended in a 25 mL wash buffer (20 mM Tris, 500 mM NaCl, and pH 7.5) inclusion body and washed for 3 times, and the supernatant was removed by centrifugation at 6000 rpm for 20 min after each resuspension. The histone solubilization buffer (6 M urea, 20 mM Tris, 500 mM NaCl, pH 7.5) was used to finally dissolve the inclusion body, and the solution was centrifuged at 10000 rpm for 30 min to remove any precipitants. The collected supernatant ran through a 1 mL GE Healthcare Ni-NTA Fast Flow column, washed with 50 mL of histone wash buffer (6 M urea, 20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.5), and finally eluted by 10 mL histone elution buffer (6 M urea, 20 mM Tris, 500 mM NaCl, and pH 7.5). The eluate was collected, concentrated, and analyzed by SDS-PAGE followed by the Western blot with anti-H3 antibody (Figure 15A).

### 3.8.3 The conversion of AcK to Kme2 in H3

A similar reaction procedure of Staudinger reduction followed by the reductive amination was applied to them for the generation of Kme2 derivatives, respectively. 5 mM TCEP was first added to each protein solution for the reduction of the AcK incorporated for 2 h, followed by the addition of 100 mM dimethyllysine and 10 mM NaCNBH<sub>3</sub> to undergo reductive amination for 8 h. The final protein products were analyzed by SDS-PAGE with both pan anti-Kme2 antibody and then anti-H3 antibody after stripping-off process. The results showed that strong bands could be detected for all the three histone variants only after the reaction steps, while the histone protein amounts remained similar (Figure 15B). Additionally, we used western blot by a site-specific antibody for H3K4me2, or anti-H3K4me2 antibody, showing a band being detected for the histone product H3K4me2 (Figure 15C). The protein products were further applied to MS/MS analysis after the trypsinization into fragments, and the site-specificity of the reductive amination reaction was confirmed as expected (Figure 15D).



**Figure 15** The conversion of AcK to Kme2 in histone H3 (A) SDS-PAGE and Western blot for the expression of three H3 variants with AcK installed; (B) Western blot for the AcK containing H3 after reductive amination with dimethylamine; (C) Western blot of H3K4me2. Adapted with permission from [87].

#### 3.8.4 Substrate preparation for LSD1 assay

To synthesize H3K4me and H3K9me, reduced H3K4AcK and H3K9AcK was mixed with 100 mM methylamine and 20 mM NaCNBH<sub>3</sub> were added for reductive amination for 8 h. The protein products were analyzed by SDS-PAGE and western blot with pan anti-Kme2/Kme antibody and anti-H3 antibody. One more thing worth mentioning that H3K4me and H3K9me had also been made with the same approach as their dimethylated derivatives. However, the products were failing to be detected by the pan-anti-Kme antibodies from commercial resources (Result not shown). After the

removal of small molecular reactants, all the three dimethyl-H3 products were dialyzed to PBS pH=7 and the supernatant as stock solution will be used as substrates of LSD1 to test its selective demethylation activity.

### **3.9 The LSD1 assay over H3K4me2, H3K9me2 and H3K36me2**

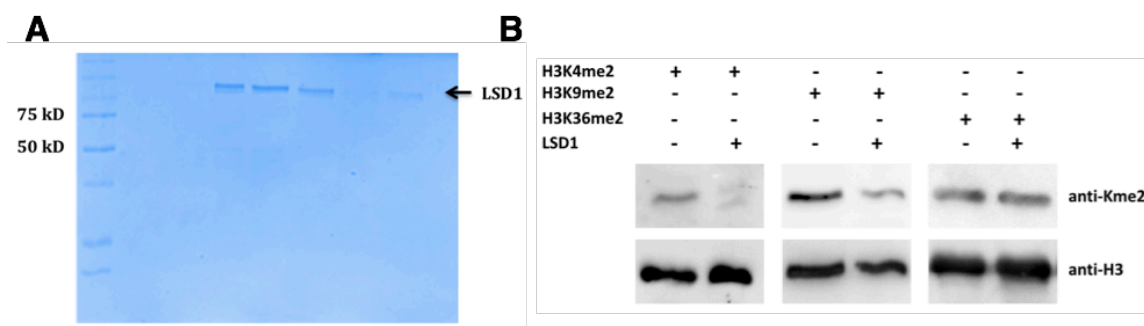
We got the pET-15b-LSD1 plasmid containing gene coding LSD1 with His<sub>6</sub>-tag from Professor Zigmund Luka at Vanderbilt University with the permission of using from Professor Yang Shi at Harvard University.

#### 3.9.1 The expression of LSD1

LSD1 expression and purification protocol was adopted from a previous publication with some modifications. The plasmid was used to transform *E. coli* BL21(DE3) cells for expression. The cells were grown in 10 ml of LB Medium with 100 µg/mL Amp for o/n, and then used to inoculate in 1 L 2YT medium with 100 µg/mL Amp for expression. When OD<sub>600</sub> reached to 1.4-1.6, 0.5 mM IPTG was added to induce the protein expression at 25 °C for 7 h. After collecting and washing the cells with cold 100 mM Tris Buffer pH 7.7, the pallets were stored at -80 °C for further purification.

Frozen cells were resuspended in 80 ml 100 mM Tris lysis buffer pH 7.5 that contained 25 mM MgCl<sub>2</sub>, 5 mM CaCl<sub>2</sub> 1 mM PMSF and 10 µg/mL DNase I. After the sonication, 14 mM β-mercaptoethanol was provided to the lysis buffer. Then, the cell debris was pelleted at 10000 rpm for 40 min. Ammonium sulfate was then added to the supernatant to precipitate the expressed LSD1 at 0 °C, and the fraction between 26% and

36% saturation was collected. To our experience, ammonium sulfate precipitation was preferred comparing with the Ni-NTA affinity purification, due to a relatively low yield of the Ni-NTA method. Then a 20 mM Tris (pH 8.1) buffer containing 50 mM NaCl, and 5 mM  $\beta$ -mercaptoethanol was used to dissolve the protein pellets, and the buffer was concentrated. During the redissolving and protein concentration process precipitation of LSD1 was observed. After redissolving the pellet, solution was dialyzed against the redissolving buffer to remove the remaining ammonium sulfate. FPLC purification was further carried out with Q Sepharose anion exchange column with a gradient from 20 mM Tris (pH 8.1) buffer A containing 50 mM NaCl and 5 mM  $\beta$ -mercaptoethanol to 20 mM Tris (pH 8.1) buffer B with 1 M NaCl and 5 mM  $\beta$ -mercaptoethanol. Fractions at around 20% of buffer B were collected and confirmed as purified LSD1. SDS-PAGE was used to analyze the purified enzyme (Figure 16A).



**Figure 16 The LSD1 enzymatic assay over histone with Kme2 installed.** (A) SDS-PAGE of LSD1 expression; (B) LSD1 assay for all three H3 variants. Adapted with permission from [87].

### 3.9.2 The setting up of LSD1 enzymatic assay

The setting-up of the enzymatic assay is following a reported literature, in which a 25  $\mu$ l reaction solution containing 10  $\mu$ M FAD and 5  $\mu$ M LSD1 was added to dimethyl-

H3 products and incubated for 4 h, respectively. The control reaction was set up without the addition of LSD1 but with the same amount of H3 and FAD. Both assays were analyzed by SDS-PAGE followed by the western blotting with pan-anti-Kme2 antibody. After the stripping-off process, anti-H3 was also applied to confirm the similar amounts with or without the enzymatic reaction.

To summarize, the result showed LSD1 could remove dimethylation at H3K4 completely and H3K9 mostly but stayed completely inert toward dimethylation at H3K36 (Figure 16B).

### **3.10 The expression of p53 and preparation of methylated p53**

#### 3.10.1 Background and general design

In the previous parts, we have showed the accessibility of our method to small insoluble proteins as histone from inclusion body, and medium-sized soluble proteins as sfGFP. To push the application of our strategy one step more, we further chose p53 as an example of relatively large soluble proteins with multiple cysteine residues to test its reliability of mono- or di-methylation installation at specific sites. As a tumor antigen transcription factor with 393 amino acids, the mutations on human p53 are associated with about 60% of known tumors. It has been reported that p53 undergoes methylation at its K372 site, which can activate its transcription activity<sup>[91]</sup>. Previous research also speculated this modification would recruit another important modification enzyme Tip60, but no more evidence has been shown. Tip60 is a histone acetyl-transferase with a chromodomain to bind methyllysine, and it can acetylate p53 at the position of K120<sup>[92]</sup>.

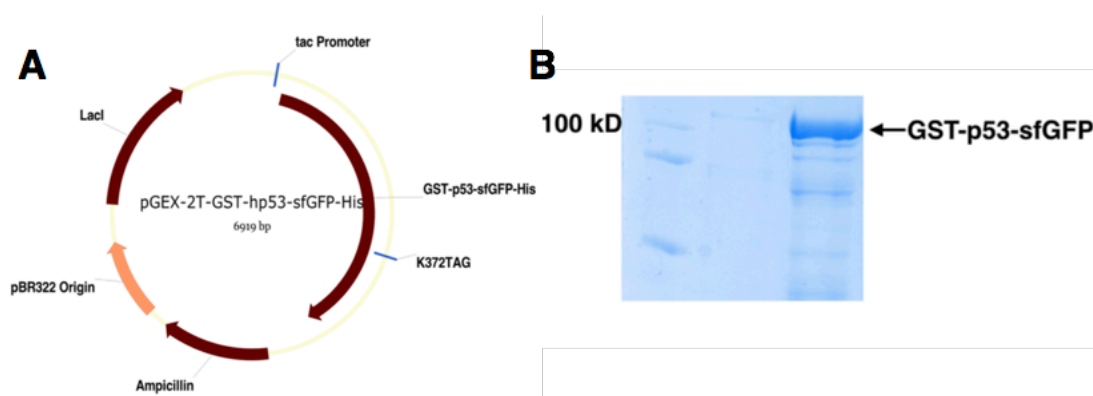
One of the possible reason is there is no easy way to get a special variant of hp53 with Kme or Kme2 at the desire site. Therefore, we wanted to further confirm this suspicion via our method to expressed p53 proteins with lysine mono- or di-methylation installed.

### 3.10.2 The expression of hp53-K372AcDK

To begin with, p53 with AcDK at K372 site was expressed in *E. coli* with a similar protocol as sfGFP-D134AcDK. We started with the trial of p53 expression in *E. coli*<sup>[93]</sup> with the plasmid pGEX-2T-GST-hp53 as a gift from Dr. Wei Gu's group in Columbia Unviersity, Due to a very low yield, we decided to fuse a sfGFP protein with a C-terminal His<sub>6</sub>-tag to the C-terminus of hp53, as well as a GST tag at its N-terminus to optimize the expression and purification. The PCR amplification of the pGEX-2T-GST-hp53backbone was carried out by using two primers (5'-GCC CAA GCT TCC GGG AAT TCA TCG TGA CTG AC-3' and 5'-ACT CGA GCT CGT CTG AGT CAG GCC CTT CTG TC-3') by Phusion DNA polymerase, and the sfGFP with His<sub>6</sub>-tag was amplified from pBAD-sfGFP as a gift from Dr. Ryan Mehl's group in Oregon State University by using two primers (5'-ACG CGA GCT CGA GAA CCT GTA TTT CCA AGG TGC CAT G-3' and 5'- CCC AAG CTT TTA ATG GTG ATG ATG ATG GTG GCT GCC T-3'). Both products after PCR were digested by enzyme combination of HindIII and SacI, and ligated together with T4 ligase according to a standard protocol to finally generate plasmid of pGEX-2T-GST-hp53-sfGFP-His (Figure 17A). After its sequence confirmation, site-directed mutagenesis was then carried out to introduce an amber mutation at the K372 site of p53. Again, PCR with Phusion DNA polymerase was carried out by using two primers



(5'-TAG AAG GGT CAG TCT ACC TCC CGC C-3' and 5'-GGA CTT CAG GTG GCT GGA GTG AG-3'), whose mutated product was phosphorylated and then ligated. The finally constructed plasmid pGEX-2T-GST- hp53(K372TAG)-sfGFP-His was sent for sequencing to confirm before its expression. Then, pGEX-2T-GST-hp53(K372TAG)-sfGFP-His, together with pEVOL-AcdKRS, is used to cotransform E. coli BL21(DE3) cells to express the p53-K372AcdK. The transformed cells were placed to plate for o/n, and single colony was picked and let to grow in the 5 mL 2YT medium. The seed culture was then used to inoculate 500 mL of 2YT media. 0.2% arabinose, 1 mM IPTG and 2 mM AcdK were added to induce the expression of p53-K372AcdK after OD reached 0.8. The expressed protein p53-K372AcdK was purified by Ni-NTA followed by the glutathione affinity resin and confirmed by SDS-PAGE (Figure 17B).



**Figure 17 The expression of hp53 with AcdK incorporated.** (A) the plasmid for the expression of hp53; (B) SDS-PAGE for the GST-p53-sfGFP expression.

### 3.10.3 The conversion of hp53-K372AcdK to Kme2

After we get purified hp53-K372AcdK at hand, we then carried out the Staudinger reduction followed by the reductive amination to generate desired p53 with mono- or

dimethylation at K372 site (p53-K372me and p53-K372me<sub>2</sub>). The purified protein was dialyzed to PBS buffer pH 7 to remove any small molecules such as glutathione or imidazole. After been incubated with 5 mM TCEP for 2 h, the reductive amination was carried out by adding 100 mM monomethylamine or dimethylamine then NaCNBH<sub>3</sub> and let to stay at 37 °C for 8 h to generate p53-K372me or p53-K372me<sub>2</sub>, respectively. The formation of p53-K372me and p53- K372me<sub>2</sub> were both similarly confirmed by western blot analysis with the anti-p53-K372me antibody and pan-anti-Kme<sub>2</sub> antibody, respectively, while the WT p53 was also used as a control with no band detected (Results combined in Figure 18).

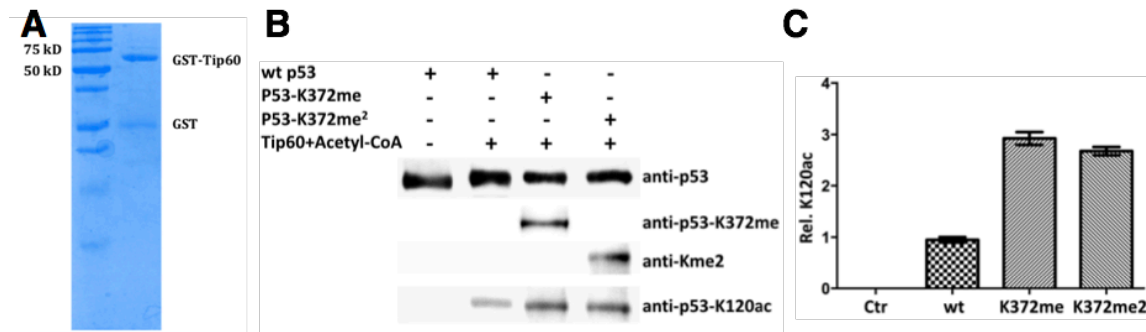
### **3.11 The expression of Tip60 and enzymatic assay**

We got the plasmid pCIN4-FH-TIP60 containing the gene for Tip60 fused with GST at its N-terminus from Dr. Wei Gu's group at Columbia University.

#### **3.11.1 The expression of Tip60**

It was used to transform *E. coli* BL21(DE3) cells. A single colony was picked for the setting up of 10 mL seed culture, and the seed culture was used to inoculate 1 L LB medium with 100 µg/mL Amp for expression. The transformed cells were grown at 37 °C in the LB medium to OD<sub>600</sub> around 0.4-0.6. The temperature was changed to 16 °C at that moment, and the expression of GST-Tip60 was induced with the addition of 0.7 mM IPTG. Next, cells were let to grow continuously at 16 °C for 12 h. After the expression, cells were collected and washed with 100 mM Tris lysis buffer pH 7.5 containing 1 mM PMSF

and 10  $\mu\text{g/mL}$  DNase I. and sonicated. The cell pallet was stored at  $-80\text{ }^{\circ}\text{C}$  for purification. 30 mL 100 mM Tris lysis buffer pH 7.5 with 1 mM PMSF and 10  $\mu\text{g/mL}$  DNase I was used to resuspend the cell pallet. The expressed GST-Tip60 was enriched in glutathione resins and eluted by 10 mM glutathione in 50 mM Tris buffer pH 8. The purified protein was detected by SDS-PAGE (Figure 18A).



**Figure 18 The expression of Tip60 and its enzymatic assay.** (A) SDS-PAGE for the expression of GST-Tip60; (B) Tip60 assay for the cross-talk between p53-K372 methylation and acetylation; (C) quantification of cross-talk between p53-K372 methylation and acetylation. Adapted with permission from [87].

### 3.11.2 The setting up of Tip60 enzymatic assay

Then, we started to test the Tip60 enzymatic assay on all three p53 protein variants with or without methylation. Each of the proteins was incubated with Tip60 with the existence of acetyl-CoA, followed by the Western blot analysis by an anti-p53-K120ac antibody. The result showed that Tip60 catalyzed the acetylation at K120 site for all three p53 protein variants, but the acetylation level was significantly enhanced by K372 methylation (Figure 18B). Quantitative analysis showed both mono- and dimethylation at K372 could increase the acetylation at K120 site by 2.5 to 3 folds, and this result provides

a first evidence proofing the direct cross-talk of p53 K372 methylation to boost the acetylation of p53 K120 by Tip60 (Figure 18C).

### 3.12 Summary

To summary, we have demonstrated a novel synthetic method with the combination of noncanonical amino acid incorporation via amber suppression with a two-step bioorthogonal reaction for the expedient preparation of proteins with site-specific lysine dimethylation incorporated. Both of the reactions, Staudinger reduction and reductive amination, are with site-specificity, convenience, and good bio-compatibility, and thus this strategy is applicable for the synthesis of proteins with site-specific lysine mono- or di-methylation for further biological studies. Besides two types of methylation mentioned above, this approach is theoretically applicable for any lysine alkylation PTMs, such as the rare hypusine found in eukaryotic translation initiation factor protein 5A (eIF5A)<sup>[94]</sup> and carboxymethyl-lysine (CML) during the metabolic of glycoproteins<sup>[95]</sup>.

Since there are only three levels of Lys methylation, it provides less challenging for the chemical biologist comparing with Lys acylation discussed in the next part. We have showed the accessibility of our method to Kme1 and Kme2, notwithstanding, the synthetic study of tri-methyllysine from protein chemistry or amber codon suppression point of view is still in abysmal ignorance. Several known examples of protein preparation with Lys tri-methylation are from chemical total synthesis. Our AcdK system is not applicable to Kme3, because allysine cannot undergo reductive amination with

trimethylamine. To develop synthetic approaches for Kme3 can advance our understanding with no doubt.

One more point worth mentioning is that the intermediate allysine itself is the first amino acid example containing an aliphatic aldehyde group incorporated into proteins. Due to its natural application as a crosslinking reagent in both elastin and collagen<sup>[88]</sup>, it can be a good electrophilic center for other protein functionalization for further protein chemical biology studies, such as labeling, bioconjugation, solidification, etc.

## CHAPTER IV

### THE SYNTHETIC INVESTIGATION OF LYSINE ACYLATION

#### 4.1 Research background

Among all the proteinogenic amino acids, Lys is the only one carrying a side chain amino group. Due to the wide reactivity of aliphatic amine, large numbers of PTMs existing on Lys have been discovered, dramatically outnumbering modifications of other amino acids<sup>[17]</sup>. Acylation as another main group of Lys PTMs has been one of the research focuses since the discovery of acetylation in histones. Besides acetylation, two other major acylation modifications are biotinylation serving as a cofactor for enzymes, and lipoylation for membrane protein or secreted protein anchoring on membrane.

With the recent development of MS/MS and proteomics, an increasing diversity and amount of novel Lys PTMs have been found widespread on cytosolic proteins, transcription factors and membrane proteins<sup>[96]</sup>. In order to shed light on the identification and functional study of these PTM associated enzymes such as “writer”, “reader” or “eraser”, one of the prerequisites is to prepare sufficient amount of native proteins with Lys acylation modifications installed at specific sites for further functional studies. Similar to Lys methylation, the direct extraction and traditionally recombinant protein expression can hardly help with modified POIs. Researchers started with the enzymatic synthesis approach<sup>[5]</sup>, nonetheless, the enzymes for most of the novel modifications are unexplored and promiscuous.

Many chemical biology strategies have been developed, whose pros and cons have been discussed in part 2, and here comes an abridge edition of introduction. NCL-based

protein total synthesis and EPL-based protein semi-synthesis can theoretically generate native proteins with natural or non-natural structures. However, chemical synthesis is generally limited to 200 AAs, and an HPLC purification step is needed that can cause protein to denature. In addition, Cys alkylation and Dha based bioorthogonal reactions can generate Lys acylation with *S*-linked or *C*-linked mimics<sup>[6]</sup>. However, the products cannot perform the same comparing with the native ones.

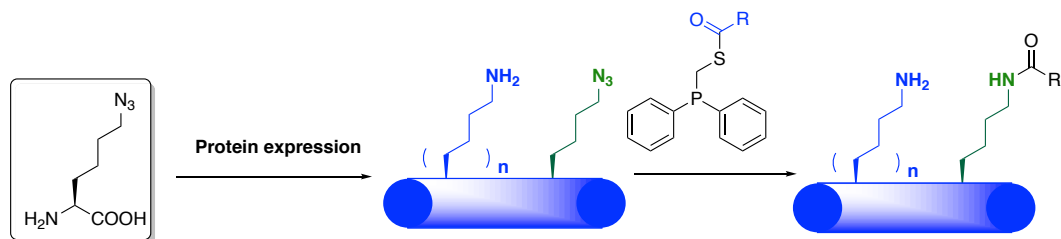
Another approach is to apply amber codon suppression for the incorporation of proteins with Lys acylations. Initiated by the Chin group, many simple PTMs have been successfully incorporated mainly based on the pyrrolysine system, such as Kac, Kpr, Kbu, Kcr, and Khib<sup>[8]</sup>. Although straightforward and convenient, the selection of a PylRS mutant necessary to recognize the particular modified Lys can be tedious and does not guarantee to be successful. For instance, Lys with charge-containing side chain such as succinylation, or large acylation modifications such as lipidation or Ubiquitination is not accessible for the direct incorporation strategy. For instance, all our trials for the incorporation of malonyllysine, succinyllysine, and glutaryllysine have turned out to be failure in the direct selection of PylRS mutants. And therefore, a more robust and versatile strategy is needed for the installation of Lys acylation modifications.

## **4.2 The design and synthesis of AznL**

### 4.2.1 The design of AznL

To solve this problem, we designed a novel strategy by coupling the amber-suppression-based mutagenesis approach with the traceless Staudinger ligation reaction.

Our idea stems from the conversion of azide group to different acylation modifications via the traceless Staudinger reaction developed by Bertozzi<sup>[97]</sup> and modified by Raines<sup>[98]</sup>. As a typical bioorthogonal reaction, traceless Staudinger reaction is capable to protein modifications. We then proposed that AznL installed in proteins could undergo a multi-step reaction with phosphinothioester to finally generate different acylated Lys with their native structures. One of the key points for this strategy is that by simply changing the phosphinothioester reagents, we can theoretically prepare any types of Lys acylation PTMs. We decided to use two examples to show the application of our strategy, the acetylation and succinylation<sup>[99]</sup>. Because acetylation is the most widely existing PTM, and succinylation is a newly discovered PTM (Figure 19).



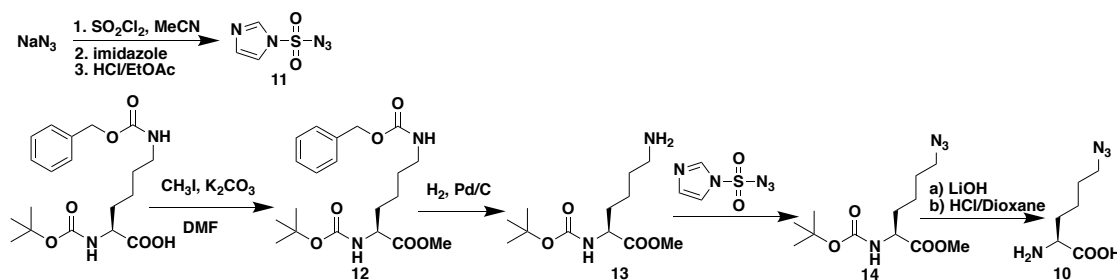
**Figure 19 The conversion of AznL to Lys acylation via traceless Staudinger ligation**

Interestingly, AznL was not a novel ncAA been incorporated into proteins: previous reports have shown the installation of AznL at the methionine positions via the residue-replacement strategy<sup>[100]</sup>. By using this strategy, AznL has been incorporated for the study of nascent cellular proteomes. However, this strategy fails to meet our requirement of site-specific acyllysine incorporation into specific POIs, due to the global substitution of Met among the entire proteome. Therefore, we must apply the amber-suppression mutagenesis method over PylRS system for the AznL incorporation.



#### 4.2.2 The synthesis of AznL

We initialized our project with the synthesis of AznL (10) from a recently reported approach. The synthetic procedure of AznL is a modified version from reference<sup>[101]</sup>, which contains four steps.



**Figure 20** Scheme for the synthesis of AznL

**The synthesis of imidazole-1-sulfonyl azide (11):** To the solution of sodium azide (1.31 g, 20.2 mmol) in anhydrous acetonitrile (30 mL) was added sulfonyl chloride solution (97% solution, 1.7 mL, 20.4 mmol) dropwise in an ice bath. The mixture was then stirred at r.t. for 8 h. Imidazole (2.78 g, 40.9 mmol) was added into the mixture in an ice bath and the afforded solution was stirred at r.t. for another 8 h. The mixture of white slurry was partitioned with EtOAc/H<sub>2</sub>O solution and the organic layer was collected and washed with H<sub>2</sub>O, saturated aqueous NaHCO<sub>3</sub>, and brine, dried over anhydrous sodium sulfate, and filtered. Hydrogen chloride solution (1 M in EtOAc, 12 mL, 12 mmol) was added into the filtrate in an ice bath. The afforded white suspension was collected by filtration. The filter cake was washed with cold ethanol and EtOAc to afford imidazole-1-sulfonyl azide hydrochloride (11) as the desired product (2.2 g, 52%) in white solid. <sup>1</sup>H NMR (D<sub>2</sub>O, 300

MHz)  $\delta$  8.88 (br, 1H), 7.87 (br, 1H), 7.44 (br, 1H). MS (Free base + H<sup>+</sup>): Calc. 174.0, Obs. 174.0362.

**The synthesis of compound methyl *N*<sup>6</sup>-((benzyloxy)carbonyl)-*N*<sup>2</sup>-(tert-butoxycarbonyl)-*L*-lysinate (12):** 9.51 g of Boc-Lys(Cbz)-OH (25 mmol, Sigma-Aldrich) was dissolved in dry DMF, followed by the addition of K<sub>2</sub>CO<sub>3</sub> (7.26 g, 52.5 mmol). 3.44 mL of MeI (55 mmol) was gradually added and was let to stir overnight at 80 °C. Then the reaction mixture was quenched by the addition of 50 mL H<sub>2</sub>O and extracted by EtOAc 30 mL for 3 times. The EtOAc layers were combined, washed with brine, and dried by anhydrous Na<sub>2</sub>SO<sub>4</sub>. The crude product was then purified by the silica gel flash column chromatography (eluted at 10% EtOAc/hexane) to give an oil product (12) (9.6 g, 97%).

**The synthesis of compound methyl (tert-butoxycarbonyl)-*L*-lysinate (13):** The solution of Boc-Lys(Cbz)-OMe (12) (3.9 g, 10 mmol) in methanol (100 mL) was added palladium on activated carbon (Pd 10%, 0.6 g, 0.6 mmol). The mixture was stirred at r.t. with hydrogen bubbled through where the reaction progress was monitored by the TLC analysis. The reaction was terminated as the starting protected amino acid on TLC disappeared. Then the reaction mixture was filtered through a celite cake packed on a Buchner funnel and the flowed-through solution was concentrated under reduced pressure to afford colorless oil that was characterized as the desired product (13). The product was directly used in the next step without further purification.

**The synthesis of compound (14):** Boc-Lys-OMe (13) (0.81 g, 3.1 mmol) from the previous step was dissolved in methanol (20 mL). CuSO<sub>4</sub> 5H<sub>2</sub>O (6.8 mg, 0.45 mmol) and K<sub>2</sub>CO<sub>3</sub> (888 mg, 6.4 mmol) were added into the solution followed by the addition of

imidazole-1-sulfonyl azide hydrochloride (11) (650 mg, 3.1 mmol) into the mixture in an ice bath. The resulting mixture was stirred at r.t. for 8 h. After addition of EtOAc (10 mL), the diluted mixture was then washed with water, saturated aqueous NaHCO<sub>3</sub>, brine, dried over anhydrous sodium sulfate, filtered, and concentrated under reduced pressure. The crude product was further purified by the silica gel flash column chromatography (eluted at 5% EtOAc/hexane) to give a clear oil as the desired product (14) (0.7 g, 79%). <sup>1</sup>HNMR (CDCl<sub>3</sub>, 300 MHz). δ 5.09 (d, 1H, J = 1.6 Hz), 4.33 (dd, 1H, J = 1.6, 1.8 Hz), 3.77 (s, 3H), 3.24 (t, 2H, J = 1.2 Hz), 1.99-1.78 (m, 2H), 1.78-1.58 (m, 2H), 1.58-1.32 (m, 12H).

**The synthesis of AznL (10):** The solution of ester from the previous step (14) (0.7 g, 2.4 mmol) in a 1:1 mixture of methanol and THF (24 mL) was added LiOH aqueous solution (0.5 M, 24 mL) and stirred at r.t. The reaction progress was monitored by the TLC analysis. After 5 h, the spot representing the starting ester on TLC plate disappeared. The reaction mixture was then diluted with EtOAc/H<sub>2</sub>O. After partition, the organic layer was discarded, and the aqueous layer was acidified by the addition of 3M HCl until pH was adjusted to 3-4. The mixture was then extracted with EtOAc twice. The organic layers were combined, washed with brine, dried over anhydrous Na<sub>2</sub>SO<sub>4</sub>, and concentrated to afford clear sticky oil as the desired product which was subjected directly to next step without chromatography purification.

The afforded acid from the previous step (0.8 g, 2.9 mmol) in dioxane (5 mL) was added into a HCl solution (4 M in dioxane, 10 mL). The mixture was stirred at r.t. The reaction progress was monitored by TLC. After reaction for 3 h, the HCl solution were removed under reduced pressure. The afforded oil was dissolved in minimal amount of

water, neutralized with aqueous sodium hydroxide and subjected to the ion exchange chromatography (Dowex 50WX4) to afford a pale white solid as the desired product AznL (10) (150 mg, 36% for two steps).  $^1\text{H}$  NMR ( $\text{D}_2\text{O}$ , 300 MHz).  $\delta$  4.17 (t, 1H,  $J = 1.4$  Hz), 3.35 (t, 2H,  $J = 1.2$  Hz), 2.10-1.79 (m, 2H), 1.78-1.30 (m, 4H). MS ( $\text{AznL} + \text{H}^+$ ): Calc. 173.1, Obs. 173.1040. (Figure 21)

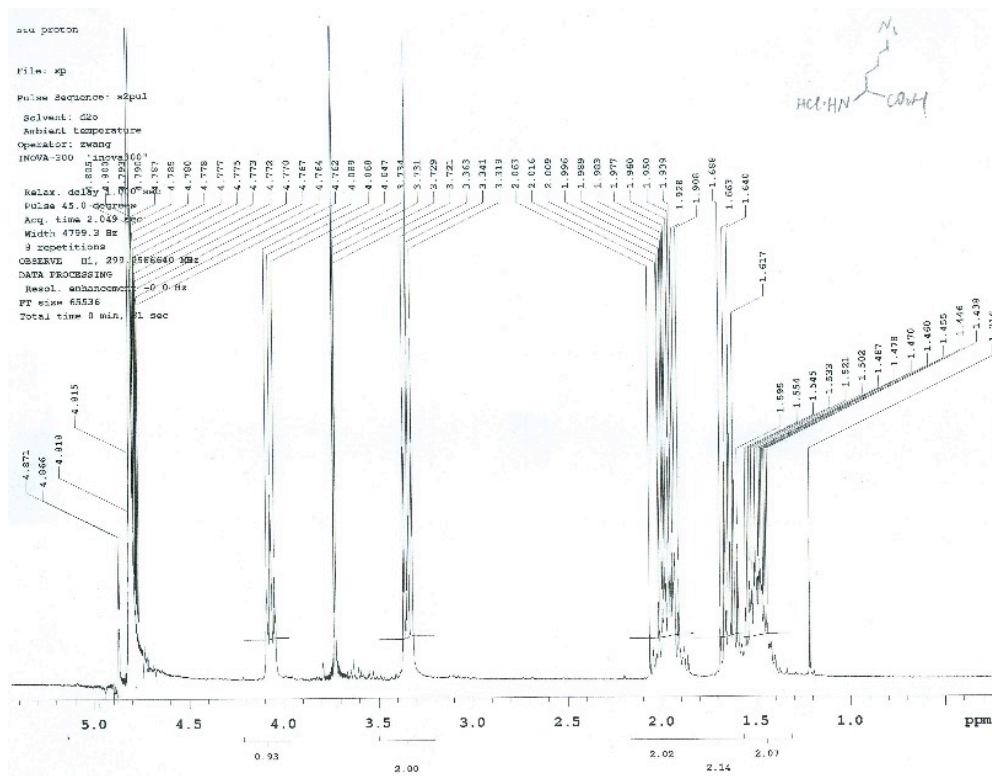


Figure 21  $^1\text{H}$ NMR spectrum of AznL

### 4.3 The selection and screening of AznL

#### 4.3.1 The selection of AznL

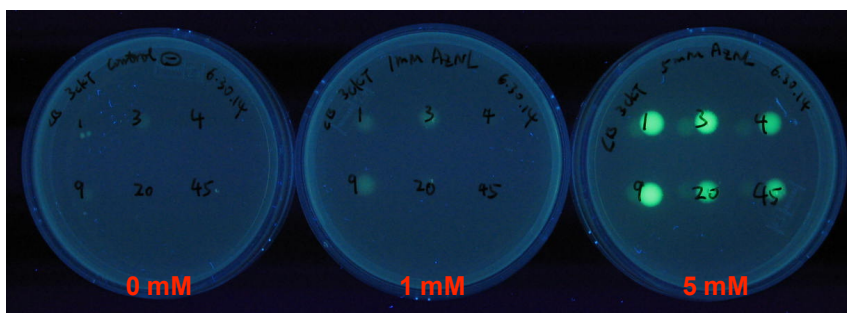
With enough amount of AznL at hand, we started to identify a specific PylRS mutant for AznL. We use the same approach for the generation of a PylRS mutant library

with randomization at five active site residues, Y306, L309, C348, Y384, and W411 as for AcdK. The DNA library was then cloned into a pBK plasmid.

A double-sieve selection was carried out with an alternative combination of three rounds of positive and two rounds of negative selections, with similar procedure as described for AcdK. Final positive selected colonies grew on LB plates with 102  $\mu\text{g/mL}$  Cm, 25  $\mu\text{g/mL}$  Kan, 12  $\mu\text{g/mL}$  Tet, and with or without 1 mM AznL to get several mutant candidates.

#### 4.3.2 The selection result

After the selection processes, we again carried out the screening by the visualization of GFP<sub>UV</sub> expression, similar to the AcdK screening. The clones picked after the third positive selection was let to grow on LB plate with 102  $\mu\text{g/mL}$  Cm, 25  $\mu\text{g/mL}$  Kan, and 12  $\mu\text{g/mL}$  Tet but with or without 1 mM AznL. By measuring the fluorescent intensity of GFP<sub>UV</sub> expression, we can select a best colony from 6 candidates after the third positive selection.



**Figure 22 Screening of the clones after the third positive selection on LB plates.** The expression of GFP can be seen on the plate with either 1 mM or 5 mM AznL but not the one without.

From Figure 22, the mutant with the best amber suppression efficiency in *E. coli* was identified as Y306L/C348I/Y384F. The gene mutant corresponding to this PylRS mutant noted as AznLRS together with the tRNA<sup>Pyl</sup> was incorporated into pEVOL plasmid as pEVOL-AznLRS-PylT and was then used to transform the *E. coli* BL21(DE3) cells for the expression of sfGFP with an amber codon at the D134 site.

**Table 1 The mutation sites for the selected PylRS candidates**

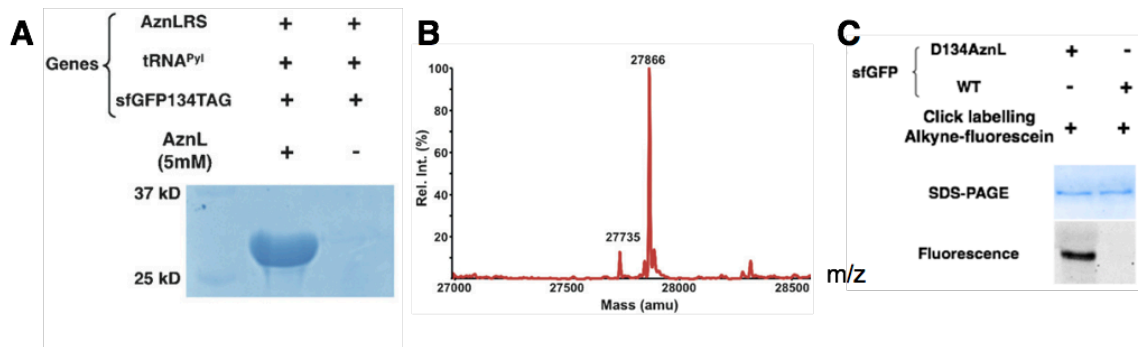
MUTANT NO.	LABEL	MUTATION SITES
1	AznLRS1	306L309L348I384F411W
2	AznLRS3	306F309M348M384F411W
3	AznLRS4	306L309L348V384F411F
4	AznLRS9	306L309L348L384F411W
5	AznLRS20	306L309L348I384F411W
6	AznLRS45	306L309L348I384F411W

#### 4.4 The expression and purification of sfGFP-D134AznL

##### 4.4.1 The incorporation of AznL

The selected plasmid pEVOL-AznLRS-PylT together with pBAD-sfGFP-D134TAG was used to transform *E. coli* BL21(DE3) cells. After picking up the single colony for seed culture, cells were grown in the 250 mL LB medium. The expression was induced with the addition of with 1 mM IPTG, 0.2 % arabinose, and 5 mM AznL when OD<sub>600</sub> reached 0.6. After another 8 h growing, the cells were collected with centrifuging

(4k 20 min, 4 °C), washed, and fully resuspended with a lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 7.5). The cells were sonicated under ice/water bath for 3 times (4 min each time with an interval of 4 min). Then the cell lysate was centrifuged (10k 40 min, 4 °C), and the precipitate was removed. The cell lysate was poured into 1 mL Ni Sepharose™ 6 Fast Flow column (GE Healthcare) column, and the flow-through was removed. 30 mL of wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 7.5) was used to wash the resin. 12 mL elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 7.5) was used to elute the protein. All the eluted fractions were collected and concentrated by Amicon Ultra-15 Centrifugal Filter Device (10k MWCO, Millipore), and buffer was exchanged to 20 mM phosphate buffer pH 8.5. Fast protein liquid chromatography (FPLC) Q-Sepharose anion exchange column was used to further purify the sfGFP-D134AznL. Buffer A is 20 mM phosphate pH 8.5 buffer, while buffer B 20 mM phosphate pH 8.5 buffer containing 1 M NaCl. The wash with 10 column volume (CV) of mobile phase containing 5 % B was followed by the elution of a gradient from 5 % to 50 % B in 10 CV mobile phase. The elution samples were collected and again concentrated by Amicon Ultra-15 Centrifugal Filter Device (10k MWCO, Millipore). The purified protein was confirmed by SDS-PAGE (Figure 23A). The buffer was further exchanged to 200 mM phosphate pH 6.5 buffer for the further Staudinger reaction. Then the expression yield of full-length sfGFP-D134AznL was about 20 mg/L. The expression of WT sfGFP followed the similar procedure without the addition of AznL for induction, which will be further used as negative control for the reaction.



**Figure 23 The expression of sfGFP-D134AznL** (A) Coomassie staining of SDS-PAGE analysis for sfGFP-D134AznL; (B) ESI-MS of sfGFP-D134AznL; (C) fluorescence labeling of sfGFP-D134AznL with azido-dye. Adapted with permission from [99].

The purified sfGFP-D134AznL was analyzed by ESI-MS showing one major peak at 27866 Da and one minor peak at 27735 Da, which agrees with the theoretical molecular weights of full-length sfGFP with AnzL installed at its 134 position and its N-terminal methionine cleavage product (Figure 23B). Notably, some side peaks in Figure 23B are representing some impurities not fully known.

#### 4.4.2 The fluorescence labeling of sfGFP-D134AznL

The existence of the azide functionality in sfGFP was further confirmed by its selective labeling with an alkyne-fluorescein dye. The dye was synthesized and donated by Erol Can Vatansever from my lab. The fluorescent labeling of sfGFP-D134AznL was carried out according to standard CuAAC procedure<sup>[102]</sup>. 50  $\mu$ M of CuSO<sub>4</sub> was mixed with 300  $\mu$ M of BTAA, followed by the addition of 500  $\mu$ M alkyne-fluorescein, and 15  $\mu$ M of sfGFP-D134AznL (all the mentioned concentrations were final concentrations). The reaction was initialized by the addition of ascorbate stock solution in PBS to a final



concentration of 2.5 mM. The same reaction was carried out for wild type sfGFP. The reaction was allowed to be incubated at 25 °C for 4 h followed by the SDS-PAGE analysis. After the fluorescence detection, the gel was then stained by Coomassie blue (Figure 23C). These combined results validated the specificity of AznLRS toward AznL and the selective incorporation of AznL at the amber codon site.

#### **4.5 The expression and purification of Ub-K48AznL, Ub-K48ac, and WT-Ub**

##### 4.5.1 The DNA and protein sequence of Ub

*DNA sequence: Ub-K48TAG-His6*

```
atgcaaatattcgtgaaaaccctaactggtaagaccatcactctcgaagtggagccgagtgacaccattgagaatgtcaaggca  
aagatccaagacaaggaaggcatcctcctgaccagcagaggttgatctttgctgggtagcagctggaagatggacgcacct  
gtctgactacaacatccagaaagagtccaccctgacttggccttaggctgagaggaggacatcacatcacatcactaa
```

*Protein sequence: Ub-K48AznL-His6*

```
MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAG(AznI)QLEDG  
RTLSD YNIQKESTLHLVLRRLRGGHHHHHH
```

##### 4.5.2 The expression of Ub-K48AznL

Although we first tested our AznL installation into sfGFP, we decided to change to ubiquitin (Ub) as model proteins for three purposes. Ub as a smaller protein with only 76 AAs is easier to characterize with MALDI-TOF instead of ESI. The wider solubility of

Ub makes it possible for higher percentage of organic solvent during chemical conversions. Last but not the least, sfGFP is supposed to be one of the most stable proteins in nature; notwithstanding, we sometimes observe an unrecognizable chopping-off of a peptide after long time storage with a remaining protein of 26 kDa molecular weight.

The plasmid of pEVOL-AznLRS-PylT mentioned before was used to co-transfer *E. coli* BL21(DE3) cells together with pET-Ub-K48TAG. The cells were placed on LB agar plate containing 100 µg/mL Amp and 34 µg/mL Cm. A single colony was picked, followed by the growth in a 5 mL LB medium culture for o/n. The overnight culture was used to inoculate 250 mL 2YT medium with the same concentration of Amp and Cm. Cells grew in a 37 °C shaker for 3 h until OD<sub>600</sub> reached 0.6. The protein expression was induced by the addition of 1 mM IPTG, 0.2% arabinose, and 5 mM Aznl. After another 8 h incubation, the cells were collected (4k 20 min, 4 °C), washed, and fully resuspended with PBS. The sonication of cells underwent under ice/water bath for 3 times (4 min each time with an interval of 4 min). Then the cell lysate was centrifuged (10k 40 min, 4 °C), and the precipitate was removed. The supernatant was adjusted pH to 3.5 with acetic acid and was incubated for 20 min before centrifuging (10k 40 min, 4 °C). After the removal of precipitation, 1 mL Ni Sepharose™ 6 Fast Flow column (GE Healthcare) was added to the supernatant, and the mixture was loaded to a column to remove the flow-through. 50 mL of wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 7.5) was used to wash the resin in three batches. Around 6 mL of elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 7.5) was used to elute the target protein. All the eluted fractions were collected and concentrated by Amicon Ultra-15 Centrifugal

Filter Device (3.5k MWCO, Millipore), and buffer was exchanged to 200 mM phosphate buffer pH 6.5 for the further traceless Staudinger ligation.

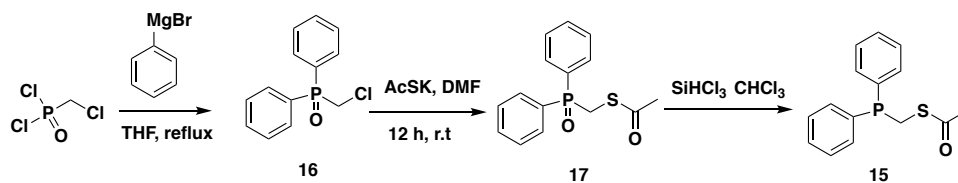
#### 4.5.3 The expression of WT-Ub and Ub-K48AcK

WT Ub with a N-terminal 6×his tag was also expressed in a 1 L 2YT culture with the addition of 100 µg/mL Amp. Cells were let to grow in a 37 °C shaker (250 r.p.m.) for 3 h until OD<sub>600</sub> reached 0.6. The protein expression was induced by the addition of 0.8 mM IPTG as the final concentration and let to incubate for another 8 h. Next, cells were collected, sonicated and the protein was purified with a similar procedure as the purification of Ub-K48AznL. For the expression of Ub-K48ac, the pEVOL-AcKRS-PylT was used instead, while 5 mM of AcK was provided while induction. All the other expression and purification steps remain the same as Ub-K48AznL.

### 4.6 The synthesis of (diphenylphosphino)methanethiol thioesters

We started with the synthesis of both (diphenylphosphino)methanethiol thioesters. The procedure design and synthesis were carried out by Dr. Yadagiri Kurra.

#### 4.6.1 The synthesis of (diphenylphosphino)methanethiol acetate (dPPMT-Ac)



**Figure 24** Scheme for the synthesis of dPPMT-Ac

**The synthesis of Compound (chloromethyl)diphenylphosphine oxide (16):**

Chloromethylphosphonic dichloride (2 g, 12 mmol) was dissolved in freshly distilled THF (30 mL). A solution of phenylmagnesium bromide (1.0 M) in THF (24 mL, 24 mmol) was added dropwise over 30 min. The resulting mixture was stirred at reflux for 24 h. The reaction was then quenched by the addition of water (20 mL), and solvent was removed under reduced pressure. The residue was taken up in CH<sub>2</sub>Cl<sub>2</sub> and washed once with water (50 mL) and once with brine (50 mL). The organic layer was dried over anhydrous MgSO<sub>4</sub> and filtered, and solvent was removed under reduced pressure. The residue was purified by flash chromatography (silica gel, 3% MeOH in CH<sub>2</sub>Cl<sub>2</sub>). Phosphine oxide (**16**) was isolated as a white solid in 63% yield. <sup>1</sup>H NMR (CDCl<sub>3</sub>, 300 MHz) δ 7.84-7.79 (m, 4H), 7.60-7.59 (m, 2H), 7.55-7.49 (m, 4H), 4.05 (d, 2H, J = 4.2 Hz). <sup>31</sup>P NMR (CDCl<sub>3</sub>, 200 MHz) δ 28.3.

**The synthesis of Compound S-((diphenylphosphoryl)methyl) ethanethioate (17).**

Phosphine oxide **16** (1.2 g, 4.8 mmol) was dissolved in DMF (20 mL). Potassium thioacetate (0.829 g, 5.76 mmol) was then added, and the reaction mixture was stirred under Ar(g) for 18 h. The solvent was then removed under reduced pressure. The resulting oil was purified by chromatography (3% v/v MeOH in CH<sub>2</sub>Cl<sub>2</sub>). Phosphine oxide **15** was isolated as a clear, colorless oil in 70 % yield. <sup>1</sup>H NMR (CDCl<sub>3</sub>, 300 MHz) δ 7.78-7.74 (m, 4H), 7.54-7.53(m, 2H), 7.49-7.45 (m, 4H), 3.77 (d, 2H, J = 4.1 Hz), 2.25 (s, 3H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 75MHz) δ 193 (d), 132.3, 132.3, 131.5, 131.1, 131.0, 130.7, 128.7, 128.6, 30.0, 27.6 (d). <sup>31</sup>P NMR (CDCl<sub>3</sub>, 200 MHz) δ 28.9.

**The synthesis of Compound *S*-((diphenylphosphaneyl)methyl) ethanethioate dPPMT-Ac (15).** Phosphine oxide **17** (1.06 g, 4 mmol) was dissolved in anhydrous chloroform (10 mL). Trichlorosilane (8 mL, 90 mmol) was added, and the resulting solution was stirred under Ar(g) for 72 h. The solvent was then removed under reduced pressure. (CAUTION: Excess trichlorosilane in the removed solvent was quenched by the slow addition of saturated sodium bicarbonate in a well-ventilated hood.) The residue was purified by flash chromatography (silica gel, 3 % v/v MeOH in CH<sub>2</sub>Cl<sub>2</sub>). Phosphine **7** was isolated as a white solid in 65%% yield. <sup>1</sup>HNMR (CDCl<sub>3</sub>, 300 MHz) δ 7.45-7.41 (m, 3H), 7.40-7.33(m, 7H), 3.52 (t, 2H, J = 1.8, 3.3 Hz), 2.29 (s, 3H). <sup>13</sup>CNMR (CDCl<sub>3</sub>, 75MHz) δ 193, 136.8, 136.6, 132.8, 132.6, 131.0, 129.1, 128.6, 128.5, 30.2, 25.9 (d). <sup>31</sup>PNMR (CDCl<sub>3</sub>, 200 MHz) δ -15.33. All the <sup>1</sup>HNMR, <sup>3</sup>CNMR and <sup>31</sup>PNMR spectra of dTTMP-Ac are listed at the end of the Chapter (Figure 34).

#### 4.6.2 The synthesis of (diphenylphosphino)methanethiol succinate (dPPMT-Su)

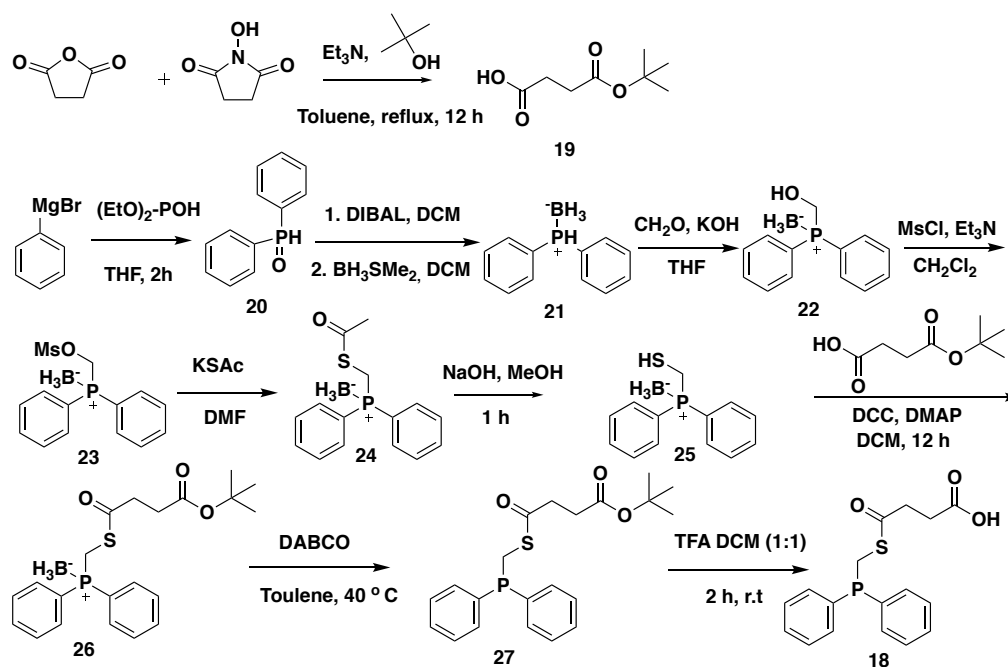


Figure 25 Scheme for the synthesis of dPPMT-Su

**The synthesis of 4-(tert-butoxy)-4-oxobutanoic acid (19):** To a mixture of succinic anhydride (3 g, 0.3 mol), N-hydroxysuccinimide (1 g, 0.009 mol), and DMAP (0.35 g, 0.003 mol) in toluene (15 mL) were added tert-butyl alcohol (3.5 mL, 0.037 mol) and  $\text{Et}_3\text{N}$  (0.009 mol, 1.25 mL). The suspension was refluxed for 24 h. The solution was cooled and diluted with EtOAc (15 mL). The reaction mixture was washed with 10% citric acid and brine, dried over  $\text{Na}_2\text{SO}_4$ , and concentrated to give a brown oil. The oil was recrystallized with ether and petroleum ether at  $-20^\circ\text{C}$  to give the desired product 8 as a white crystal (4.0 g, 78%).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ , 300 MHz):  $\delta$  2.62 (t, 2H,  $J = 3.3$  Hz), 2.54 (t, 2H,  $J = 1.2$  Hz), 1.44 (s, 9H),  $^{13}\text{CNMR}$  ( $\text{CDCl}_3$ , 75MHz): 178.5, 171.4, 81.0, 30.0, 29.1, 28.0.

**The synthesis of phosphine oxide diphenylphosphine oxide (20):** A 250 mL round bottom flask equipped with an addition funnel was evacuated and Ar filled for 3 times, charged with 50 mL 1 M PhMgBr/THF (71.7 mL, 71.7 mmol, 3.3 eq.), and the solution cooled to 0 °C under Ar. A solution of 3 g diethylphosphite (21.7 mmol, 1.0 eq) in 10 mL THF was then added dropwise over 15 min. The mixture was let stand for 15 min at 0 °C, then the bath was removed, and the mixture was stirred for two hours at ambient temperature and then cooled again to 0 °C. 75 mL 0.1 N HCl was added dropwise over 20 min, then 75 mL MTBE was added, and the mixture was filtered through a celite pad followed by washing the pad with CH<sub>2</sub>Cl<sub>2</sub>. The filtrate phases were separated, and the organic phase combined with the first organic phase, dried by MgSO<sub>4</sub>, and the solvents removed in vacuo. The residue was purified by flash column chromatography to give 20 (3.0 g, 70%) <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ: 8.87 (s, 1H), 7.73-7.50 (m, 4H), 7.49-7.45 (m, 6H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ: 132.5, 132.5, 130.6, 130.5, 128.9, 128.7. <sup>31</sup>P NMR (500 MHz, CDCl<sub>3</sub>): δ: 21.4.

**The synthesis of Phosphine-Borane Complex (21):** A solution of phosphine oxide 20 (2.00 g, 9.90 mmol) in anhydrous CH<sub>2</sub>Cl<sub>2</sub> (30 mL) was added dropwise slowly to a solution of DIBAL (1 M in CH<sub>2</sub>Cl<sub>2</sub>, 39.6 mL, 39.6 mmol) under Ar(g) in a flame-dried three-neck round-bottom flask. The resulting solution was stirred for 20 min, and then cooled to 0 °C with an ice bath. The solution was then diluted with CH<sub>2</sub>Cl<sub>2</sub> (30 mL), and a sparge needle of Ar(g) was allowed to blow through the solution for 5 min. A solution of 2 N NaOH (20 mL) was added dropwise slowly to the reaction mixture (Caution of gas evolution!) followed by a saturated solution of Rochelle's salt (20 mL) to dissipate the

emulsion that forms. The resulting biphasic solution was transferred to a separatory funnel, and the organic layer was separated, dried over anhydrous MgSO<sub>4</sub>, filtered, and concentrated under reduced pressure. (1.8 g 100%). Borane-dimethylsulfide (3 M solution in THF, 6.45 mL, 19.35 mmol) was added to a round-bottom flask with freshly distilled THF (20 mL). To this solution, the crude diphenylphosphine (1.8 g, 9.67 mmol) was added in one portion. The solution was stirred under nitrogen for 2 h and was quenched slowly with ice (~10 g). The resulting solution was diluted with brine (20 mL) and extracted with EtOAc (3 x 20 mL). The combined organic layers were washed with brine (2 x 10 mL), dried over sodium sulfate, filtered, and the solvent was removed under reduced pressure and the crude oil was purified by flash chromatography to yield 21 (1.1 g, 60% yield). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>) δ: 7.73-7.66 (m, 4H), 7.54-7.45 (m, 6H). <sup>31</sup>P NMR (500MHz, CDCl<sub>3</sub>): δ: 22.1.

**The synthesis of diphenylphosphino(borane)methane alcohol (22):** 21 (1.0 g, 5.0 mmol) was dissolved in a mixture of aqueous formaldehyde (37% w/w, 5 mL) and THF (10 mL). To this solution, KOH (700 mg, 10.0 mmol) was added. The solution was stirred for 2 h, and the volatile solvent was removed under reduced pressure. The resulting aqueous solution was extracted with EtOAc (3 x 10 mL), washed with brine (1 x 10 mL), dried over sodium sulfate, filtered, and the solvent was removed under reduced pressure and the crude product was purified by flash chromatography to yield 22 as a colorless liquid (0.861 g, 70%). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 300 MHz): δ 7.76-7.72 (m, 4H), 7.55-7.53(m, 2H), 7.50-7.47 (m, 4H), 4.45 (s, 3H), 2.05 (br, 1H), 1.46-0.52 (m, 3H); <sup>13</sup>C NMR (CDCl<sub>3</sub>, 75 MHz) δ



132.8, 132.7, 131.7, 131.7, 129.0, 128.9, 126.9, 126.4, 60.5, 60.2. <sup>31</sup>P NMR (CDCl<sub>3</sub>, 300 MHz) δ 25.7.

**The synthesis of diphenylphosphino(borane)methyl methanesulfonate (23):** 22 (1.1 g, 5.0 mmol) was dissolved in freshly distilled methylene chloride (10 mL), followed by addition of trimethylamine (1.4 mL, 10.1 mmol). The solution was cooled to 0 °C with ice-water bath. To this solution, methanesulfonyl chloride (580.0 μL, 7.4 mmol) was added dropwise. The solution was stirred under nitrogen for 15 h. The solvent was removed under reduced pressure. The crude product was purified by flash chromatography (3:1 Hexane:EtOAc) to yield 23 as a white solid (1.0 g, 3.9 mmol, 70%). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 300 MHz) δ 7.76-7.72 (m, 4H), 7.59-7.57 (m, 2H), 7.53-7.49 (m, 4H), 4.91 (d, 2H, J=1.2 Hz), 2.89 (s, 3H). <sup>31</sup>P NMR (CDCl<sub>3</sub>, 300 MHz) δ 18.2.

**The synthesis of diphenylphosphino(borane)methanethiol acetate (24):** Potassium thioacetate (0.737 g, 5.12 mmol) was added to a solution of 12 (0.8 g, 2.58 mmol) in anhydrous DMF (15 mL) under Ar(g). The resulting solution was stirred overnight at room temperature, after which the solvent was removed under reduced pressure. The residue was dissolved in EtOAc (20 mL), and the resulting solution was washed with water and brine. The combined organic extracts were dried over anhydrous MgSO<sub>4</sub>, filtered, and the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (4:1 hexanes:ethyl acetate) to yield 24 as a white solid (0.487 g, 65%). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 300 MHz) δ 7.71-7.67 (m, 4H), 7.51-7.49 (m, 2H), 7.46-7.43 (m, 4H), 3.72 (d, 2H, J=4.5 Hz), 2.24 (s, 3H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 75 MHz) δ 193.2, 132.4, 132.4,

131.8, 131.7, 128.9, 128.8, 127.8, 127.3, 30.0, 23.9 (d). <sup>31</sup>P NMR (CDCl<sub>3</sub>, 300 MHz) δ 19.0 (d).

**The synthesis of diphenylphosphino(borane)methanethiol-mono-*t*-butyl succinate ester (26):** 24 (0.4 g, 1.38 mmol) was dissolved in freshly distilled methanol (5 mL), followed by addition of NaOH (66 mg, 1.66 mmol). After stirring under nitrogen for 10 min, the solution was neutralized by 1 N HCl, and was extracted with EtOAc (2 × 30 mL). The combined organic layers were washed by brine (1 × 10 mL), dried over sodium sulfate, filtered, and concentrated under reduced pressure to obtain 25. The crude thiol of 25 was dissolved in freshly distilled CH<sub>2</sub>Cl<sub>2</sub> (10 mL), followed by the addition of 19 (0.352 g, 2.0 mmol), 1,3-dicyclohexylcarbodiimide (0.5 g, 2.42 mmol), and cat. *N,N'*-dimethyl-4-aminopyridine. The solution was stirred under nitrogen for 4 h, and the solution was filtered through a pad of celite. The solvent was removed under reduced pressure. The crude product was purified by flash chromatography (3:1 Hexane:EtOAc) to yield the thioester 26 as a colorless oil (0.33 g, 60%). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 300 MHz) δ 7.72-7.68 (m, 4H), 7.53-7.45 (m, 6H), 3.75 (d, 2H, J = 3.9 Hz), 2.75 (t, 2H, J = 3.9 Hz), 2.48 (t, 2H, J = 3.3 Hz), 1.43 (s, 9H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 75 MHz) δ 195 (d), 170.6, 132.5, 132.4, 131.8, 131.7, 128.9, 128.8, 127.8, 127.3, 38.4, 30.3, 28.0, 23.5 (d). <sup>31</sup>P NMR (CDCl<sub>3</sub>, 75 MHz) δ 41.5.

**The synthesis of diphenylphosphinomethanethiol-mono-*t*-butyl succinate ester (27):** Phosphine-borane 26 (0.25 g, 0.621 mmol) and 1,4-diazabicyclo[2.2.2]octane (DABCO) (0.139 g, 1.24 mmol) was dissolved in freshly distilled toluene (2 mL). The solution was submerged in an oil-bath pre-heated to 60 °C. The solution was stirred under argon

for 1 h. After cooling to room temperature, the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (4:1 Hexane:EtOAc) under nitrogen to yield 27 as a colorless oil (0.168 g, 70%). <sup>1</sup>H NMR (CDCl<sub>3</sub>) δ 7.46-7.43 (m, 4H), 7.37-7.36 (m, 6H), 3.55 (d, 2H, J = 2.1 Hz), 2.80 (t, 2H, J = 4.2 Hz), 2.55 (t, 2H, J = 4.5 Hz), 1.45 (s, 9H). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 75MHz) δ 196 (d), 170.9, 136.8, 136.7, 132.8, 132.6, 129.1, 128.6, 128.5, 80.9, 38.5, 30.5, 28.0, 25.6, 25.4 (d). <sup>31</sup>P NMR (CDCl<sub>3</sub>, 300 MHz) δ - 15.3.

**The synthesis of diphenylphosphinomethanethiol-succinic acid (18):** To a solution of ester 27 (0.15 g, 0.45 mmol) in CH<sub>2</sub>Cl<sub>2</sub> (3 mL) was added TFA (3 mL). After being stirred at r.t. for 1 h, the reaction mixture was concentrated in vacuum to give a colorless oil 18 as the final product (0.098 g, 73%). <sup>1</sup>H NMR (CD<sub>3</sub>OH, 300 MHz) δ 7.44-7.42 (m, 4H), 7.41-7.36 (m, 6H), 3.56 (d, 2H, J=2.4Hz), 2.82 (t, 2H, J=4.2Hz), 2.58 (t, 2H, J=4.2Hz). <sup>13</sup>C NMR (CDCl<sub>3</sub>, 75 MHz) δ 196.9 (d), 174.0, 136.8, 132.5, 132.3, 128.8, 128.3, 128.2, 37.8, 28.4, 24.8 (d). <sup>31</sup>P NMR (CDCl<sub>3</sub>, 300 MHz) δ - 14.9. All the <sup>1</sup>H NMR, <sup>13</sup>C NMR and <sup>31</sup>P NMR spectra of dTTMP-Su are listed at the end of the Chapter (Figure 35).

## 4.7 Methodology investigation on Ub-K48AznL

### 4.7.1 Reaction setting up of dPPMT-Ac

With the synthesized diphenylphosphinothioester reagents at hand, we started to investigate our methodology over Ub-K48AznL as our model protein in a time-dependent view (Figure 26A). Because the traceless-Staudinger ligation reaction has been known as a multi-step reaction with a complicated mechanism<sup>[103]</sup>, we decided to elongate the

reaction to two days. Around 5 mg of dPPMT-Ac was dissolved in pure DMSO to a final concentration of 100 mM as the stock solution. The Ub-K48AznL, wild type Ub, and Ub-K48ac were all prepared in 200 mM phosphate buffer pH 6.0 with concentrations between 10 and 20  $\mu$ M. The phosphine reagents in DMSO stock solution were slowly added into the Ub-K48AznL ( $\sim$ 20  $\mu$ M) solution to reach to a final concentration of 5 mM. Reactions were gently pipetted to make solutions well mixed, and potential precipitations were removed by centrifuging. Then, the reaction mixture was let to incubate at 37  $^{\circ}$ C for one to two days. A same amount of WT Ub with the addition of same amount of phosphine reagent was used as the negative control, while Ub-K48ac without phosphine was used as the positive control. Samples of reaction mixtures were taken 12 h intervals, followed by the addition of 5  $\mu$ L azidoethanol to each sample as a quencher, and then diluted with 10 $\times$  PBS buffer. The quenched samples were centrifuged at 14K for 5 min to remove any precipitations, dialyzed over 10 $\times$  PBS buffer for 3 times, and then concentrated to small volumes.

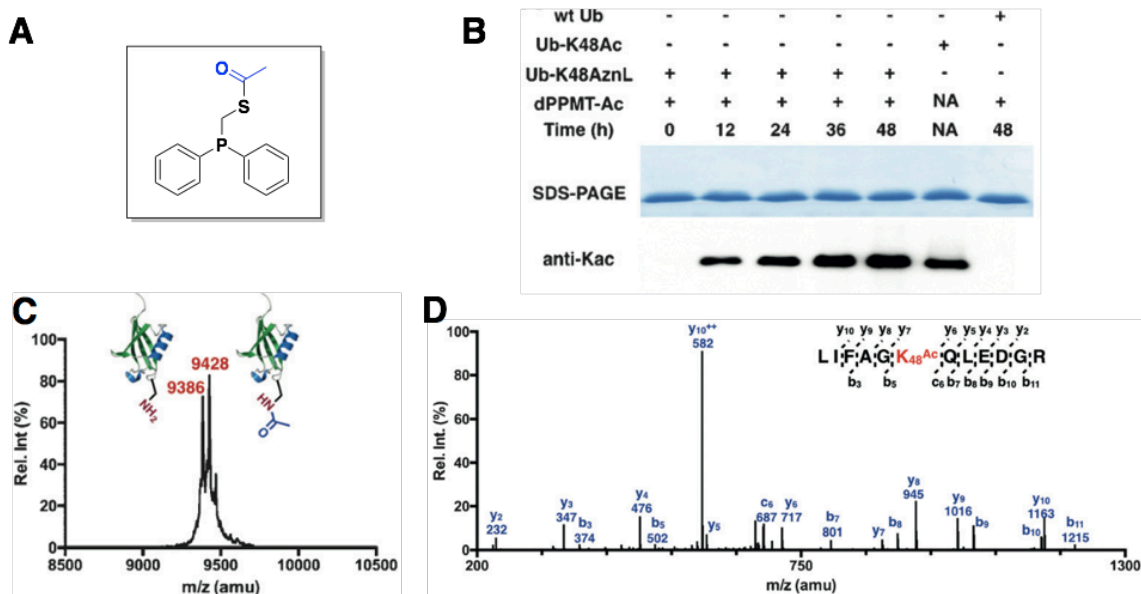
#### 4.7.2 Reaction characterization

Different samples were taken with a time interval of 12 h. After mixing with a SDS loading buffer, final samples were applied to two 15 % SDS-PAGE gel separately. The first gel was stained with Coomassie blue and then directly applied for imaging. The second gel was transferred to a nitrocellulose membrane for Western blot analysis. The pan anti-Kac antibody from PTM BioLabs Inc. was first applied and then followed by blotting with secondary antibody and visualization. The result showed that an increased

acetylation level with the elongation of reaction time (Figure 26B). The reaction is not close to completion until at the time point of 48 h, since the acetylation level was not significantly higher than the one at 36 h. The negative control WT Ub did not have any detectable acetylation band, in contrast to the intense acetylation band for the Ub-K48AznL reaction group.

We further applied MALDI-TOF-MS to exam the product from Ub-K48AznL to its corresponding acetylation product Ub-K48ac. Reaction sample after 48-h incubation was applied to Ziptip (Reversed-phase, pipette tips for sample preparation, Millipore Corporation) sample preparation, followed by sending to MALDI-TOF-MS (Applied Biosystems Voyager-DE STR) with data explorer software from the same company. The spectrum showed two main peaks at 9386 and 9428Da, suggesting the corresponding traceless Staudinger ligation product Ub-K48ac and a side product Ub-K48 with native Lys, whose theoretical molecular weights are 9387 and 9429 Da, respectively (Figure 26C). The generation of side-product will be further discussed in the next part, while some other peaks are representing some unknown impurities or potential side reactions not fully studied. We can see from the quantification that the desired product Ub-K48ac has slightly higher percentage over the by-product WT Ub in the final product mixture.

To confirm the site-specificity of the reaction, the 48-h reaction product was trypsinized and applied to the MS/MS analysis. The K48ac-containing peptide fragment could be clearly observed, indicating the acetylation at K48 site (Figure 26D).

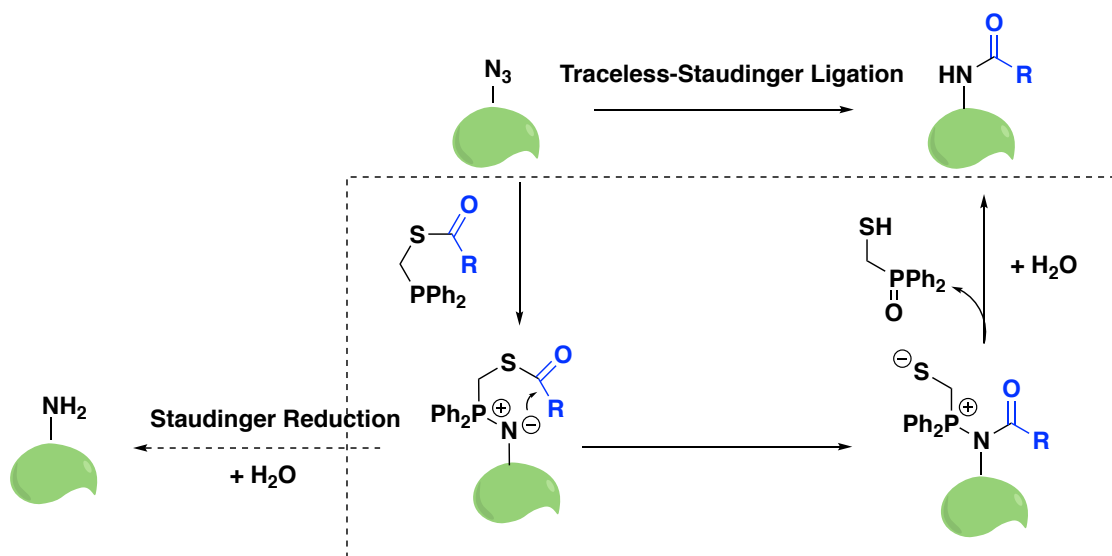


**Figure 26 Methodology examination on UbK48AznL** (A) the structure of dPPMT-Ac; (B) the gel-based time dependent assay for Lys acetylation; (C) MALDI-TOF of UbK48AznL after traceless Staudinger acetylation reaction for 48 h; (D) MS/MS analysis of UbK48AznL after traceless Staudinger acetylation reaction for 48 h. Adapted with permission from [99].

#### 4.7.3 Result analysis

Although as thioester, the phosphinothioester did not show much non-specific interaction with native Lys to generate Kac everywhere, due to the nature of aliphatic thioester. From another point of view, although traceless Staudinger ligation owns a very slow kinetics, the Western blot result together with MS/MS spectrum shows a much slower non-specific intermolecular S-to-N acyl transfer between dMMPT-Ac and free Lys on Ub.

For the generation of native Lys peak, we need to consider about the mechanism of traceless-Staudinger ligation reaction, because it has been noted that phosphinothioester can reduce azide back to amine due to its intrinsic reductive nature<sup>[104]</sup>.



**Figure 27 Mechanism of traceless Staudinger ligation reaction on proteins**

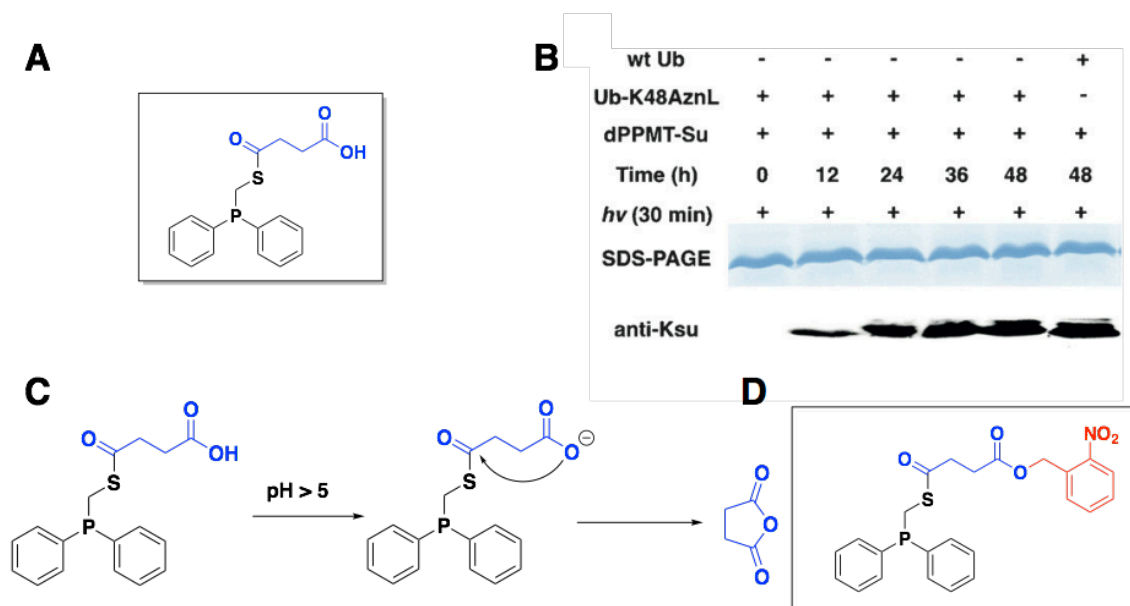
To start with, the phosphine (III) of the phosphinothioester attacks the terminal azide nitrogen atom, and with the leaving of the nitrogen gas, a key N-P ylide intermediate forms. The nitrogen with partial negative charge will attack the adjacent carbonyl group of the thioester to form an amide bond, which can undergo water attacking to give out the final acylated product (Figure 27). However, the N-P ylide intermediate can directly undergo hydrolysis to release native Lys side chain. Therefore, the overall result for the side reaction is to reduce the azide group back to an amino group, which shares exactly the same process as other trialkylphosphine reagents such as TCEP. Although this is inevitable process, theoretically we should be able to separate the Lys from the acylated products by using an immunoaffinity column with acylated Lys specific antibody. But for this research, we did not take further efforts for the product purification, because the native Lys will not give out any signal for enzymatic deacylation assay.

## 4.8 Methodology examination and optimization for Lys succinylation

### 4.8.1 Reaction test of dPPMT-Su

Encouraged by the result of acetylation, we decided to move ahead to succinylation, with the synthesized dPPMT-Su at hand (Figure 28A). Around 5 mg of dPPMT-Su and was dissolved in DMSO solution to reach 100 mM as stock solution. The Ub-K48AznL and WT Ub in 200 mM phosphate buffer pH 6.0 with was added with the phosphine reagents in DMSO stock solution slowly to reach to a final concentration of 5 mM. Reactions were gently pipetted to make solutions well mixed and then let to stand at 37 °C. A same amount of WT Ub with the addition of same amount of phosphine reagent was used as the negative control. Samples of reaction mixtures were taken 12 h intervals, followed by mixturing with 5  $\mu$ L azidoethanol to each sample as a quencher. The quenched samples were centrifuged at 14K for 5 min to remove any precipitations, dialyzed over 10 $\times$  PBS buffer for 3 times, and then concentrated to small volumes. The samples were mixed with a SDS loading buffer, and then applied to two 15 % SDS-PAGE gel separately. The first gel was stained with Coomassie blue and then directly applied for imaging. The second gel was transferred to a nitrocellulose membrane for Western blot analysis. The pan anti-Ksu antibody from PTM BioLabs Inc. was first applied and then followed by blotting with secondary antibody and visualization. The result showed that a band with increased intensity with the elongation of reaction time. However, the negative control WT Ub showed a strong succinylation band as well after mixing with dPPMT-Su (Figure 28B).





**Figure 28 Methodology examination and optimization of Lys succinylation** (A) the structure of dPPMT-Su; (B) the gel-based time dependent assay for Lys succinylation; (C) the potential side reaction of non-specific Lys succinylation; (D) the structure of dPPMT-NB-Su. Adapted with permission from [99].

#### 4.8.2 Result analysis

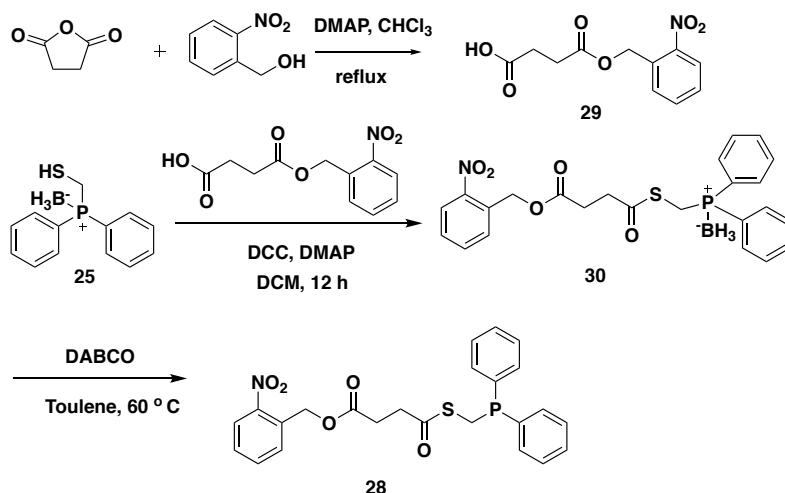
From the result above, we found that the dPPMT-Su showed a very strong non-specific acylation reaction to provide succinylation over native Lys residues. The first explanation comes from an increased electrophilicity of the succinyl thioester due to the electron-withdrawing effect of the second carboxyl group comparing with the acetyl group. However, even though the activity of the thioester bond increases, we cannot expect such high level of by product formation.

Then we proposed another potential mechanism. Different from the dPPMT-Ac, the dPPMT-Su owns a free carboxyl group on the side chain of succinyl part. This free carboxyl group can attack the thioester warhead followed by the hydrolysis to generate

succinyl anhydride. This reaction has been shown to undergo in aqueous condition with neutral pH, especially during long-time storage. Succinic anhydride is a very active compound that can label Lys without any selection (Figure 28C). One simple way to overcome it is to provide an orthogonal protecting group over the free carboxyl group, which can be efficiently removed after traceless Staudinger ligation reaction. Therefore, we decided to choose the *O*-nitrobenzyl photocaging group (Figure 28D).

#### 4.8.3 The synthesis of photocaged (diphenylphosphino)thiol succinate

We have decided to use the *O*-nitrobenzyl protecting group (NB) to protect the free carboxyl group, and thus we labeled our new dPPMT-Su edition as dPPMT-NB-Su.



**Figure 29** Scheme for the synthesis of dPPMT-NB-Su

**The synthesis of 4-((2-nitrobenzyl)oxy)-4-oxobutanoic acid (29):** *O*-nitrobenzyl alcohol (4.0 g, 26.12 mmol), succinic anhydride (5.23 g, 52.24 mmol), and DMAP (1.60 g, 13.06 mmol) were dissolved completely in dried CHCl<sub>3</sub> (86 mL) and refluxed under a

nitrogen atmosphere for 24 h. After removing partially  $\text{CHCl}_3$  under reduced pressure, the mixture was washed three times with 10 % HCl and then extracted with saturated  $\text{NaHCO}_3$  solution. The basic aqueous phase was washed with ether and acidified to pH 5.0 with 10% HCl. The white solid precipitate was collected and dried in vacuo at 40 °C overnight to give 29 (6.08 g, 92%).  $^1\text{H NMR}$  ( $\text{CDCl}_3$ ):  $\delta$  8.09 (dd, 1H,  $J = 0.4, 1.8$  Hz), 7.66–7.59 (m, 2H), 7.50–7.47 (m, 1H), 5.55 (s, 2H), 2.76–2.72 (m, 4H).

**The synthesis diphenylphosphino(borane)methanethiol-2-nitrobenzylsuccinate ester (30):** 25 (0.4 g, 1.38 mmol) from the synthesis of dPPMT-Ac was dissolved in freshly distilled methanol (5 mL), followed by addition of NaOH (66 mg, 1.66 mmol). After stirring under nitrogen for 10 min, the solution was neutralized by 1 N HCl, and was extracted with EtOAc (30 mL) twice. The combined organic layers were washed by 10 mL brine, dried over sodium sulfate, filtered, and concentrated under reduced pressure. The crude thiol (14) was dissolved in freshly distilled  $\text{CH}_2\text{Cl}_2$  (10 mL), followed by addition of 18 (0.408 g, 1.58 mmol), DCC (0.4 g, 0.97 mmol), and cat. *N,N'*-dimethyl-4-aminopyridine. The solution was stirred under nitrogen for 4 h, and the solution was filtered through a pad of celite. The solvent was removed under reduced pressure. The crude product was purified by flash chromatography (3:1 Hexane:EtOAc) to yield the thioester 30 as a colorless oil (0.432g, 65%).  $^1\text{HNMR}$  ( $\text{CDCl}_3$ , 300MHz)  $\delta$  8.10 (t, 1H,  $J = 1.2$  Hz), 7.85–7.83 (m, 1H), 7.82–7.80 (m, 4 H), 7.69–7.43 (m, 8H), 5.61 (s, 2H), 3.73 (d, 2H,  $J = 1.2$  Hz), 2.85 (t, 2H,  $J = 4.2$  Hz), 2.67 (t, 2H,  $J = 4.2$  Hz).  $^{13}\text{CNMR}$  ( $\text{CDCl}_3$ , 75 MHz)  $\delta$  195 (d), 170, 147.4, 133.9, 132.4, 132.4, 132.0, 131.8, 131.4, 131.3, 129.0, 128.9,

128.9, 128.8, 128.7, 128.6, 127.6, 127.2, 125.0, 63.3, 37.9, 29.0, 23.6 (d).  $^{31}\text{P}$  NMR ( $\text{CDCl}_3$ , 300 MHz)  $\delta$  40.4;

**The synthesis of diphenylphosphinomethanethiol-2-nitrobenzylsuccinate ester (28):**

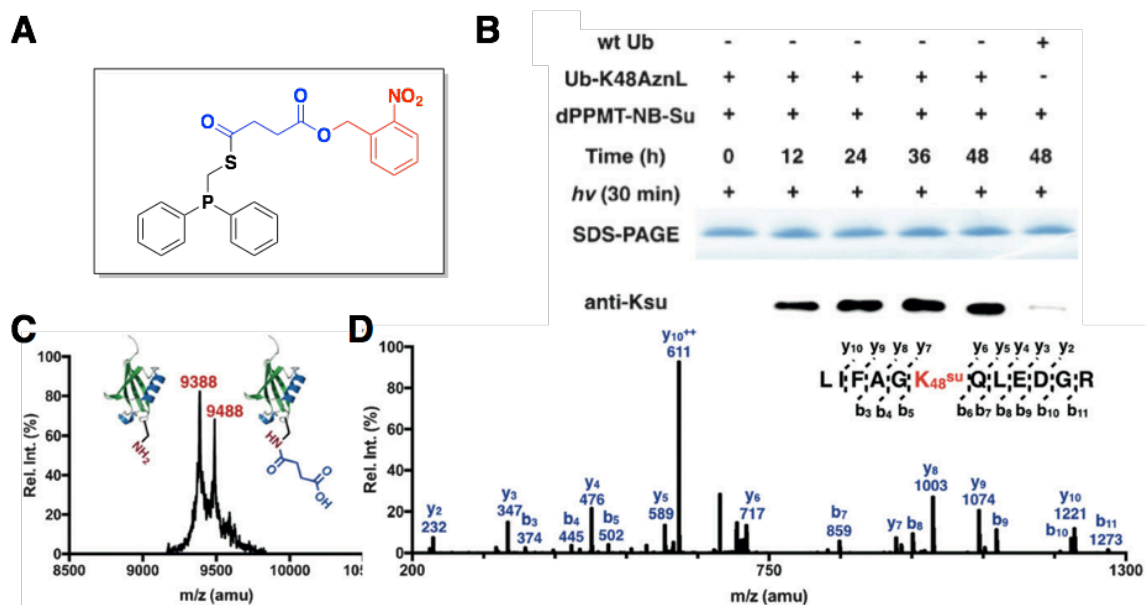
30 (0.2 g, 0.41 mmol) and DABCO (0.139 g mg, 1.24 mmol) was dissolved in freshly distilled toluene (2 mL). The solution was submerged in an oil-bath pre-heated to 60 °C. The solution was stirred under argon for 1 h. After cooling to r.t., the solvent was removed under reduced pressure. The crude product was purified by flash chromatography (3:1 Hexane:EtOAc) under nitrogen to yield 28 as a colorless oil (0.15 g, 80%).  $^1\text{H}$ NMR ( $\text{CDCl}_3$ )  $\delta$  8.10 (dd, 1 H,  $J = 0.6, 4.8$  Hz), 7.66-7.63 (m, 2H), 7.46-7.42 (m, 5H), 7.36-7.34 (m, 6H), 5.52 (s, 2H), 3.53 (d, 2H,  $J = 2.1$  Hz), 2.89 (t, 2H,  $J = 4.2$  Hz), 2.73 (t, 2H,  $J = 4.2$  Hz).  $^{13}\text{C}$ NMR ( $\text{CDCl}_3$ , 75 MHz)  $\delta$  196 (d), 171.2, 147.4, 136.7, 136.6, 133.8, 132.8, 132.6, 131.9, 129.2, 128.9, 128.8, 128.7, 128.6, 128.5, 120.0, 63.2, 38.1, 29.1, 25.7 (d).  $^{31}\text{P}$  NMR ( $\text{CDCl}_3$ , 300 MHz)  $\delta$  - 15.1. All the  $^1\text{H}$ NMR,  $^{13}\text{C}$ NMR and  $^{31}\text{P}$ NMR spectra of dTTMP-NB-Su are listed at the end of the Chapter (Figure 36).

#### 4.8.4 Reaction test of photocaged succinyl diphenylphosphinothioester

With dPPMT-NB-Su at hand, the reaction setting up is exactly the same as the non-protected edition, which WT-Ub was still used as a negative control (Figure 30A). After incubated under 48 h, the reaction mixture was quenched. The quenched samples were treated with UV light at 365 nm wavelength for 30 min on ice bath. After mixing with a SDS loading buffer, final samples were applied to two 15 % SDS-PAGE gel separately. The first gel was stained with Coomassie blue and applied for imaging. The

second gel was then transferred to nitrocellulose membrane for the Western blot analysis. The pan-anti-Ksu antibody from PTM BioLabs Inc. was first applied and then followed by the blotting with secondary antibody and visualization. To our surprise, the control reaction with wild type Ub did show detectable lysine succinylation in a very low level (Figure 30B). Although we did not observe any side reaction with dPPMT-Ac, it is possible that the electronwithdrawing inductive effect of the photocaged carbonyl group in dPPMT-NB-Su actually still plays a role in the activation of its thioester. Therefore, the non-specific S-to-N acyl transfer becomes more favored, and a succinylation over native Lys side chain was then observed.

To further confirm the formation of Ub-K48su, the sample after 48-h incubation and UV treatment was treated with Ziptip as mentioned above and then sent for MALDI-TOF-MS analysis. In the spectrum, two major peaks showed up as one at 9388 Da and another at 9488 Da, agreeing with the MW of the corresponding WT Ub and Ub-K48su, respectively (Figure 30C). Interestingly, the side product with native Lys is a little more abundant than the desired product of Ub-K48su, which does not agree with the increased electrophilicity of the succinyl carbonyl group. We explain it as an increased steric effect interrupting the attacking of the N-P ylid to the thioester, due to the introduction of the photo caging group at the side chain carbonyl group. Notably, some side peaks in Figure 30C are representing some potential side reactions not fully studied.



**Figure 30 Methodology examination of Lys succinylation via dPPMT-NB-Su.** (A) the structure of dPPMT-NB-Su; (B) the gel-based time dependent assay for Lys succinylation; (C) MALDI-TOF of UbK48AznL after traceless Staudinger succinylation reaction for 48 h; (D) MS/MS analysis of UbK48AznL after traceless Staudinger succinylation reaction for 48 h. Adapted with permission from [99].

Next, the final product was trypsinized and sent for MS/MS analysis to confirm the succinylation at the K48 position. A peptide fragment containing K48su was detected, showing the succinylation at K48 (Figure 30D).

## 4.9 The expression of histone H3-K4AznL, WT-H3, and H4-His6-SUMO

### 4.9.1 The DNA and protein sequence of histone H3

*DNA sequence: WT-H3(C110A):*

gctcgcaccaaacagactgctcgttaagtccactggcggtaaagcgcgcgtaaacagctggcaaccaagatggcgcgtaaa  
 agcgcctcagctactggcggcgtgaagaagccgcaccgttatcgccgggtactgtggctctgcgtgaaatccgccgctacc

agaaaagcaccgaactgctgattcgcaaactgccattcaacgtctggttcgcaaatgctcaggatttcaaaaccgacctgcg  
ctccagtctagcgtgtgatggcactgcaagaggcgtctgaggcatatctggttgccctgttcgaagataccaacctgtgcga  
atccatgcaaagcgtgtaaccattatgccgaaagacatccaactggctcgtcgtatccgtggtgagcgtgcgtga

*Protein sequence: His6-TEV-H3-K4AznL(C110A):*

MGSSHHHHHSQDPENLYFQART(AznL)QTARKSTGGKAPRKQLATKAARKSA  
PATGGVKKPHRYRPGTVALREIRRYQKSTELLIRKLPFQRLVREIAQDFKTDLRF  
QSSAVMALQEASEAYLVGLFEDTNLAAIHAKRV TIMPKDIQLARRIGERA

#### 4.9.2 The purification of H3K4AznL

Encouraged by the acetylation and succinylation result on Ub, we decided to move ahead to histone succinylation and other higher structures such as tetramer or nucleosome. In other methodology for PTM study, different histone variants with site-specific Lys PTMs must be prepared separated and then assembled to tetramer. By combine our AznL incorporation and traceless Staudinger ligation reaction, we can install a large variety of Lys acylation modifications into proteins by simply changing the acyl group in the phosphinothioester reagent we use, in a “one-size-fits-all” manner. And we started our methodology investigation over histone H3 with amber codon at K4 position.

The plasmid of pEvol-PylT-AznLRS mentioned before was used together with pETduet-H3K4TAG to co-transform *E. coli* BL21(DE3) cells. The cells were placed on LB agar plate containing 100 µg/mL Amp and 34 µg/mL Cm. A single colony was picked, followed by the growth in a 5 mL LB medium for o/n. The overnight culture was further

inoculated into a 250 mL 2YT medium with the same concentration of Amp and Cam. Cells were grown in a 37 °C shaker (250 r.p.m.) for 3 h until OD<sub>600</sub> reached 0.6. The protein expression was induced by the addition of 1 mM IPTG, 0.2 % arabinose and 5 mM AznL. After another 8 h incubation, the cells were harvested by centrifugation (4k 20 min, 4 °C), washed, and fully resuspended in a 40 mL lysis buffer (20 mM Tris-HCl, 500 mM NaCl, 0.1% Triton X-100, 0.1% NaN<sub>3</sub>, pH 7.5). The sonication of cells underwent under ice/water incubation for 3 times (4 min each time with an interval of 4 min). Then the cell lysate was centrifuged (6k 20 min, 4 °C), and the supernatant was removed. The precipitate was washed with 30 mL lysis wash buffer (20 mM Tris-HCl, 500 mM NaCl, 0.1% NaN<sub>3</sub>, pH 7.5) twice. The supernatant was removed again, and the inclusion body was dissolved in 6 M urea histone solubilization buffer (6 M urea, 20 mM Tris, 500 mM NaCl, pH 7.5). The solution was centrifuged at 10 k for 40 min to remove all the precipitate. The supernatant was loaded to 1 mL Ni Sepharose<sup>TM</sup> 6 Fast Flow column (GE Healthcare) and the flow-through was removed. 15 mL of wash buffer (6 M urea, 20 mM Tris, 500 mM NaCl, 20 mM imidazole, pH 7.5) was used to wash the protein-bound resin in three batches. Around 6 mL elution buffer (6 M urea, 20 mM Tris, 500 mM NaCl, pH 7.5) was used to elute the target protein. All the eluted fractions were then collected and concentrated by Amicon Ultra-15 Centrifugal Filter Device (3 k MWCO, Millipore), and buffer was exchanged to histone solubilization buffer for the further traceless Staudinger ligation reaction.

Both WT histone H3 and H4 proteins with N-terminal his<sub>6</sub> tags were expressed in a 1 L 2YT media with the addition of 100 µg/mL Amp. Cells were let to grow in a 37 °C



for 3 h until OD<sub>600</sub> reached 0.6. The protein expression was induced by the addition of 0.8 mM IPTG. After incubated for 8 h, cells were collected and sonicated. The further purification processes were the same as the purification of H3-K4AznL.

#### **4.10 General process of time-dependent traceless-Staudinger ligation reaction of H3K4AznL**

##### 4.10.1 The setting up of time-dependent Succinylation assay

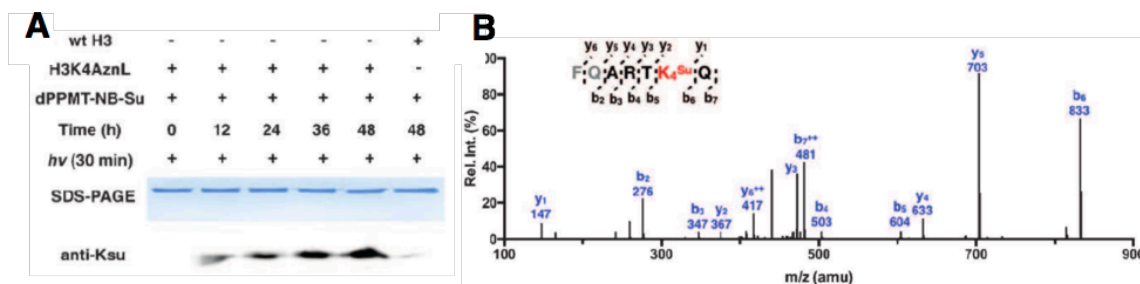
The setting up of traceless Staudinger ligation reaction for H3K4AznL was similar to the setting up of Ub, except for H3K4su, the degassed histone solubilization buffer with 6 M urea at pH 6.0 instead of a PBS buffer was used when it was necessary. The photocaged phosphine reagents dPPMT-NB-Su in DMSO stock solution were added to reach to a final concentration of 5 mM. Reactions were gently pipetted to make solutions well mixed and then let to incubate at 37 °C. A same amount of WT H3 with the addition of same amount of phosphine reagent was used as the negative control. Samples of reaction mixtures were taken with a 12 h interval, followed by the addition of 5 µL of azidoethanol to each sample as the quencher. The quenched samples were centrifuged at 14K for 5 min to remove any precipitations, dialyzed over PBS with 6 M urea for 3 times, and then concentrated to small volumes. The dialyzed samples were treated with UV light at 365 nm for 30 min and applied to two 15 % SDS-PAGE gel separately. The first gel was stained with Coomassie blue and applied for imaging. The second gel was further transferred to nitrocellulose membrane for Western blot. The pan-anti-Ksu antibody from

PTM BioLabs Inc. was first applied and then followed by the blotting with secondary antibody and visualization (Figure 31A).

#### 4.10.2 The MS/MS analysis of H3K4AznL

We further confirmed K4 succinylation with the MS/MS analysis, which clearly showed a Ksu-containing peptide fragment at K4 position (Figure 31B). Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS) was applied to identify the specific modifications to Ub-K48 and histone H3-K4 proteins. The protein digestion, LC-MS/MS analysis, and data analysis were conducted following the method described previously. 10 µg of each protein was used for the analysis. Briefly, protein sample was denatured by 8 M urea dissolved in Tris buffer (50 mM Tris-HCl, 10mM CaCl<sub>2</sub>, pH 7.6), supplemented with 5 mM dithiothreitol (DTT) for reduction of potential disulphide bonds. The denatured protein was diluted into a final concentration of 1 M urea with the same Tris buffer, followed by peptide digestion. Ubiquitin was digested by Mass Spectrometry Grade Trypsin Gold (Promega, Madison, WI) with 1:50 w/w at 37 °C overnight, while histone H3 was digested by chymotrypsin (Promega, Madison, WI) at 25 °C. The digested peptides were then cleaned up using a Sep-Pak Plus C18 column (Waters Corporation, Milford, MA), followed by loading onto a biphasic 2D reversed phase (RP)/strong cation exchange (SCX) capillary column with a pressure cell. The column was then washed with 100% aqueous solvent (99.9 % H<sub>2</sub>O, 0.1 % formic acid) for 5 min, followed by a 1-hour wash by ramping up the solvent to 100% organic solvent (0.1 % formic acid, 80% acetonitrile). This washing step migrates all the peptides into the SCX phase, where

peptides were eluted and separated in a 15-cm-long 100  $\mu\text{m}$ -ID C18 capillary column connected to the 2D column. LTQ ion trap mass spectrometer (Thermo Finnegan, San Jose, CA) was used to analyze the peptides. The full scan was set to the range of 300-1700  $m/z$ , in which the top 5 abundant peptides in each scan were subjected to collision induced dissociation (CID) fragmentation for tandem mass spectrometry (MS/MS) analysis. Peptide identity was processed by the in-house pipeline as described previously. Peptide searching was based on ProLuCID (version 1.0) search algorithm. For the peptide modification search, mass increase was added in order to identify the modified peptides. More specifically, 42.0367 was added for lysine acetylation modification, and 100.0648 was added for lysine succinylation modification.



**Figure 31 Preparation of histone with Lys succinylation via dPPMT-NB-Su** (A) the gel-based time dependent assay for succinylation of H3; (B) MS/MS analysis of H3 succinylation product. Adapted with permission from [99].

## 4.11 The assembly of H3K4su-H4 tetramer

### 4.11.1 The preparation of H3K4su

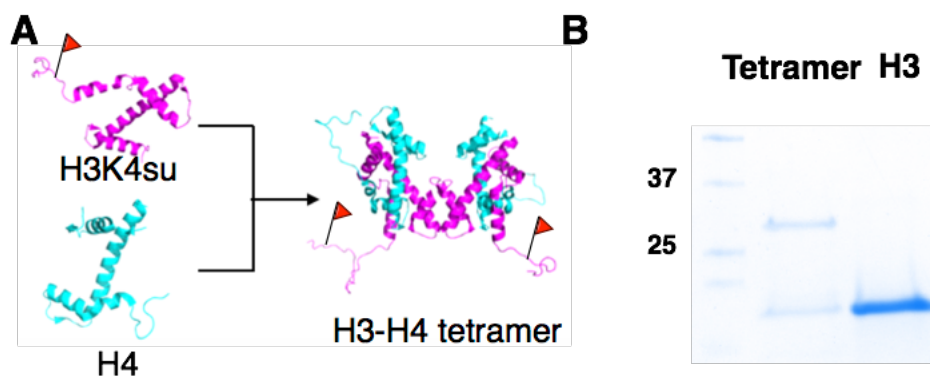
Since the time-dependent assay showed a successful conversion of H3K4AznL to H3K4su, we wanted to further test whether the product we have could be directly used as substrate for enzymatic assays. As mentioned in the previous parts, lysine succinylation

reverses the charge state of lysine, and thus its occurrence in histones can play a significant role to impact the interactions between the histone octamer and its associated DNA. Therefore, we decided to use the approach we developed to synthesize histone H3 with succinylation at K4 and further assembly it into tetramer with WT H4 as a substrate for enzymatic desuccinylation test. AznL was incorporated into H3 as H3K4AznL, followed by the reaction with dMMPT-NB-Su similarly as the setting up of time-dependent assay described above. The product after UV radiation was dialyzed and confirmed with Western blot analysis showing a strong band with the pan-anti-Ksu antibody. WT H3 was again used as negative control resulting in a very low level of lysine succinylation. After concentrating to lower volume, the reaction product was then treated with UV light for 30 min, followed by a second dialysis against histone solubilization buffer pH 7.5 for 3 times. The solution was concentrated again down to 100  $\mu$ l and stored for the next step.

#### 4.11.2 The assembly of H3K4su-H4 tetramer

After confirming the succinylation site, we directly used the product after 48-h incubation and UV-treatment to assemble an H3K4su-H4 tetramer as a substrate to test Sirt5 activity (Figure 32A). Both histone proteins H3 and WT H4-SUMO with a N-terminal His<sub>6</sub> tag were dissolved in the histone solubilization buffer (6 M urea, 20 mM Tris, 500 mM NaCl, pH 7.5) with equimolar ratios. The protein mixture was sequentially dialyzed against 500 mL refolding buffer 1 (20 mM Tris, 2 M NaCl, pH 7.5) for 30 min once, 500 mL refolding buffer 2 (20 mM Tris, 1 M NaCl, pH 7.5) for 30 min once, and 500 mL refolding buffer 3 (20 mM Tris, 500 mM NaCl, pH 7.5) for 30 min twice with

Slide-A-Lyzer MINI Dialysis devices (3,500-Da cutoff). After the dialysis process to refold, tetramers were put into a microcentrifuge tube and let to spin down by 14000 rpm for 1 min to remove all the possible precipitation during the refolding process. The tetramer product was measured with a concentration of about 16 mg/mL. SDS-PAGE gel was used to confirm the tetramer folding showing both bands of H3 and H4, and the tetramer was confirmed with SDS-PAGE and then stored on ice for further usages (Figure 32B).



**Figure 32 Preparation of H3K4su-H4 tetramer** (A) scheme of the assembly of H4-H3K4su tetramer; (B) SDS-PAGE analysis of H4-H3K4su tetramer.

## 4.12 Sirt5 enzymatic assay

### 4.12.1 The expression of Sirt1 and Sirt5

Sirtuin 5 (Sirt5) is a NAD-dependent histone deacetylase family enzyme, which has been reported to remove succinylation from histones and regulate chromatin functions<sup>[105]</sup>. However, previous research has been based only on histone N-terminal peptides, and no evidence has been shown whether Sirt5 performs active over histone or other higher

organizations such as tetramer, octamer, or nucleosome. Therefore, we would like to apply the tetramer we have to show the desuccinylation activity of Sirt5 but no Sirt1.

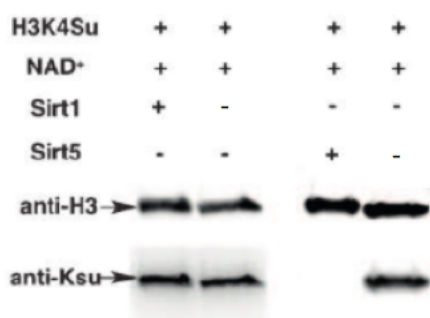
Both the Sirt1 and Sirt5 enzymes were expressed according to the procedure reported before<sup>[106]</sup>. Both enzymes were expressed in *E. coli* BL21(DE3) cells in 1 L LB media at 30 °C for o/n, induced by 0.1 mM IPTG. After collecting and sonicating cells, the precipitate was removed by centrifugation (6k 20 min, 4 °C). The supernatant was purified with the GST resin similar to Tip60. The mixture was directly loaded to a column and the flow-through was removed. 50 mL PBS washing buffer (200 mM NaH<sub>2</sub>PO<sub>4</sub>, 100 mM NaCl, pH 7.4) was used to wash the protein-bound resin in three batches. 10 mL of elution buffer (15 mM Glutathione in reduced form, 200 mM NaH<sub>2</sub>PO<sub>4</sub>, 100 mM NaCl, pH 7.4) was used to elute the target protein. After concentration measurement, the enzymes were split into different 50 uL portions with the addition of glycerol to a final concentration of 40 %. The final enzyme solution was frozen immediately by ethanol-dry ice mixture and stored into -80 °C freezer for future usage.

#### 4.12.2 The setting up of Sirt1 and Sirt5 enzymatic assay

The H3K4su-H4 tetramer was freshly assembled according to Part 10 at concentration of about 16 mg/mL or 180 μM. Sirtuin enzyme stock solution was added to each sample as the final concentration of 500 nM for Sirt1 or Sirt5. After 4 h incubation at 37 °C, the samples were subjected to the SDS-PAGE analysis. SDS-PAGE separated proteins were transferred to nitrocellulose membranes for Western blot. The pan anti-Ksu

antibody was used for the membrane blotting, followed by the Stripping-off of the antibody for probing with a generally used anti-H3 antibody (Abcam).

As shown in Figure 33, when Sirt5 and its cofactor NAD<sup>+</sup> were provided to the H3K4su-H4 solution, the succinylation was actively removed from H3. On the contrary, a control reaction with Sirt1 showed no succinylation removal from H3. These combined results unequivocally approve show that Sirt5 but not Sirt1 is a histone desuccinylase.



**Figure 33 The setting up of Sirt1 and Sirt5 enzymatic assay of H3K4su-H4 tetramer**  
Adapted with permission from [99].

#### 4.13 Summary

To summarize, we reported a versatile chemical biology approach for the site-specific installation of various Lys acylations into proteins. With the combination of ncAA and traceless-Staudinger ligation reaction, Lys acetylation and succinylation were successfully installed into proteins in a site-specific manner. Many proteins in many different organisms undergoing Lys succinylation are metabolic enzymes. Lys succinylation is a recently discovered Lys acylation found and profiled in many different species such as *E. coli*, yeast (*Saccharomyces cerevisiae*), human cells, mouse liver tissues, and even human pathogens as *Mycobacterium tuberculosis*<sup>[107]</sup>. These acidic modification

modes share one key feature as the reversion of charge states for Lys. Especially happen on histone, Lys succinylation can repulsion charge interaction between histone and DNA to loosen the core nucleosome structure. It still remains to be studied whether cells choose Lys succinylation as a powerful weapon to regulate the chromatin structure and gene transcription function, and whether Lys succinylation is regulated by Sirt5.

To address these interesting questions, we can first install Ksu at a desired position in proteins with the new strategy we developed. And with the combination of typical biochemistry or chemical biology techniques, we can further study the role of Lys succinylation and its regulation. In addition, the strategy provides an avenue to a myriad of other novel acylation PTMs. Although we did not carry out any test for the other two acidic modifications as malonylation and glutarylation in the current study, but we are almost certain about it. For fatty acylations not explored here such as myristoylation, the method we reported needs to be further improved to increase the reaction yield and water-solubility. On the one hand, although surprisingly the phosphinothioester reagents performed much better than the phosphinophenyl ester, there is still space for the optimization of reaction yield by attempting to avoid the hydrolysis side reaction. On the other hand, the water-soluble version of phosphinothioester reagents can be applied with high concentrations to accelerate the reaction<sup>[104]</sup>. To design and add different substitution groups to both phenyl groups from a physical organic chemistry point of view<sup>[108]</sup> will be a potential way to increase the reaction efficiency and to reduce the side reaction of N-P ylide. We are looking forward the advance of new generations of phosphinothioester reagents.



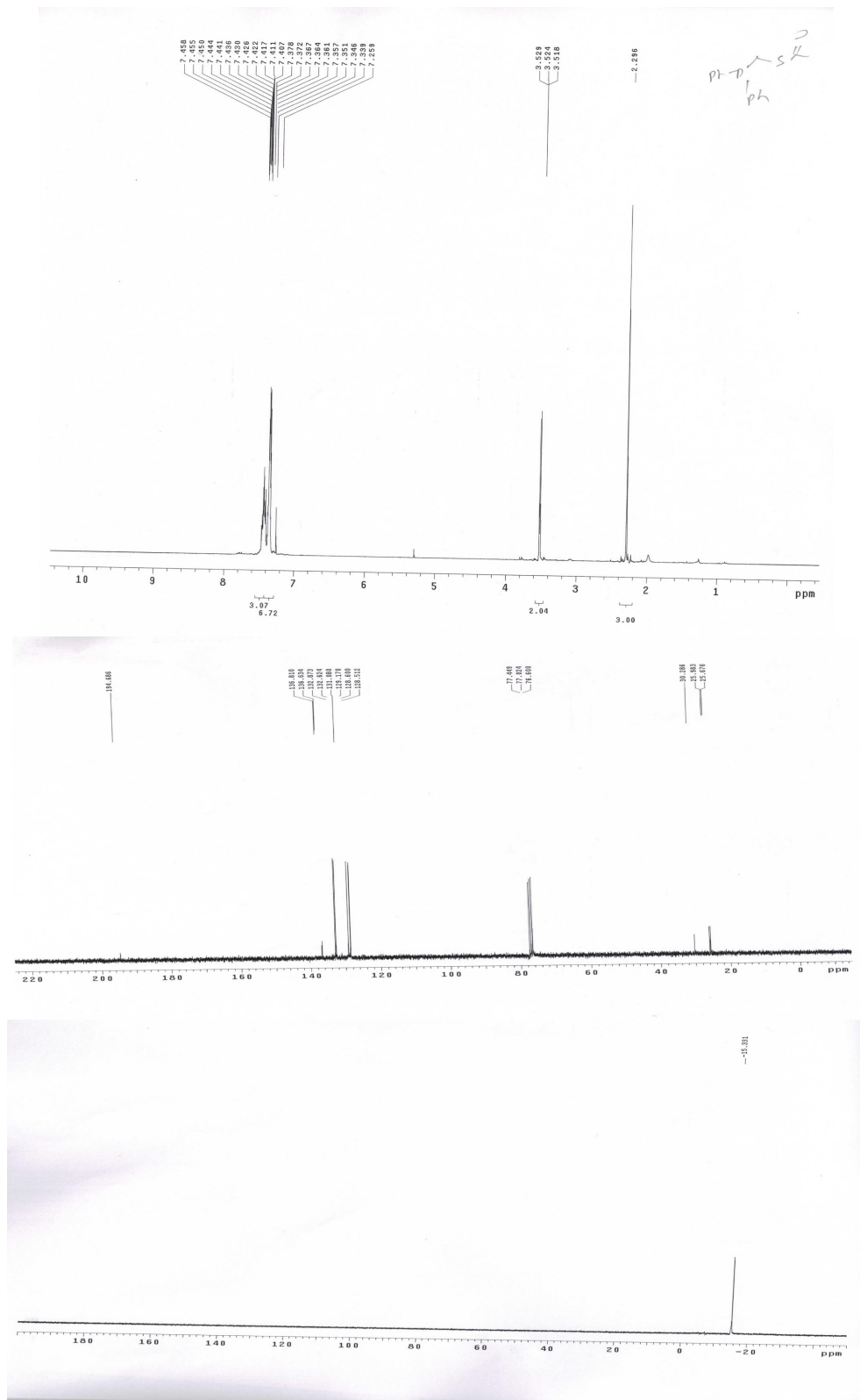
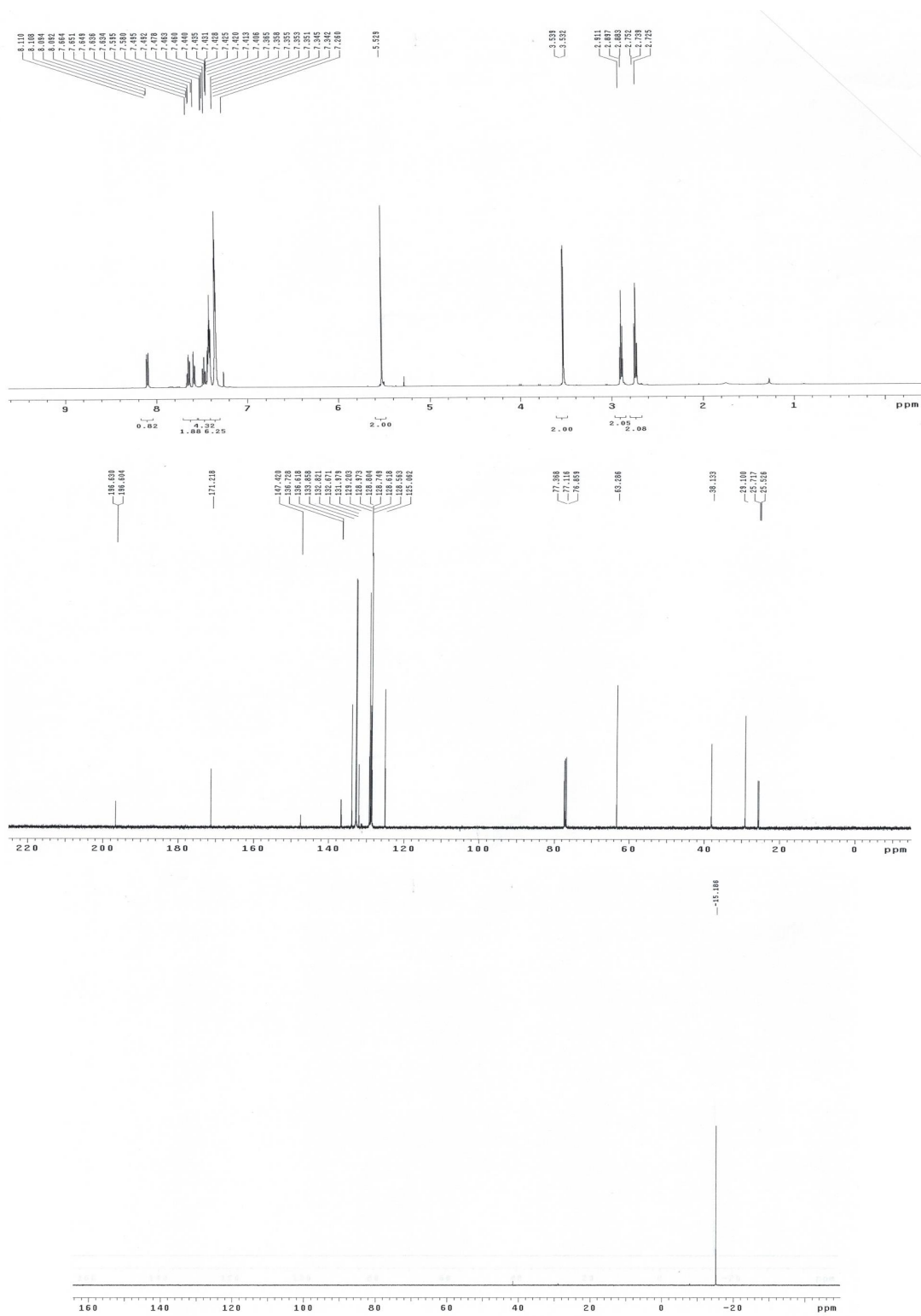


Figure 34 <sup>1</sup>H NMR, <sup>13</sup>C NMR and <sup>31</sup>P NMR of dPPMT-Ac





CHAPTER V  
THE SYNTHETIC INVESTIGATION OF LYSINE ACYLATION VIA KAHN  
REACTION

### 5.1 Background

The synthetic study of proteins with site-specific acyllysine installed is of great importance to life science study. However, none of the current strategies is ideal or robust enough when facing novel Lys PTMs with various structures. Advantages and disadvantages have been analyzed in detail for several chemical biology methods in the previous parts, such as the enzymatic synthesis, protein total and semi-synthesis, bioorthogonal reactions, direct incorporation via amber codon suppression, etc. One new pace to provide potential solution is to combine the noncanonical amino acid incorporation with protein chemistry. A typical example reported is the installation of AznL for the following traceless Staudinger ligation to generate acyllysine. By using this method, our group has shown the preparation of different proteins with site-specific Kac and Ksu, such as water-soluble protein Ub and non-water-soluble protein histone. The protein product can be used for the assembly of H3-H4 tetramer in vitro, and further applied for enzymatic assay directly.

Although as the first installation of Ksu at a specific site into proteins, the approach suffers from several disadvantages that may significantly hinder its further usage. Due to a slow kinetics of traceless Staudinger ligation reaction, the conversion from AznL to acyllysine takes at least 48 h in order to reach decent yield. More importantly, the reductive essence of the phosphinothioester reagents used in it can cause the Staudinger reduction

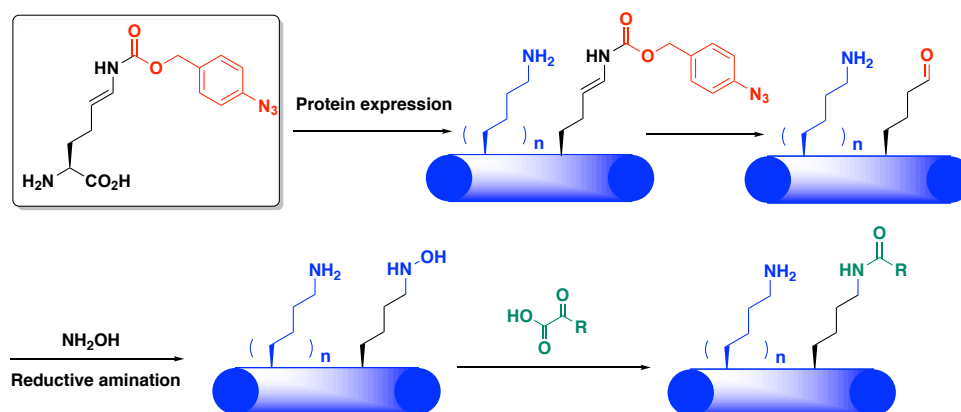
as an inevitable side reaction for the generation of native Lys with a free amino group. Although Lys may not give out signal under western blot by using PTM antibodies, the heterogeneous nature of the product can still raise concern for further biological studies. Last but not the least, the synthesis of different acyl phosphinothioester reagents requires special expertise, which is time consuming and tedious. Therefore, we are still looking forward to the development of new synthetic approaches for the incorporation of acyllysine to overcome those challenges.

## 5.2 The Design of acylation from AcdK

One direction for the development of protein chemistry is to introduce new reactions from organic chemistry exploration or peptide chemistry onto proteins<sup>[109]</sup>. Literature study showed a novel ligation reaction between  $\alpha$ -ketoacids and *N*-alkylhydroxylamine reported by the Bode group<sup>[110]</sup>, which has been used for the peptide ligation under mild aqueous condition<sup>[111]</sup>. Instead of non-substituted type-I *N*-alkylhydroxylamine, type-II *N,O*-dialkylhydroxylamine or its cyclic derivatives could also undergo the so-called KAHA reaction and showed better ligation yield as well as higher tolerance against water<sup>[112]</sup>. They further applied the ligation product with an aspartyl aldehyde group for further functionalization, such as reacting with hydroxylamine for protein labeling or reductive amination to generate hydroxylamine side chain. The product as hydroxylamine-containing side chain can undergo another KAHA ligation with external small molecular ketoacids<sup>[113]</sup>.

Inspired by this approach, we came up with a combinatorial methodology of KAHA ligation with our ncAA technique of AcdK. As previously mentioned, AcdK as an allysine precursor can undergo reductive amination to generate Lys alkylation after being incorporated into POIs. Kme and Kme2 are both examples to demonstrate the application. By simply changing the methylamine to hydroxylamine, we are able to generate a similar hydroxylamine-containing side chain to undergo KAHA ligation and finally generate acyllysine.

We expect this new protein chemistry approach to endow several advantages comparing with the traceless-Staudinger ligation strategy. First, the  $\alpha$ -ketoacid reagents used in KAHA are easy-accessible from commercial resources. Second, there is no known reported side reaction between ketoacids with the native amino acid side chains. For instance, the  $\alpha$ -ketoglutarate as the only reactant for Lys succinylation is purchasable from Sigma-Aldrich, which is also a widely existing natural product for the transamination procedure of glutamate. Similar to the traceless Staudinger ligation approach, we can generate theoretically any different acyllysine PTMs by simply changing the  $\alpha$ -ketoacid moiety.



**Figure 37 The combination of AcdK incorporation and KAHA ligation reaction for the installation of Lys acylation**

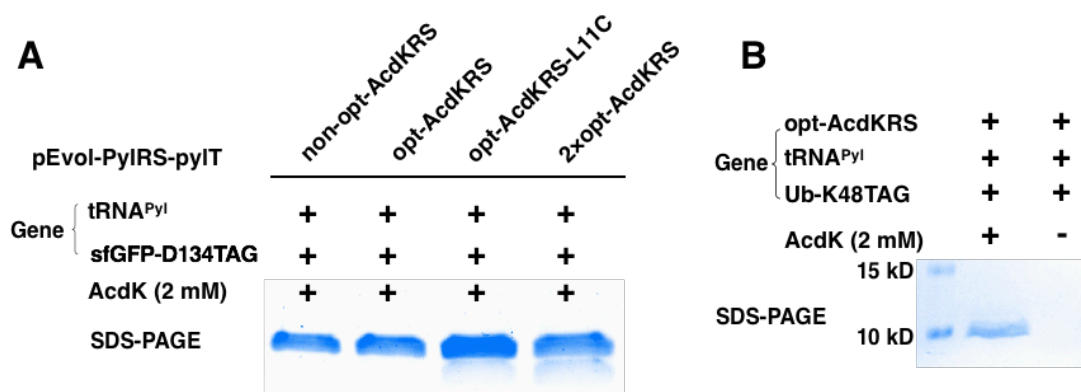
### 5.3 The optimization of AcdKRS plasmid

#### 5.3.1 The cloning of opt-AcdKRS plasmid

Before our investigation on protein chemistry part, we hope to optimize the PylRS mutant for better incorporation. The selected original plasmid for AcdK, pEvol-non-opt-AcdKRS, was codon optimized to generate pEvol-opt-AcdKRS. Three site-directed mutagenesis procedures are carried out sequentially to introduce three mutations to the pEvol-opt-WT-PylRS plasmid as L309T/D348G/Y384F. Phusion DNA polymerase was used for the PCR procedure. Next, we make the insertion of L11C into the optimized plasmid to be pEvol-opt-AcdKRS-L11C. Finally, a second copy of opt-AcdKRS was installed to the pEvol-opt-AcdKRS yielding pEvol-2×opt-AcdKRS.

Each of the three constructed plasmids, together with pBAD-sfGFP-D134TAG, was then used to co-transform *E. coli* BL21(DE3) cells. A single colony was picked and let to grow in 50 mL 2YT culture for 8 h, induced by 1 mM IPTG, 0.2% arabinose and 2 mM AcdK after OD reached 0.4. The sfGFP from each sample was purified from via 1

mL Ni Sepharose™ 6 Fast Flow column (GE Healthcare) and the elution was concentrated. The purified protein was analyzed and quantified by SDS-PAGE with Commassie staining (Figure 38A). The plasmid pEvol-opt-AcdKRS-L11C shows the best expression with an improvement of 15 % comparing with the original non-optimized strain, which was further used for protein expression.



**Figure 38 The optimization of AcdKRS by the comparison of sfGFP expression with different optimized AcdKRS mutants (A) the GFP expression with optimized AcdKRS; (B) SDS-PAGE of the expression of Ub-K48AcK.**

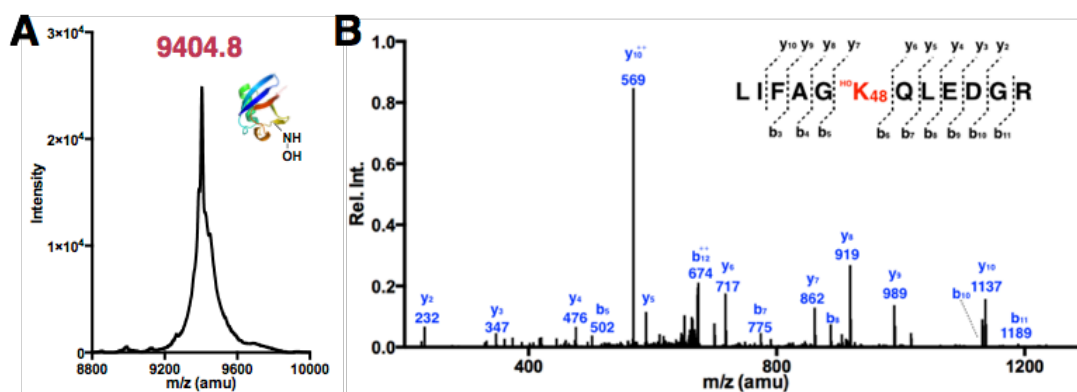
### 5.3.2 The expression of UbK48AcK

We decided to test the reaction condition with the AcdK-containing Ub at K48 position (Ub-K48AcK). The best optimized plasmid pEvol-opt-AcdKRS-L11C was used to co-transform *E. coli* BL21(DE3) cells together with pEtduet-UbK48TAG with a C-terminal His<sub>6</sub> tag. Ub-K48AcK was expressed and purified following a similar process as Ub-K48AznL (Figure 38B).



#### 5.4 Methodology investigation of the reductive amination with hydroxylamine

After the routine conversion into Ub-K48A1K by incubating with 5 mM TCEP, a two-step reductive amination procedure was carried out according to the reported literature<sup>[113]</sup>: Hydroxylamine hydrochloride salt was first dissolved in water to make 100 mM stock solution. 10 mM hydroxylamine was then added to the solution of Ub-K48A1K. The solution mixture was let to incubate with hydroxylamine for 4 h at 37 °C, and then dialyzed to 100 mM oxalic acid buffer pH 2.5 containing 10 mM hydroxylamine. Because the reduction process needs to undergo in acidic condition, the pH of the solution was measured with micro-pH meter to guarantee the acidity. Then, the reaction solution was treated by excess 20 mM NaBH<sub>3</sub>CN in two batches with an interval of 1 h. Then, the final reaction mixture was let to incubate at 37 °C for o/n. After been dialyzed to pure water to completely remove the small molecular reactants, the hydroxylated lysine product Ub-K48K<sub>OH</sub> was dried with lyophilizer to give out its powder form for the next step. Before that, MALDI-TOF-MS was used to analyze the product, which gave out a single peak at 9404 Da, agreeing with the calculated hydroxylated product after reductive amination (Figure 39A). Notably, Figure 39A may contain some side peaks that are not fully studied. Similar to our previous researches, MS/MS was used to analyze the trypsinized product Ub-K48K<sub>OH</sub>, and a 48K<sub>OH</sub> containing peptide fragment was detected (Figure 39B).



**Figure 39** The reductive amination of hydroxylamine on UbK48AIK (A) MALDI-TOF-MS spectrum of product Ub-K48K<sub>OH</sub>; (B) MS/MS analysis of 48K<sub>OH</sub>-containing trypsinized peptide fragments.

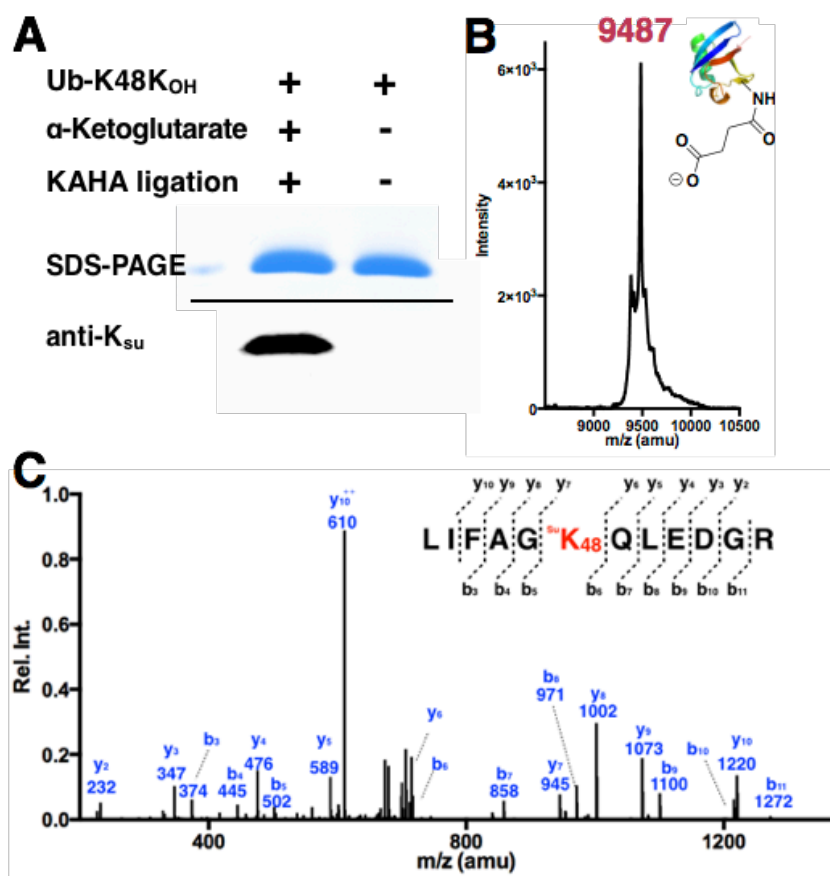
## 5.5 Methodology investigation of the KAHA ligation reaction with $\alpha$ -ketoglutarate

### 5.5.1 The reaction setting up for Lys succinylation

After the successful conversion of Ub-K48AIK to Ub-K48K<sub>OH</sub>, we start the investigation of KAHA ligation reaction with  $\alpha$ -ketoglutarate, and its generation of Ksu can be easily compared with the traceless-Staudinger ligation strategy. Because type-I *N*-alkylhydroxylamine needs to react under high DMSO condition, the Ub-K48K<sub>OH</sub> powder was directly used for the reaction setting up. A reaction solution was prepared by dissolving a final concentration of 200 mM oxalic acid and 200 mM  $\alpha$ -ketoglutarate in a pH 2.0 solvent mixture with 9:1 DMSO:H<sub>2</sub>O. Ub-K48K<sub>OH</sub> in its powder was dissolved in the reaction mixture to reach a final protein concentration of  $\sim$ 20  $\mu$ M, followed by the gentle pipetting to make the reaction solution well-mixed. After nitrogen protecting, the reaction tube was incubated under 37  $^{\circ}$ C for 36 h.

### 5.5.2 The reaction working up and characterization

The reaction was quenched by a 1:1 volume ratio water dilution, and potential protein precipitate was removed by centrifuging at 14K for 5 min. The supernatant was then dialyzed to a 20 mM phosphate pH 7 buffer. After mixing the sample with the SDS loading buffer, the final product Ub-K48Ksu was analyzed by 15 % SDS-PAGE gel for Coomassie blue staining and western blot analysis. The first gel showed clear bands for both samples with or without reaction. Pan-anti-Ksu antibody was used to blot the nitrocellulose membrane after the protein transfer from gel to membrane. After blotting with secondary antibody and visualization, a strong band could be observed for the reaction group while none for the one without reaction (Figure 40A). Then, we sent the product sample Ub-K48Ksu for MALDI-TOF-MS analysis after the Ziptip (Reversed-phase, pipette tips for sample preparation, Millipore Corporation) sample preparation process., and the result showed a single major peak is 9487 Da, same as the calculated molecular weight (Figure 40B). No other significant peaks can be seen besides a slight amount of unreacted starting material. The existence of lysine succinylation at K48 site was further confirmed by MS/MS analysis (Figure 40C).



**Figure 40 Methodology examination of Lys succinylation** (A) after reacting with  $\alpha$ -ketoglutarate, the protein product Ub-K48K<sub>Su</sub> was analyzed by SDS-PAGE and the Western blot probed by a pan-anti-succinyllysine antibody; (B) MALDI-TOF-MS spectrum of the KAHA reaction product as Ub-K48K<sub>Su</sub>; (C) MS/MS analysis of the trypsinized 48K<sub>Su</sub>-containing trypsinized peptide fragments.

### 5.5.3 Analysis

With the test of succinylation on Ub, we want to compare the KAHA strategy with traceless-Staudinger ligation reaction for the purpose of acyllysine installation especially for K<sub>Su</sub>, the KAHA ligation reaction approach can reach to higher conversion within shorter time period. Additionally, multiple advantages stand out for the ketoacid reactant used here over the phosphinothioester reagents, such as the easy-accessibility, no side-reaction, higher stability against water/air for long-term storage, etc.

## 5.6 Methodology investigation of the KAHA ligation reaction with oxoadipic acid

### 5.6.1 The reaction setting up for Lys glutarylation

Encouraged by the result of Lys succinylation initial trial by the KAHA ligation reaction approach, we wanted to test another Lys PTM, Lys glutarylation, which has not been reported yet. Oxoadipic acid was purchased from Sigma-Aldrich as the corresponding  $\alpha$ -ketoacid reagent, and the reaction was set up according to a similar process as described in the previous part. To reach a decent yield for the reaction, we carried out a concentration scanning for both oxalic acid and oxoadipic acid. With a complicated mechanism for the KAHA ligation reaction<sup>[114]</sup>, the optimized concentration for glutarylation is with to dissolve 200 mM oxalic acid and 400 mM oxoadipic acid in 9:1 DMSO:H<sub>2</sub>O solvent mixture to ensure a strong acidic reaction condition pH 2.3. The Ub-K48K<sub>OH</sub> was dissolved to the reaction mixture in its powder form and the reaction was let to incubate at 37 °C for 36 h under the nitrogen protecting condition.

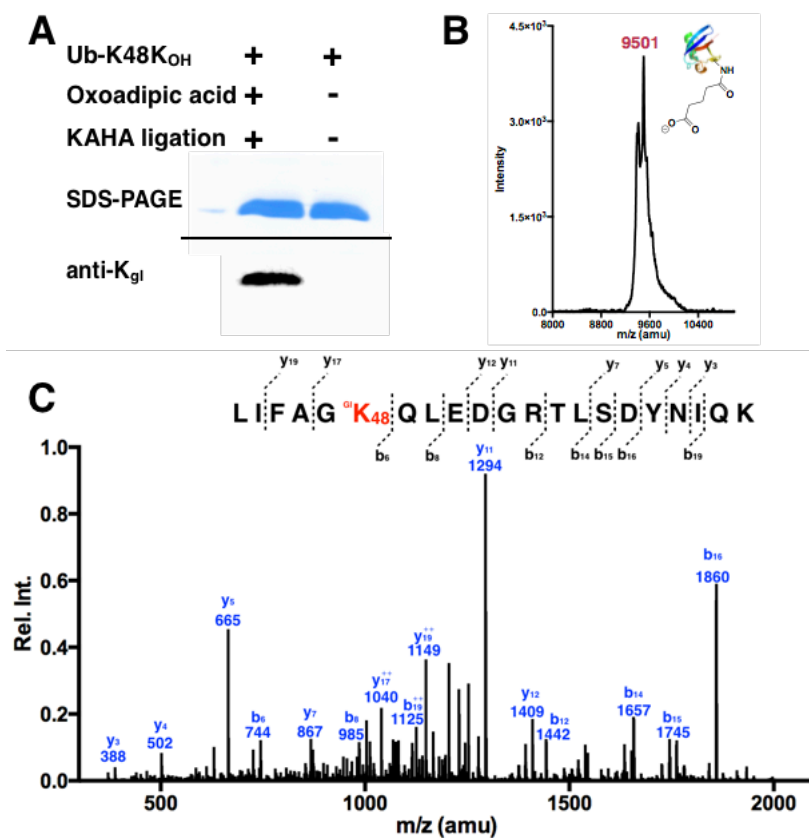
### 5.6.2 The reaction working up and characterization

After being quenched by the water dilution with a 1:1 volume ratio, the reaction product was analysed by SDS-PAGE. Western blot analysis probed by pan-anti-Kgl antibody showed a strong band for the reaction group while none for the control without reaction (Figure 41A). MALDI-TOF-MS was then carried out to test the conversion of Ub-K48K<sub>OH</sub> to Ub-K48Kgl, and the only major peak 9501 Da agreed with the calculated

molecular weight of Ub-K48Kgl (Figure 41B). Notably, some minor peaks can be seen representing unreacted starting material and unknown impurities. The product sample Ub-K48Kgl was trypsinized and then applied for MS/MS analysis. A peptide fragment with 48Kgl was found to guarantee the site-specific installation of lysine glutarylation (Figure 41C).

### **5.7 KAHA ligation reaction on H3K9**

After the success of Lys succinylation and glutarylation on Ub, we wanted to apply this KAHA ligation approach to install either Ksu or Kgl into H3K9 forming H3K9Ksu or H3K9Kgl, respectively. Both of them have been found on human histone proteins. Comparing with the common Lys acetylation, succinylation and glutarylation can reverse the charge state of lysine from positive to negative, which can greatly impact the interaction between histone octamer and DNA. We have shown the Ksu could be regulated by Sirt5, but no direct data provided for Kgl. The products we got from our approach may be directly applied for Sirt5 activity assay.



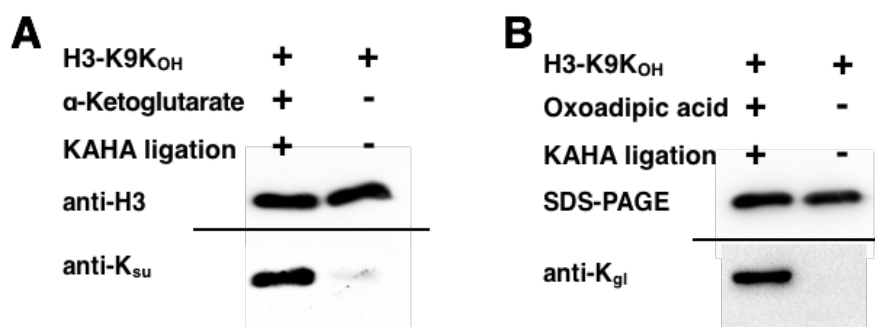
**Figure 41 Methodology examination of Lys glutarylation** (A) After reaction with oxoadipic acid, the protein product Ub-K48K<sub>gl</sub> was analyzed by SDS-PAGE and the Western blot probed by a pan-anti-glutaryllysine antibody; (B) MALDI-TOF-MS spectrum of the KAHA reaction product as Ub-K48K<sub>gl</sub>; (C) MS/MS analysis of the trypsinized 48K<sub>gl</sub>-containing trypsinized peptide fragments.

The expression and purification of H3K9AIK was following a similar process as described before. To install AcdK into H3K9 position, the optimized plasmid pEvol-opt-AcdKRS mentioned above was used to co-transform *E. coli* BL21(DE3) cells, together with pETduet-H3K9TAG. After the elution from 1 mL Ni Sepharose™ 6 Fast Flow column (GE Healthcare) was collected and concentrated, TCEP was added to a final concentration of 5 mM and incubated for 20 min. The final protein H3K9AIK was

exchanged to histone reaction buffer (6 M urea, 20 mM Phosphate, pH 7.4) and stored for further usage.

A similar two-step reaction procedure was then applied to the stored H3K9AIK. After been incubated with hydroxylamine solution at 37 °C for 4 h, the reaction buffer was then dialyzed to a pH 2.5 100 mM oxalic acid buffer containing 6 M urea and 10 mM hydroxylamine. Next, 20 mM NaBCNH<sub>3</sub> was added in two batches with an interval of 1 h, and the final reaction solution was incubated at 37 °C for o/n. The reaction mixture was dialyzed to pure water, lyophilized to remove all the solvent and stored as H3K9K<sub>OH</sub> in its powder form. A 9:1 DMSO:H<sub>2</sub>O reaction solution pH 2.0 with 200 mM oxalic acid and 200 mM  $\alpha$ -ketoglutarate for succinylation, or 400 mM oxoadipic acid for glutarylation was used to dissolve H3K9K<sub>OH</sub>, separately. The reaction mixture was well mixed and let to incubate under 37 °C for 36 h under the nitrogen protecting condition. After quenching by water and dialyzed to pure water, the reaction solution containing H3K9K<sub>su</sub> or H3K9K<sub>gl</sub> was applied to SDS-PAGE gel and then Western blot. The pan-anti-K<sub>su</sub> antibody or pan-anti-K<sub>gl</sub> antibody followed by the anti-H3 antibody from PTM BioLabs Inc. was applied. The visualization showed a strong band for the sample after KAHA reaction, suggesting the successful installation of Lys succinylation (Figure 42A) and glutarylation (Figure 42B) as expected. The histone final products as H3K9<sub>su</sub> and H3K9<sub>gl</sub> were both saved for the further Sirt5 enzymatic assay.



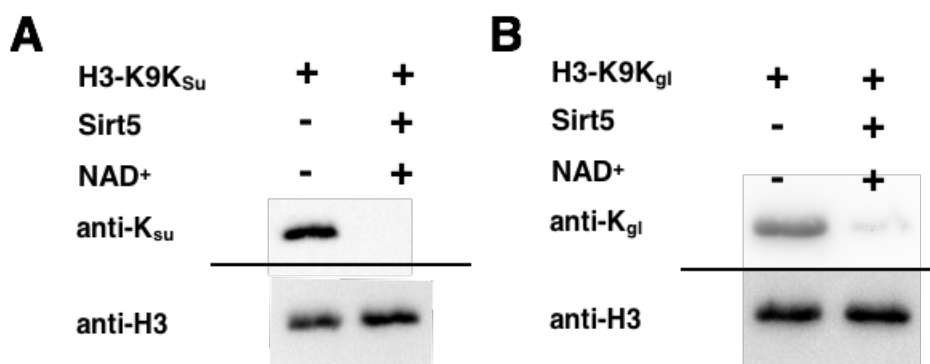


**Figure 42** After reaction with oxoadipic acid, the protein product Ub-K48Kgl was analyzed by SDS-PAGE and the Western blot probed by a pan-anti-glutaryllysine antibody (A) After reaction with  $\alpha$ -ketoglutarate, the protein product H3K9Ksu was analyzed by the Western blot probed by a pan-anti-succinyllysine antibody and an anti-H3 antibody; (B) After reaction with oxoadipic acid, the protein product H3K9Kgl was analyzed by the Western blot probed by a pan-anti-glutaryllysine antibody and an anti-H3 antibody.

### 5.8 Sirt5 assay over H3K9Ksu and H3K9Kgl

As a member of the NAD-dependent protein deacylase family, Sirt5 is a histone deacylase known to remove lysine succinylation over histone peptide substrates. We have shown that Sirt5 can catalyze desuccinylation process over the synthetic H3-H4 tetramer prepared by traceless-Staudinger ligation approach over AznL. But no enzymatic assay results *in vitro* have been reported for the deglutarylation activity over histone proteins. By using our KAHA ligation strategy, we can site-specifically install Ksu and Kgl into H3K9 site. The products stored in powder form were dissolved in 20 mM Tris-HCl pH 7.5 buffer containing 1 mM NAD<sup>+</sup> and 1 mM DTT, and 3  $\mu$ l of Sirt5 enzyme stock solution was added to each sample to reach a final concentration of 540 nM. The reaction mixture was let to incubate at r.t. for 4 h, and then applied to SDS-PAGE and western blot with pan-anti-Ksu antibody, pan-anti-Kgl antibody or anti-H3 antibody. After visualization,

lysine succinylation and glutarylation can both be removed completely from H3K9 (Figure 43A&B), which agrees with the previous reports for lysine glutarylation.



**Figure 43 Sirt5 enzymatic assay over H3-H4 tetramer** (A) Desuccinylation of the H3K9K<sub>Su</sub> by incubating H3K9K<sub>Su</sub> with or without 0.5  $\mu$ M Sirt5 in the presence of 1 mM NAD<sup>+</sup> and 1 mM dithiothreitol (DTT) for 4 h. After reaction, proteins were analyzed with western blot analysis by a pan-anti-succinyllysine antibody and anti-H3 antibody; (B) Desglutarylation of the H3K9K<sub>gl</sub> under the same condition. After reaction, proteins were analyzed with western blot analysis by a pan-anti-K<sub>gl</sub> antibody and anti-H3 antibody.

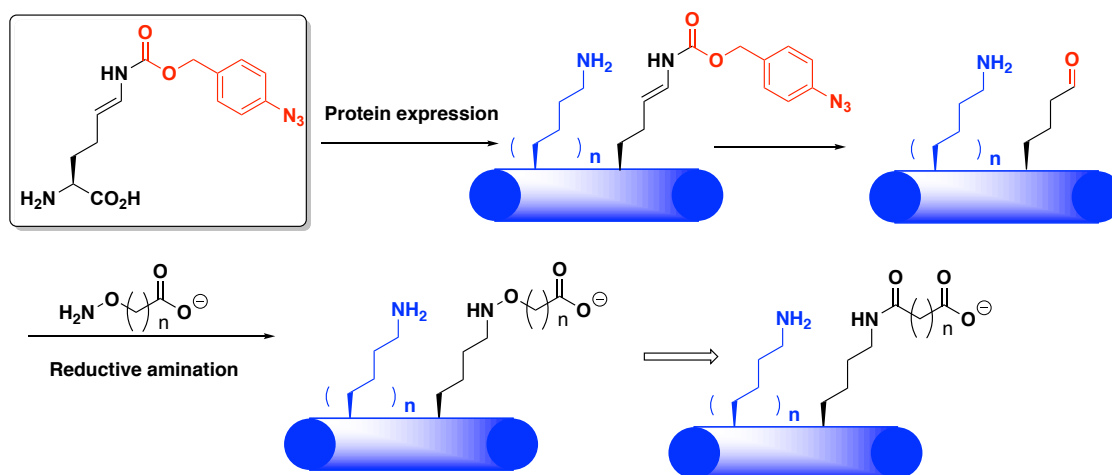
## 5.9 Other Lys acylation strategies investigation based on AcDK

In this part, several trial-and-error tests will be briefly discussed for a similar goal of methodology development for Lys acylation. Due to limited time and space, these unfinished projects will be briefly summarized mainly as proof of concepts. We feel them helpful because they are meaningful extension of our AcDK system based on rational design.

### 5.9.1 The succinylation analogs

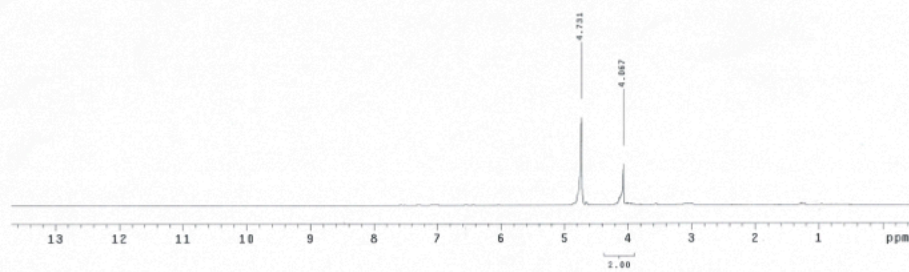
Although with the traceless Staudinger ligation reaction we can generate Lys acidic PTMs such as succinylation, the yield is not good due to the inevitable Staudinger

reduction. Additionally, the reaction in 40 % DMSO may raise the concern of protein unfolding. Another design was inspired by the high efficiency of reductive amination over hydroxylamine and its derivatives. After the conversion from AcdK to allysine, the protein can undergo reductive amination reaction to generate acidic PTM analogs, more specifically,  $n = 1$  for Kmal,  $n = 2$  for Ksu, and  $n = 3$  for Kgl.



**Figure 44 Scheme for the installation of acidic PTM analogs**

Std Proton parameters  
File: home/lls/zhang/vnmrsys/data/2016-11-25-NH2OCH2COOH-H.fid  
Pulse Sequence: s2pul  
Solvent: d2o  
Ambient temperature  
Operator: zhang  
File: 2016-11-25-NH2OCH2COOH-H  
INOVA-300 "Inova300"  
Relax. delay 1.000 sec  
Pulse 45.0 degrees  
Acq. time 1.938 sec  
Width 4789.5 Hz  
0 repetitions  
OBSERVE H1 299.8150507 MHz  
DATA PROCESSING  
FT size 32768  
Total time 0 min, 24 sec



Std Carbon experiment  
File: home/lls/zhang/vnmrsys/data/2016-11-25-NH2OCH2COOH-C.fid  
Pulse Sequence: s2pul  
Solvent: d2o  
Ambient temperature  
Operator: zhang  
File: 2016-11-25-NH2OCH2COOH-C  
INOVA-300 "Inova300"  
Relax. delay 1.000 sec  
Pulse 45.0 degrees  
Acq. time 1.300 sec  
Width 16985.2 Hz  
256 repetitions  
OBSERVE C13 75.4136895 MHz  
DECOUPLE H1 299.8165489 MHz  
Power 37 dB  
continuously on  
WALTZ-16 modulated  
DATA PROCESSING  
Line broadening 0.5 Hz  
FT size 43536  
Total time 9 min, 51 sec

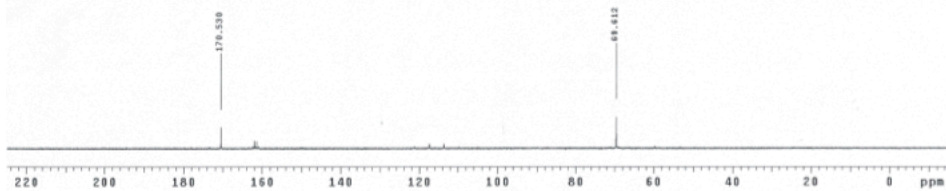
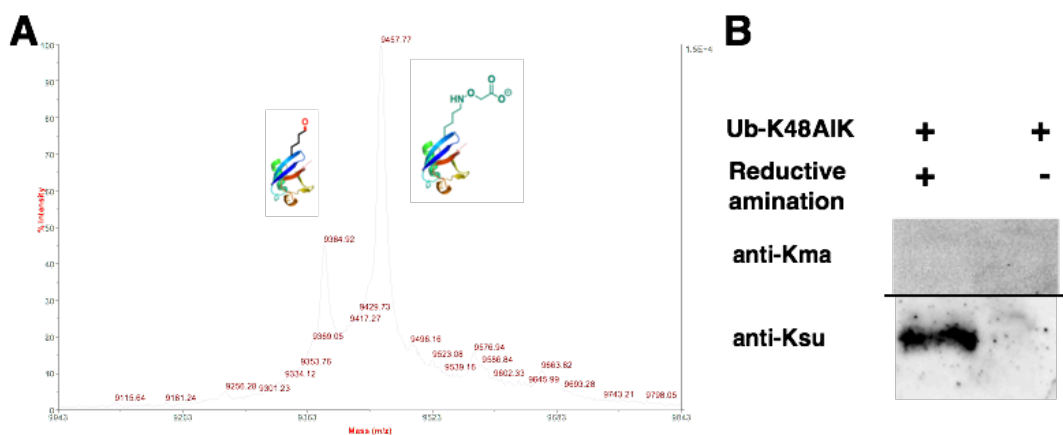


Figure 45 <sup>1</sup>H NMR and <sup>13</sup>C NMR of 2-(aminoxy)acetate

The 2-(aminoxy)acetate (n = 1) was synthesized according to a modified protocol, which is omitted here. After being analyzed by <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O) with δ 4.067 (s, 2H), and <sup>13</sup>C NMR (75 MHz, D<sub>2</sub>O) with δ: 170.5, 69.6, the product was applied for reductive amination reaction with UbK48AIK, which is further confirmed with MALDI-TOF (Figure 46A). Then we loaded the sample after reaction to SDS-PAGE followed by the Western blot analysis (Figure 46B). All the trials of Western blot by pan-anti-Kmal antibody turned out to be failure. However, to our surprise, a clear band could be detected by pan-anti-Ksu antibody.

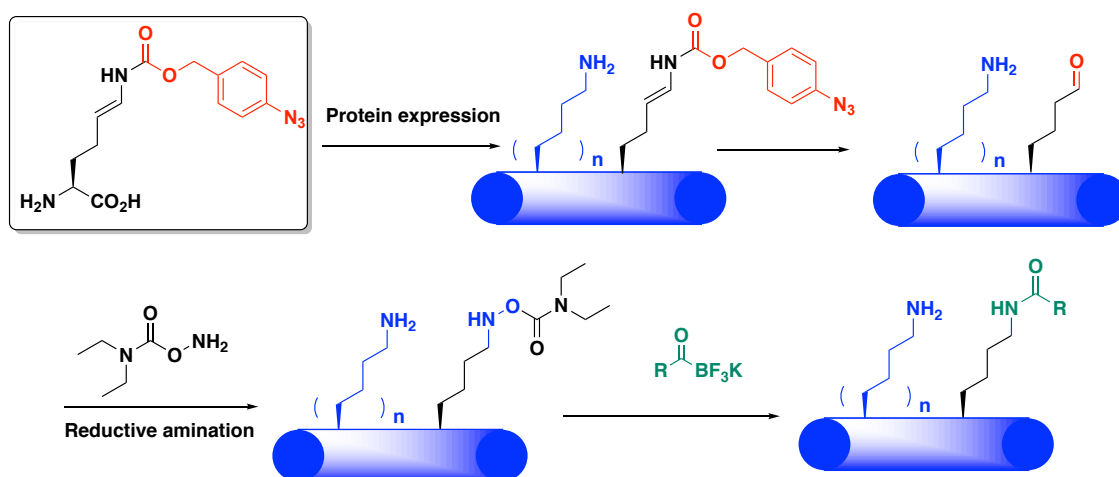


**Figure 46 Reductive amination investigation of acidic PTM analog** (A) the MALDI-TOF of the product after reductive amination with 2-(aminoxy)acetate; (B) Western blot analysis of the product after reductive amination with 2-(aminoxy)acetate.

### 5.9.2 The KAT reaction strategy

The KAHA ligation reaction discussed in this Chapter can generate theoretically any acylation modifications. However, the reaction needs near pure DMSO solution with a slow kinetics. Inspired by the Bode group for the newly discovered potassium acyltrifluoroborate (KAT) amide formation reaction<sup>[115]</sup> and its related reaction<sup>[116]</sup>, we

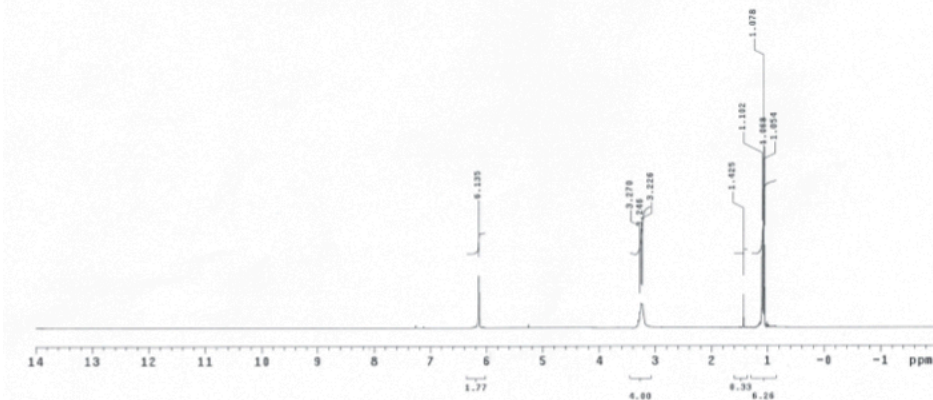
proposed an alternative as below (Figure 47). By replacing the type I hydroxylamine with acyl hydroxylamine, the reaction can undergo in water/butanol solvent mixture with a very fast kinetics.



**Figure 47 Scheme for the Lys acylation installation with KAT reaction**

To start with, *N*-((aminooxy)carbonyl)-*N*-ethylethanamine was synthesized according to a protocol not mentioned here. The product was analyzed by <sup>1</sup>HNMR (300 MHz, CDCl<sub>3</sub>) with  $\delta$  6.13 (s, 2H), 3.25 (m, 4H), 1.07 (m, 6H), and <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) with  $\delta$ : 170.5, 42.4, 40.9, 28.0, 26.4 (Figure 48). Surprisingly, the product *N*-((aminooxy)carbonyl)-*N*-ethylethanamine for reductive amination was very unstable. After the deprotection of phthalimide group, the product is prone to degradation even under r.t. to release diethylamine. And therefore, the trials of reductive amination with UbK48AIK turned out to be hydroxylamine adduct due to the removal of the acyl group.

Std Proton parameters  
 File: home/jlu/zwang/vnmrsys/data/2016-12-8-NH2OCONE12-H-final.fid  
 Pulse Sequence: s2ps1  
 Solvent: cdcl3  
 Ambient Temperature  
 Operator: zwang  
 File: 2016-12-8-NH2OCONE12-H-final  
 INOVA-300 "inova300"  
 Relax. delay 1.000 sec  
 Pulse 45.0 degrees  
 Acq. time 1.350 sec  
 Width 4739.5 Hz  
 3 repetitions  
 OBSERVE H1, 299.9142799 MHz  
 DATA PROCESSING  
 FT size 32768  
 Total time 9 min, 24 sec



Std Carbon experiment  
 File: home/jlu/zwang/vnmrsys/data/2016-12-8-NH2OCONE12-C-final.fid  
 Pulse Sequence: s2ps1  
 Solvent: cdcl3  
 Ambient Temperature  
 Operator: zwang  
 File: 2016-12-8-NH2OCONE12-C-final  
 INOVA-300 "inova300"  
 Relax. delay 1.000 sec  
 Pulse 45.0 degrees  
 Acq. time 1.350 sec  
 Width 18083.2 Hz  
 250 repetitions  
 OBSERVE C13, 75.4135057 MHz  
 DECOUPLE H1, 299.9157781 MHz  
 Power 32 dB  
 continuously on  
 WALTZ-16 modulated  
 DATA PROCESSING  
 Line broadening 0.5 Hz  
 FT size 65536  
 Total time 9 min, 51 sec

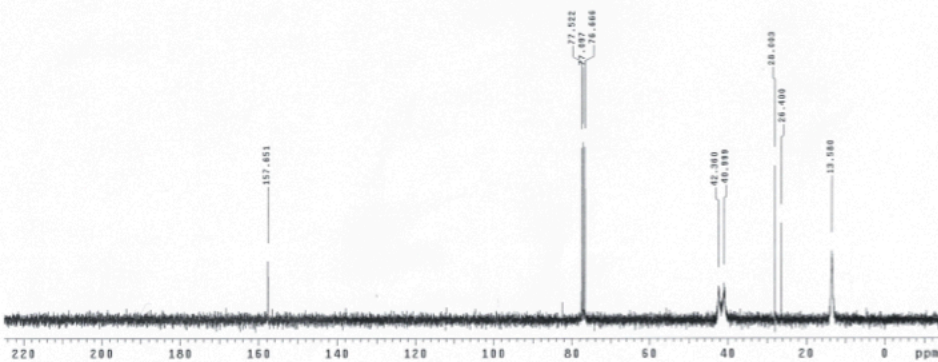
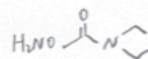
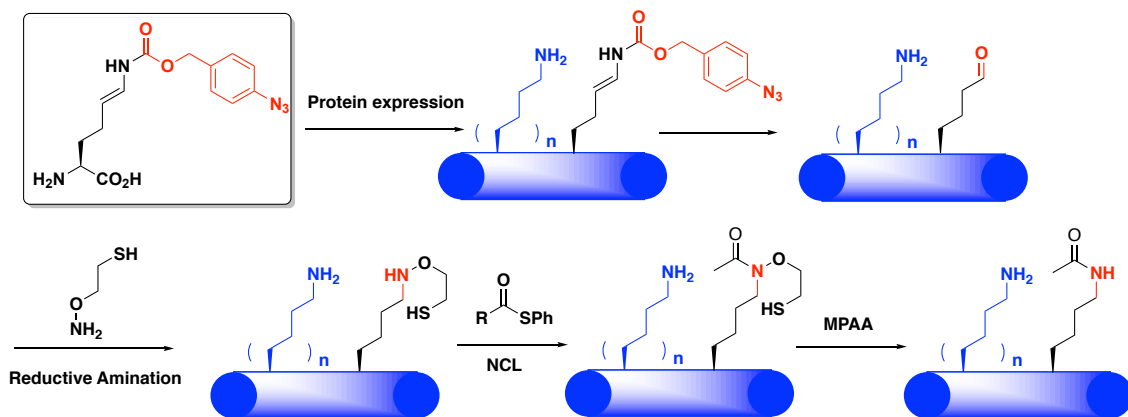


Figure 48 <sup>1</sup>H NMR and <sup>13</sup>C NMR of N-((aminoxy)carbonyl)-N-ethylethanamine

### 5.9.3 The N-O band cleavage strategy

The two rest approaches are all based on NCL as the most widely used reaction in protein chemical biology field. Inspired by the N-O cleavage reaction developed by the Chatterjee group<sup>[117]</sup>, we came up with the tendon reaction strategy. The UbK48AIK will first undergo reductive amination with 2-(aminoxy)ethanethiol to generate a  $\gamma$ -thiol group to facilitate the NCL with theatrically any acyl thioester. By using the 4-mercaptophenylacetic acid (MPAA), the N-O bond can be selectively cleaved to release native amide structures (Figure 49).

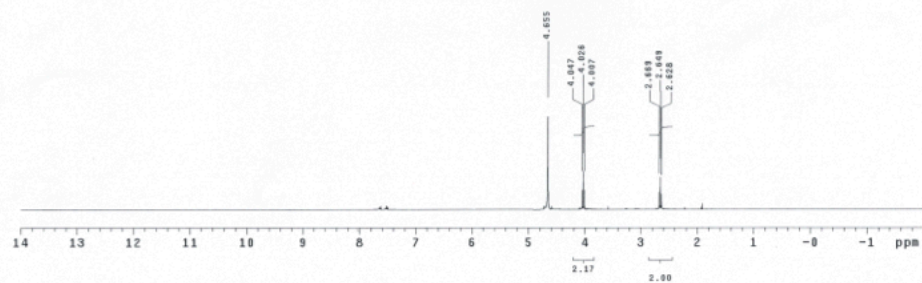


**Figure 49** Scheme of the Lys acylation installation with N-O cleavage strategy

We first synthesized 2-(aminoxy)ethanethiol with started phthalimide protection strategy. The product was then analyzed by  $^1H$ NMR (300 MHz,  $D_2O$ ) with  $\delta$  4.03 (t, 2H), 2.65 (t, 2H), and  $^{13}C$ NMR (75 MHz,  $CDCl_3$ ) with  $\delta$ : 76.1, 21.3 (Figure 50). The reductive amination was carried out with Ub-K48AIK and desired product can be produced. However, further optimization of the N-O cleavage is still needed after NCL.



Std Proton parameters  
 File: home/liu/zhang/vnmrsys/data/2017-2-16-NH2CH2CH2SH-H-f1n-2.fid  
 Pulse Sequence: s2pul  
 Solvent: d2o  
 Ambient temperature  
 Operator: zwang  
 File: 2017-2-16-NH2CH2CH2SH-H-f1n-2  
 INOVA-300 "inova300"  
 Relax. delay 1.000 sec  
 Pulse 45.0 degree  
 Acq. time 1.956 sec  
 Width 4788.5 Hz  
 3 repetitions  
 OBSERVE H1, 299.9150507 MHz  
 DATA PROCESSING  
 FT size 32768  
 Total time 9 min, 24 sec



Std Carbon experiment  
 File: home/liu/zhang/vnmrsys/data/2017-2-16-NH2CH2CH2SH-C.fid  
 Pulse Sequence: s2pul  
 Solvent: d2o  
 Ambient temperature  
 Operator: zwang  
 File: 2017-2-16-NH2CH2CH2SH-C  
 INOVA-300 "inova300"  
 Relax. delay 1.000 sec  
 Pulse 45.0 degree  
 Acq. time 1.380 sec  
 Width 18080.2 Hz  
 256 repetitions  
 OBSERVE C13, 75.4136985 MHz  
 DECOUPLE H1, 299.9150488 MHz  
 Power 37 dB  
 Continuously on  
 WALTZ-16 Modulated  
 DATA PROCESSING  
 Line broadening 0.5 Hz  
 FT size 85536  
 Total time 9 min, 51 sec

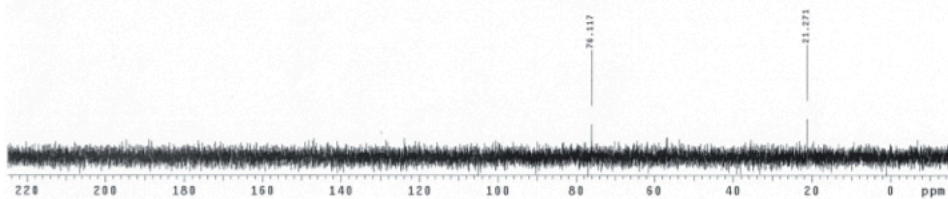
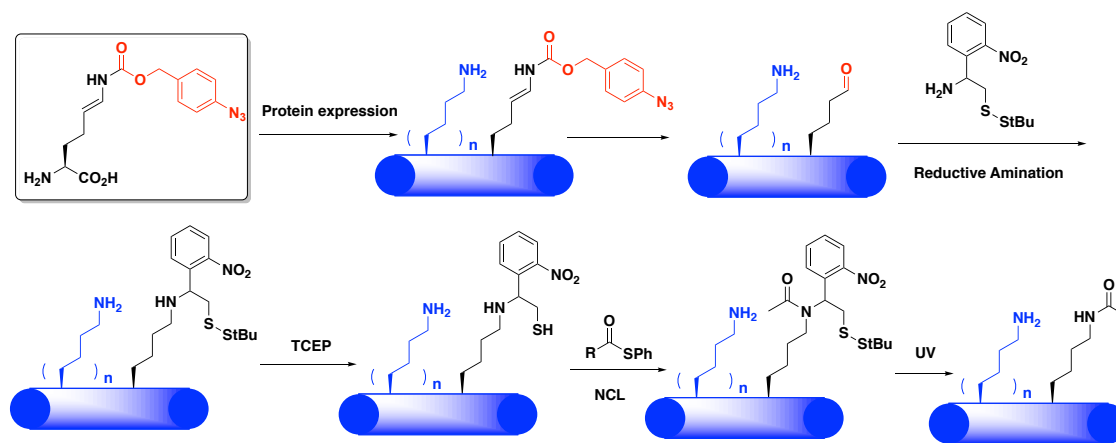


Figure 50 <sup>1</sup>H NMR and <sup>13</sup>C NMR of 2-(aminoxy)ethanethiol

#### 5.9.4 The auxiliary group-assisting strategy

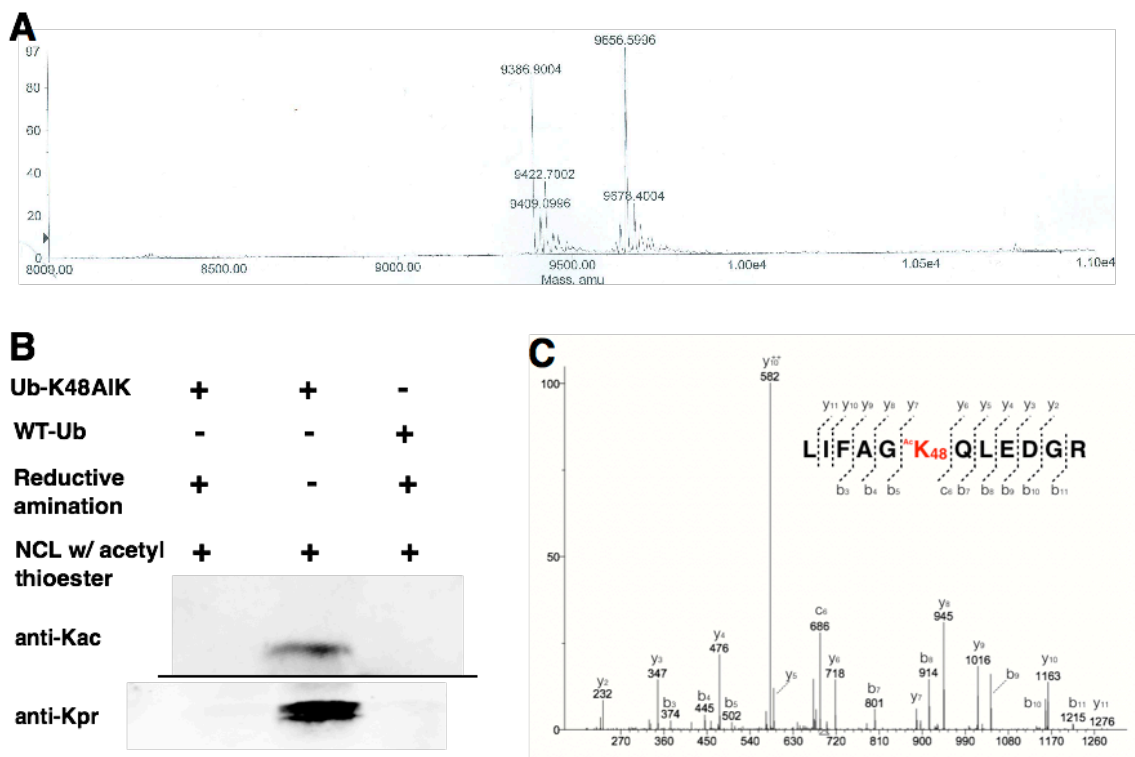
Another approach we came up with is to use an auxiliary group for to assist the NCL, which can be further removed by UV irradiation<sup>[118]</sup>. After the removal of TCEP for the conversion of AcdK to AIK, the amino compound with disulfide bond synthesized by Dr. Xinqiang Fang was used to do the reductive amination in water/DMSO solvent mixture. The disulfide bond was then cleaved by 10 mM TCEP to release the warhead for NCL. After undergoing NCL with different acyl thioester, the *O*-nitrophenol group can be removed under UV light for 1 h on ice (Figure 51).



**Figure 51** Scheme for the Lys acylation installation with the auxiliary group-assisting strategy

MALDI-TOF showed the desired reductive amination reaction as the peak of 9656 Da with about 60 % yield (Figure 52A). Then, we carried out the NCL with different acyl thioester. Both acetyl thioester and propionyl thioester synthesized by Dr. Xinqiang Fang were used for NCL. However, due to a long procedure with multiple reaction steps followed by dialysis after each reaction, a significant loss of final product made it hard to be visualized on MALDI-TOF. Both Western blot analysis showed a strong band only

with both reductive amination and NCL (Figure 52B). MS/MS further confirmed the existence of K48ac after the reaction (Figure 52C).

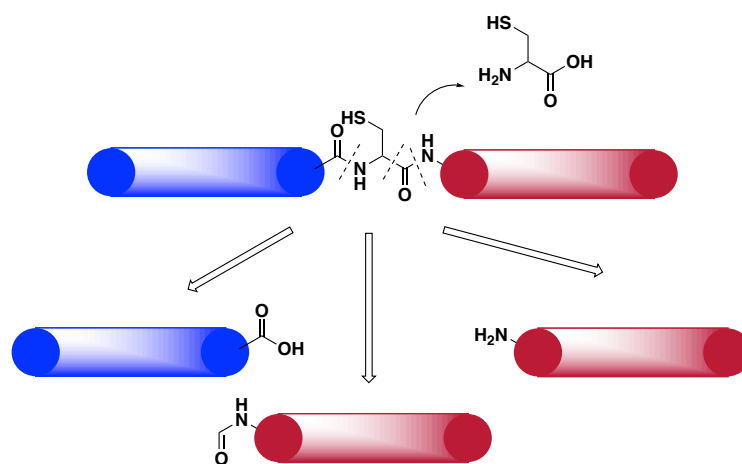


**Figure 52 Methodology investigation of the Lys acylation installation with the auxiliary group-assisting strategy (A) MALDI-TOF of reductive amination; (B) Western blot for the acetylation and propionylation; (C) MS/MS of acetylation product.**

### 5.9.5 The Ub cleavage with TCEP

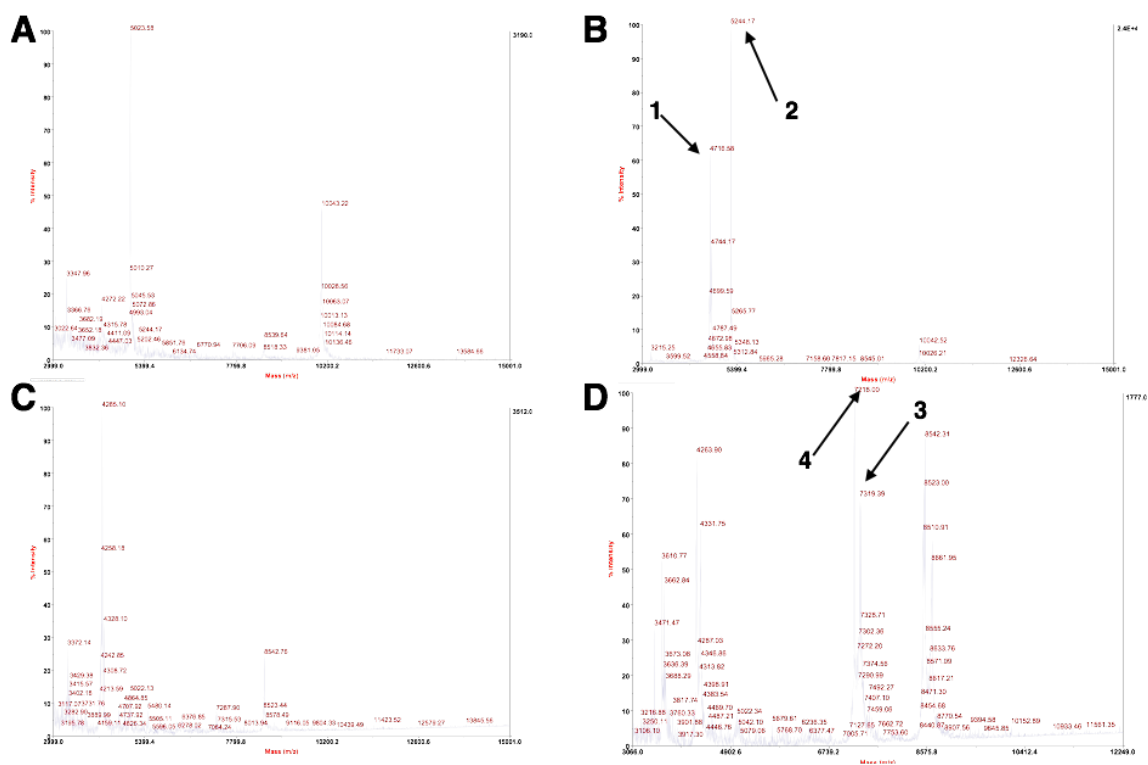
During the TCEP treatment of Ub with Cys mutations, I observed an unknown phenomenon (Figure 53). By incubating 20 mM TCEP with either UbK48C or UbK11C, a cleavage around the Cys position could be observed from MALDI-TOF. For UbK48C, the cleavage happened on both the C-site and N-site of Cys to completely eliminate Cys in between (Figure 54A&B). For UbK11C, the cleavage was not complete. Additionally, with a very complicated mechanism, the cleavage can also happen at the carbon-carbon

bond to leave a formylation instead of a free *N*-terminal amine. For UbK11C, a cleavage at the T12 site is also cleaved incompletely (Figure 54C&D). For both K11 and K48 sites, it happened to be -GC- sequence, which has been known in equilibrium with its thioester form<sup>[119]</sup>. When we extended the test to K27 and K29 sites, a similar cleavage process happened with different yields.



**Figure 53 Scheme for Ub with Cys mutations cleaved by TCEP**

We varied the reaction condition including reaction pH, buffer components, and temperature, but no change had been observed. However, with a decreased concentration of TCEP, the reaction yields significantly dropped. At a concentration of 5 mM, almost no reaction can be observed. We originally thought about a radical mechanism. However, the addition of 10 mM DTT could not protect the peptide to be chopped.



**Figure 54 MALDI-TOF of Ub with Cys mutations cleaved by TCEP (A) UbK48C with His<sub>6</sub> tag as MW 10044 Da before reaction (...FAG-C-QLE...); (B) UbK48C with His<sub>6</sub> tag was cleaved by 20 mM TCEP to generate fragment 1 4716 Da (QLE...), and fragment 2 5244 Da (...FAG), and the Cys was eliminated; (C) UbK11C without His<sub>6</sub> tag as MW 8542 Da before reaction (...LTG-C-TIT...); (D) UbK11C without His<sub>6</sub> tag was partially cleaved by 20 mM TCEP to generate fragment 3 MW 7319 Da (TIT...) and a cleavage at T site to generate fragment 4 MW 7218 Da (IT...), while the C-terminal fragment was too small to be detected.**

## 5.10 Summary

To summarize, with the combination of amber codon suppression and KAHA ligation reaction, a novel approach has been reported for the site-specific lysine acylation in POIs. The goal of Lys succinylation and glutarylation have been achieved with this approach in both ubiquitin as a water-soluble protein and histone proteins from inclusion

body. The acylated protein products can be directly applied for further enzymatic desuccinylation or deglutarylation assays.

Ksu and Kgl are recently discovered new lysine acylation types, which have been shown widely existing in the proteomes of several different species<sup>[120]</sup>, however, their related PTM enzymes “Writer, eraser, and reader” are remaining largely unknown. A recent paper reported a potential histone H3 succinyltransferase<sup>[22]</sup>, but no literature exists for the identification of lysine glutaryltransferase. With the current method in this part, many lysine succinylation and glutarylation related questions can be further addressed.

Additionally, since our “AcdK to allysine” system provides a very important chemical biology handle as aliphatic aldehyde group, which can be used to achieve many protein chemistry designs. To demonstrate its potential applications, we listed several trials or discoveries in the end of this chapter. Although some of them may not work so far, we are still looking forward to further optimization or condition screening. We are also looking forward to applying more chemical reactions from glass beakers<sup>[73]</sup> to protein contents<sup>[109]</sup>, and to a broader biological substrates<sup>[121]</sup>.

## CHAPTER VI

### THE CHEMICAL LABELING AND INVESTIGATION OF LYSINE LIPIDATION

#### 6.1 Background

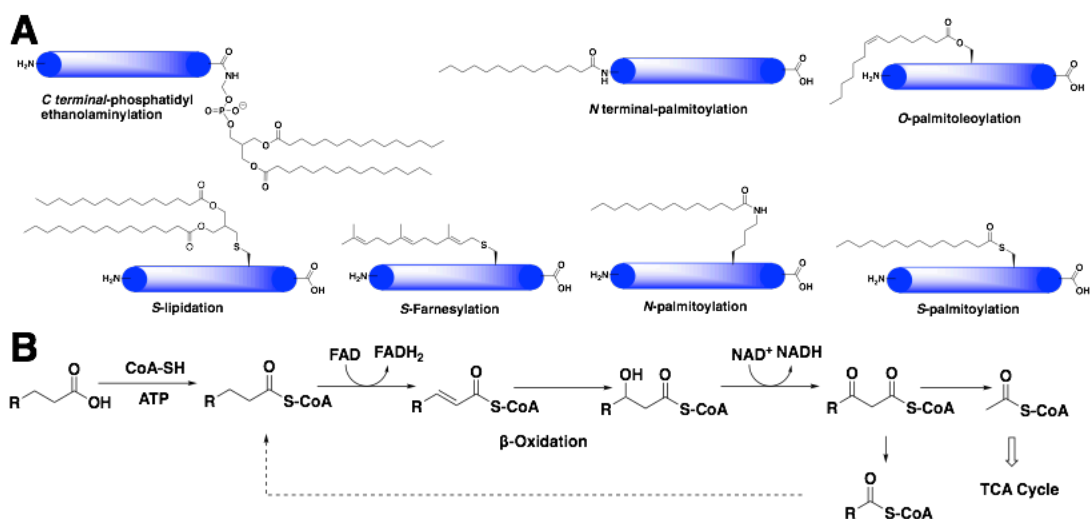
Besides the Lys mono-/di-methylation discussed in Part III and Lys acylation in Part IV/V, lipidation is another important Lys modification mode widely existing with high structural and functional diversity. In this chapter, we hope to apply our ncAA technique combining with the chemical biology and biochemistry methods to study Lys fatty acylation modes from a proteomic view.

Lipids we intake can modify our proteome with various structures, such as the C-terminal phosphatidylethanoamination and S-farnesylation. Another main group of protein lipidation is fatty acylation, which can be divided as *N*-terminal modification, *N*-modification on Lys, *S*-modification on Cys, *O*-modification on Ser or Thr four groups (Figure 55A).

In daily diet, we absorb a high amount of fatty acids from all kinds of food, including livestock and poultry meats, eggs, fruits, vegetables, nuts, milk and other dairy products, etc<sup>[122]</sup>. Dietary fatty acids are saturated or unsaturated aliphatic acids with short, medium, and long hydrocarbon chains. Typically, the fatty acids in food mentioned above are rich in fatty acids with variable total carbon number from 3 to 23. As a main family of lipids, fatty acids with long hydrocarbon chains own carbon atoms in low oxidation states and thus with very high energy comparing with other biomolecules such as saccharin and protein<sup>[42]</sup>.

After being taken into human body, some of the dietary fatty acids are used as building blocks for many cellular lipids and membrane, while others are processed different metabolic degradation to release the energy for the synthesis of other intracellular energy carrier molecules such as ATP. As a main source with highly condensed energy, a sophistic metabolic network has been established to gradually degrade dietary fatty acids into smaller components and release the energy inside. For long chain fatty acids with even carbon number, they will be first converted into corresponding fatty acyl-CoA enzymatically to further undergo  $\beta$ -oxidation pathway repeatedly to produce acetyl-CoA, which will then enter the tricarboxylic acid (TCA) cycle as a key intermediate (Figure 55B). Also known as the Krebs cycle, TCA has been regarded as the hub of cellular metabolism. For fatty acids with odd carbon number, a final residual propionyl-CoA will generate, which is then converted to succinyl-CoA and enters the TCA cycle and be metabolically degraded. Specially, acetic acid and propanoic acid, as two small fatty acids, are directly converted to acetyl-CoA and propionyl-CoA by their corresponding acyl-CoA synthetases to enter TCA cycle, respectively.





**Figure 55 Protein fatty acylation and fatty acid metabolism (A) examples of protein lipidation; (B)  $\beta$ -oxidation pathway for fatty acid metabolism.**

During the  $\beta$ -oxidation pathway, many different acyl-CoA intermediates generate during the metabolic degradation of fatty acids. These intermediates with a large variety of structures are high-energy labile thioesters that can undergo nucleophilic acyl substitution reactions with nucleophiles on proteins such as amino group, hydroxyl group, and thiol group. These nucleophilic attacking processes can be either enzymatically or chemically to potentially introduce diverse protein fatty acylation modes. Although the  $\beta$ -oxidation pathway and the TCA cycle generally happen in mitochondria, these attacking reactions between fatty acyl-CoAs and protein nucleophiles are not just in mitochondria, because the fatty acyl-CoA synthetases have been found in different subcellular localizations.

Till now, protein fatty acylation has been discovered mainly on the N-terminal amino group and two amino acid residues as Lys and Cys<sup>[123]</sup>. Recently, some other nucleophilic residues are also reported with fatty acylation modifications. Many important

proteins have been found to undergo fatty acylation include histones, G proteins, transcription factors, and metabolic enzymes, which plays vital regulatory roles in their own structures and functions and in epigenetic control of gene expression.

We will still focus on Lys fatty acylation modifications. Current identified protein Lys fatty acylation types include short chain ones as acetylation, propionylation, butyrylation, crotonylation, long chain ones as myristoylation, and palmitoylation<sup>[124]</sup>. Many researches have been done for short chain modifications. However, little has been known about long chain modifications, and thus many fundamental questions are remaining unclear<sup>[39]</sup>. For instance, since the modification enzymes for acetylation, myristoylation, and palmitoylation are using fatty acyl-CoAs, we wonder whether these dietary fatty acids and their metabolic intermediates can modify Lys under non-enzymatic manner. For fatty acids with different carbon numbers, will they lead to distinct modification pattern or different “fatty acylomes”, and how can these modifications be regulated and removed? Solving these questions, we can generate more understanding of the impact of dietary fatty acids to human life from an epigenetic point of view.

## **6.2 Previous research and disadvantages**

Although many important questions remaining unclear, not many progresses have been made due to the lack of biochemistry approaches to study long chain fatty acylations. Even primarily focusing on the saturated fatty acids with the simplest structures, only a few methods are available to profile proteins with fatty acylation so far. Small acylation modifications such as acetylated proteins can be enriched by pan-anti-Kac antibody for

further proteomic analysis<sup>[125]</sup>. However, concerns exist for the selectivity and specificity of antibody prepared with normal injection to animal due to the heterogeneous nature, especially for the long chain fatty acids.

### 6.2.1 Surrogates for acetylation

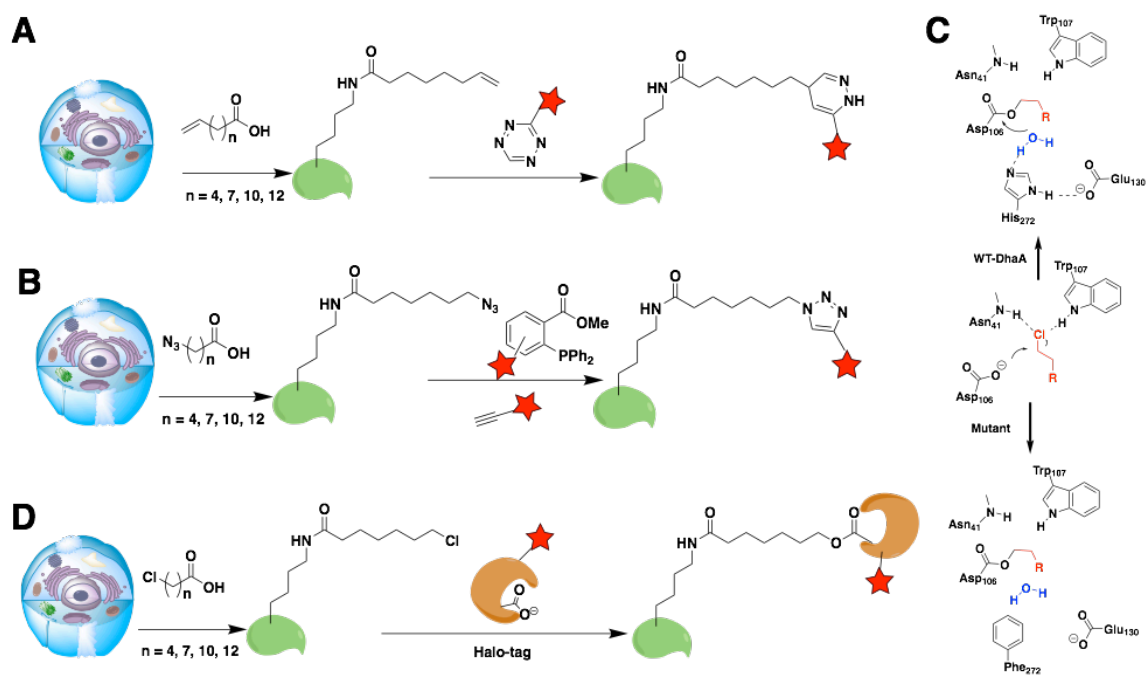
In the previous reports, different functionalized fatty acids have been developed as fatty acid surrogates, which are applied to profile fatty acylation proteome to improve the selectivity<sup>[126]</sup>. These surrogates generally have certain chemical functionalities for further modifications as chemical labeling or bioconjugation formation to add purification tags such as biotin for the purpose of acylated protein enrichment. Both azido group and alkyne group are most widely used warheads to arm the carbohydrate chains, due to the easy handling Staudinger reaction or “Click reaction”<sup>[127]</sup>.

For small acylation modification type as acetylation, two clickable small molecules have been designed and synthesized as 4-pentynoic acid and 3-azidopropionic acid for the substrate identification of an important histone acetyltransferase (HAT) family enzyme p300. After being added to certain proteins, the terminal-alkyne group on 4-pentynoic acid, for instance, can conjugate with Biotin for further pull-down and proteomic analysis<sup>[128]</sup>. Although these two surrogates can be alternative cofactors for p300, their structures are quite different from the acetic acid or its CoA conjugate. Another acetyllysine surrogate design is the chloroacetyl-CoA as a cofactor for HAT family enzyme GCN5 to be added to protein substrates<sup>[129]</sup>. Although chloroacetic acid can be a

better mimic of acetic acid to profile acetylome, its toxicity and high reactivity over thiol group raise concern when applying directly to cellular study.

### 6.2.2 Surrogates for medium and long chain fatty acylation

With a similar design, terminal alkyne or azide-conjugated long-chain fatty acid surrogates have also been developed to study long-chain fatty acylation modifications and the corresponding acylomes. The azide-conjugated myristic acid and palmitic acid are used as chemical probes for myristoylation and palmitoylation, respectively. After the Staudinger ligation reaction (Figure 56A), the proteins with azidofatty acylations are conjugated with biotin and visualized with Streptavidin<sup>[126]</sup>. Although robust and effective, these surrogates are still structurally different from native fatty acids, especially at the terminal region. Additionally, long chain fatty acids with terminal olefin have also been developed to undergo tetrazine labeling for proteomic study (Figure 56B).



**Figure 56 Surrogates of fatty acids and their corresponding detection strategies (A)** terminal olefin group labeled by tetrazine; **(B)** terminal azido group labeled by Click reaction or Staudinger ligation; **(C)** the basic mechanism of dehalogenase enzyme; **(D)** design of terminal chloride labeled by HaloTag protein.

Neither the alkyne nor the azido group can be a good mimic for a methyl group, and therefore they are not ideal to mimic native fatty acids for further protein acylome analysis about fatty acylation distribution and function. Therefore, we are looking forward to developing novel and robust metabolic labeling strategies for the profiling of protein fatty acylation. This study can facilitate our understanding of dietary fatty acids and their high-energy intermediates that are potentially able to acylate cell proteome in an enzymatic or non-enzymatic manner.

## 6.3 Our design and potential advantages

### 6.3.1 Project initiation

As mentioned in the previous part, fatty acids with a terminal chloride atom are good surrogates for their native forms. However, the chloroacetic acid together with its corresponding chloroacetyl-CoA are not applicable due to the high electrophilicity of its  $\alpha$ -carbon atom, which will undergo nucleophilic attack in cell. We envisioned that medium and long chain fatty acids with terminal chloride atom would not have such a problem with the electron withdrawing effect from the adjacent carbonyl group. We realized a popular chemical biology tool named as “Halo-tag” technique. This is an enzyme mutated from dehalogenase enzyme by changing a key His in the acid-base catalyzing to Phe (Figure 56C). Labeling as  $\omega$ -chlorofatty acids, we can further apply the “Halo-tag” technique for the further labeling or bioconjugation formation (Figure 56D).

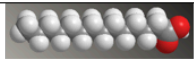

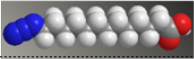

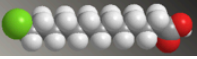
### 6.3.2 Advantage analysis

To start with, we want to compare the  $\omega$ -chlorofatty acids with other existing fatty acids with distinct terminal functional groups. The  $\omega$ -chlorofatty acids own many advantages comparing with their corresponding  $\omega$ -alkyl fatty acids,  $\omega$ -azido fatty acids, and  $\omega$ -olefin fatty acids. We can barely divide them into three aspects. As an enzymatic procedure, Halo-tag recognize and react with terminal chloride atom with high selectivity, high specificity, and fast kinetics with a secondary rate constant of  $10^6 \text{ M}^{-1}\text{s}^{-1}$ , comparing with other typical bioorthogonal reaction such as Click reaction or tetrazine-ene reaction. As for selectivity, no side reaction has been reported to for  $\omega$ -chloro fatty acids in cell,

while other alternatives are not as stable. For instance, the  $\omega$ -azidofatty acids may undergo reduction reaction under the reductive condition inside of cell.

Next, the terminal chloride atom is a good mimic for the methyl group as regard to both size and hydrophobicity. The  $\omega$ -chlorofatty acid with one less carbon number comparing with the corresponding natural fatty acid is much similar in topology comparing with a terminal alkyne or azide or olephin. We are using myristic acid and its four surrogates with different terminal functional groups as example. From the 3D structure models in Table 6.1, the ones with either azido group or alkyne group are much longer than the native one, while the one with olefin group has a plain topology at its terminal region. In addition, surrogates with the terminal alkyne and azide are also different in hydrophobicity comparing with the ethyl group they are trying to mimic. Therefore, only the  $\omega$ -chlorofatty acid can be the best analogy to substitute the native fatty acid for our further metabolic study.

**Table 2 The comparison of fatty acids with different terminal functional groups**  
**Our Design: a better mimic of CH<sub>3</sub>- group**

X-(CH <sub>2</sub> ) <sub>11</sub> COOH	Detection	Rate	Stability	Structures
CH <sub>3</sub> CH <sub>2</sub> -	NA	NA	Very stable	
HCC-	Click reaction	10-100 M <sup>-1</sup> s <sup>-1</sup>	Potential thiol addition	
N <sub>3</sub> CH <sub>2</sub> -	Click/Staudinger reaction	10-100 M <sup>-1</sup> s <sup>-1</sup>	Thiol reduction	
CH <sub>2</sub> =CH <sub>2</sub> -	Tetrazine-ene reaction	1-100 M <sup>-1</sup> s <sup>-1</sup>	Background labeling	
ClCH <sub>2</sub> -	Halo-tag	10 <sup>6</sup> M <sup>-1</sup> s <sup>-1</sup>	Very stable	

### 6.3.3 Introduction to Halo-tag and our strategy

Halo-tag is one of the most widely used protein tag techniques, in which an engineered enzyme was used to react with  $\omega$ -chlorofatty acids<sup>[61]</sup>. Starting from a haloalkane dehalogenase, the His272 necessary for the acid-base catalyzing mechanism was engineered to Phe (Figure 56C), which makes the enzyme covalently bind to the chloride containing substrate without falling-off<sup>[130]</sup>. The modified enzyme was then applied for the binding to synthetic ligands with long chlorohydrocarbon chains. The follow-up research has extended the Halo-tag ligands into any molecules with a long chain functionalized with a terminal chloride atom, which makes it better applicable to a wider chemical biology fields. By applying the Halo-tag technique, we can use Halo protein with fluorescence labeling or biotin labeling to react with proteins with Lys long chain acylation and further visualized them (Figure 56D).

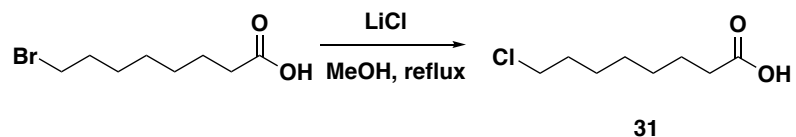
Since we are trying to study dietary fatty acids and their metabolic intermediates with long, medium, and short chain length, we want to confirm whether Halo protein can recognize all of them and with what kinds of kinetics. Therefore, we selected four different  $\omega$ -chlorofatty acids with various carbon number as 5-chlorovaleric acid representing short chain, 8-chlorooctanoic acid representing medium chain, 11-chloroundecanoic acid and 13-chlorotridecanoic acid representing long chains. A gel-based kinetics assay will be carried out by using fluorescence labeled Lys derivatives as small molecular probes to react with HaloTag protein.



## 6.4 The synthesis of chlorofatty<sub>n</sub>-acids (n = 4, 7, 10, 12) and FITC-Lys(chloroacyl)-OMe

To start with, we need to synthesize the ω-chlorofatty<sub>n</sub>-acids (n = 4, 7, 10, 12) with different carbon numbers. More specifically, we made three different compounds as 8-chlorooctanoic acid, 11-chloroundecanoic acid and 13-chlorotridecanoic acid, while the 5-chloropentanoic acid is commercially available. To further study the binding and reaction kinetics, we want to synthesize the corresponding fluorescein labeled Lys derivatives.

### 6.4.1 The synthesis of chlorofatty acids



**Figure 57 Scheme for the synthesis of 8-bromooctanoic acid**

**The synthesis of 8-chlorooctanoic acid (31):** 8-bromooctanoic acid (1 eq, 22.4 mmol, 5 g) was dissolved in methanol (2 ml/0.3 mmol), with the addition of LiCl powder (20 eq, 450 mmol, g). The reaction mixture was allowed to reflux overnight under stirring condition, followed by the evaporation of methanol over vacuum. The remaining product was extracted by EtOAc/H<sub>2</sub>O system to the EtOAc layer. The organic phase was then collected, washed with Brine, and dried over sodium sulfate anhydrous. After the evaporation of EtOAc, an oil product 8-chlorooctanoic acid (31) was generated with 97 % yield, which was further verified by NMR. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ 3.55 (t, 2H,

$J=1.2$  Hz), 2.36 (t, 2H,  $J=1.2$  Hz), 1.79-1.34 (m, 10H),  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ )  $\delta$  180.4, 45.0, 34.0, 32.5, 28.8, 28.5, 26.6, 24.5

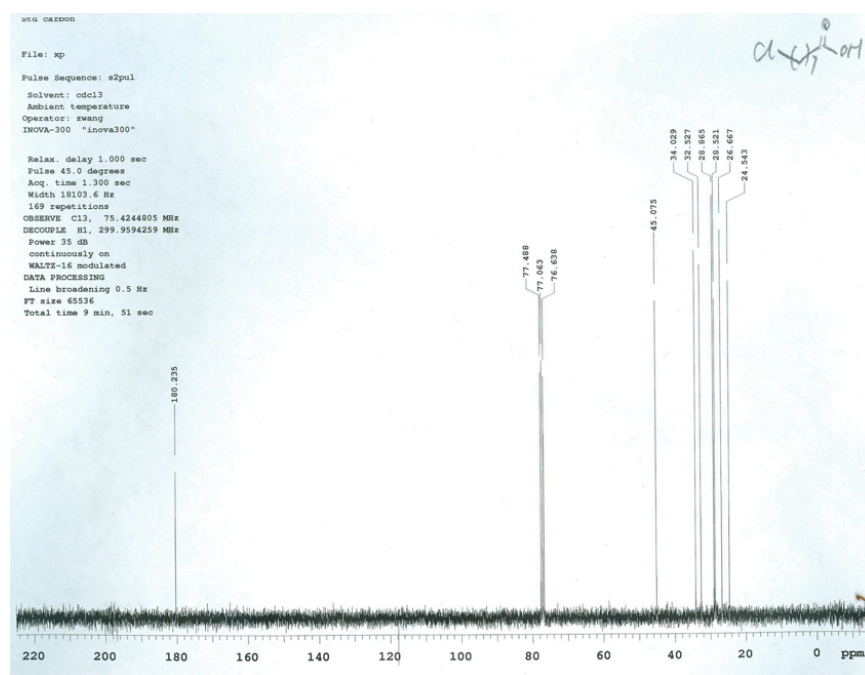
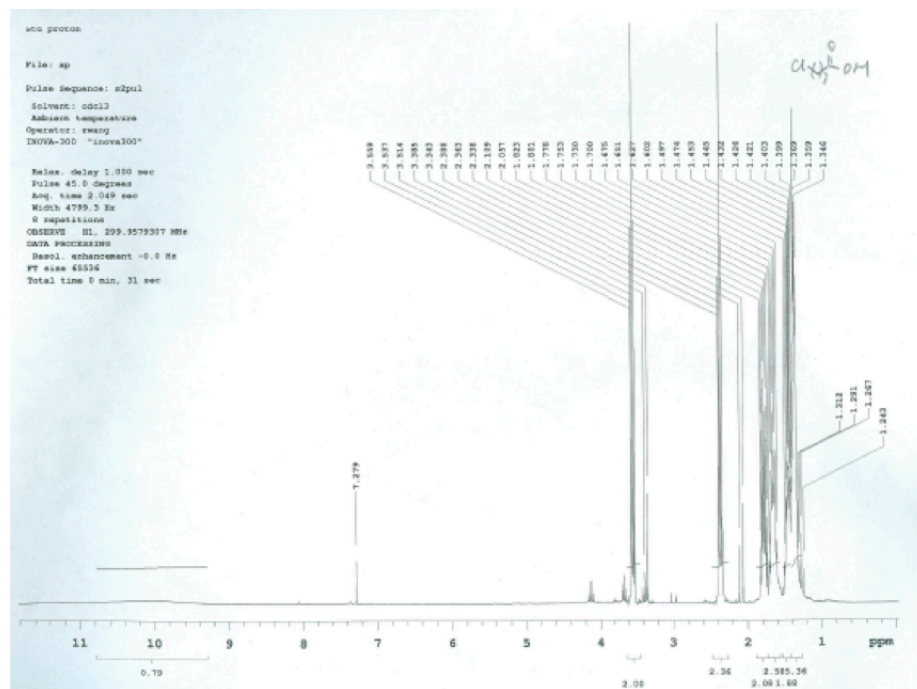
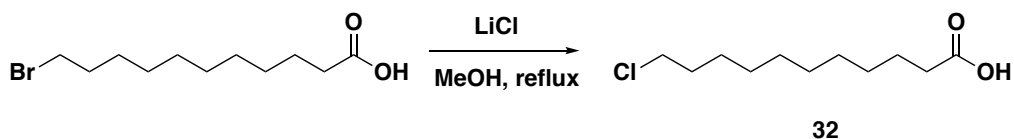


Figure 58  $^1\text{H}$ NMR and  $^{13}\text{C}$ NMR of 8-bromooctanoic acid



**Figure 59 Scheme for the synthesis of 11-chloroundecanoic acid**

**The synthesis of 11-chloroundecanoic acid (32):** 11-bromoundecanoic acid (1 eq, 18.9 mmol, 5 g) was dissolved in methanol (2 ml/0.3 mmol), with the addition of LiCl powder (20 eq, 380 mmol, 63 g). The reaction mixture was allowed to reflux overnight under stirring condition, followed by the same procedure of extraction by EtOAc/H<sub>2</sub>O for 2 times. The EtOAc phase was collected, washed with Brine and evaporated to give out yellow solid as 11-chloroundecanoic acid (2) with 95 % yield, which was further verified by NMR. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ 3.50 (t, 2H, J=1.2 Hz), 2.28 (t, 2H, J=1.2 Hz), 1.76-1.02 (m, 16H), <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ 180.8, 45.4, 34.3, 32.8, 29.6, 29.5, 29.4, 29.2, 29.0, 27.0, 24.8.

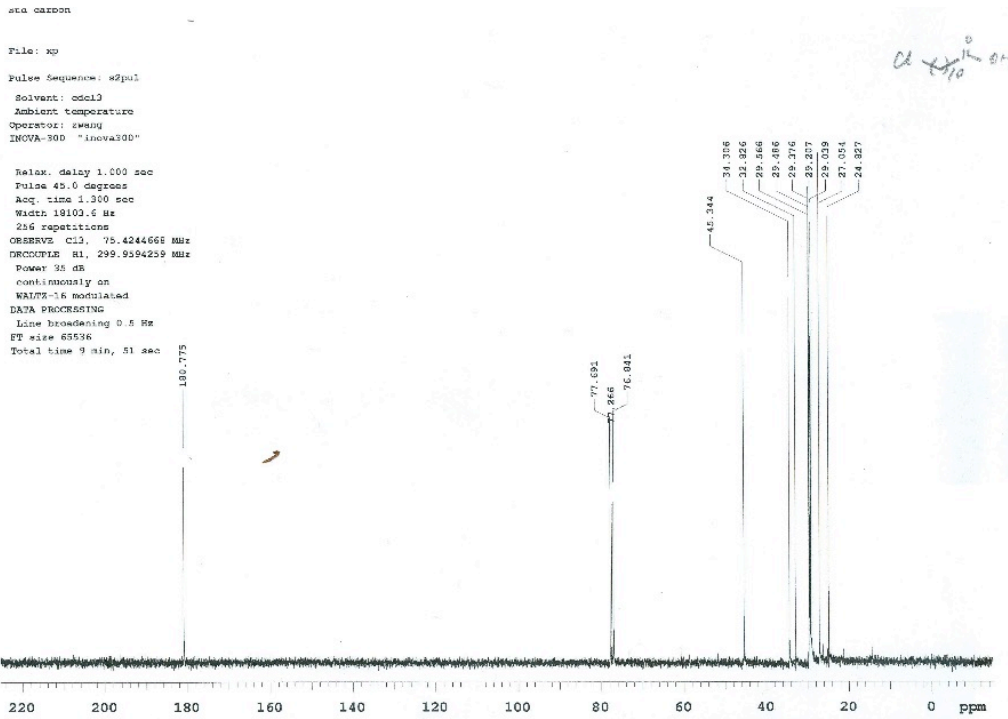
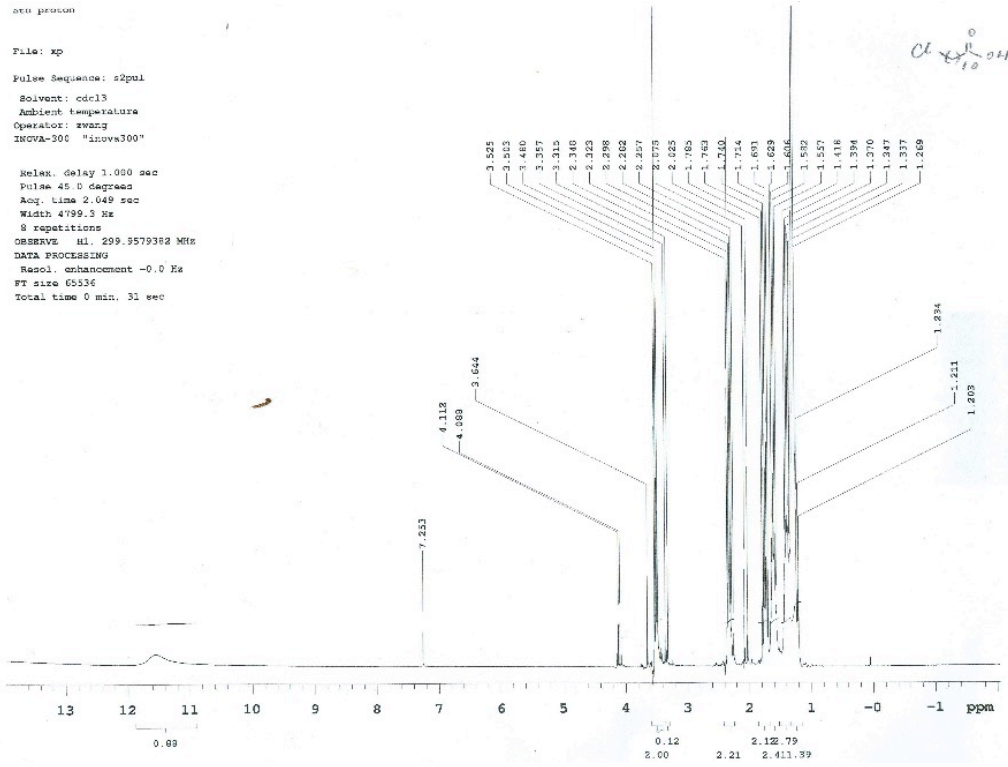
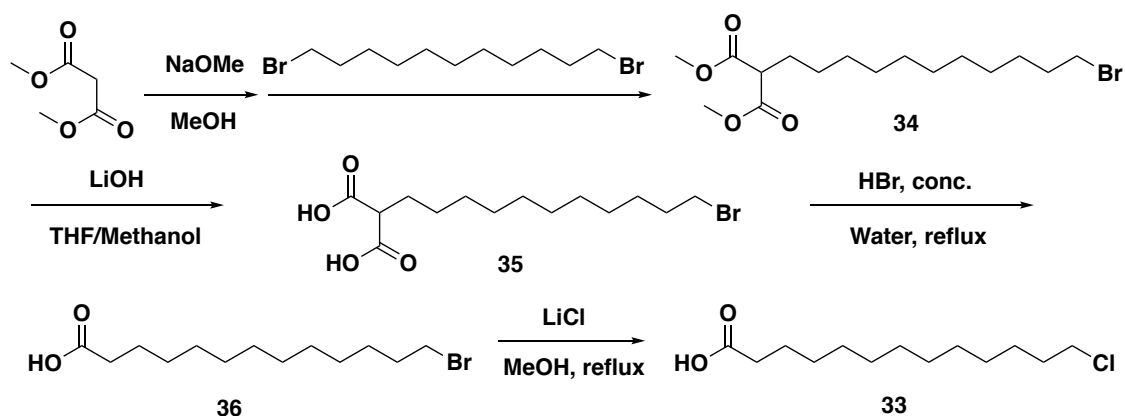


Figure 60  $^1\text{H}$ NMR and  $^{13}\text{C}$ NMR of 11-chloroundecanoic acid



**Figure 61 Scheme for the synthesis of 13-chlorotridecanoic acid**

**Dimethyl 2-(11-bromoundecyl)malonate (34):** Sodium methoxide was freshly made by the slow addition of 23 ml of methanol into sodium metal (2.3 g, 100 mmol) over ice bath and refluxing the solution for 2 h until all the sodium metal dissolved. The sodium methoxide generated was then added slowly to a methanol solution of dimethylmalonate (1.38 g, 1.2 ml, 1 eq, 10 mmol), followed by the precipitation of a white solid. After more methanol was added, the white solid could redissolve to give out a clear solution. Then, add this clear solution dropwisely to a methanol/THF solution of 1,11-dibromoundecane (3.07 g, 2.3 ml, 10 mmol) and let it to reflux while stirring for o/n. After the evaporation of solvent, the reaction was directly load to silica column chromatography and dimethyl 2-(11-bromoundecyl)malonate (3) was purified as a 1.8 g colorless oil with a 45 % yield.  $^1\text{H NMR}$  (300 MHz,  $\text{CDCl}_3$ )  $\delta$  3.71 (s, 6H), 3.38 (m, 4H), 1.83 (m, 4H), 1.40-1.29 (m, 20H),  $^{13}\text{C NMR}$  (75 MHz,  $\text{CDCl}_3$ )  $\delta$  170.1, 52.5, 51.8, 34.1, 32.9, 29.8, 29.7, 29.6, 29.56, 29.54, 29.51, 29.4, 29.3, 28.9, 28.8, 28.3.

**2-(11-bromoundecyl)malonic acid (35):** To a THF/ $\text{H}_2\text{O}$  1:1 mixt solution of dimethyl 2-(11-bromoundecyl)malonate (100 mg, 0.27 mmol), lithium hydroxide (398.6 mg, 19 ml,

0.5 M) was added. The reaction mixture was let to stir at r.t. for o/n. After the evaporation of THF, direct EtOAc/H<sub>2</sub>O extraction was carried out for 2 times. The organic layer was collected, washed with Brine, dried, and evaporated via vacuum. No further purification was made, and the product was directly used for the next step.

**13-bromotridecanoic acid (36):** 2-(11-bromoundecyl)malonic acid (0.27 mmol) was dissolved in 2 ml of concentrated HBr aqueous solution (~8.9 M). The reaction solution was let to reflux under stirring condition for o/n and to release carbon dioxide as by-product. EtOAc was used to extract the water phase for 3 times. The organic layer was collected, washed with Brine, dried, and evaporated via vacuum to finally generate a brown colored oil. The crude product was purified by silica chromatography first with pure hexane then with 9:1 hexanes:ethyl acetate to generate 13-bromotridecanoic acid (6) as a 13 mg white solid with an overall yield of 14 %. A larger scaled reaction was then set up to prepare enough amounts for the next step. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ 3.40 (t, 2H, J=1.2 Hz), 2.34 (t, 2H, J=1.2 Hz), 1.90-1.26 (m, 20H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>) δ: 179.7, 33.86, 33.81, 32.62, 29.30, 29.28, 29.21, 29.18, 29.01, 28.83, 28.54, 27.97, 24.45.

**13-chlorotridecanoic acid (33):** 13-bromoundecanoic acid (560 mg, 1 eq, 1.9 mmol) was dissolved in methanol/THF 1:1 (5 ml), with the addition of LiCl powder (15 eq, 28.5 mmol, 1.2 g). The reaction mixture was let to reflux under stirring condition for o/n. After the evaporation of methanol followed by the extraction of EtOAc/H<sub>2</sub>O system, the organic layer was collected, washed with Brine, and dried. The evaporation of EtOAc via vacuum could finally give out a 299.6 mg white solid with 63 % yield, which was further verified by NMR. <sup>1</sup>H NMR (300 MHz, CDCl<sub>3</sub>) δ 3.51 (t, 2H, J=1.2 Hz), 2.29 (t, 2H, J=1.2 Hz),

1.79-1.25 (m, 20 H);  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ )  $\delta$  180.4, 45.1, 34.1, 32.6, 29.50, 29.48, 29.43, 29.21, 29.12, 29.02, 28.9, 26.9, 24.6.

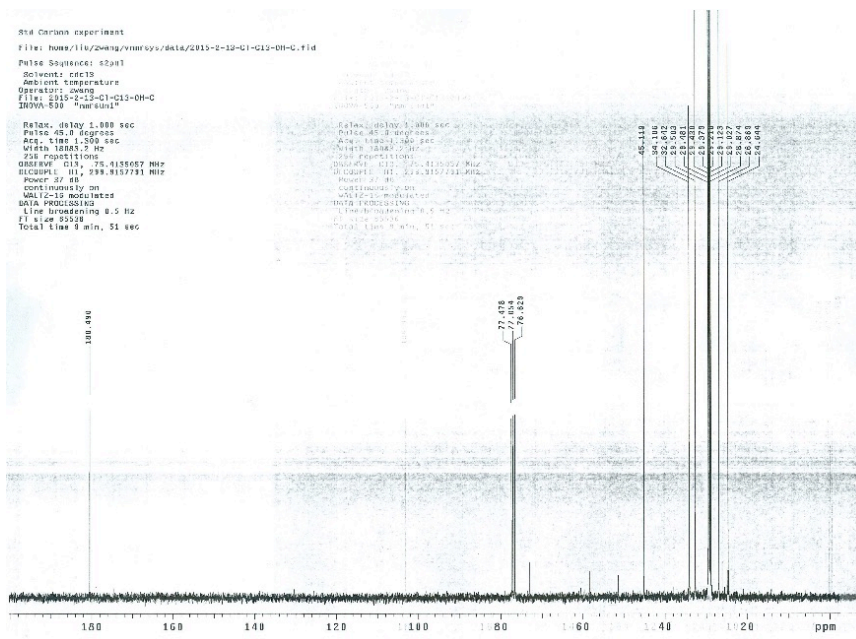
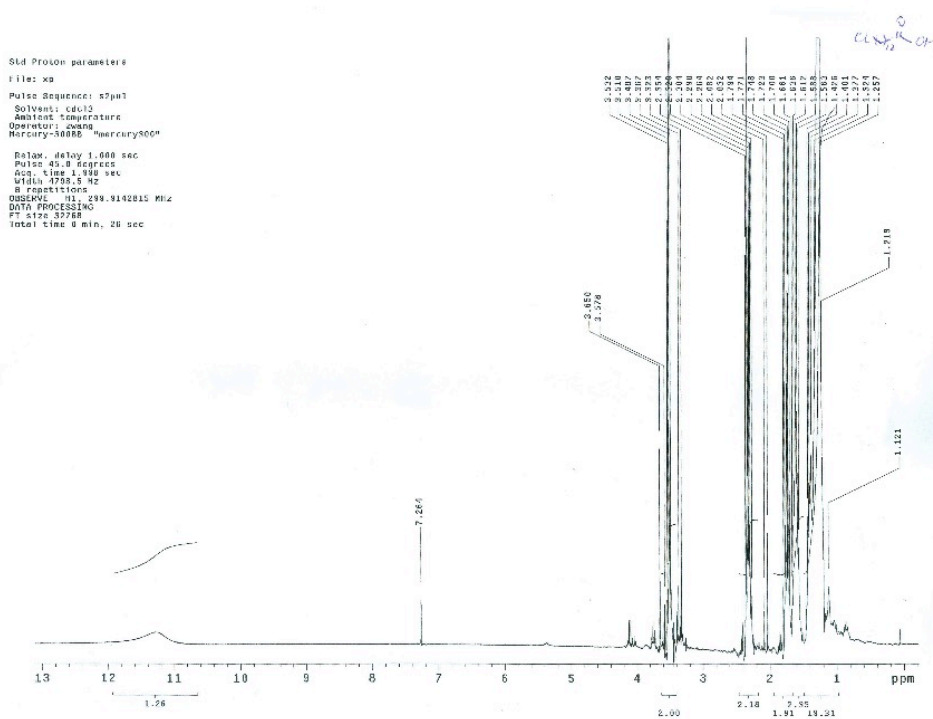
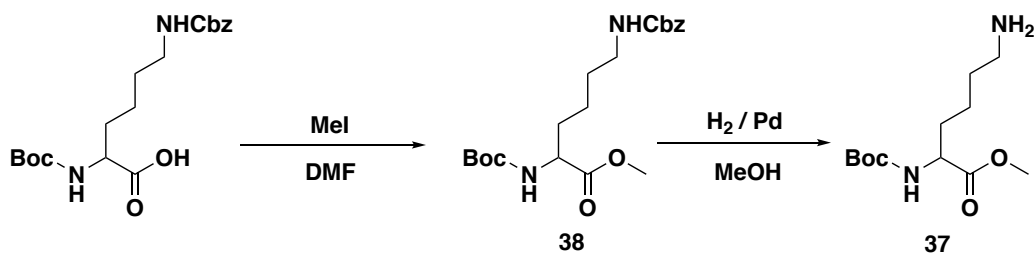


Figure 62  $^1\text{H}$ NMR and  $^{13}\text{C}$ NMR of 13-chlorotridecanoic acid

#### 6.4.2 The synthesis of Boc-Lys-OMe intermediate



**Figure 63 Scheme for the synthesis of Boc-Lys-OMe intermediate**

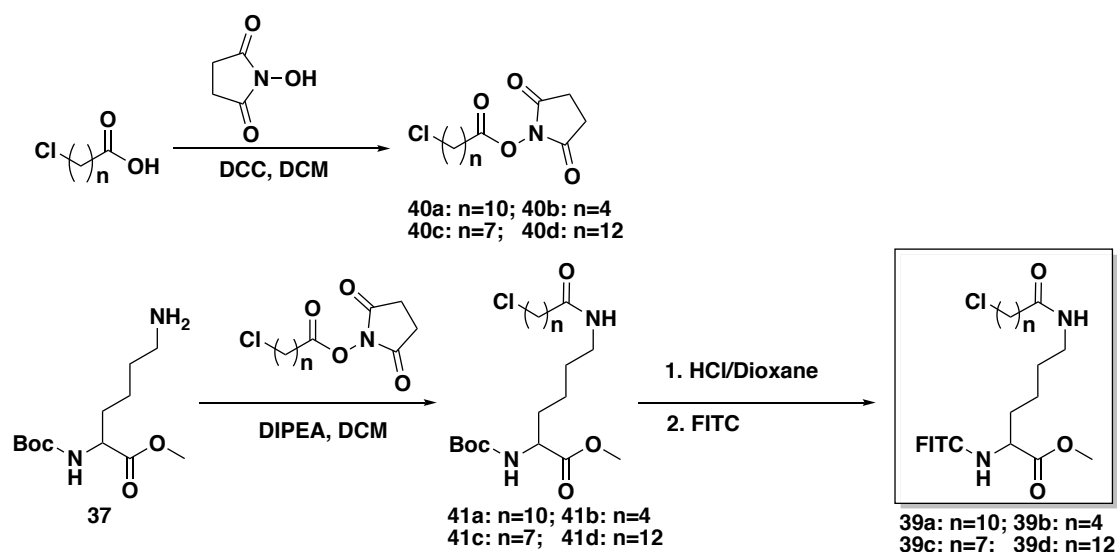
**Boc-Lys(Cbz)-OMe (38):** Boc-Lys(Cbz)-OH (2 g, 1 eq, 5.3 mmol) and K<sub>2</sub>CO<sub>3</sub> (1.38 g, 2 eq, 10.6 mmol) were both dissolved in DMF 15 ml, followed by the addition of MeI (495  $\mu$ L, 1.5 eq, 7.9 mmol). The reaction was let to stir at r.t. for 30 min. Direct extraction of EtOAc/H<sub>2</sub>O was carried out by the addition of 100 mL H<sub>2</sub>O and 100 mL EtOAc for three times. The organic layer was collected, washed with brine, and evaporated via vacuum. The oil product was directly used for the next step without further purification.

**Boc-Lys-OMe (37):** 10 mol% Pd/C was put in a round bottle and purged by nitrogen gas for 30 min. Boc-Lys(Cbz)-OMe (1.5 g, 1 eq, 5.3 mmol) was dissolved in MeOH (30 mL), and added to the round bottle. A hydrogen gass balloon was then set up, and the reaction mixture was let to stir for 4 h monitored by TLC. After the starting material fully disappeared, filtration was then carried out for the removal of Pd. Then, MeOH was evaporated via vacuum. The oil product was directly used for the next steps without further purification.



### 6.4.3 The synthesis of FITC-Lys(chloroacyl)-OMe

In this part, we will use the Boc-Lys-OMe as starting material to make all four kinds of dyes. 11-chloroundecanoic acid (32) will be used as example for the synthesis of FITC-Lys(11-chloroundecanyl)-OMe (39a). For the other three FITC-Lys( $\omega$ -chloroacyl)-OMe derivatives, a corresponding  $\omega$ -chloroaliphatic acid is used with the similar synthesis procedure, respectively. The NMR for Boc- Lys( $\omega$ -chloroacyl)-OMe (41a,b,c,d) and FITC-Lys( $\omega$ -chloroacyl)-OMe (39a,b,c,d) are provided.



**Figure 64** Scheme for the synthesis of Boc-Lys-OMe intermediate

**11-chloroundecanoic NHS ester (40a):** 11-chloroundecanoic acid (676.7 mg, 1 eq, 3.8 mmol) and *N*-hydroxysuccinimide (NHS) (524 mg, 1.2 eq, 4.6 mmol) were both dissolved in DCM. A solution of *N,N'*-Dicyclohexylcarbodiimide (DCC) (939 mg, 1.2 eq, 4.6 mmol) in DCM was added dropwisely with ice bath. White precipitation was observed in several minutes. Then, ice bath was removed, and the reaction was let to stir at r.t. for o/n. Filtration was carried out, followed by the solvent evaporation via vacuum. The product

in oil form was directly used for the next step without further purification.  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ )  $\delta$  3.56 (t, 2H,  $J=1.2$  Hz), 2.86 (s, 4H), 2.62 (t, 2H,  $J=1.2$  Hz), 1.98-1.24 (m, 16H)

**Boc-Lys(11-chloroundecanyl)-OMe (41a):** Boc-Lys-OMe (1.4 g, 1 eq, 5.3 mmol) was dissolved in DCM, followed by the addition of DIPEA (1.0 mL, 1.2 eq, 6.3 mmol) and incubation on ice bath. 11-chloroundecanoic NHS ester (1.7 g 1 eq, 5.3 mmol) dissolved in DCM was added slowly, and the mixture was let to stir under r.t. for o/n. The reaction mixture was then extracted with 1 M Citric acid pH=2 buffer/EtOAc twice. The organic layer was collected, washed with Brine, and evaporated. The crude product was purified by flash chromatography (2:1 hexanes:ethyl acetate) to yield Boc-Lys(8-chlorooctanyl)-OMe as white powders (1.2 g, 66 %).  $^1\text{H}$  NMR (300 MHz,  $\text{CDCl}_3$ )  $\delta$  4.31 (m, 1H), 3.74 (s, 3H), 3.56 (t, 2H,  $J=1.2$  Hz), 3.25 (q, 2H,  $J=0.8$  Hz), 2.16 (t, 2H,  $J=1.2$  Hz), 2.05-1.24 (m, 22H).  $^{13}\text{C}$  NMR (75 MHz,  $\text{CDCl}_3$ )  $\delta$ : 172.9, 164.6, 155.1, 79.4, 52.7, 51.8, 44.7, 38.5, 36.3, 32.2, 31.9, 28.9, 28.9, 28.8, 28.5, 28.3, 27.8, 26.4, 25.3, 22.2.

**FITC-Lys(11-chloroundecanyl)-OMe (39a):** Boc-Lys(11-chloroundecanyl)-OMe (250 mg, 1 eq) was dissolved in 20 mL HCl (1 M in dioxane) and let to stir for 8 h, monitored by TLC. After the completely removal of solvent, 160 mg white powder of H-Lys(11-chloroundecanyl)-OMe in HCl salt form was prepared without further purification. H-Lys(11-chloroundecanyl)-OMe (118 mg, 0.22 mmol 1.2 eq) was dissolved in a solution of DIPEA (41  $\mu\text{L}$ , 0.22 mmol, 1.2 eq) in DMF (2 mL). The reaction mixture was cooled to  $0^\circ\text{C}$  before the addition of fluorescein 5-isothiocyanate (FITC) (70 mg, 0.18 mmol, 1 eq). The reaction mixture was allowed to warm to r.t. and let to stir for o/n. The solvent

was evaporated and the crude oil was directly applied to silica column chromatography using a gradient of DCM:acetone:MeOH (80:10:10 to 0:90:10) to afford compound FITC-Lys(11-chloroundecanyl)-OMe (12a) (82 mg, 62%) as an orange solid. <sup>1</sup>H NMR (MeOH-*d*<sub>4</sub>) δ: 8.30 (s, 1H), 7.36 (d, 1H, J=1.2 Hz), 7.32(d, 1H, J=4.8 Hz), 6.71-6.67 (m, 4H), 6.59-6.54 (m, 2H), 5.11 (t, 1H, J=1.2Hz), 3.75 (s, 3H), 3.54 (t, 2H, J= 3.3Hz), 3.20 (m, 2H), 2.16 (m, 2H), 2.02-1.98(m, 2H), 1.74-1.71(m, 2H), 1.60-1.41(m, 24H). <sup>1</sup>HNMR and <sup>13</sup>CNMR for all the four compounds are listed at the end of the chapter (Figure 76-79).

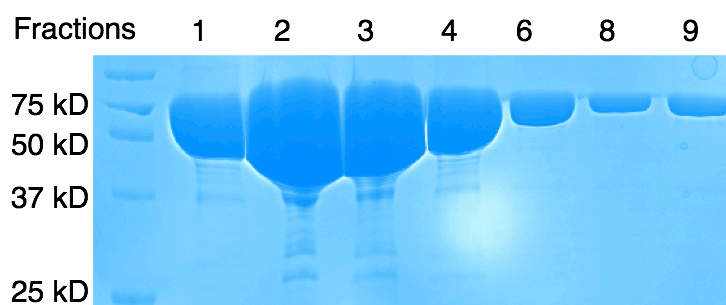
## 6.5 Kinetics study of Halo-tag reaction

With all the four FITC-Lys( $\omega$ -chloroacyl<sub>n</sub>)-OMe (n = 4, 7, 10, 12) in hand, we carried out gel-based kinetics study to test whether Halo-tag protein can react with all the long medium and short chained chlorofatty acids we chose.

### 6.5.1 The expression of Halo protein

We first started with the expression of Halo protein (MW: 61940 Da). We got the plasmid pGEX-GST-Halo containing the gene of Halo-tag protein fused with GST tag at its N-terminus from Dr. Eric Kool's group at Stanford University. The plasmid was used to transform *E. coli* BL21(DE3) cells, and a single colony was picked for the setting up of 10 mL seed culture. The seed culture was then used to inoculate 1 L LB medium with 100  $\mu$ g/mL Amp for the expression of Halo-tag protein. The cells were let to grow at 37 °C in the 2YT medium to OD<sub>600</sub> reaching 0.4, and the expression of GST-HaloTag protein was induced with the addition of 0.5 mM IPTG. Next, cells were let to grow continuously at

16 °C for 12 h, collected and washed with 100 mM Tris lysis buffer pH 7.5 containing 1 mM PMSF. The cell pallet was stored at -80 °C for purification. 30 mL 100 mM Tris lysis buffer pH 7.5 with 1 mM PMSF was used to resuspend the cell pallet. The expressed GST-HaloTag was loaded to a column with 1 mL of GST resin, washed by 50 mM Tris buffer pH 8 and finally eluted by 10 mM glutathione in 50 mM Tris buffer pH 8 in different fractions, which are confirmed by SDS-PAGE (Figure 65). The protein was dialyzed against a PBS pH 7.4, and aliquoted into different tubes. The concentration of Halo-tag protein was measured as 3.12 mg/mL, which can be calculated as 50.3  $\mu$ M.



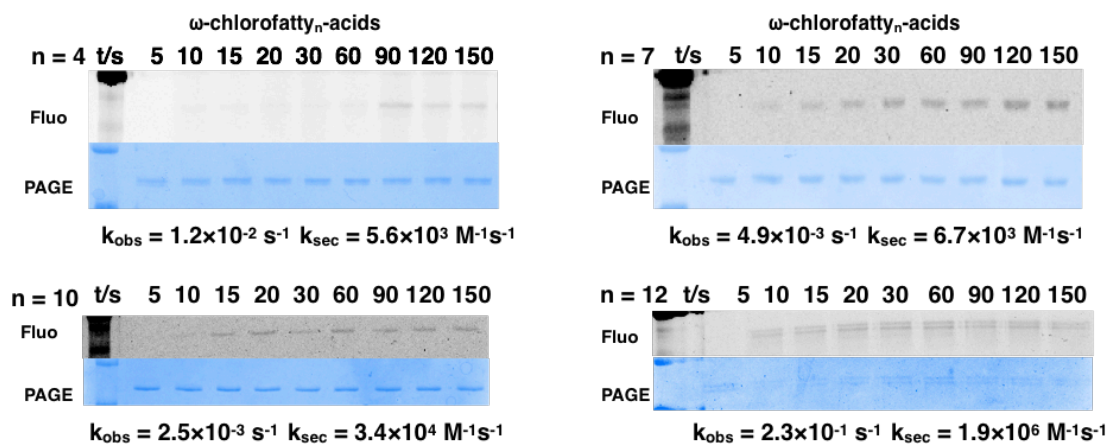
**Figure 65 The purification of GST-HaloTag protein**

#### 6.5.2 Gel-based kinetics study

All the four FITC-Lys( $\omega$ -chloroacyl<sub>n</sub>)-OMe (n = 4, 7, 10, 12) dyes were dissolved in DMSO to a concentration between 0.1-1  $\mu$ M range as original solutions. UV absorbance was used to precisely quantify the concentration, followed by the dilution to DMSO to make stock solutions. The concentration of each of them are 900 nM for n=4, 710 nM for n=7, 900 nM for n=10, and 890 nM for n=12.

Each of the dyes was added to a reaction mixture containing 10 nM Halo-tag protein to a final concentration of 1 nM. The final volume of the reaction was adjusted to 20  $\mu$ L with PBS, and the solution was pipetted to mix well. After reacting for a certain period time, 4  $\mu$ L of 6 $\times$ loading buffer was added followed by heating up the reaction solution to boil. After several trials, we finally select the time interval as 5 s, 10 s, 15 s, 20 s, 30 s, 60 s, 90 s, 120 s, and 150 s. The boiled samples were loaded to 15 % SDS-PAGE. The gel was visualized under UV irrigation before staining with Coommassie blue (Figure 66).

The reaction is following pseudo-first order kinetics while the enzyme is about 10 folds excessive. We use the gel quantification software from Bio-Rad company, and the fluorescence intensity after normalization was used to plot and calculate the observed rate constant and the second order rate constant. Interestingly, we found Halo-tag could recognize short chained fatty acylation even with only 4 carbons in between the carbonyl group and the chloride atom.



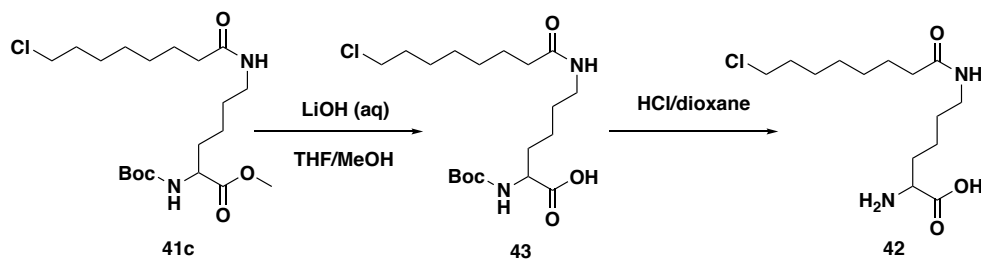
**Figure 66 Gel-based kinetics study of Lys conjugated dye reacting with HaloTag protein**

From the gel result, we can clearly see that all the four dyes with various carbon chain number can all react with Halo-tag protein. This result is surprising to us since no one has shown that Halo-tag protein can recognize short chain with only 5 carbon atom length. We also observed that with the increase of the chain length the Halo-tag recognition speeds up. Especially, the reaction between FITC-Lys( $\omega$ -chloroacyl<sub>12</sub>)-OMe and Halo-tag protein showed a second order rate constant  $10^6 \text{ M}^{-1}\text{s}^{-1}$  agreeing with the original result<sup>[130]</sup>.

## 6.6 Test on proteins

With the kinetics data at hand, we then carried out the test of HaloTag recognition on protein levels to get an optimized condition for further membrane blotting. We decided to use the fatty acids with medium chain length as a model to test, because with our ncAA technique, we can easily install a non-natural structure *N*<sup>ε</sup>-(8-chlorooctanyl)lysine into desired site. Also noted as H-Lys(8-chlorooctanyl)-OH, We first carried out the synthesis of the *N*<sup>ε</sup>-(8-chlorooctanyl)lysine (CloK).

### 6.6.1 The synthesis of H-Lys(8-chlorooctanyl)-OH (42)



**Figure 67** Scheme for the synthesis of H-Lys(8-chlorooctanyl)-OH

**Boc-Lys(8-chlorooctanyl)-OH (43):** Boc-Lys(8-chlorooctanyl)-OMe (41c from Part 6.4.3) (1 eq, 2.4 mmol, 1 g) was dissolved in methanol/THF 1:1 (15 ml), with the addition of LiCl powder (15 eq, 36 mmol, 1.5 g). The reaction mixture was let to reflux under stirring condition for o/n. After the evaporation of solvent followed by the extraction of EtOAc/H<sub>2</sub>O system, the organic layer was collected, washed with Brine, and dried. The evaporation of EtOAc via vacuum could finally give out 0.9 g of colorless oil with 95 % yield.

**H-Lys(8-chlorooctanyl)-OH (13):** Boc-Lys(8-chlorooctanyl)-OH (1 eq, 2.3 mmol, 910 mg) was dissolved in 10 mL HCl (1 M in dioxane) and let to stir for 8 h. The reaction was monitored by TLC. After the completely removal of solvent, 750 mg white powder of H-Lys(8-chlorooctanyl)-OH in HCl salt form was prepared without further purification. <sup>1</sup>H NMR (300 MHz, D<sub>2</sub>O)  $\delta$  3.94 (t, 1H, J=10 Hz), 3.38 (t, 2H, J=12 Hz), 3.01 (m, 2H), 2.06 (t, 2H, J=1.2 Hz), 1.79 (t, 2H, J=1.2 Hz), 1.58-1.13 (m, 14H). <sup>13</sup>C NMR (75 MHz, CDCl<sub>3</sub>)  $\delta$ : 175.9, 171.6, 66.4, 52.5, 45.3, 38.8, 35.6, 32.1, 29.3, 28.4, 28.0, 26.2, 25.4, 21.7.

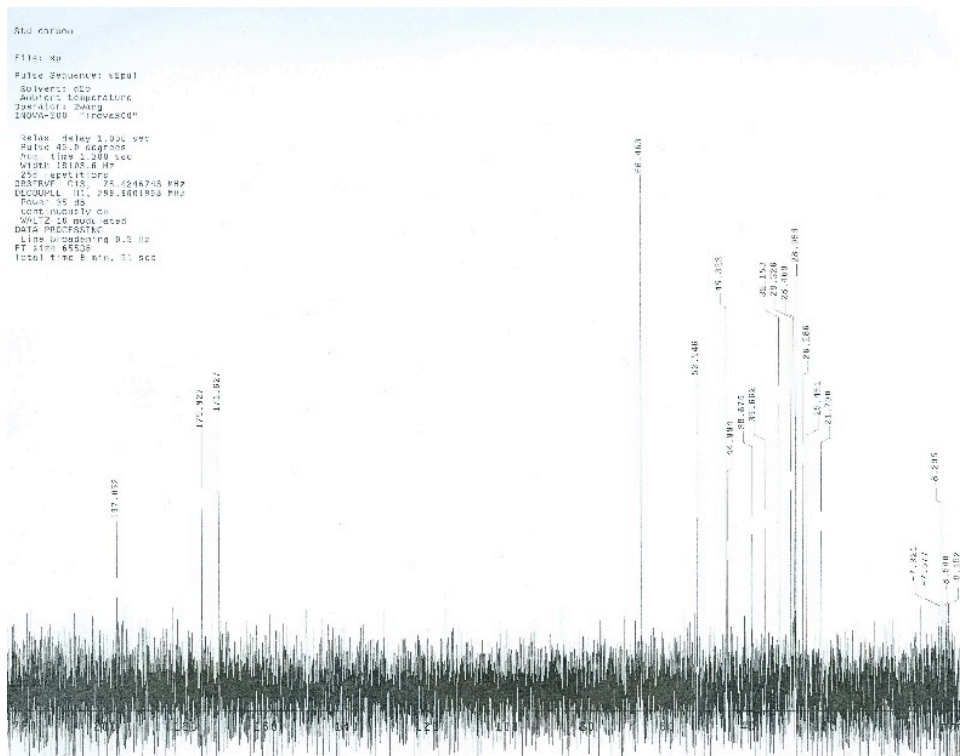
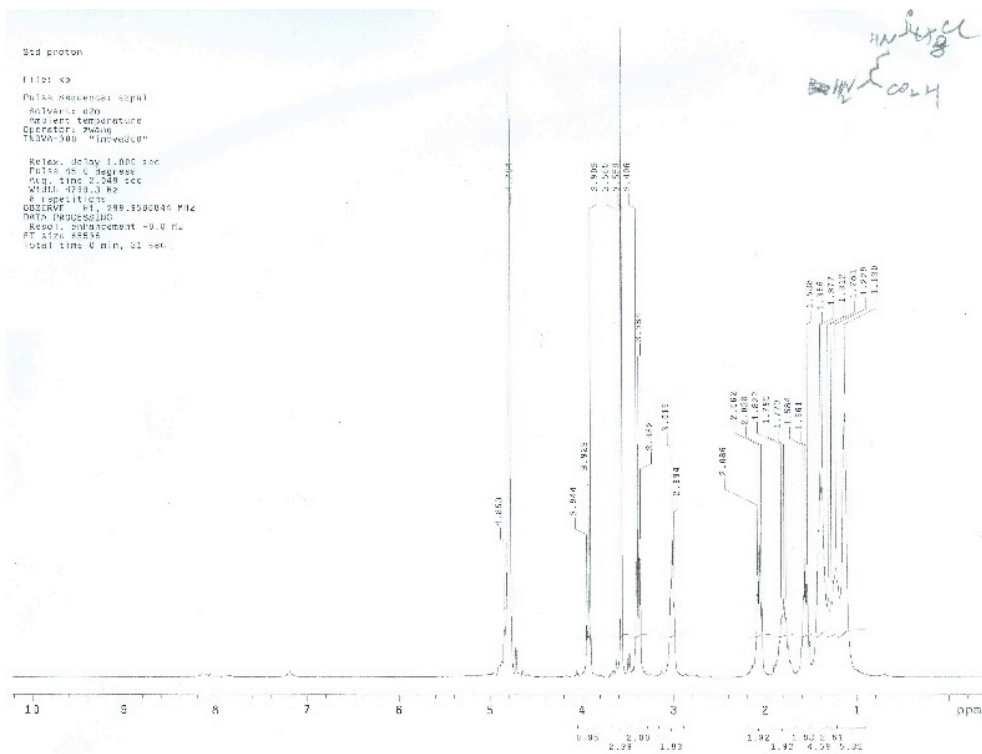


Figure 68  $^1\text{H}$ NMR and  $^{13}\text{C}$ NMR of H-Lys(8-chlorooctanyl)-OH

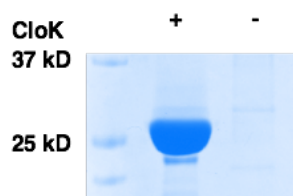


### 6.6.2 The incorporation of H-Lys(8-chlorooctanyl)-OH (13)

We had previously selected a specific PylRS mutant for the incorporation of Lys with side chain fatty acylation modifications<sup>[131]</sup>. By applying the same mmOckRS to co-transform *E. coli* BL21(DE3) cells together with another plasmid pBAD-sfGFP-D134TAG containing a gene for the expression of the sfGFP with an amber mutation at the D134 position as well as a His<sub>6</sub> tag at the C-terminus. A single colony was picked and let to grow in 5 mL LB medium with 100 mg/mL Amp and 34 mg/mL Cm at 37 °C for o/n. Then, the transformed cells were used to inoculate 50 mL 2YT medium with 100 mg/mL Amp and 34 mg/mL Cm, which can express full-length sfGFP with CloK at its D134 position (sfGFP-D134CloK) after been induced by 0.2% arabinose and 0.5 mM IPTG.

After growing at 37 °C for 6 h, the cells were collected by centrifugation (6000 rpm for 15 min). After been resuspended in a 25 mL lysis buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 10 mM imidazole, pH 8.0) and sonicated in an ice water bath three times (4 min each, with a 5 min interval between each run), the cell lysate was precipitated by centrifugation (10000 rpm for 30 min). The supernatant was added to 1 mL Ni Sepharose™ 6 Fast Flow column (GE Healthcare) and the flow-through was then removed. The resin was washed with a wash buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 20 mM imidazole, pH 8.0) and finally eluted with an elution buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, 300 mM NaCl, 250 mM imidazole, pH 8.0). After been concentrated with Amicon Ultra-15 Centrifugal Filter Devices (10000 MWCO cut, Millipore), the purified sfGFP-D134AcdK was analyzed by SDS-PAGE. A clear band can be clearly seen only with 2 mM CloK afforded in the media,

while none for the one without CloK added (Figure 69). The yield of sfGFP-D134AcidK is 7 mg/L with the presence of 0.5 mM CloK, determined by the BCA method.

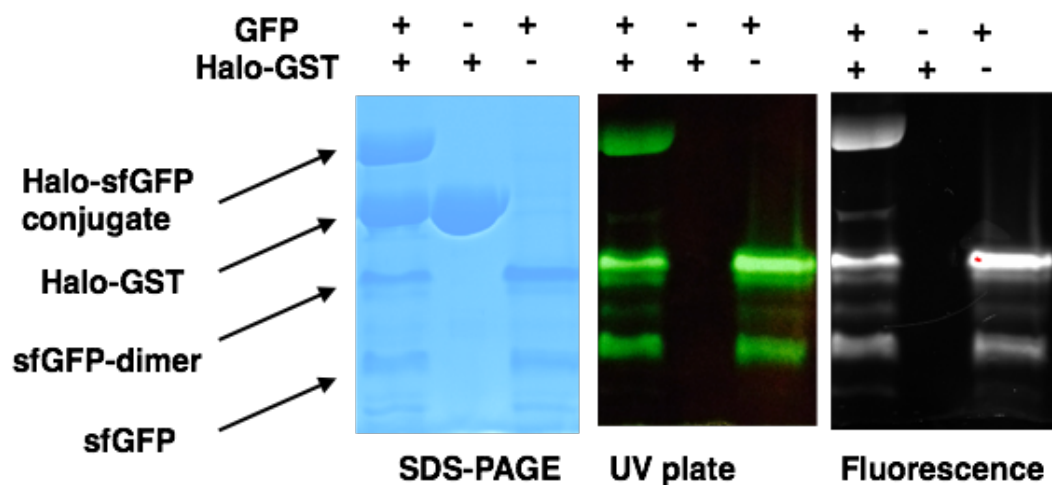


**Figure 69** The expression of sfGFP-D134CloK

### 6.6.3 The binding test of HaloTag protein

With the kinetics data at hand, we further carried out the HaloTag-substrate protein binding assay. To get an optimized condition for further membrane blotting, we first tried to directly incubate HaloTag protein with sfGFP-D134CloK for some time to let them react under aqueous condition. A volume of 20  $\mu$ L reaction mixture containing 22.2  $\mu$ g sfGFP-D134CloK and 27.4  $\mu$ g HaloTag protein was pipetted to mix well. After 1 h incubation, the sample was loaded to 15 % SDS-PAGE. Two negative controls with only HaloTag protein or with only sfGFP-D134CloK were also loaded. After finishing running, the protein sample was first visualized under UV plates and then stained with Coomassie blue.

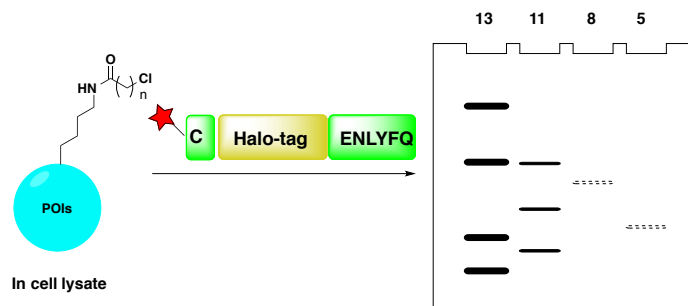
The three gel figures of a same gel from Coomassie blue staining, UV plate and fluorimeter are listed in parallel (Figure 70). In lane 1, a mixed sfGFP-D134CloK with HaloTag protein sample was loaded, while control for HaloTag and sfGFP-D134CloK was loaded to lane 2 and lane 3, respectively. A new band with higher molecular weight as the conjugate can be observed only in lane 1.



**Figure 70** The binding test of HaloTag protein with sfGFP-D134CloK

### 6.7 HaloTag protein labeling

For the purpose of metabolic labeling and proteomic study, we hope to label HaloTag protein with fluorescence dye or Biotin. In our experimental design, the mammalian cell lysate will be loaded to run the whole protein SDS-PAGE gel, which is fed with synthetic chlorofatty acids. Then we can apply the labeled HaloTag protein to blot the membrane serving as an artificial first antibody for long chain fatty acids (Figure 71).



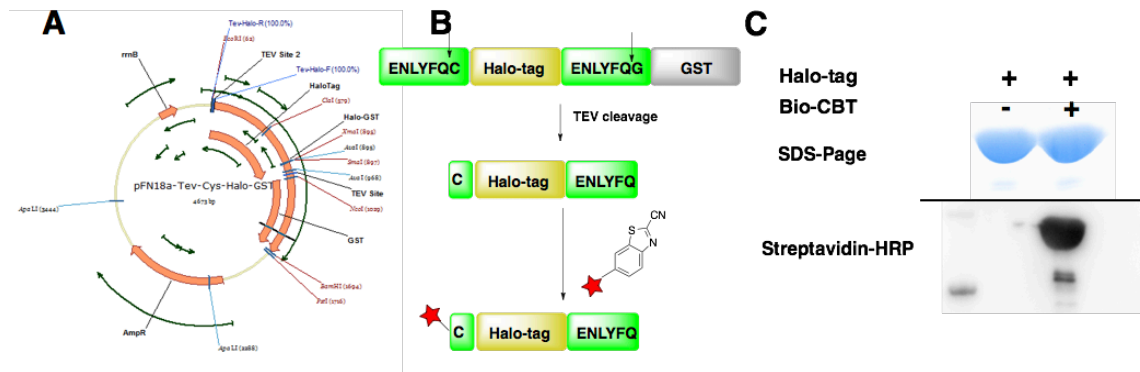
**Figure 71 Screen for the metabolic labeling with biotin conjugated HaloTag protein**

### 6.7.1 The labeling of HaloTag protein

The initial trial of labeling was based on the commercially available biotin-OSu active ester. However, the removal of free excessive dye turned out to be a problem. Inspired by the 2-Cyanobenzothiazole (CBT)-Cys bioorthogonal reaction, CBT-dye labeling is a selective reaction with high yield and no background<sup>[132]</sup>.

We carried out the molecular cloning to insert another TEV cleavage site to the N-terminal of HaloTag protein, besides the one in between HaloTag and GST tag. The original TEV recognition sequence as Glu-Asn-Liu-Tyr-Phe-Gln-Gly, while the last Gly was mutated to Cys for our purpose of labeling (Figure 72A). The novel plasmid was used to transform *E. coli* BL21(DE3) cells, and the target protein was purified with a similar process as reported in Part 6.5.1. Before column elution, an on-resin TEV cleavage was carried out for the removal of GST tag and to release a N-terminal Cys residue for the CBT labeling (Figure 72B). A concentration of 20  $\mu$ M TEV in 1 mL PBS was added to the column and let to incubate for 24 h. With the removal of GST tag, the protein will fall off the GST-resin to solution. The final enzyme was aliquoted and stored at -20  $^{\circ}$ C.

A final concentration of 10 mM CBT-Biotin made by Dr Yadagiri Kurra was added to HaloTag protein and let to incubate at 37 °C for 12 h. After labeling, the HaloTag protein was analyzed with SDS-PAGE, transferred to membrane, and then visualized with Streptavidin-HRP (Figure 72C). Only after the labeling, a clear dark band can be observed for the biotin conjugated HaloTag protein.



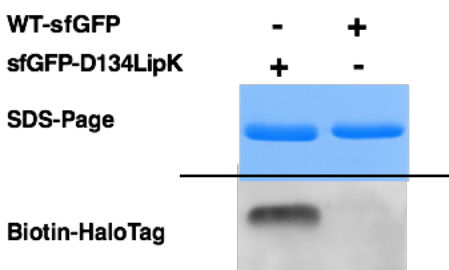
**Figure 72 The biotin labeling of HaloTag protein and its gel characterization** (A) the insertion of TEV protease site at the C-terminal of HaloTag; (B) the TEV cleavage followed by the CBT-dye labeling of HaloTag; (C) the SDS-PAGE and Western blot of labeled HaloTag by Streptavidin-HRP.

### 6.7.2 The blotting examination of HaloTag protein on membrane

Before the final metabolic labeling on real cell lysate, we hope to confirm its potential application as a first antibody with purified protein containing chlorofatty acids. The HaloTag protein after labeling was dialyzed to completely remove the small molecular reactants.

Both WT-GFP and sfGFP-D134CocK previously mentioned were loaded to SDS-PAGE followed by the transfer to membrane. The membrane was first blotted with milk buffer (PBS with 15 % milk powder) at r.t. for 1 h, and then incubated with HaloTag

solution with a final concentration of around 10  $\mu$ M. After blotting at 4  $^{\circ}$ C for 2 h, the membrane was directly visualized by the commercially available Streptavidin-Horseradish peroxidase (HRP). A clear band can be detected only for the sfGFP-D134CocK, while the same of protein amounts for both lanes provided (Figure 73). This demonstrated the successful membrane blotting with Biotin conjugated HaloTag.



**Figure 73 The membrane blotting with Biotin conjugated HaloTag protein**

## 6.8 Metabolic labeling on HEK293T cells

After the methodology demonstration, we started to move ahead to our final application as the metabolic labeling of mammalian cells. We decided to choose HEK293T cell line as the starting point and do proteomic analysis over the entire proteome after feeding chlorofatty acids with different chain length. Noted that this part is not fully finished when this manuscript is prepared.

### 6.8.1 Cell culture of HEK293T cell

The original HEK293 cell prepared by previous lab member Sasha Chihak was stored in liquid Nitrogen. The tubes with HEK293T cell was placed in a 60  $^{\circ}$ C water bath for 2 minutes until thawed. The cells were carefully transferred from the vial to a 15 mL

conical tube with 37 °C culture medium. After centrifuging to remove the supernatant (1 min, 100 rpm), 15 ml DMEM with 10 % FBS was added and the cell was placed on a 100 mm culture dish. The dish was put in the incubator and let to grow under 37 °C with 5% CO<sub>2</sub> for o/n. The cell growth was monitored by a microscope.

The HEK293T cells on dish were passaged when reaching 80% confluent. The medium was removed by gently pipetting without touching cell surface, followed by washing with PBS. After carefully removing the PBS, trypsin buffer was then added to the dish and let to incubate in 37 °C incubator for about 2 min. The cell aspiration was monitored by microscope to guarantee the release of cells from the plate bottom. Then 10 mL DMEM was added to inactivate the trypsin inside, followed by the resuspend of the cells by careful pipetting. The cell suspension was added to new sterile dishes and incubated under 37 °C with 5% CO<sub>2</sub>.

#### 6.8.2 Cell culture with chlorofatty acids

After passaging the HEK293T cells for 3 generations, we started to feed the cells with the synthetic chlorofatty acids. During the last passaging process, the cell was transferred to a 4 cm 6-well plate with 3 mL each and let to grow in the incubator under 37 °C with 5% CO<sub>2</sub> till confluent reached 50 %.

All four  $\omega$ -chlorofatty<sub>n</sub>-acids (n = 4, 7, 10, 12) were first dissolved in sterile DMSO after filtration to a final concentration of 50 mM. Then the stock solution was diluted with DMEM/FBS to a final concentration of 0.1 mM, respectively. Mammalian cells are trypsinized and quenched with the DMEM with chlorofatty acids. The solution was added

to another 4 cm 6-well plate with 3 mL each and let to grow in the incubator under 37 °C with 5% CO<sub>2</sub>. The cell growth was regularly monitored with microscope.

The cell showed lower growing rate generally with the addition of fatty acids. After 24 h, the cell confluency reached 50 %. Then the medium was removed followed by washing with PBS once for further lysis.

### 6.8.3 Cell lysis and sample preparation

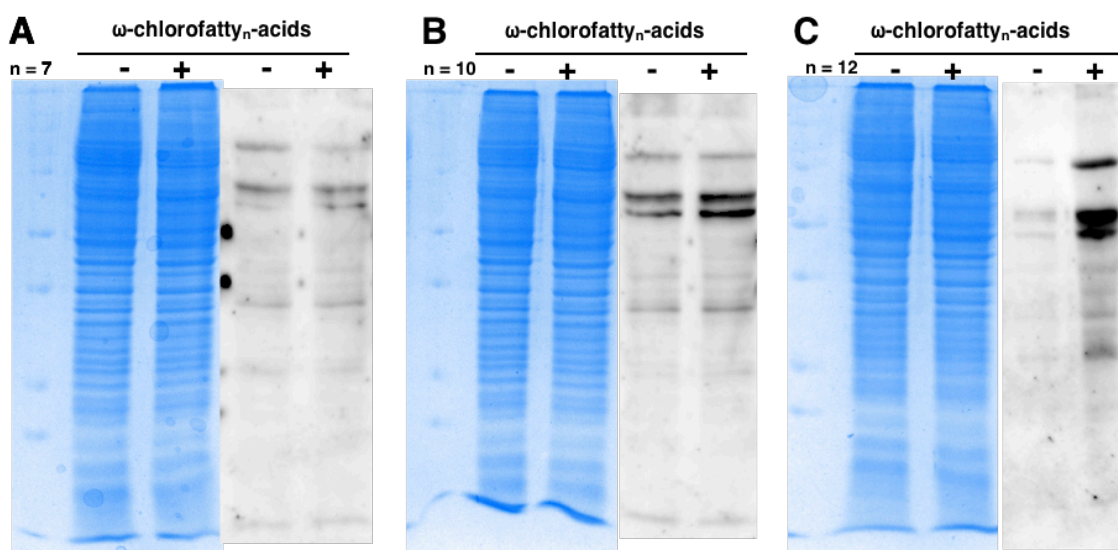
For each well on the 6-well plate, 100 µL ice-cold RIPA lysis buffer with protease inhibitor cocktail was added and let to incubate on ice for 5 min. Then, the cells are carefully gathered together by a scraper, and the cell lysate was collected in 1.5 mL Eppendorf tube. The cell lysate was quickly frozen over dry ice-acetone bath, and let to thaw in 37 °C water bath, followed by the vigorous shaking by vortexing for 1 min. Repeat this fast freeze-thaw cycle for 5 times to complete break the aggregations or chunks in the cell lysate, and the tube was centrifuged for 15 min over 14000 rpm. The supernatant was kept and whole cell lysate BSA assay was carried out to determine the entire protein concentration. The readings of all five including control were between 1.0-5.0 µg/mL. A standard acetone precipitation was carried out to precipitate all the proteins in the cell lysate. Each sample was aliquoted as 100 µg cell lysate dissolving in 20 µL PBS and then stored at -20 °C for further gel analysis.



#### 6.8.4 Gel analysis

After normalization, the cell lysates with or without  $\omega$ -chlorofatty<sub>n</sub>-acids (n = 4, 7, 10, 12) diet were analyzed by two gradient SDS-PAGE gels. One of them was stained by Coomassie blue and the other was transferred to membrane for the Western blot analysis. After standard blotting with milk, Biotin conjugated HaloTag protein was used as first antibody to blot the membrane.

For the protein myristoylation, several darker bands could be observed for the experimental group fed with 13-chlorotridecanoic acid (n = 12) as the myristic acid mimetic (Figure 74C), while the entire protein amount and pattern remained almost identical. Similar situations happened for 11-chloroundecanoic acid (n = 10) and 8-chlorooctanoic acid (n = 7) (Figure 74A&B), which suggesting a wide acylation of cellular proteome by the dietary fatty acids. But almost no difference can be seen for 5-chloropentanoic acid (n = 4), which may result from a fast metabolic degradation kinetics due to the fact of short chain length. Notably, there are many essence Biotin conjugated proteins in cells, and thus even without the addition of  $\omega$ -chlorofatty<sub>n</sub>-acids, some of the cellular proteins could still be visualized with Streptavidin-HRP.



**Figure 74 The metabolic labeling of HEK293T cell lysate via biotin labeled HaloTag protein (A) SDS-PAGE and Western blot of cells fed with  $\omega$ -chlorofatty-acid  $n = 7$ ; (B) SDS-PAGE and Western blot of cells fed with  $\omega$ -chlorofatty-acid  $n = 10$ ; (C) SDS-PAGE and Western blot of cells fed with  $\omega$ -chlorofatty-acid  $n = 12$ .**

## 6.9 Metabolic labeling on HEK293T cells

Since the fatty acylation can be modified over either Lys or Cys side chain or the N-terminus, we want to further identify them and study their distribution.

### 6.9.1 Background design

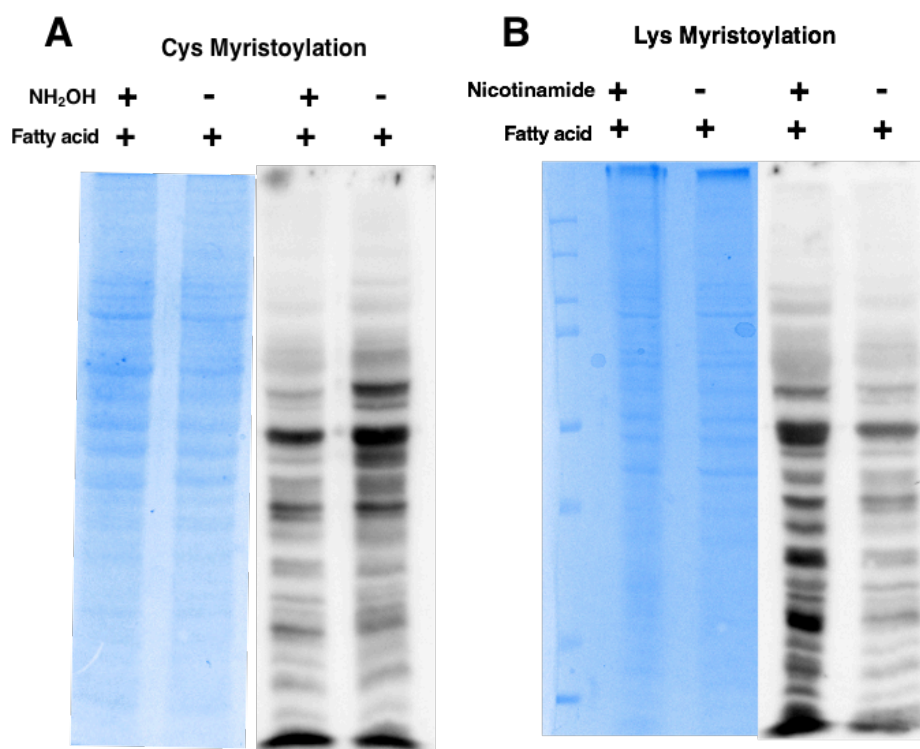
For Cys acylation, since the thioester linkage is not as stable as amide bond, we can use nucleophilic reagents such as hydroxylamine to remove them, which can be further detected via the Western blot analysis. On the other hand, Sirtuin family enzymes are known to remove Lys acylation modifications, which can all be inhibited by nicotinamide. Then we hypothesized the addition of nicotinamide into the culture could inhibit the Sirtuin activity. Thus, the bands on the Western blot with increasing amount of intensity

will be acylations on Lys residues. To demonstrate the principle, only the 13-chlorotridecanoic acid will be used in the next part as a mimetic for myristic acid.

### 6.9.2 Cys myristoylation

To start with, we wanted to separate the Cys myristoylation modifications from the Lys ones. The cells fed with 13-chlorotridecanoic acid were lysed with the same process as previously mentioned. After been transferred to membrane and blotted by milk buffer, PBS with 10 % hydroxylamine was first added to blot the membrane for 1 h, followed by the blotting with Biotin conjugated HaloTag protein as first antibody. And cells blotting with PBS were used as negative control that could still show Cys myristoylation (Figure 75A).

From the Western blot result we can clearly tell that after the treating with hydroxylamine, a lighter band pattern was observed, while the entire protein amount was kept the same. This showed that by nucleophilic attacking the thioester bond of Cys myristoylation, less labeled proteins can be detected, and this further confirmed the wide existence of myristoylation on Cys residues.



**Figure 75** The metabolic labeling of HEK293T cell lysate under different conditions via biotin labeled HaloTag protein (A) SDS-PAGE and Western blot of cells with hydroxylamine treatment to remove Cys myristoylation; (B) SDS-PAGE and Western blot of cells fed with nicotinamide to inhibit Sirtuin enzymes.

### 6.9.3 Lys myristoylation

The next step, we wanted to distinguish the Lys myristoylation by inhibiting Sirtuin family enzymes. A final concentration of 100 mM nicotinamide was added to the medium together with 13-chlorotridecanoic acid to feed the cells with the same process as previously mentioned. After transferring the gel with cell lysate to membrane, standard milk blotting followed by the blotting with Biotin conjugated HaloTag protein was then carried out.

The Western blot result showed that after the addition of nicotinamide, many new bands can be detected in a different pattern, while the entire protein amount was kept the same (Figure 75B). This demonstrated that by inhibiting Sirtuin enzymes we can boost up the fatty acylation level in mammalian cells, and we are also able to clearly see the myristoylation on Lys residues.

## 6.10 Summary

Many questions remain for the study of Lys lipidation especially for Lys fatty acylation modifications. For protein fatty acylation, modifications have been discovered mainly on the two amino acid residues as Lys and Cys and the N-terminus. Although fatty acylation has been found on many important protein substrates which plays vital regulatory roles in epigenetic control of structures and functions, their detailed molecular mechanisms are remaining unclear. To solve the question of lacking reporting groups, we hope to introduce a novel design based on the HaloTag and chlorofatty acid interaction.

We demonstrated the wide application of HaloTag in recognition of fatty acids with long, medium, or even short chains. With this technique, proteins with various modifications of dietary fatty acids and their derivatives can be distinguished and collected for further identification and characterization in combination of proteomic techniques. By applying biotin labeled HaloTag protein as a probe, we can directly visualize whole cell lysate and find out proteins with fatty acylations. With the new technique, a deeper understanding of the impact of dietary fatty acids to human life will be generated from an epigenetic point of view.

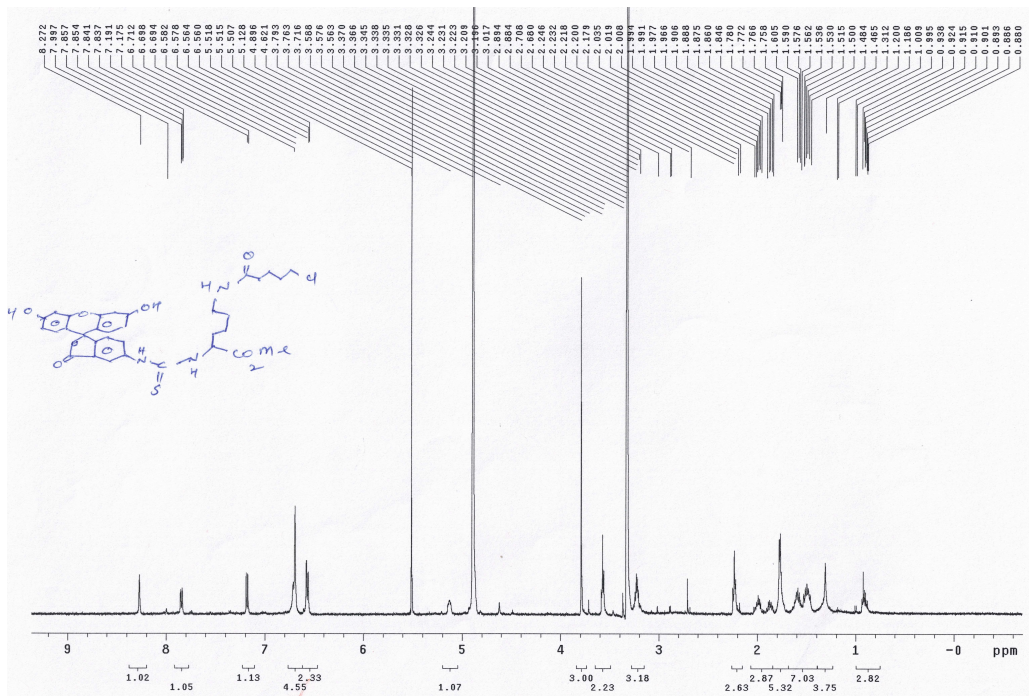


Figure 76 <sup>1</sup>H NMR of FITC-Lys(5-chloropentanyl)-OMe

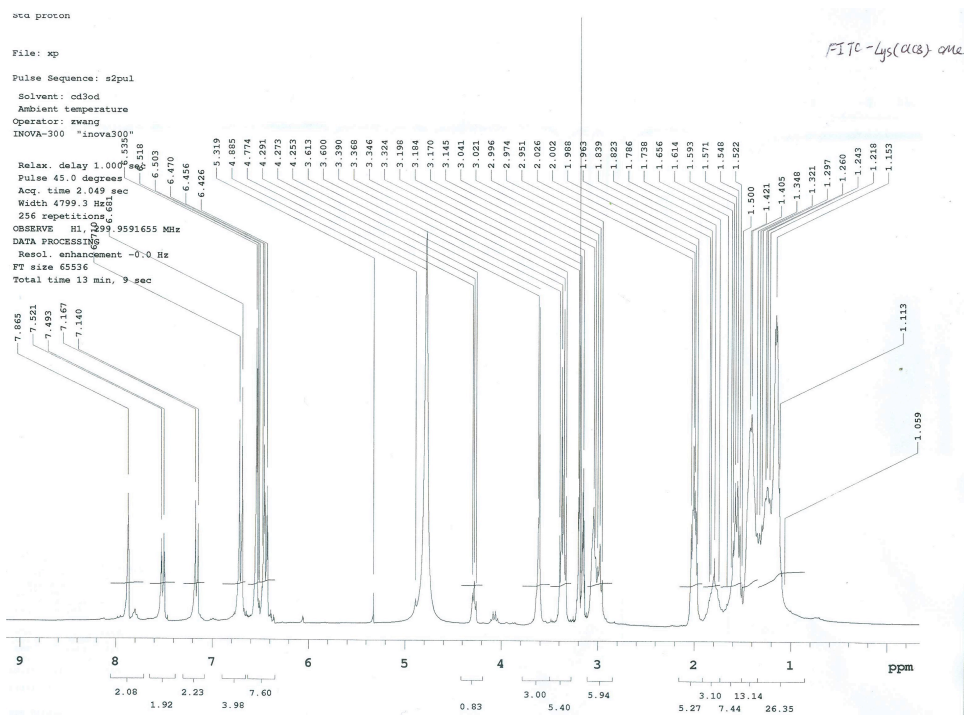


Figure 77 <sup>1</sup>H NMR of FITC-Lys(8-chlorooctanyl)-OMe

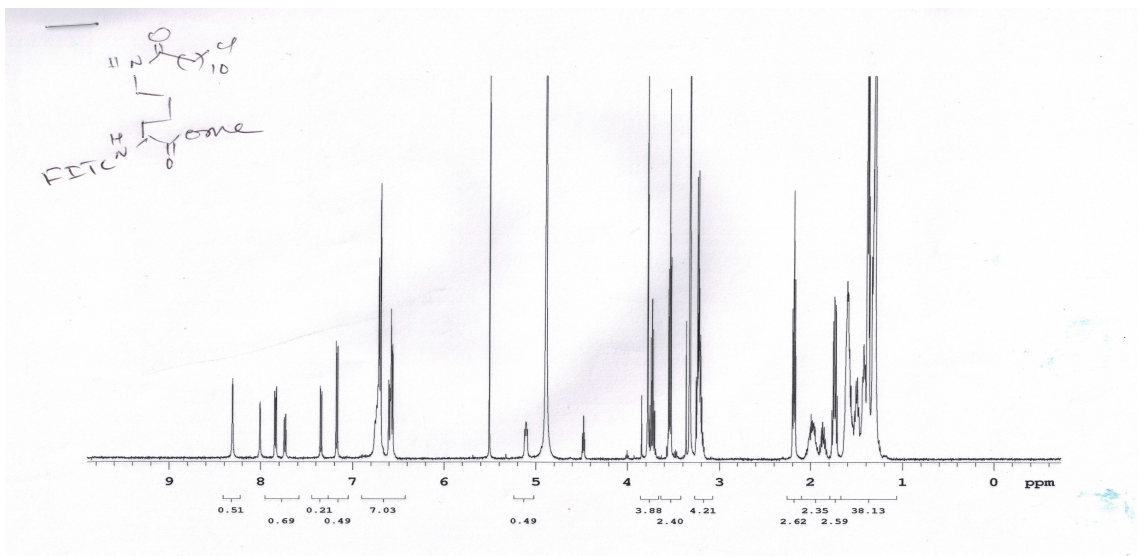


Figure 78  $^1\text{H}$ NMR of FITC-Lys(11-chloroundecanyl)-OMe

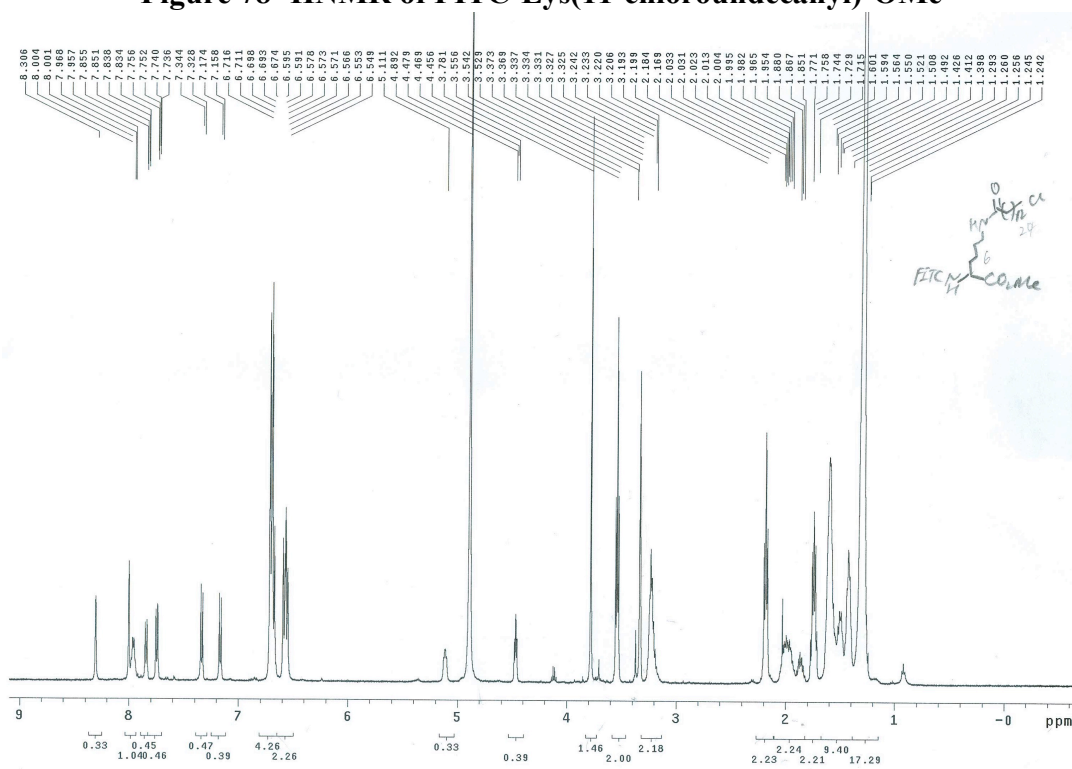


Figure 79  $^1\text{H}$ NMR of FITC-Lys(13-chlorotridecanyl)-OMe

## CHAPTER VII

### CONCLUSION AND OUTLOOK

#### 7.1 Conclusion

Over the past decade, an increasing amount of protein PTMs have been gradually discovered due to the development of analytical instruments, which are of great significance to many cellular activities. Especially for Lys residue, many novel PTM types with a large variety of structures have been characterized widely existing in human proteome. However, the detailed molecular mechanisms are remaining unclear. One of the main factors that are greatly harmful for the functional study is to gain sufficient amounts of target proteins with site-specific PTMs. Recombinant protein expression can, no doubt, achieve most of the goals when facing different target proteins. But it can hardly be applied to proteins with PTMs. Chemical biology has enabled us to have a fine control of theoretically any proteins at atomic level. However, many previous strategies are not robust or versatile enough when facing more and more challenges.

In this dissertation, a new combinatorial strategy with ncAA and bioorthogonal reaction has been described. More specifically, two ncAA based Lys PTM installation systems are mainly introduced, followed by the direct application of the protein products with site-specific methyllysine or acyllysine into enzymatic assays. With high selectivity and wide applicability of bioorthogonal reactions, ubiquitin, sfGFP, and hp53 were applied as water-soluble protein models with small, medium, and large volume, respectively, while histone H3 as a water non-soluble protein. All of our designs are proved to work as expected on these model proteins to generate site specific Lys PTMs.



The protein products are submitted to many important epigenetic enzymatic assays as LSD1, Tip60, Sirt5, etc.

Importantly, this strategy is also providing different functional groups for other chemical biology handling. For AcdK system, the AIK generated contains an aldehyde group which can serve as an electrophilic center with high reactivity over different nucleophiles. AznL, on the other hand, is also providing an azido group as a key chemical biology function group. Many important reactions can undergo with this azido group, such as “Click reaction”, Staudinger reduction, Staudinger ligation, traceless Staudinger ligation, and so on. Therefore, both AcdK and AznL can be further applied for a wider chemical biology functionalization purposes.

## **7.2 Outlook**

Besides the challenges mentioned in each chapter, there are several overall points worth mentioning here. To start with, many PTM targets are still unachievable with our strategy. For instance, lysine trimethylation Kme3 carrying a positive charge cannot be directly incorporated by PylRS, while there is no way to generate a tertiary amine from reductive amination reaction. All the reported examples of site-specific installation of Kme3 come from protein chemical synthesis. The development of current organic synthesis especially the metal-catalyzed reaction is providing our new insights into the building up of those common structures such as amide bond<sup>[109]</sup>. To introduce them into protein chemical biology field may help to set up new avenues in our trials of solving those cumbersome problems<sup>[133]</sup>.

The optimization of bioorthogonal reaction conditions are definitely one of the main developing directions. The modification at a single amino acid residue makes the product very similar with the starting material. Currently widely used reactions cannot reach full conversion easily, which brings the concern of product purification. Another major challenge is to modify biomolecules in higher levels not just purified protein or protein complex<sup>[121]</sup>. How to apply more and more chemical reactions really into the biological systems are still not applicable in most cases, not only in cell lysate or living cells<sup>[134]</sup>, but also in higher organs such as tissue or entire animal level. To achieve this goal, more robust and versatile reactions are welcome to be introduced to the field.

## REFERENCES

- [1] C. C. Liu, M. C. Jewett, J. W. Chin, C. A. Voigt, *Nat Chem Biol* **2018**, *14*, 103-106.
- [2] K. W. Barber, J. Rinehart, *Nat Chem Biol* **2018**, *14*, 188-192.
- [3] Z. Fu, E. R Gilbert, D. Liu, *Current Diabetes Reviews* **2013**, *9*, 25-53.
- [4] M.-Q. Xu, M. W. Southworth, F. B. Mersha, L. J. Hornstra, F. B. Perler, *Cell* **1993**, *75*, 1371-1377.
- [5] M. Unoki, A. Masuda, N. Dohmae, K. Arita, M. Yoshimatsu, Y. Iwai, Y. Fukui, K. Ueda, R. Hamamoto, M. Shirakawa, *J Biol Chem* **2013**, *288*, 6053-6062.
- [6] A. N. Glazer, *Annu Rev Biochem* **1970**, *39*, 101-130.
- [7] T. Kouzarides, *Cell* **2007**, *128*, 693-705.
- [8] B. A. Garcia, S. B. Hake, R. L. Diaz, M. Kauer, S. A. Morris, J. Recht, J. Shabanowitz, N. Mishra, B. D. Strahl, C. D. Allis, D. F. Hunt, *J Biol Chem* **2007**, *282*, 7641-7655.
- [9] J. A. Latham, S. Y. Dent, *Nat Struct Mol Biol* **2007**, *14*, 1017-1024.
- [10] Y. Zhao, O. N. Jensen, *Proteomics* **2009**, *9*, 4632-4641.
- [11] C. A. Musselman, M. E. Lalonde, J. Cote, T. G. Kutateladze, *Nat Struct Mol Biol* **2012**, *19*, 1218-1227.
- [12] M. Shogren-Knaak, H. Ishii, J.-M. Sun, M. J. Pazin, J. R. Davie, C. L. Peterson, *Science* **2006**, *311*, 844-847.
- [13] R. Hamamoto, V. Saloura, Y. Nakamura, *Nature Reviews Cancer* **2015**, *15*, 110-124.

- [14] E. L. Greer, Y. Shi, *Nature Reviews Genetics* **2012**, *13*, 343-357.
- [15] S. Chuikov, J. K. Kurash, J. R. Wilson, B. Xiao, N. Justin, G. S. Ivanov, K. McKinney, P. Tempst, C. Prives, S. J. Gamblin, *Nature* **2004**, *432*, 353-360.
- [16] Z.-P. Wang, Y.-H. Wang, G.-C. Chu, J. Shi, Y.-M. Li, *Curr Org Synth* **2015**, *12*, 150-162.
- [17] C. Choudhary, B. T. Weinert, Y. Nishida, E. Verdin, M. Mann, *Nat Rev Mol Cell Bio* **2014**, *15*, 536-550.
- [18] A. J. de Ruijter, A. H. Van Gennip, H. N. Caron, K. Stephan, A. B. Van Kuilenburg, *Biochem J* **2003**, *370*, 737-749.
- [19] H. Jing, H. Lin, *Chem Rev* **2015**, *115*, 2350-2375.
- [20] M. D. Hirshey, Y. Zhao, *Molecular & cellular proteomics : MCP* **2015**, *14*, 2308-2315.
- [21] H. Huang, S. Lin, B. A. Garcia, Y. Zhao, *Chem Rev* **2015**, *115*, 2376-2418.
- [22] Y. Wang, Y. R. Guo, K. Liu, Z. Yin, R. Liu, Y. Xia, L. Tan, P. Yang, J.-H. Lee, X.-j. Li, *Nature* **2017**, *552*, 273-277.
- [23] D. Komander, M. Rape, *Annu Rev Biochem* **2012**, *81*, 203-229.
- [24] A. G. van der Veen, H. L. Ploegh, *Annu Rev Biochem* **2012**, *81*, 323-357.
- [25] S. Tang, L. J. Liang, Y. Y. Si, S. Gao, J. X. Wang, J. Liang, Z. Mei, J. S. Zheng, L. Liu, *Angew Chem Int Ed* **2017**, *56*, 13333-13337.
- [26] H. Xiao, W. Xuan, S. Shao, T. Liu, P. G. Schultz, *ACS Chem Biol* **2015** *10*, 1599-1603.

- [27] Z. Xie, D. Zhang, D. Chung, Z. Tang, H. Huang, L. Dai, S. Qi, J. Li, G. Colak, Y. Chen, C. Xia, C. Peng, H. Ruan, M. Kirkey, D. Wang, L. M. Jensen, O. K. Kwon, S. Lee, S. D. Pletcher, M. Tan, D. B. Lombard, K. P. White, H. Zhao, J. Li, R. G. Roeder, X. Yang, Y. Zhao, *Mol Cell* **2016**, *62*, 194-206.
- [28] J. Jin, B. He, X. Zhang, H. Lin, Y. Wang, *J Am Chem Soc* **2016**, *138*, 12304-12307.
- [29] S. R. Bhaumik, E. Smith, A. Shilatifard, *Nat Struct Mol Biol* **2007**, *14*, 1008-1016.
- [30] P. A. Cole, *Nat Chem Biol* **2008**, *4*, 590.
- [31] B. D. Strahl, C. D. Allis, *Nature* **2000**, *403*, 41-45.
- [32] Y. T. Kwon, A. Ciechanover, *Trends Biochem Sci* **2017**, *42*, 873-886.
- [33] R. Yau, M. Rape, *Nature Cell Biology* **2016**, *18*, 579.
- [34] S. Park, D. T. Krist, A. V. Statsyuk, *Chem Sci* **2015**, *6*, 1770-1779.
- [35] J. Liang, L. Zhang, X. L. Tan, Y. K. Qi, S. Feng, H. Deng, Y. Yan, J. S. Zheng, L. Liu, C. L. Tian, *Angew Chem Int Ed* **2017**, *129*, 2788-2792.
- [36] M. Hackett, L. Guo, J. Shabanowitz, D. F. Hunt, E. L. Hewlett, *Science* **1994**, *266*, 433-435.
- [37] F. T. Stevenson, S. L. Bursten, C. Fanton, R. M. Locksley, D. H. Lovett, *P Natl Acad Sci USA* **1993**, *90*, 7245-7249.
- [38] F. Stevenson, S. Bursten, R. Locksley, D. Lovett, *Journal of Experimental Medicine* **1992**, *176*, 1053-1062.
- [39] H. C. Hang, J. P. Wilson, G. Charron, *Acc Chem Res* **2011**, *44*, 699-708.

- [40] H. Jiang, S. Khan, Y. Wang, G. Charron, B. He, C. Sebastian, J. Du, R. Kim, E. Ge, R. Mostoslavsky, H. C. Hang, Q. Hao, H. Lin, *Nature* **2013**, *496*, 110-113.
- [41] J.-S. Zheng, Y. He, C. Zuo, X.-Y. Cai, S. Tang, Z. A. Wang, L.-H. Zhang, C.-L. Tian, L. Liu, *J Am Chem Soc* **2016**, *138*, 3553–3561.
- [42] M. Broncel, R. A. Serwa, P. Ciepla, E. Krause, M. J. Dallman, A. I. Magee, E. W. Tate, *Angew Chem Int Ed* **2015**, *54*, 5948-5951.
- [43] R. Anand, R. Marmorstein, *J Biol Chem* **2007**, *282*, 35425-35429.
- [44] R. Brabham, M. A. Fascione, *ChemBioChem* **2017**, *18*, 1973 – 1983.
- [45] A. El-Faham, F. Albericio, *Chem Rev* **2011**, *111*, 6557-6602.
- [46] J.-M. Kee, R. C. Oslund, D. H. Perlman, T. W. Muir, *Nat Chem Biol* **2013**, *9*, 416-421.
- [47] J.-M. Kee, R. C. Oslund, A. D. Couvillon, T. W. Muir, *Org Lett* **2014**.
- [48] W. W. Hsu, B. Wu, W. R. Liu, *ACS Chem Biol* **2016**, *11*, 792–799.
- [49] P. E. Dawson, T. W. Muir, I. Clark-Lewis, S. B. Kent, *Science* **1994**, *266*, 776-779.
- [50] T. W. Muir, D. Sondhi, P. A. Cole, *P Natl Acad Sci USA* **1998**, *95*, 6705-6710.
- [51] S. B. H. Kent, *Chem Soc Rev* **2009**, *38*, 338-351.
- [52] T. W. Muir, *Annu Rev Biochem* **2003**, *72*, 249-289.
- [53] J.-S. Zheng, S. Tang, Y.-K. Qi, Z.-P. Wang, L. Liu, *Nat Protoc* **2013**, *8*, 2483-2495.
- [54] X. Chen, S. Tang, J.-S. Zheng, R. Zhao, Z.-P. Wang, W. Shao, H.-N. Chang, J.-Y. Cheng, H. Zhao, L. Liu, *Nat Commun* **2015**, *6*, 1-9.

- [55] S. He, D. Bauman, J. S. Davis, A. Loyola, K. Nishioka, J. L. Gronlund, D. Reinberg, F. Y. Meng, N. Kelleher, D. G. McCafferty, *P Natl Acad Sci USA* **2003**, *100*, 12033-12038.
- [56] T. Kawakami, Y. Akai, H. Fujimoto, C. Kita, Y. Aoki, T. Konishi, M. Waseda, L. Takemura, S. Aimoto, *B Chem Soc Jpn* **2013**, *86*, 690-697.
- [57] J. Li, Y. Li, Q. He, Y. Li, H. Li, L. Liu, *Org Biomol Chem* **2014**, *12*, 5435-5441.
- [58] U. T. Nguyen, L. Bittova, M. M. Müller, B. Fierz, Y. David, B. Houck-Loomis, V. Feng, G. P. Dann, T. W. Muir, *Nature methods* **2014**, *11*, 834-840.
- [59] M. Jbara, N. Guttmann-Raviv, S. K. Maity, N. Ayoub, A. Brik, *Bioorg Med Chem* **2017**.
- [60] P. Siman, S. V. Karthikeyan, M. Nikolov, W. Fischle, A. Brik, *Angew Chem Int Ed Engl* **2013**, *52*, 8059-8063.
- [61] Z. Wang, X. Ding, S. Li, J. Shi, Y. Li, *RSC Advances* **2014**, *4*, 7235-7245.
- [62] E. M. Sletten, C. R. Bertozzi, *Angew Chem Int Ed* **2009**, *48*, 6974-6998.
- [63] C. R. Bertozzi, *Acc Chem Res* **2011**, *44*, 651-653.
- [64] X. Chen, Y.-W. Wu, *Org Biomol Chem* **2016**.
- [65] J. Li, P. R. Chen, *Nat Chem Biol* **2016**, *12*, 129-137.
- [66] M. D. Simon, F. Chu, L. R. Racki, C. Cecile, A. L. Burlingame, B. Panning, G. J. Narlikar, K. M. Shokat, *Cell* **2007**, *128*, 1003-1012.
- [67] E. M. Valkevich, R. G. Guenette, N. A. Sanchez, Y. C. Chen, Y. Ge, E. R. Strieter, *J Am Chem Soc* **2012**, *134*, 6916-6919.

- [68] Z. U. Wang, Y.-S. Wang, P.-J. Pai, W. K. Russell, D. H. Russell, W. R. Liu, *Biochemistry* **2012**, *51*, 5232-5234.
- [69] G. J. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, *J Am Chem Soc* **2008**, *130*, 5052-5053.
- [70] J. M. Chalker, S. B. Gunnoo, O. Boutureira, S. C. Gerstberger, M. Fernández-González, G. J. Bernardes, L. Griffin, H. Hailu, C. J. Schofield, B. G. Davis, *Chem Sci* **2011**, *2*, 1666-1676.
- [71] A. Yang, S. Ha, J. Ahn, R. Kim, S. Kim, Y. Lee, J. Kim, D. Söll, H.-Y. Lee, H.-S. Park, *Science* **2016**, *354*, 623-626.
- [72] T. H. Wright, B. J. Bower, J. M. Chalker, G. J. Bernardes, R. Wiewiora, W.-L. Ng, R. Raj, S. Faulkner, M. R. J. Vallée, A. Phanumartwiwath, B. G. Davis, *Science* **2016**, *354*, aag1465.
- [73] M. Imiołek, G. Karunanithy, W.-L. Ng, A. J. Baldwin, V. E. Gouverneur, B. G. Davis, *J Am Chem Soc* **2018**, *140*, 1568–1571.
- [74] W. S. Glenn, S. E. Stone, S. H. Ho, M. J. Sweredoski, A. Moradian, S. Hess, J. Bailey-Serres, D. A. Tirrell, *Plant Physiology* **2017**, *173*, 1543-1553.
- [75] T. Hoshaka, D. Kajihara, Y. Ashizuka, H. Murakami, M. Sisido, *J Am Chem Soc* **1999**, *121*, 34-40.
- [76] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498-500.
- [77] A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, *Chem Sci* **2015**, *6*, 50-69.
- [78] H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat Chem Biol* **2008**, *4*, 232-234.



- [79] M. J. Gattner, M. Vrabel, T. Carell, *Chemical communications (Cambridge, England)* **2013**, *49*, 379-381.
- [80] C. H. Kim, M. Kang, H. J. Kim, A. Chatterjee, P. G. Schultz, *Angew Chem Int Ed Engl* **2012**, *51*, 7246-7249.
- [81] D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *J Am Chem Soc* **2009**, *131*, 14194-14195.
- [82] H. W. Ai, J. W. Lee, P. G. Schultz, *Chem Commun (Camb)* **2010**, *46*, 5506-5508.
- [83] a Y.-S. Wang, B. Wu, Z. Wang, Y. Huang, W. Wan, W. K. Russell, P.-J. Pai, Y. N. Moe, D. H. Russell, W. R. Liu, *Molecular BioSystems* **2010**, *6*, 1557-1560; b D. Groff, P. R. Chen, F. B. Peters, P. G. Schultz, *Chembiochem* **2010**, *11*, 1066-1068.
- [84] W. Wan, J. M. Tharp, W. R. Liu, *Biochim Biophys Acta (BBA)-Proteins and Proteomics* **2014**, *1844*, 1059-1070.
- [85] D. P. Nguyen, M. M. G. Alai, S. Virdee, J. W. Chin, *Chem Biol* **2010**, *17*, 1072-1076.
- [86] J. Luo, Q. Y. Liu, K. Morihira, A. Deiters, *Nat Chem* **2016**, *8*, 1027-1034.
- [87] Z. A. Wang, Y. Zeng, Y. Kurra, X. Wang, J. M. Tharp, E. C. Vatansever, W. W. Hsu, S. Dai, X. Fang, W. R. Liu, *Angew Chem Int Ed* **2017**, *56*, 212–216.
- [88] M. Yamauchi, M. Sricholpech, *Essays Biochem* **2012**, *52*, 113-133.
- [89] Y.-S. Wang, X. Fang, A. L. Wallace, B. Wu, W. R. Liu, *J Am Chem Soc* **2012**, *134*, 2950-2953.
- [90] G. Srinivasan, C. M. James, J. A. Krzycki, *Science* **2002**, *296*, 1459-1462.

- [91] J. K. Kurash, H. Lei, Q. Shen, W. L. Marston, B. W. Granda, H. Fan, D. Wall, E. Li, F. Gaudet, *Mol Cell* **2008**, *29*, 392-400.
- [92] Y. Tang, J. Luo, W. Zhang, W. Gu, *Mol Cell* **2006**, *24*, 827-839.
- [93] E. Seto, A. Usheva, G. P. Zambetti, J. Momand, N. Horikoshi, R. Weinmann, A. J. Levine, T. Shenk, *P Natl Acad Sci USA* **1992**, *89*, 12028-12032.
- [94] M. H. Park, *Journal of biochemistry* **2006**, *139*, 161-169.
- [95] C. Delgado-Andrade, *Food & Function* **2016**, *7*, 46-57.
- [96] J. R. Wiśniewski, A. Zougman, M. Mann, *Nucleic Acids Res* **2008**, *36*, 570-577.
- [97] E. Saxon, J. I. Armstrong, C. R. Bertozzi, *Org Lett* **2000**, *2*, 2141-2143.
- [98] B. L. Nilsson, L. L. Kiessling, R. T. Raines, *Org Lett* **2000**, *2*, 1939-1941.
- [99] Z. A. Wang, Y. Kurra, X. Wang, Y. Zeng, Y.-J. Lee, V. Sharma, H. Lin, S. Y. Dai, W. R. Liu, *Angew Chem Int Ed* **2017**, *56*, 1643-1647.
- [100] I. C. Tanrikulu, E. Schmitt, Y. Mechulam, W. A. Goddard, 3rd, D. A. Tirrell, *P Natl Acad Sci USA* **2009**, *106*, 15285-15290.
- [101] A. J. Link, M. K. S. Vink, D. A. Tirrell, *Nat Protoc* **2007**, *2*, 1879-1883.
- [102] C. Besanceney - Webler, H. Jiang, T. Zheng, L. Feng, D. Soriano del Amo, W. Wang, L. M. Klivansky, F. L. Marlow, Y. Liu, P. Wu, *Angew Chem Int Ed* **2011**, *50*, 8051-8056.
- [103] M. B. Soellner, B. L. Nilsson, R. T. Raines, *J Am Chem Soc* **2006**, *128*, 8820-8828.
- [104] A. Tam, M. B. Soellner, R. T. Raines, *J Am Chem Soc* **2007**, *129*, 11421-11430.

- [105] J. Du, Y. Zhou, X. Su, J. J. Yu, S. Khan, H. Jiang, J. Kim, J. Woo, J. H. Kim, B. H. Choi, H. Lin, *Science* **2011**, *334*, 806-809.
- [106] J. Du, H. Jiang, H. Lin, *Biochemistry* **2009**, *48*, 2878-2890.
- [107] M. Yang, Y. Wang, Y. Chen, Z. Cheng, J. Gu, J. Deng, *Molecular & cellular proteomics : MCP* **2015**, *14.4*, 796-811.
- [108] A. Tam, R. T. Raines, *Bioorg Med Chem* **2009**, *17*, 1055-1063.
- [109] V. R. Pattabiraman, J. W. Bode, *Nature* **2011**, *480*, 471-479.
- [110] J. W. Bode, R. M. Fox, K. D. Baucom, *Angew Chem Int Ed* **2006**, *45*, 1248-1252.
- [111] J. Wu, J. Ruiz-Rodríguez, J. M. Comstock, J. Z. Dong, J. W. Bode, *Chem Sci* **2011**, *2*, 1976-1979.
- [112] V. R. Pattabiraman, A. O. Ogunkoya, J. W. Bode, *Angew Chem Int Ed* **2012**, *51*, 5114-5118.
- [113] C. E. Murar, F. Thuaud, J. W. Bode, *J Am Chem Soc* **2014**, *136*, 18140-18148.
- [114] I. Pusterla, J. W. Bode, *Angew Chem Int Ed* **2012**, *124*, 528-531.
- [115] H. Noda, G. b. Erős, J. W. Bode, *J Am Chem Soc* **2014**, *136*, 5611-5614.
- [116] H. Noda, J. W. Bode, *Chem Sci* **2014**, *5*, 4328-4332.
- [117] C. E. Weller, A. Dhall, F. Ding, E. Linares, S. D. Whedon, N. A. Senger, E. L. Tyson, J. D. Bagert, X. Li, O. Augusto, C. Chatterjee, *Nat Commun* **2016**, *7*, 12979.
- [118] C. Chatterjee, R. K. McGinty, J. P. Pellois, T. W. Muir, *Angew Chem Int Ed* **2007**, *119*, 2872-2876.

- [119] B. Cowper, L. Shariff, W. Chen, S. M. Gibson, W.-L. Di, D. Macmillan, *Org Biomol Chem* **2015**, *13*, 7469-7476.
- [120] B. T. Weinert, C. Schölz, S. A. Wagner, V. Iesmantavicius, D. Su, J. A. Daniel, C. Choudhary, *Cell reports* **2013**, *4*, 842-851.
- [121] Y. Amamoto, Y. Aoi, N. Nagashima, H. Suto, D. Yoshidome, Y. Arimura, A. Osakabe, D. Kato, H. Kurumizaka, S. A. Kawashima, *J Am Chem Soc* **2017**, *139*, 7568–7576.
- [122] T. Peng, E. Thinon, H. C. Hang, *Curr Opin Chem Biol* **2016**, *30*, 77-86.
- [123] Z. Liu, T. Yang, X. Li, T. Peng, H. C. Hang, X. D. Li, *Angew Chem Int Ed* **2015**, *54*, 1149-1152.
- [124] J. S. Yount, M. M. Zhang, H. C. Hang, *Curr Opin Chem Biol* **2013**, *17*, 27-33.
- [125] P. Henriksen, S. A. Wagner, B. T. Weinert, S. Sharma, G. Bačinskaja, M. Rehman, A. H. Juffer, T. C. Walther, M. Lisby, C. Choudhary, *Molecular & cellular proteomics : MCP* **2012**, *11*, 1510-1522.
- [126] H. C. Hang, E.-J. Geutjes, G. Grotenbreg, A. M. Pollington, M. J. Bijlmakers, H. L. Ploegh, *J Am Chem Soc* **2007**, *129*, 2744-2745.
- [127] M. Grammel, H. C. Hang, *Nat Chem Biol* **2013**, *9*, 475-484.
- [128] Y. Yu-Ying, G. Markus, H. C. Howard, *Bioorg Med Chem* **2011**, *21*, 4976-4979.
- [129] M. Yu, L. P. S. De Carvalho, G. Sun, J. S. Blanchard, *J Am Chem Soc* **2006**, *128*, 15356-15357.

- [130] G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, *ACS Chem Biol* **2008**, *3*, 373-382.
- [131] W. W. Wang, Y. Zeng, B. Wu, A. Deiters, W. R. Liu, *ACS Chem Biol* **2016**, *11*, 1973-1981.
- [132] C. P. Ramil, P. An, Z. Yu, Q. Lin, *J Am Chem Soc* **2016**, *138*, 5499-5502.
- [133] B. Shen, D. M. Makley, J. N. Johnston, *Nature* **2010**, *465*, 1027-1032.
- [134] X. Fang, Y. Fu, M. J. C. Long, J. A. Haegele, E. J. Ge, S. Parvez, Y. Aye, *J Am Chem Soc* **2013**, *135*, 14496-14499.