INVESTIGATE THE SUBSTRATE SPECIFICITY OF PROTEIN POST-TRANSLATIONAL MODIFICATIONS WITH BIOORTHOGONAL CHEMICAL PROBES

by

JIABAO SONG

(Under the Direction of Y. George Zheng)

ABSTRACT

Protein post translational modifications (PTMs) refer to the modification by adding a specific moiety to the amino acid side chain after their biosynthesis. Owing to the advances of high-sensitivity mass spectrometry, more than 400 different types of PTMs have been identified, including phosphorylation, ubiquitylation, glycosylation, methylation and acetylation. Biological investigations of the functionalities of different PTMs have revealed their participation in different crucial cellular processes. Dysregulation of PTMs or corresponding regulatory enzymes could result in dysfunction of critical biological processes and lead to occurrence and progression of various diseases. Despite many efforts to disclose protein modifications in cells by quantitative proteomics, the substrate specificity for individual PTM or its regulatory enzyme is still less investigated, which limits our understanding of how the PTM contributes to a specific cellular process. Therefore, there is an urgent need to profile and identify the protein substrates for each single PTM. Here, we reported our work in the study of Histone acetyltransferase 1 (HAT1) substrates, lysine methacrylylation substrates and cysteine (S)-2-carboxypropylation substrates. First, we created a series of engineered HAT1 and found one could efficiently catalyze the acylation and label its protein substrates combined with a clickable acyl-CoA bioorthogonal

reporter, 3-azidopropanoyl CoA (3AZ-CoA). We identified hundreds of novel protein substrates of HAT1 by chemoproteomic profiling. Second, we developed a sensitive and robust chemical probe which could selectively label and identify protein methacrylylation substrates. Proteomic identification revealed hundreds of sodium methacrylate and valine dependent methacrylylated proteins as well as HAT1 dependent methacrylylated proteins. Last, we designed and synthesized a bioorthogonal chemical probe which was used to discover a new protein modification, (S)-2-carboxypropylation. Through chemoproteomic profiling, we successfully identified hundreds of protein substrates as well as 120 modification sites. Together, the current work expands the researchers' toolbox to study protein PTMs, provides valuable insights in understanding the mechanisms of how PTMs regulate cellular processes and offers new directions for the development of therapeutical methods for PTMs related diseases.

INDEX WORDS: Lysine acetylation, Lysine methacrylylation, S-2-carboxypropylation, Lysine acetyltransferase, Bioorthogonal chemical probes, Protein substrates, Proteomics, Click chemistry

INVESTIGATE THE SUBSTRATE SPECIFICITY OF PROTEIN POST-TRANSLATIONAL MODIFICATIONS WITH BIOORTHOGONAL CHEMICAL PROBES

by

JIABAO SONG

B.S., China Pharmaceutical University, China, 2017

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2023

INVESTIGATE THE SUBSTRATE SPECIFICITY OF PROTEIN POST-TRANSLATIONAL MODIFICATIONS WITH BIOORTHOGONAL CHEMICAL PROBES

by

JIABAO SONG

Major Professor:	Y. George Zheng
Committee:	Arthur Roberts
	Vladimir Popik
	Eugene Douglass

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
May 2023

DEDICATION

The dissertation is dedicated to my family and friends, who have supported me throughout my education. To my beloved parents, who continually provide their moral, spiritual, and financial support. To my girlfriend, Zhe Yang, who has always been a source of happiness, strength, and motivation. I am grateful to have them, who are the most important people in my life.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my major advisor, Dr. Y. George Zheng for his careful patience, unwavering support and expertise to guide me during this journey. He is an excellent scientist who not only has a deep knowledge in the field of epigenetics, but stays curious about every interesting scientific problem, which largely inspired me for my Ph.D. study. I am also thankful for my graduate committee, Dr. Arthur Roberts, Dr. Vladimir Popik and Dr. Eugene Douglass for their helpful advice and constructive feedback.

I want to thank my collaborators Dr. Ronghu Wu and Mr. Kejun Yin for their support on the two projects with the quantitative proteomic analysis and data processing. I am also grateful for the help from all the Zheng laboratory members. I want to thank Dr. Zhen Han, Dr. Liza Ngo and Dr. Kun Qian for their generous help on my projects. I am also thankful for the helpful suggestions on the synthetic procedures form Mr. Tyler Brown. I would like to thank Dr. Melody Fulton, Dr. Bo Zhou, Dr. Zhesi Zhu, Ms. Angela Bae, Mr. Terry Nguyen, Ms. Mengtong Cao, Ms. Qianyun Fu, Ms. Tran Dang, Mr. Nolan Swain, Mr. Bill Trieshmann and all other lab members for the helpful discussions.

Lastly, I want to thank the staff in the department of Pharmaceutical and Biomedical Sciences and Dr. Dennis Phillips and Dr. Chau-Wen Chou from the Proteomics and Mass Spectrometry Facility to help me in my study.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

This work is adapted with permission from the review article: Song, J.; Zheng, Y. G. Bioorthogonal Reporters for Detecting and Profiling Protein Acetylation and Acylation. SLAS Discov 2020, 25 (2), 148-162. DOI: 10.1177/2472555219887144

1.1 Protein Lysine Acetylation and KATs

Lysine acetylation of proteins is an essential posttranslational modification (PTM) that regulates plethoric biological processes from gene transcription, cell cycle, apoptosis, metabolism, signal transduction, to cell differentiation.[1-8] Although nonenzymatic processes are present, the site-specific acetylation reaction is dominantly catalyzed by protein lysine acetyltransferases (KATs) that transfer the acetyl group from acetyl–coenzyme A (acetyl-CoA, Ac-CoA) to the ε-amino group of specific lysine residues in proteins (**Figure 1.1A**). A dozen KAT members in mammalian cells have been identified and characterized both genetically and biochemically, which include the GNAT representative members GCN5, PCAF, and HAT1; five MYST family members (MOF, TIP60, MORF, MOZ, HBO1); p300 and CBP.[9-16] A few other proteins such as CLOCK, NAT10 and NCOAT, though mentioned as KAT members, remain poorly characterized and are sometimes considered as orphan or noncanonical KAT members.[16, 17] Importantly, dysfunction of KATs were found to associate with cancer incidence, progression, and metastasis.[15, 18, 19] For example, it has been found that the expression of PCAF, which can reduce the hepatocellular carcinoma (HCC) growth by acetylating histone H4, is decreased in most of the HCC cells.[20] Upregulation of GCN5 has been detected in human colon adenocarcinoma tissues and suppression

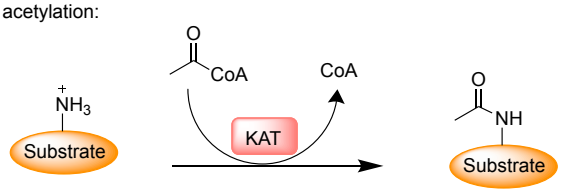
of GCN5 expression inhibits the tumor cell growth, which could serve as a potential therapeutical target.[21] Mechanisms for regulation of human lung cancer cells by HAT1 have also been investigated and the studies suggested that HAT1 can promote lung cancer cell apoptosis via up-regulation the expression of Fas.[22] Moreover, MOF, as a representative KAT in MYST family, has also been found to play important roles in tumor suppressor functions. The decreased expression or mutation of MOF was detected in both breast cancer and medulloblastoma, which regulates the cancer progression via acetylation of histone H4 lysine 16.[23] Additionally, dysfunctions of p300 and CBP are also involved in various malignancies.[24] High expression of p300 and CBP has been detected in colorectal cancer cells.[25] Studies in human colon cancer cells showed that p300 activity is critical for p53 response after DNA damage and suppression of p300 activity in cancer cells aids in chemotherapy.[26] Although the dysfunctions or mutations of different KATs in cancer cells have been elucidated, the knowledge of the specific mechanisms, i.e., the protein substrates regulated by individual KATs, are still largely unknown, which hindered the further development of therapeutical methods. On the other end of the acylation spectrum, long chain fatty acylation have also gained increasing interests in recent years, of which the most extensively studied modifications are *N*-myristoylation on the N-terminal glycine residues catalysed by *N*-myristoyltransferases (NMTs) and *S*-palmitoylation on the cysteine residues via a thioester bond catalysed by palmitoyltransferases (PATs).[27-29] Side chain myristoylation of lysine residues is also identified but no corresponding acyltransferase has been reported so far.

Proteomic studies have revealed the presence of hundreds to thousands of acetylated proteins and acetylation sites throughout the cell, which suggests the prevalence of lysine acetylation in nearly every facet of cell physiology.[30-35] Since KAT enzymes are key regulators for cancer biology, high-throughput assays have been designed to screen drug-lead inhibitors for

specific KATs with the goal of developing new anticancer drugs.[16, 36-38] While the importance of KATs in physiology and disease is widely recognized, functional annotation of KAT enzymes in regulating key biological pathways is far from completely understood. Especially, how the acetylome of individual KAT enzymes distinguishes from one another and how the substrate distribution of KATs is affected by various intracellular or environmental stimuli demand further investigation. A clear biochemical, structural, and proteomic understanding of KAT substrate specificity and the impact of individual KATs in (patho)physiology regulation is greatly needed. Elucidation of molecular targets of KATs represents a leading step toward fully dissecting the roles of KATs in gene regulation and their functions beyond the chromatin biology realm. In this regard, mass spectrometry (MS)-based profiling has provided a plethora of information about acetylated proteins[5, 39, 40], albeit with limited information on enzyme-substrate correlations. Protein microarray is also an effective method in KAT substrate identification on a large scale.[41] Using yeast proteome microarrays, Lin et al. identified many nonchromatin substrates of the nucleosome acetyltransferase of H4 (NuA4) complex.[42]

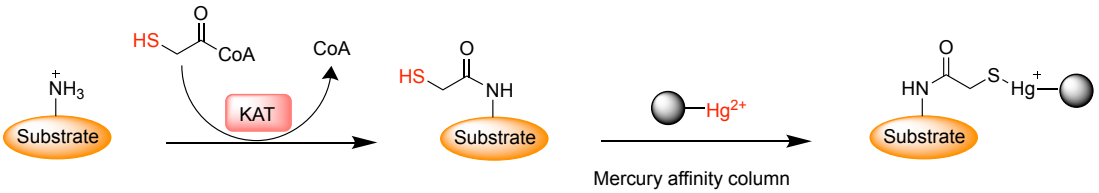
Lysine acetylation:

A

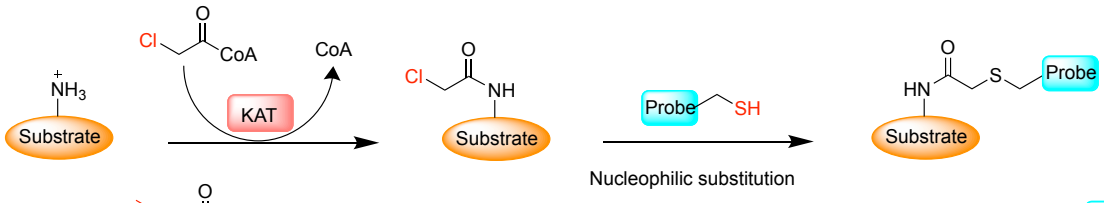


Functionalized acylation:

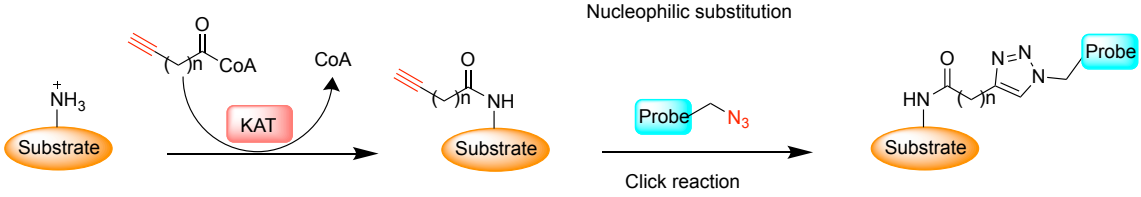
B



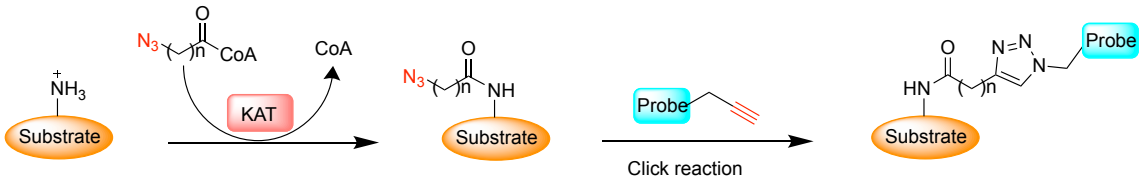
C



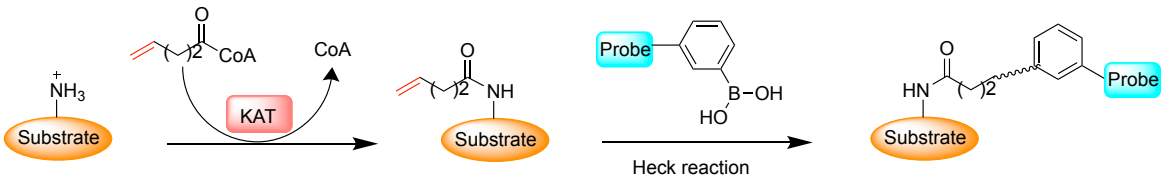
D



E



F



G

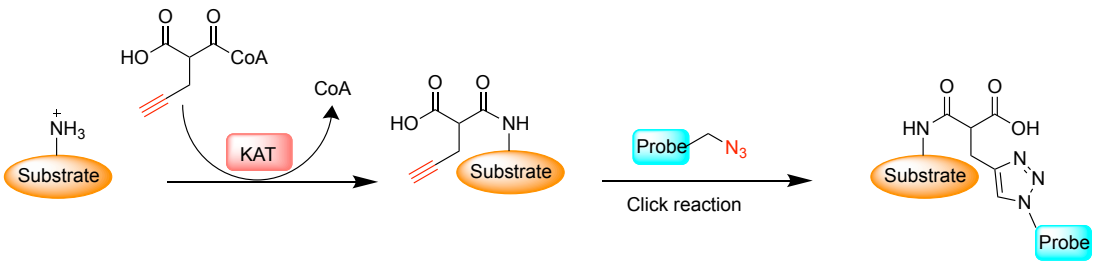


Figure 1.1 Schematic of protein lysine acetylation and its reformulated bioorthogonal acylation using functionalized acyl-CoA surrogates. (A) Normal acetylation of protein substrates is catalyzed by KATs. (B) Mercaptoacetyl-CoA can be utilized by KATs and mercaptoacetate-labeled proteins are isolated with mercury affinity column. (C) Chloroacetyl-CoA can be utilized by KATs and chloroacetylated proteins react with thiol-containing reagents for further detection. (D) Alkyne-labeled CoA surrogates can be utilized by KATs, and the labeled proteins undergo downstream click reaction to conjugate with a probe. (E) Azido-labeled CoA surrogates can be utilized by KATs, and the labeled proteins undergo downstream click reaction to conjugate with a probe. (F) Alkene-labeled proteins can be connected with probes by Heck reaction. (G) Alkyne-labeled malonyl-CoA surrogates are used to detect protein malonylation.

1.2 Protein lysine acylations and cysteine modifications

Except for lysine acetylation, a growing number of short chain lysine acylations have drawn a lot of attention in recent years. For example, there are propionylation, butyrylation, isobutyrylation, crotonylation, methacrylylation, lactylation, 2-hydroxyisobutyrylation, benzoylation, malonylation, succinylation and glutarylation.[43-46] Evidence have shown that these short chain acylations can also be generated enzymatically by above mentioned KATs or formed non-enzymatically.[47] Additionally, it has been noted that some of the KATs possess cofactor promiscuity, which means that they not only catalyze protein acetylation, but also catalyze other protein acylations by transferring the short chain acyl groups from Acyl-CoA to the corresponding protein substrates. For instance, p300 has been reported to work as a transferase for lysine acetylation, propionylation, butyrylation, crotonylation and 2-hydroxyisobutyrylation, etc..[48-51] MOF can catalyze protein lysine acetylation, crotonylation and propionylation.[51, 52] HBO1 is able

to catalyze protein lysine acetylation and benzoylation.[53] Also, HAT1 can regulate protein lysine acetylation, isobutyrylation, and methacrylylation as a transferase.[43, 45] Importantly, mounting evidence have shown that different acylations on protein substrates result in different biological readouts. Wei et al. showed that crotonylation of HDAC1 can attenuate its enzymatic activities on histone substrates.[54] β-hydroxybutyrylation of p53 could regulate p53 acetylation level and reduce cell growth by decreasing p53 activity.[55] Lactylation of poly(ADP-ribose) polymerase 1 can influence its ADP-ribosylation activity and may potentially regulate DNA repair.[56] Besides various lysine acylations, cysteine modifications such as itaconation and S-succination have also been identified and it has demonstrated that the modifications greatly influence cellular inflammatory pathways by modifying some essential proteins such as serine/threo-nine-protein kinase 3 (RIPK3), and ATP-citrate synthase (ACLY).[57, 58] Therefore, the identification of protein substrates for different modifications would be extremely important for understanding their roles in cell regulation.

1.3 Functionalized Chemical Reporters for Labeling Protein Substrates

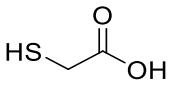
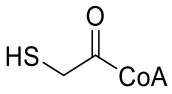
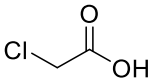
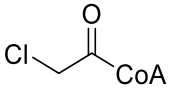
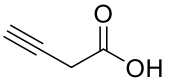
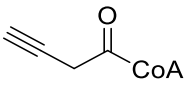
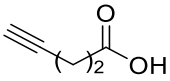
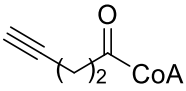
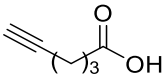
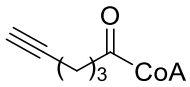
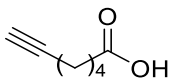
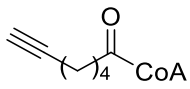
A technical challenge in the biochemical study of lysine acetylation lies in that the transferred acetyl group is chemically inert, which renders that direct chemical detection of KAT-medicated protein acetylation to be practically difficult. Standard methods of acetylation detection typically rely on radioisotope labeled acetyl-CoA or antibody recognition of acetylated lysine residues.[16, 38, 59, 60] Application of high-throughput mass spectrometry techniques for label-free detection and quantitation of lysine acetylation reaction products are also used to characterize KAT enzyme activities and inhibitors.[61] All the methods have satisfying sensitivity and reliability in acetylation detection, but they suffer from such issues as lacking substrate specificity, loss of tempo-spatial information, operational discontinuity, and high costs. In order to gain amenable

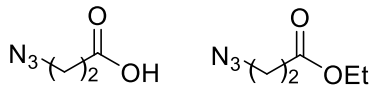
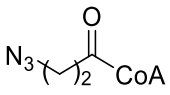
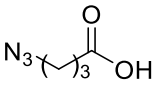
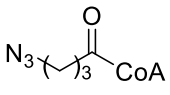
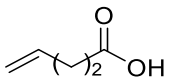
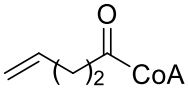
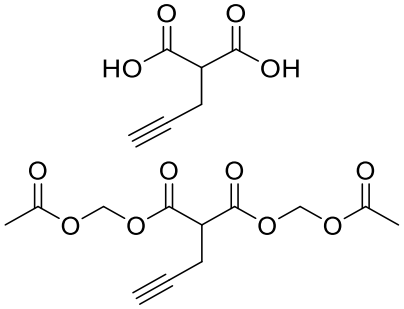
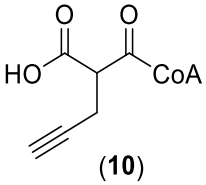
signals for detection and capture, a number of research efforts have been made to explore acetyl or fatty acyl derivatives that containing small-size bioorthogonal functional reporter groups to make acetyl-CoA and acyl-CoA surrogates that can be used by KATs for substrate labeling.[16, 59, 62-64] These bioorthogonal approaches provide a new dimension of technical accessibility to explore the landscape of lysine acetylation and acylation. The evident benefit is that the chemical reporter moiety on fatty acyl groups can be utilized for conjugation with fluorescent probes or affinity tags via a chemoselective chemical ligation to detect acylated proteins (**Figure 1.1**). For in vitro studies, functionalized short chain fatty acyl-CoAs are directly used to replace acetyl-CoA for substrate labeling. For cellular studies, functionalized short chain fatty acids (SCFAs) are typically used to circumvent cell permeability issue, which is based on the principle that inside cells SCFAs are converted by endogenous acyl-CoA synthetases into functionalized fatty acyl-CoAs which are used by respective KATs to label their acetylated substrates.

With 2-mercaptoacetate as an acetate surrogate, Sterner and Allfrey first reported a cell-permeable chemical reporter for monitoring post-synthetic acetylation to target acetylated proteins in duck erythrocytes.[65] They proposed that the incorporated acetate analog would be metabolically converted to mercaptoacetyl-CoA that subsequently underwent histone mercaptoacetylation (**Table 1.1**). The study indeed showed that the incorporated sulfhydryl groups can be utilized by endogenous KATs to be introduced into chromosomal proteins such as histones and HMG proteins. Furthermore, mercaptoacetylated proteins were selectively recovered from the biological mixture by mercury-affinity column chromatographic techniques and then subjected to MS analysis (**Figure 1.1B**). The shortage of this method is that it is limited for the labeling and isolation of KAT substrates lacking sulfhydryl amino acids. For most proteins the sulfhydryl group

present on the side chain of cysteine residues will interfere with and complicate labeling efficiency and specificity.

Table 1.1 Acetyl-CoA surrogates combined with wild-type or mutant KATs to label their substrates. With the treatment of acetate analog precursors, cells can generate corresponding acetyl-CoA surrogates which can be then recognized by respective KAT enzymes.

Acetate analogs	acetyl-CoA surrogates	KAT Enzymes
	 (**1**)	Unspecified KATs
	 (**2**)	RimL HAT1
	 (**3**)	Unspecified KATs
	 (**4**)	p300 GCN5-T612G MOF-I317A
	 (**5**)	p300 GCN5-L531A GCN5-T612G GCN5-T612G/F622A GCN5-T612G/L531A MOF-I317A MOF-H273A MOF-H273G
	 (**6**)	GCN5-T612G MOF-I317A MOF-V314G

	 (**7**)	p300 GCN5-L531A GCN5-T612G MOF-I317A MOF-V314G
	 (**8**)	MOF-I317A GCN5-T612G GCN5-F622A
	 (**9**)	Unspecified KATs
	 (**10**)	Unspecified KATs

The use of chloroacetyl-coenzyme A (ClAcCoA) (**Table 1.1**) to identify protein substrates for some members of the GNAT superfamily was reported by the Blanchard group.[66] One such GNAT member is the histone acetyltransferase 1 (HAT1), the first identified HAT enzyme.[67] Using ClAcCoA as an acetyl-CoA surrogate in junction with sulfydryl functionalized fluorophores, Blanchard et al. developed this orthogonal method to label yeast HAT1 substrates (**Figure 1.1C**).[66] Histone H4, known substrates of HAT1, can be chloroacetylated by HAT1 rapidly and selectively in vitro. Moreover, RimL, which is an acetyltransferase in *E. coli* also showed a good activity with the reporter. The chloroacetylated products are stable enough with a slow rate of hydrolyzation, whereas reactive enough to be conjugated with thiol nucleophiles such as TAMRA-cysteamine for fluorescent visualization. Potentially, thiol-containing reagents can be

designed with affinity handles for LC-MS/MS analysis to make the proteome-wide identification of substrates for the GNAT members. A unique property of ClAcCoA is that, during KAT catalysis, the chloroacetyl group transferred to the substrate can react with the other product, CoASH, to generate an CoA-acetyl-substrate conjugate. When [3'-^{32}P]-ClAcCoA is used, the product of the reaction becomes radiolabeled, and for protein substrates, can be observed after SDS-PAGE and autoradiography.[66]

Biomolecular labeling with azide or alkyne-functionalized chemical reporters and subsequent bioorthogonal ligation have gained prominent recognition in the chemical biology field.[64, 68] The incorporated alkyne or azide reporter group serves effectively as a chemical warhead to selectively react with azide/alkyne-containing probes via copper-catalyzed azide−alkyne cycloaddition (CuAAC) chemistry.[69-71] Owing to the much stronger chemoselectivity and bioorthogonality of CuAAC reaction compared with thiol-halide nucleophilic substitution reactions, utilization of clickable reporter groups for biomolecular labeling is a prime choice in bioorganic chemistry. Design of clickable acyl-CoA surrogates bearing an alkyne functional group (**Table 1.1**) to identify KAT substrates were first explored by Hang group to provide a chemoproteomic strategy to profile the acetylome of KAT members p300/CBP.[72, 73] p300 showed a good activity towards 4-pentynoyl CoA (4PY-CoA)) (**Table 1.1**) as the cofactor for substrate labeling which suggests that 4PY-CoA serves as a bioorthogonal chemical reporter to study p300 activity.[73] In kinetic characterization, 4PY-CoA exhibited a 10-fold lower k_{cat}/K_M value as compared with acetyl-CoA.[74] Consistent with the data reported by the Hang group, we found p300 have great cofactor promiscuity toward alkyne-functionalized acetyl-CoA analogs 4PY-CoA, 5-hexynoyl CoA (5HY-CoA), and 6-heptynoyl CoA (6HY-CoA) (**Figure 1.1D**).[75] Additionally, we found that the azido-containing analog 3-azidopropanoyl CoA (3AZ-CoA) (**Table 1.1**) is strongly

recognized by p300 HAT domain with comparable or stronger activity than Ac-CoA based on kinetic characterization. Also, there is around 8-fold higher specificity for 3AZ-CoA toward p300 than that of 4PY-CoA suggesting that 3AZ-CoA is another excellent acetyl-CoA surrogate to identify protein targets of the p300/CPB KATs (**Figure 1.1E**).[74] Importantly, the labeling activities of 3AZ-CoA for the other wild-type KATs such as MOF, Tip60, MOZ, MORF, GCN5, and PCAF were quite weak, which demonstrates the unique pairing property between 3AZ-CoA and p300.[74, 75]

Although several short chain clickable acyl-CoAs are active for wild-type p300/CBP to label their substrates, most KATs cannot efficiently take clickable acyl-CoAs for substrate labeling. The major obstacle is that the active pocket of most KATs is too small to accommodate the acyl head group of acyl-CoA. To overcome this problem, we performed experiments to engineer the active site of KATs GCN5 and MOF to expand their cofactor binding capability to accommodate the bulkier synthetic cofactors.[75] As such, the engineered KATs will be able to be used in junction with synthetic acetyl-CoA surrogates to establish bioorthogonal probes to investigate cellular substrates of these KAT enzymes. In our study, we tested five functionalized acetyl-CoA analogs containing alkynyl functional groups, including 4PY-CoA, 5HY-CoA, 6HY-CoA and two azide functionalized acyl-CoA molecules, 3AZ-CoA and 4-azidobutanoyl CoA (4AZ-CoA) (**Table 1.1**). The installed alkynyl and azido functional group is expected to facilitate downstream detection and characterization of labeled KAT substrates.

For GCN5 KAT, we analyzed its active site structure and identified several conserved bulky residues surround the acetyl warhead group of acetyl-CoA: L531, M534, I576, F578, T612, F622, and Y645.[75] To expand the cofactor binding pocket for it to accommodate bigger size acyl groups, we replaced each of these residues with smaller ones, i.e. alanine or glycine. To identify

the engineered enzyme forms that are active to the Ac-CoA substitutes, the entire panel of acetyl-CoA analogs was screened in histone modification reactions catalyzed by both the wild-type and engineered GCN5 proteins. As expected, wt-GCN5 exhibited a strong activity toward acetyl-CoA, the cognate cofactor of KATs, but was inert toward all the acetyl-CoA substitutes. On the other hand, several engineered GCN5, e.g. GCN5-L531A/G, -I576A, -T612A/G, and -F622A, exhibited appreciable activities to the synthetic analogs at varied degrees. In particular, the single mutant GCN5-T612G was active toward all the tested cofactors with 4AZ-CoA being the weakest.[74] Further, GCN5-T612G/F622A and -T612G/L531A mutants exhibited excellent activity toward 5HY-CoA.

We also engineered the active site of the MYST member MOF.[75] The crystal structure of MOF─Acetyl-CoA complex[76] shows that the acetyl moiety of the cofactor is surrounded by several bulky residues: V314, I317, I333, P349, P352 and L353. H273 is another potential residue that may affect the enzyme activity[77-79]. We carried out mutation of each residue to Ala or Gly to expand the enzyme active site for acyl-CoA binding. Activities of each MOF mutant toward the synthetic acetyl-CoA analogs were screened through enzymatic modification of the histone H4 tail peptide. Among the tested MOF mutants, H273A and H273G recognized 5HY-CoA, and V314G recognized 6HY-CoA and 3AZ-CoA. Strikingly, MOF-I317A was active toward all the acetyl-CoA substitutes. The double mutant MOF-I317A/H273A retained a reasonably good activity to 5HY-CoA. More details of the work can be found in reference [75].

Except alkyne/azide functionalized acyl-CoA as chemical reporters for probing protein acylation, Dekker and coworkers introduced sodium 4-pentenoate as a bioorthogonal chemical reporter to label endogenous acylated proteins followed by Heck coupling.[80] In their studies, incubation RAW 264.7 cells with sodium 4-pentenoate to form intracellular 4-pentenoyl-CoA

(**Table 1.1**) to establish 4-pentenoyl-labeled proteins. The labeled histones were then coupled with fluorescein or biotin-labeled phenylboronic acid via oxidative Heck reaction using EDTA-Pd^{II} as a catalyst (**Figure 1.1F**).[81] Fluorescence-based imaging showed that the 4-pentenoyl labeling of histones was dose-dependent and a strong fluorescent signal was seen for histone H3. Compared with alkynic reporters, the authors suggested that the lack of histone H4 labeling was due to the poor cell-permeability of the alkenic reporter. This study provides an alternative orthogonal reporter method for metabolic labeling of KAT substrates.

1.4 Chemoproteomics for Identification of Protein Substrates

Bioorthogonal acyl-CoA reporters provide unprecedented opportunities to establish a powerful chemical biology protocol to detect and study protein lysine acetylation in ways that cannot be accomplished by classic molecular biology methods. The added chemoselective functionality on the acyl group in substrates can be utilized to achieve various applications. First, fluorescent and other optical probes can be chemoselectively linked to the clickable reporter group to image acylated products and their subcellular location without recourse to the reliance on antiacetyl antibody or radioisotope. In principle, the labeling and imaging can be performed in live cells, which will provide spatially resolved images of KAT substrate distribution in different organelles. Second, the acylated substrates can be further labeled with affinity handles for their pulldown enrichment from complex biological mixtures. This methodology coupling with modern tandem MS allows for proteomic profiling of KAT substrates. Since the chemical modifying group is drastically different from the native acetyl group, it is technically more facile to identify acetylated substrates caused by overexpression or knockdown of a particular KAT of study. Third, the chemoreactive group in acetylated substrates can potentially be linked to any other biomolecules of interest that bear complementary clickable moieties to investigate KAT-mediated

protein-protein interactions and other biological processes. Last but not least, currently available high-throughput assays for KAT drug discovery are mostly antibody-based detection of the acetyl product or spectroscopic detection of the side product CoA.[16, 38] Bioorthogonal linking of acylated KAT substrates with fluorescent or bioluminescent probes could generate new protocols for KAT drug screening.[82, 83]

The methodology of using bioorthogonal acyl-CoA reporters for KAT substrate identification and profiling typically involves two steps: the first step is protein substrate labeling by acyl reporters which are enzymatically acted by respective KATs; then the labeled targets are detected with an imaging or affinity probe via a chemoselective bioorthogonal chemical conjugation.[84] In practice, the reporter labeling can be carried out in two different protocols which depend on the biological contexts to be studied. In one protocol, functionalized acyl-CoA and respective recombinant KATs are used directly to install acyl reporters onto proteins in purified forms or complex cell lysates. This is technically desirable because there is no requirement of plasma membrane penetration. The limitation, however, is that KATs are not present in their native forms and cell lysate proteomes do not retain their intact cellular milieu. Therefore, there is a higher chance of introducing artificial interference. The other protocol is through metabolic labeling in living cells. Acyl-CoA are widely believed to have no or little membrane permeability; therefore they cannot be directly applied to interrogate protein acetylation and acylation in cells. To circumvent the membrane passage issue, functionalized short chain fatty acids (SCFAs) are used as precursors of respective acyl-CoA and added in tissue culture media. A number of studies have shown that functionalized short chain fatty acids can cross plasma membrane to enter cells. Once inside the cells, these molecules are converted to respective acyl-CoA derivatives by the

cell's own biosynthetic machinery. Acyl-CoA are eventually used as cofactors by KATs to acylate their substrates.

The KAT-mediated protein labeling followed by site-selective chemical conjugation provides ample opportunities for detection, visualization, and further characterization of KAT targets. Most commonly, labeled proteins are either be visualized by coupling to fluorescent dyes such as fluorescein or tetramethylrhodamine (TAMRA)for in-gel fluorescence imaging, or isolated and enriched by specific affinity pull-down experiments (**Figure 1.2**).[85] Tandem MS proteomic profiling of the bioorthogonally labeled of KAT substrates is of great significance to provide insights into function of KATs. To facilitate affinity enrichment of labeled protein substrates for proteomic study, the Hang group developed a highly efficient $Na_2S_2O_4$-cleavable azobenzene–biotin tags (azido-azo–biotin) that allow pulldown of labeled proteins on streptavidin beads and meanwhile selective elution of affinity attached proteins for MS-based protein identification (**Figure 1.2**).[86]

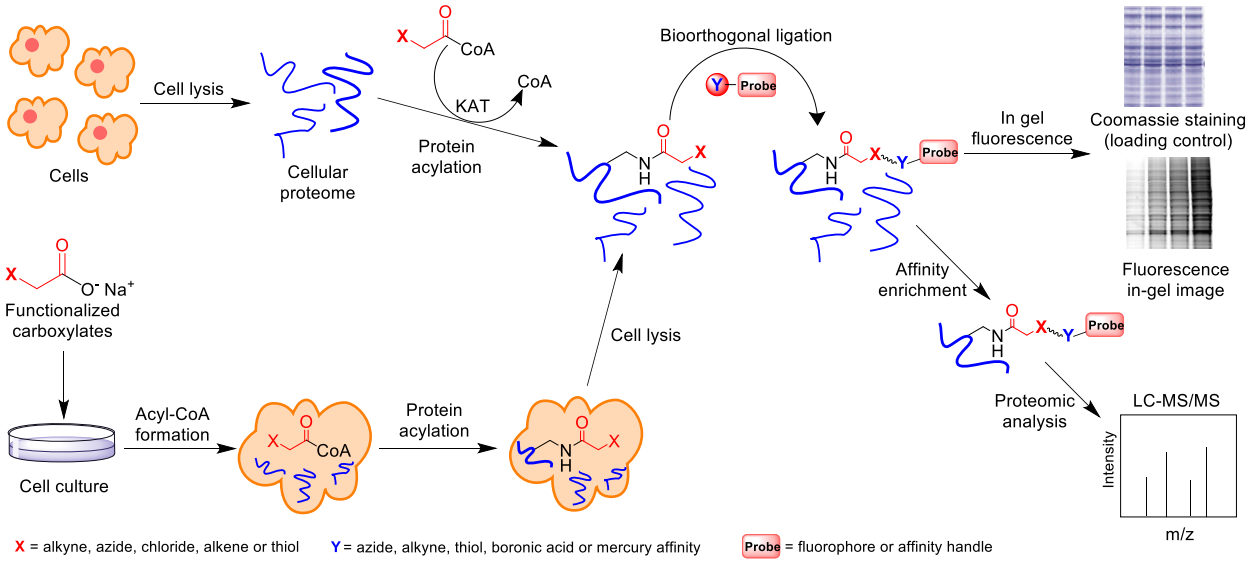


Figure 1.2 General procedures to profile KAT substrates using bioorthogonal reporters. The reporter labeling can be carried out in two different ways: one, acyl-CoA and respective

recombinant KATs are used directly to install acyl reporters onto proteins in purified forms or complex cell lysates. Two, functionalized short chain fatty acids are used as precursors of respective acyl-CoA to treat cells in tissue culture. Once entering the cells, these precursor molecules are converted to respective acyl-CoA derivatives by the cell's own biosynthetic machinery. Following reporter labeling, modified substrates are conjugated via bioorthogonal ligation to install either fluorophore probes for fluorescent imaging or affinity tags for proteomic enrichment and tandem MS profiling.

Several research groups have demonstrated substrate labeling with functionalized acyl-CoA, bioorthogonal click conjugation, and downstream protein detection and analysis. Hang et al. detected the feasibility of cell-permeable bioorthogonal chemical reporters to profile protein lysine acetylation.[87] Culturing Jurkat T cells with 3-butynoate, 4-pentynoate and 5-hexynoate followed by CuAAC-fluorescent detection, all of the three reporters showed the metabolic labeling was time- and dose-dependent. The application of sodium 4-pentynoate to Jurkat T cells identified 194 modified proteins as KATs substrate candidates through affinity enrichment by reacting with a cleavable azido-biotin tag followed by LC-MS/MS analysis.[87] Furthermore, the MS analysis also revealed the modification sites on protein by 4-pentynote moiety were the same as known acetylation sites on histones H2B, H3, and H4 in cells, which provides validating evidence that these short-chain alkyne-acyl carboxylates can monitor protein acetylation in living cells as an efficient chemical reporters (**Figure 1.2**). In addition, the same group applied 4PY-CoA together with purified KAT p300 to profile and identify p300 substrates from nuclear extracts of HeLa cells.[73] This chemical proteomics experiment identified several known protein substrates of p300, several new candidate p300 substrates, and the sites of modification. These results are an excellent

demonstration that bioorthogonal chemical proteomics allows the rapid substrate identification of individual protein acetyltransferases in vitro.

We applied 3AZ-CoA as a bioorthogonal surrogate of acetyl-CoA for understanding the substrate profile difference between KAT p300 and GCN5, two representative KATs in mammalian cells.[74] In this study, protein extracts of human embryonic kidney 293T cells were subjected to 3AZ-CoA with wild type p300 and GCN5-T612G mutant. The 3-azidopropionylated substrates were subsequently labeled with a sodium dithionite cleavable alkyne-azo-biotin through the CuAAC reaction. Following protein enrichment on streptavidin-coated resin, we conducted LC-MS/MS studies from which more than four hundred proteins were identified as GCN5 or p300 substrate candidates. The identified proteins are either p300- or GCN5-unique or shared by the two KATs, and are extensively involved in various biological events including gene expression, cell cycle, and cellular metabolism. We validated two novel substrates of GCN5 by co-transfection studies, i.e. IQGAP1 and SMC1. These results reveal extensive engagement of GCN5 and p300 in cellular pathways and provide insights in understanding their functional redundancy and distinction in biological pathway regulation.

To boost cell permeability of functionalized SCFAs, Meier and coworkers developed a set of ester modified SCFA pro-metabolites inspired by masking the polar groups in analogy of pro-drugs.[88] After the treatment of 3azidopropionyl-ester (**Table 1.1**) to the cell, LC-MS metabolomics analysis confirmed the formation of corresponding 3AZ-CoA. The modified proteins can be conjugated with fluorescent probes with CuAAC reaction visualized through the fluorescent gel-based assay. Again, the pro-metabolic labeling was also time- and dose-dependent. Interestingly, inhibition of p300 activity by a small molecule inhibitor decreased the labeling efficiency of some but not all proteins, which suggests that p300 work with other KATs to establish proteins acylation.

After conjugated to alkyne biotin via CuAAC and enriched by streptavidin-agarose, the labeled proteins were subjected to LC-MS/MS analysis, the results showed that 41% of the enriched proteins have been identified by Hang and coworkers, and some new acetylated marks were discovered.

Relatedly, Li and co-workers developed an alkyne functionalized malonyl ester reporter to study lysine malonylation.[89] This alkynyl diacetoxy-methyl pro-drug (MalAM-yne) (**Table 1.1**) can get into cells, release the corresponding functionalized malonic acid, and be converted into alkyl malonyl-CoA, which is subsequently utilized by hypothetic malonyltransferases (note that no *bona fide* malonyltransferases have been authenticated) to label their substrates (**Figure 1.1G**).[89]

1.5 Conclusions

In this chapter, we provided some important perspectives regarding the biochemical mechanisms of protein acetylation and acylations, the regulatory functions of KATs, the development of functionalized chemical reporters, and the progress of chemoproteomic methods for protein substrates identification. We discussed the important roles that protein acetylations and acylations play in cellular functional regulation. We also discussed the cofactor promiscuity of the different KATs which shows the complicated mechanisms that KATs possess to mediate protein activities. Additionally, we discussed the progress of the development of functionalized chemical reporters and the associated chemoproteomic methods, which enable the substrates labeling, imaging, and identification. These important research outcomes have laid a solid foundation for further investigation of the functionalities of different KATs, and identification of substrates for both KATs and PTMs. However, there are still some critical problems to be tackled.

Although we have mentioned that in our and other people's previous work, the clickable bioorthogonal chemical reporters have been designed and successfully applied to label and identify the protein substrates for individual KATs such as p300, GCN5 and MOF, the cell permeability issue of the chemical reporters limits their further utilization in detecting activities of KAT enzymes in intracellular contexts. Applying functionalized fatty carboxylates as precursors of acyl-CoA has proven to be a practical approach, but metabolic conversion of fatty acid to fatty acyl-CoA could be limited by the cellular level of acyl-CoA synthetases.[90] Therefore, more efforts need to be made to develop more efficient and robust cell permeable bioorthogonal chemical reporters. In this way, the labeling and identification of protein substrates for individual KATs can be performed intracellularly, which will provide valuable information for KATs' functions in native cellular context. Moreover, among different families of KATs, HAT1 is the first discovered but poorly studied enzyme. Dysregulation of HAT1 has been found to be related to lots of pathological processes. Thus, using the cell-permeable bioorthogonal chemical reporters to identify the protein substrates for HAT1 will be critical for understanding its biological functions.

Additionally, as aforementioned, lots of short chain acylations have been identified, but lack of information of the regulatory functions for each single PTM still hinders our understanding of its role in cells. To solve the problem, we believe the substrates specificity for the individual PTM is critical to decipher its function. Methacrylylation is a recent identified protein modification and is the isomer of crotonylation, which is already a well-studied protein modification. To date, the protein methacrylylation is only discovered on histone proteins. Questions like whether methacrylylation could also be expanded on other non-histone proteins like crotonylation and what is the functional roles of methacrylylation in cell regulation remain to be answered. Therefore,

designing functionalized chemical reporters for labeling and identification of cellular methacrylylated protein substrates would be significantly important.

The development of chemical reporters could not only aid in the identification of the protein substrates for the known protein modifications but help with the discovery of the new PTMs. Different protein modifications lead to different biological functions, so it is important to identify novel PTMs intracellularly. Since methacrylylation has been identified and it has known that HAT1 can catalyze protein methacrylylation by transferring methacrylyl group from Methacryl-CoA (MC-CoA) to its protein substrates.[43] Additionally, the methacryl group of MC-CoA possesses a double bond, leading to its strong electrophilicity. Based on these findings, we suspected that the methacryl group might conjugate with nucleophiles such as cystine intracellularly to form new modifications. Therefore, it is necessary to design functionalized chemical reporters to identify uncovered new modifications as well as their protein substrates.

1.6 Goal for the work

As reviewed in this chapter, PTMs are important for regulation of diverse cellular functions by modifying specific proteins. KATs are essential regulatory enzymes to mediate the corresponding PTMs. Unregulated PTMs or dysfunction of KATs are associated with multiple diseases. Although previous studies have elucidated many different types of acylations, corresponding KATs and protein substrates, the substrates specificity for individual PTMs or KATs haven't been thoroughly investigated, which greatly hinders our understanding for their functions in cell regulation. Therefore, in this work, we sought to design functionalized chemical reporters to identify the protein substrates for individual KATs and PTMs. Our goals are to develop cell-permeable chemical reporters coupled with engineered HAT1 to label and identify novel non-histone protein substrates for HAT1; design and synthesize chemical probes to capture cellular

methacrylylated proteins and HAT1 dependent methacrylylated proteins, which will then be identified through chemoproteomics; and generate chemical probes to uncover novel PTM which is potentially formed by MC-CoA and identify the corresponding protein substrates.

CHAPTER 2

CHEMOPROTEOMIC PROFILING OF PROTEIN SUBSTRATES OF A MAJOR LYSINE ACETYLTRANSFERASE IN THE NATIVE CELLULAR CONTEXT

Song, J.; Ngo, L.; Bell, K.; Zheng, Y. G. *ACS Chem Biol* 2022, *17* (5), 1092-1102

Abstract

The family of lysine acetyltransferases (KATs) regulates epigenetics and signaling pathways in eukaryotic cells. So far, knowledge of different KAT members contributing to the cellular acetylome is limited, which limits our understanding of biological functions of KATs in physiology and disease. Here, we found that a clickable acyl-CoA reporter, 3-azidopropanoyl CoA (3AZ-CoA), presented remarkable cell permeability and effectively acylated proteins in cells. We rationally engineered the major KAT member, histone acetyltransferase 1 (HAT1), to generate its mutant forms that displayed excellent bio-orthogonal activity for 3AZ-CoA in substrate labeling. We were able to apply the bio-orthogonal enzyme−cofactor pair combined with SILAC proteomics to achieve HAT1 substrate targeting, enrichment, and proteomic profiling in living cells. A total of 123 protein substrates of HAT1 were disclosed, underlining the multifactorial functions of this important enzyme than hitherto known. This study demonstrates the first example of utilizing bio-orthogonal reporters as a chemoproteomic strategy for substrate mapping of individual KAT isoforms in the native biological contexts.

2.1 Introduction

Amongst the diverse post-translational modifications (PTMs), lysine acetylation is versatile and has been linked to various aspects of cellular processes from chromatin remodeling, DNA repair, signal transduction, to cellular metabolism.[8] High resolution mass spectrometry-based proteomic analysis has disclosed thousands of acetylation targets and acetylation sites in all subcellular organelles, which include the nucleus, cytosol, mitochondrion, and endoplasmic reticulum, highlighting the widespread engagement of lysine acetylation in regulating cell physiology.[88, 91] Site-selective lysine acetylation in protein substrates is dominantly catalyzed by protein lysine acetyltransferases (KATs), which transfer the acetyl group from acetyl-coenzyme A

(acetyl-CoA, Ac-CoA) to the epsilon-amino group of lysine residues.[92] Lysine acetylation marks can be recognized by "reader" proteins, connecting this modification to downstream signaling networks.[93] A dozen of KAT members have been identified and classified into different major families including the GNAT family, the MYST family, and the p300/CBP family. HAT1, which belongs to the GNAT superfamily, was one of the first discovered KATs in mid-1990s.[67] Renewed attention to HAT1 was witnessed in recent years owing to its functional association with diverse oncologic processes.[94]

Biochemical and cellular studies show that HAT1 acetylates newly synthesized histone H4 at K5 and K12 sites,[95] as well as histone H2A at K5 site in the cytoplasm.[96] Take into account the thousand-sized acetylome and the extensive correlation of HAT1 with disease processes, it is proposed that the substrate profile of HAT1 could be much broader than merely nuclear histones, a scope beyond the chromatin regulation.[97] To fully illuminate the significance of lysine acetylation and HAT1 functions in regulating epigenetics, signaling cascades, and disease pathways, it is critical to identify the molecular targets of HAT1 on the proteomic scale in the cellular context.

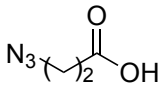
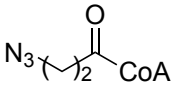
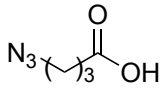
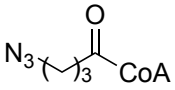
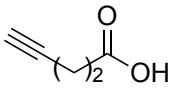
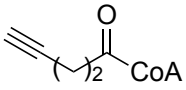
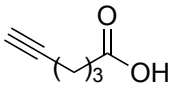
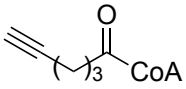
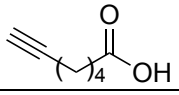
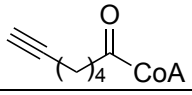
Earlier approaches to profiling individual KAT substrates typically rely on immunoprecipitation of acetylated proteins or peptides from cells pretreated with genetic knockdown or selective KAT inhibitors.[49, 98, 99] For instance, Garcia et al. recently identified 65 HAT1-dependent acetylation proteins with 84 sites by comparing immortalized $HAT1^{+/+}$ and $HAT1^{-/-}$ mouse embryonic fibroblast cell lines (MEFs).[99] These methods are effective but have some inherent drawbacks such as the functional redundancy of many HATs. One might expect that when a particular HAT member is knocked out but another redundant HAT acetylates its substrates, those substrates will be invisible in knockout experiments. Importantly, the PTM

antibody-based immunoprecipitation is often compromised by issues such as limited sequence diversity, promiscuous cross-reactivity, and batch- and source-caused disparities, which may lead to incomplete enrichment and/or high background signals.[100] Moreover, many KATs possess promiscuous acyltransferase activities, not merely acetylation.[101, 102] Besides lysine acetylation, HAT1 was recently found to transfer succinyl, propionyl, isobutyryl, and methacrylyl groups to its substrates.[43, 45, 103] It would be projected that incomplete substrates are detected when solely using an acetyllysine antibody to enrich protein substrates of KATs. As such, it is of great demand to develop new and alternative approaches for the identification and profiling of KAT substrates.

In recent years, functionalized acyl-CoA reporters, particularly those bearing small-sized and chemoselective warhead groups, have been developed and applied for the identification of lysine acylated substrates.[16, 92] Yu et al. used chloroacetyl-CoA as an orthogonal probe to label HAT1 substrates in vitro, but low efficiency and specificity as well as poor cell-permeability limit its further application on global elucidation of HAT1 substrates in cellular context.[66] Lyu et al.[104] recently used fluoroacetyl-CoA in combination with reactive thiol agents to label KAT substrates. We and others have made and applied acyl-CoA reporters for KAT substrates discovery that bear azido or alkynyl functional groups, such as 3-azidopropanoyl CoA (3AZ-CoA), 4-azidobutanoyl CoA (4AZ-CoA), 4-pentynoyl CoA (4PY-CoA), 5-hexynoyl CoA (5HY-CoA) and 6-heptynoyl CoA (6HY-CoA) (**Table 2.1**). Several of these acyl-CoA reporters were shown to be excellent Ac-CoA surrogates to label protein substrates of the KATs either in recombinant proteins or in the complex proteomes of whole cell lysates.[74, 75, 87] However, the lack of native cellular contexts poses a restriction on the accuracy and physiological relevance of substrate discovery. Taking advantage of the bump-and-hole approach,[105, 106] herein we attempted to create engineered HAT1 (eng-HAT1) forms that are capable to use bioorthogonal acyl-CoA reporters to label HAT1 substrates

with an azido or alkynyl-containing acyl group, and then the labeled substrates are conjugated with biotin tag for further optical detection or proteomic analysis **(Figure 2.1)**. In the process, we investigated cell permeability of a panel of acyl-CoA bioorthogonal reporters to identify membrane penetrating ones for intracellular use. Combined, we applied the eng-HAT1 with its paired cell-permeable bioorthogonal reporter to label and detect HAT1 substrates in the native cellular environment. We successfully identified hundreds of HAT1 substrates, most of which are previously unknown.

Table 2.1 Structures of the azide or alkyne-functionalized fatty acids and acyl-CoAs that were used for the cell permeability detection.

Structures of acetate analogs	Names of the acetate analogs	Structures of corresponding acyl-CoA analogs	Abbreviated names of the acyl-CoA analogs
	3-azidopropanoic acid (3AZ-acid)		3AZ-CoA
	4-azidobutyric acid (4AZ-acid)		4AZ-CoA
	4-pentynoic acid (4PY-acid)		4PY-CoA
	5-hexynoic acid (5HY-acid)		5HY-CoA
	6-heptynoic acid (6HY-acid)		6HY-CoA

2.2 Results and Discussion

Generating eng-HAT1 forms conferring new activities for bioorthogonal acyl-CoA reporters

Clickable acyl-CoA compounds containing azido or alkynyl groups (e.g. 3AZ-CoA, 4AZ-CoA, 4PY-CoA, 5HY-CoA, 6HY-CoA) have been recently designed as cofactors to label peptide

and protein substrates for different KATs.[74, 75, 87] In order to identify the potential protein substrates of HAT1 with bioorthogonal reporters, we wondered whether wild-type (wt) HAT1 can utilize these unnatural acyl-CoAs for substrate labeling. After analyzing the crystal structure of HAT1-

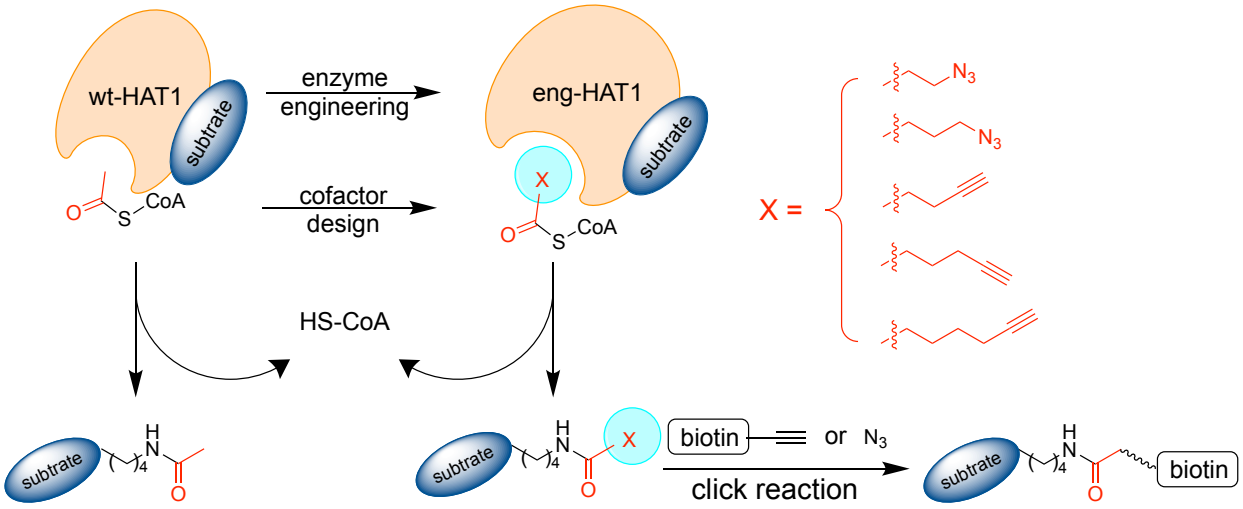


Figure 2.1 Schematic description of eng-HAT1 combined with acyl-CoA analogs to label HAT1 substrates. The Ac-CoA binding pocket of wild type HAT1 was rationally expanded by mutation of selected amino acid residues to accommodate bulkier acyl-CoA analogs. The transferrable azido or alkyne moiety can further be detected via click reaction-induced biotin probe.

Ac-CoA complex (**Figure 2.2A**, PDB: 2P0W),[107] we surmise that the active pocket is too narrow to accommodate the bulky acyl group of acyl-CoA. Therefore, we adopted a bump-and-hole approach to mutate the residues bordering the Ac-CoA binding site of HAT1 to expand the space surrounding the acetyl group (**Figure 2.1**), with the goal of obtaining some mutant forms that can accommodate the synthetic acyl-CoA reporters. We identified seven bulky amino acid residues in the Ac-CoA binding pocket of HAT1, i.e., M222, V238, M241, A275, P278, S279, and Y282. Each of these residues were mutated to a less sterically hindered amino acid, i.e., alanine or glycine

(**Figure 2.2A**). It is our hope that, as a result of the mutation, the expanded binding pocket will be able to tolerate the larger size of acyl-CoA analogs, leading to deposition of azido- or alkynyl- acyl moiety to the substrates of HAT1. Therefore, by using the pET28a-HAT1 (20-341) plasmid (Addgene, plasmid# 25239) as a template, we mutated each of the selected residues to alanine or glycine using the QuikChange site-directed mutagenesis protocol (**Supplementary Table S2.1**). By protein expression in *Escherichia coli* BL21(DE3) cells and followed by Ni-NTA affinity purification, we successfully obtained the recombinant wt-HAT1 and different eng-HAT1 proteins (SDS-PAGE of the expressed proteins were shown in **Supplementary Figure S2.1**).

(A) (B)

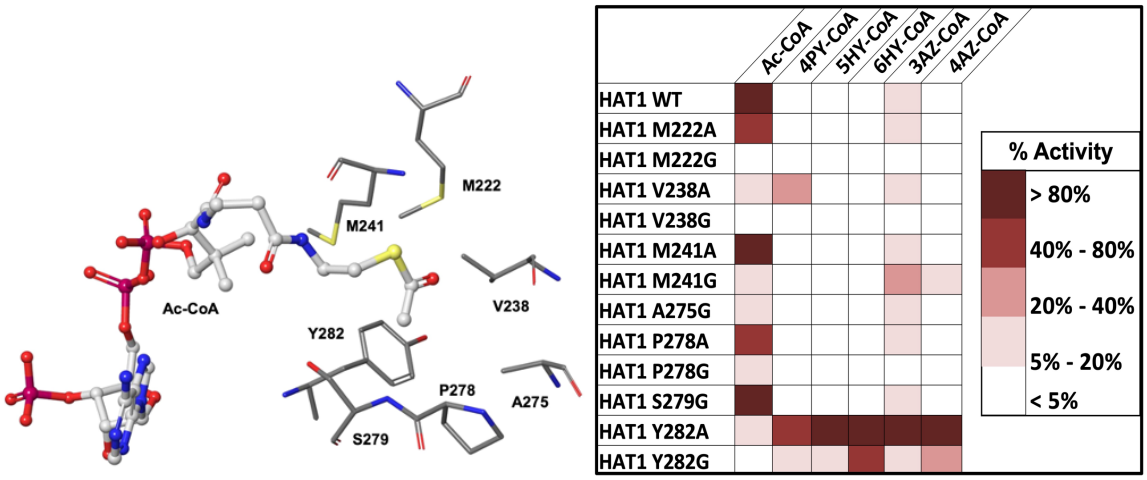


Figure 2.2 Investigation of wt or eng-HAT1 activities towards different acyl-CoA analogs. (**A**) Targeted hydrophobic bulky residues surrounding the Ac-CoA active site of HAT1-catalyzed-acetylation for HAT1 engineering. (PDB file: 2P0W). (**B**) Enzymatic activity of HAT1 and its various mutants towards Ac-CoA and its analogs.

Screening acyltransferase activities of wt- and eng-HAT1 towards acyl-CoA reporters

The acyl transfer activities of wt-HAT1 and the HAT1 mutants were measured by quantifying the side product CoA using the fluorogenic probe 7-diethylamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM).[108, 109] In a typical procedure, reactions containing 0.04 μM of the wt- or eng-HAT1 enzyme, incubated with 40 μM of the N-terminal 20 amino acid H4 peptide (H4-20), and individual acyl-CoA analogs (20 μM). The reactions were quenched with CPM (60 μM) in DMSO and then incubated in total darkness for 20 min at room temperature. Fluorescence was measured at an excitation and emission wavelength of 392 nm and 482 nm, respectively. The data were summarized in the heat map format (**Figure 2.2B and Supplementary Figure S2.2**).

As expected, wt-HAT1 showed desired acetyl transfer activity towards Ac-CoA but almost no detectable activities for other acyl-CoA cofactors. This testifies our premise that the active site of HAT1 is too narrow to accommodate the bulky size acyl groups. In contrast, different HAT1 mutants exhibited varied activities towards azide/alkyne acyl-CoAs. When P278 was mutated to Gly, there was only 10% activity for Ac-CoA and none for the other cofactors. However, when replaced with an Ala, the activity for Ac-CoA increased to 40% and there was 10% activity for 3AZ-CoA. This suggests that a residue with a hydrophobic side chain is needed at this position to aid in the stabilization of the acyl group binding.[107] The same phenomenon observed was seen with HAT1-Y282: the methyl side chain of Ala in HAT1-Y282A was observed to increase the acyltransferase activity greatly in comparison to HAT1-Y282G. HAT1-Y282G had 1% activity with Ac-CoA, less than 17% activity with 4PY-CoA, 5HY-CoA, as well as 3AZ-CoA, about 20% activity with 4AZ-CoA, and the highest activity at 50% with 6HY-CoA. In comparison, HAT1-Y282A had about 6% activity with Ac-CoA, 61% activity with 4PY-CoA, and greater than 80%

activity with most of the acyl-CoA reporters such as 5HY-CoA, 6HY-CoA, 3AZ-CoA, and 4AZ-CoA. Remarkably, HAT1-Y282A was almost inert to Ac-CoA compared to its high activity towards the synthetic acyl-CoA cofactors. Given this orthogonal feature, we chose HAT1-Y282A mutant for further substrate detection for HAT1.

To confirm HAT1-Y282A's activities towards the acyl-CoA reporters in histone H4 acylation, the mixtures of the enzymatic reaction of HAT1-Y282A, acyl-CoA analogs and H4 peptide (H4-(1-20)) were analyzed by MALDI-MS. As shown in **Supplementary Figure S2.3**, the H4-(1-20) was clearly acylated by the acyl-CoA analogs, which is consistent with the above CPM assay results. Next, we quantitatively measured the steady-state kinetic rates of HAT1-Y282A in H4 labeling with the acyl-CoA analogs at a series of concentrations by using the CPM assay. The kinetic constants k_{cat} and K_m were determined by fitting the acyl-CoA concentration—catalytic rate to the Michaelis-Menten equation. k_{cat}/K_m was calculated to evaluate the catalytic efficiency for each acyl-CoA analog **(Table 2.2, Supplementary Figure S2.4).** In particular, HAT1-Y282A exhibited extremely high activities towards 6HY-CoA and 3AZ-CoA with k_{cat}/K_m values of 1.40 and 1.08 $min^{-1}\mu M^{-1}$, respectively. Moderate activities were shown against 5HY-CoA, 4AZ-CoA and 4PY-CoA (k_{cat}/K_m: 0.84, 0.76 and 0.30 $min^{-1}\mu M^{-1}$), yet HAT1-Y282A was almost inert towards the natural cofactor Ac-CoA (k_{cat}/K_m: 0.03 $min^{-1}\mu M^{-1}$). Collectively, the results support that the engineered enzyme HAT1-Y282A had obtained excellent activities and orthogonal preference towards the synthetic azide/alkyne acyl-CoA reporters for HAT1 substrates labeling.

Table 2.2. Kinetic analysis of HAT1-Y282A activity against acyl-CoA analogs. Kinetics in the presence of varying concentrations of CoA analogs was obtained by fitting data points to the Michaelis-Menten equation. To determine the kinetics of HAT1-Y282A against various acyl-CoA analogs, reactions containing 40 nM of HAT1-Y282A, 200 μM H4-(1-20) peptide, and 0–200 μM of various acyl-CoA analogs were incubated for 15 min at 30°C.

Cofactor	k_{cat} (min^{-1})	K_m (μM)	k_{cat}/K_m (min^{-1}*μM^{-1})
6HY-CoA	33.2 ± 2.3	23.7 ± 5.2	1.40
3AZ-CoA	11.1 ± 0.3	10.3 ± 1.2	1.08
5HY-CoA	9.8 ± 0.8	11.7 ± 2.6	0.84
4AZ-CoA	13.15 ± 0.6	17.4 ± 2.6	0.76
4PY-CoA	25.4 ± 0.8	85.9 ± 5.9	0.30
Ac-CoA	NA	> 200	0.03*

In-gel imaging of labeled substrates by HAT1-Y282A and acyl-CoA reporters

We sought to further validate the activities of HAT1-Y282A towards the azide/alkyne acyl-CoA cofactors by examining the labeling efficiency on histone H4 using an imaging mode. Recombinant histone H4 protein was incubated with 6HY-CoA, 3AZ-CoA and 4AZ-CoA respectively in the presence of HAT1-Y282A. The mixture was then reacted with azido- or alkynyl-biotin by copper-catalyzed azide−alkyne cycloaddition (CuAAC) reaction followed by SDS-PAGE gel separation. The biotinylated H4 was visualized via streptavidin-HRP

chemiluminescent imaging. The biotinylated bands of H4 only showed up with the mixture containing HAT1-Y282A and acyl-CoA analogs, but barely any labeling was seen from the groups where HAT1-Y282A was absent (**Figure 2.3**). To determine if HAT1-Y282A acylates histone H4 in a site-dependent manner, we synthesized three peptides, e.g., H4 (1-20), H4 (1-22) and H4 (15-38) for labeling detection. As shown in **Supplementary Figure S2.5**, strong fluorescence was clearly seen within H4 (1-20) and H4 (1-22), but no detectable fluorescence was found for peptide H4 (15-38). This result was consistent with the fact that HAT1 acetylates H4 at K5 and K12 sites.[67, 95] These in-gel imaging data further support the bioorthogonal activities of HAT1-Y282A to the acyl-CoA reporters and demonstrate the feasibility and robustness to use HAT1-Y282A paired with acyl-CoA reporters to identify and image HAT1 cellular substrates.

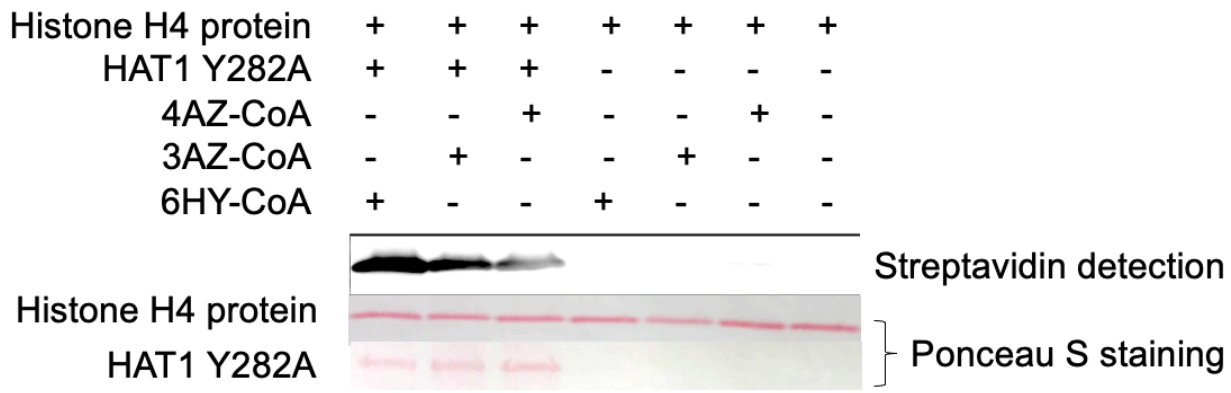


Figure 2.3 Streptavidin detection of HAT1-Y282A substrate histone H4. Recombinant histone H4 was incubated with HAT1-Y282A paired with 6HY-CoA, 3AZ-CoA or 4AZ-CoA respectively, followed by conjugation with alkyne-biotin with CuAAC reaction, and imaged by streptavidin-HRP.

Identification of 3AZ-CoA as a cell-permeable bioorthogonal reporter

Previously, some engineered KATs combined with bioorthogonal acyl-CoA reporters have been reported to identify KATs substrates from cell lysates.[74] Lack of native cellular environment,

however, potentially compromises substrate labeling efficiency and accuracy. As such, having validated the activity of the HAT1-Y282A mutant towards acyl-CoA reporters on the known HAT1 substrate H4, we attempted to investigate potential cell-permeable acyl-CoA reporters that can be applied to profile HAT1 substrates in living cells. Short chain fatty acids are known to enter the mammalian cells and metabolically be transformed to their corresponding acyl-CoA which are utilized by endogenous KATs to install the functional groups to their protein substrates.[87, 88] However, the membrane-penetrating efficiency of short chain fatty acids is typically quite low.[88] Interestingly, a recent study shows that extracellular CoA can adjust intracellular CoA levels.[110] We thus conducted a comparative study of a panel of azide or alkyne-containing short chain fatty acids and acyl-CoA reporters to evaluate their cell membrane permeability and cellular protein labeling sensitivity (**Table 2.1**).

Individual compounds were added to the culture medium of human embryonic kidney (HEK) 293T cells at 2.5 mM and incubated for 12 h. Cells were washed with PBS, harvested and lysed in M-PER buffer with sonication and subjected to CuAAC reaction to conjugate with alkyne-biotin or azide-biotin. The proteins were resolved by SDS-PAGE, and labeled proteins were visualized with streptavidin-HRP **(Figure 2.4A)**. To our surprise, among all the tested compounds, 3AZ-CoA turned out to show the most intense protein labeling, and 4AZ-CoA had moderate labeling, while other acyl-CoA analogs and the SCFA compounds exhibited very low levels of protein labeling (**Figure 2.4B**). The poor protein labeling by SCFA coincides with the previous study.[88] This result suggests the potential for 3AZ-CoA to serve as a cell-permeable bioorthogonal reporter for KAT substrate labeling. To validate cell permeability of 3AZ-CoA, we extracted acyl-CoAs from the cell lysate and used LC-MS/MS to detect the intracellular 3AZ-CoA. Indeed, the signal of intracellular 3AZ-CoA was clearly detected in the cell sample treated with 3AZ-CoA

(Supplementary Figure S2.6). By contrast, we cannot detect any MS/MS signal of 3AZ-CoA formation from 3AZ-Na treatment. This result confirms the excellent cell permeability of 3AZ-CoA to HEK293T cell.

To gain further insight into the cellular protein labeling by 3AZ-CoA, a dose-dependent treatment of HEK293T cell with 3AZ-CoA was conducted. We clearly observed that protein labeling grew stronger when 3AZ-CoA concertation increased from 0, 1, 2.5, to 5 mM (**Figure 2.4C**). This result again supports the excellent cell permeability of 3AZ-CoA as a bioorthogonal reporter for KATs. We also evaluated the cytotoxicity of 3AZ-CoA, and only a minor level of cell toxicity was seen at concentration of 2.5 mM (**Supplementary Figure S2.7**). Next, we sought to test the generality of 3AZ-CoA as a cell permeable reporter in different kinds of cell lines. HeLa cells, HCT 116 cells, and MEF cells were used to test the cellular protein labeling (**Figure 2.4D**). Interestingly, in all the tested cell lines, 2.5 mM 3AZ-CoA treatment induced much stronger protein labeling compared to either the negative controls (no treatment) or the 2.5 mM 3AZ-Na treatment. Clearly, the cell permeability of 3AZ-CoA is outstanding and versatile. Of note is that there appeared to be different patterns of protein labeling for each cell line, which suggests distinctive cellular acylations in different cells. Together, our data demonstrates the capability of 3AZ-CoA to be a cell-permeable reporter for KAT substrate labeling.

(A)

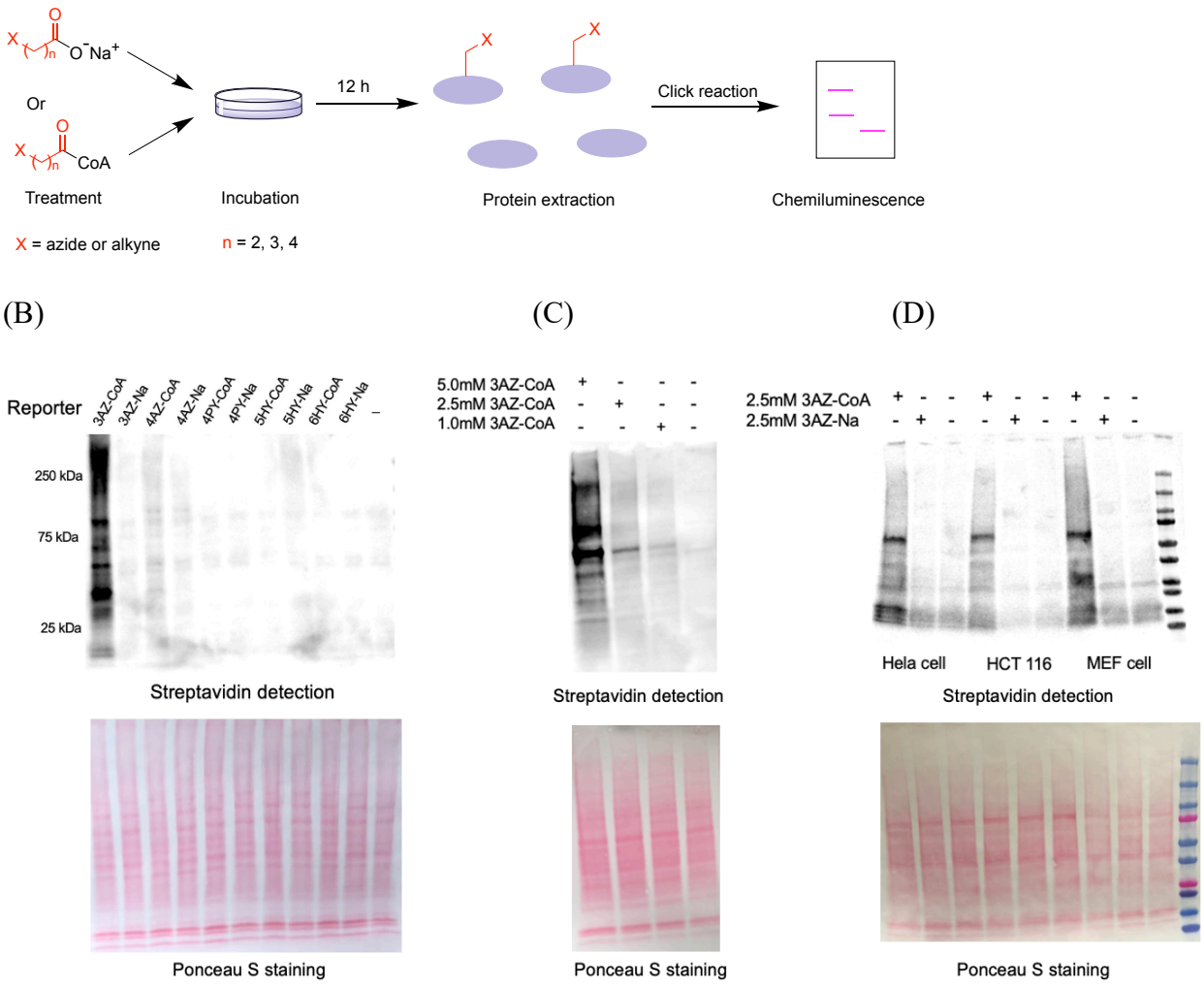


Figure 2.4 Evaluation of potential cell-permeable bioorthogonal reports for in vitro protein labeling. (**A**) Scheme for cellular analysis of cell permeability of bioorthogonal reporters. (**B**) Chemoproteomic cellular protein labeling by acetate analogs and acyl-CoA analogs in HEK293T cell. Cells were incubated with different analogs (2.5mM, 12 h), followed by CuAAC reaction. (**C**) Dose dependent labeling of HEK293T cell by 3AZ-CoA. Cells were incubated with 3AZ-CoA as indicated concentration for 12 h. (**D**) Various cell lines labeling by 3AZ-CoA and 3AZ-Na. Different cells were incubated with 2.5mM 3AZ-CoA or 3AZ-Na for 12 h.

Having successfully demonstrated 3AZ-CoA as a cell permeable acylation reporter, we then proceeded to test cellular imaging of protein labeling using the HeLa cell model. Herein, an

azide-sensitive turn-on fluorescence probe was chosen to induce live cell imaging, which was reported previously by us.[82] In this method, the copper-free click reaction would allow conjugation between the azide moiety on the protein substrates and the alkynyl-O-NBD probe, which subsequently triggers the intramolecular nucleophilic substitution by free amines of neighboring lysine residues to turn on the fluorescence of the N-NBD group. Briefly, after incubating HeLa cells with 2.5 mM 3AZ-CoA for 12 h, the cells were washed by PBS and then incubated with 50 μM alkynyl-O-NBD probe (PN-3) for another 2 h. The cells were rinsed to remove any excess reagents, and then subjected to fluorescence imaging on a Zeiss LSM 710 Confocal Microscope. As shown in **Figure 2.5**, a green fluorescence signal was detected with the treatment of 3AZ-CoA and the fluorescence probe PN-3. In contrast, incubation with 3AZ-CoA or PN-3 probe alone resulted in no detectable fluorescence. These exciting results were consistent with the in-gel streptavidin-HRP chemiluminescence imaging shown in **Figure 2.4B, 2.4C, and 2.4D**. These observations further corroborate that 3AZ-CoA can be applied as a cell-permeable bioorthogonal reporter to label intracellular proteins in living cells.

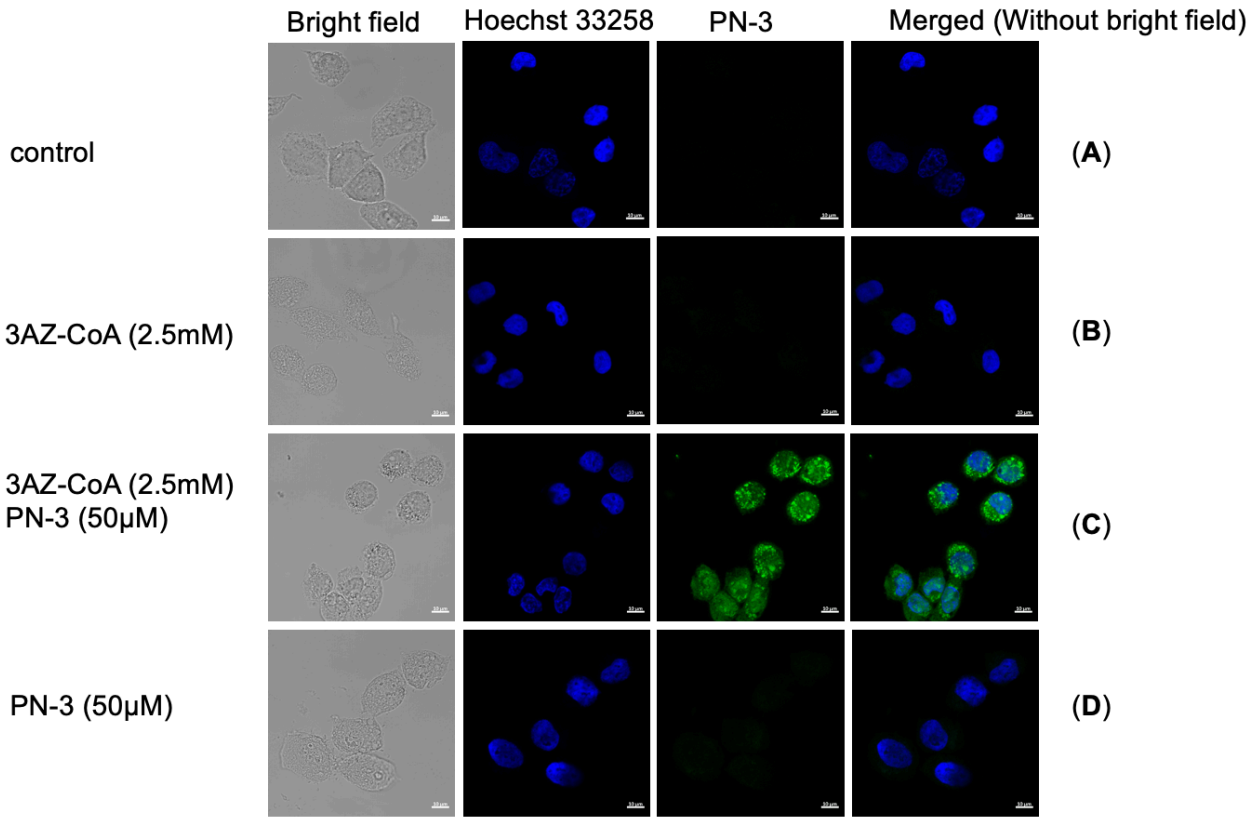


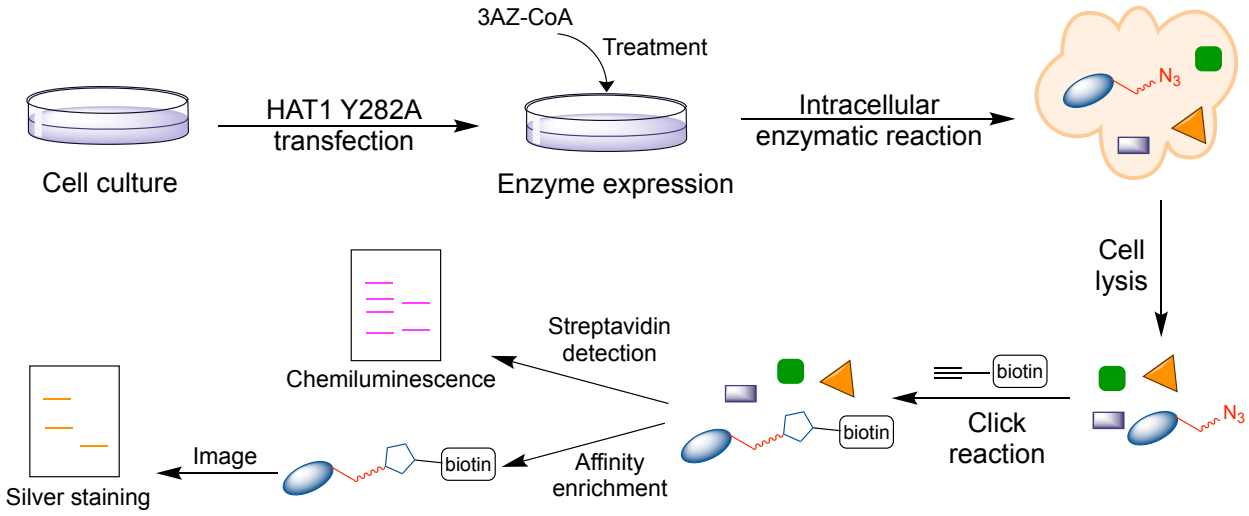
Figure 2.5 Confocal fluorescence images of HeLa cells with indicated concentration of 3AZ-CoA and PN-3. (**A**) Control group. (**B**) Cells were only treated with 2.5mM 3AZ-CoA for 12 h without PN-3. (**C**) Cells were pretreated with 2.5mM 3AZ-CoA for 12 h and subsequently treated with 50μM PN-3 for 2 h. (**D**) Cells were only treated with 50μM PN-3 for 2 h. Fluorescence images were obtained with a Zeiss LSM 710 Confocal Microscope. Scale bar: 10 μm.

Labeling of HAT1 substrates in living cells

Encouraged by the biochemical and cellular data that showed the bioorthogonal histone H4 labeling activity of HAT1-Y282A with 3AZ-CoA, and that 3AZ-CoA exhibited excellent cell membrane permeability, we then investigated if HAT1-Y282A matched with 3AZ-CoA could label HAT1 substrates under the native cellular environment. The working scheme is depicted in

Figure 2.6A: first, HEK293T cells were transiently transfected with full-length HAT1-Y282A plasmid in expression vector pReceiver-M11 (GeneCopoeia). After 36 h of incubation to allow the

(A)



(B) (C)

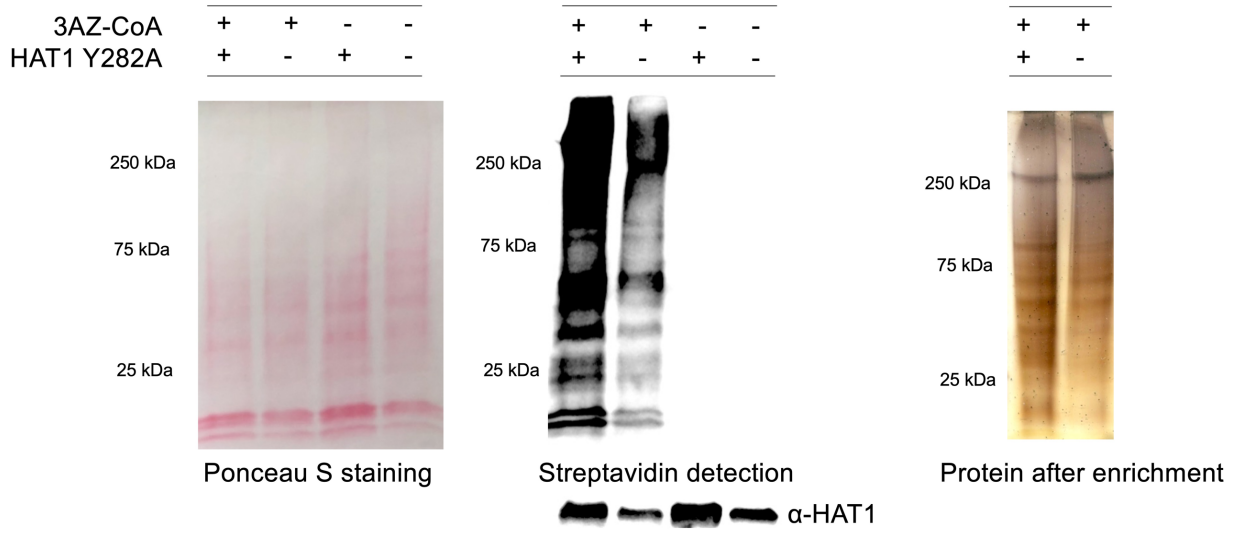


Figure 2.6 Label HAT1 substrates under native cellular environment and subsequent protein enrichment. (**A**) Schematic description of labeling intracellular HAT1 substrates by enzyme overexpression, bioorthogonal reporter treatment, click reaction and affinity enrichment. (**B**)

Streptavidin detection of HAT1 cellular substrates in HEK293T cells with HAT1-Y282A overexpression and cell-permeable bioorthogonal reporter 3AZ-CoA treatment followed by reaction with alkyne-biotin. (**C**) Enrichment of biotinylated protein by streptavidin beads followed by $Na_2S_2O_4$ cleavage. The protein eluents were resolved on SDS-PAGE and imaged by silver staining.

cells to express the mutant HAT1, the cells were treated with 2.5 mM 3AZ-CoA and grew for another 12 h. The cells were then lysed and the cell lysate was extracted to conjugate with alkyne-biotin by the CuAAC reaction. Next, the biotinylated proteins were visualized by streptavidin-HRP (**Figure 2.6B**). No labeling was seen from the control cells which had no reporter treatment. Importantly, compared with 3AZ-CoA treatment alone, multiple darker bands showed up when 3AZ-CoA treatment was combined with HAT1-Y282A overexpression. This indicates that the HAT1-Y282A catalyzed the enzymatic reaction to transfer the 3-azidopropanoyl group from 3AZ-CoA to its corresponding protein substrates under native cellular environment. This prompted us to enrich the labeled proteins from the cell lysate mixture to identify HAT1 substrates. For this purpose, a cleavable alkyne-diazo-biotin probe was used to conduct the CuAAC reaction, connecting labeled proteins to the biotin handle. After the pull-down experiment on streptavidin-coated beads, the enriched proteins were eluted with sodium dithionite, resolved on SDS-PAGE and imaged by silver staining (**Figure 2.6C**). As expected, much more proteins were enriched from cells with the HAT1-Y282A overexpression and 3AZ-CoA co-treatment, while 3AZ-CoA treatment alone led to less enriched proteins. This demonstrates that our strategy of engineering HAT1 paired with its cell-permeable bioorthogonal reporter is highly effective for substrate labeling and identification from native cellular environment.

Proteomic profiling of HAT1 substrates by SILAC-MS/MS and functional annotation

To discover HAT1 substrates in their native cellular environment, we adopted a quantitative chemoproteomic strategy by using stable isotope labeling of amino acids in cell culture (SILAC) in combination with high-resolution mass spectrometry. The workflow is illustrated in **Figure 2.7A**. HEK293T cells were cultured in "light" (medium containing natural arginine and lysine) and "heavy" (medium containing ^{13}C, ^{15}N-substituted arginine and lysine) medium separately. In the "light" medium, the cells were transfected with full-length HAT1-Y282A plasmid, whereas in the "heavy" medium, the cells were incubated with the empty-vector as a negative control. After 36 h of incubation, both cells were treated with 2.5 mM 3AZ-CoA and incubated for another 12 h. Next, the cells were lysed and the cell lysates were pooled together. The proteins were captured by aforementioned CuAAC reaction and streptavidin pulldown. After trypsin digestion of the enriched proteins, the collected peptides were analyzed by liquid chromatography with tandem mass spectrometry (LC-MS/MS). To ensure the reliability of the results, two biological replicate experiments were performed. Proteins were considered as HAT1 substrates following the ratio light/heavy > 2. Among the total identified 1391 protein targets (**Supplementary Table S2.2**), 123 proteins (**Supplementary Table S2.3**) were meeting the threshold requirement. Histone H4 was identified among the labeled proteins, which is consistent with our knowledge that HAT1 can acetylate histone H4 at lysine residues 5 and 12.[67, 95] Recent study has found that HAT1 is capable to acetylate histone H2A at lysine 5, and in our protein list of HAT1 substrates, the histone H2A was also highly enriched.[96] These identified proteins were also compared with a previously published work that revealed HAT1-dependent protein targets. Garcia et al. identified 65 proteins from embryonic fibroblast cells by HAT1 knockout and used antibody-based enrichment to capture the acetylated proteins that are dependent on HAT1.[99] Of

note, 12 of the 65 proteins were found exactly the same or belonging to the same family in our work including H4, H2A, H2B, TPR, PPP2R2A, HSD17B, HMG, HSP, DDX, ATP6V, hnRNP and NCL (**Figure 2.7B**). Importantly, we identified 111 new distinct protein substrates. This could be owing to differences in cell line type and higher sensitivity of our method.

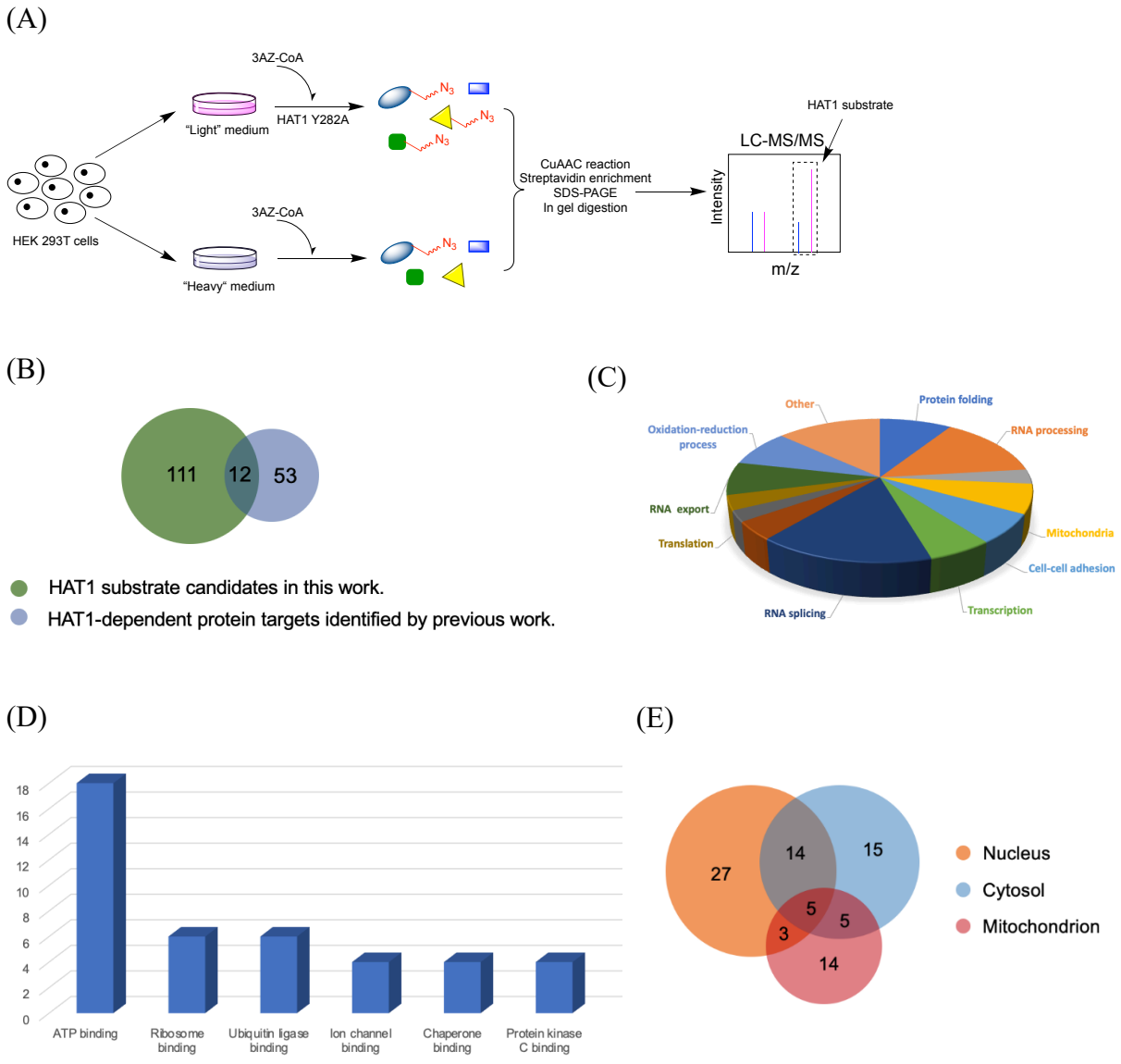


Figure 2.7 Identification and functional annotation of HAT1 substrates. (**A**) Schematic illustration of the SILAC proteomics study to identify HAT1 substrates. (**B**) Identified HAT1 substrates were compared with previous identified HAT1-dependent protein targets. (**C**)

Representative biological processes annotation for HAT1 substrates. (**D**) HAT1 substrates show different binding activities. (**E**) Venn diagram shows cellular compartment distribution of HAT1 substrates.

To confirm the HAT1 substrates identified in our proteomic results, we chose two undefined HAT1 substrates for further validation through immunoprecipitation and western blotting: Human high mobility group box 2 (HMGB2) and human serine/arginine-rich splicing factor 1 (SRSF1). Each plasmid (Myc-DDK-HMGB2, Myc-DDK-SRSF1) was transiently transfected to HEK293T cells and overexpressed with or without HAT1 overexpression. The DDK tagged proteins were immunoprecipitated on Protein G PLUS-agarose resin, and the acetylation levels of the proteins were detected with Western blot using pan anti-acetyllysine antibody. As shown in **Supplementary Figure S2.8,** the acetylation signals in both HMGB2 and SRSF1 proteins increased with the presence of HAT1 overexpression, which validates that HAT1 acetylates these two proteins in cells.

The broad-spectrum profiling of HAT1 substrates prompted us to examine its biological involvement and annotate the physiological functions. To this end, the 123 HAT1 substrates were subject to bioinformatics analysis using the DAVID web tool.[111] The gene ontology analysis of the HAT1 substrates shows that HAT1 is involved in a plethora of cellular pathways including transcription, translation, RNA splicing, RNA processing, protein folding, oxidation-reduction process, mitochondrial regulation, etc (**Figure 2.7C**). It is not surprising that canonical cellular functions of KATs such as translation initiation, RNA processing and transcriptional regulation were found in the functional annotation. For example, EIF5B and EIF2S1 play important roles in translation initiation[112], and acetylation of these proteins may alter their activities. Of note, RNA

splicing and RNA processing are among of the prominent cellular functions of identified HAT1 substrates.

Through interacting with their binding partners, proteins can carry out their specific functions. We thus analyzed the interaction network of the identified HAT1 substrates (**Figure 2.7D**). Consistent with the pathway analysis in **Figure 2.7C**, a large majority of the identified HAT1 substrates are binding with proteins, or RNA with 96 and 64 protein substrates, respectively. Interestingly, we found that 18 of the HAT1 substrates are binding with ATP such as ATP2A2 and DDX, suggesting that HAT1 might play a role in energy metabolism to drive intracellular biochemical reactions. Ubiquitin is an important small regulatory protein, which can trigger protein degradation, DNA repair, apoptosis, and signal transduction.[113] Alteration of protein ubiquitylation often leads to severe pathological conditions.[114] Importantly, we found that 6 of the HAT1 substrates including CUL1, HSPA5, HSPA9, HSPD1, SCAMP3 and UBE2O binding with ubiquitin protein ligase enzyme, suggesting the involvement of HAT1 in protein ubiquitination. We also found 4 HAT1 substrates binding with protein kinase C, e.g., C1QBP, DSP, GLRX3 and PRKCSH. RPKCSH is a known substrate for protein kinase C, which is associated with polycystic liver disease.[115] The finding of RPKCSH as a HAT1 substrate may provide another possible mechanism for regulatory of liver disease.

We also performed a cellular localization analysis of the HAT1 substrates (**Figure 2.7E**). We found that 49 of the HAT1 substrates are localized in the nucleus, 39 are in cytosol and 27 are distributed in mitochondrion with some of them overlapping. Therefore, functions of HAT1 clearly go beyond nuclear cellular processing. Overall, these functional annotations suggest that HAT1 substrates have a wide distribution for cellular regulation, which further demonstrates that HAT1

is not only involved in histone modification and chromatin regulation but also many other important cellular biological functions.

Coenzyme A and acyl-CoA are generally considered of lacking cell-permeability and eukaryotic cells obtain CoA from their extracellular precursors such as vitamin B5.[116] After the uptake of vitamin B5, the molecule will be metabolized and converted into CoA via an evolutionarily conserved five-step enzymatic reaction.[117] Nevertheless, Srinivasan et al. demonstrated that exogenous CoA could also serve as a source for mammalian cells or organisms to adjust intracellular CoA levels: one possible mechanism is that CoA could be hydrolyzed into a more stable molecule, 4'-phosphopantetheine extracellularly near the cell membrane, which then penetrates cell membrane by passive diffusion. Once entering the cell, 4'-phosphopantetheine is enzymatically converted to CoA by CoA synthases.[110] Later, the same group further demonstrated that acetylated 4'-phosphopantetheine is also stable and possesses cell permeability to serve as a source for intracellular Ac-CoA level.[118] These findings provide an important understanding for CoA compounds entering the cell and may suggest a possible pathway for 3AZ-CoA and 4AZ-CoA entering the cell, which subsequently label cellular protein substrates of HATs (**Supplementary Figure S2.9**). However, mechanism studies will be needed to investigate the details how the two acyl-CoA reporters get into cells. In principle, experiments would be required to determine whether these reporters can be hydrolyzed in the cell culture medium and enter the cells by passive diffusion or whether there is any specific transporter to facilitate their entry across the cell membrane. Compared to the alkyne acyl-CoA, 3AZ-CoA and 4AZ-CoA showed a better labeling efficiency (**Figure 2.4B**). The reason for this distinction in cellular protein labeling is still elusive at this stage and warrant mechanistic investigation.

We have identified more than one hundred protein substrates of HAT1 by combining bioorthogonal labeling and SILAC proteomics, with 12 being previously known and 111 first-time disclosed. We also validated two novel substrates of HAT1: HMGB2 and SRSF1 with transfection experiments. These findings greatly broaden the scope of the biological function of HAT1 which is far beyond histone modification and chromatin regulation. The separated binding sites in HAT1 for Ac-CoA and peptide substrates[107] ensure that a mutation in the Ac-CoA binding pocket of HAT1 has low possibility to change the binding specificity of protein substrates. From our previous work, some wild type HATs such as p300 can also utilize 3AZ-CoA to transfer the azido group to its protein substrates.[74, 75] Therefore, some of the labeled proteins appearing in the absence of HAT1-Y282A overexpression may be substrates of other HATs. Future research is needed to characterize detailed acetylation sites in the found HAT1 substrates and study their biochemical outcome in protein function regulation.

2.3 Conclusion

In summary, we have established a bioorthogonal proteomic methodology to quickly and reliably detect protein substrates of an individual KAT in living cells. We generated HAT1 mutants (e.g. HAT1-Y282A) through rational engineering that can accommodate clickable acyl-CoA reporters for substrate labeling. Meanwhile, we identified 3AZ-CoA to be a cell permeable bioorthogonal reporter for protein acylation. Piecing the KAT—acyl-CoA bioorthogonal labeling pair together with SILAC proteomics, we successfully profiled more than one hundred substrates of HAT1 in the native cellular environment. This finding that HAT1 targets numerous non-histone cellular protein substrates greatly broadens the scope of biological modalities regulated by HAT1 which is far beyond the chromatin realm. The bioorthogonal chemoproteomic strategy

demonstrated here would be complementary and advantageous for the detection and profiling of cellular protein substrates of other KAT members in the native biological contexts.

2.4 Methods and materials

Reagents

Unless otherwise noted, all chemical reagents were purchased from Sigma–Aldrich and used without further purification. Fmoc-protected amino acids and preloaded Wang resin were purchased from NovaBiochem. Recombinant Histone H4 protein was purchased from New England Biolabs. *E. coli* strains BL21-CodonPlus (DE3)-RIL were obtained from Stratagene. Protein dual color standards and precast protein gels (4%–20%) were purchased from Bio-Rad Laboratories. Thiol-reactive fluorescence dye 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin (CPM) were purchased from Invitrogen. Click chemistry reagents alkyne biotin (catalog# TA105), azide biotin (catalog# AZ104) and alkyne diazo biotin (catalog# 1042) were purchased from Click Chemistry Tools. High sensitivity streptavidin-HRP was purchased from Thermo Fisher Scientific (catalog# 21130). High-capacity streptavidin agarose was purchased from Thermo Fisher Scientific (catalog# 20359). HEK293T, Hela and HCT 116 cells were purchased from American Type Culture Collection (ATCC). MEF cell was a gift from Dr. Mark Parthun's lab. The anti-HAT1 antibody (catalog# 11432) was purchased from Proteintech. The anti-rabbit-HRP linked antibody (catalog# 7074) was purchased from Cell Signaling.

Mutation of HAT1 DNA and protein expression and purification.

The wild type pET28a-HAT1(20-341) plasmid was obtained from Addgene (plasmid# 25239). The wild type DNA plasmid was used as a template to mutate all the selected sites. HAT1 mutants were produced by first designing forward and reverse primers containing the desired single point mutation. The forward and reverse primers used to generate the single point mutants

are listed in Supplementary Table S1. Next, by using the QuikChange procedure (Stratagene), the target site-directed mutation was introduced into the HAT1 plasmid. Once all of the ingredients were added, the sample was preheated at 95°C for 5 min, then the polymerase chain reaction (PCR) method was followed: First, the temperature increased to 95°C for 1 min for the denaturation of the double stranded plasmid. Next, the annealing of the primers occurred when the temperature decreased to 55°C for 1 min. Finally, the extension occurred when the temperature was at 68°C for 16.5 min. The cycle repeated 17 times, and then the template DNA was digested with Dpn I restriction enzyme. Transformation was then carried out using the PCR product and *E. coli* XL1-Blue competent cells. The next day, colonies were selected and grown at 37°C overnight in LB media supplemented with kanamycin (0.125 mg/mL). Plasmids were purified using the Promega Wizard plus Miniprep system. DNA sequencing confirmed all intended mutations occurred as desired.

The expression and purification of HAT 1 (20-341) and HAT1 mutants were done following the method described by Hong Wu et al. [107] Briefly, the proteins were expressed in Escherichia coli and purified using the Ni-NTA resin. Transformation was done in *E. coli* BL21-CodonPlus (DE3)-RIL competent cells using the heat-shock method, and then the cells were spread on agar plates containing antibiotics kanamycin and chloramphenicol. Protein expression was induced by the addition of 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and shaken for 16 h at 16°C. The cells were collected and suspended in the lysis buffer (50 mM Na-phosphate (pH 7.4), 250 mM NaCl, 5 mM imidazole, 5% glycerol, 2 mM β-mercaptoethanol, and 1 mM phenylmethanesulfonyl fluoride (PMSF)) and then disrupted using the Microfluidics cell disruptor. The supernatant was passed through a column containing Ni-NTA resin equilibrated with column washing buffer (20 mM HEPES pH 8, 250 mM NaCl, 5% glycerol, 30 mM imidazole,

1 mM PMSF) and the resin was washed with column washing buffer. Next, the resin was washed with the buffer containing a higher concentration of imidazole (20 mM HEPES pH 8, 250 mM NaCl, 5% glycerol, 50 mM imidazole, 1 mM PMSF). Lastly, HAT 1 was eluted with elution buffer (20 mM HEPES pH 8, 250 mM NaCl, 5% glycerol, 500 mM imidazole, 1 mM PMSF). The eluent fractions were further checked by SDS-PAGE, and the enzyme concentrations were determined by Bradford assay. Enzymes were aliquoted and stored at -80°C for future use.

Fluorogenic CPM assay to screen HAT1 mutant activity.

The acylation activity of HAT1 and HAT1 mutants was analyzed using the CPM assay. To screen the various HAT1 mutants against the acyl-CoA analogs, reactions contained 0.04 μM enzyme, H4(1-20) peptide (40 μM), and acyl-CoA analogs (20 μM). Samples were incubated for 1h at 30°C. Evaluation of HAT1-Y282A kinetics was also done using the CPM assay with reactions carried out at 30°C for 15 min and sample conditions were as follows: 0.04 μM enzyme, H4(1-20) peptide (200 μM), and acyl-CoA analogs (0-200 μM). All reactions were quenched by the addition of excessive amount of 7-Diethylamino-3-(4'-Maleimidylphenyl)-4-Methylcoumarin (CPM) dissolved in DMSO then incubated in total darkness for 20 min at room temperature. For the kinetics assay, 203 μM (final) of CPM was used when acyl-CoA analogs concentrations varied from 0-200 μM. When the acyl-CoA analogs concentration was fixed at 20 μM, 30 μM (final) of CPM was used. Fluorescence was measured at an excitation and emission wavelength of 392 nm and 482 nm, respectively.

Normal and SILAC HEK293T cell culture.

For normal HEK293T cell culture, cells were cultured in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Life Technologies) and 1% penicillin-streptomycin (15140122, Thermo Fisher Scientific). For SILAC HEK293T cell culture, cells were

cultured in DMEM (arginine and lysine free, Life Technologies) supplemented with 10% dialyzed FBS (Life Technologies), 1% penicillin-streptomycin containing either 22 mg/L and 55 mg/L non-labeled arginine and lysine or 22 mg/L $^{13}C_6{}^{15}N_4$-arginine and 68 mg/L $^{13}C_6{}^{15}N_2$-lysine (Cambridge Isotope Laboratories). The SILAC cells were cultured for at least 8 passages before any experiments. All the cells were maintained at 37 ℃ with 5% CO_2.

Transient transfection of HAT1-Y282A and bioorthogonal reporter treatment for HAT1 substrates labeling.

Full-length HAT1-Y282A-encoding sequence was inserted into the M11 vector to generate HAT1-Y282A overexpression plasmid for mammalian cell. HEK293T cells were cultured to reach around 60% confluence stage, and transient transfection of the plasmid HAT1 Y282A was conducted by Lipofectamine 3000 (ThermoFisher SCIENTIFIC, Product #L3000008) according to the manufacture's protocol followed by 36 h of incubation. Next, cells were treated with 2.5mM 3AZ-CoA to induce the protein labeling. After another 12 h of incubation, the cells were washed by ice-cold PBS and harvested and lysed by ice-cold M-PER buffer containing 1% protease inhibitor cocktail. The protein was then collected and the protein concentration was determined as described above. The protein solution was used for further click reaction and streptavidin detection.

Protein labeling by CuAAC reaction and streptavidin detection.

30 μg cell lysate protein mixture was mixed with click cocktail containing 50 μM alkyne-biotin or diazo biotin alkyne, 2.5mM sodium ascorbate, 0.5 mM copper sulfate and 0.25 mM ligand BTTAA. After incubation at room temperature for 1 h, 1 x loading dye was added and boiled for 5 minutes. The samples were then resolved on 4-20% SDS-PAGE gradient gel. The separated proteins were transferred to nitrocellulose membrane and blocked with 5% non-fat milk for 1 h.

The membrane was incubated with streptavidin-HRP for another 1 h and scanned by chemiluminescence scanner.

Streptavidin affinity enrichment and sample preparation for mass spectrometry.

600 μg cell lysate was incubated with click cocktail as described above to conduct biotinylation of target protein. Excess diazo biotin alkyne was removed by spin dialysis. High-capacity streptavidin agarose was first equilibrated by PBS and then incubated with sample for 1 h with gentle agitation. Next, the resins were collected by centrifugation at 5000g for 5 minutes and washed with PBS supplemented with 0.2% (w/v) SDS, PBS supplemented with 0.1% (w/v) of SDS and 6M urea and 50 mM of NH_4HCO_3 supplemented with 0.1% of SDS. Then, the resins were incubated with 30 mM of $Na_2S_2O_4$ in 50 mM NH_4HCO_3 for 1 h to elute the protein off the beads. The mixture was then centrifuged at 5000g for 5 minutes and the supernatant was collected and dried by SpeedVac for further analysis. The dried protein sample was dissolved in 30 μL water followed by incubation with 20 mM DTT at 75 ℃ for 15 minutes and 200 mM of iodoacetamide at room temperature for 20 minutes.

Synthesis of acyl-CoA analogs and PN-3

3-azidopropionyl CoA (3AZ-CoA), 4-azidobutanoyl CoA (4AZ-CoA), 4-pentynoyl CoA (4PY-CoA), 5-hexynoyl CoA (5HY-CoA) and 6-heptynoyl CoA (6HY-CoA) were synthesized as previously described by CoA-SH reacting with carboxylic anhydride.[75] PN-3 was synthesized as described by previous work.[82]

Synthesis of N-terminal H4 peptide H4 (1-20), H4 (1-22) and H4 (15-38)

The standard solid peptide synthesis (SPPS) was used for synthesis of H4 peptides including H4 (1-20), H4 (1-22) and H4 (15-38), which was purified by C-18 RP-HPLC and confirmed by MALDI-MS. Specifically, The Fmoc-Lysine(Boc) preloaded Wang resin was used

at a 0.1 mmol scale, and 0.4 mmol of each amino acid was weighed out. Individual amino acids were coupled using activation reagent HCTU [O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate]. The peptide was cleaved off the resin using 95% trifluoroacetic acid (TFA), 2.5% H2O, and 2.5% triisopropylsilane (TIS) for 4 h. The sequence of the peptide H4 (1-20) is Ac-SGRGKGGKGLGKGGAKRHRK. The sequence of the peptide H4 (1-22) is Ac-SGRGKGGKGLGKGGAKRHRKVL. The sequence of the peptide H4 (15-38) is Ac-AKRHRKVLRDNIQGITKPAIRRLA. The peptide of H4 (1-20) was used for mass spectrum analysis of HAT1 and mutant HAT1 activity, peptides labeling as well as fluorogenic CPM assay. The peptides of H4 (1-22) and H4 (15-38) were used for peptide labeling assay.

Mass spectrometric analysis of modified Histone H4 peptide

The same conditions used to screen the acyl-CoA analogs were used to prepare the samples for MALDI. Briefly, 40 μM of H4(1-20) peptide was added to 20 μM of the acyl CoA analog and incubated for 5min at 30°C. Next, 40 nM of enzyme was added, and the reaction incubated again for 1 h at 30°C. The reaction was quenched by the addition of 10% TFA (final) (trifluoroacetic acid) and the samples were submitted for MALDI.

Histone H4 protein labeling by CuAAC reaction and streptavidin detection.

In the presence of 2 μ M HAT1-Y282A, 1.5 μg histone H4 protein was reacted with 20 μ M 6HY-CoA, 3AZ-CoA and 4AZ-CoA, respectively. Reactions in samples 1-3 contained 2 μM HAT1-Y282A with 1.5 μg histone H4 and 20 μM of 6HY-CoA, 3AZ-CoA, or 4AZ-CoA respectively. Samples in lane 4-6 contained the same components except HAT1-Y282A was not added. All reactions were carried out at 30°C for 1h before click cocktail was added to quench the reaction. For samples containing 6HY-CoA, the click cocktail used included: 50 μM azide-biotin, 2.5 mM sodium ascorbate, 0.25 mM BTTAA (2-(4-((bis((1-(tert-butyl)-1H-1,2,3-triazol-4-

yl)methyl)amino)methyl)-1H-1,2,3-triazol-1-yl)acetic acid), and 0.5 mM copper sulfate. When the cofactor 3AZ-CoA, or 4AZ-CoA was in the reaction the same click cocktail was used except 50 μM alkyne-biotin was used in place of the 50 μM azide-biotin. Post running the samples on an SDS-PAGE, the bands were transferred to a membrane and bands containing a biotin moiety were detected using Streptavidin-Horseradish Peroxidase (HRP) Conjugate.

Bioorthogonal reporter treatment and extraction of cellular protein for cell permeability test.

HEK293T cells were plated and cultured to adhere overnight. Acetate analogs were prepared as sodium salt and the pH of acyl-CoA analogs were adjusted to neutral followed by directly adding the stock solution to DMEM to achieve indicated concentration with gentle agitation and incubation for 12 h. Cells were then scrapped and harvested by washing with ice-cold PBS for 3 times. Cells were collected by centrifugation (5000 rpm x 5 min, 4 ℃) and were suspended with ice-cold M-PER mammalian protein extraction reagent (ThermoFisher SCIENTIFIC, Product #78501) containing 1% protease inhibitor cocktail (ThermoFisher SCIENTIFIC, Product #78438). After sitting on ice for 15 minutes, to release the whole cellular protein, cells were sonicated with 30% amplitude. The cellular proteins were collected by centrifugation for 20 minutes (13200 rpm x 20 min, 4 ℃). The protein concentration was determined by Bradford assay.

Cell toxicity assay

The cytotoxicity of 3AZ-CoA was evaluated by resazurin assay. HEK293T cells were incubated in 96-well plate at a density of 6000 cells per well. Cells were maintained in DMEM medium as described above. Next, cells were treated with dose dependent 3AZ-CoA, and incubated for another 12 h. The resazurin was then added in the medium at a concentration of 44

mM and cultured for 4 h. Cell viability was then determined by measuring fluorescence emission at 590 nm. Fluorescence intensities from the wells without cells were used as blanks (Fl_{blank}). Fluorescence intensities from the cells without treatment were used as 100% cell viability ($Fl_{control}$). All the measurements were performed for three times and the cell viability was determined as follows:

Cell viability (%) = 100 x $\frac{Fl_{sample} - Fl_{blank}}{Fl_{control} - Fl_{blank}}$

Western blotting analysis

The HAT1 Y282A level was detected by anti-HAT1 antibody. The protein mixture was subjected to SDS-PAGE and transferred to nitrocellulose membrane followed by blocking with 5% non-fat milk for 1 h. Next, the membrane was incubated with the primary anti-HAT1 antibody at 4 ℃ overnight. After washing with TBST (10 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween 20) for 3 times (5 minutes for each), the membrane was incubated with secondary antibody at room temperature for 1 h. The membrane was then washed three times (5 minutes for each) with TBST and developed by enhanced chemiluminescence detection system according to the manufacturer's protocols and scanned by chemiluminescence scanner.

Fluorescence imaging with confocal laser scanning microscopy.

Hela cells were cultured in DMEM medium supplemented with 10% fetal bovine serum and 1% penicillin-streptomycin and maintained at 37 ℃. The cells were treated with 2.5mM 3AZ-CoA and incubated for 12 h. Then the medium was removed and the cells were washed with PBS for three times. Next, fresh medium was added along with 50μM PN-3 (ex/em = 485/535 nm) and the cells were incubated for another 2 h. Hoechst 33258 (ex/em = 350/450 nm) was then added and the cells were incubated for 0.5 h. After washing the cells for three times with PBS, the cells

were fixed with 4% paraformaldehyde. The cells were then imaged with a Zeiss LSM 710 Confocal Microscope.

Immunoprecipitation and acetylation detection of HMGB2 and SRSF1

HEK293T cells were cultured to reach around 70% confluence stage, and transient co-transfection of the plasmid HAT1 (Genecopoeia, EX-I0105-M11) and Myc-DDK-HMGB2 (Origene, cat# RC200750) or HAT1 and Myc-DDK-SRSF1 (Origene, cat# RC201636) were conducted by Lipofectamine 3000 (ThermoFisher SCIENTIFIC, Product# L3000008) according to the manufacture's protocol followed by 48 h of incubation. The cells were then harvested and lysed by ice-cold M-PER buffer containing 1% protease inhibitor cocktail. The proteins were then collected and the protein concentration was determined as described above. The HAT1 expression level was detected using anti-HAT1 antibody. 100 μg proteins were subjected to immunoprecipitation for HMGB2 and SRSF1 enrichment by Protein G Plus-agarose (Santa Cruz Biotechnology, Product# sc-2002) following the standard protocol. The expression levels of HMGB2 and SRSF1 were detected by anti-DDK monoclonal antibody (Origene, cat# TA50011-100). The acetylation level of the enriched proteins was detected by pan anti-acetyllysine antibody (PTM BioLabs, PTM-105).

CoA extraction and mass spectrometry characterization of 3AZ-CoA formation in cell.

The cellular acyl-CoA was extracted using previously described method.[119] Briefly, the cell culture media was aspirated off after incubation with bioorthogonal reporters for 6 h. The dish was washed by ice-cold PBS, placed on dry ice and 1 mL of extraction buffer (80% methonal/water, v/v) was added and left in -80 ℃ for 15 minutes. Next, the cells were scraped and isolated by centrifugation (13200 rmp x 15 min, 4 ℃). The supernatant was dried down by Speed Vac and stored in -80 ℃ for further HPLC-MS analysis.

The solvent was taken linearly from 100% B (5mM NH_4OAc in 95% Acetonitrile and 5% water) to 10% B in 90% A (5 mM NH_4OAc in water) in 3 minutes and held at 10% B for 12 minutes, at which time 5 ul of the sample was injected onto the column. The gradient was linearly ramped to 100% B over 25 minutes and held at 100% B for 10 minutes. The flow rate was 50 ul/min. The column used is a Keystone Scientific, Inc (Now ThermoScientific) Deltabond Octyl. 1 x 250 mm, 5 um particle size and 300 A pore size. Mass analysis was performed using positive electrospray on Micromass (Waters) Q-TOF Micro. The capilary was set to 3,500 V and the cone voltage at 45 V. The instrument performed the survey scan from 200-900 m/z at 2 sec per scan. The scan range for MS/MS was 100-900 m/z at 2.5 sec. Per scan and a collision energy of 25 V. The mass resolution was 4800 on peak 865 m/z at half peak width.

Proteomic data processing and statistical analysis.

Quantitative analysis of the SILAC experiments was performed simultaneously to protein identification using Proteome Discoverer 2.5 software. Before mixing light and heavy samples, heavy samples were checked for Heavy labelling efficiency (>95% median efficiency). The precursor and fragment ion mass tolerances were set to 10 ppm, 0.02 Da, respectively. Enzyme was Trypsin with a maximum of 2 missed cleavages and Uniprot Human proteome FASTA file was used in SEQUEST searches. The following settings were used to search the data; dynamic modifications; Oxidation / +15.995Da (M), Deamidated / +0.984 Da (N, Q), 13C(6)15N(2) / +8.014 Da (K), 13C(6)15N(4) / +10.008 Da (R) and static modification Carbamidomethyl +57.021 (C). Scaffold Q+ (version Scaffold_5.0.0, Proteome Software Inc., Portland, OR) was used to quantitate peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 64.0% probability to achieve an FDR less than 1.0% by the Percolator posterior error probability calculation (Käll, L et al, Bioinformatics, 24(16): i42-i48, Aug 2008).

Protein identifications were accepted if they could be established at greater than 5.0% probability to achieve an FDR less than 1.0% and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii, Al et al Anal. Chem. 2003;75(17):4646-58). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

CHAPTER 3

CHEMOPROTEOMIC PROFILING OF PROTEIN LYSINE METHACRYLYLATION WITH A CHEMICAL PROBE

Song, J.; Yin, K.; Bae, A; Wu, R.; Zheng, Y. G. To be submitted to J. Am. Chem. Soc

Abstract

Lysine methacrylylation is a recent identified histone modification and is an isomer of crotonylation. Biological and physiological influence of crotonylation have been investigated in global cellular proteomes extensively, which shows its great potential as a therapeutic target. As a structural isomer of crotonylation, methacrylylation also possesses, even stronger, an electrophilic chemical group, which may bring its unique biological activities. However, as a new PTM mark, the substrates specificity in the cellular proteome are still unknown. The biological difference of methacrylylation versus crotonylation is also unclear. Therefore, to fill this knowledge gap, we developed a phosphine based selective chemical probe for quantitative chemoproteomic profiling of lysine methacrylylation. We first demonstrated the probe selectivity targeted to lysine methacrylylation but not crotonylation. We then applied our probe to various cellular models to show the broad involvement of lysine methacrylylation in diverse cellular proteins. Furthermore, we found histone acetyltransferase 1 (HAT1) could promote lysine methacrylylation in cellular proteins beyond histones. Since valine is the intracellular source of methacrylyl-CoA (MC-CoA) for protein methacrylylation, we compared the valine and sodium methacrylate induced protein methacrylylation and found the distinct cellular functional regulation for the two metabolites. Importantly, with quantitative proteomic analysis, we successfully disclosed 256 valine-dependent substrates and 254 sodium methacrylate dependent substrates and 1002 HAT1 dependent substrates for lysine methacrylylation in eukaryotes. Our selective chemical probe for lysine methacrylylation is therefore providing a versatile tool for investigation of this newly discovered PTM mark.

3.1 Introduction

Protein post translational modifications (PTMs) are emerging mechanisms for cell biological regulation. Owing to the advances of high-sensitivity mass spectrometry, hundreds of different PTMs have been identified, including phosphorylation, glycosylation, methylation and acylation.[120] Among of them, a group of short chain lysine acylations such as propionylation, butyrylation, isobutyrylation and crotonylation have been investigated and mounting evidence show that these modifications are functionally different from the widely distributed lysine acetylation.[45, 102, 121] These modifications were initially discovered on histones and dynamically modulated by histone acetyltransferases (HATs) and histone deacetylases (HDACs) to act as positive regulators for gene transcription.[45, 51, 122] Later, the study of these modifications was spread to non-histone proteins and a large number of acylated non-histone proteins were identified and characterized, which are associated with numerous biological events such as translational initiation, RNA processing, chromatin remodeling and mitochondria regulation.[54, 88, 123] Recently, lysine methacrylylation (Kmea) was reported as a histone modification and dynamically regulated by HAT1 and SIRT2.[43] To date, the modification was only found on histone proteins. However, considering the intracellular source of lysine methacrylylation, MC-CoA, which is from mitochondrial catabolism of valine, and that HAT1 plays an important role in regulation of mitochondrial proteins, we speculate that Kmea exist in cellular proteome beyond histone proteins.[99, 123, 124] Therefore, globally revealing the Kmea in cellular proteins will contribute to the elucidation of the important cellular functions of this under-investigated PTM.

As a structure isomer of Kmea, lysine crotonylation (Kcr) was discovered earlier on histone proteins as well as thousands of other cellular proteins and has been well-characterized in previous studies.[122, 125] Distinct from other short chain acylations, Kcr has drawn a large attention due to its

unique property of α, β unsaturated acyl group contributing to its planar orientation. The C=C bond of crotonyl group on histone has been demonstrated to mediate binding with the "reader" proteins such as YEATS domain containing proteins through a π-π-π stacking, which enhances its binding affinity.[126, 127] Emerging evidence suggest that cellular crotonylome participate in multiple cellular functions. For example, crotonylation of HDAC1 is able to decrease its deacetylase activity on histone substrates.[54] The interaction between replicative protein A1 (RPA1) and ssDNA can be promoted through Kcr of RPA1, which can regulate DNA repair.[128] Similar to Kcr, Kmea was also found on all histone proteins including H1, H2A, H2B, H3 and H4 with some sites overlapping for these two modifications.[43] Additionally, methacrylate group also possesses a double bond, leading to its electrophilicity to conjugate with suitable nucleophiles. Although the Kmea has been described on histones in current stage, the question of whether methacrylylation, like crotonylation, can also be expanded on non-histone proteins and what are the functional differences of methacrylylation versus crotonylation in biological regulation remains a mystery.

Traditionally, antibody-based immunoprecipitation was applied to study the acylated cellular proteins, but the issues such as promiscuous cross-reactivity and batch-caused disparities are still big challenges for protein substrates enrichment and identification.[129] For example, Jiang et al. showed that anti lysine benzoylation (K_{bz}) antibody could also recognize lysine picolinylation, lysine nicotinylation, and lysine isonicotinylation.[130] Our group demonstrated that anti lysine butyrylation antibody is also not specific and could recognize its isomer, lysine isobutyrylation.[45] Additionally, previous work showed that with the treatment of crotonate in Hela cell, the signal of histone protein could be increased with the detection of anti-methacrylylation antibody[43] and our data showed that with the treatment of methacrylate, the signal of histone protein could also be promoted with detection by anti-crotonylation antibody (**Figure 3.2B**). These

data showed that the antibody-based immunoprecipitation has a great specificity issue, especially for the protein modification with the similar structure, which would make them hard to differentiate. Hence, there is a great demand to develop new strategies to specifically profile the Kmea.

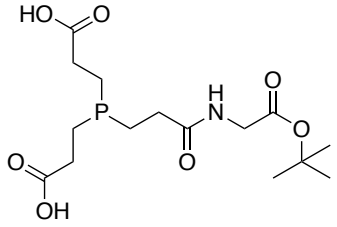
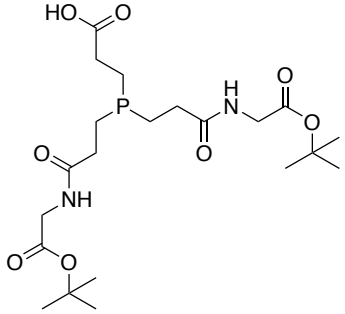
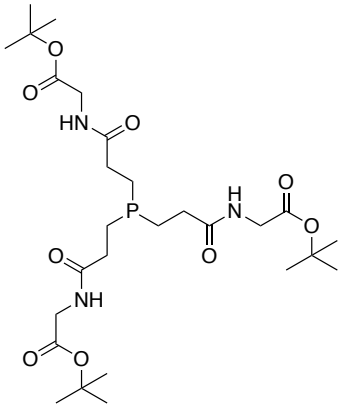
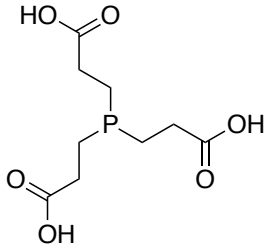
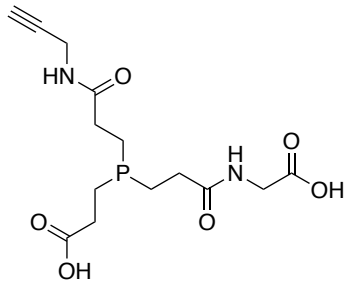
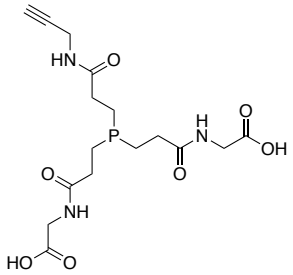
Taking advantage of the bioorthogonal chemistry[92, 131] and the different electrophilicity of the crotonylate and methacrylate group, we sought to develop a chemical probe to specifically capture cellular methacrylylated proteins but not crotonylated proteins. Recently, a phosphine-mediated phospha-Michael addition was introduced to occur between tris(2-carboxyethyl)phosphine (TCEP) and unsaturated modification on protein residues.[132-134] Here, we describe a series of TCEP analogs, especially one named gly-TCEP-AK, that enables selectively labeling and enrichment of cellular methacrylated proteins in a mild condition (**Table 3.1**). Previous study has demonstrated sodium methacrylate as a precursor to boost cellular MC-CoA level, which could induce histone lysine methacrylylation.[43] Additionally, valine is the intracellular source for MC-CoA formation. Using the probe combined with quantitative proteomic analysis, we successfully identified sodium methacrylate and valine dependent methacrylylated proteins, respectively. Furthermore, we applied the probe to profile and identify HAT1 regulated protein lysine methacrylylation. By comparing the methacrylylated proteins with the previous identified crotonylated proteins, we also found the common and unique cellular functions for protein lysine methacrylylation.

3.2 Results and Discussion

Development of a selective chemical probe for lysine methacrylylation

Phosphorus compounds mediated phospha-Michael addition has been introduced for detection of α, β-unsaturated amides on cellular proteins due to its excellent nucleophilicity, water solubility and biocompatibility. For instance, previous studies have reported that acrylated, crotonylated and dehydrobutyrine modified protein could be labeled by the probe TCEP-biotin for visualization.[132-134] However, none of them have applied this TCEP based chemical probe for intact protein identification at proteomic level, probably due to the steric hindrance, non-selectivity and

Table 3.1 Names and structures of TCEP analogs.

(**1**)	(**2**)	(**3**)
t-butyl-Gly-TCEP	2t-butyl-Gly-TCEP	3t-butyl-Gly-TCEP
		
(**4**)	(**5**)	(**6**)
TCEP	Gly-TCEP-AK	2Gly-TCEP-AK
		

bulky adducts (biotin moiety) introduced on TCEP as well as the non-cleavable property. Therefore, to overcome the steric hindrance issue and further apply the chemical probe to quantitatively identify cellular methcarylylated protein by LC-MS/MS, we developed a series of TCEP analogs (**Table 3.1**, **1-6**) to investigate the reactivity with crotonylated and methacrylylated protein aiming to gain selectivity towards cellular methacrylylation but not crotonylation. To this end, we first utilized sodium methacrylate as a phospha-Michael acceptor to react with the TCEP analogs (**1-3**); the reaction between sodium methacrylate and TCEP analogs were carried out in aqueous Tris buffer at pH 8.0 and 37 ℃ and monitored by RP-HPLC and ESI-MS. The results showed that 71.3% of **1** was converted to the product and the rest was oxidized **(Supplementary Figure S3.1A and D)**. Similar results were observed for the TCEP analogs **2** and **3** with 72.2% and 74.1% conversion, respectively (**Supplementary Figure S3.1B**, **C and D**). The data showed that limited influence of the tert-butyl glycine conjugation for TCEP reactivity towards the Michael acceptor.

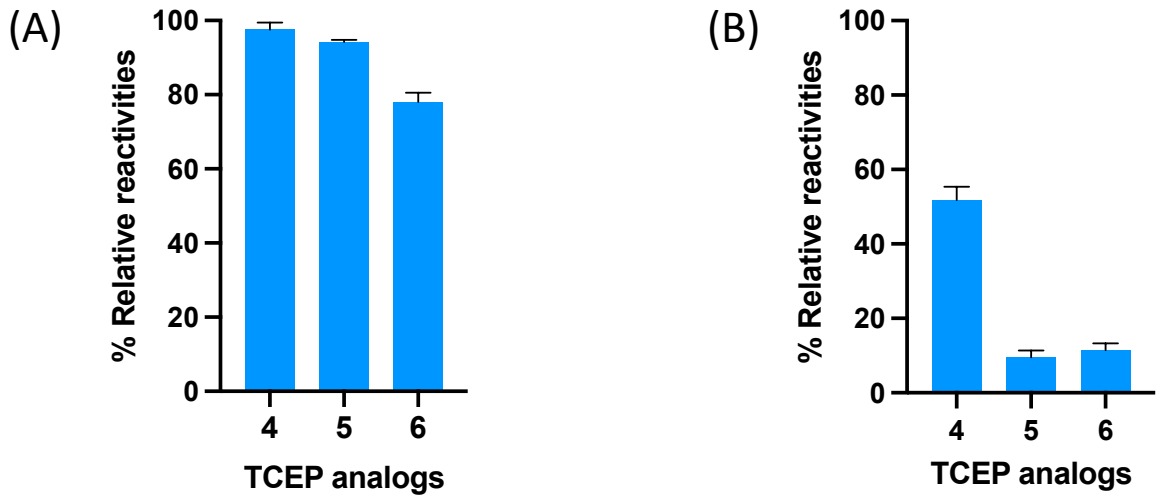


Figure 3.1 Study the reactivity between TCEP analogs (4-6) and methacrylylated peptide h4(2-8)k5metha or crotonylated peptide h4(1-20)k5cro. (A) The column graph shows the percentage conversion of the TCEP analogs **4-6** after the reactions between TCEP analogs and

methacrylylated peptide. (B) The column graph shows the percentage conversion of the TCEP analogs **4-6** after the reactions between TCEP analogs and crotonylated peptide.

Next, to develop a selective chemical probe for labeling the methacrylylated protein with minimum steric hindrance influence, we decided to introduce an alkyne functional group instead of the previous used biotin group. Although the previous study has shown that the alkyne could react with the phosphorus on TCEP under the desulfurization condition[135], we found that with the coupling reagents EDC and HOAt, the carboxylate group on TCEP could immediately conjugate with propargylamine by forming a peptide bond achieving around 43% yield within 30 mins. The available alkyne could then undergo click reaction for further imaging or enrichment for proteomic identification. Additionally, previous study has demonstrated that the negatively charged carboxylate groups on TCEP are essential for its nucleophilicity, which is probably due to the stabilization of the phosphonium cation from the product.[133] Therefore, we decided to further modify the carboxylate group aiming to gain more selectivity (**Table 3.1**, **4-6**). We also synthesized a model compound Boc-cro-lysine for the activity exploration. With the aforementioned reaction condition in aqueous Tris buffer, the TCEP and TCEP analogs were reacted with excess model compound and the reactions were monitored by HPLC and ESI-MS. As shown in **Supplementary Figure S3.2**, based on the ratio of the relative peak area of products and the left Boc-cro-lysine after the reaction, the TCEP analogs **5** and **6** showed less reactivity towards the model compound compared with TCEP (**4**). Furthermore, we tested the reactivity of the probes with methacrylylated and crotonylated peptides. Through the MALDI-TOF mass analysis and quantification of the mass peak intensity, we found TCEP treatment to the methacrylylated peptide could lead to 98.9% conversion to the product, and the probes **5** and **6** could achieve 94.6% and

79.8% conversion, respectively (**Figure 3.1 A and Supplementary Figure S3.3**). However, with the treatment to the crotonylated peptide, we found a dramatically decreased conversion especially for the probe **5** and **6**, with 10.8% and 12.8%, respectively (**Figure 3.1B and Supplementary Figure S3.4**). These data demonstrated the different reactivity for the methacrylate and crotonylate group with the TCEP analogs and the probe **5** showed a high reactivity towards methacrylylation but minimum reactivity for crotonylation. Therefore, we decided to use probe **5** for further profiling and identification of the cellular protein methacrylylation.

Evaluation of gly-TCEP-AK for cellular methacrylylated protein labeling

To test whether gly-TCEP-AK could be applied to the complex cellular proteome for labeling of the methacrylylated protein and also investigate the probe selectivity, we treated HEK293T cells with 20 mM sodium methacrylate and sodium crotonylate respectively, which as noted up-regulates cellular MC-CoA and crotonyl-CoA levels in previous study[43, 136]. After 24 hours' incubation, we first extracted histone proteins which were then incubated with 4 mM gly-TCEP-AK in Tris buffer at pH 8.0 at 37 ℃ for 16 hours. The proteins were then dialyzed to remove excess probe. Thereafter, the histone proteins were further conjugated with azide-biotin (Click Chemistry Tools) for visualization through copper catalyzed azide alkyne cycloaddition (CuAAC) followed by SDS-PAGE gel separation and streptavidin-HRP detection (**Figure 3.2A**). The cells treated with 20 mM sodium methacrylate showed increased signal (**Figure 3.2B**), especially for

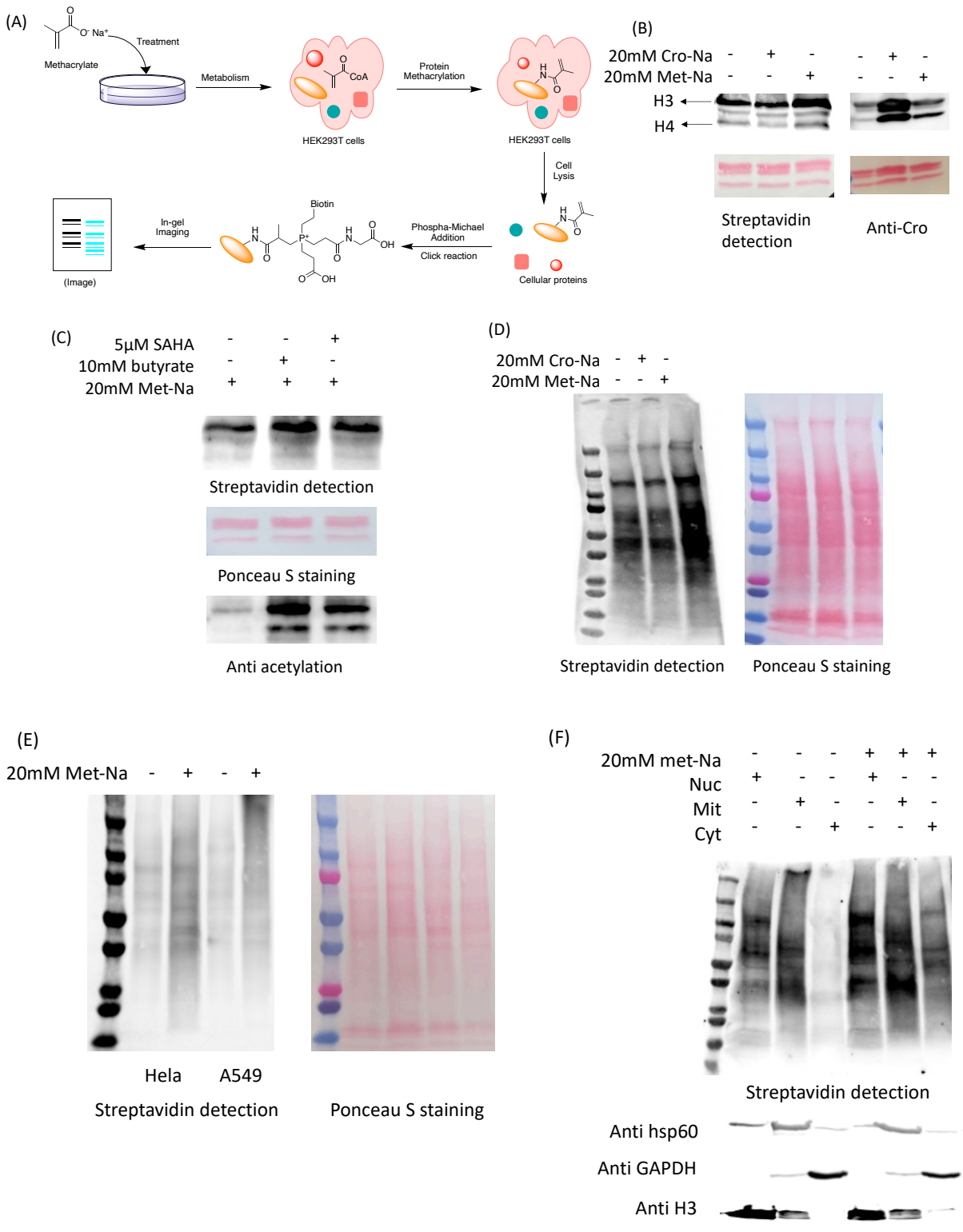


Figure 3.2 Labeling and imaging of the cellular methacrylylated proteins by the probe Gly-TCEP-AK. (A) Workflow for labeling and imaging of cellular methacrylylated proteins with

sodium methacrylate treatment and bioorthogonal labeling. (B) Detection of methacrylylated histones in HEK293T cells. HEK293T cells were treated with 20 mM sodium methacrylate or sodium crotonylate, and the histones were extracted, followed by incubation with 4 mM Gly-TCEP-AK for 16 hrs and click reaction. (C) Evaluation of histone methacrylylation with inhibitor treatment. The cells were treated with HDAC inhibitors and also sodium methacrylylate, and then the histone proteins were labeled by the probe. (D) Examination of methacrylylation on cellular non-histone proteins. HEK293T cells were incubated with 20 mM sodium methacrylylate and 20 mM sodium crotonylate for 24 hrs and the whole cellular proteins were extracted, labeled and imaged. (E) Detection of protein methacrylylation in different cell lines. Hela and A549 cells were incubated with 20 mM sodium methacrylate for 24 hrs, and the extracted whole cellular proteins were labeled by the probe. (F) Detection of cellular protein methacrylylation in different cellular compartments. HEK293T cells were incubated with 20 mM sodium methacrylylate for 24 hrs and then the cellular proteins were extracted from nucleus, cytosol, and mtochondrion.

histone H3 and histone H4, which is consistent with the previous results detected by anti-methacrylylation antibody.[43] However, with 20mM sodium crotonylate treatment, we didn't observe increased signal for any histone band, whereas a decreased signal on histone H4 was shown. This unexpected decreased labeling is probably due to the competition on the same modification sites between the crotonylation and methacrylylation on histone H4.[43] With anti-crotonylation antibody detection, the signals on all the histone bands for 20 mM sodium crotonylate treatment were higher than the control (**Figure 3.2B**).

Lysine deacetylases such as HDAC3, SIRT1 and SIRT2 have been demonstrated to enzymatically remove the methacryl group from lysine residues.[43] Therefore, we incubated

HEK293T cells with deacetylases inhibitors, e.g., 5μM SAHA or 10mM sodium butyrate for 12 hours and then treated the cell with additional 20mM sodium methacrylate to increase the cellular methacrylylation. As expected, we clearly saw increased signals on histone bands with the inhibitor treatment (**Figure 3.2C**). Next, we extracted the whole cellular protein to explore the protein methacrylylation on other proteins beyond histones. As shown in **Figure 3.2D**, treatment of 20 mM sodium methacrylate led to a substantially increased signal with multiple proteins. Again, incubation with the sodium crotonylate couldn't change the labeling signal indicating that limited crotonylated proteins could react with the probe. Taken together, these results demonstrated that the probe Gly-TCEP-AK could be used for selective labeling of the cellular methacrylylated proteins including histones and other proteins in complex proteomes but with minimum reactivity towards crotonylated proteins.

Next, we applied the probe to study the cellular protein methacrylylation. First, we incubated the HEK293T cells with 20 mM sodium methacrylate to boost the celluar protein methacrylylation and then the extracted cellular proteins were reacted with the probe gly-TCEP-AK through a time-dependent manner followed by CuAAC reaction for visualization. The protein labeling was increased with prolonged incubation time and reached maximum over 12 h (**Supplementary Figure S3.5A**). Moreover, we observed an increased methacrylylation with a dose-dependent sodium methacrylate treatment to HEK293T cells (**Supplementary Figure S3.5B**). We also tested the reaction between the cellular methacrylylated proteins with the probe in different pH condition and found that the reaction had a wide range of tolerance over the pH 6-9 (**Supplementary Figure S3.5C**). Furthermore, we detected the cellular protein methacrylylation in different cell lines and found that with 20 mM sodium methacrylate incubation, the protein methacrylylation was clearly increased in both Hela and A549 cells (**Figure 3.2E**). Interestingly,

previous studies have identified a large number of crotonylated proteins in both Hela and A549 cells and showed that non-histone crotonylated proteins participated in numerous biological functions.[54, 137] The results in current study demonstrated that multiple proteins could also be methacrylylated in cancer cells and showed a great potential to regulate critical cellular functions and served as possible therapeutical targets. Although it has shown that sodium methacrylate could increase cellular MC-CoA level and further promote the protein methacrylylation, little is known which part of the protein in cells could be affected. Therefore, we incubated the HEK293T cells with sodium methacrylate and extracted cellular proteins from different compartments including cytosol, mitochondrion, and nucleus. Intriguingly, although the protein labeling from all the compartments was enhanced by sodium methacrylate, there is a dramatically increase in cytosolic proteins (**Figure 3.2F**). This data shows that sodium methacrylate, as a source of cellular MC-CoA, mainly induces the protein methacrylylation in cytosolic compartment.

Identification of the sodium methacrylate and valine induced methacrylylated proteome

Except for sodium methacrylate, previous studies have shown that valine catabolism is also an important source to generate endogenous MC-CoA.[124] Additionally, our previous work have demonstrated that the cellular biosynthesis of isobutyryl-CoA, a precursor of MC-CoA in valine catabolism pathway, could be stimulated by feeding cells with valine.[45] We therefore proceeded to test whether valine could also be able to increase the cellular protein methacrylylation level by our probe (**Figure 3.3A**). First, HEK293T cells were incubated with 20 mM sodium methacrylate or valine for 24 hours. The histone proteins were then collected, reacted with gly-TCEP-AK and further conjugated with biotin through CuAAC reaction as aforementioned. The results showed that with the incubation of either sodium methacrylate or valine, the methacrylylation on histone H3 could be increased (**Figure 3.3B**). Interestingly, we observed a much more increased

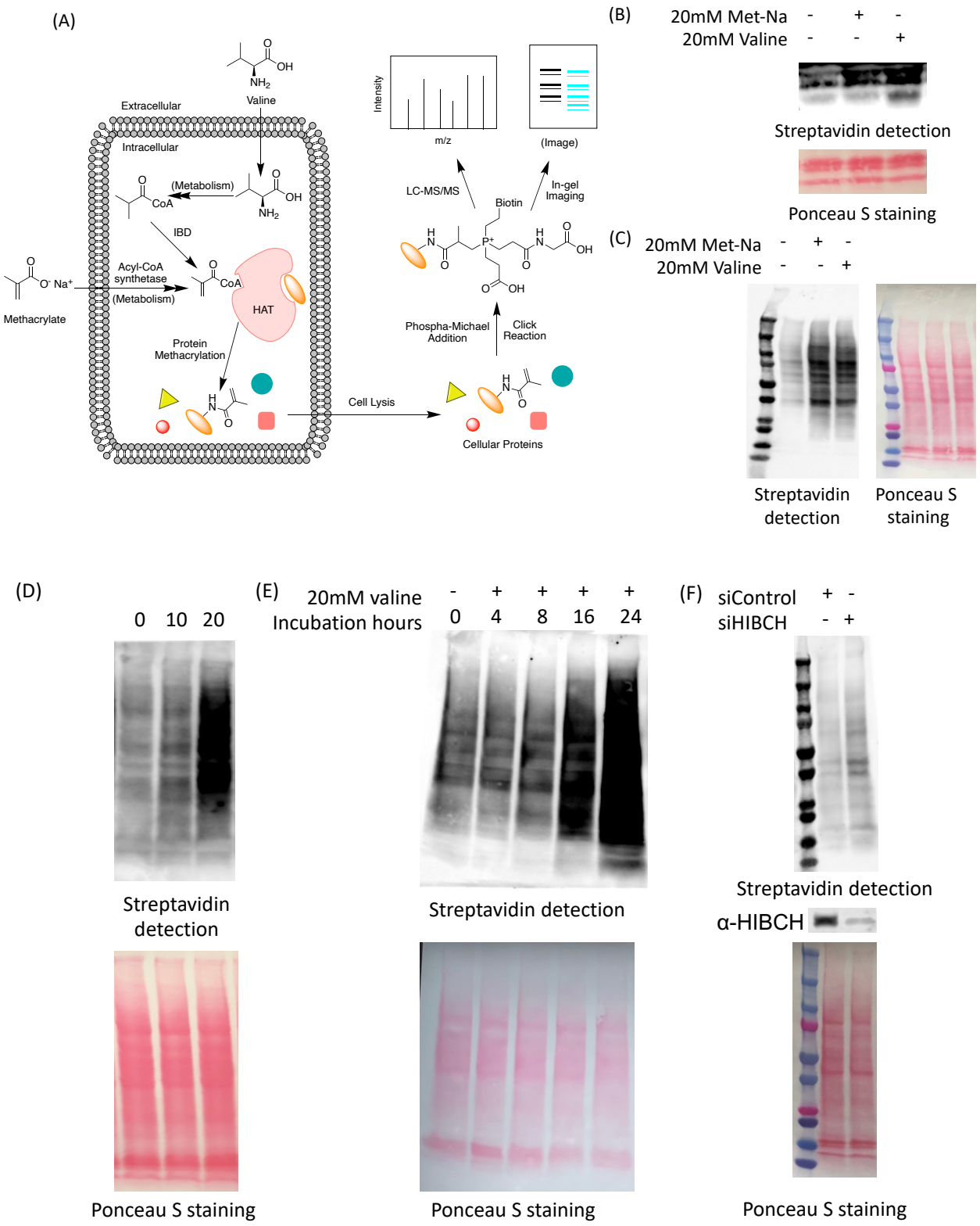


Figure 3.3 Detection of valine metabolism dependent protein methacrylylation. (A) Schematic

graph to identify the methacrylylated proteins dependent on valine and sodium methacrylate metabolism by Gly-TCEP-AK. (B) Detection of valine dependent methacrylylated histones. HEK293T cells were incubated with 20 mM Met-Na or 20 mM valine, and then the histone proteins were extracted and labeled with Gly-TCEP-AK. (C) Detection of valine dependent methacrylylated whole cellular proteins. (D) Dose-dependent valine treatment to HEK293T cells to induce cellular protein methacrylylation. Cells were incubated with 10 or 20 mM valine for 24 hrs. (E) Time-dependent valine treatment to HEK293T cells to induce protein methacrylylation. Cells were incubated with 20 mM valine for varied time. (F) Detection of cellular protein methacrylylation with knock-down of the enzyme HIBCH.

methacrylylation on histone H4 as compared with control and sodium methacrylate treatment, implying the strong connection between valine mitochondrial metabolism and nuclear regulation. We also extracted the whole cellular protein to detect the influence of valine treatment on other proteins. As shown in **Figure 3.3C**, the methacrylylation on multiple proteins were increased by incubating the cell with valine. These data showed that valine could not only induce methacrylylation on histone proteins but also many other proteins, suggesting a broad involvement of valine catabolism in cellular regulation via this important PTM. Additionally, the dose-dependent protein methacrylylation by valine treatment was also observed (**Figure 3.3D**).

Moreover, we incubated the cells with 20 mM valine through a time-dependent manner to explore the dynamic regulation of valine catabolism towards protein methacrylylation. Specifically, we didn't observe strongly increased protein methacrylylation with 4- or 8-hours' incubation. However, with prolonged incubation time, the methacrylylation on multiple proteins were greatly enhanced (**Figure 3.3E**). We further studied the influence of valine catabolism

towards protein methacrylylation by knockdown of the protein, 3-Hydroxyisobutyryl-CoA hydrolase (HIBCH), a key enzyme in valine catabolism pathway. Previous studies have demonstrated that HIBCH mutation could cause the accumulation of MC-CoA, which could further lead to neurological diseases by reacting with some sulfhydryl-containing enzymes, but the detailed mechanisms and direct evidence are still largely unknown.[138-140] Therefore, we transfected the siRNA of HIBCH into HEK293T cells, and after 48 hours' incubation, the lysate proteins were reacted with gly-TCEP-AK and further conjugated with biotin by CuAAC reaction as above described. With the knockdown of HIBCH, the methacrylylation level of certain proteins were clearly increased (**Figure 3.3F**). These studies demonstrated an important valine metabolic route to induce cellular protein methacrylylation and provide an alternative mechanism of MC-CoA accumulation accounting for HIBCH-deficiency induced neurodevelopmental disorders.

Although from the imaging results, both sodium methacrylate and valine could promote the cellular protein methacrylylation, specific protein profiles remain to be elucidated. Therefore, we next sought to identify the cellular methacrylylated proteins induced by the two different metabolites through quantitative LC-MS/MS analysis. As illustrated in **Figure 3.4A**, HEK293T cells were incubated with 20 mM sodium methacrylate or 20 mM valine for 24 hr, respectively. We also included another group with PBS incubation as the negative control. The collected protein samples were further subjected to Phospha-Michael addition and then conjugated with biotin through CuAAC reaction as aforementioned. After removing the extra azide-biotin by precipitating the protein samples with excess acetone, the biotin tagged proteins were further enriched on streptavidin beads and then subjected to on-beads digestion. The obtained peptides were then labeled by tandem mass tags (TMTs) for multiplexed quantitative proteomic analysis. The TMT tagged peptides were able to generate different intensities of reporter ions from different

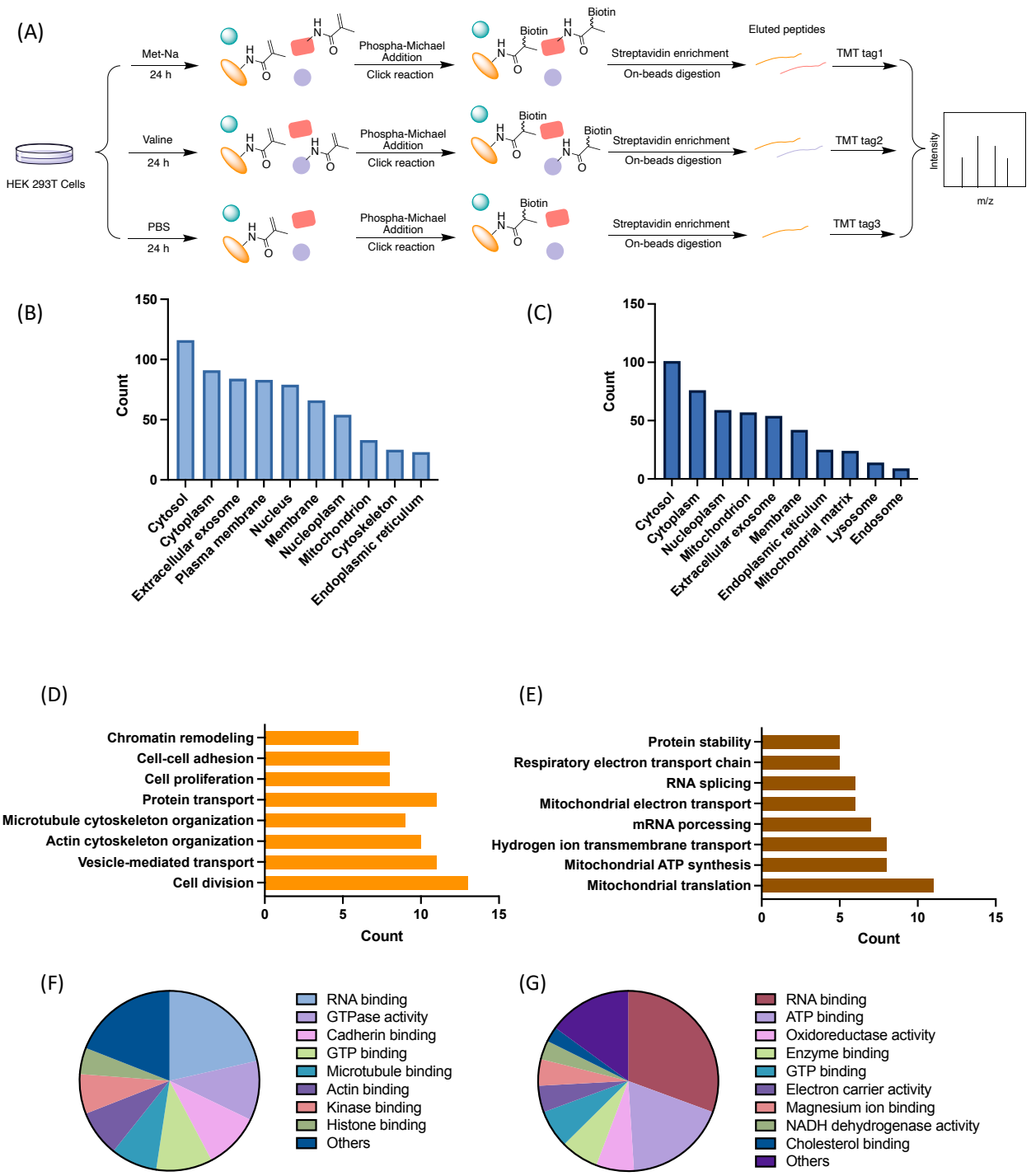


Figure 3.4 Quantitative chemoproteomic profiling of sodium methacrylate and valine dependent methacrylylated proteins by Gly-TCEP-AK in HEK293T cells. (A) Workflow for proteomic identification of sodium methacrylate and valine dependent methacrylylated proteins. (B) Cellular component analysis of Met-Na dependent methacrylylated proteins. (C) Cellular

component analysis of valine dependent methacrylylated proteins. (D) Biological process functional annotation of Met-Na dependent methacrylylated proteins. (E) Biological process functional annotation of valine dependent methacrylylated proteins. (F) Representative molecular function analysis of Met-Na dependent methacrylylated proteins. (G) Representative molecular function analysis of Valine dependent methacrylylated proteins.

channels in the tandem MS, which could be used for quantifying peptides. To ensure the reliability of the results, we performed replicate experiments. By averaging the corresponding intensities from the reporter ions, we got two quantitative ratios: the Meth/Ctrl and Val/Ctrl. We further applied cutoff ratio of 1.2 for sodium methacrylate dependent proteins and 1.1 for valine dependent proteins and got 254 and 256 proteins, respectively and totally found 48 overlapped proteins (**Supplementary Figure S3.6**).

To better understand the biological functions of the newly identified methacrylylated proteins and study the different regulatory mechanisms of the two metabolites dependent methacrylylated proteins, we performed Gene Ontology (GO) analysis through the Database for Annotation, Visualization and Integrated Discovery (DAVID).[111] Cellular component analysis showed that the majority of both of the sodium methacrylate and valine dependent methacrylylated proteins are localized in cytosol. However, more of the valine induced methacrylylated proteins are found in mitochondrion (**Figure 3.4B and 3.4C**). These results indicated that different metabolites might induce protein methacrylylation from different cellular compartments and valine dependent methacrylylation could have more influence on mitochondrion proteins compared with sodium methacrylate induced protein methacrylylation. In addition, biological process analysis showed that sodium methacrylate induced methacrylylated proteins are involved

in lots of important cellular functions such as cell division, vesicle-mediated transport, actin and microtubule cytoskeleton organization, protein transport and cell proliferation (**Figure 3.4D and 3.4E**). Again, we found that valine induced methacrylylated proteins have more effect on mitochondrial functionalities, i.e., mitochondrial translation, mitochondrial ATP synthesis, and mitochondrial electron transport. Molecular function analysis demonstrated that sodium methacrylate induced methacrylylated proteins possess RNA binding, cadherin binding and GTP binding activities, whereas valine dependent methacrylylated proteins mainly mediate RNA binding, ATP binding and enzyme binding activities (**Figure 3.4F and 3.4G**). Taken together, these functional survey results show that both metabolites induced methacrylylated proteins are involved in diverse biological processes, regulating cellular functions. Specifically, valine has more influence on the methacrylylation of mitochondrial proteins.

Identification of HAT1 dependent methacrylylated proteome

Previous study has demonstrated that HAT1 is able to catalyze protein methacrylylation on histone H4 in vitro and in mammalian cells.[43] Our data shows that hundreds of HAT1 protein substrates are localized in different cellular compartments.[123] Therefore, our next step is to apply the probe Gly-TCEP-AK to examine whether HAT1 could catalyze methacrylylation on cellular proteins beyond histone H4. We first incubated the mouse embryonic fibroblasts (MEF) cell line (36T) with sodium methacrylate through a dose-dependent manner for 24 hr. The collected cell lysate proteins were reacted with the probe gly-TCEP-AK and then conjugated with biotin as above mentioned. As a result, the sodium methacrylate could promote protein methacrylylation on diverse cellular proteins in 36T cells, with a strong labeling signal for 20 mM treatment (**Figure 3.5A**). Then we knocked out the HAT1 enzyme and obtained a $HAT^{-/-}$ cell line (33T) to study the

HAT1 dependent methacrylylated proteins. As shown in **Figure 3.5B**, with the incubation of sodium methacrylate, the methacrylylation on cellular proteins were decreased with HAT1 knock-

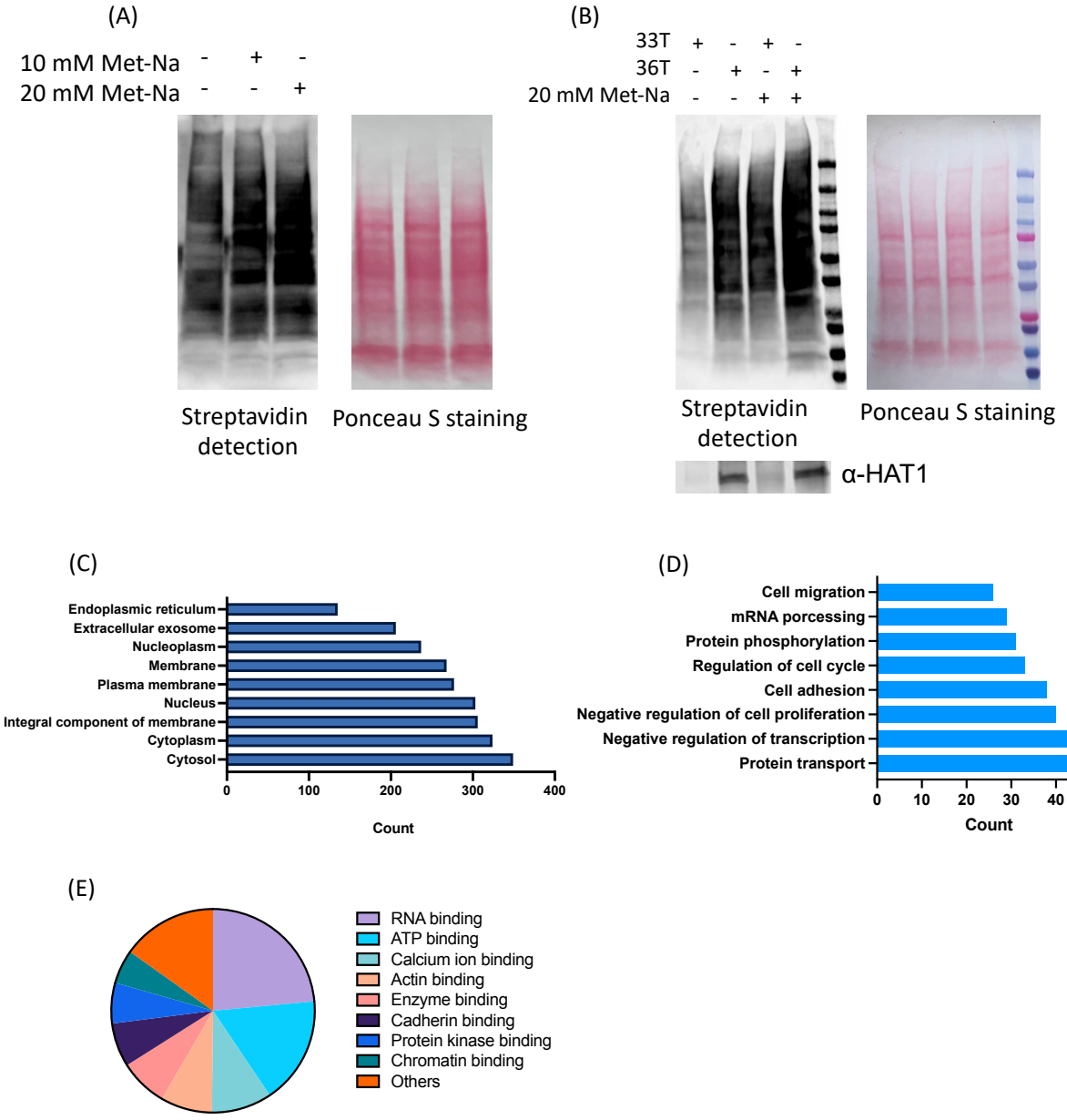


Figure 3.5 labeling and proteomic profiling of HAT1 dependent methacrylylated proteins. (A) Labeling of methacrylylated proteins in 36T cells in a dose-dependent manner. 36T cells were incubated with sodium methacrylate in varied concentrations for 24 hours. (B) Detection of HAT1 dependent methacrylylated proteins. 33T and 36T cells were incubated with 20 mM Met-Na for

24 hours. (C) Cellular component analysis of HAT1 dependent methacrylylated proteins. (D) Biological process analysis of HAT1 dependent methacrylylated proteins. (E) Molecular function analysis of HAT1 dependent methacrylylated proteins.

out. Importantly, there is still an increased protein methacrylylation with the sodium methacrylate treatment in 33T cells, indicating that HAT1 is not the only catalyzing enzyme for protein methacrylylation. We also incubated MEF cells with 20 mM valine and found that knock-out of HAT1 could also decrease valine dependent protein methacrylylation (**Supplementary Figure S3.7**). These data demonstrate that sodium methacrylate and valine can also induce protein methacrylylation in MEF cells and HAT1 is one of the main regulatory enzymes for protein methacrylylation, which can catalyze protein methacrylylation on diverse cellular proteins except for histone H4.

Next, we applied the probe Gly-TCEP-AK to globally identify the HAT1 dependent methacrylylated proteins. To this end, 33T and 36T cells were either incubated with or without 20 mM sodium methacrylate for 24 hr. The resulted cell lysate proteins were reacted with Gly-TECP-Ak, conjugated with biotin, and enriched on streptavidin beads. Then the proteins were digested by trypsin and the released peptides were further subjected to TMT labeling as aforementioned for LC-MS/MS analysis. We totally identified 6567 proteins and then by averaging the corresponding intensities of the peptides from the reporter ions, we got the quantitative ratios for identifying the HAT1 dependent methacrylylated proteins: 36T(20Met)/33T(20Met). With further applying the cutoff ratio at 1.5, we finally selected 1002 proteins as our prominent targets. We then analyzed these proteins by Gene Ontology (GO) through the Database for Annotation, Visualization and Integrated Discovery (DAVID) to study the important biological functions.[111] Specifically, cellular

compartment analysis revealed that the HAT1 dependent methacrylylated proteins are mostly localized in cytosol, cytoplasm, integral component of membrane, nucleus and plasma membrane (**Figure 3.5C**). The analysis on biological process showed that the proteins are associated with protein transport, transcription, cell proliferation, cell adhesion, cell cycle, protein phosphorylation, mRNA processing, and cell migration, indicating a broad involvement of the HAT1 dependent methacrylylated proteins in critical cellular biological functions (**Figure 3.5D**). In addition, molecular function analysis suggested that many of the proteins possess RNA, ATP, calcium ion, actin, enzyme, cadherin and chromatin binding activities (**Figure 3.5E**). Overall, these functional investigation results indicate that the HAT1 dependent methacrylylated proteins are highly enriched in various biological processes and reveal a novel link between HAT1 and protein methacrylylation.

3.3 Conclusion

To summarize, we have developed a highly reactive and selective chemical probe, Gly-TCEP-AK, for labeling and profiling protein methacrylylation with mammalian cells. The probe exhibits great reactivity both with peptides and proteins and shows selectivity towards methacrylylation but not crotonylation. The protein imaging data showed that the methacrylylation exist on multiple cellular proteins beyond histones. Importantly, we showed that the protein methacrylylation exist in varied cell lines including HEK293T, Hela, A549 and MEF cells. It is therefore worth noting that further investigations are still needed for the study of how methacrylylation affect protein functions in different cells especially in cancer cells, which might contribute on cancer progression. Except for sodium methacrylate, our data showed that valine is another important source to promote cellular protein methacrylylation. Additionally, we found by knock-down of HIBCH, a critical enzyme in the valine metabolic pathway, the protein

methacrylylation can be promoted, which provides an important mechanism to study HIBCH related diseases. We further applied the probe to globally identify the sodium methacrylate and valine dependent methacrylylated proteins through quantitative chemoproteomics and found hundreds of new methacrylylated proteins. To the best of our knowledge, this is the first proteome analysis to show the methacrylylated proteins in mammalian cells. Notably, we found the different protein profiles for the two metabolites, indicating that sodium methacrylate and valine may influence different cellular functions through protein methacrylylation. Moreover, HAT1 has been proved to catalyze protein methacrylylation on histone H4 in previous study. Our study on HAT1 activities suggests that HAT1 can catalyze methacrylylation on multiple cellular proteins besides histone H4. Importantly, we identified over one thousand HAT1 dependent methacrylylated proteins, which expands the substrate scope of both protein methacrylylation and HAT1, aiding in further study of this newly identified PTM and the regulatory role of HAT1 in cellular functions. Overall, our study provides a powerful chemical probe, which enables labeling and identification of methacrylylated proteins in different biological contexts. Additionally, our protein set offers a valuable resource for further detection of the function of protein methacrylylation. With the probe readily applicable for labeling and enrichment of the cellular methacrylylated proteins, we envision that more functional roles of this under-investigated PTM will be revealed in the future.

3.4 Methods and materials

Reagents and materials

Unless otherwise noted, all chemicals and solvents were purchased from Sigma–Aldrich and used without further purification. Fmoc-protected amino acids and preloaded Wang resin were purchased from NovaBiochem. Click chemistry reagents biotin-azide (catalog# 1265) was purchased from Click Chemistry Tools. Protein dual color standards and precast protein gels (4%–

20%) were purchased from Bio-Rad Laboratories. High sensitivity streptavidin-HRP was purchased from Thermo Fisher Scientific (catalog# 21130). High-capacity streptavidin agarose was purchased from Thermo Fisher Scientific (catalog# 20359). The anti-HAT1 antibody (catalog# 11432) was purchased from Proteintech. The anti-Histone H3 antibody (catalog# sc-517576) was purchased from Santa Cruz. The anti-acetyllysine antibody (catalog# PTM-105) was purchased from PTM BIO LABS. The anti-GAPDH antibody (catalog# 2118) was purchased from Cell Signaling. The anti-Hsp60 antibody (catalog# ab59457) was purchased from Abcam. The anti-rabbit-HRP linked antibody (catalog# 7074) was purchased from Cell Signaling. The anti-mouse-HRP linked antibody (catalog# 7076) was purchased from Cell Signaling. HEK293T, Hela and A549 cells were purchased from American Type Culture Collection (ATCC). MEF cell line was a gift from Dr. Mark Parthun's lab.

Synthesis of N^{α}-Boc,N^{ε}-crotonyl lysine

N^{α}-Boc,N^{ε}-crotonyl lysine was synthesized as described by previous work.[133]

Synthesis of N-terminal H4 peptide H4(2-8)K5metha and H4 peptide H4(1-20)K5cro.

Peptides were synthesized by standard Fmoc solid-phase peptide synthesis (SPPS) protocol on an AAPPTec Focus XC synthesizer using a Wang resin (Fmoc-Lys (Boc)-Wang, 100–200 mesh, ChemPep). The peptides were purified by C-18 RP-HPLC and confirmed by MALDI-MS. Synthesis scale was 0.1 mmol. Specifically, Individual amino acids were coupled using activation reagent HCTU, HOBt, and DIPEA. Lys(DDE) was deprotected with 2% hydrazine in DMF for 2 h at room temperature, drained, and repeated with fresh 2% hydrazine in DMF for 2 h at room temperature. After washing the resin with DMF, the crotonylation and methacrylylation reaction was conducted with 10 eq of crotonic acid or methacrylic acid in DMF, 10 eq HCTU in DMF, and 40 eq NMM in DMF for overnight. The Fmoc deprotection reactions were performed with 20%

piperidine in dimethylformamide. The peptide was cleaved off the resin using 95% trifluoroacetic acid (TFA), 2.5% H2O, and 2.5% triisopropylsilane (TIS) for 4 h. Peptides were precipitated in cold diethyl ether and pelleted by centrifugation. The peptides were precipitated in cold diethyl ether and lyophilized to get solid form.

Synthesis of t-butyl-Gly-TCEP

t-butyl-glycine (59 mg, 352 μmol, 1 eq), EDC (67.4 mg, 352 μmol, 1 eq.) and 95.8 mg (704 μmol, 1mL, 2 eq.) 1-hydroxy-7-azabenzotriazole (HOAt) were dissolved in 2 mL DMF (degassed) under an inert atmosphere. This solution was added to TCEP (300 mg, 1056 μmol, 3eq.) dissolved in 2 mL degassed DMF containing 296 μL (1408 μmol, 4 eq.) DIPEA under an inert atmosphere. The reaction was stirred at room temperature for 1hr. Yield: 62.1mg 48.5%

Synthesis of 2t-butyl-Gly-TCEP

T-but-glycine (14.8 mg, 88 μmol, 1 eq), EDC (16.9 mg, 88 μmol, 1 eq.) and 1-hydroxy-7-azabenzotriazole (HOAt) (24 mg, 176 μmol, 2 eq.) were dissolved in 2 mL DMF (degassed) under an inert atmosphere. This solution was added to 1-t-but-gly-TCEP (96mg, 264 μmol, 3eq.) dissolved in 1.5 mL degassed DMF containing 74 μL (352 μmol, 4 eq.) DIPEA under an inert atmosphere. The reaction was stirred at room temperature for 1hr. Yield: 19.8 mg, 47.2%

Synthesis of 3t-butyl-Gly-TCEP

T-but-glycine (118 mg, 704 μmol, 4 eq), EDC (134.8 mg, 704 μmol, 4 eq.) and 1-hydroxy-7-azabenzotriazole (HOAt) (143.7 mg, 1056 μmol, 6 eq.) were dissolved in 3 mL DMF (degassed) under an inert atmosphere. This solution was added to TCEP (50mg, 176 μmol, 1eq.) dissolved in 1 mL degassed DMF containing 55.5 μL (264 μmol, 1.5 eq.) DIPEA under an inert atmosphere. The reaction was monitored by RP-HPLC and stirred at room temperature for 4 hr. Yield: 38.1 mg, 36.7%

Synthesis of Gly-TCEP-AK

Propargyl amine 11.3 µL (9.68mg, 176 µmol, 1 eq), EDC (33.7 mg,176 µmol, 1 eq.) and 1-hydroxy-7-azabenzotriazole (HOAt) (47.9 mg, 352 µmol, 2 eq.) 1-hydroxy-7-azabenzotriazole (HOAt) were dissolved in 1 mL DMF (degassed) under an inert atmosphere. This solution was added to t-butyl-Gly-TCEP (528 µmol, 191.7 mg, 3 eq.) dissolved in 2 mL degassed DMF containing 148 µL (704 µmol, 4 eq.) DIPEA under an inert atmosphere. The reaction was stirred at room temperature for 30 min to achieve t-butyl-Gly-TCEP-AK. Yield: 30.6 mg, 43.4%

40 mg t-butyl-Gly-TCEP-AK were dissolved in 75% TFA/water (4 ml) and stirred under inert atmosphere for 3 hours in room temperature. Yield: 33.2 mg, 97%

Synthesis of 2Gly-TCEP-AK

Propargyl amine 26µL (19.36 mg, 352 µmol, 2 eq), EDC (67.4 mg, 352 µmol, 2 eq.) and 72 mg (880µL, 528 µmol, 3 eq.) 1-hydroxy-7-azabenzotriazole (HOAt) were dissolved in 1 mL DMF (degassed) under an inert atmosphere. This solution was added to 2t-butyl-Gly-TCEP (176 µmol, 83.8 mg, 1 eq.) dissolved in 1 mL degassed DMF containing 74 µL (352 µmol, 2 eq.) DIPEA under an inert atmosphere. The reaction was stirred at room temperature for 45 min to achieve 2t-butyl-Gly-TCEP-AK. Yield: 37.5 mg, 41.6%

35 mg 2t-butyl-Gly-TCEP-AK were dissolved in 75% TFA/water (4 ml) and stirred under inert atmosphere for 3 hours in room temperature. Yield: 25.7 mg, 94%

Reactions between TCEP Analogs and sodium methacrylate or Boc-crotonyl-lysine

50 mM sodium methacrylate were incubated with 10 mM t-butyl-Gly-TCEP, 2t-butyl-Gly-TCEP or 3t-butyl-Gly-TCEP in 50 mM Tris-HCl buffer, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS.

50 mM Boc-crotonyl-lysine were incubated with 10 mM TCEP, Gly-TCEP-Ak or 2Gly-TCEP-Ak in 50 mM Tris-HCl buffer, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS.

Cell culture and preparation of whole cellular protein

For HEK293T, Hela, A549 and MEF cell culture, cells were grown in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Life Technologies) and 1% penicillin-streptomycin (15140122, Thermo Fisher Scientific) at 37 ℃ in a humidified cell incubator with 5% CO_2.

For preparation of the whole cellular proteins, the cells were plated and cultured to around 90% confluence and then scrapped and harvested after washing through PBS for three times. Cells were collected through centrifugation (5000 rpm x 5 min, 4 ℃) and were resuspended by ice-cold M-PER mammalian protein extraction reagent (ThermoFisher SCIENTIFIC, Product #78501) containing 1% protease inhibitor cocktail (ThermoFisher SCIENTIFIC, Product #78438). After putting the lysate on ice for 15 minutes, to release the whole cellular proteins, cells were sonicated with 30% amplitude. The cellular proteins (supernatant) were collected by centrifugation for 20 minutes (13200 rpm x 20 min, 4 ℃). The protein concentration was determined by Bradford assay.

Western blotting analysis

The HAT1 level, HIBCH level, hsp60 level, GAPDH level, crotonylation level and histone H3 level were detected by the respective antibody. The protein mixture was subjected to SDS-PAGE and transferred to nitrocellulose membrane followed by blocking with 5% non-fat milk for 1 h. Next, the membrane was incubated with the primary antibody at 4 °C overnight. After washing with TBST (10 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween 20) for 3 times (5 minutes for each), the membrane was incubated with secondary antibody at room temperature for 1 h. The membrane

was then washed three times (5 minutes for each) with TBST and developed by enhanced chemiluminescence detection system according to the manufacturer's protocols and scanned by chemiluminescence scanner.

Transfection of siRNA for HIBCH knockdown

HEK293T cells were cultured to reach around 60% confluence stage, and transfection of siRNA of HIBCH was conducted by Lipofectamine 3000 (ThermoFisher SCIENTIFIC, Product #L3000008) according to the manufacture's protocol followed by 48 h of incubation. Next, the cells were washed with ice-cold PBS and harvested and lysed by ice-cold M-PER buffer containing a 1% protease inhibitor cocktail. The protein was then collected and the protein concentration was determined by Bradford assay.

Protein labeling by the probe Gly-TCEP-AK, CuAAC reaction, and streptavidin detection

The methods for labeling, CuAAC reaction and streptavidin detection were adapted from the previous experimental procedures.[129] Specifically, 50 μg cellular proteins (10 μg for histone proteins) were incubated with the probe Gly-TCEP-AK as indicated concentrations at 37 ℃. The whole proteins were then precipitated with 5-fold ice-cold acetone and the precipitated proteins were collected by centrifugation at 5000 g for 10 min at 4 ℃, washed by ice-cold methanol for three times. The resulting proteins were resuspended by PBS with 0.4% SDS. The protein solutions were then subjected to CuAAC reaction. For CuAAC reaction, the proteins were mixed with click cocktail containing 50 μM biotin-azide, 2.5 mM sodium ascorbate, 0.5 mM copper sulfate and 0.25 mM ligand BTTAA. After 1 h incubation at room temperature, the solutions were further mixed with protein loading dye and then boiled for 5 min. The samples were then loaded on 4-20% SDS-PAGE gradient gel. Thereafter, the proteins were transferred to nitrocellulose membrane

and blocked with 5% non-fat milk for 1 h. The membrane was incubated with streptavidin-HRP for another 1 h and scanned by chemiluminescence scanner.

Protein enrichment and protein digestion

Lyophilized lysates were transferred to new tubes and resuspended in 100 mM DPBS. The streptavidin beads were washed in 100 mM DPBS twice and incubated with cell lysates for 2 h at room temperature. The beads were then washed by 100 mM DPBS ten times in a spin column to remove unspecific binding proteins.

The beads were resuspended in the digestion buffer containing 50 mM HEPES pH=8.1, 1.6 M urea, and enriched proteins were digested with 50:1 (w/w) trypsin (Promega) for 16 h at 37 ˚C with shaking. The digestion was quenched by adding trifluoroacetic acid (TFA) to a final concentration of 0.4% after removing beads with spin columns. Digests were centrifuged and desalted using 50 mg SepPak tC18 cartridges (Waters).

TMT labeling and peptide fractionation by HPLC

Purified peptides were lyophilized and resuspended in 100 µL of 100 mM HEPES pH = 8.5, and 30 µL ACN. For each cell type, the samples were labeled with the TMT6plex reagent (Thermofisher). After a one-hour reaction, the excessive amount of the TMT reagents was quenched by adding 10 µL of 5% hydroxylamine. Labeled peptides were pooled, purified, and fractionated into three fractions using 20/50/80% ACN. The collected fractions were further purified by the StageTip method.

LC-MS/MS analysis

Dried peptides were resuspended in 6 µL loading solution containing 5% ACN and 4% formic acid (FA), and 3 µl was loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 5 µm, 200 Å, 75 µm x 16 cm) by a Dionex WPS-3000TPLRS autosampler

(UltiMate 3000 Thermostatted Pulled Loop Rapid Separation Wellplate Sampler). Peptides were separated by reversed-phase HPLC using an UltiMate 3000 binary pump with a 150-minute gradient designed for the TMT-labeled peptides. The full MS and MS2 were detected in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, Thermo Scientific, with Xcalibur 3.0.63 software) using a data-dependent Top15 method. Each cycle included one full MS scan (resolution: 60,000) in the Orbitrap at the automatic gain control (AGC) target of $1*10^6$, followed by up to 15 MS/MS for the most intense ions. Selected ions were excluded from further analysis for 90 s. Ions with a single or unassigned charge were not sequenced. MS2 scans were activated by HCD at 30% normalized collision energy with an isolation width of 1.2 m/z and detected in the Orbitrap cell with a resolution of 15,000.

Database searching, data filtering, and protein quantification

Raw data files recorded from the mass spectrometer were converted into the mzXML format. Mass spectra were searched using the SEQUEST algorithm (version 28) (Eng et al., 1994) against a human proteome (Homo sapiens) database encompassing sequences of all proteins downloaded from UniProt (https://www.uniprot.org/taxonomy/9606). Each protein sequence was listed in both forward and reversed orientations to estimate the false discovery rate (FDR) of peptide and protein identifications. The following parameters were used for the search: 10 ppm precursor mass tolerance; 0.025 Da product ion mass tolerance; fully digested with trypsin; up to two missed cleavages; variable modifications: oxidation of methionine (+15.9949); fixed modifications: TMT labeling of lysine and the N-termini (+229.1630).

The target-decoy method (Elias and Gygi, 2007) was employed to evaluate and control the FDRs of peptide and protein identifications. Linear discriminant analysis (LDA) was used to distinguish correct and incorrect peptide identifications using numerous parameters such as XCorr,

ΔCn, and precursor mass error. After scoring, peptides with fewer than seven amino acids in length were discarded, and peptides were filtered to a less than 1% FDR based on the number of decoy sequences in the final data set.

The TMT reporter ion intensities in the MS2 spectra were used to quantify peptides. The isotopic information provided by Thermo was utilized to correct the ion intensities. The summed intensity for each protein was calculated based on all unique peptides from this protein in each sample.

Bioinformatic analysis

Protein functional annotation information was obtained from UniProt (UniProt, 2019) (https://www.uniprot.org/) and analyzed using the Database for Annotation, Visualization and Integrated Discovery (DAVID (Huang *et al.*, 2009), https://david.ncifcrf.gov/).

CHAPTER 4

DISCOVERY AND PROFILING OF PROTEIN CYSTEINE *S*-2-CARBOXYPROPYLATION

Song, J.; Yin, K.; Wu, R.; Zheng, Y. G. Submitted to J. Am. Chem. Soc, 3/8/2023

Abstract

Methacrylyl-CoA is a key metabolic intermediate in valine catabolic pathway and its accumulation has been found to be cytotoxic and causative to oncological and neurological diseases. Nevertheless, detailed biological effects of methacrylyl-CoA and methacrylate in human physiology and pathology are poorly understood. We propose that the electrophilicity of the alkene bond in the methacrylyl group can react with cysteine residues in proteins resulting in an unexplored protein post-translational modification, cysteine *S*-2-carboxypropylation (C2cp). To test and validate this mechanistic hypothesis, in this study, we detected and profiled *S*-2-carboxypropylated proteins from the complex cellular proteome with the design and application of a bioorthogonal chemical probe, N-propargyl methacrylamide. We tested the probe in different mammalian cell models and demonstrated its versatility and sensitivity to protein cysteine *S*-2-carboxypropylation. We established quantitative chemical proteomics for global and site-specific profiling of protein *S*-2-carboxypropylation, which successfully identified 403 *S*-2-carboxypropylated proteins and 120 cysteine modification sites from HEK293T cells. Through bioinformatic analysis, we found that C2cp-modified proteins were involved in a variety of critical cellular functions including translation, RNA splicing and protein folding. Our chemoproteomic studies demonstrating the proteome-wide distribution of cysteine *S*-2-carboxypropylation provide a new biochemical mechanism for functional investigation of methacrylyl-CoA and understanding valine metabolic disorders.

4.1 Introduction

Branched-chain amino acids (BCAAs) that include leucine, isoleucine, and valine are essential building blocks for protein synthesis in all living organisms. BCAAs account for protein anabolic functionalities, energy production, signaling transmission, etc.[45, 141, 142] Administration of

leucine to food-deprived rats has been shown to effectively stimulate protein synthesis in skeletal muscles.[143] Feeding the rat muscle cells with leucine suppresses the activity of AMP-activated protein kinase (AMPK), which is known to maintain energy homeostasis in various tissues.[144] Isoleucine or valine deficiency reduces the fat mass in mice through promoting energy expenditure and modulating fat metabolism.[145] Recent studies on the isoleucine metabolism in mammalian cells suggest that it induces β-defensins and shows immunotherapy effect in infectious disease.[146] Importantly, dysregulation of BCAAs has been found to associate with multiple diseases such as liver cirrhosis, diabetes, immune system disorders and urine diseases.[147, 148] For instance, BCAAs regulate the mTOR signaling pathway, a key therapeutic target in cancer biology.[149] Emerging roles of BCAAs in insulin resistance and strong correlation with obesity have also been revealed.[150, 151] Biochemically, the catabolic pathways of BCAAs consist of multiple enzyme-mediated metabolic steps including reversible transamination, irreversible oxidative decarboxylation, and dehydrogenation.[152, 153] The major end products (acetyl-CoA, succinyl-CoA) of BCAA metabolism further enter the Krebs cycle for energy production or participate in protein post-translational modifications.[92, 154, 155]

Methacrylyl-CoA (MC-CoA) is a key intermediate metabolite generated in the valine degradation pathway. High levels of MC-CoA are toxic to cells and have been implicated in multiple valine metabolism-associated disorders.[138, 156] Studies on the toxicity of MC-CoA can be traced back to 1980s when it was found that the inherited mutation of 3-hydroxyisobutyryl-CoA hydrolase (HIBCH) enzyme led to the physical malformation and inborn error.[139, 157] HIBCH deficiency or loss of the short-chain enoyl-CoA hydratase (ECHS1) activity generates excess MC-CoA which is a causing factor for the metabolic diseases such as liver dysfunction, impaired ATP production and lactic acidosis.[138, 157, 158] It was further found that activities of MC-CoA hydratase

and HIBCH decreased in livers with cirrhosis or hepatocellular carcinoma[140], indicating decreased capability for detoxifying MC-CoA. The defect of ECHS1 in leigh disease was detected, which led to the accumulation of MC-CoA and brain pathology.[158] From the chemical perspective, the double bond on MC-CoA owns a strong electrophilic nature and can undergo the Michael addition to conjugate with a free thiol group.[156] Clinical studies showed that N-acetyl-*S*-(2-carboxypropyl)cysteine, which is likely produced from MC-CoA, is a useful biomarker for the diagnosis of the ECHS1 deficiency (ECHS1D).[159, 160] By incubating the fibroblasts with either radioisotope-labeled valine or cysteine, elevated secretion of *S*-2-carboxypropyl-cysteamine and *S*-2-carboxypropyl-cysteine molecules from patients with HIBCH mutations has been identified.[139] However, direct evidence for *S*-2-carboxypropylated proteins are unexplored. Recently, we and colleagues found that as a key epigenetic enzyme, histone acetyltransferase 1 (HAT1) serves as a mitochondrial modulator[123] and catalyzes protein methacrylylation by transferring the methacrylic group from MC-CoA to histone proteins[43]. The connections among valine metabolism, MC-CoA, and protein modifications warrant further investigation.

The goal of this study is to pursue the first investigation of cysteine *S*-2-carboxypropylation (C2cp) in the human proteome. Being an unexplored protein PTM, no antibody is available for C2cp detection in proteins using standard western blotting or immunoprecipitation methods. To circumvent this technical barrier, we designed N-propargyl methacrylamide (PMAA) as a bioorthogonal chemical probe to set up an antibody-free chemoproteomic platform for the detection and profiling of C2cp marks in the human proteome. Through quantitative and site-specific proteomic analysis, we have successfully identified more than 400 C2cp-containing protein targets and 100 modified cysteine sites. Functional annotation revealed that cysteine *S*-2-carboxypropylated proteins are involved in diverse essential cellular processes ranged from gene

transcription, protein translation, RNA splicing, protein folding, to energy production. This study establishes C2cp as a new protein PTM and discloses an important molecular mechanism for understanding valine defect-induced metabolic and neurological diseases.

4.2 Results and discussion

Development of chemoproteomic probes to study cysteine *S*-2-carboxypropylation

On account of the chemical reactivity of the methacrylyl group, we hypothesized that either MC-CoA or sodium methacrylate (Met-Na) could non-enzymatically react with cysteine residues in proteins to result in a hitherto unstudied protein modification, cysteine *S*-2-carboxypropylation (C2cp) (**Figure 4.1A**). To investigate the possible modification of cysteine residues by methacrylate group on cellular proteins, we examined the reactions of MC-CoA with several thiol-containing biological compounds (cysteine, cysteamine, and coenzyme A). Product detection with MALDI-MS showed that MC-CoA was efficiently conjugated with all of these molecules (**Supplementary Figure S4.1**). The result was consistent with a previous report through detection of free sulfhydryl groups by incubating the equimolar concentrations of MC-CoA with varied sulfhydryl compounds.[157] Furthermore, we tested the reactivity between MC-CoA and a cysteine-containing tripeptide, glutathione, and found it was able to be modified by MC-CoA (**Supplementary Figure S4.2**). These results suggest that direct chemical modification of proteins by MC-CoA is highly likely. In contrast, for the reactions between Met-Na and the above-mentioned thiol-containing compounds, none or very little signals of adduct formation were detected through mass spectrometry (data not shown). Therefore, reactivity of MC-CoA was much higher than Met-Na. Furthermore, we sought to synthesize a methacrylic acid analogue, N-propargyl methacrylamide (PMAA) as a bioorthogonal and competitive chemical probe to identify and profile *S*-2-carboxypropylated proteins (**Figure 4.1B**). We introduced a short amide bond in

the probe for better biocompatibility and less steric hindrance. For comparison, we also synthesized a control compound, N-propargyl isobutyrylamide (PIBA), and used it to compete for protein labeling by the probe PMAA (**Figure 4.1B**). PMAA and PIBA were synthesized following the previously reported methods.[161, 162]

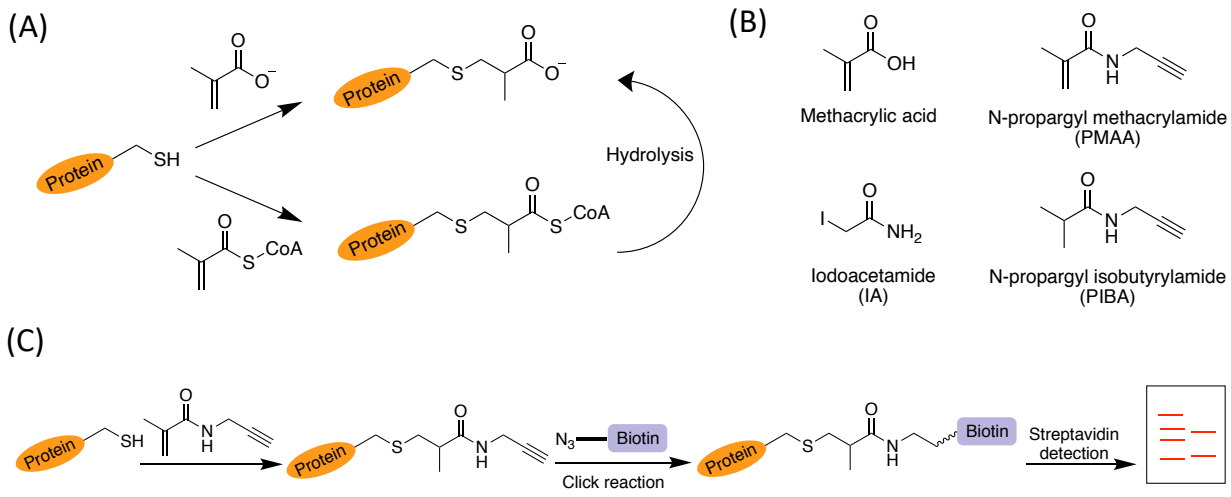


Figure 4.1 Design of chemical probes for *S*-2-carboxypropylated protein profiling. (A) Proposed pathway for the formation of cellular protein *S*-2-carboxypropylation. The protein can either directly conjugate with methacrylate or react with MC-CoA first and then be hydrolyzed to form *S*-2-carboxypropylation. (B) The structures of methacrylic acid, PMAA, IA and PIBA. (C) Scheme for the probe PMAA to capture cellular *S*-2-carboxypropylated proteins. The cellular proteins were reacted with PMAA and further conjugated with biotin through click reaction, after which the labeled proteins can be visualized through streptavidin detection.

Evaluation of the PMAA probe for protein cysteine modification

Next, we investigated how cellular proteins can be labeled by PMAA. Our idea is that PMAA reacts with cysteine residues on complex protein mixtures, after which the modified proteins can be biotin tagged through the alkyne click handle by conjugating with biotin-azide and

then be detected by chemiluminescence (**Figure 4.1C**). In the experiment, cultured HEK293T cells were lysed and the whole lysate proteins were incubated with PMAA with a dose-dependent and a time-dependent manner, respectively. The labeled proteins were then reacted with biotin-azide (Click Chemistry Tools, catalog# 1265) through copper-catalyzed azide-alkyne cycloaddition (CuAAC), resolved on SDS-PAGE, and imaged by streptavidin-HRP. As shown in **Figure 4.2A**, the whole cell lysates were effectively labeled with PMAA. The labeling became stronger with increasing concentrations of the probe and labeling signals were detectable at as low concentration as 0.2 mM. Also, the labeling level increased with prolonged incubation times in the range of 1-16 h (**Figure 4.2B**). Combined, incubation with 2 mM PMAA for 12 h would reach a saturating level of labeling on whole lysate proteins. We then proceeded to test whether the labeling was resulted from the conjugation between the cysteine residues on cellular proteins and the alkene bond of the PMAA probe. As shown in **Figure 4.2C**, the PMAA labeling signal was significantly abolished by the presence of iodoacetamide (IA), a commonly used alkylating agent that reacts with the thiol group of cysteines. In contrast, when either PIBA or sodium isobutyrate which did not have a thiol-reactive functional group was co-incubated in the mixture, there was no influence on the PMAA labeling (**Supplementary Figure S4.3A and S4.3B**). These comparative results demonstrate that proteins were labeled by PMAA specifically through the Michael addition reaction with the side chain sulfhydryl group of cysteine residues (**Figure 4.1C**). We also tested whether PMAA could be used as a robust bioorthogonal probe to label proteins in different cellular systems. After incubating the probe with different types of cellular proteomes, strong protein labeling bands were observed in the whole lysate proteins from several cell lines tested, including mouse embryonic fibroblast cell line (36T), colon cancer cell line (HCT116) and human epithelial cell line (HeLa) (**Supplementary Figure S4.4**). Varied patterns of labeled proteins were seen from

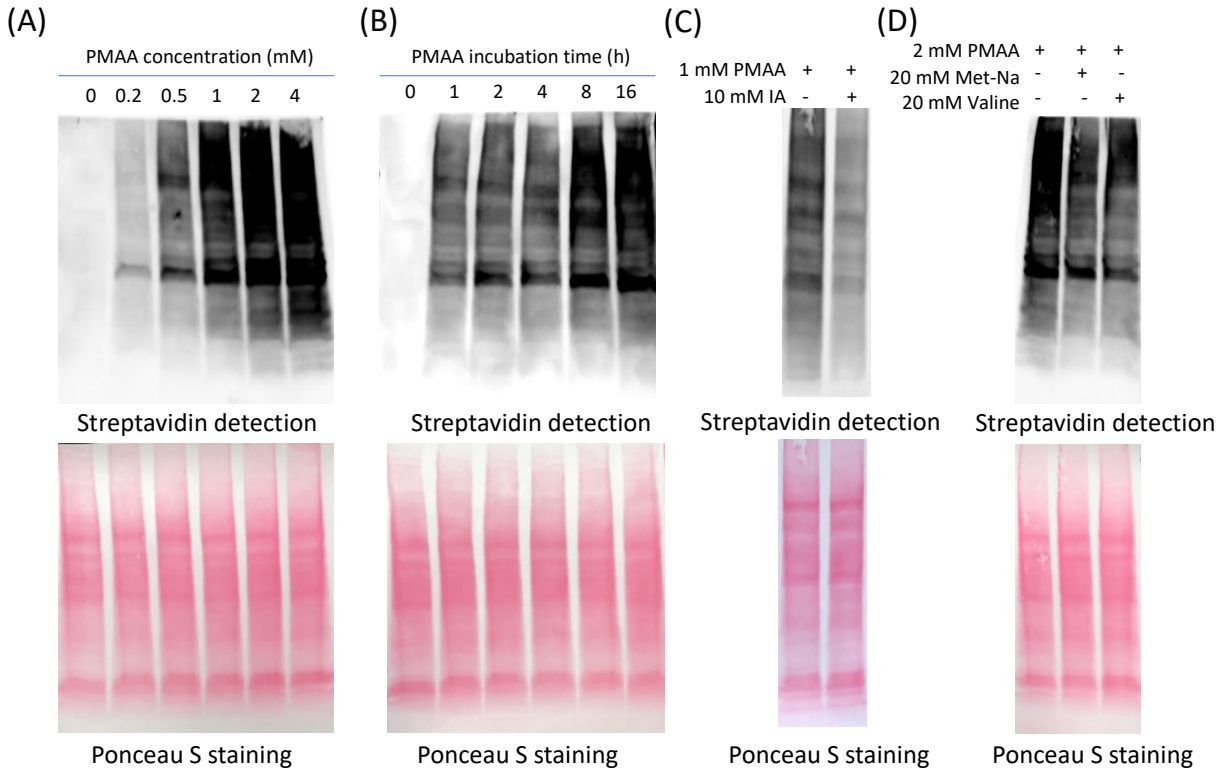


Figure 4.2 Evaluation of the probe PMAA labeling to detect *S*-2-carboxypropylation in native proteomes. (A) Concentration-dependent labeling of PMAA with cellular proteomes. HEK293T cellular proteins were incubated with probe PMAA with indicated concentration at 37 ℃ for 12 h, followed by CuAAC reaction to conjugate with biotin and detected by streptavidin-HRP. (B) Time-dependent labeling of PMAA with cellular proteomes. HEK293T cellular proteins were incubated with 2 mM PMAA at 37 ℃ for indicated time, followed by CuAAC reaction. (C) Using iodoacetamide (IA) as a competitive probe to block free cysteine. The HEK293T cellular proteins were pretreated with 10 mM IA for 2 h and then incubated with 1 mM PMAA for another 4 h (37 ℃), followed by CuAAC reaction. (D) Competition of PMAA labeling by endogenous *S*-2-carboxypropylation induced by sodium methacrylate (Met-Na) and valine. HEK293T cells were incubated with 20 mM Met-Na or 20 mM valine separately, and the resulted cellular protein were reacted with 2 mM PMAA at 37 ℃ for 12 h, followed by CuAAC reaction.

different cell lines, which suggests that protein C2cp modification profiles vary in different cells. To probe endogenous levels of protein *S*-2-carboxypropylation, we cultured HEK293T cells in the presence of 20 mM Met-Na or valine to boost cellular MC-CoA levels, and then the probe PMAA was used to label the cellular proteins to compete for the endogenous C2cp in proteins. As revealed in **Figure 4.2D** and **Supplementary Figure S4.5**, incubation of the cells with either 20 mM Met-Na or 20 mM valine significantly decreased PMAA-driven labeling signals on the cellular proteins as compared with the cells with no treatment of Met-Na or valine. These results can be explained by the fact that Met-Na and valine promoted endogenous cellular levels of methacrylate or MC-CoA, which led to enhanced protein C2cp modifications. Such enhanced protein C2cp modifications in cells reciprocally reduced the numbers of those cysteine sites that are accessible to PMAA labeling. These data demonstrate that PMAA can be applied as a competitive bioorthogonal chemical probe to profile C2cp substrates on cellular proteins.

Identification of *S*-2-carboxypropylated protein substrates in the human cellular proteome

We next applied PMAA as a chemoproteomic probe to identify *S*-2-carboxypropylated proteins in HEK293T cells. The workflow was depicted in **Figure 4.3A**: HEK293T whole lysate proteins were prepared and then incubated with either 4 mM PMAA or 4 mM PIBA for 16 h at 37 °C, respectively. The protein samples were then precipitated with excess acetone and washed with ice-cold methanol to remove unreacted chemical probes. Thereafter, the labeled proteins were subjected to CuAAC click reaction for conjugation with azide-diazo-biotin (Click Chemistry Tools, catalog# 1041), and then enriched on streptavidin beads. The affinity enriched proteins were cleaved from the beads with sodium dithionite ($Na_2S_2O_4$) and then resolved on SDS-PAGE gel and imaged by silver staining. As expected, much more protein bands were shown in the group with

PMAA incubation compared with the group with PIBA incubation (**Supplementary Figure S4.6**). Next, we sought to globally identify the PMAA labeled proteins using MS-based proteomics. To this end, the affinity enriched proteins were subjected to on-bead trypsin digestion, and the resulting peptides were further labeled by tandem mass tags (TMTs) for multiplexed quantitative proteomic analysis (**Figure 4.3A**). The peptides labeled by the TMT reagents from different channels generated a unique reporter ion in the tandem MS, and the intensities of the reporter ions were used for quantifying peptides. To ensure the reliability of the results, we performed the replicate experiments and obtained a total of 1113 proteins. Further setting the enrichment ratio cutoff at 1.4 and then selecting the proteins which must have been identified by more than 2 unique peptides, we finally nailed down to 403 proteins to be highly confident *S*-2-carboxypropylated protein substrates.

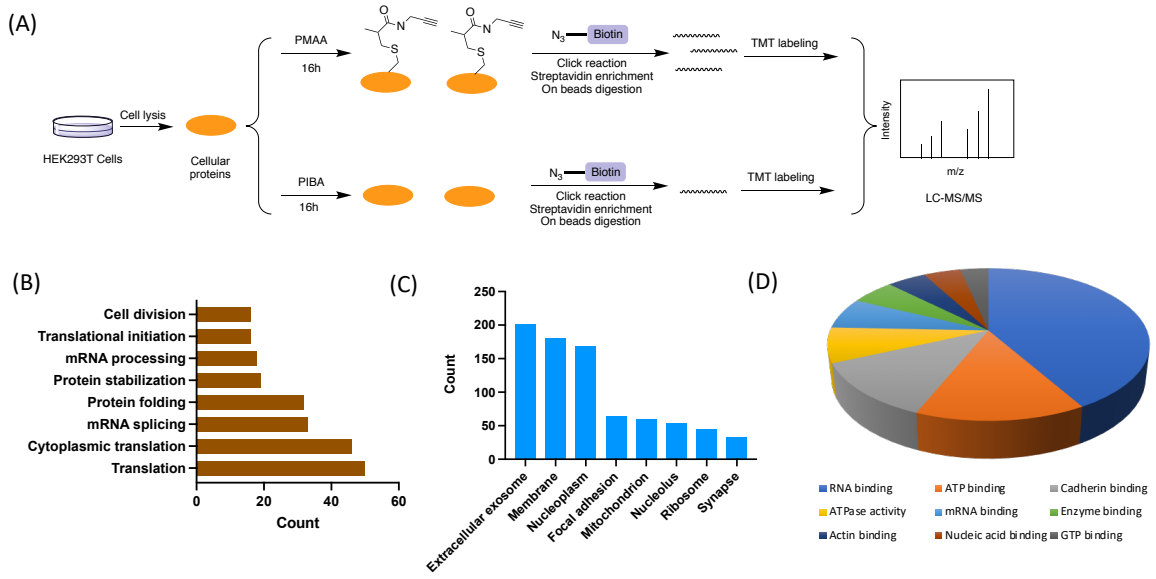


Figure 4.3 Quantitative proteomic profiling of *S*-2-carboxypropylated proteins by PMAA. (A) Schematic description of TMT-based multiplexed proteomic profiling of *S*-2-carboxypropylation in HEK293T cellular proteins by PMAA. (B) Biological process analysis of *S*-2-carboxypropylated proteins by gene ontology. (C) Cellular component analysis of *S*-2-

carboxypropylated proteins in HEK293T cells. (D) Representative molecular function analysis of *S*-2-carboxypropylated proteins.

To understand biological involvements and physiological functions of the identified C2cp proteins, we used the Database for Annotation, Visualization and Integrated Discovery (DAVID) to perform Gene Ontology (GO) analysis of the identified 403 proteins.[111] The results showed that the PMAA labeled proteins are involved in a plethora of biological processes including protein translation, mRNA splicing, protein folding, protein stabilization, mRNA processing, translational initiation and cell division (**Figure 4.3B**). Cellular component analysis showed that majority of the proteins are localized in extracellular exosome, membrane, nucleoplasm, focal adhesion and mitochondrion (**Figure 4.3C**), demonstrating a broad distribution of C2cp-modified proteins. Specifically, 59 proteins were found in the mitochondrion, including ALDH5A1, AIFM1, CS, HSPA9 and ACAT1, suggesting that C2cp may have multiple regulatory impacts on cellular metabolism through modifying mitochondrial proteins. In addition, GO analysis based on molecular function indicated that a large number of the identified proteins mediate molecular bindings of RNA, ATP, cadherin, kinases, ATPases, actin, etc. (**Figure 4.3D**). Collectively, these functional annotation results suggest that protein C2cp modifications have a huge impact on diverse cellular pathways across all the major cellular organelles, which can be a rich source for further exploration of molecular mechanisms of MC-CoA and C2cp regulated biological processes and pathological disorders.

Chemoproteomic profiling of C2cp sites in the cellular proteins

Next, we utilized the PMAA probe to globally identify C2cp sites on cellular proteins. The schematic is illustrated in the flow chart of **Figure 4.4A**: the cultured HEK293T cells were lysed

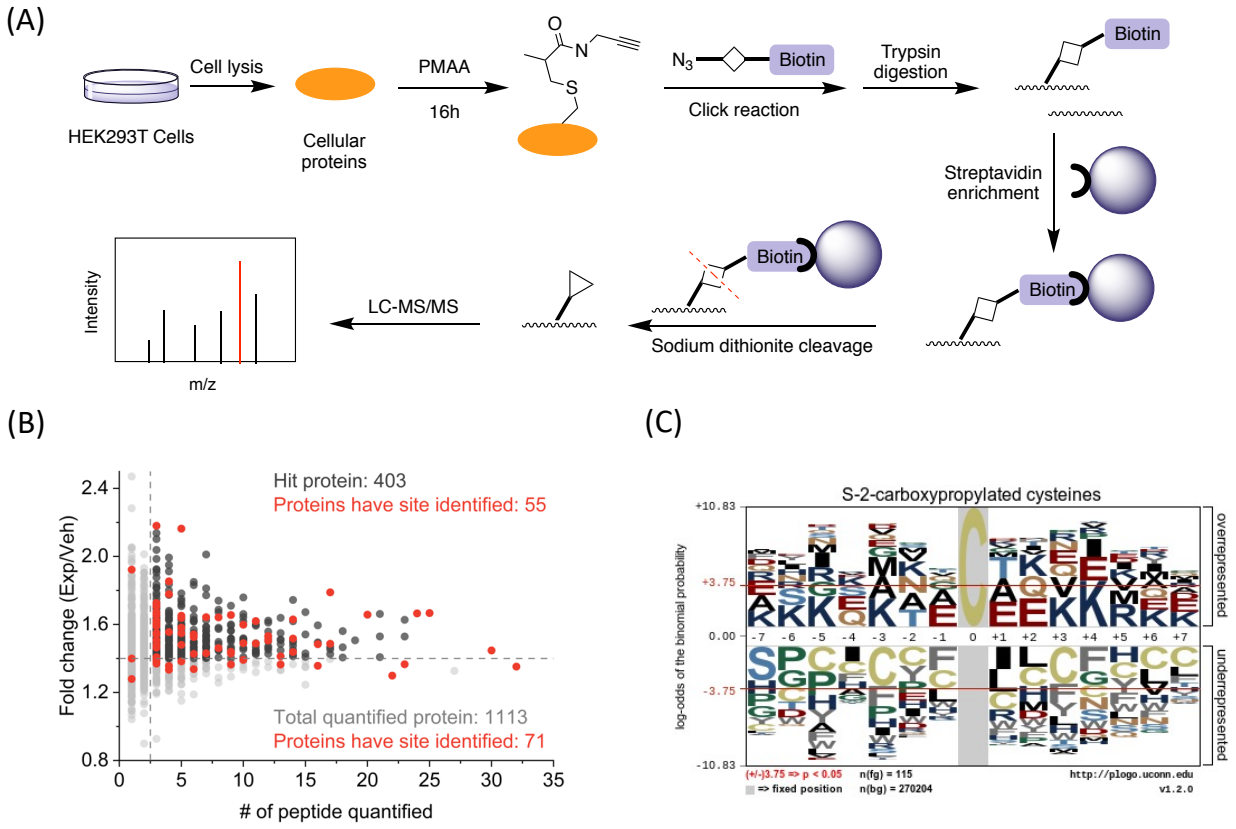


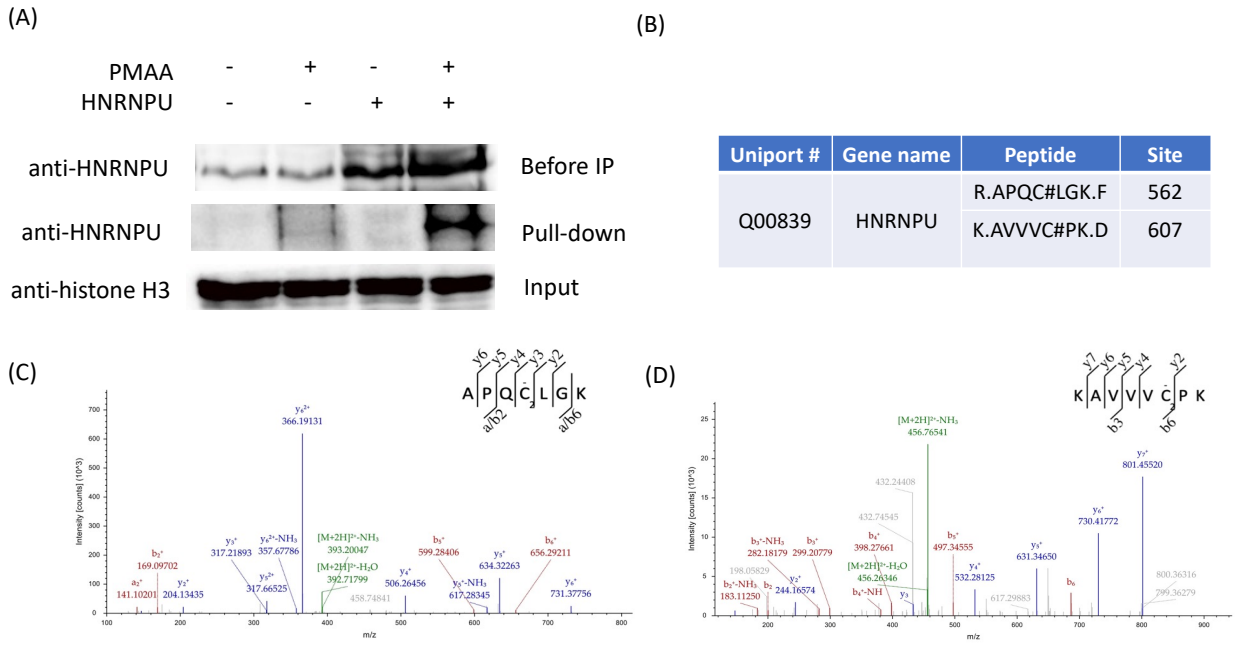
Figure 4.4 Identification of *S*-2-carboxypropylated cysteines by PMAA. (A) The working scheme for LC-MS/MS identification of *S*-2-carboxypropylation sites in HEK293T cells. (B) Identified *S*-2-carboxypropylated proteins and cysteine residues in HEK293T cells. The proteins which have Exp/Veh ratios > 1.4 and must be identified by more than 2 unique peptides are considered as hit proteins. (C) Sequence motif analysis of protein *S*-2-carboxypropylation. Images were generated with pLogo. The red horizontal lines on the pLogo plots denote P = 0.05 thresholds.

and the whole lysate proteins were labeled by PMAA, conjugated with azide-diazo-biotin through CuAAC click reaction, and then subjected to trypsin digestion. The digested peptides were enriched using streptavidin agarose beads and eluted by sodium dithionite as aforementioned, which were then analyzed by LC-MS/MS for site identification. From two biological replicates,

we totally obtained 120 cysteine residues containing C2cp marks. Through further cross-checking with the above mentioned 403 highly-confident proteins, we found 55 of the proteins having C2cp sites identified (**Figure 4.4B**). We also found 8 proteins containing more than 1 site, including G3P, RS2, FLNA, STIP1, RS20, HNRNPU, TBB3 and FUBP2. To explore the structural features for C2cp, we analyzed the flanking amino acid sequences of the identified *S*-2-carboxypropylated cysteine residues by pLogo algorithm.[163] Specifically, we found positively charged lysine was significantly overrepresented at the -7, -6, -5, -3, +3, +4, +6 and +7 positions of *S*-2-carboxypropylation sites, whereas negatively charged glutamate was significantly overrepresented at the -1, +1 and +2 positions ($p < 0.05$, **Figure 4.4C**). In addition, glutamine, threonine and arginine were highly enriched at the -4, -2 and +5 positions, respectively. The deprotonated Glu might be possible to reduce the adjacent cysteine pKa through hydrogen bond formation, whereas a nearby arginine could elevate the cysteine pKa.[164] By contrast, we found Cys was generally underrepresented in the flanking sequences of the *S*-2-carboxypropylated sites. Therefore, the cysteine reactivity is greatly influenced by the local surrounding amino acids. It is noted that cysteine modification sites were also found in 40 proteins which were not identified through globally identification of the PMAA labeled proteins by MS-based proteomics. These likely are additional C2cp substrates.

Validation of C2cp modification on a select protein

To corroborate *S*-2-carboxypropylated proteins identified by our proteomic analysis, we experimentally validated C2cp on a selected protein from our proteomic data list. It has been found that the dysfunction of Heterogeneous Nuclear Ribonucleoprotein U (HNRNPU) is related to neurodevelopmental syndrome and excess MC-CoA can also cause neurological disorder.[156, 165] Therefore, we selected HNRNPU as a protein candidate for biochemical validation. Specifically,



Uniport #	Gene name	Peptide	Site
Q00839	HNRNPU	R.APQC#LGK.F	562
		K.AVVVC#PK.D	607

Figure 4.5 Validation of cysteine *S*-2-carboxypropylation on the selected protein, HNRNPU. (A) Verification of the PMAA labeled protein (HNRNPU) by western blot. HEK293T cell lysate proteins with or without HNRNPU overexpression were incubated with 2 mM PMAA for 12 h at 37 ℃, and then the proteins were conjugated with azide-diazo-biotin through click reaction. After pull-down by streptavidin-beads, the eluted proteins were further detected by western blot. (B) Identified C2cp sites on the protein HNRNPU by PMAA probe. (C) MS/MS peptide spectrum shows the C2cp modification on the C562 residue on HNRNPU. (D) MS/MS peptide spectrum shows the C2cp modification on the C607 residue on HNRNPU.

the HNRNPU plasmid was transiently transfected into HEK293T cells and then the lysate protein mixture was labeled by PMAA, conjugated with azide-diazo-biotin, and enriched by streptavidin beads. Thereafter, the labeled proteins were eluted by sodium dithionite. Enhanced HNRNPU intensity by western blotting detection was observed when both HNRNPU overexpression and PMAA labeling were applied, which demonstrates that the intracellularly expressed protein

HNRNPU indeed was labeled by PMAA and then be pull-down by streptavidin beads (**Figure 4.5A**). Furthermore, we sought to confirm the identified C2cp sites on HNRNPU and investigated whether the modification was induced by Met-Na or MC-CoA. Based on our site-specific proteomic data, we identified two modification sites on HNRNPU by the probe PMAA, which were Cys562 and Cys607 (**Figure 4.5B**). We incubated the HNRNPU-overexpressed cell lysate proteins with either 1mM Met-Na or 1mM MC-CoA and then enriched the HNRNPU with anti-HNRNPU antibody. The obtained proteins were digested and further analyzed through LC-MS/MS. Using this approach, we successfully confirmed the two modification sites Cys562 and Cys607 on HNRNPU with the treatment of MC-CoA (**Figure 4.5C and 4.5D**). On the other hand, with Met-Na incubation, we could not identify any C2cp sites, which coincides with the above observation that methacrylate is less reactive than MC-CoA. These results demonstrate that PMAA is an excellent probe for profiling cellular protein C2cp, and the metabolite MC-CoA but not Met-Na was able to induce protein C2cp modifications.

4.3 Conclusion

In this work, we have developed a highly efficient bioorthogonal chemical probe, PMAA, for identification and profiling of *S*-2-carboxypropylated proteins in mammalian cells through chemical labeling, affinity enrichment, and proteomic identification. Using the probe, we showed that mammalian cellular proteins were extensively *S*-2-carboxypropylated and the modification was present in different cell lines such as 36T, HCT116 and HeLa cells. Using multiplexed quantitative proteomic analysis and site-specific profiling, we totally identified 403 *S*-2-carboxypropylated proteins and 120 cysteine modification sites from HEK293T cells. To the best of our knowledge, this is the first direct demonstration and profiling of protein cysteine *S*-2-carboxypropylation on mammalian proteins. Previously, using itaconate-alkyne, Qin et al.

identified 1926 protein itaconated protein targets [58], whereas less number of identified protein targets for *S*-2-carboxypropylated proteins in our experiment. This result as well as different protein lists suggests the different specificity for the two modifications and C2cp modification is likely more specific on cellular proteins compared with itaconation. Furthermore, we experimentally validated cysteine *S*-2-carboxypropylation on the protein HNRNPU and confirmed the modification sites identified from the chemoproteomic profiling with the PMAA probe. Interestingly, we found that the C2cp mark was far more effectively induced by the metabolite MC-CoA rather than Met-Na, which demonstrates that MC-CoA is a more reactive metabolite than Met-Na in protein modification. Since MC-CoA is a key metabolic intermediate in valine degradation, C2cp modification in proteins represents a new regulatory mechanism for valine associated metabolic diseases. Further studies are needed to investigate detailed mechanisms for the influence of the C2cp modification on valine-pertinent disease pathways. Overall, our work detected cysteine *S*-2-carboxypropylation in mammalian proteins and mapped out its proteomic distribution in the proteome. The PMAA chemical probe we developed here would serve as a useful tool for studying other cysteine-participating protein modifications.

4.4 Methods and materials

Reagents and materials

Unless otherwise noted, all chemicals and solvents were purchased from Sigma–Aldrich and used without further purification. Click chemistry reagents biotin-azide (catalog# 1265) and azide-diazo-biotin (catalog# 1041) were purchased from Click Chemistry Tools. Protein dual color standards and precast protein gels (4%–20%) were purchased from Bio-Rad Laboratories. High sensitivity streptavidin-HRP was purchased from Thermo Fisher Scientific (catalog# 21130). High-capacity streptavidin agarose was purchased from Thermo Fisher Scientific (catalog#

20359). Plasmid hnRNPU (pcDNA3.1-hnRNPU-V5) was a gift from Susana Valente (Addgene plasmid # 35974; http://n2t.net/addgene:35974; RRID: Addgene_35974).[166] The anti-hnRNPU antibody (catalog# sc-32315) was purchased from Santa Cruz. The anti-Histone H3 antibody (catalog# sc-517576) was purchased from Santa Cruz. The anti-mouse-HRP linked antibody (catalog# 7076) was purchased from Cell Signaling. HEK293T, Hela and HCT 116 cells were purchased from American Type Culture Collection (ATCC). MEF cell line (36T cells) was a gift from Dr. Mark Parthun's lab.

Synthesis of PMAA, PIBA and MC-CoA

The probe N-propargyl methacrylamide (PMAA) was synthesized as previously described.[161] The probe N-propargyl isobutyrylamide (PIBA) was synthesized following the previous protocol.[162] Methacrylyl-CoA (MC-CoA) was synthesized as described in our previous publication.[43]

MALDI-mass analysis of the reactions between MC-CoA and thiol-containing biological compounds

1mM MC-CoA was used to react with 1 mM cysteine, cysteamine, or coenzyme A at 37 ℃ for 4 h in PBS, respectively. The resulted samples were sent for MALDI mass detection.

The same condition was used for the reaction between MC-CoA and cysteine-containing peptide, glutathione.

Cell culture and preparation of whole cellular protein

For HEK293T, Hela, 36T, HCT116 cell culture, cells were grown in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Life Technologies) and 1% penicillin-streptomycin (15140122, Thermo Fisher Scientific) at 37 ℃ in a humidified cell incubator with 5% CO_2.

For preparation of the whole cellular proteins, the cells were plated and cultured to around 90% confluence and then scrapped and harvested after washing through PBS for three times. Cells were collected through centrifugation (5000 rpm x 5 min, 4 ℃) and were resuspended by ice-cold M-PER mammalian protein extraction reagent (ThermoFisher SCIENTIFIC, Product #78501) containing 1% protease inhibitor cocktail (ThermoFisher SCIENTIFIC, Product #78438). After putting the lysate on ice for 15 minutes, to release the whole cellular proteins, cells were sonicated with 30% amplitude. The cellular proteins (supernatant) were collected by centrifugation for 20 minutes (13200 rpm x 20 min, 4 ℃). The protein concentration was determined by Bradford assay.

Protein labeling by the probe PMAA, CuAAC reaction, and streptavidin detection

The methods for labeling, CuAAC reaction and streptavidin detection were adapted from the previous experimental procedures.[129] Specifically, 50 μg cellular proteins were incubated with the probe PMAA as indicated concentrations at 37 ℃ for 12 h. The whole proteins were then precipitated with 5-fold ice-cold acetone and the precipitated proteins were collected by centrifugation at 5000 g for 10 min at 4 ℃, washed by ice-cold methanol for three times. The resulting proteins were resuspended by PBS with 0.4% SDS. The protein solutions were then subjected to CuAAC reaction. For CuAAC reaction, the proteins were mixed with click cocktail containing 50 μ M biotin-azide, 2.5 mM sodium ascorbate, 0.5 mM copper sulfate and 0.25 mM ligand BTTAA. After 1 h incubation at room temperature, the solutions were further mixed with protein loading dye and then boiled for 5 min. The samples were then loaded on 4-20% SDS-PAGE gradient gel. Thereafter, the proteins were transferred to nitrocellulose membrane and blocked with 5% non-fat milk for 1 h. The membrane was incubated with streptavidin-HRP for another 1 h and scanned by chemiluminescence scanner.

Protein enrichment and protein digestion for protein identification

Lyophilized lysates were transferred to new tubes and resuspended in 100 mM DPBS. The streptavidin beads were washed in 100 mM DPBS twice and incubated with cell lysates for 2 h at room temperature. The beads were then washed by 100 mM DPBS ten times in a spin column (Thermo) to remove unspecific binding proteins.

The beads were resuspended in the digestion buffer containing 50 mM HEPES pH=8.1, 1.6 M urea, and enriched proteins were digested with 50:1 (w/w) trypsin (Promega) for 16 h at 37 ˚C with shaking. The digestion was quenched by adding trifluoroacetic acid (TFA) to a final concentration of 0.4% after removing beads with spin columns. Digests were centrifuged and desalted using 50 mg SepPak tC18 cartridges (Waters).

Protein digestion and peptide enrichment for sites identification

Lyophilized lysates were transferred to new tubes and resuspended in the digestion buffer containing 50 mM HEPES pH=8.1, 1.6 M urea, and proteins were digested with 50:1 (w/w) trypsin (Promega) for 16 h at 37 ˚C with shaking. The digestion was quenched by adding trifluoroacetic acid (TFA) to a final concentration of 0.4%. Digests were centrifuged and desalted using 50 mg SepPak tC18 cartridges (Waters) before lyophilization.

For each sample, 50 µL Neutravidin beads (Thermo) were used to enrich biotinylated peptides. Beads were washed 3 times by 100 mM DPBS, then incubated with peptides dissolved in 500 µL 100 mM DPBS for 2 h. Unbounded peptides were removed through 10 times washes with 100 mM DPBS, and tagged peptides were eluted in 25 mM sodium dithionite solution for 30 min twice. Two elutes were combined and desalted as mentioned above.

TMT labeling and peptide fractionation by HPLC

Purified peptides were lyophilized and resuspended in 100 µL of 100 mM HEPES pH = 8.5, and 30 µL ACN. The samples were labeled with the TMT6plex reagent (Thermo). After a

one-hour reaction, the excessive amount of the TMT reagents was quenched by adding 10 µL of 5% hydroxylamine. Labeled peptides were pooled, purified, and fractionated into three fractions using 20/50/80% ACN. The collected fractions were further purified by the StageTip method.

LC-MS/MS analysis for proteomic identification

Dried peptides were resuspended in 6 µL loading solution containing 5% ACN and 4% formic acid (FA), and 3 µl was loaded onto a microcapillary column packed with C18 beads (Magic C18AQ, 5 µm, 200 Å, 75 µm x 16 cm) by a Dionex WPS-3000TPLRS autosampler (UltiMate 3000 Thermostatted Pulled Loop Rapid Separation Wellplate Sampler). Peptides were separated by reversed-phase HPLC using an UltiMate 3000 binary pump with a 150-minute gradient designed for the TMT-labeled peptides. The full MS and MS2 were detected in a hybrid dual-cell quadrupole linear ion trap – Orbitrap mass spectrometer (LTQ Orbitrap Elite, Thermo Scientific, with Xcalibur 3.0.63 software) using a data-dependent Top15 method. Each cycle included one full MS scan (resolution: 60,000) in the Orbitrap at the automatic gain control (AGC) target of 1*106, followed by up to 15 MS/MS for the most intense ions. Selected ions were excluded from further analysis for 90 s. Ions with a single or unassigned charge were not sequenced. MS2 scans were activated by HCD at 30% normalized collision energy with an isolation width of 1.2 m/z and detected in the Orbitrap cell with a resolution of 15,000.

Database searching, data filtering, and protein quantification

Raw data files recorded from the mass spectrometer were converted into the mzXML format. Mass spectra were searched using the SEQUEST algorithm (version 28) (Eng et al., 1994) against a human proteome (Homo sapiens) database encompassing sequences of all proteins downloaded from UniProt (https://www.uniprot.org/taxonomy/9606). Each protein sequence was listed in both forward and reversed orientations to estimate the false discovery rate (FDR) of

peptide and protein identifications. The following parameters were used for the search: 10 ppm precursor mass tolerance; 0.025 Da product ion mass tolerance; fully digested with trypsin; up to two missed cleavages; variable modifications: modification of cysteine (+301.153874); fixed modifications: TMT labeling of lysine and the N-termini (+229.1630).

The target-decoy method (Elias and Gygi, 2007) was employed to evaluate and control the FDRs of peptide and protein identifications. Linear discriminant analysis (LDA) was used to distinguish correct and incorrect peptide identifications using numerous parameters such as XCorr, ΔCn, and precursor mass error. After scoring, peptides with fewer than seven amino acids in length were discarded, and peptides were filtered to a less than 1% FDR based on the number of decoy sequences in the final data set.

The TMT reporter ion intensities in the MS2 spectra were used to quantify peptides. The isotopic information provided by Thermo was utilized to correct the ion intensities. The summed intensity for each protein was calculated based on all unique peptides from this protein in each sample.

In order to evaluate the site localization confidence, a ModScore was calculated for each labeled cysteine site[167], which applies a probabilistic algorithm that considers all cysteines in a peptide and uses the presence of MS/MS fragment ions unique to each cysteine site to evaluate the confidence of localization when the best site match is compared with the next best one. Sites with ModScore > 13 ($p < 0.05$) were considered to be confidently localized. For quantification, only peptides with at least one confidently localized site were used. Because cysteine is a relatively rare amino acid, most peptides have only one site identified.

Bioinformatic analysis

Biological process, cellular component, and molecular function analysis of the identified *S*-2-carboxypropylated proteins were carried out through DAVID bioinformatics.[111] Sequence preference motif was generated by pLogo algorithm.[163]

Western blotting analysis

The protein HNRNPU and histone H3 were detected by antibody. The enriched sample or input sample were subjected to SDS-PAGE and transferred to nitrocellulose membrane followed by blocking with 5% non-fat milk for 1 h. Next, the membrane was incubated either with the primary anti-HNRNPU antibody or anti-Histone H3 antibody at 4 ℃ overnight. After washing with TBST (10 mM Tris, pH 8.0, 150 mM NaCl, 0.5% Tween 20) for 3 times (5 minutes for each), the membranes were then incubated with secondary antibody at room temperature for 1 h. The membranes were then washed for three times (5 minutes for each) with TBST and developed by enhanced chemiluminescence detection system according to the manufacturer's protocols and scanned by chemiluminescence scanner.

Transfection, PMAA labeling and enrichment of the protein HNRNPU

HEK293T cells were plated and cultured until 70% confluence and the transient transfection of the plasmid hnRNPU (pcDNA3.1-hnRNPU-V5, plasmid# 35974, addgene) was conducted by Lipofectamine 3000 (ThermoFisher SCIENTIFIC, Product #L3000008) according to the manufacture's protocol followed by 48 h of incubation. The whole cellular proteins were prepared following the aforementioned procedure.

PMAA labeling of the protein HNRNPU followed the above-described protocol. Briefly, 100 μg cellular proteins were incubated with 2 mM PMAA at 37 ℃ for 12 h. The whole proteins were then precipitated with 5-fold ice-cold acetone and the precipitated proteins were collected by centrifugation at 5000 g for 10 min at 4 ℃, washed by ice-cold methanol for three times. The

resulting proteins were resuspended by PBS with 0.4% SDS. The protein solutions were then subjected to CuAAC reaction. For CuAAC reaction, the proteins were mixed with click cocktail containing 50 μ M azide-diazo-biotin, 2.5 mM sodium ascorbate, 0.5 mM copper sulfate and 0.25 mM ligand BTTAA. After 1 h incubation at room temperature, the proteins were further precipitated to remove excess azide-diazo-biotin and then redissolved in PBS with 0.2% SDS.

High-capacity streptavidin agarose was first equilibrated by PBS and then incubated with the protein sample for 1 h with gentle agitation. Next, the resins were collected by centrifugation at 5000g for 5 minutes and washed with PBS supplemented with 0.2% (w/v) SDS, PBS supplemented with 0.1% (w/v) of SDS and 6M urea and 50 mM of NH_4HCO_3 supplemented with 0.1% of SDS. Then, the resins were incubated with 25 mM of $Na_2S_2O_4$ in 50 mM NH_4HCO_3 for 1 h to elute the protein off the beads. The mixture was then centrifuged at 5000g for 5 minutes and the supernatant was collected and dried by SpeedVac and redissolved in 100 μL SDS loading buffer. The solution was then subjected to SDS-PAGE running and western blotting analysis.

Protein reaction and Immunoprecipitation of the protein HNRNPU

HNRNPU overexpression and the whole cellular protein extraction were following the aforementioned procedures. 100 μg HNRNPU overexpressed cellular proteins were incubated with either 1 mM sodium methacrylate or 1mM MC-CoA at 37 ℃ for 12 h. The resulted samples were subjected to immunoprecipitation for HNRNPU enrichment by Protein G Plus-agarose (Santa Cruz Biotechnology, Product# sc-2002) following the standard protocol. The enriched proteins were separated by SDS-PAGE and the gel bands were cut for LC-MS/MS analysis.

In gel digestion, LC-MS/MS analysis of the modification sites on the protein HNRNPU

The gel bands were sliced into small pieces, and then rinsed with 50% acetonitrile/20 mM ammonium bicarbonate (~pH 7.5-8) twice. The gel pieces were dehydrated by adding 100% of

acetonitrile and dried out by a SpeedVac. A various amount of trypsin solution (0.01 μg/μL in 20 mM ammonium bicarbonate) was added until the gel pieces totally absorb the trypsin solution. The tubes were placed in an incubator at 37°C overnight. The tryptic peptides were extracted from gel pieces by incubating with 50% acetonitrile/0.1% formic acid twice. The extracts were dried down by a SpeedVac.

The mass spectrometry analyses were performed on a Thermo-Fisher LTQ Orbitrap Elite Mass Spectrometer coupled with a Proxeon Easy NanoLC system (Waltham, MA). The enzymatic peptides were loaded into a reversed-phase column (self-packed column/emitter with 200 Å 5 μM Bruker MagicAQ C18 resin), then directly eluted into the mass spectrometer. Briefly, the two-buffer gradient elution at the flow rate 450 nL/min (0.1% formic acid as buffer A and 99.9% acetonitrile with .0.1% formic acid as buffer B) starts with 0% B, holds at 0%B for 2 minutes, then increases to 30% B in 50 minutes, to 50% B in 10 minutes, and to 95% B in 10 minutes. The data-dependent acquisition (DDA) method was used to acquire MS data. A survey MS scan was acquired first, and then the top 10 ions in the MS scan were selected for following CID MS/MS analysis. Both MS and MS/MS scans were acquired by Orbitrap at the resolutions of 120,000 and 15,000, respectively. Data were acquired using Xcalibur software (version 2.2, Thermo Fisher Scientific). Proteins identification and modification characterization were performed using Thermo Proteome Discoverer (version 1.4) with Mascot (Matrix Science 2.7) and Uniprot database. The spectra of possible modified peptides were inspected further to verify the accuracy of the assignments.

CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

Protein posttranslational modifications such as lysine acetylation and acylations as well as cysteine modifications are important processing events, which regulate cellular processes by reversibly or non-reversibly adding modifying groups on protein substrates. The modifying groups can change the properties of the proteins and then widely influence the various biological processes such as transcriptional regulation, cell cycle, DNA repair, cell metabolism, signal transduction and inflammatory pathways. [8, 57, 91, 168] Aberrant modifications of cellular proteins correlates with lots of disease states, such as tumorigenesis, cardiovascular diseases, diabetes, inflammation and neurodegenerative diseases.[169-171] Additionally, some modifications, e.g. protein acetylation and acylations, are modulated by their writer proteins such as lysine acetyltransferases (KATs) and the eraser proteins such as lysine deactylases (KDACs). The dysregulation of these regulatory enzymes can cause abnormal state of the modifications on their protein substrates, which can further lead to a variety of diseases as above mentioned. Therefore, the identification of the protein substrates for different protein PTMs and also the protein substrates for the regulatory enzymes, e.g. KATs, is greatly important for understanding the function of the protein PTMs and also KATs, which can further contribute to the development of effective therapeutic methods to treat related disorders. In this work, we sought to identify the protein substrates for one of the most important KATs, HAT1, elucidate the protein substrates for recent identified protein methacrylylation, and uncover a novel PTM, (*S*)-2-carboxypropylation, as well as its protein substrates.

The use of high-resolution tandem mass spectrometry (MS/MS) represents a significant improvement in the investigation of lysine acetylation. This method has enabled researchers to identify numerous sites of lysine acetylation, as well as thousands of protein targets that undergo this modification.[172] However, the information for protein substrates of individual KAT is still limited. The major obstacle that makes the identification of the protein substrates difficult is the chemical inert property of acetyl group. Therefore, to overcome the problem, we synthesized a series of functionalized Ac-CoA reporters, including 3-azidopropanoyl CoA (3AZ-CoA), 4-azidobutanoyl CoA (4AZ-CoA), 4-pentynoyl CoA (4PY-CoA), 5-hexynoyl CoA (5HY-CoA), and 6-heptynoyl CoA (6HY-CoA) to label protein substrates of KATs. We mutated several bulky amino acids surrounding the acetyl group binding sites of HAT1 to smaller ones (i.e. alanine and glycine) by bump-and-hole strategy. Interestingly, we found that one mutant (e.g., HAT1-Y-282A) can accommodate all the functionalized chemical reporters that we generated and has no activities towards Ac-CoA. Additionally, we also found that 3AZ-CoA is a cell-permeable chemical probe which can be used for intracellular protein labeling in living cells. Combine the mutant HAT1 (HAT1Y282A) and 3AZ-CoA for bioorthogonal labeling, we successfully identified more than 100 protein substrates for HAT1 in the native cellular context. Functional annotation of the identified proteins shows that the HAT1 protein substrates are largely involved in various cellular pathways including transcription, translation, RNA splicing, RNA processing, protein folding, oxidation-reduction process, and mitochondrial regulation. The work laid a solid foundation for future investigation of the HAT1 function in the different cellular pathways. Additionally, the cell-permeable bioorthogonal chemical probe, 3AZ-CoA, that we identified here, will expand the researchers' toolbox for future elucidation of the protein substrates for other KATs.

When we worked on the identification of HAT1 protein substrates, our lab together with Zhao's lab discovered a new type of histone modification, lysine methacrylylation.[43] Interestingly, we found that HAT1 is one of the regulatory enzymes that can catalyze histone methacrylylation by transferring the methacrylyl group form MC-CoA to the histone proteins. However, the information still lacks regarding the methacrylylated proteins. Specifically, there is no evidence to show whether non-histone protein can be methacrylylated and also whether HAT1 can catalyze protein methacrylylation on non-histone proteins. And the antibody is hard to efficiently differentiate the two modifications. To solve the problems, we designed and synthesized a selective chemical probe gly-TCEP-AK targeted to methacrylylated protein but not its isomer, crotonylated proteins. The probe gained great selectivity on both peptides level and also complex cellular protein level. By using the probe, we found that there are a lot of non-histone proteins that could be labeled and imaged, which demonstrated that there are various non-histone methacrylylated proteins in cells. Since the valine is the intracellular source of MC-CoA, we also examined the influence of valine metabolism towards protein methacrylylation. Specifically, we found that valine treatment could also induce cellular protein methacrylylation and knockdown of a key regulatory enzyme, HIBCH, in valine metabolic pathway led to increased methacrylylation on several non-histone proteins. LC-MS/MS identification revealed 254 sodium methacrylate dependent methacrylylated proteins and 256 valine dependent methacrylylated proteins. Functional survey of the proteins showed that both of the sodium methacrylate and valine dependent proteins are involved in various biological processes but valine dependent proteins have more influence in mitochondrial effect. Furthermore, we found that HAT1 could also induce protein methacrylylation on multiple non-histone proteins. And proteomic profiling analysis showed that more than 1000 proteins are prominent targets for HAT1-dependent methacrylylated

substrates. GO analysis showed that the proteins are largely involved in multiple biological processes including protein transport, transcription, and cell proliferation. These findings suggested that methacrylylated proteins exist on multiple cellular proteins, which may regulate cellular functions in a broader aspect than sole chromatin. It is also found that HAT1 carried out protein methacrylylation on a wide range of cellular proteins beyond histones, suggesting that HAT1 could also regulate cellular proteins through protein methacrylylation. However, future studies should focus on the detailed mechanism of the influence of methacrylylation towards individual protein substrate. Our chemical probe, gly-TCEP-AK, can serve as a powerful tool to aid in further analysis.

In the course of our investigation of the substrates' specificity of protein lysine methacrylylation, we found that the double bond on MC-CoA has strong electrophilicity and could potentially conjugate with nucleophiles in a mild condition. Given that MC-CoA is an important metabolite in valine metabolic pathway in mitochondrion, we proposed that the MC-CoA may conjugate with cysteine on cellular proteins and form an unexplored protein modification, *S*-2-carboxypropylation (C2cp). As an unexplored protein PTM, no antibody is available for the detection. To tackle this challenge, we synthesized the probe PMAA, and demonstrated that the probe can be used to profile C2cp modification on complex cellular proteins. Through proteomic analysis, we, for the first time, identified 403 protein targets as well as 120 cysteine residues for the C2cp modification. Bioinformatic analysis showed that the proteins are involved in a variety of biological processes including protein translation, mRNA splicing, protein stabilization and cell division. These important findings demonstrated that C2cp modification is also an important protein PTM, which can regulate cellular functions in a broad range. Moreover, we also validated the C2cp modification on the protein HNRNPU, which was induced by MC-CoA. The dysfunction

of the HNRNPU has been found to be related to neurological disorders. Therefore, this finding could provide an alternative mechanism for further study of the HNRNPU-related disorders. Although the C2cp modification has been detected, it is still unknown how it impacts the cellular functions and what the detailed mechanism is. In the future, it would be valuable to examine the C2cp modification in different cellular context, especially in cancer cells. it would be also valuable to determine which modification sites are more important in protein functional regulation.

In summary, we have generated various engineered HAT1 forms and identified a robust cell permeable bioorthogonal reporter, 3AZ-CoA. We found one mutant, HAT1Y282A, can efficiently label the protein substrates combined with 3AZ-CoA. The subsequent proteomic analysis revealed over 100 proteins for HAT1. We also developed a bioorthogonal chemical probe, gly-TCEP-AK, which can be used to selectively label cellular methacrylylated proteins and identified hundreds of methacrylylated proteins. Moreover, we demonstrated the metabolite MC-CoA can form *S*-2-carboxypropylation as a new protein PTM on multiple cellular proteins. Together, this work expands our understanding of protein PTMs in cellular functional regulation and provides valuable tools for further mechanism investigation.

REFERENCES

(1) Selvi, R. B.; Kundu, T. K. Reversible acetylation of chromatin: implication in regulation of gene expression, disease and therapeutics. *Biotechnol. J.* **2009**, *4* (3), 375-390, Research Support, Non-U.S. Gov't Review. DOI: 10.1002/biot.200900032.

(2) Smith, K. T.; Workman, J. L. Introducing the acetylome. *Nat. Biotechnol.* **2009**, *27* (10), 917-919. DOI: nbt1009-917 [pii]10.1038/nbt1009-917.

(3) Strahl, B. D.; Allis, C. D. The language of covalent histone modifications. *Nature* **2000**, *403* (6765), 41-45.

(4) Yang, X. J.; Seto, E. Lysine acetylation: codified crosstalk with other posttranslational modifications. *Mol. Cell* **2008**, *31* (4), 449-461. DOI: S1097-2765(08)00457-7 [pii] 10.1016/j.molcel.2008.07.002.

(5) Glozak, M. A.; Sengupta, N.; Zhang, X. H.; Seto, E. Acetylation and deacetylation of non-histone proteins. *Gene* **2005**, *363*, 15-23. DOI: 10.1016/j.gene.2005.09.010.

(6) Spange, S.; Wagner, T.; Heinzel, T.; Kramer, O. H. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int. J. Biochem. Cell Biol.* **2009**, *41* (1), 185-198. DOI: S1357-2725(08)00347-6 [pii] 10.1016/j.biocel.2008.08.027.

(7) Voss, A. K.; Thomas, T. Histone Lysine and Genomic Targets of Histone Acetyltransferases in Mammals. *BioEssays* **2018**, *40* (10), e1800078. DOI: 10.1002/bies.201800078.

(8) Ali, I.; Conrad, R. J.; Verdin, E.; Ott, M. Lysine Acetylation Goes Global: From Epigenetics to Metabolism and Therapeutics. *Chem. Rev.* **2018**, *118* (3), 1216-1252. DOI: 10.1021/acs.chemrev.7b00181.

(9) Marmorstein, R.; Trievel, R. C. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim. Biophys. Acta* **2009**, *1789* (1), 58-68. DOI: S1874-9399(08)00157-0 [pii] 10.1016/j.bbagrm.2008.07.009.

(10) Vetting, M. W.; LP, S. d. C.; Yu, M.; Hegde, S. S.; Magnet, S.; Roderick, S. L.; Blanchard, J. S. Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* **2005**, *433* (1), 212-226. DOI: S0003-9861(04)00503-X [pii] 10.1016/j.abb.2004.09.003.

(11) Lee, K. K.; Workman, J. L. Histone acetyltransferase complexes: one size doesn't fit all. *Nat. Rev. Mol. Cell Biol.* **2007**, *8* (4), 284-295.

(12) Yang, X. J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res.* **2004**, *32* (3), 959-976.

(13) Berndsen, C. E.; Denu, J. M. Catalysis and substrate selection by histone/protein lysine acetyltransferases. *Curr. Opin. Struct. Biol.* **2008**, *18* (6), 682-689.

(14) Steunou, A.-L.; Rossetto, D.; Côté, J. Regulating chromatin by histone acetylation. In *Fundamentals of chromatin*, Springer, 2014; pp 147-212.

(15) Audia, J. E.; Campbell, R. M. Histone Modifications and Cancer. *Cold Spring Harb Perspect Biol* **2016**, *8* (4), a019521. DOI: 10.1101/cshperspect.a019521.

(16) He, M.; Han, Z.; Liu, L.; Zheng, Y. G. Chemical Biology Approaches for Investigating the Functions of Lysine Acetyltransferases. *Angew Chem Int Ed Engl* **2018**, *57* (5), 1162-1184. DOI: 10.1002/anie.201704745 From NLM Medline.

(17) Montgomery, D. C.; Sorum, A. W.; Meier, J. L. Defining the orphan functions of lysine acetyltransferases. *ACS Chem. Biol.* **2015**, *10* (1), 85-94. DOI: 10.1021/cb500853p.

(18) Zheng, Y. G.; Wu, J.; Chen, Z.; Goodman, M. Chemical regulation of epigenetic modifications: Opportunities for new cancer therapy. *Med. Res. Rev.* **2008**, *28* (5), 645-687.

(19) Di Cerbo, V.; Schneider, R. Cancers with wrong HATs: the impact of acetylation. *Brief. Funct. Genomics* **2013**, *12* (3), 231-243, Research Support, Non-U.S. Gov't Review. DOI: 10.1093/bfgp/els065.

(20) Zheng, X.; Gai, X.; Ding, F.; Lu, Z.; Tu, K.; Yao, Y.; Liu, Q. Histone acetyltransferase PCAF up-regulated cell apoptosis in hepatocellular carcinoma via acetylating histone H4 and inactivating AKT signaling. *Mol Cancer* **2013**, *12* (1), 96. DOI: 10.1186/1476-4598-12-96 From NLM Medline.

(21) Yin, Y. W.; Jin, H. J.; Zhao, W.; Gao, B.; Fang, J.; Wei, J.; Zhang, D. D.; Zhang, J.; Fang, D. The Histone Acetyltransferase GCN5 Expression Is Elevated and Regulated by c-Myc and E2F1 Transcription Factors in Human Colon Cancer. *Gene Expr* **2015**, *16* (4), 187-196. DOI: 10.3727/105221615X14399878166230 From NLM Medline.

(22) Han, N.; Shi, L.; Guo, Q.; Sun, W.; Yu, Y.; Yang, L.; Zhang, X.; Zhang, M. HAT1 induces lung cancer cell apoptosis via up regulating Fas. *Oncotarget* **2017**, *8* (52), 89970-89977. DOI: 10.18632/oncotarget.21205 From NLM PubMed-not-MEDLINE.

(23) Pfister, S.; Rea, S.; Taipale, M.; Mendrzyk, F.; Straub, B.; Ittrich, C.; Thuerigen, O.; Sinn, H. P.; Akhtar, A.; Lichter, P. The histone acetyltransferase hMOF is frequently downregulated in primary breast carcinoma and medulloblastoma and constitutes a biomarker for clinical outcome in medulloblastoma. *Int J Cancer* **2008**, *122* (6), 1207-1213. DOI: 10.1002/ijc.23283 From NLM Medline.

(24) Iyer, N. G.; Ozdag, H.; Caldas, C. p300/CBP and cancer. *Oncogene* **2004**, *23* (24), 4225-4231. DOI: 10.1038/sj.onc.1207118 From NLM Medline.

(25) Ishihama, K.; Yamakawa, M.; Semba, S.; Takeda, H.; Kawata, S.; Kimura, S.; Kimura, W. Expression of HDAC1 and CBP/p300 in human colorectal carcinomas. *J Clin Pathol* **2007**, *60* (11), 1205-1210. DOI: 10.1136/jcp.2005.029165 From NLM Medline.

(26) Iyer, N. G.; Chin, S. F.; Ozdag, H.; Daigo, Y.; Hu, D. E.; Cariati, M.; Brindle, K.; Aparicio, S.; Caldas, C. p300 regulates p53-dependent apoptosis after DNA damage in colorectal cancer cells by modulation of PUMA/p21 levels. *Proc Natl Acad Sci U S A* **2004**, *101* (19), 7386-7391. DOI: 10.1073/pnas.0401002101 From NLM Medline.

(27) Farazi, T. A.; Waksman, G.; Gordon, J. I. The biology and enzymology of protein N-myristoylation. *J Biol Chem* **2001**, *276* (43), 39501-39504. DOI: 10.1074/jbc.R100042200.

(28) Linder, M. E.; Deschenes, R. J. Palmitoylation: policing protein stability and traffic. *Nat Rev Mol Cell Biol* **2007**, *8* (1), 74-84. DOI: 10.1038/nrm2084.

(29) Jiang, H.; Zhang, X.; Chen, X.; Aramsangtienchai, P.; Tong, Z.; Lin, H. Protein Lipidation: Occurrence, Mechanisms, Biological Functions, and Enabling Technologies. *Chem. Rev.* **2018**, *118* (3), 919-988. DOI: 10.1021/acs.chemrev.6b00750.

(30) Spange, S.; Wagner, T.; Heinzel, T.; Kramer, O. H. Acetylation of non-histone proteins modulates cellular signalling at multiple levels. *Int. J. Biochem. Cell Biol.* **2009**, *41* (1), 185-198. DOI: S1357-2725(08)00347-6 [pii]10.1016/j.biocel.2008.08.027.

(31) Zhang, K.; Tian, S.; Fan, E. Protein lysine acetylation analysis: current MS-based proteomic technologies. *Analyst* **2013**, *138* (6), 1628-1636, Research Support, Non-U.S. Gov't Review. DOI: 10.1039/c3an36837h.

(32) Li, T.; Du, Y.; Wang, L.; Huang, L.; Li, W.; Lu, M.; Zhang, X.; Zhu, W. G. Characterization and prediction of lysine (k)-acetyl-transferase specific acetylation sites. *Mol. Cell. Proteomics* **2012**, *11* (1), M111.011080. DOI: 10.1074/mcp.M111.011080.

(33) Zhang, J.; Sprung, R.; Pei, J.; Tan, X.; Kim, S.; Zhu, H.; Liu, C. F.; Grishin, N. V.; Zhao, Y. Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol. Cell. Proteomics* **2009**, *8* (2), 215-225, Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: 10.1074/mcp.M800187-MCP200.

(34) Basu, A.; Rose, K. L.; Zhang, J.; Beavis, R. C.; Ueberheide, B.; Garcia, B. A.; Chait, B.; Zhao, Y.; Hunt, D. F.; Segal, E.; et al. Proteome-wide prediction of acetylation substrates. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (33), 13785-13790, Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: 10.1073/pnas.0906801106.

(35) Choudhary, C.; Kumar, C.; Gnad, F.; Nielsen, M. L.; Rehman, M.; Walther, T. C.; Olsen, J. V.; Mann, M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **2009**, *325* (5942), 834-840. DOI: 1175371 [pii]10.1126/science.1175371.

(36) Falk, H.; Connor, T.; Yang, H.; Loft, K. J.; Alcindor, J. L.; Nikolakopoulos, G.; Surjadi, R. N.; Bentley, J. D.; Hattarki, M. K.; Dolezal, O.; et al. An efficient high-throughput screening method for MYST family acetyltransferases, a new class of epigenetic drug targets. *J Biomol Screen* **2011**, *16* (10), 1196-1205. DOI: 10.1177/1087057111421631.

(37) Ngo, L.; Wu, J.; Yang, C.; Zheng, Y. G. Effective Quenchers Are Required to Eliminate the Interference of Substrate: Cofactor Binding in the HAT Scintillation Proximity Assay. *Assay Drug Dev. Technol.* **2015**, *13* (4), 210-220. DOI: 10.1089/adt.2015.636.

(38) Luan, Y., Ngo, L., Han, Z., Wang, X., Qu, M., Zheng, Y. G. Histone Acetyltransferases: Enzymes, Assays, and Inhibitors. In *Epigenetic Technological Applications*, Zheng, Y. G. Ed.; Academic Press, 2015; pp 291-317.

(39) Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; et al. Regulation of cellular metabolism by protein lysine acetylation. *Science* **2010**, *327* (5968), 1000-1004. DOI: 327/5968/1000 [pii]10.1126/science.1179689.

(40) Wang, Q.; Zhang, Y.; Yang, C.; Xiong, H.; Lin, Y.; Yao, J.; Li, H.; Xie, L.; Zhao, W.; Yao, Y.; et al. Acetylation of metabolic enzymes coordinates carbon source utilization and metabolic flux. *Science* **2010**, *327* (5968), 1004-1007. DOI: 327/5968/1004 [pii] 10.1126/science.1179687.

(41) Lu, J. Y.; Lin, Y. Y.; Boeke, J. D.; Zhu, H. Using functional proteome microarrays to study protein lysine acetylation. *Methods Mol. Biol.* **2013**, *981*, 151-165, Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: 10.1007/978-1-62703-305-3_12.

(42) Lin, Y. Y.; Lu, J. Y.; Zhang, J.; Walter, W.; Dang, W.; Wan, J.; Tao, S. C.; Qian, J.; Zhao, Y.; Boeke, J. D.; et al. Protein acetylation microarray reveals that NuA4 controls key metabolic target regulating gluconeogenesis. *Cell* **2009**, *136* (6), 1073-1084. DOI: S0092-8674(09)00081-6 [pii]10.1016/j.cell.2009.01.033.

(43) Delaney, K.; Tan, M.; Zhu, Z.; Gao, J.; Dai, L.; Kim, S.; Ding, J.; He, M.; Halabelian, L.; Yang, L.; et al. Histone lysine methacrylation is a dynamic post-translational modification regulated by HAT1 and SIRT2. *Cell Discov* **2021**, *7* (1), 122. DOI: 10.1038/s41421-021-00344-4.

(44) Zhao, S.; Zhang, X.; Li, H. Beyond histone acetylation-writing and erasing histone acylations. *Curr Opin Struct Biol* **2018**, *53*, 169-177. DOI: 10.1016/j.sbi.2018.10.001 From NLM Medline.

(45) Zhu, Z.; Han, Z.; Halabelian, L.; Yang, X.; Ding, J.; Zhang, N.; Ngo, L.; Song, J.; Zeng, H.; He, M.; et al. Identification of lysine isobutyrylation as a new histone modification mark. *Nucleic Acids Res* **2021**, *49* (1), 177-189. DOI: 10.1093/nar/gkaa1176 From NLM Medline.

(46) Zhang, D.; Tang, Z.; Huang, H.; Zhou, G.; Cui, C.; Weng, Y.; Liu, W.; Kim, S.; Lee, S.; Perez-Neut, M.; et al. Metabolic regulation of gene expression by histone lactylation. *Nature* **2019**, *574* (7779), 575-580. DOI: 10.1038/s41586-019-1678-1 From NLM Medline.

(47) Simic, Z.; Weiwad, M.; Schierhorn, A.; Steegborn, C.; Schutkowski, M. The varepsilon-Amino Group of Protein Lysine Residues Is Highly Susceptible to Nonenzymatic Acylation by Several Physiological Acyl-CoA Thioesters. *Chembiochem* **2015**, *16* (16), 2337-2347. DOI: 10.1002/cbic.201500364 From NLM Medline.

(48) Dancy, B. M.; Cole, P. A. Protein lysine acetylation by p300/CBP. *Chem Rev* **2015**, *115* (6), 2419-2452. DOI: 10.1021/cr500452k From NLM Medline.

(49) Huang, H.; Tang, S.; Ji, M.; Tang, Z.; Shimada, M.; Liu, X.; Qi, S.; Locasale, J. W.; Roeder, R. G.; Zhao, Y.; et al. p300-Mediated Lysine 2-Hydroxyisobutyrylation Regulates Glycolysis. *Mol Cell* **2018**, *70* (4), 663-678 e666. DOI: 10.1016/j.molcel.2018.04.011 From NLM Medline.

(50) Han, X.; Xiang, X.; Yang, H.; Zhang, H.; Liang, S.; Wei, J.; Yu, J. p300-Catalyzed Lysine Crotonylation Promotes the Proliferation, Invasion, and Migration of HeLa Cells via Heterogeneous Nuclear Ribonucleoprotein A1. *Anal Cell Pathol (Amst)* **2020**, *2020*, 5632342. DOI: 10.1155/2020/5632342 From NLM Medline.

(51) Chen, Y.; Sprung, R.; Tang, Y.; Ball, H.; Sangras, B.; Kim, S. C.; Falck, J. R.; Peng, J.; Gu, W.; Zhao, Y. Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol Cell Proteomics* **2007**, *6* (5), 812-819. DOI: 10.1074/mcp.M700021-MCP200 From NLM Medline.

(52) Liu, X.; Wei, W.; Liu, Y.; Yang, X.; Wu, J.; Zhang, Y.; Zhang, Q.; Shi, T.; Du, J. X.; Zhao, Y.; et al. MOF as an evolutionarily conserved histone crotonyltransferase and transcriptional

activation by histone acetyltransferase-deficient and crotonyltransferase-competent CBP/p300. *Cell Discov* **2017**, *3*, 17016. DOI: 10.1038/celldisc.2017.16 From NLM PubMed-not-MEDLINE.

(53) Tan, D.; Wei, W.; Han, Z.; Ren, X.; Yan, C.; Qi, S.; Song, X.; Zheng, Y. G.; Wong, J.; Huang, H. HBO1 catalyzes lysine benzoylation in mammalian cells. *iScience* **2022**, *25* (11), 105443. DOI: 10.1016/j.isci.2022.105443 From NLM PubMed-not-MEDLINE.

(54) Wei, W.; Mao, A.; Tang, B.; Zeng, Q.; Gao, S.; Liu, X.; Lu, L.; Li, W.; Du, J. X.; Li, J.; et al. Large-Scale Identification of Protein Crotonylation Reveals Its Role in Multiple Cellular Functions. *J Proteome Res* **2017**, *16* (4), 1743-1752. DOI: 10.1021/acs.jproteome.7b00012 From NLM Medline.

(55) Liu, K.; Li, F.; Sun, Q.; Lin, N.; Han, H.; You, K.; Tian, F.; Mao, Z.; Li, T.; Tong, T.; et al. p53 beta-hydroxybutyrylation attenuates p53 activity. *Cell Death Dis* **2019**, *10* (3), 243. DOI: 10.1038/s41419-019-1463-y From NLM Medline.

(56) Sun, Y.; Chen, Y.; Peng, T. A bioorthogonal chemical reporter for the detection and identification of protein lactylation. *Chem Sci* **2022**, *13* (20), 6019-6027. DOI: 10.1039/d2sc00918h From NLM PubMed-not-MEDLINE.

(57) Murphy, M. P.; O'Neill, L. A. J. Krebs Cycle Reimagined: The Emerging Roles of Succinate and Itaconate as Signal Transducers. *Cell* **2018**, *174* (4), 780-784. DOI: 10.1016/j.cell.2018.07.030 From NLM Medline.

(58) Qin, W.; Zhang, Y.; Tang, H.; Liu, D.; Chen, Y.; Liu, Y.; Wang, C. Chemoproteomic Profiling of Itaconation by Bioorthogonal Probes in Inflammatory Macrophages. *J Am Chem Soc* **2020**, *142* (25), 10894-10898. DOI: 10.1021/jacs.9b11962 From NLM Medline.

(59) Yang, Y. Y.; Hang, H. C. Chemical approaches for the detection and synthesis of acetylated proteins. *ChemBioChem* **2011**, *12* (2), 314-322, Review. DOI: 10.1002/cbic.201000558.

(60) Berndsen, C. E.; Denu, J. M. Assays for mechanistic investigations of protein/histone acetyltransferases. *Methods* **2005**, *36* (4), 321-331.

(61) Rye, P. T.; Frick, L. E.; Ozbal, C. C.; Lamarr, W. A. Advances in label-free screening approaches for studying histone acetyltransferases. *J. Biomol. Screen.* **2011**, *16* (10), 1186-1195. DOI: 10.1177/1087057111418653.

(62) Ourailidou, M. E.; Zwinderman, M. R.; Dekker, F. J. Bioorthogonal metabolic labelling with acyl-CoA reporters: targeting protein acylation. *MedChemComm* **2016**, *7* (3), 399-408.

(63) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **2009**, *48* (38), 6974-6998.

(64) Thinon, E.; Hang, H. C. Chemical reporters for exploring protein acylation. *Biochem. Soc. Trans.* **2015**, *43* (2), 253-261, Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: 10.1042/BST20150004.

(65) Sterner, R.; Allfrey, V. G. In vitro mercaptoacetylation of chromosomal proteins. Selective recovery of newly modified protein molecules. *J Biol Chem* **1982**, *257* (22), 13872-13876.

(66) Yu, M.; de Carvalho, L. P.; Sun, G.; Blanchard, J. S. Activity-based substrate profiling for Gcn5-related N-acetyltransferases: the use of chloroacetyl-coenzyme A to identify protein substrates. *J. Am. Chem. Soc.* **2006**, *128* (48), 15356-15357. DOI: 10.1021/ja066298w.

(67) Kleff, S.; Andrulis, E. D.; Anderson, C. W.; Sternglanz, R. Identification of a gene encoding a yeast histone H4 acetyltransferase. *J. Biol. Chem.* **1995**, *270* (42), 24674-24677.

(68) Sletten, E. M.; Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem.* **2009**, *48* (38), 6974-6998. DOI: 10.1002/anie.200900942.

(69) Meldal, M.; Tornoe, C. W. Cu-catalyzed azide-alkyne cycloaddition. *Chem. Rev.* **2008**, *108* (8), 2952-3015, Review. DOI: 10.1021/cr0783479.

(70) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regioselective "ligation" of azides and terminal alkynes. *Angew Chem Int Ed Engl* **2002**, *41* (14), 2596-2599. DOI: 10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4.

(71) Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. Bioconjugation by copper(I)-catalyzed azide-alkyne [3 + 2] cycloaddition. *J Am Chem Soc* **2003**, *125* (11), 3192-3193. DOI: 10.1021/ja021381e.

(72) Yang, Y. Y.; Ascano, J. M.; Hang, H. C. Bioorthogonal chemical reporters for monitoring protein acetylation. *J. Am. Chem. Soc.* **2010**, *132* (11), 3640-3641, Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: 10.1021/ja908871t.

(73) Yang, Y. Y.; Grammel, M.; Hang, H. C. Identification of lysine acetyltransferase p300 substrates using 4-pentynoyl-coenzyme A and bioorthogonal proteomics. *Bioorg. Med. Chem. Lett.* **2011**, *21* (17), 4976-4979, Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: 10.1016/j.bmcl.2011.05.060.

(74) Han, Z.; Chou, C. W.; Yang, X.; Bartlett, M. G.; Zheng, Y. G. Profiling Cellular Substrates of Lysine Acetyltransferases GCN5 and p300 with Orthogonal Labeling and Click Chemistry. *ACS Chem. Biol.* **2017**, *12* (6), 1547-1555. DOI: 10.1021/acschembio.7b00114.

(75) Yang, C.; Mi, J.; Feng, Y.; Ngo, L.; Gao, T.; Yan, L.; Zheng, Y. G. Labeling lysine acetyltransferase substrates with engineered enzymes and functionalized cofactor surrogates. *J. Am. Chem. Soc.* **2013**, *135* (21), 7791-7794. DOI: 10.1021/ja311636b.

(76) Kadlec, J.; Hallacli, E.; Lipp, M.; Holz, H.; Sanchez-Weatherby, J.; Cusack, S.; Akhtar, A. Structural basis for MOF and MSL3 recruitment into the dosage compensation complex by MSL1. *Nat. Struct. Mol. Biol.* **2011**, *18* (2), 142-149.

(77) Yuan, H.; Rossetto, D.; Mellert, H.; Dang, W.; Srinivasan, M.; Johnson, J.; Hodawadekar, S.; Ding, E. C.; Speicher, K.; Abshiru, N.; et al. MYST protein acetyltransferase activity requires active site lysine autoacetylation. *Embo J.* **2012**, *31* (1), 58-70.

(78) Sun, B.; Guo, S.; Tang, Q.; Li, C.; Zeng, R.; Xiong, Z.; Zhong, C.; Ding, J. Regulation of the histone acetyltransferase activity of hMOF via autoacetylation of Lys274. *Cell Res.* **2011**, *21* (8), 1262-1266.

(79) Yang, C.; Wu, J.; Sinha, S. H.; Neveu, J. M.; Zheng, Y. G. Autoacetylation of the MYST Lysine Acetyltransferase MOF. *J. Biol. Chem.* **2012**, *287*, 34917-34926.

(80) Ourailidou, M. E.; Dockerty, P.; Witte, M.; Poelarends, G. J.; Dekker, F. J. Metabolic alkene labeling and in vitro detection of histone acylation via the aqueous oxidative Heck reaction. *Org Biomol Chem* **2015**, *13* (12), 3648-3653. DOI: 10.1039/c4ob02502d.

(81) Ourailidou, M. E.; van der Meer, J. Y.; Baas, B. J.; Jeronimus-Stratingh, M.; Gottumukkala, A. L.; Poelarends, G. J.; Minnaard, A. J.; Dekker, F. J. Aqueous oxidative Heck reaction as a protein-labeling strategy. *Chembiochem* **2014**, *15* (2), 209-212. DOI: 10.1002/cbic.201300714.

(82) He, M.; Han, Z.; Qiao, J.; Ngo, L.; Xiong, M. P.; Zheng, Y. G. A bioorthogonal turn-on fluorescent strategy for the detection of lysine acetyltransferase activity. *Chem. Commun. (Camb.)* **2018**, *54* (44), 5594-5597. DOI: 10.1039/c8cc02987c.

(83) Han, Z.; Luan, Y.; Zheng, Y. G. Integration of Bioorthogonal Probes and Q-FRET for the Detection of Histone Acetyltransferase Activity. *ChemBioChem* **2015**, *16*, 2605-2609. DOI: 10.1002/cbic.201500427.

(84) Grammel, M.; Hang, H. C. Chemical reporters for biological discovery. *Nat. Chem. Biol.* **2013**, *9* (8), 475-484. DOI: 10.1038/nchembio.1296.

(85) Presolski, S. I.; Hong, V. P.; Finn, M. G. Copper-Catalyzed Azide-Alkyne Click Chemistry for Bioconjugation. *Curr Protoc Chem Biol* **2011**, *3* (4), 153-162. DOI: 10.1002/9780470559277.ch110148.

(86) Yang, Y. Y.; Grammel, M.; Raghavan, A. S.; Charron, G.; Hang, H. C. Comparative analysis of cleavable azobenzene-based affinity tags for bioorthogonal chemical proteomics. *Chem. Biol.* **2010**, *17* (11), 1212-1222, Comparative Study Research Support, N.I.H., Extramural Research Support, Non-U.S. Gov't. DOI: 10.1016/j.chembiol.2010.09.012.

(87) Yang, Y. Y.; Ascano, J. M.; Hang, H. C. Bioorthogonal chemical reporters for monitoring protein acetylation. *J Am Chem Soc* **2010**, *132* (11), 3640-3641. DOI: 10.1021/ja908871t.

(88) Sinclair, W. R.; Shrimp, J. H.; Zengeya, T. T.; Kulkarni, R. A.; Garlick, J. M.; Luecke, H.; Worth, A. J.; Blair, I. A.; Snyder, N. W.; Meier, J. L. Bioorthogonal pro-metabolites for profiling short chain fatty acylation. *Chem Sci* **2018**, *9* (5), 1236-1241. DOI: 10.1039/c7sc00247e.

(89) Bao, X.; Zhao, Q.; Yang, T.; Fung, Y. M.; Li, X. D. A chemical probe for lysine malonylation. *Angew. Chem.* **2013**, *52* (18), 4883-4886. DOI: 10.1002/anie.201300252.

(90) Montgomery, D. C.; Sorum, A. W.; Guasch, L.; Nicklaus, M. C.; Meier, J. L. Metabolic Regulation of Histone Acetyltransferases by Endogenous Acyl-CoA Cofactors. *Chem. Biol.* **2015**, *22* (8), 1030-1039, Research Support, N.I.H., Intramural. DOI: 10.1016/j.chembiol.2015.06.015.

(91) Kim, S. C.; Sprung, R.; Chen, Y.; Xu, Y.; Ball, H.; Pei, J.; Cheng, T.; Kho, Y.; Xiao, H.; Xiao, L.; et al. Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol Cell* **2006**, *23* (4), 607-618. DOI: 10.1016/j.molcel.2006.06.026.

(92) Song, J.; Zheng, Y. G. Bioorthogonal Reporters for Detecting and Profiling Protein Acetylation and Acylation. *SLAS Discov* **2020**, *25* (2), 148-162. DOI: 10.1177/2472555219887144.

(93) Marmorstein, R.; Zhou, M. M. Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb Perspect Biol* **2014**, *6* (7), a018762. DOI: 10.1101/cshperspect.a018762.

(94) Poziello, A.; Nebbioso, A.; Stunnenberg, H. G.; Martens, J. H. A.; Carafa, V.; Altucci, L. Recent insights into Histone Acetyltransferase-1: biological function and involvement in pathogenesis. *Epigenetics* **2021**, *16* (8), 838-850. DOI: 10.1080/15592294.2020.1827723.

(95) Verreault, A.; Kaufman, P. D.; Kobayashi, R.; Stillman, B. Nucleosomal DNA regulates the core-histone-binding subunit of the human Hat1 acetyltransferase. *Curr. Biol.* **1998**, *8* (2), 96-108.

(96) Tafrova, J. I.; Tafrov, S. T. Human histone acetyltransferase 1 (Hat1) acetylates lysine 5 of histone H2A in vivo. *Mol. Cell. Biochem.* **2014**, *392* (1-2), 259-272. DOI: 10.1007/s11010-014-2036-0.

(97) Parthun, M. R. Hat1: the emerging cellular roles of a type B histone acetyltransferase. *Oncogene* **2007**, *26* (37), 5319-5328. DOI: 10.1038/sj.onc.1210602.

(98) Weinert, B. T.; Narita, T.; Satpathy, S.; Srinivasan, B.; Hansen, B. K.; Scholz, C.; Hamilton, W. B.; Zucconi, B. E.; Wang, W. W.; Liu, W. R.; et al. Time-Resolved Analysis Reveals Rapid Dynamics and Broad Scope of the CBP/p300 Acetylome. *Cell* **2018**, *174* (1), 231-244 e212. DOI: 10.1016/j.cell.2018.04.033.

(99) Agudelo Garcia, P. A.; Nagarajan, P.; Parthun, M. R. Hat1-Dependent Lysine Acetylation Targets Diverse Cellular Functions. *J. Proteome Res.* **2020**, *19* (4), 1663-1673. DOI: 10.1021/acs.jproteome.9b00843.

(100) Shaw, P. G.; Chaerkady, R.; Zhang, Z.; Davidson, N. E.; Pandey, A. Monoclonal antibody cocktail as an enrichment tool for acetylome analysis. *Anal. Chem.* **2011**, *83* (10), 3623-3626. DOI: 10.1021/ac1026176.

(101) Huang, H.; Wang, D. L.; Zhao, Y. Quantitative Crotonylome Analysis Expands the Roles of p300 in the Regulation of Lysine Crotonylation Pathway. *Proteomics* **2018**, *18* (15), e1700230. DOI: 10.1002/pmic.201700230.

(102) Han, Z.; Wu, H.; Kim, S.; Yang, X.; Li, Q.; Huang, H.; Cai, H.; Bartlett, M. G.; Dong, A.; Zeng, H.; et al. Revealing the protein propionylation activity of the histone acetyltransferase MOF (males absent on the first). *J Biol Chem* **2018**, *293* (9), 3410-3420. DOI: 10.1074/jbc.RA117.000529 From NLM Medline.

(103) Yang, G.; Yuan, Y.; Yuan, H.; Wang, J.; Yun, H.; Geng, Y.; Zhao, M.; Li, L.; Weng, Y.; Liu, Z.; et al. Histone acetyltransferase 1 is a succinyltransferase for histones and non-histones and promotes tumorigenesis. *EMBO Rep.* **2021**, *22* (2), e50967. DOI: 10.15252/embr.202050967.

(104) Lyu, Z.; Zhao, Y.; Buuh, Z. Y.; Gorman, N.; Goldman, A. R.; Islam, M. S.; Tang, H. Y.; Wang, R. E. Steric-Free Bioorthogonal Labeling of Acetylation Substrates Based on a Fluorine-Thiol Displacement Reaction. *J. Am. Chem. Soc.* **2021**, *143* (3), 1341-1347. DOI: 10.1021/jacs.0c05605.

(105) Islam, K. The Bump-and-Hole Tactic: Expanding the Scope of Chemical Genetics. *Cell Chem Biol* **2018**, *25* (10), 1171-1184. DOI: 10.1016/j.chembiol.2018.07.001.

(106) Radziwon, K.; Weeks, A. M. Protein engineering for selective proteomics. *Curr. Opin. Chem. Biol.* **2021**, *60*, 10-19. DOI: 10.1016/j.cbpa.2020.07.003.

(107) Wu, H.; Moshkina, N.; Min, J.; Zeng, H.; Joshua, J.; Zhou, M. M.; Plotnikov, A. N. Structural basis for substrate specificity and catalysis of human histone acetyltransferase 1. *Proc Natl Acad Sci U S A* **2012**, *109* (23), 8925-8930. DOI: 10.1073/pnas.1114117109.

(108) Gao, T.; Yang, C.; Zheng, Y. G. Comparative studies of thiol-sensitive fluorogenic probes for HAT assays. *Anal Bioanal Chem* **2013**, *405* (4), 1361-1371. DOI: 10.1007/s00216-012-6522-5.

(109) Trievel, R. C.; Li, F. Y.; Marmorstein, R. Application of a fluorescent histone acetyltransferase assay to probe the substrate specificity of the human p300/CBP-associated factor. *Anal. Biochem.* **2000**, *287* (2), 319-328.

(110) Srinivasan, B.; Baratashvili, M.; van der Zwaag, M.; Kanon, B.; Colombelli, C.; Lambrechts, R. A.; Schaap, O.; Nollen, E. A.; Podgorsek, A.; Kosec, G.; et al. Extracellular 4'-phosphopantetheine is a source for intracellular coenzyme A synthesis. *Nat Chem Biol* **2015**, *11* (10), 784-792. DOI: 10.1038/nchembio.1906.

(111) Huang da, W.; Sherman, B. T.; Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **2009**, *4* (1), 44-57. DOI: 10.1038/nprot.2008.211.

(112) Smith, R. C. L.; Kanellos, G.; Vlahov, N.; Alexandrou, C.; Willis, A. E.; Knight, J. R. P.; Sansom, O. J. Translation initiation in cancer at a glance. *J Cell Sci* **2021**, *134* (1). DOI: 10.1242/jcs.248476.

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

(114) Deng, L.; Meng, T.; Chen, L.; Wei, W.; Wang, P. The role of ubiquitination in tumorigenesis and targeted drug discovery. *Signal Transduct Target Ther* **2020**, *5* (1), 11. DOI: 10.1038/s41392-020-0107-0.

(115) Li, A.; Davila, S.; Furu, L.; Qian, Q.; Tian, X.; Kamath, P. S.; King, B. F.; Torres, V. E.; Somlo, S. Mutations in PRKCSH cause isolated autosomal dominant polycystic liver disease. *Am J Hum Genet* **2003**, *72* (3), 691-703. DOI: 10.1086/368295.

(116) Sibon, O. C.; Strauss, E. Coenzyme A: to make it or uptake it? *Nat Rev Mol Cell Biol* **2016**, *17* (10), 605-606. DOI: 10.1038/nrm.2016.110.

(117) Leonardi, R.; Zhang, Y. M.; Rock, C. O.; Jackowski, S. Coenzyme A: back in action. *Prog Lipid Res* **2005**, *44* (2-3), 125-153. DOI: 10.1016/j.plipres.2005.04.001.

(118) Di Meo, I.; Colombelli, C.; Srinivasan, B.; de Villiers, M.; Hamada, J.; Jeong, S. Y.; Fox, R.; Woltjer, R. L.; Tepper, P. G.; Lahaye, L. L.; et al. Acetyl-4'-phosphopantetheine is stable in serum and prevents phenotypes induced by pantothenate kinase deficiency. *Sci Rep* **2017**, *7* (1), 11260. DOI: 10.1038/s41598-017-11564-8.

(119) Huang, H.; Zhang, D.; Wang, Y.; Perez-Neut, M.; Han, Z.; Zheng, Y. G.; Hao, Q.; Zhao, Y. Lysine benzoylation is a histone mark regulated by SIRT2. *Nat. Commun.* **2018**, *9* (1), 3374. DOI: 10.1038/s41467-018-05567-w.

(120) Jensen, O. N. Interpreting the protein language using proteomics. *Nat Rev Mol Cell Biol* **2006**, *7* (6), 391-403. DOI: 10.1038/nrm1939.

(121) Sabari, B. R.; Zhang, D.; Allis, C. D.; Zhao, Y. Metabolic regulation of gene expression through histone acylations. *Nat Rev Mol Cell Biol* **2017**, *18* (2), 90-101. DOI: 10.1038/nrm.2016.140.

(122) Tan, M.; Luo, H.; Lee, S.; Jin, F.; Yang, J. S.; Montellier, E.; Buchou, T.; Cheng, Z.; Rousseaux, S.; Rajagopal, N.; et al. Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **2011**, *146* (6), 1016-1028. DOI: 10.1016/j.cell.2011.08.008.

(123) Song, J.; Ngo, L.; Bell, K.; Zheng, Y. G. Chemoproteomic Profiling of Protein Substrates of a Major Lysine Acetyltransferase in the Native Cellular Context. *ACS Chem Biol* **2022**, *17* (5), 1092-1102. DOI: 10.1021/acschembio.1c00935 From NLM Medline.

(124) Bachhawat, B. K.; Coon, M. J.; Kupiecki, F. P.; Nagle, R.; Robinson, W. G. Coenzyme A thiol esters of isobutyric, methacrylic, and beta-hydroxyisobutyric acids as intermediates in the enzymatic degradation of valine. *J Biol Chem* **1957**, *224* (1), 1-11.

(125) Xu, W.; Wan, J.; Zhan, J.; Li, X.; He, H.; Shi, Z.; Zhang, H. Global profiling of crotonylation on non-histone proteins. *Cell Res* **2017**, *27* (7), 946-949. DOI: 10.1038/cr.2017.60.

(126) Li, Y.; Sabari, B. R.; Panchenko, T.; Wen, H.; Zhao, D.; Guan, H.; Wan, L.; Huang, H.; Tang, Z.; Zhao, Y.; et al. Molecular Coupling of Histone Crotonylation and Active Transcription by AF9 YEATS Domain. *Mol Cell* **2016**, *62* (2), 181-193. DOI: 10.1016/j.molcel.2016.03.028.

(127) Li, X.; Li, X. M.; Jiang, Y.; Liu, Z.; Cui, Y.; Fung, K. Y.; van der Beelen, S. H. E.; Tian, G.; Wan, L.; Shi, X.; et al. Structure-guided development of YEATS domain inhibitors by targeting pi-pi-pi stacking. *Nat Chem Biol* **2018**, *14* (12), 1140-1149. DOI: 10.1038/s41589-018-0144-y.

(128) Yu, H.; Bu, C.; Liu, Y.; Gong, T.; Liu, X.; Liu, S.; Peng, X.; Zhang, W.; Peng, Y.; Yang, J.; et al. Global crotonylome reveals CDYL-regulated RPA1 crotonylation in homologous recombination-mediated DNA repair. *Sci Adv* **2020**, *6* (11), eaay4697. DOI: 10.1126/sciadv.aay4697.

(129) Song, J.; Han, Z.; Zheng, Y. G. Identification and Profiling of Histone Acetyltransferase Substrates by Bioorthogonal Labeling. *Curr Protoc* **2022**, *2* (7), e497. DOI: 10.1002/cpz1.497 From NLM Medline.

(130) Jiang, Y.; Li, Y.; Liu, C.; Zhang, L.; Lv, D.; Weng, Y.; Cheng, Z.; Chen, X.; Zhan, J.; Zhang, H. Isonicotinylation is a histone mark induced by the anti-tuberculosis first-line drug isoniazid. *Nat Commun* **2021**, *12* (1), 5548. DOI: 10.1038/s41467-021-25867-y.

(131) Nguyen, S. S.; Prescher, J. A. Developing bioorthogonal probes to span a spectrum of reactivities. *Nat Rev Chem* **2020**, *4*, 476-489. DOI: 10.1038/s41570-020-0205-0.

(132) Lee, Y. J.; Kurra, Y.; Liu, W. R. Phospha-Michael Addition as a New Click Reaction for Protein Functionalization. *Chembiochem* **2016**, *17* (6), 456-461. DOI: 10.1002/cbic.201500697.

(133) Bos, J.; Muir, T. W. A Chemical Probe for Protein Crotonylation. *J Am Chem Soc* **2018**, *140* (14), 4757-4760. DOI: 10.1021/jacs.7b13141 From NLM Medline.

(134) Chambers, K. A.; Abularrage, N. S.; Hill, C. J.; Khan, I. H.; Scheck, R. A. A Chemical Probe for Dehydrobutyrine. *Angew Chem Int Ed Engl* **2020**, *59* (19), 7350-7355. DOI: 10.1002/anie.202003631.

(135) Qi, Y. K.; Chang, H. N.; Pan, K. M.; Tian, C. L.; Zheng, J. S. Total chemical synthesis of the site-selective azide-labeled [I66A]HIV-1 protease. *Chem Commun (Camb)* **2015**, *51* (78), 14632-14635. DOI: 10.1039/c5cc04846j.

(136) Sabari, B. R.; Tang, Z.; Huang, H.; Yong-Gonzalez, V.; Molina, H.; Kong, H. E.; Dai, L.; Shimada, M.; Cross, J. R.; Zhao, Y.; et al. Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Mol Cell* **2015**, *58* (2), 203-215. DOI: 10.1016/j.molcel.2015.02.029.

(137) Wu, Q.; Li, W.; Wang, C.; Fan, P.; Cao, L.; Wu, Z.; Wang, F. Ultradeep Lysine Crotonylome Reveals the Crotonylation Enhancement on Both Histones and Nonhistone Proteins by SAHA Treatment. *J Proteome Res* **2017**, *16* (10), 3664-3671. DOI: 10.1021/acs.jproteome.7b00380.

(138) Sharpe, A. J.; McKenzie, M. Mitochondrial Fatty Acid Oxidation Disorders Associated with Short-Chain Enoyl-CoA Hydratase (ECHS1) Deficiency. *Cells* **2018**, *7* (6). DOI: 10.3390/cells7060046 From NLM PubMed-not-MEDLINE.

(139) Truscott, R. J.; Malegan, D.; McCairns, E.; Halpern, B.; Hammond, J.; Cotton, R. G.; Mercer, J. F.; Hunt, S.; Rogers, J. G.; Danks, D. M. Two new sulphur-containing amino acids in man. *Biomed. Mass Spectrom.* **1981**, *8* (3), 99-104. DOI: 10.1002/bms.1200080304.

(140) Ishigure, K.; Shimomura, Y.; Murakami, T.; Kaneko, T.; Takeda, S.; Inoue, S.; Nomoto, S.; Koshikawa, K.; Nonami, T.; Nakao, A. Human liver disease decreases methacrylyl-CoA hydratase and beta-hydroxyisobutyryl-CoA hydrolase activities in valine catabolism. *Clin Chim Acta* **2001**, *312* (1-2), 115-121. DOI: 10.1016/s0009-8981(01)00597-6 From NLM Medline.

(141) Nair, K. S.; Short, K. R. Hormonal and signaling role of branched-chain amino acids. *J Nutr* **2005**, *135* (6 Suppl), 1547S-1552S. DOI: 10.1093/jn/135.6.1547S.

(142) Pedroso, J. A.; Zampieri, T. T.; Donato, J., Jr. Reviewing the Effects of L-Leucine Supplementation in the Regulation of Food Intake, Energy Balance, and Glucose Homeostasis. *Nutrients* **2015**, *7* (5), 3914-3937. DOI: 10.3390/nu7053914.

(143) Crozier, S. J.; Kimball, S. R.; Emmert, S. W.; Anthony, J. C.; Jefferson, L. S. Oral leucine administration stimulates protein synthesis in rat skeletal muscle. *J Nutr* **2005**, *135* (3), 376-382. DOI: 10.1093/jn/135.3.376 From NLM Medline.

(144) Saha, A. K.; Xu, X. J.; Lawson, E.; Deoliveira, R.; Brandon, A. E.; Kraegen, E. W.; Ruderman, N. B. Downregulation of AMPK accompanies leucine- and glucose-induced increases in protein synthesis and insulin resistance in rat skeletal muscle. *Diabetes* **2010**, *59* (10), 2426-2434. DOI: 10.2337/db09-1870 From NLM Medline.

(145) Du, Y.; Meng, Q.; Zhang, Q.; Guo, F. Isoleucine or valine deprivation stimulates fat loss via increasing energy expenditure and regulating lipid metabolism in WAT. *Amino Acids* **2012**, *43* (2), 725-734. DOI: 10.1007/s00726-011-1123-8 From NLM Medline.

(146) Rivas-Santiago, C. E.; Rivas-Santiago, B.; Leon, D. A.; Castaneda-Delgado, J.; Hernandez Pando, R. Induction of beta-defensins by l-isoleucine as novel immunotherapy in experimental murine tuberculosis. *Clin Exp Immunol* **2011**, *164* (1), 80-89. DOI: 10.1111/j.1365-2249.2010.04313.x From NLM Medline.

(147) Holecek, M. Branched-chain amino acids in health and disease: metabolism, alterations in blood plasma, and as supplements. *Nutr Metab (Lond)* **2018**, *15*, 33. DOI: 10.1186/s12986-018-0271-1.

(148) Zhang, S.; Zeng, X.; Ren, M.; Mao, X.; Qiao, S. Novel metabolic and physiological functions of branched chain amino acids: a review. *J Anim Sci Biotechnol* **2017**, *8*, 10. DOI: 10.1186/s40104-016-0139-z.

(149) Goberdhan, D. C.; Ogmundsdottir, M. H.; Kazi, S.; Reynolds, B.; Visvalingam, S. M.; Wilson, C.; Boyd, C. A. Amino acid sensing and mTOR regulation: inside or out? *Biochem Soc Trans* **2009**, *37* (Pt 1), 248-252. DOI: 10.1042/BST0370248 From NLM Medline.

(150) Newgard, C. B. Interplay between lipids and branched-chain amino acids in development of insulin resistance. *Cell Metab* **2012**, *15* (5), 606-614. DOI: 10.1016/j.cmet.2012.01.024 From NLM Medline.

(151) Lynch, C. J.; Adams, S. H. Branched-chain amino acids in metabolic signalling and insulin resistance. *Nat Rev Endocrinol* **2014**, *10* (12), 723-736. DOI: 10.1038/nrendo.2014.171 From NLM Medline.

(152) Manoli, I.; Venditti, C. P. Disorders of branched chain amino acid metabolism. *Transl Sci Rare Dis* **2016**, *1* (2), 91-110. DOI: 10.3233/TRD-160009.

(153) Nie, C.; He, T.; Zhang, W.; Zhang, G.; Ma, X. Branched Chain Amino Acids: Beyond Nutrition Metabolism. *Int J Mol Sci* **2018**, *19* (4). DOI: 10.3390/ijms19040954.

(154) Neinast, M.; Murashige, D.; Arany, Z. Branched Chain Amino Acids. *Annu Rev Physiol* **2019**, *81*, 139-164. DOI: 10.1146/annurev-physiol-020518-114455 From NLM Medline.

(155) Sivanand, S.; Vander Heiden, M. G. Emerging Roles for Branched-Chain Amino Acid Metabolism in Cancer. *Cancer Cell* **2020**, *37* (2), 147-156. DOI: 10.1016/j.ccell.2019.12.011 From NLM Medline.

(156) Ferdinandusse, S.; Waterham, H. R.; Heales, S. J.; Brown, G. K.; Hargreaves, I. P.; Taanman, J. W.; Gunny, R.; Abulhoul, L.; Wanders, R. J.; Clayton, P. T.; et al. HIBCH mutations can cause Leigh-like disease with combined deficiency of multiple mitochondrial respiratory chain enzymes and pyruvate dehydrogenase. *Orphanet J. Rare Dis.* **2013**, *8*, 188. DOI: 10.1186/1750-1172-8-188.

(157) Brown, G. K.; Hunt, S. M.; Scholem, R.; Fowler, K.; Grimes, A.; Mercer, J. F.; Truscott, R. M.; Cotton, R. G.; Rogers, J. G.; Danks, D. M. beta-hydroxyisobutyryl coenzyme A deacylase deficiency: a defect in valine metabolism associated with physical malformations. *Pediatrics* **1982**, *70* (4), 532-538.

(158) Peters, H.; Buck, N.; Wanders, R.; Ruiter, J.; Waterham, H.; Koster, J.; Yaplito-Lee, J.; Ferdinandusse, S.; Pitt, J. ECHS1 mutations in Leigh disease: a new inborn error of metabolism affecting valine metabolism. *Brain* **2014**, *137* (Pt 11), 2903-2908. DOI: 10.1093/brain/awu216.

(159) Yamada, K.; Aiba, K.; Kitaura, Y.; Kondo, Y.; Nomura, N.; Nakamura, Y.; Fukushi, D.; Murayama, K.; Shimomura, Y.; Pitt, J.; et al. Clinical, biochemical and metabolic characterisation

of a mild form of human short-chain enoyl-CoA hydratase deficiency: significance of increased N-acetyl-S-(2-carboxypropyl)cysteine excretion. *J Med Genet* **2015**, *52* (10), 691-698. DOI: 10.1136/jmedgenet-2015-103231 From NLM Medline.

(160) Huffnagel, I. C.; Redeker, E. J. W.; Reneman, L.; Vaz, F. M.; Ferdinandusse, S.; Poll-The, B. T. Mitochondrial Encephalopathy and Transient 3-Methylglutaconic Aciduria in ECHS1 Deficiency: Long-Term Follow-Up. *JIMD Rep* **2018**, *39*, 83-87. DOI: 10.1007/8904_2017_48 From NLM PubMed-not-MEDLINE.

(161) Yu, F.; Wang, Y. Z.; Hang, Y.; Tang, W. M.; Zhao, Z. F.; Oupicky, D. Synthesis and biological characterization of clicked chloroquine copolymers as macromolecular inhibitors of cancer cell migration. *J Polym Sci Pol Chem* **2019**, *57* (22), 2235-2242. DOI: 10.1002/pola.29512.

(162) Krzywik, J.; Nasulewicz-Goldeman, A.; Mozga, W.; Wietrzyk, J.; Huczynski, A. Novel Double-Modified Colchicine Derivatives Bearing 1,2,3-Triazole: Design, Synthesis, and Biological Activity Evaluation. *Acs Omega* **2021**, *6* (40), 26583-26600. DOI: 10.1021/acsomega.1c03948.

(163) O'Shea, J. P.; Chou, M. F.; Quader, S. A.; Ryan, J. K.; Church, G. M.; Schwartz, D. pLogo: a probabilistic approach to visualizing sequence motifs. *Nat Methods* **2013**, *10* (12), 1211-1212. DOI: 10.1038/nmeth.2646 From NLM Medline.

(164) Witt, A. C.; Lakshminarasimhan, M.; Remington, B. C.; Hasim, S.; Pozharski, E.; Wilson, M. A. Cysteine pKa depression by a protonated glutamic acid in human DJ-1. *Biochemistry* **2008**, *47* (28), 7430-7440. DOI: 10.1021/bi800282d.

(165) Taylor, J.; Spiller, M.; Ranguin, K.; Vitobello, A.; Philippe, C.; Bruel, A. L.; Cappuccio, G.; Brunetti-Pierri, N.; Willems, M.; Isidor, B.; et al. Expanding the phenotype of HNRNPU-related

neurodevelopmental disorder with emphasis on seizure phenotype and review of literature. *Am J Med Genet A* **2022**, *188* (5), 1497-1514. DOI: 10.1002/ajmg.a.62677.

(166) Mediouni, S.; Jablonski, J.; Paris, J. J.; Clementz, M. A.; Thenin-Houssier, S.; McLaughlin, J. P.; Valente, S. T. Didehydro-cortistatin A inhibits HIV-1 Tat mediated neuroinflammation and prevents potentiation of cocaine reward in Tat transgenic mice. *Curr HIV Res* **2015**, *13* (1), 64-79. DOI: 10.2174/1570162x13666150121111548 From NLM Medline.

(167) Beausoleil, S. A.; Villen, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat Biotechnol* **2006**, *24* (10), 1285-1292. DOI: 10.1038/nbt1240 From NLM Medline.

(168) Narita, T.; Weinert, B. T.; Choudhary, C. Functions and mechanisms of non-histone protein acetylation. *Nat Rev Mol Cell Biol* **2019**, *20* (3), 156-174. DOI: 10.1038/s41580-018-0081-3.

(169) Yang, M.; Zhang, Y.; Ren, J. Acetylation in cardiovascular diseases: Molecular mechanisms and clinical implications. *Biochim Biophys Acta Mol Basis Dis* **2020**, *1866* (10), 165836. DOI: 10.1016/j.bbadis.2020.165836 From NLM Medline.

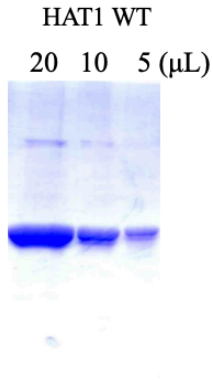
(170) Xu, H.; Wang, Y.; Lin, S.; Deng, W.; Peng, D.; Cui, Q.; Xue, Y. PTMD: A Database of Human Disease-associated Post-translational Modifications. *Genomics Proteomics Bioinformatics* **2018**, *16* (4), 244-251. DOI: 10.1016/j.gpb.2018.06.004 From NLM Medline.

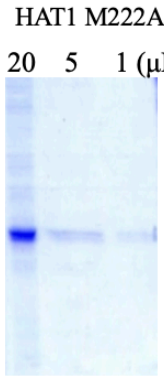
(171) Peace, C. G.; O'Neill, L. A. The role of itaconate in host defense and inflammation. *J Clin Invest* **2022**, *132* (2). DOI: 10.1172/JCI148548 From NLM Medline.

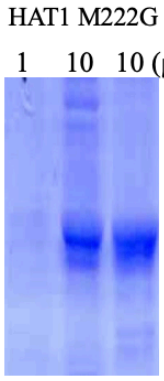
(172) Huang, H.; Lin, S.; Garcia, B. A.; Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem Rev* **2015**, *115* (6), 2376-2418. DOI: 10.1021/cr500491u From NLM Medline.

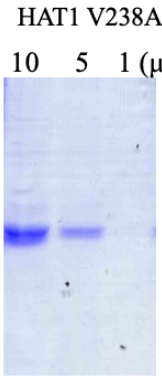
APPENDICS

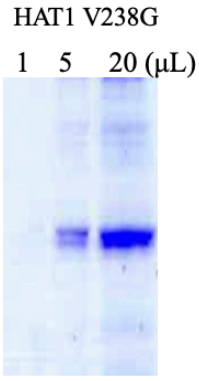
A Supporting information for Chapter 2

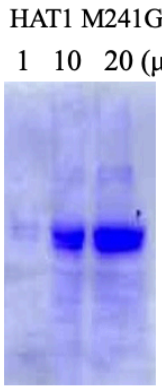


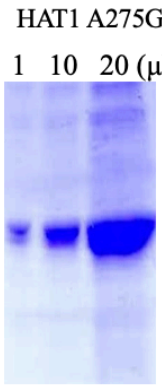


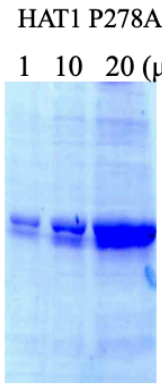


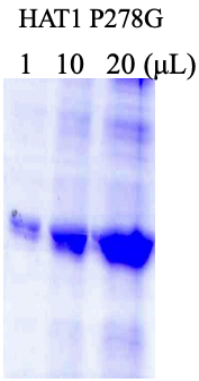


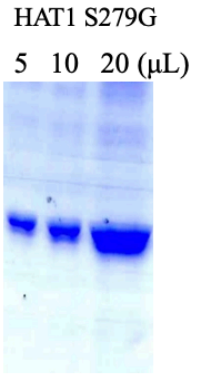


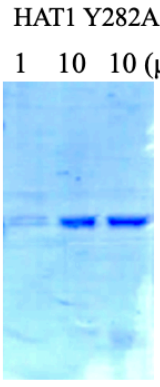


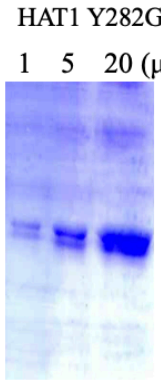












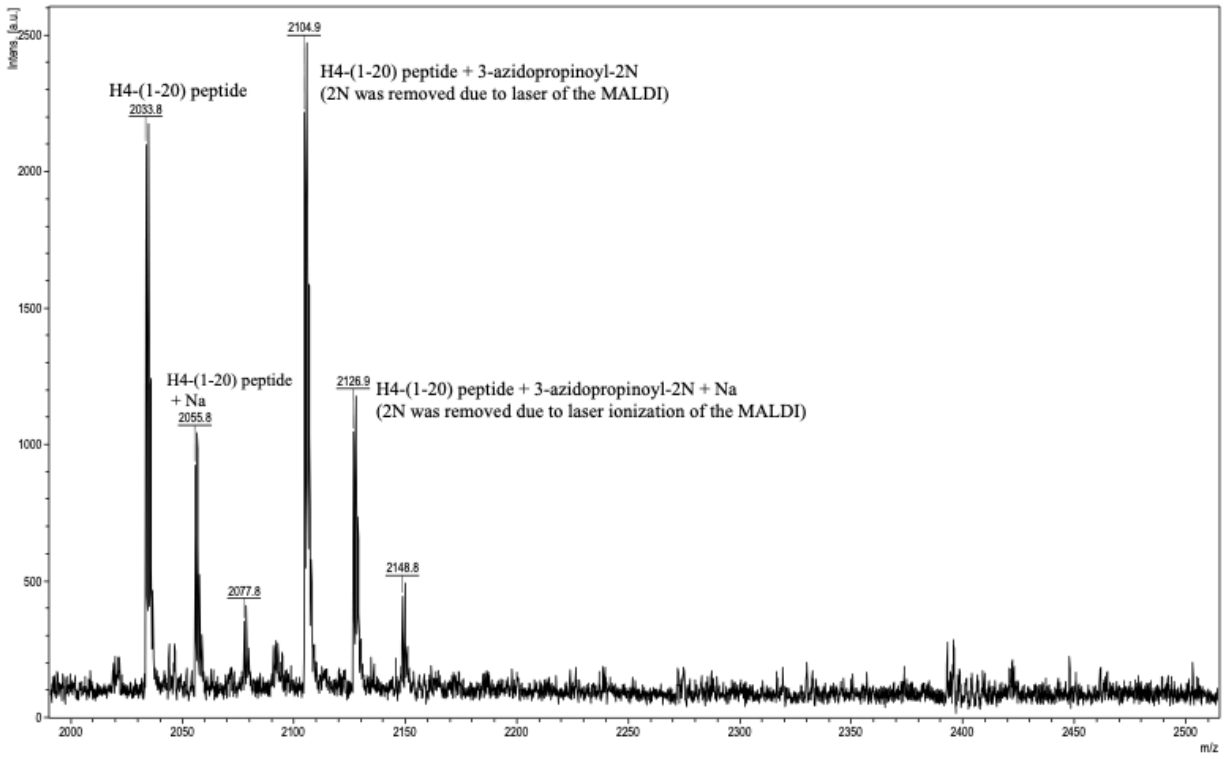
Supplementary Figure S2.1 Examination of the recombinant wt- and eng-HAT1 proteins by SDS-PAGE. The eluted proteins were loaded on the gel with different volume and the gel was stained by Coomassie Blue staining.

	Ac-CoA	4PY-CoA	5HY-CoA	6HY-CoA	3AZ-CoA	4AZ-CoA
HAT1 WT	1.0	0.0	0.0	0.0	0.1	0.0
HAT1 M222A	0.8	0.0	0.0	0.0	0.1	0.0
HAT1 M222G	0.0	0.0	0.0	0.0	0.0	0.0
HAT1 V238A	0.1	0.3	0.0	0.0	0.1	0.0
HAT1 V238G	0.0	0.0	0.0	0.0	0.0	0.0
HAT1 M241A	0.9	0.0	0.0	0.0	0.1	0.0
HAT1 M241G	0.1	0.0	0.0	0.0	0.2	0.1
HAT1 A275G	0.2	0.0	0.0	0.0	0.1	0.0
HAT1 P278A	0.4	0.0	0.0	0.0	0.1	0.0
HAT1 P278G	0.1	0.0	0.0	0.0	0.0	0.0
HAT1 S279G	1.7	0.0	0.0	0.0	0.1	0.0
HAT1 Y282A	0.1	0.6	1.0	2.4	1.3	1.1
HAT1 Y282G	0.0	0.1	0.1	0.5	0.2	0.2

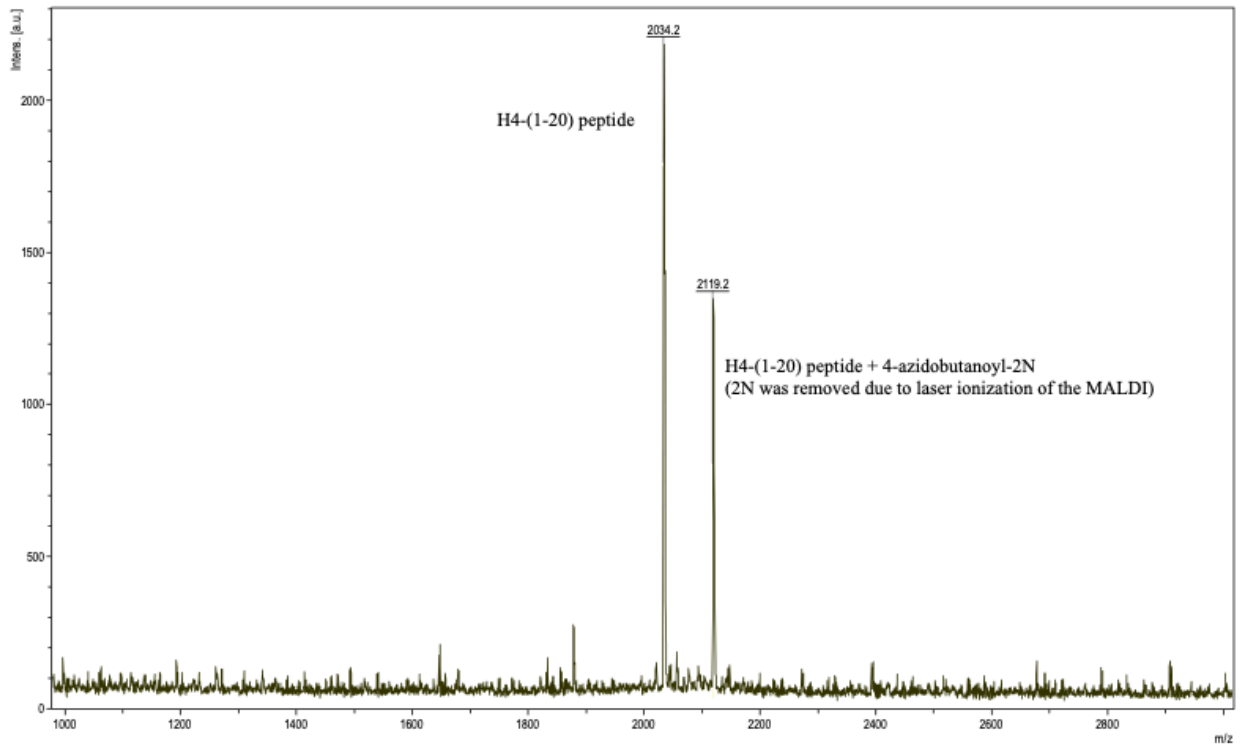
> 0.8
0.4 - 0.8
0.2 - 0.4
0.05 - 0.2
< 0.05

Supplementary Figure S2.2 Summarized data for the enzymatic activities of wt-HAT and HAT1 mutants for different Ac-CoA analogs. (The heat map **Fig. 2B** is the same except that the numbers were removed for clarity purpose.)

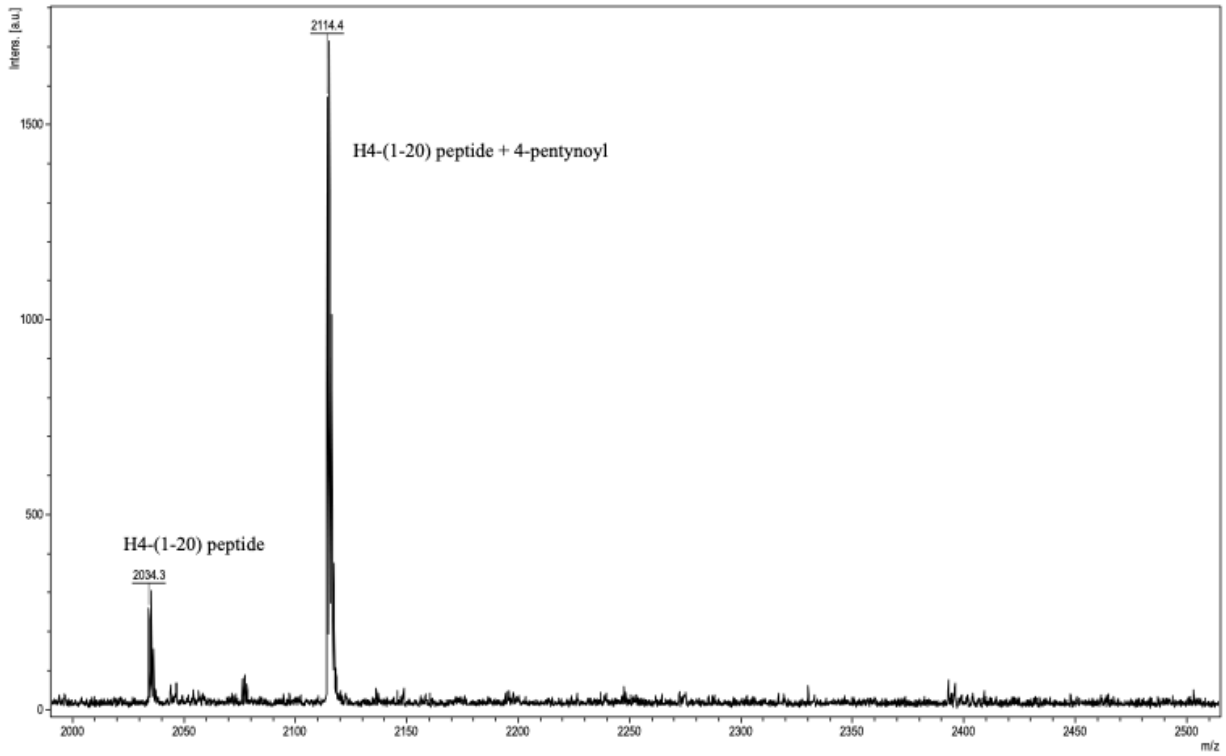
(**A**) H4-(1-20) peptide modified by HAT1-Y282A and 3AZ-CoA



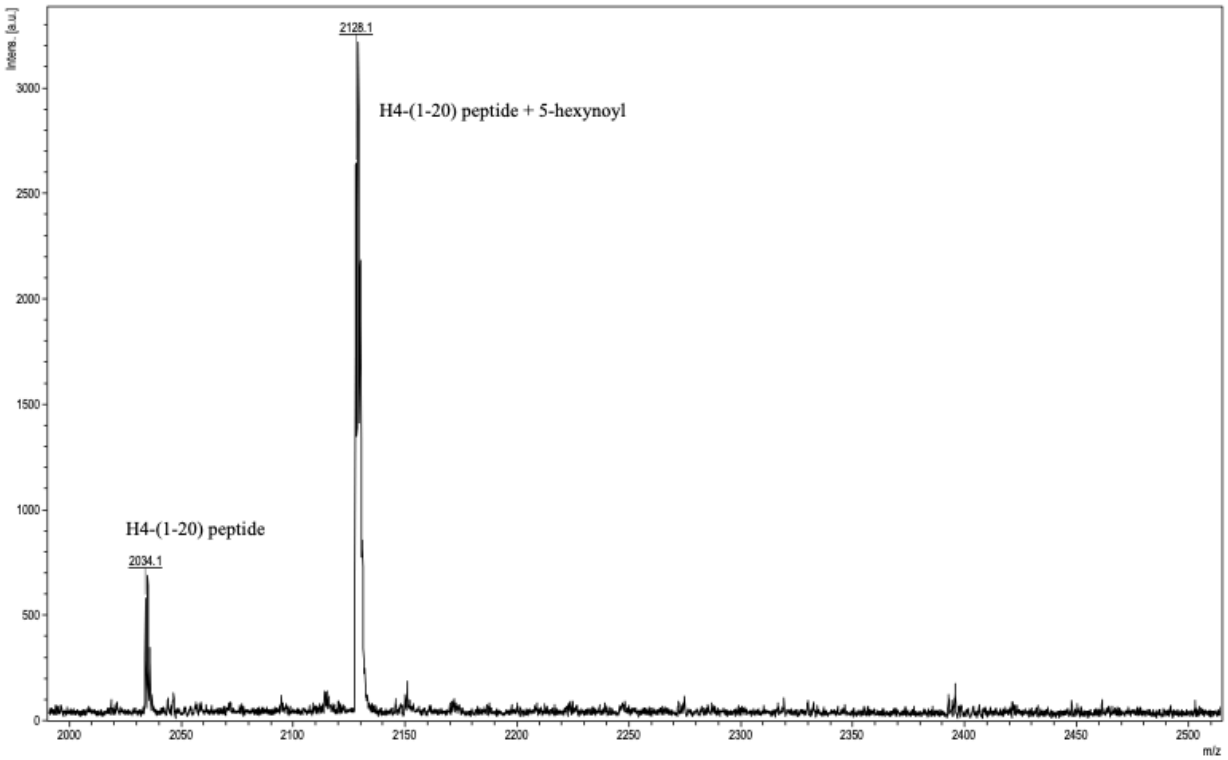
(**B**) H4-(1-20) peptide modified by HAT1-Y282A and 4AZ-CoA



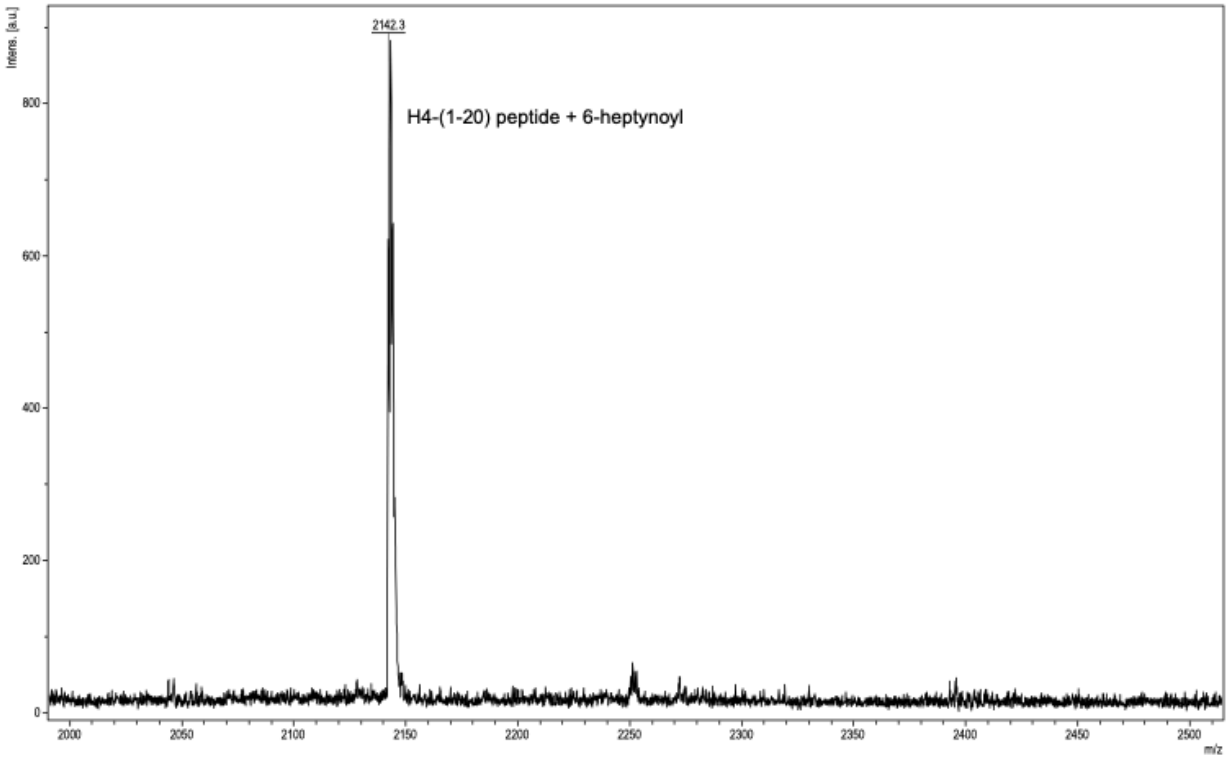
(**C**) H4-(1-20) peptide modified by HAT1-Y282A and 4PY-CoA



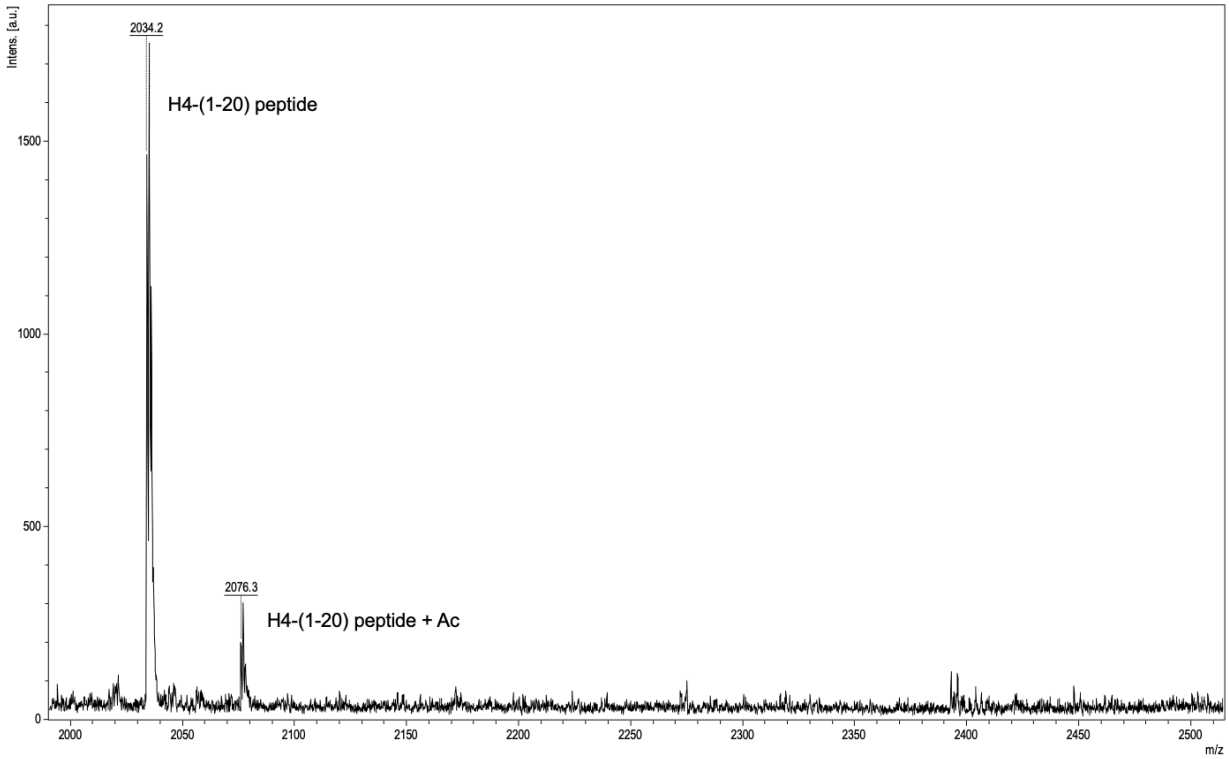
(**D**) H4-(1-20) peptide modified by HAT1-Y282A and 5HY-CoA



(**E**) H4-(1-20) peptide modified by HAT1-Y282A and 6HY-CoA

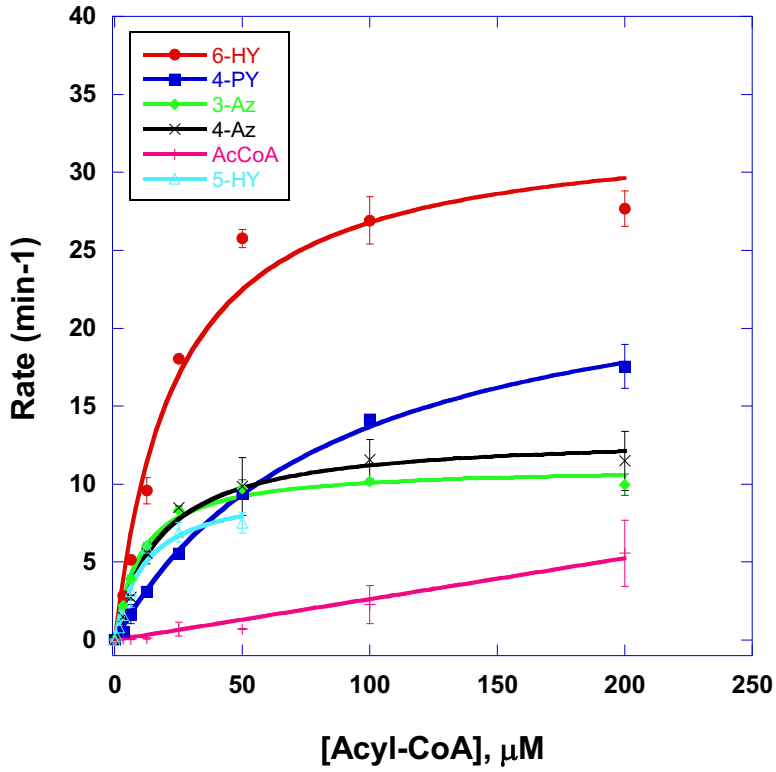


(**F**) H4-(1-20) peptide modified by HAT1-Y282A and Ac-CoA

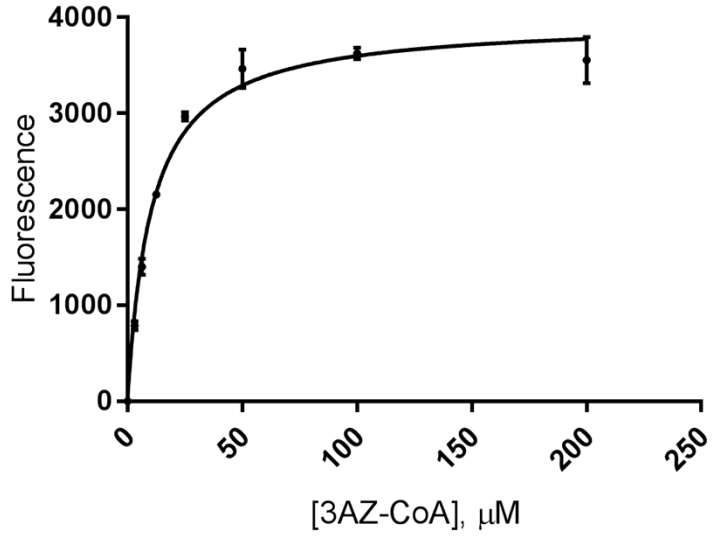


Supplementary Figure S2.3 Mass spectra of H4-20 peptide modified by HAT1-Y282A and Ac-CoA analogs. (**A**) H4-(1-20) peptide was modified by HAT1-Y282A and 3AZ-CoA. (2N was removed due to laser of MALDI (**B**) H4-(1-20) peptide was modified by HAT1-Y282A and 4AZ-CoA. (2N was removed due to laser of MALDI) (**C**) H4-(1-20) peptide was modified by HAT1-Y282A and 4PY-CoA. (**D**) H4-(1-20) peptide was modified by HAT1-Y282A and 5HY-CoA. (**E**) H4-(1-20) peptide was modified by HAT1-Y282A and 6HY-CoA. (**F**) H4-(1-20) peptide was modified by HAT1-Y282A and Ac-CoA.

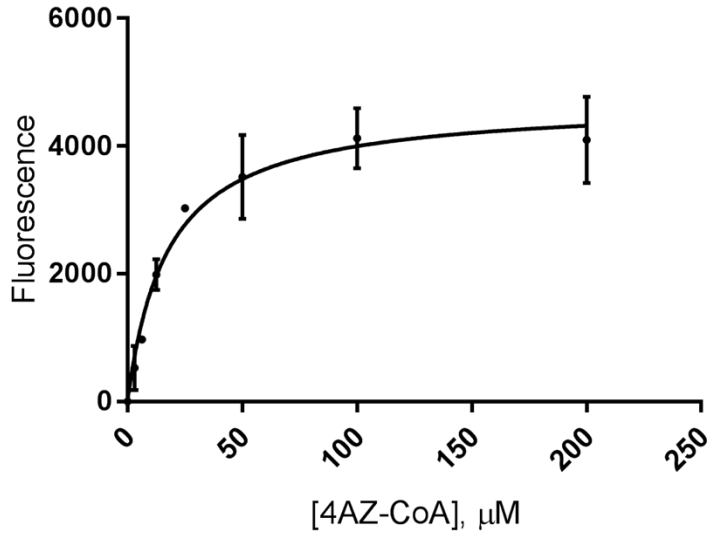
(**A**)



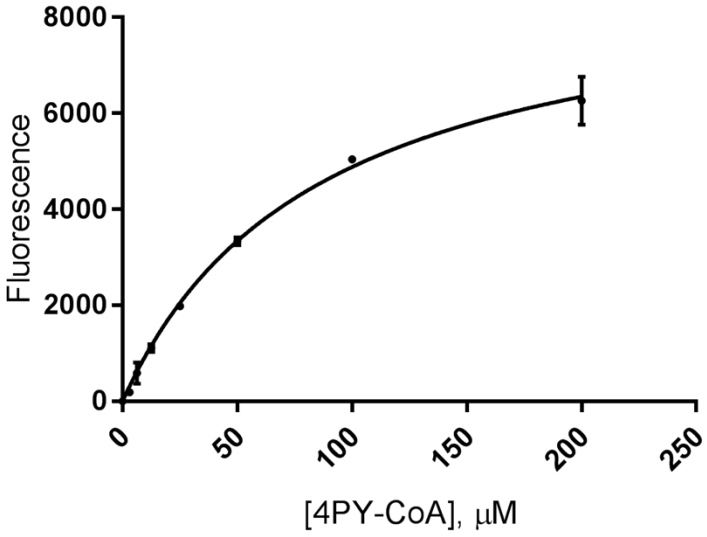
(**B**)


4000
3000
2000
1000
0
Fluorescence
0
50
100
150
200
250
[3AZ-CoA], μM

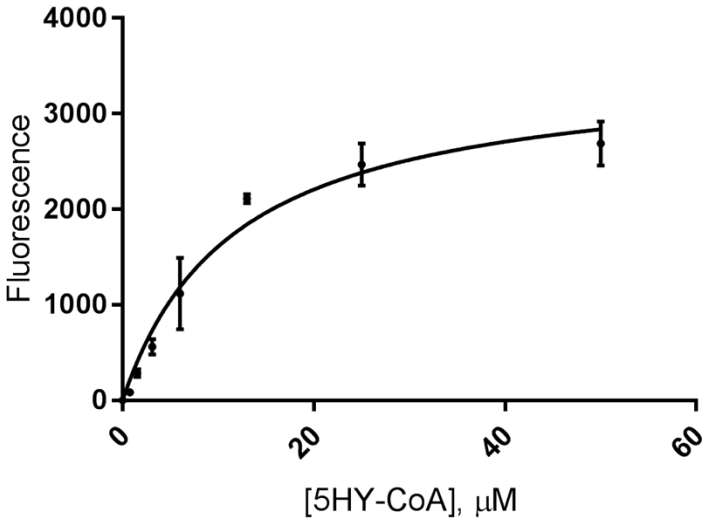
(**C**)


6000
4000
2000
0
Fluorescence
0
50
100
150
200
250
[4AZ-CoA], μM

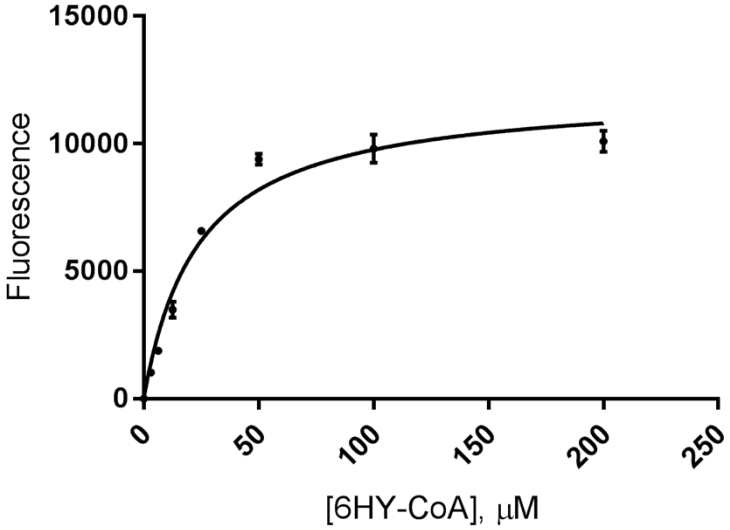
(**D**)


8000
6000
4000
2000
0
Fluorescence
0
50
100
150
200
250
[4PY-CoA], μM

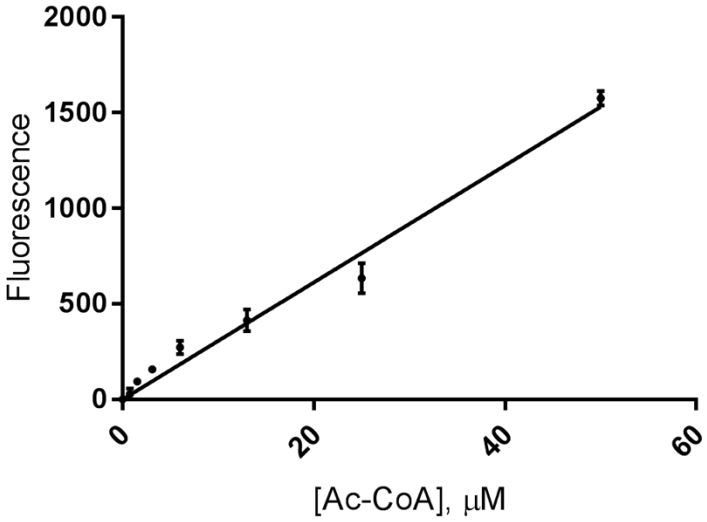
(**E**)


4000
3000
2000
1000
0
Fluorescence
0
20
40
60
[5HY-CoA], μM

(**F**)

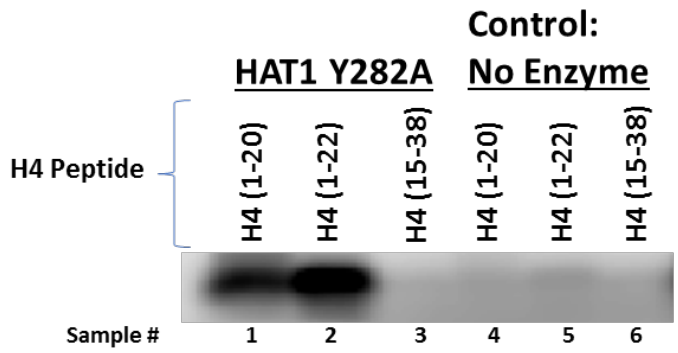


(**G**)

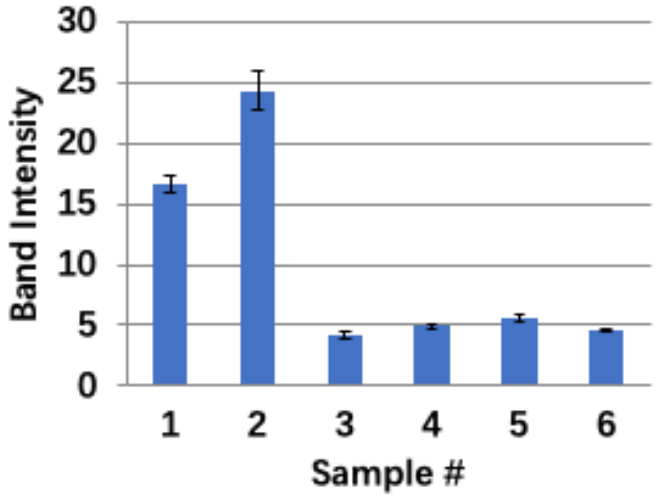


Supplementary Figure S2.4 Kinetic analysis of HAT1-Y282A activity with several acyl-CoA analogs. (**A**) Graphical representation of HAT1-Y282A activity fitted to the Michaelis–Menten equation. (**B**) Fluorescence detection of HAT1-Y282A to 3AZ-CoA. (**C**) Fluorescence detection of HAT1-Y282A to 4AZ-CoA. (**D**) Fluorescence detection of HAT1-Y282A to 4PY-CoA. (**E**) Fluorescence detection of HAT1-Y282A to 5HY-CoA. (**F**) Fluorescence detection of HAT1-Y282A to 6HY-CoA. (**G**) Fluorescence detection of HAT1-Y282A to Ac-CoA.

(**A**)

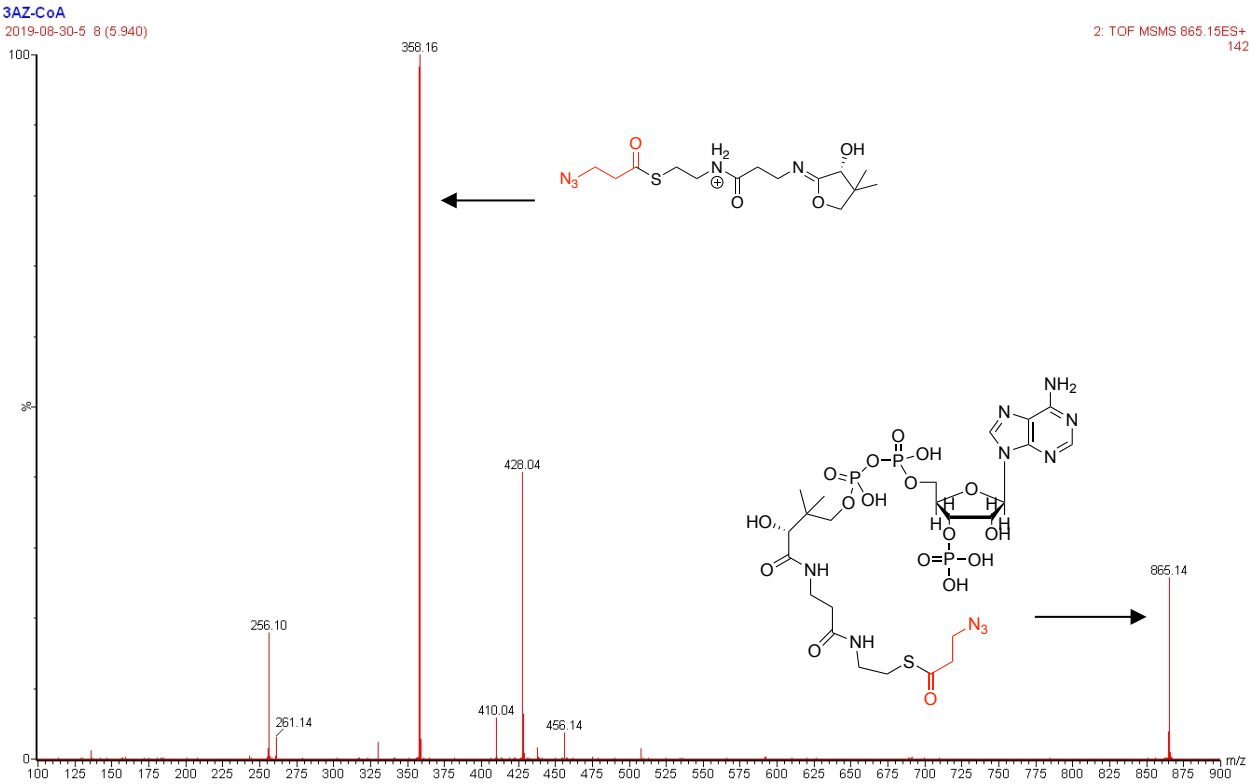


(**B**)

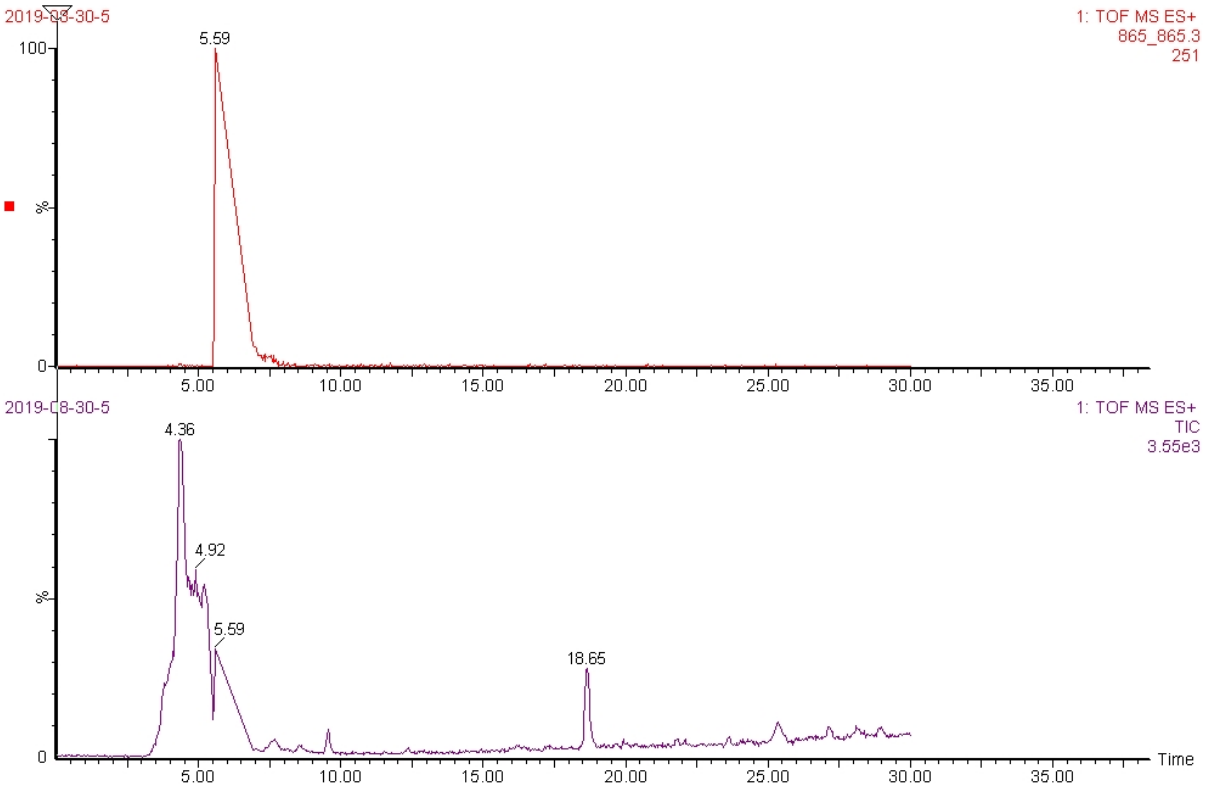


Supplementary Figure S2.5 Labeling of various H4-peptides using HAT1-Y282A. (**A**) Reactions containing 20 µM 3AZ-CoA, with or without 0.2 µM HAT1Y282A, and 40 µM H4-peptide were incubated at 30°C for 1h and then the peptides were conjugated with TAMRA alkyne (catalog# TA108) by CuAAC reaction. Gel was de-stained overnight. Gel was imaged at excitation of 532nm and emission of 580nm. (**B**) Quantification of the fluorescent histone H4 peptide bands.

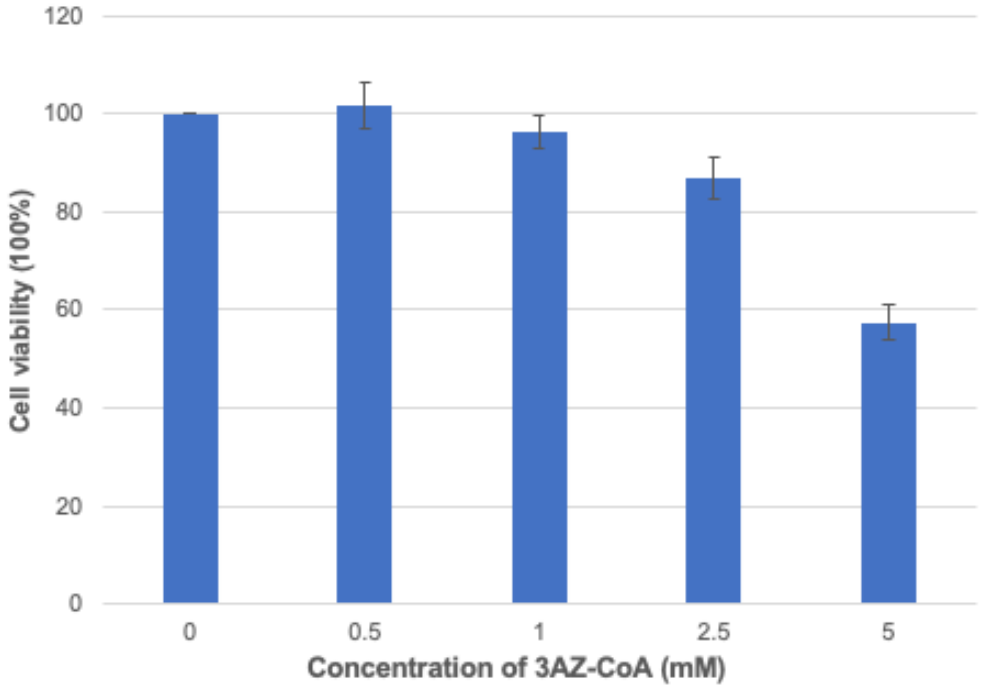
(A)


3AZ-CoA
2019-08-30-5 8 (5.940)
2: TOF MSMS 865.15ES+
142
358.16
428.04
256.10
261.14
410.04
456.14
865.14
%
m/z

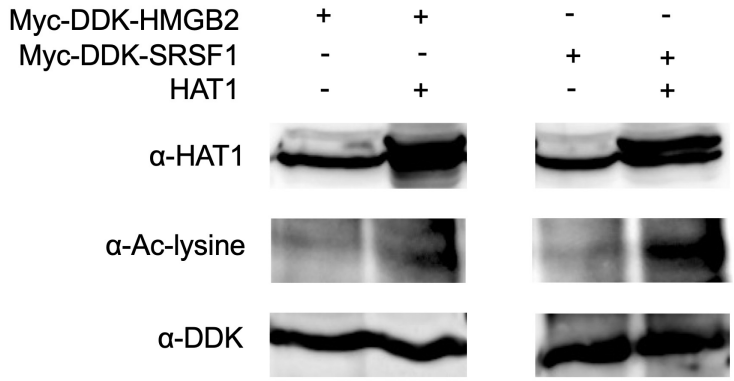
(B)



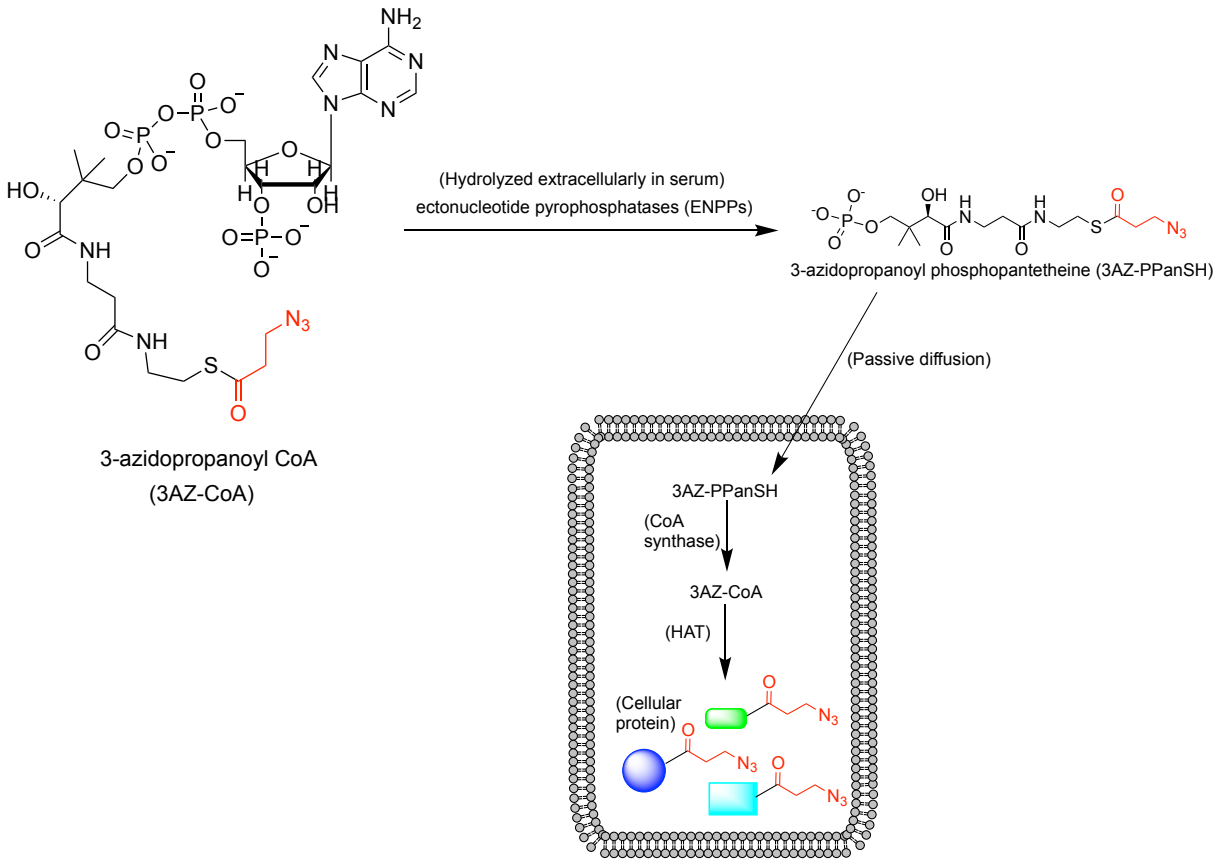
Supplementary Figure S2.6 LC-MS/MS analysis for the formation of 3AZ-CoA in HEK293T cells. Cells were treated with 2.5mM 3AZ-CoA and incubated for 6 h. (**A**) Product ions derived from 3AZ-CoA treatment of cell extract (865.16 m/z corresponding to 3AZ-CoA). (**B**) Extracted ion chromatogram (top) and total ion Chromatogram (bottom) of the cell extract with 3AZ-CoA treatment. (3AZ-CoA elution at 5.59 mins)



Supplementary Figure S2.7 Toxicity of 3AZ-CoA in HEK293T cells by resazurin assay. The cells were incubated with indicated concentration of 3AZ-CoA for 12 h. With 1mM and 2.5mM 3AZ-CoA incubation, the cell deaths can be minimized.



Supplementary Figure S2.8 Validation of HAT1 substrates by immunoprecipitation and western blot. The Myc-DDK-HMGB2 DNA and Myc-DDK-SRSF1 DNA were separately transiently transfected into HEK293T cell line with or without wt-HAT1 cotransfection. After 48 h of incubation, cells were lysed and the target proteins were immunoprecipitated with the anti-DDK antibody. The acetylation level of HMGB2 and SRSF1 were examined with Western blot by using pan anti-acetyllysine antibody.



Supplementary Figure S2.9 Possible mechanism of 3AZ-CoA getting into cells. 3AZ-CoA was hydrolyzed to 3AZ-PPanSH extracellularly in serum. 3AZ-PPanSH can enter cells by passive diffusion. After getting into cells, 3AZ-PPanSH can be metabolically synthesized to become 3AZ-CoA which can be utilized by cellular HAT enzymes.

Supplementary Table S2.1 Sequences of the forward and reverse PCR Primers of HAT1 mutagenesis

Primers	DNA Sequences from 5' to 3'
HAT1 M222A forward	GCTCTTTGCGACCGTAGGCTAC**GCC**ACAGTCTATAATTAC
HAT1 M222A reverse	GTAATTATAGACTGT**GGC**GTAGCCTACGGTCGCAAAGAGC
HAT1 M222G forward	GCTCTTTGCGACCGTAGGCTAC**GGG**ACAGTCTATAATTAC
HAT1 M222G reverse	GTAATTATAGACTGT**CCC**GTAGCCTACGGTCGCAAAGAGC
HAT1 V238A forward	GACAAAACCCGGCCACGT**GCC**AGTCAGATGC
HAT1 V238A reverse	GCATCTGACT**GGC**ACGTGGCCGGGTTTTGTC
HAT1 V238G forward	GACAAAACCCGGCCACGT**GGG**AGTCAGATGC
HAT1 V238G reverse	GCATCTGACT**CCC**ACGTGGCCGGGTTTTGTC
HAT1 M241A forward	CGTGTAAGTCAG**GCC**CTGATTTTGACTCCATTTCAAGG
HAT1 M241A reverse	CCTTGAAATGGAGTCAAAATCAG**GGC**CTGACTTACACG
HAT1 M241G forward	CGTGTAAGTCAG**GGG**CTGATTTTGACTCCATTTCAAGG
HAT1 M241G reverse	CCTTGAAATGGAGTCAAAATCAG**CCC**CTGACTTACACG
HAT1 A275G forward	CCTACAGTTCTTGATATTACA**GGG**GAAGATCCATCCAAAAGC
HAT1 A275G reverse	GCTTTTGGATGGATCTT**CCC**CTGTAATATCAAGAACTGTAGG
HAT1 P278A forward	GATATTACAGCGGAAGAT**GCC**TCCAAAAGC
HAT1 P278A reverse	GCTTTTGGA**GGC**ATCTTCCGCTGTAATATC
HAT1 P278G forward	GATATTACAGCGGAAGAT**GGG**TCCAAAAGC
HAT1 P278G reverse	GCTTTTGGA**CCC**ATCTTCCGCTGTAATATC
HAT1 S279G forward	CGGAAGATCCA**GGG**AAAAGCTATGTGAAATTACGAG
HAT1 S279G reverse	CTCGTAATTTCACATAGCTTTT**CCC**TGGATCTTCCG
HAT1 Y282A forward	CGGAAGATCCATCCAAAAGC**GCC**GTGAAATTACG
HAT1 Y282A reverse	CGTAATTTCAC**GGC**GCTTTTGGATGGATCTTCCG

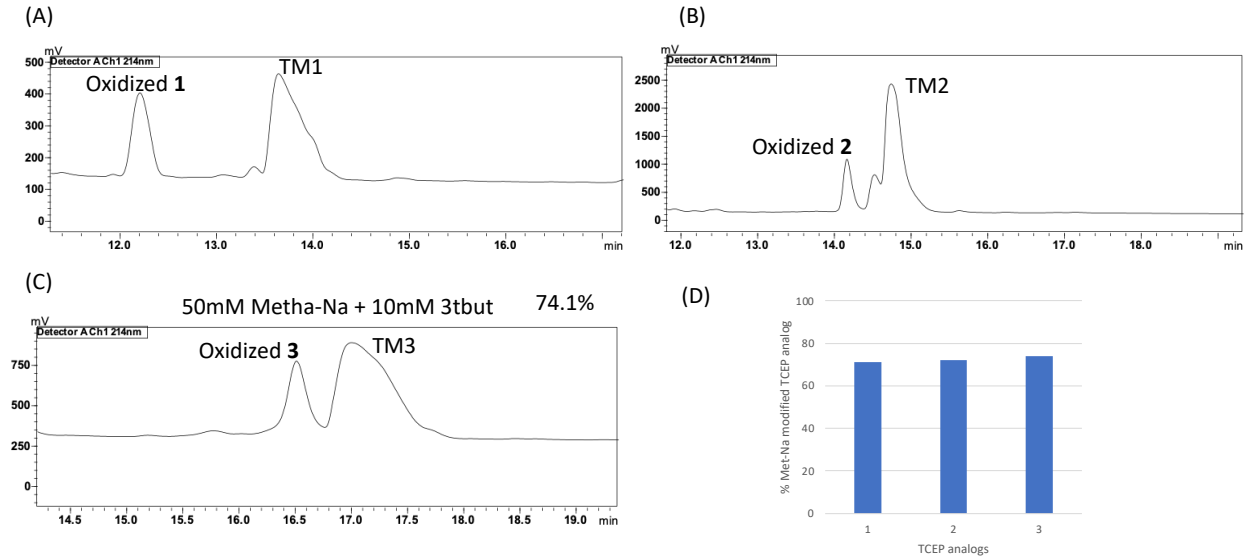
HAT1 Y282G forward	CGGAAGATCCATCCAAAAGC**GGG**GTGAAATTACG
HAT1 Y282G reverse	CGTAATTTCAC**CCC**GCTTTTGGATGGATCTTCCG

The mutated codon is highlighted in yellow color.

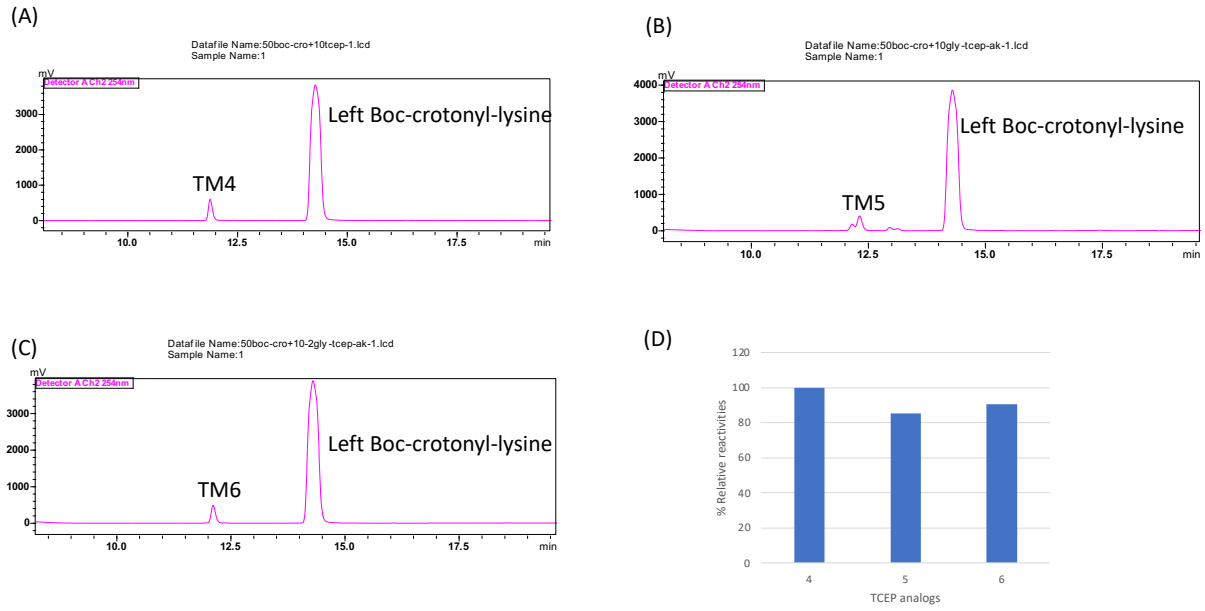
Additional Supplementary Tables can be found online at:

https://pubs.acs.org/doi/full/10.1021/acschembio.1c00935

B Supporting information for Chapter 3

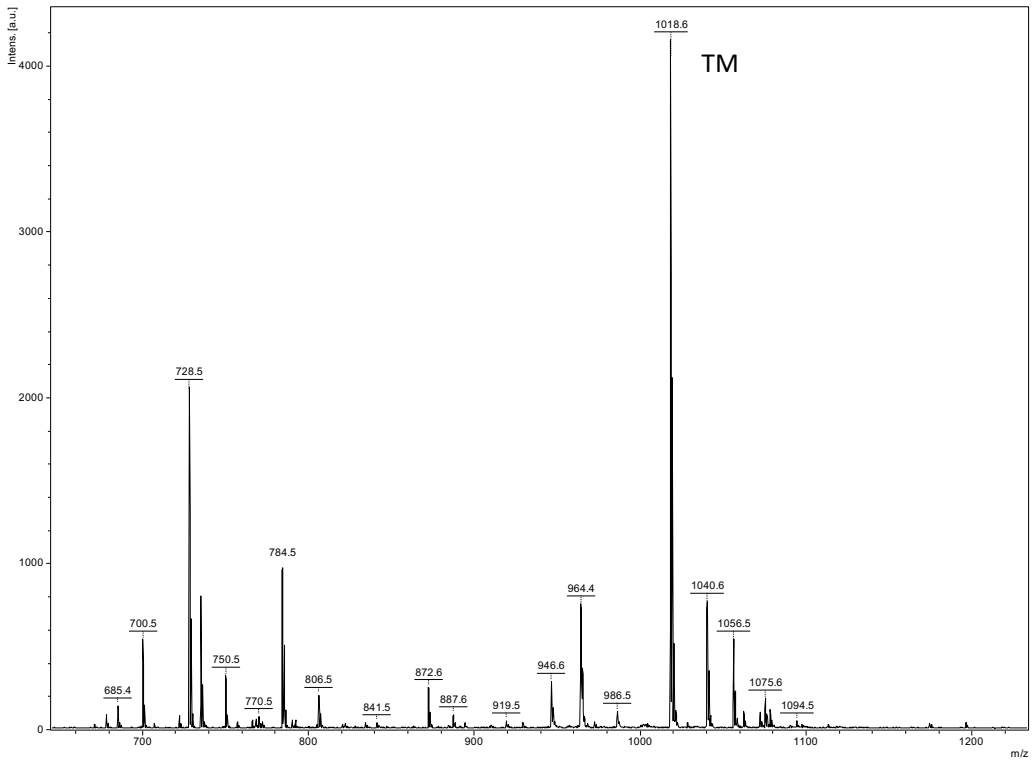


Supplementary Figure S3.1 Study the reactivity between TCEP analogs (1-3) and sodium methacrylate. (A) 50 mM sodium methacrylate reacted with 10 mM t-butyl-Gly-TCEP, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS. 71.3% of the t-butyl-Gly-TCEP became the target molecule1, and the rest was oxidized. (B) 50 mM sodium methacrylate reacted with 10 mM 2t-butyl-Gly-TCEP, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS. 72.2% of the 2t-butyl-Gly-TCEP became the target molecule2, and the rest was oxidized. (C) 50 mM sodium methacrylate reacted with 10 mM 3t-butyl-Gly-TCEP, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS. 74.1% of the 3t-butyl-Gly-TCEP became the target molecule3, and the rest was oxidized. (D) The column graph shows the percentage conversion of the TCEP analogs.

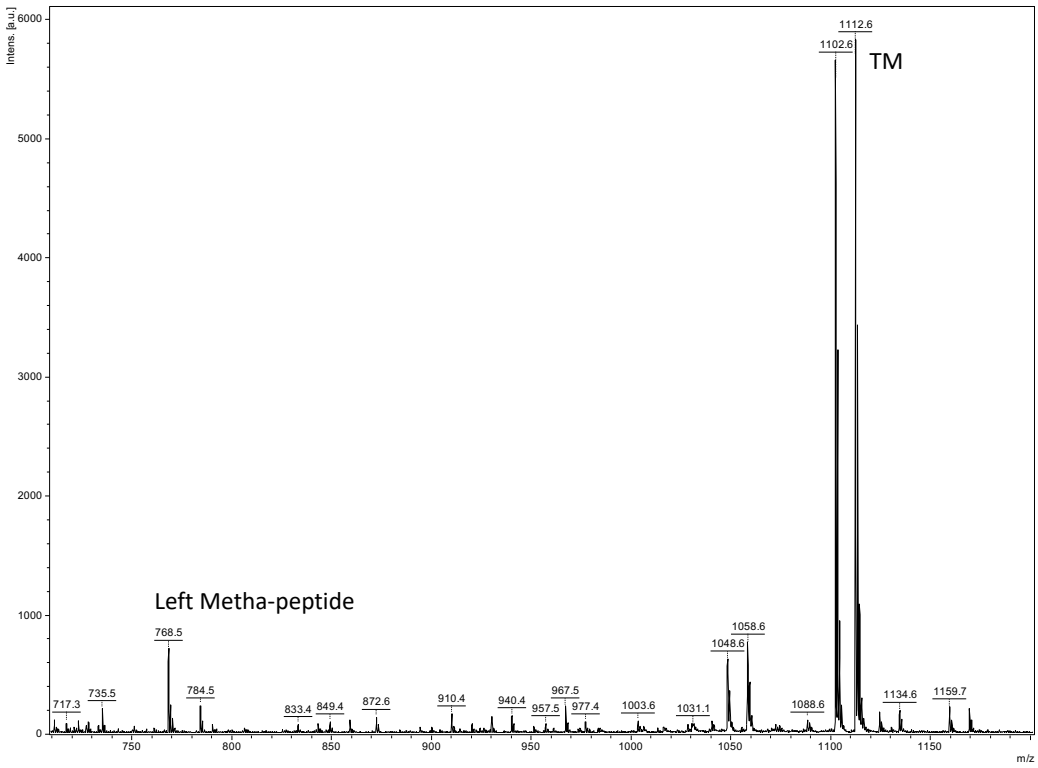


Supplementary Figure S3.2 Study the reactivity between TCEP probes (4-6) and Boc-crotonyl-lysine. (A) 50 mM Boc-crotonyl-lysine reacted with 10 mM TCEP, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS. The left peak represents the target molecule (TM4) and the right peak represents the left Boc-crotonyl-lysine. (B) 50 mM Boc-crotonyl-lysine reacted with 10 mM Gly-TCEP-AK, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS. The left peak represents the target molecule (TM5) and the right peak represents the left Boc-crotonyl-lysine. (C) 50 mM Boc-crotonyl-lysine reacted with 10 mM 2Gly-TCEP-AK, and after overnight incubation, the reaction mixture was analyzed by HPLC and ESI-MS. The left peak represents the target molecule (TM6) and the right peak represents the left Boc-crotonyl-lysine. (D) The column graph shows the percentage conversion of the TCEP probes.

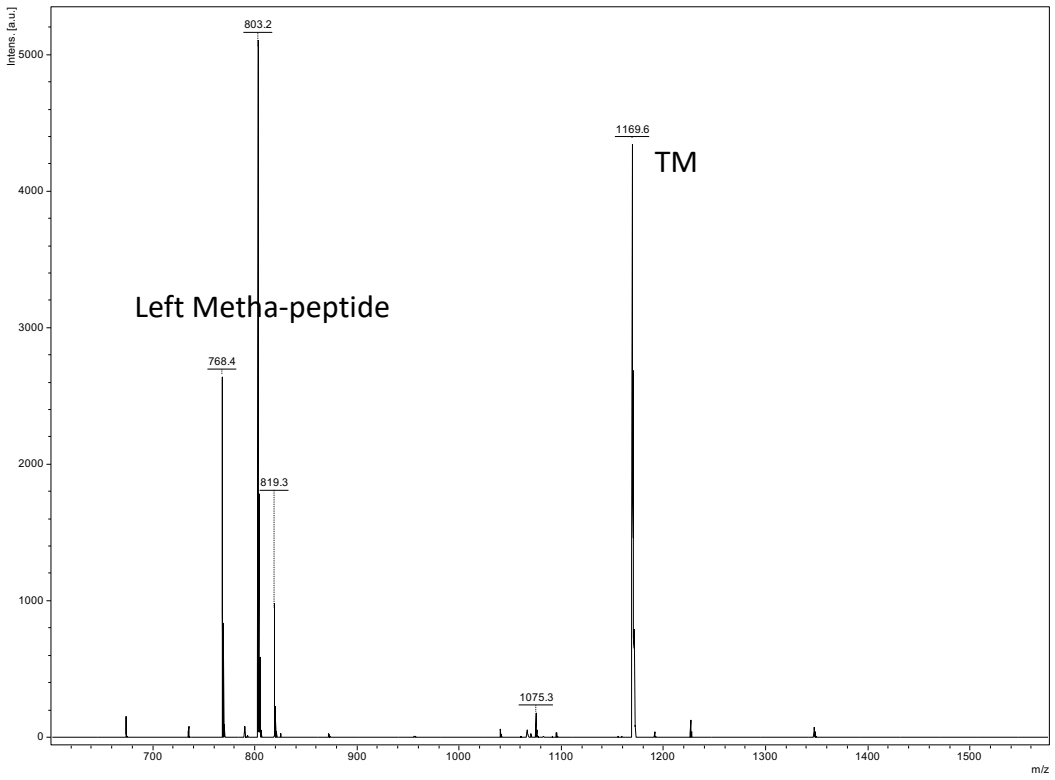
(A)


Intens. [a.u.]
TM
1018.6
728.5
784.5
964.4
1040.6
700.5
1056.5
750.5
946.6
872.6
806.5
685.4
887.6
986.5
1075.6
770.5
919.5
841.5
1094.5
m/z

(B)

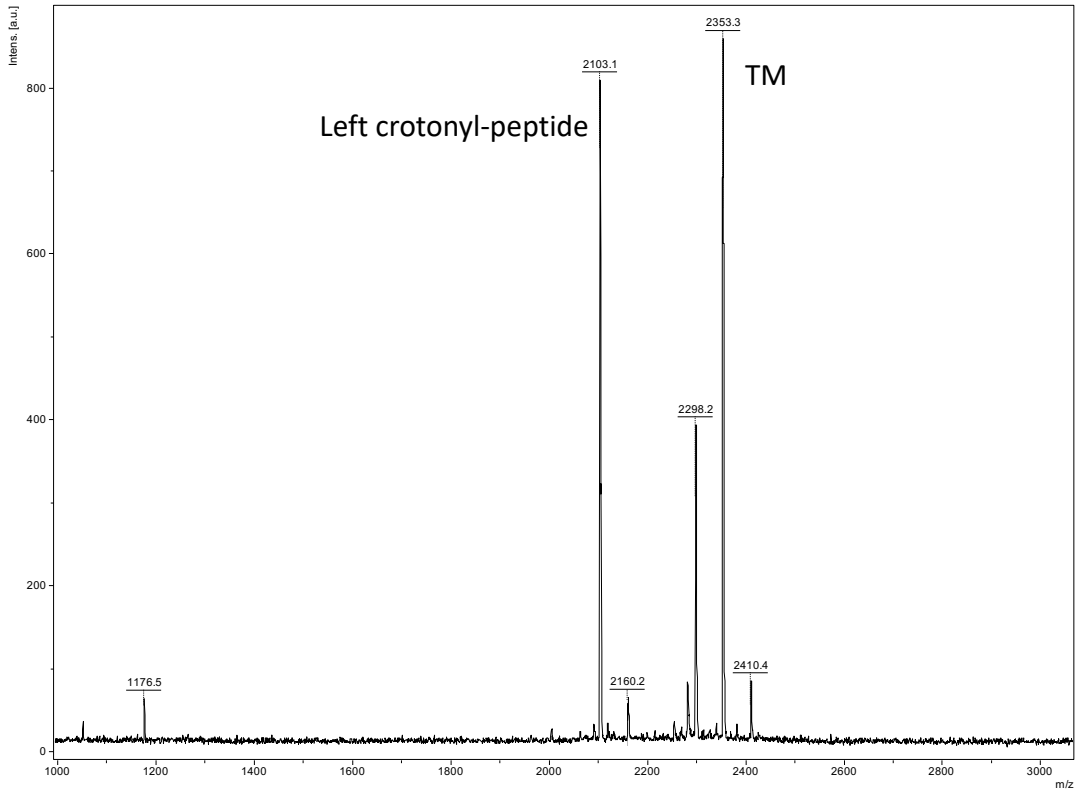

Intens. [a.u.]
1112.6
1102.6
TM
Left Metha-peptide
768.5
1058.6
1048.6
784.5
735.5
717.3
872.6
910.4
940.4
967.5
833.4
849.4
957.5
977.4
1003.6
1031.1
1088.6
1134.6
1159.7
m/z

(C)

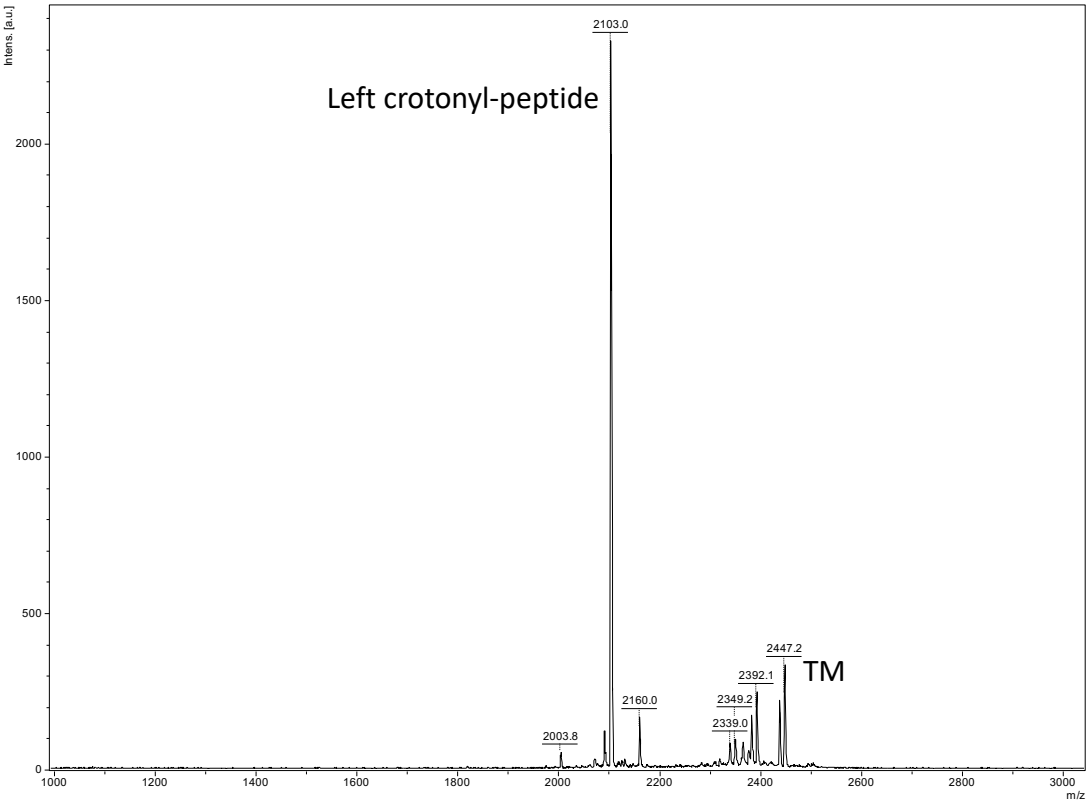


Supplementary Figure S3.3 MALDI-mass spectra of the reaction mixtures between TCEP analogs (4-6) and methacrylylated peptide h4(2-8)k5metha. (A) 1 mM methacrylated peptide (h4(2-8)k5metha) reacted with 4 mM probe **4** at 37℃ in 100mM Tris buffer (pH 8) for 16 hrs. (B) 1 mM methacrylated peptide (h4(2-8)k5metha) reacted with 4 mM probe **5** at 37℃ in 100mM Tris buffer (pH 8) for 16 hrs. (C) 1 mM methacrylated peptide (h4(2-8)k5metha) reacted with 4 mM probe **6** at 37℃ in 100mM Tris buffer (pH 8) for 16 hrs.

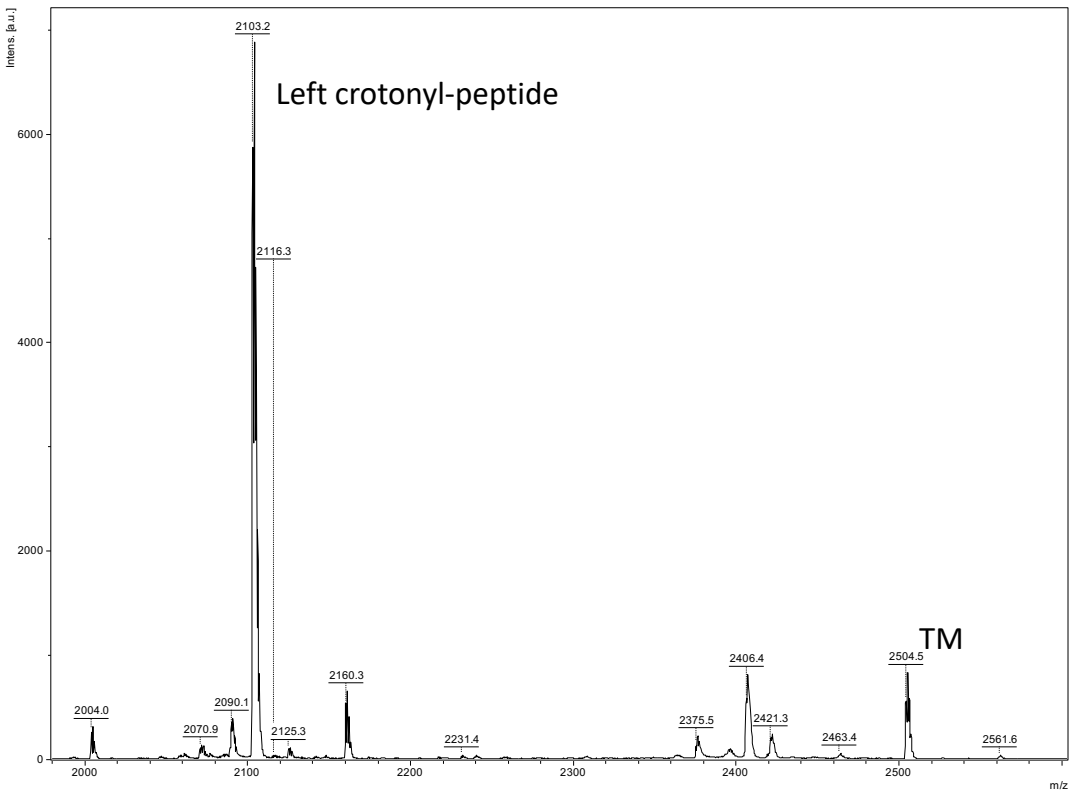
(A)


Intens. [a.u.]
2353.3
2103.1
TM
Left crotonyl-peptide
2298.2
1176.5
2160.2
2410.4
m/z

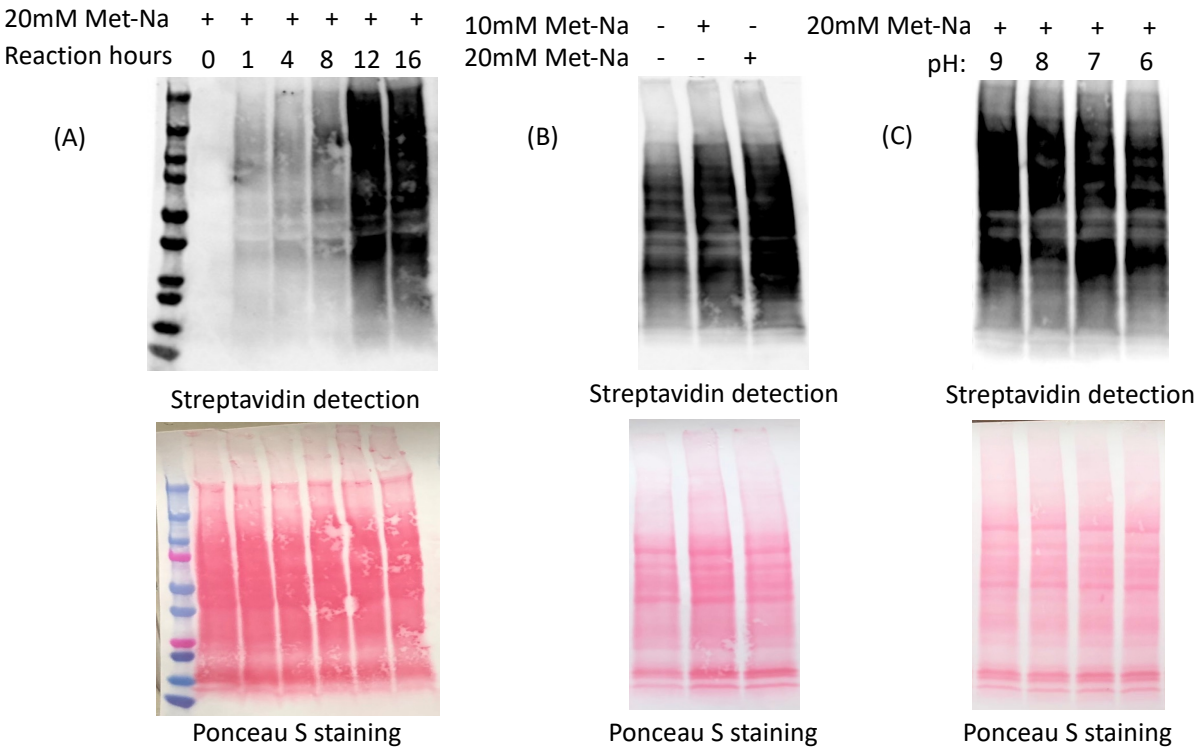
(B)


Intens. [a.u.]
2103.0
Left crotonyl-peptide
2447.2
2392.1
TM
2160.0
2349.2
2339.0
2003.8
m/z

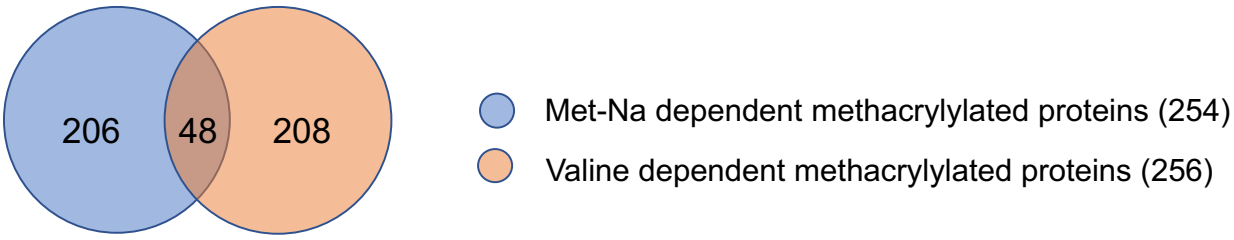
(C)



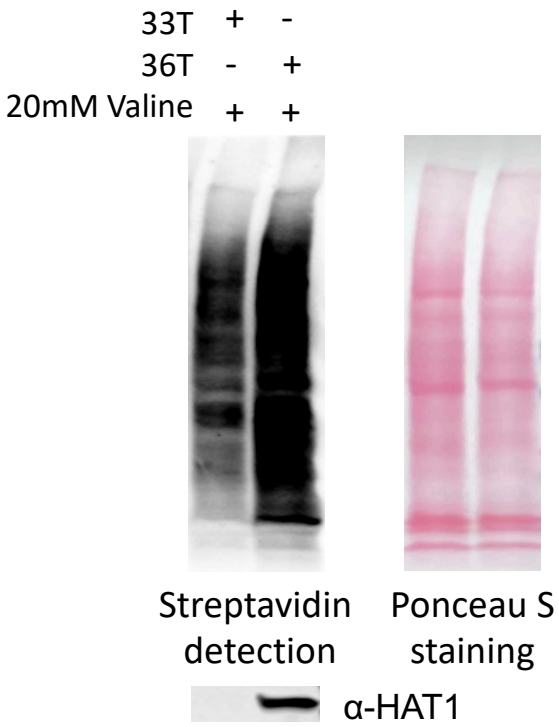
Supplementary Figure S3.4 MALDI-mass spectra of the reaction mixtures between TCEP analogs (4-6) and crotonylated peptide h4(1-20)k5cro. (A) 1 mM crotonylated peptide h4(1-20)k5cro reacted with 4 mM probe **4** at 37°C in 100mM Tris buffer (pH 8) for 16 hrs. (B) 1 mM crotonylated peptide h4(1-20)k5cro reacted with 4 mM probe **5** at 37°C in 100mM Tris buffer (pH 8) for 16 hrs. (C) 1 mM crotonylated peptide h4(1-20)k5cro reacted with 4 mM probe **6** at 37°C in 100mM Tris buffer (pH 8) for 16 hrs.



Supplementary Figure S3.5 Study the labeling activity of Gly-TCEP-AK with cellular methacrylylated proteins. (A) Detection of methacrylylated protein by Gly-TCEP-AK in a time dependent manner. HEK293T cells were incubated with 20 mM sodium methacrylate for 24 hrs and the lysate proteins were reacted with 4 mM Gly-TCEP-AK with varied time, followed by click reactions. (B) Dose dependent sodium methacrylate treatment to HEK293T cells for protein methaccrylylation. HEK293T cells were incubated with 10 and 20 mM sodium methacrylate for 24 hrs and the extracted proteins were labeled by Gly-TCEP-AK. (C) Detection of the pH tolerance of the reaction between Gly-TCEP-AK and methacrylylated proteins. HEK293T cells were incubated with 20 mM sodium methacrylate for 24 hrs and the cell lysate proteins were then reacted with Gly-TCEP-AK in varied pH.

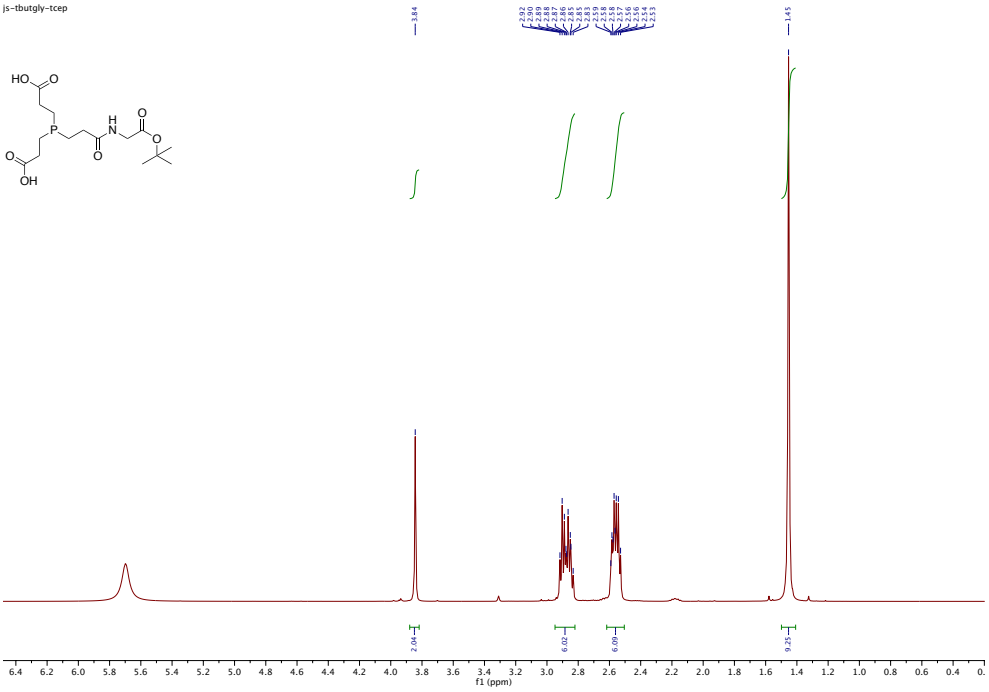


Supplementary Figure S3.6 Overlapped proteins for Met-Na dependent and valine dependent methacrylylated proteins.

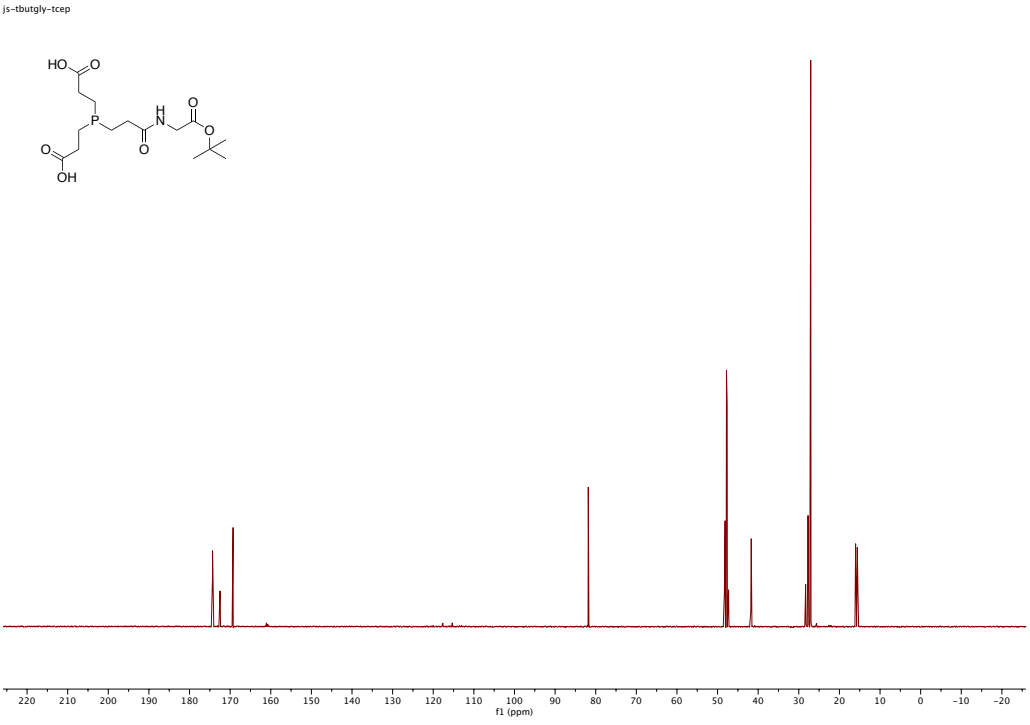


Supplementary Figure S3.7 Detection of valine dependent methacrylylation in MEF cells. 33T and 36T cells were incubated with 20 mM valine for 24 hrs. The whole cellular proteins were labeled with Gly-TCEP-AK. HAT1 expression level was determined by anti-HAT1 antibody.

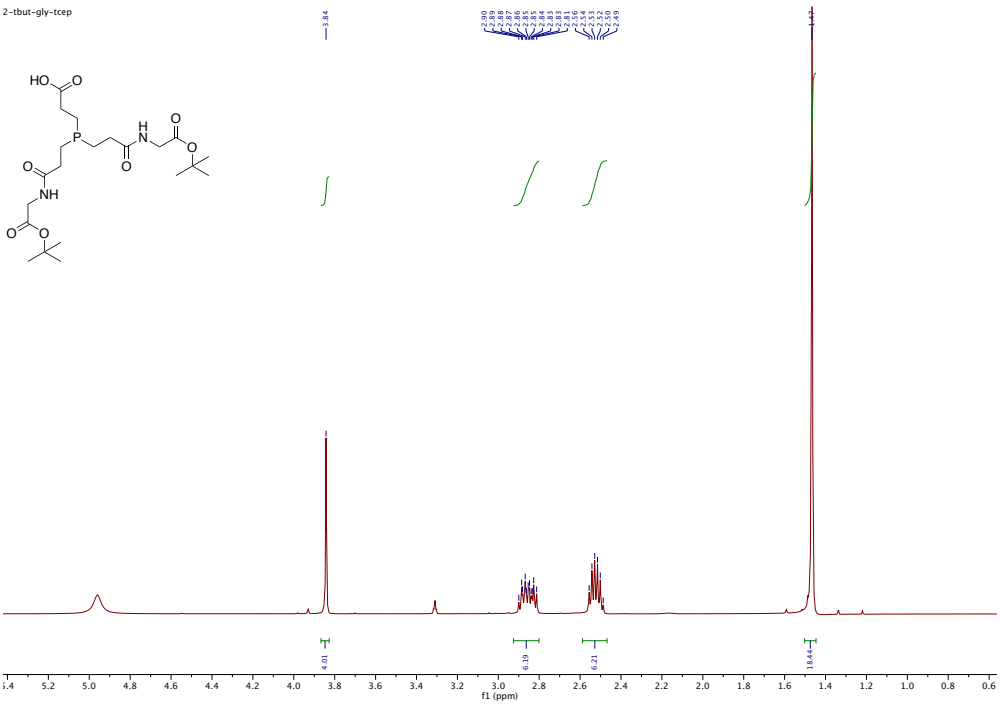
^{1}H-NMR (MeOD, 500 MHz) **1**



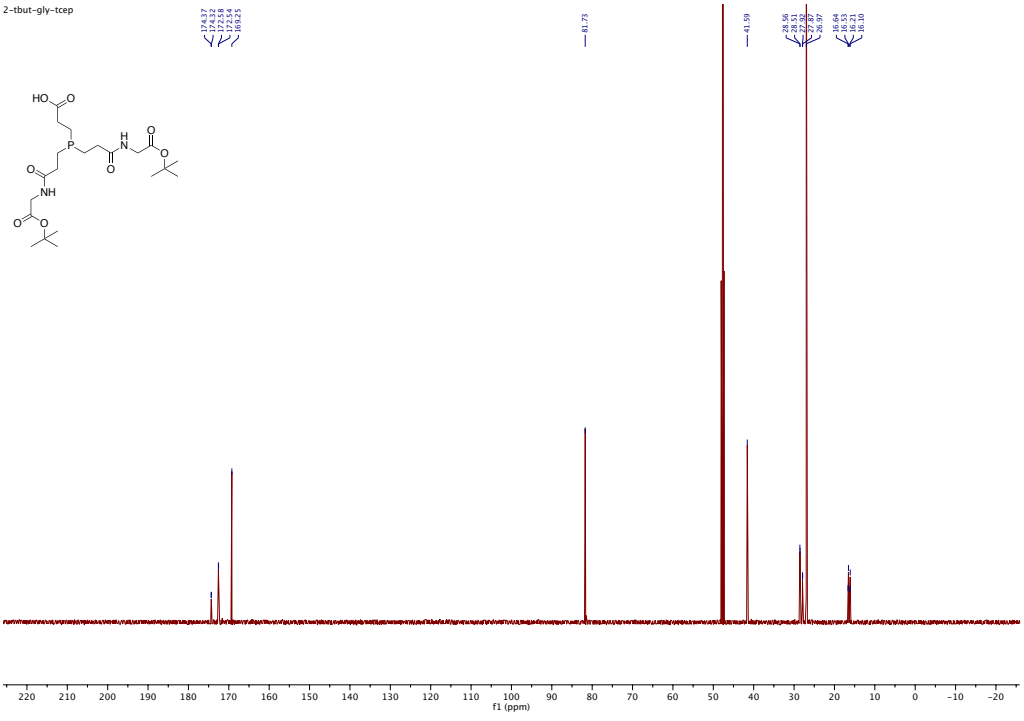
^{13}C-NMR (MeOD, 500 MHz) **1**



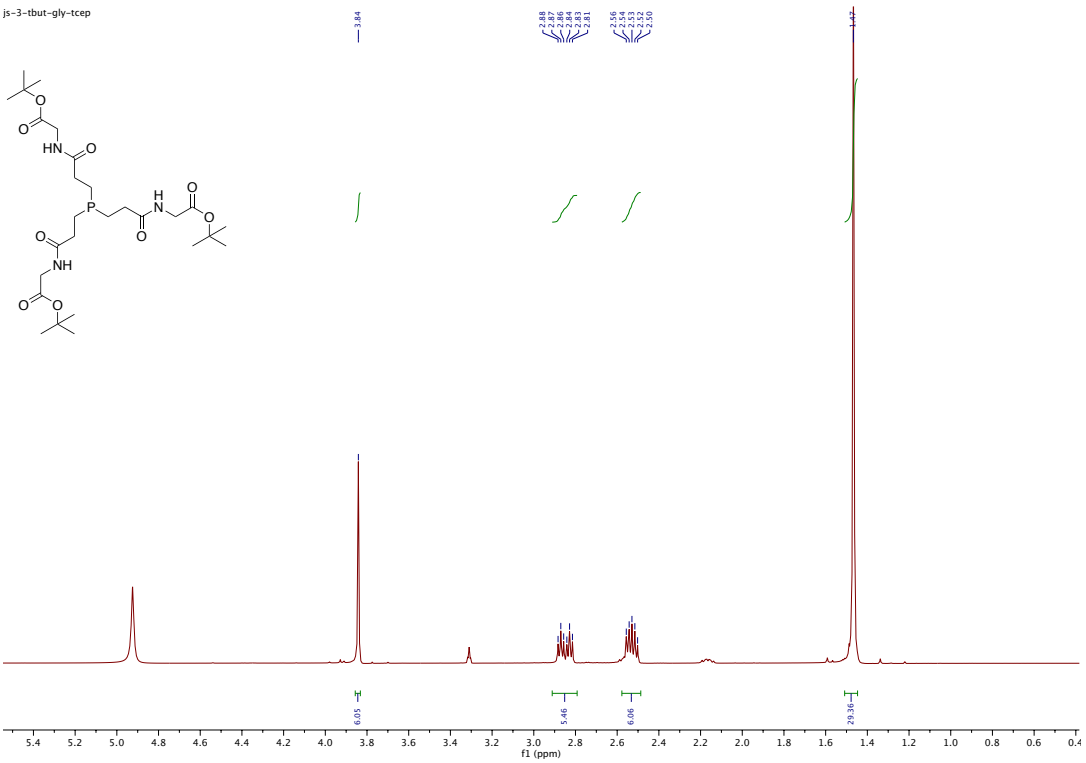
^{1}H-NMR (MeOD, 500 MHz) **2**



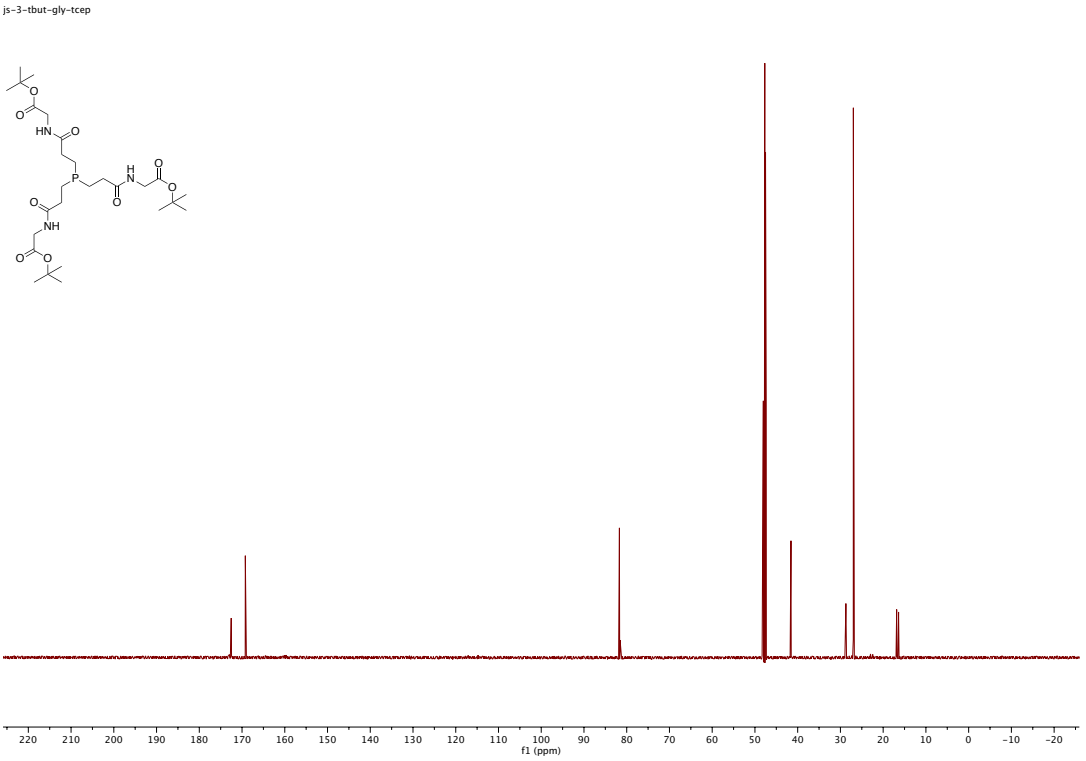
^{13}C-NMR (MeOD, 500 MHz) **2**



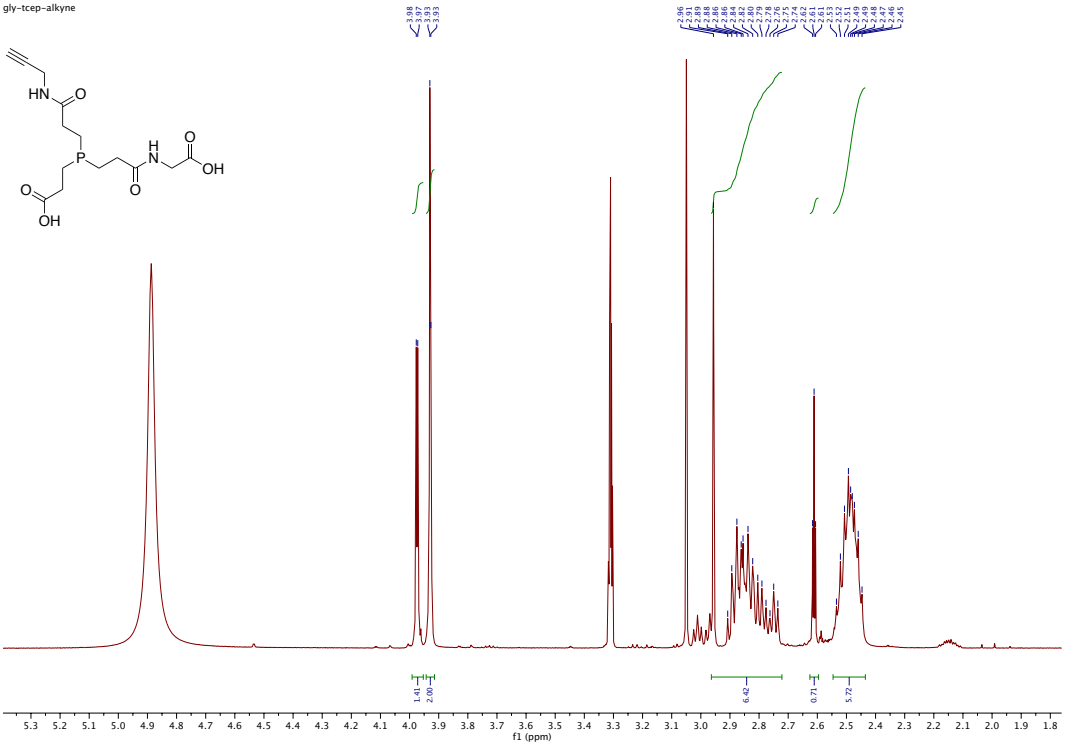
^{1}H-NMR (MeOD, 500 MHz) **3**



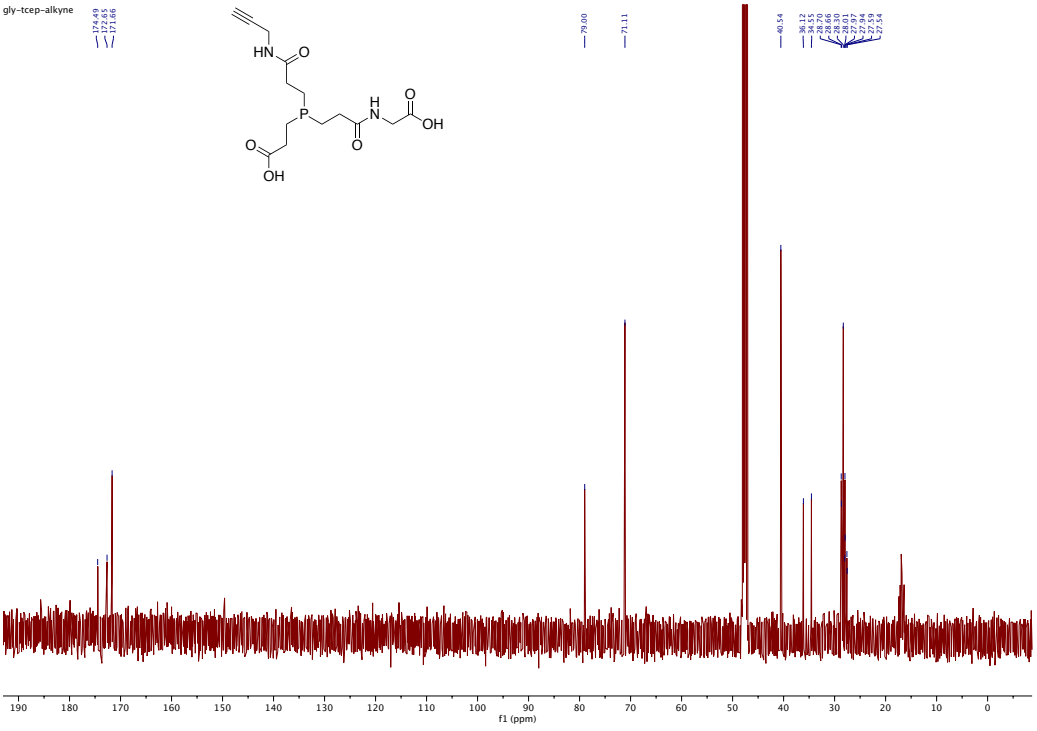
^{13}C-NMR (MeOD, 500 MHz) **3**



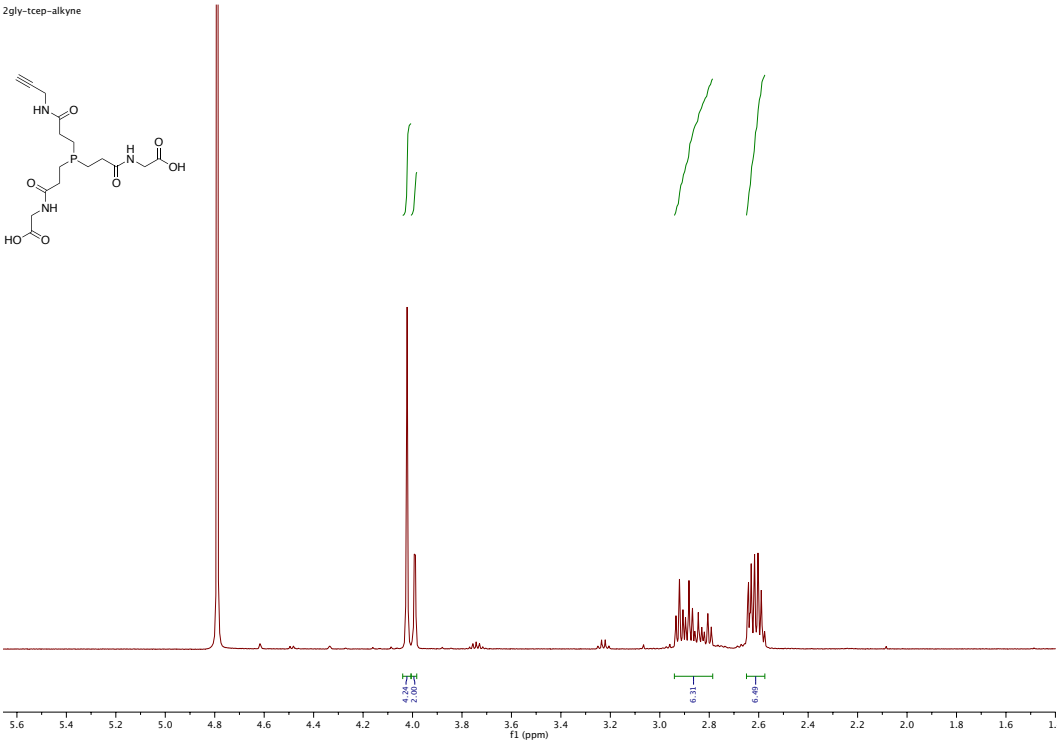
^{1}H-NMR (MeOD, 500 MHz) **5**



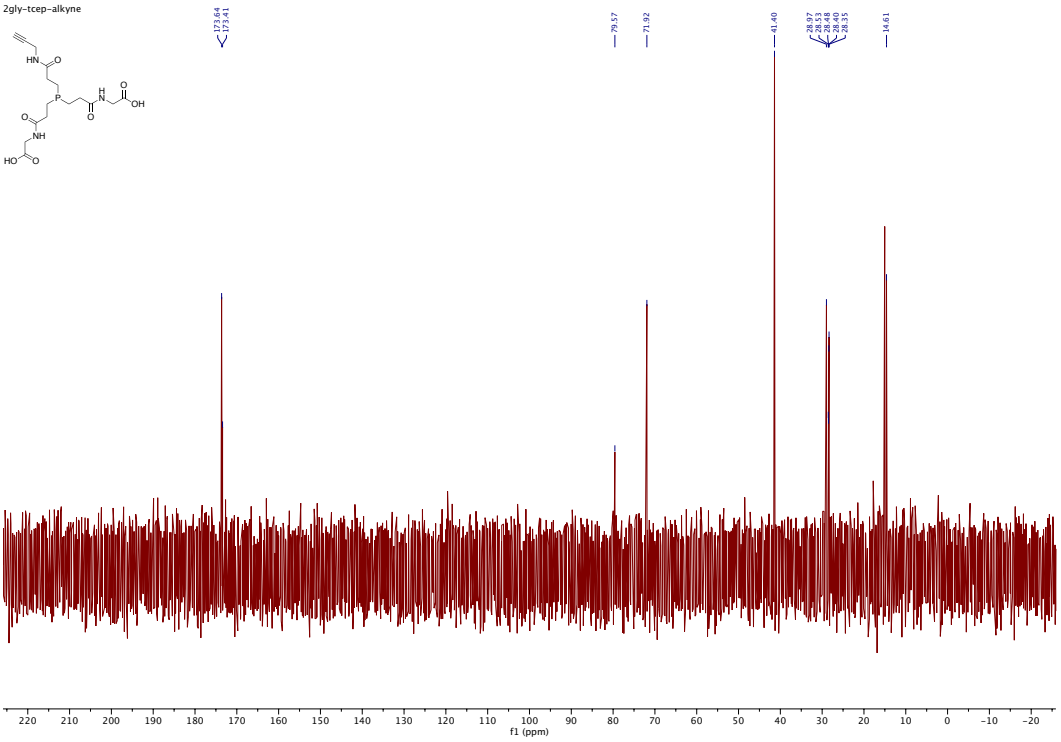
^{13}C-NMR (MeOD, 500 MHz) **5**



^{1}H-NMR (MeOD, 500 MHz) **6**

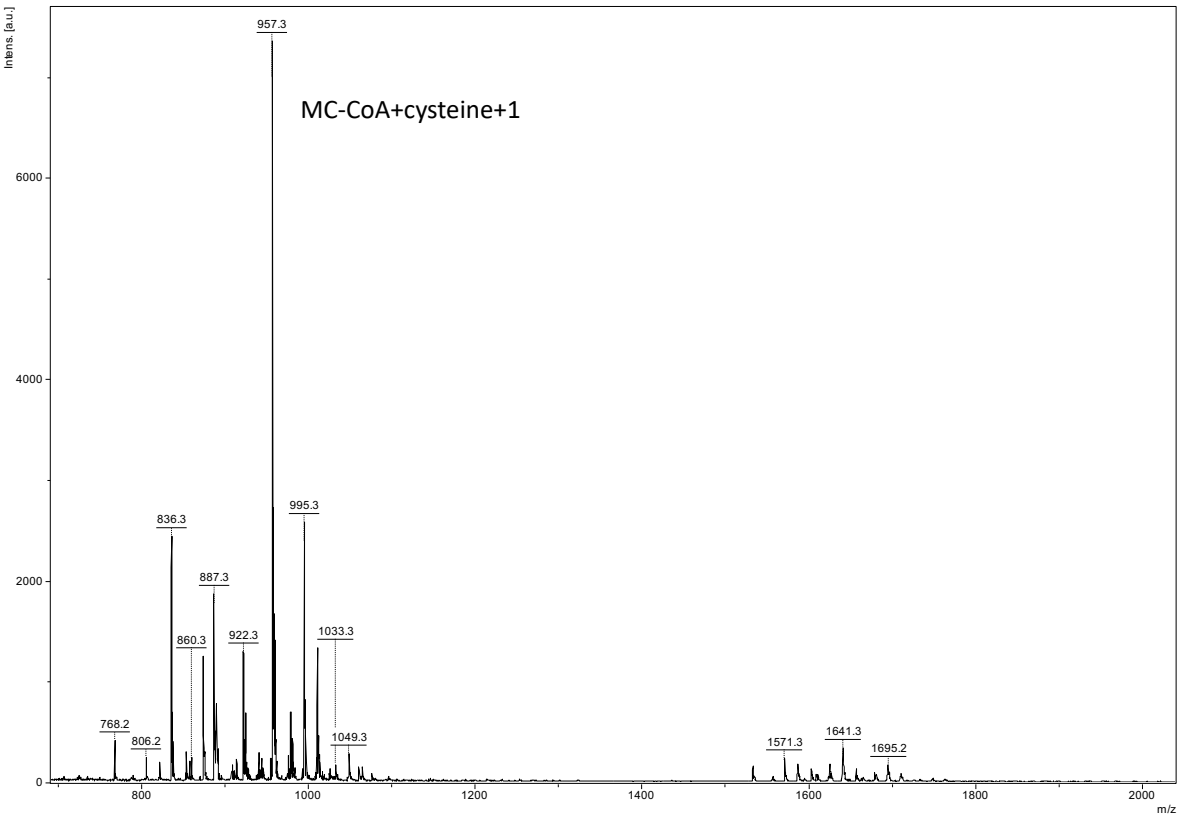


^{13}C-NMR (MeOD, 500 MHz) **6**

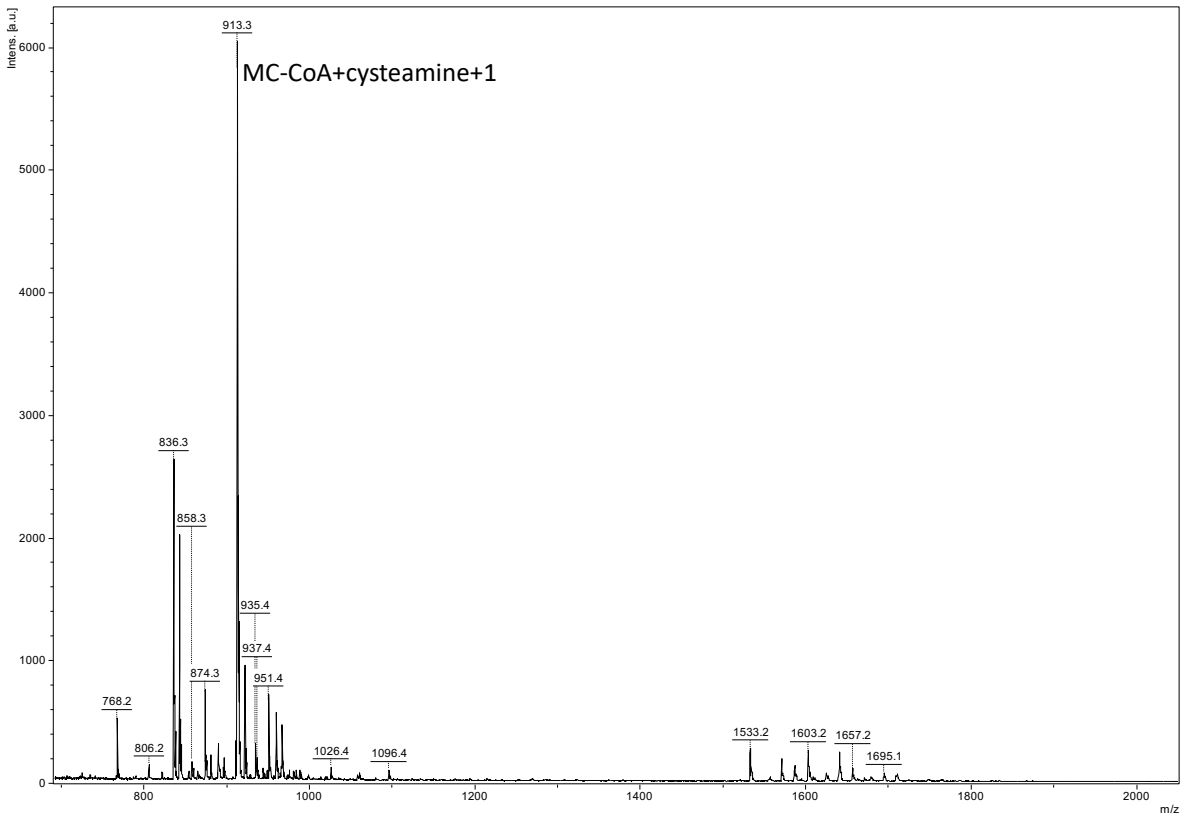


C Supporting information for Chapter 4

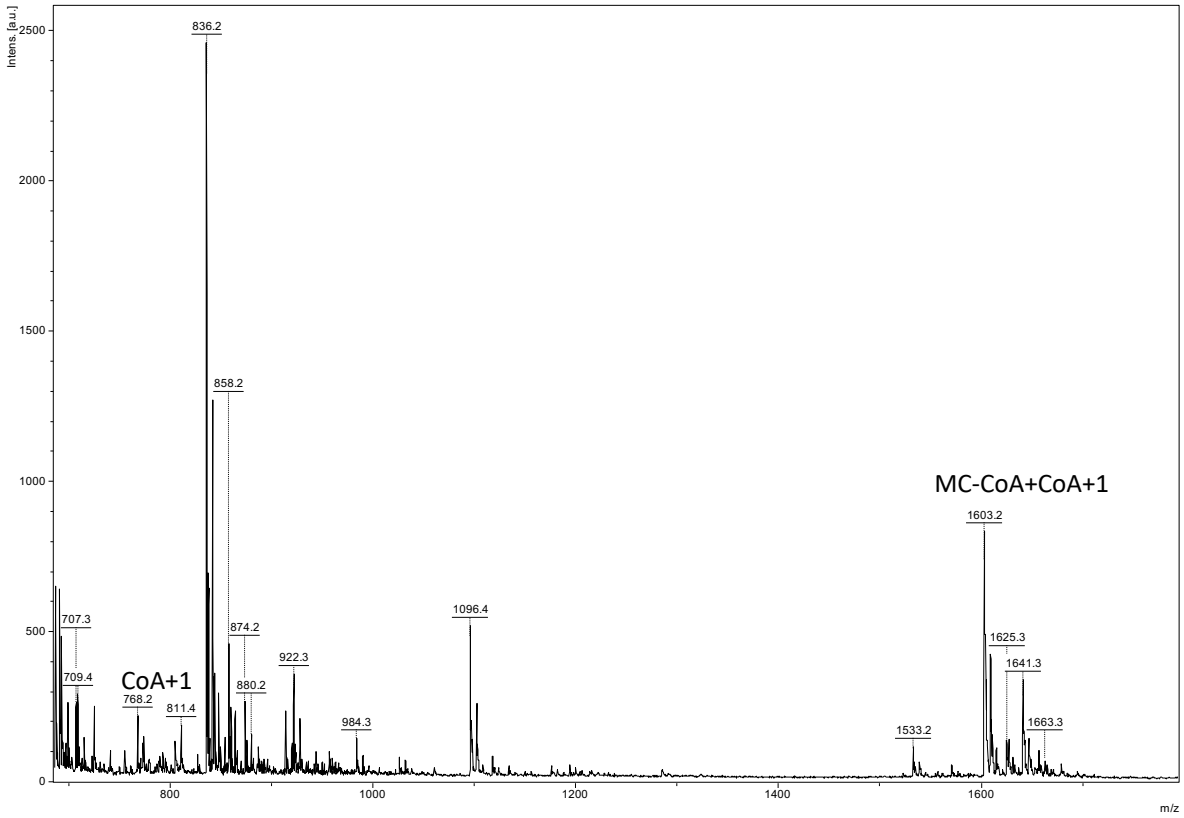
(A)



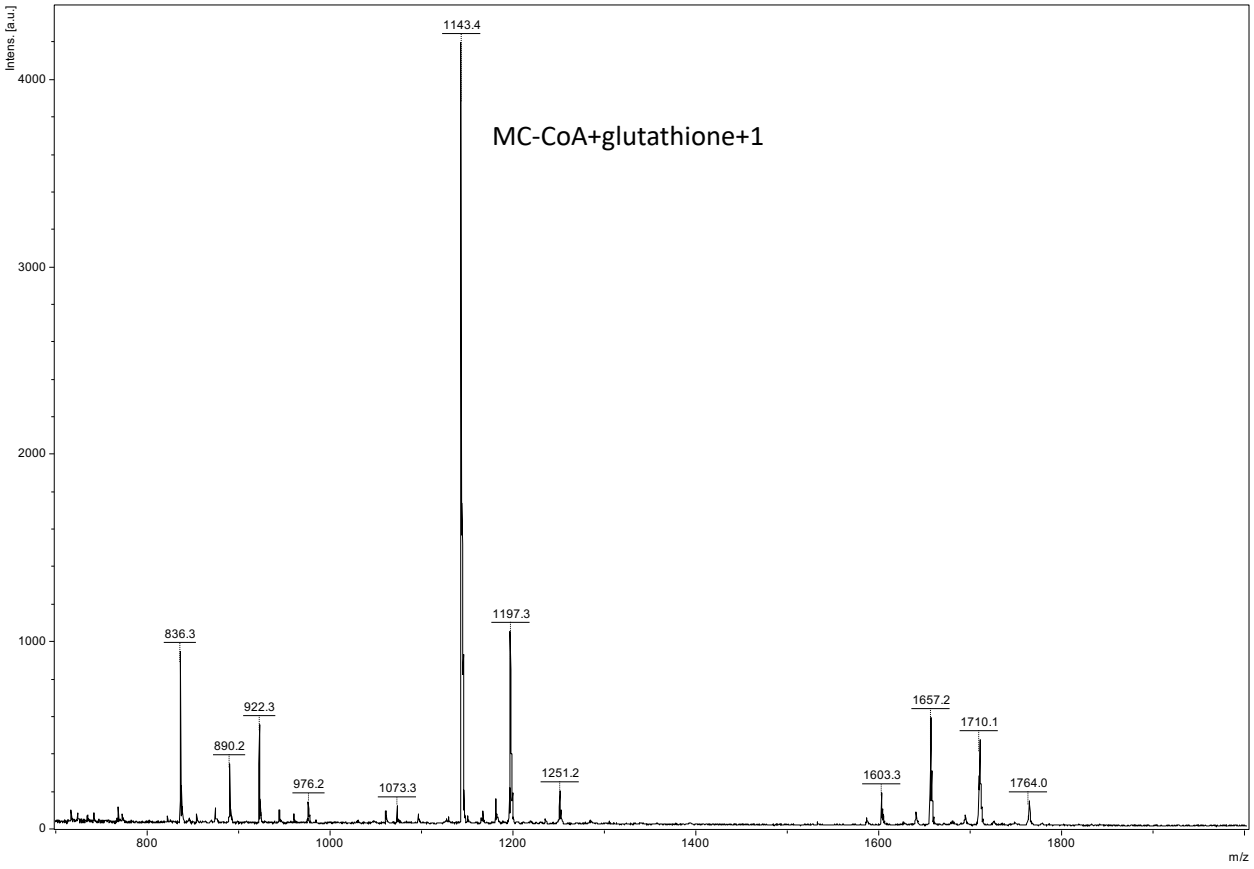
(B)


Intens. [a.u.]
MC-CoA+cysteamine+1
913.3
836.3
858.3
935.4
937.4
874.3
951.4
768.2
806.2
1026.4
1096.4
1533.2
1603.2
1657.2
1695.1
m/z

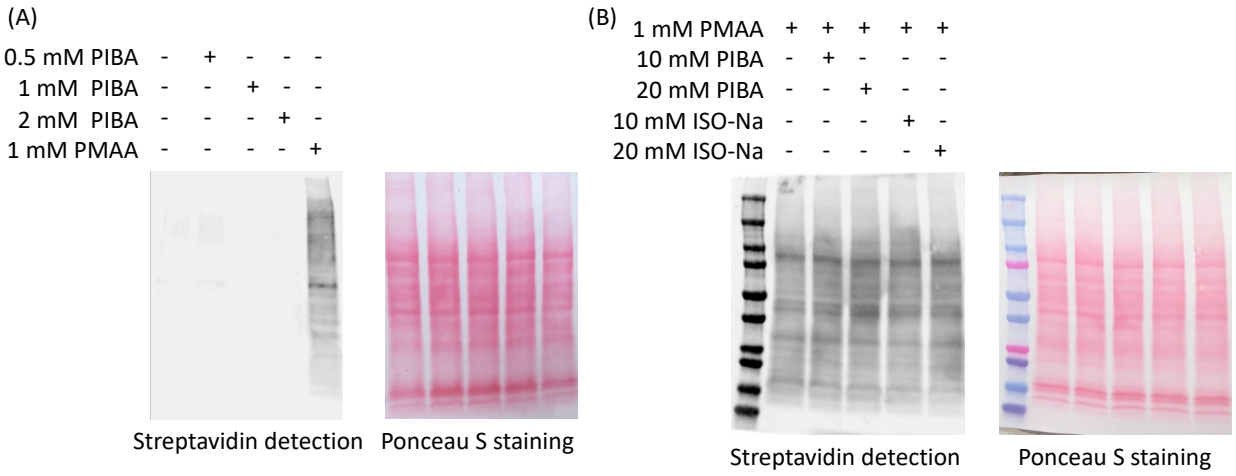
(C)


Intens. [a.u.]
MC-CoA+CoA+1
836.2
858.2
1603.2
1096.4
707.3
874.2
922.3
1625.3
1641.3
709.4
CoA+1
768.2
880.2
811.4
984.3
1533.2
1663.3
m/z

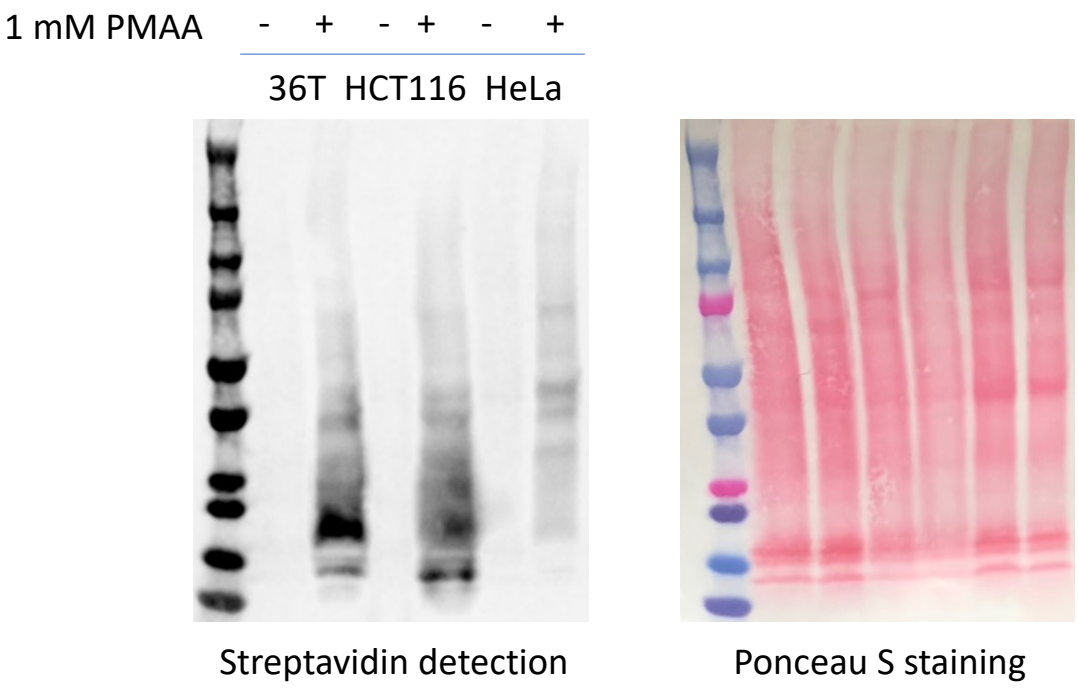
Supplementary Figure S4.1 Mass spectra of MC-CoA modified by cysteine, cysteamine, and coenzyme A. (A) 1 mM MC-CoA reacted with 1 mM cysteine at 37 ℃ for 4 h. (B) 1mM MC-CoA reacted with 1mM cysteamine at 37 ℃ for 4 h. (C) 1mM MC-CoA reacted with 1mM coenzyme A at 37 ℃ for 4 h.



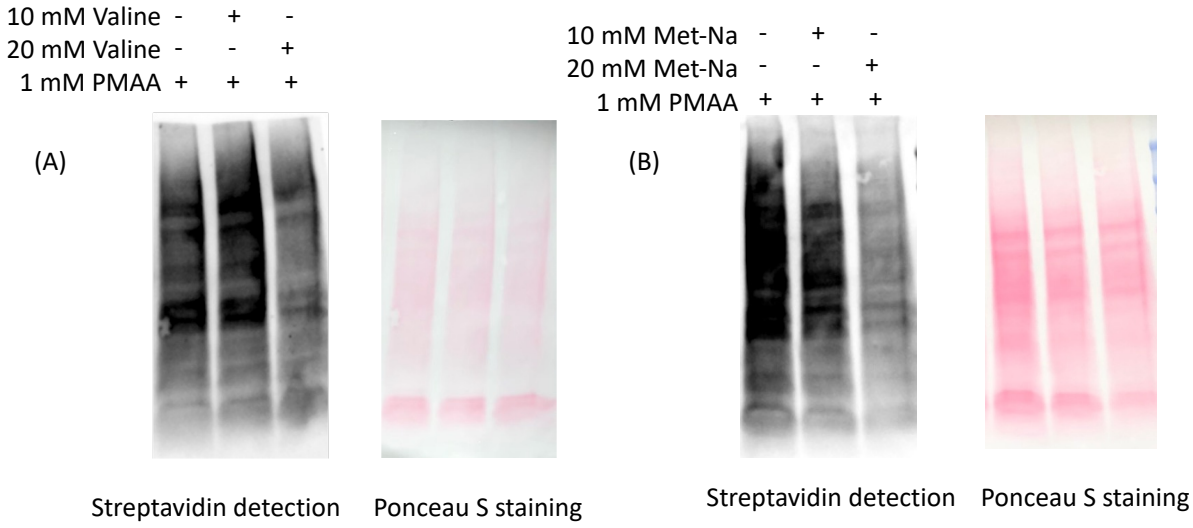
Supplementary Figure S4.2 Mass spectra of MC-CoA modified by a cysteine-containing peptide, glutathione. 1mM MC-CoA reacted with 1mM glutathione at 37 ℃ for 4 h.



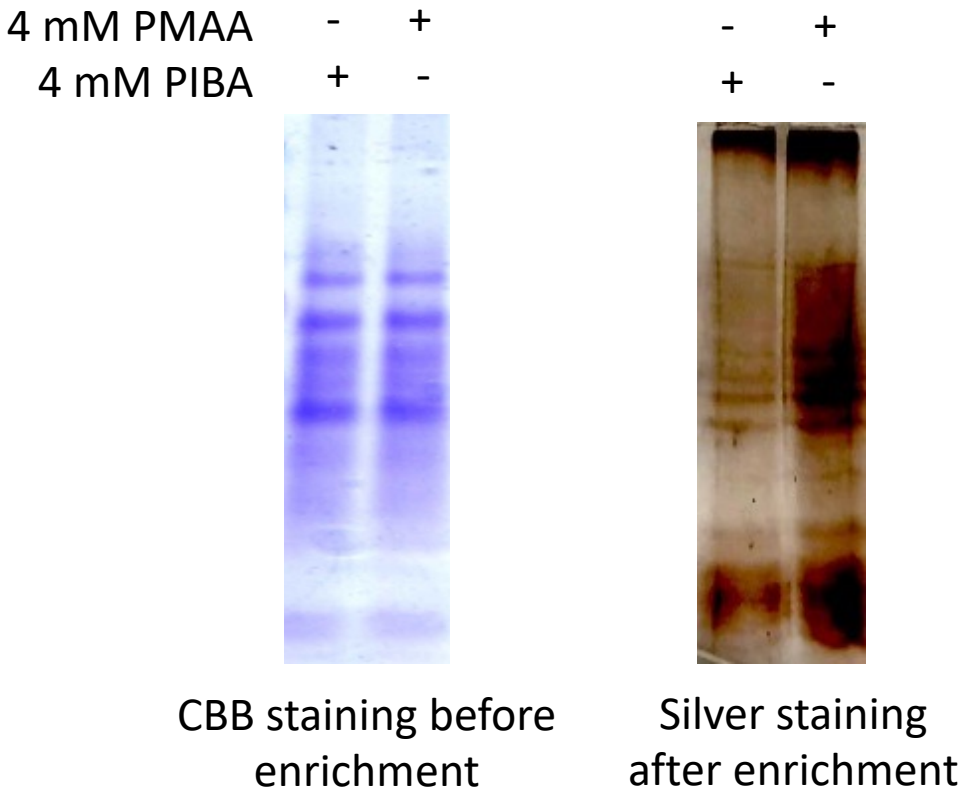
Supplementary Figure S4.3 Competitive labeling of PMAA by PIBA and sodium isobutyrate. (A) The HEK293T cell lysate protein was incubated with indicated concentration of the probe PIBA or PMAA at 37 ℃ for 4 h. (B) The HEK293T cell lysate protein was pretreated with PIBA or sodium isobutyrate for 2 h and then incubated with 1 mM PMAA at 37 ℃ for another 4 h.



Supplementary Figure S4.4 Evaluation of the probe PMAA labeling in different cell lines. The cellular protein extracted from 36T, HCT116 and Hela cells were reacted with 1 mM PMAA at 37 ℃ for 12 h.



Supplementary Figure S4.5 Competitive labeling of PMAA by the endogenous *S*-2-carboxypropylation. (A) HEK293T cells were incubated with 10 mM or 20 mM valine at 37 ℃ for 24 h, and then the cellular proteins were reacted with 1 mM PMAA at 37 ℃ for 12 h. (B) HEK293T cells were incubated with 10 mM or 20 mM sodium methacrylate at 37 ℃ for 24 h, and then the cellular protein were reacted with 1 mM PMAA at 37 ℃ for 12 h.



Supplementary Figure S4.6 Enrichment of the PMAA labeled protein. The HEK293T cellular proteins were reacted with 4 mM PMAA or 4 mM PIBA at 37 ℃ for 16 h, and then the labeled proteins were further reacted with azide-diazo-biotin and pull-down by streptavidin beads. The eluted proteins were then imaged by silver staining.