INVESTIGATION OF ACTIVITIES AND FUNCTIONS OF THE LYSINE ACETYLTRANSFERASES

by

ZHEN HAN

(Under the Direction of Y. George Zheng)

ABSTRACT

Chemically diverse acylations on protein lysines have emerged as important posttranslational modifications (PTMs) that regulate many cellular processes such as gene transcription, cell cycle, and apoptosis. Lysine acylations are driven by lysine acetyltransferases (KATs) that can covalently deposit acyl group to the ε-amino group of lysine residues from acyl-CoA molecules. Many studies have demonstrated that dysregulation of lysine acylation due to the dysfunction of KATs or abnormal fluctuation of acyl-CoA level may lead to the occurrence and progression of various diseases. Elucidation of the regulatory mechanisms of lysine acylations and their biological outcomes has profound significance to understand the pathophysiological mechanisms of the related diseases and develop effective therapeutic approaches towards them.

Here, we report our efforts in the studies of novel functions and activities of KAT enzymes. We designed and synthesized an acetyl-CoA surrogate 3-azidopropionyl-CoA (3AZ-CoA) that can specifically react with the wild type KAT p300 and the mutant KAT GCN5-T612G. Based on the application of 3AZ-CoA, we developed an activity based protein profiling (ABPP) approach for valid profiling of KAT sub-acylome and identified hundreds of substrates for p300

and GCN5 KAT enzymes. By welding the application of 3AZ-CoA with fluorescence resonance energy transfer (FRET), we designed and validated a mix-and-read method for KAT activity measurement. This method can directly quantify the production of acylated protein in a fast, sensitive, and accurate way. We also discovered the novel activity of KAT enzymes from the MYST family as lysine propionyltransferases. The western blot and proteomic data suggested that one of the MYST members, MOF is able to propionylate both histone and non-histone proteins and has a partially distinct propionylome than other lysine propionyltransferases. The crystal structure of MOF bound with propionyl-CoA provides structural insights of how MYST KATs carry out lysine propionylation. Last but not least, we defined lysine isobutyrylation as a novel PTM on histone lysines and HAT1 as a potential isobutyryltransferase. Together, the present work provides valuable insights in understanding the regulatory mechanism of lysine acylations and the functions of KATs.

INDEX WORDS: Lysine acetyltransferase, Lysine acylation, Activity based protein profiling, Click chemistry, Lysine propionylation, Lysine isobutyrylation

INVESTIGATION OF ACTIVITIES AND FUNCTIONS OF THE LYSINE ACETYLTRANSFERASES

by

ZHEN HAN

B.S., Shandong University, China, 2012

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

2017

INVESTIGATION OF ACTIVITIES AND FUNCTIONS OF THE LYSINE ACETYLTRANSFERASES

by

ZHEN HAN

Major Professor:	Y. George Zheng
Committee:	Author Roberts
	Eileen J. Kennedy
	Houjian Cai
	Vladimir Popik

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

DEDICATION

I would like to dedicate this dissertation to my parents and my girlfriend Silei Li, who gave me great spiritual and financial supports along my Ph.D. study. You are and will always be the most important people in my life.

ACKNOWLEDGEMENTS

Firstly, I would like to express my deepest appreciation to my advisor, Dr. Y. George Zheng. Dr. Zheng has given me so many academic guidance, encouragement, and support through my Ph.D. study. I have learned a lot from him, not only the way to intelligently pursue scientific research but also his excellent work ethnics, which largely inspired me to keep exploring science. I would like to acknowledge my dissertation committee members, Dr. Author Roberts, Dr. Eileen Kennedy, Dr. Houjian Cai, and Dr. Vladimir Popik for their full support and guidance in my research. I am also thankful to many faculty members and staffs who helped me in my Ph.D. study.

I want to express my appreciation to my lab mates for their support and help in the past several years. I want to thank Dr. Maomao He, Dr. Yepeng Luan, and Liza Ngo for their generous help on my projects. I am also grateful to Dr. Jing Zhang, Dr. Hao Hu, Kun Qian, Hui Xu, Melody Fulton, Mitchell Wolfe, Zhesi Zhu, Jiabao Song and all other group members for their discussions and helps in lab research.

Last but not least, I would like to thank National Institutes of Health, American Heart Association, and National Science Foundation for the funding support on my research. I would like to thank the Chinese Scholarship Council for the financial support on my daily life.

TABLE OF CONTENTS

CHAPTER 1 Introduction and Literature Review

Posttranslational modifications (PTMs) in both prokaryotes and eukaryotes such as protein phosphorylation, acetylation, methylation, and ubiquitination diversify protein functions and activities by altering protein stability, localization, and protein-protein interaction, etc.[1-3] Lysine acetylation is an essential PTM that is associated with many cellular processes such as transcription activation, cell cycle, signal transduction, and cellular metabolism.[4-7] Lysine acetylation was first identified in 1964 by Allfrey et al., who demonstrated that lysine acetylation alleviated the inhibitory effects of histone proteins on RNA synthesis.[5] Since then, several mechanistic studies revealed that lysine acetylation activates gene transcription through two mechanisms: First, histone acetylation neutralizes the positive charges on histone lysine residues, which weakens the charge-charge interaction between lysine residues and nucleotides. This leads to the disassembly of the highly compacted nucleosomal structure, which enables the binding of transcription activators to the unwind DNA and initiation of gene transcription.[8] Second, lysine acetylation alters chromatin structure and recruits reader proteins that specifically recognize acetylated histones via specialized structural folds such as bromodomain proteins.[9] With one or two mechanisms working individually or together, histone lysine acetylation was known as an important mechanism to regulate chromatin dynamics and gene transcription **(Figure 1.1)**.

Lysine acetylation is driven by histone/lysine acetyltransferases (HATs/KATs), also known as the "writers", which transfer the acetyl group from acetyl-Coenzyme A (Ac-CoA) to the ε-amino group of lysines with the formation of an amide bond. This process can be reversed by the "erasers", histone/lysine deacetylases (HDACs/KDACs) **(Figure 1.2)**. Since the first KAT

was discovered in 1995, more than ten KAT enzymes were identified with *bona fide* lysine acetyltransferase activity in the past decades. According to the amino acid sequence and domain organization, KAT enzymes are divided into three major families including the MYST family, the GCN5/PCAF family, and the p300/CBP family **(Table 1.1)**.[10] The discovery and further functional investigation of KATs, in turn, facilitated the study of the biological functions of lysine acetylation. Particularly, lysine acetylation not only regulates chromatin dynamics through histone modification but also ubiquitously exist in non-histone proteins and is related with various biological processes beyond the chromatin realm.[6,11,12] The homeostasis of lysine acetylation is co-regulated by KATs and KDACs and dysfunction of these enzymes due to genetic deficiency or overexpression is tightly associated with multiple diseased forms.[7,13-15] For examples, overexpression of GCN5 has been detected in leukemic cells. GCN5 overexpression leads to hyperacetylation and stabilization of the oncogenic protein E2A-PBX1, which induces the progression of acute lymphoblastic leukemia (ALL).[16] p300 and CBP also play important roles in various cancers. It is reported that p300 gene and CBP gene are mutated in more than 85% of microsatellite instable (MSI)+ colon cancer cell lines and primary tumors.[17] In addition, p300 activity is associated with the growth and differentiation of human melanocytes and CBP can potentiate colorectal cancers by regulating the activity of the WNT/β-catenin pathway.[18] Moreover, the MYST KAT MOF can prevent diet-induced obesity in mice by activating hypothalamic polysialylation. Silencing of MOF gene would induce the increase of body weight in adult mice.[19] Although some pathological pathways have been well elucidated, the mechanistic investigation and therapeutic development targeted on KAT-related diseases are largely hindered by limited knowledge about the sub-acetylome of individual KAT enzymes and lack of efficient approaches for KAT activity measurement. To solve these problems, we

developed a bioorthogonal labeling system for KAT substrate profiling and KAT activity measurement, respectively. The detailed methodology, experimental design, and results are discussed in chapter 2 and chapter 3.

In the past decade, more than ten new acylations besides acetylation have been discovered on protein lysines due to the development of high resolution mass spectrometry. These include lysine propionylation, butyrylation, glutarylation, crotonylation, succinylation, and 2-hydroxyisobutyrylation, etc.[3,20-23] With different length and charge state on the acyl chains, different lysine acylations lead to different biological readouts.[24,25] Meanwhile, some of the KAT enzymes are reported to possess great cofactor promiscuity that they not only act as acetyltransferases but also carry out lysine propionylation, butyrylation, and crotonylation, etc.[26-30] Identification of novel acylations and novel dual-/multiple- enzymatic activity of KAT enzymes indicated more complicated mechanisms regarding pathological processes caused by dysregulation of KAT mediated acylations. In chapter 4 and chapter 5 of this dissertation, we reported the novel activity of MYST KATs as lysine propionyltransferases and a novel lysine modification isobutyrylation, respectively. These discoveries provided valuable information on KAT acylation activity.

Figures and Captions

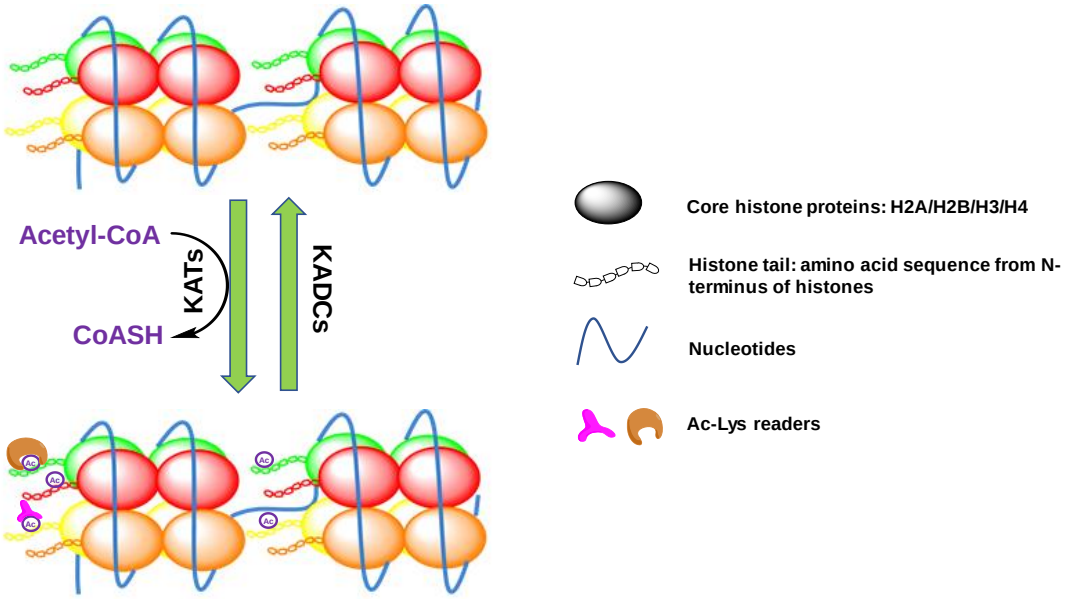


Figure 1.1. KATs and KDACs regulate gene transcription by altering histone lysine acetylation.

Lysine acetylation can activate gene transcription through two mechanisms: 1). Acetylation of histone tail lysines blocks the charge-charge interaction between lysines and nucleotides, making the DNA accessible to transcription activating proteins; 2). Acetylated lysines recruit reader proteins, which can initiate gene transcription directly or through the recruitment of other transcription activating proteins.

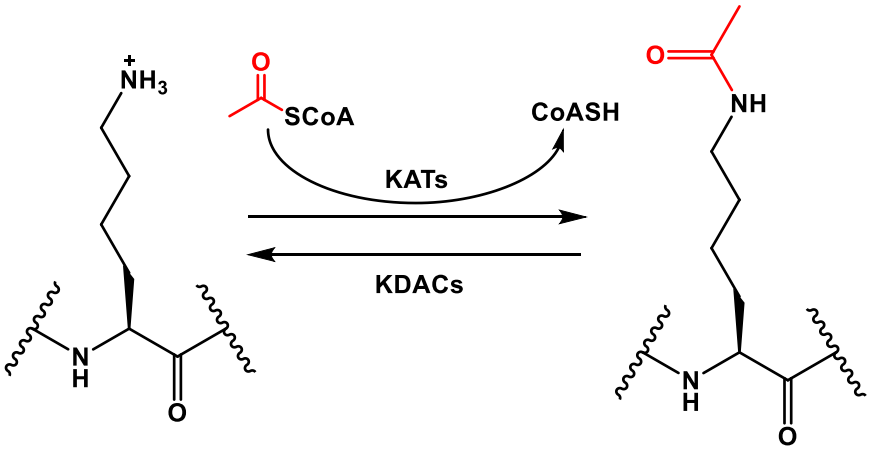


Figure 1.2. Lysine acetylation and deacetylation are driven by KATs and KDACs.

Lysine acetyltransferases (KATs) transfer the acetyl group from the cofactor acetyl-CoA to the ε-amino group of specific lysines with the formation of an amide bond. This process is reversed by lysine deacetylases (KDACs).

Table 1.1 Major KAT enzymes in eukaryotes

KAT Family	Members
MYST	MOF, Tip60, MOZ, MORF, HBO1
GCN5/PCAF	GCN5, PCAF, HAT1
p300/CBP	p300, CBP

CHAPTER 2. Profiling Cellular Substrates of Lysine Acetyltransferases GCN5 and p300 with Orthogonal Labeling and Click Chemistry

---

Han, Z., Chou, C., Yang X., Michael, G.B., and Zheng, G.Y. 2017. *ACS Chemical Biology*. 12 (6), 1547–1555.

Abstract

p300 and GCN5 are two representative histone/lysine acetyltransferase (HAT/KAT) enzymes in mammalian cells. It was recently reported that they possess multiple acyltransferase activities including acetylation, propionylation, and butyrylation of the ε-amino group of lysine residues of histone and non-histone protein substrates. Although thousands of acetylated substrates and acetylation sites have been identified by mass spectrometry-based proteomic screening, our knowledge about protein acylation, especially the causative connection between individual KAT members and their substrates remain limited. Herein, we applied a chemical probe 3-azidopropionyl CoA (3AZ-CoA) as a surrogate of acetyl-, propionyl- and butyryl-CoA for KAT substrate identification. We successfully attached the azide tag to cellular substrates of wild type p300 and engineered GCN5; subsequently biotin affinity tag was attached to labeled substrates through copper-catalyzed azide-alkyne cycloaddition (CuAAC). After 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. These proteins are either p300- or GCN5-unique or shared by the two KATs, and are extensively involved in various biological events such as gene expression, cellular metabolism, influenza infection, and cell cycle, etc. We also experimentally validated two novel substrates of GCN5, i.e. IQGAP1 and SMC1. These results demonstrate extensive engagement of GCN5 and p300 in cellular pathways and are highly valuable to understand their functions in particular biological processes.

Introduction

Acetylation of the epsilon amino group of lysine residues is one of the most important posttranslational modifications that diversify protein functions by changing protein stability, nucleic acid binding, protein-protein interaction, etc.[8,31] This process is mediated by histone

acetyltransferases (HATs, also known as lysine acetyltransferases, KATs) which were first identified in 1995.[32] In the past twenty years, more than ten KATs have been discovered with *bona fide* activities and many of them own significant disease relevance. Acetylation of histones and transcription regulatory proteins by KATs has been extensively studied with classic biological and biochemical techniques such as immunoblotting, chromatin immunoprecipitation, and radiometric assays and has demonstrated the key regulatory roles of KATs in chromatin dynamics and transcriptional activation.[8,33] Recently, high resolution mass spectrometry (MS) has resulted in the identification of thousands of acetylated proteins and lysine residues, suggesting that the effects of lysine acetylation and KATs go far beyond the chromatin kingdom.[34-36] Despite the progress made on KAT functional studies and acetylation substrate identification, there remains a missing link connecting the compositions of the cellular acetylome to different KAT enzyme members. Moreover, recent studies have demonstrated that some KATs not only have acetyltransferase activity but also propionyl- or butyryl-transferase activities, implicating even broader functions of KATs.[26-28] Therefore, methods that can efficiently elucidate functions of individual KATs and their distinct sub-acylomes are greatly needed.

There are three major families of KATs grouped according to their sequence homology and domain organizations, which include the MYST family, GCN5/PCAF family and p300/CBP family. p300 and GCN5 are well studied members of KAT families. Both enzymes acetylate histone and non-histone proteins and have extensive involvement in multiple cellular processes including cell cycle, apoptosis, DNA repair, etc.[37-39] Aberrant activity induced by abnormal expression or genetic mutation of these two enzymes is related with diseases varying from inflammation to cancers.[16,40] Detailed molecular mechanisms accounting for these physiological and pathological events are largely unknown. Earlier studies clearly demonstrate that

dysregulation of the activities of various KATs leads to dynamic and functional changes to chromatins which constitutes an epigenetic etiology to human pathology. Accumulating evidence in recent years reveals that diverse non-histone acetylations also play indispensable roles in disease incidence.[16,37] Importantly, a few studies have demonstrated propionyl- and butyryltransferase activity of GCN5 and p300, which further entangles the study of GCN5 and p300 functions.[26,28] So far there are no proteomic studies that distinguish substrate profile differences between p300 and GCN5.

It has been a technical challenge to define the unique substrate profiles of a particular KAT due to the ubiquitous nature of acetylation and the redundancy of cellular KAT activities. New tools and probes are highly needed to track and identify substrates of a KAT of interest and will significantly advance this field. Recently, we and others have explored synthetic acetyl-CoA surrogates to identify KAT substrates, which provides a powerful chemical biology strategy to interrogate the acetylome of particular KATs.[35,41,42] In particular, our group created engineered KATs matching with a specific, functionalized Ac-CoA substitutes to establish bioorthogonal pairs which has great promise to label cellular substrates of any particular KATs [42]. In this study, we applied an acyl-CoA surrogate 3-azidopropionyl coenzyme A (3AZ-CoA) to profile the sub-acylomes of p300 and GCN5, respectively. By using a consecutive labeling strategy with enzymatic reaction and copper-catalyzed alkyne-azide cycloaddition (CuAAC) followed by protein enrichment and proteomic analysis, we have specifically labeled and enriched hundreds of p300 and GCN5 substrate candidates from the cellular proteomic context **(Figure 2.1)**. We have also annotated these proteins in respect to different cellular pathways, which lays a great foundation to decipher p300 and GCN5 functions in cellular biology.

Result and Discussion

3AZ-CoA is an acetyl-CoA surrogate specific to wild type p300 and GCN5-T612G mutant.

Several acyl-CoA compounds including 4-pentynoyl CoA (4PY-CoA), 5-hexynoyl CoA (5HY-CoA), 6-heptynoyl CoA (6HY-CoA) and 3-azidopropionyl CoA (3AZ-CoA) were synthesized as described previously to probe KAT activity and substrates.[42] p300 exhibited a great degree of cofactor promiscuity and can recognize some acetyl-CoA surrogates containing alkyne or azide motifs. On the other hand, other wild-type KATs such as GCN5 and MOF have extremely lower activity with these orthogonal acyl donors, most likely because of their limited acyl-CoA binding pocket. Therefore, we mutated the large amino acid residues in the acetyl-CoA binding pocket of GCN5 to a smaller residue (Ala or Gly) by site-directed mutagenesis. Several GCN5 mutants with enlarged acetyl-CoA binding space display higher activities toward the synthetic acetyl-CoA surrogates. [42] In particular, we found that the mutant with Thr612 to Gly (GCN5-T612G) is able to activate a series of alkynyl and azido acyl-CoA for histone acylation (ref.[42] and **Table 2.1**). Of note, Thr612 is conserved in the PCAF/GCN5 family, so it is likely that the same mutation in PCAF could render similar activity as GCN5 toward the alkyne- or azide-functionalized cofactors. We first quantitatively determined the activities of the KATs for each acyl-CoA.[42] The kinetic constants of these acetyl-CoA surrogates with GCN5-T612G and p300 were measured with an *in vitro* fluorogenic assay.[43] Histone peptide H3-20 or H4-20 (the first 20 amino acid sequence from the amino terminus of histone H3 or H4) was used as an acyl acceptor substrate. Kinetic constants k_{cat} and K_m were determined by fitting the cofactor concentration─catalytic rate data curves to the Michaelis-Menten equation **(Figure 2.2)**. k_{cat}/K_m was used to evaluate the specificity of KAT for each cofactor. Among all the acyl-CoA cofactors tested, 3AZ-CoA turned out to be the most active one for both p300 and GCN5-T612G. p300

possesses similar activities for acetyl-CoA and 3AZ-CoA. The k_{cat}/K_m value of p300 to 3AZ-CoA is 9 times that of p300 to 4PY-CoA, which was used in a previous study for p300 substrate labeling.[36] The other two acyl-CoA molecules 5HY- and 6HY-CoA are quite inert for p300, with little activity detected under initial velocity conditions. In terms of GCN5, the GCN5-T612G mutant is significantly more active to 3AZ-CoA than the natural cofactor Ac-CoA, with k_{cat}/K_m being 44.5 and 9.6 $min^{-1}.\mu M^{-1}$ respectively. Also, the k_{cat}/K_m value of GCN5-T612G to 3AZ-CoA is higher than that of wild type GCN5 to Ac-CoA (k_{cat}/K_m 12.8 $min^{-1}.\mu M^{-1}$). We also tested the kinetic constants of wild type GCN5 to 3AZ-CoA in which the k_{cat}/K_m value was found to be 0.04 $min^{-1}.\mu M^{-1}$, hundreds of folds weaker than that of wild type GCN5 and GCN5-T612G to acetyl-CoA and GCN5-T612G to 3AZ-CoA **(Table 2.1 and Figure 2.2)**. Collectively, 3AZ-CoA is a sensitive and specific probe for both p300 and GCN5-T612G with comparable or even stronger activity than Ac-CoA. At this point, the structural mechanism underlying 3AZ-CoA being a strong acyl-donor in p300 activity is not clear. Nevertheless, that p300 uses 3AZ-CoA as an equally strong cofactor as Ac-CoA is in good accordance with the fact that p300 also acts as propionyl- and butyryl-transferases. After the *in vitro* kinetic studies proved that 3AZ-CoA is a highly active acyl-CoA surrogate for p300 and GCN5-T612G, we next performed in-gel fluorescence studies to check the labeling efficiency of the two KATs with 3AZ-CoA on the cellular proteome. The human embryonic kidney (HEK) 293T cell lysates were incubated with KAT and 3AZ-CoA, followed by CuAAC reaction to tag azidopropionylated proteins with alkyne tetramethylrhodamine. Treatment of the cell lysate with p300 and GCN5-T612G resulted in strong fluorescence signals, indicating efficient labeling of protein substrates with the 3AZ group. In contrast, treatment of the cell lysate with 3AZ-CoA and other wild-type enzymes such as MOF, Tip60, MOZ, MORF, GCN5, and PCAF resulted in very weak fluorescence that was

close to the negative control, which only had 3AZ-CoA treatment **(Figure 2.3)**. These data support that labeling using 3AZ-CoA coupled with the CuAAC click chemistry is unique and specific for wild type p300 and GCN5-T612G.

GCN5 and p300 substrate enrichment with biotin-streptavidin pull down.

The in-gel fluorescence data demonstrated that the conjugation of enzymatic reaction and click reaction effectively incorporated 3-azidopropionyl groups to target proteins and enabled subsequent detection with fluorescent reporters. Next, we used a sodium dithionite cleavable alkyne azo biotin (Ak-Azo-BTN) probe to replace Ak-TAMRA and expected that target proteins would be efficiently biotinylated using the same labeling strategy. The efficiency of protein biotinylation via KAT reaction and CuAAC click reaction was verified with chemiluminescent imaging **(Figure 2.4a)**. In this experiment, streptavidin-HRP was used for biotin detection. Treating the cell lysate with only 3AZ-CoA resulted in very weak chemiluminescence, indicating that endogenous KATs barely catalyzed the acylation reaction with 3AZ-CoA. On the contrary, exogenous p300 and GCN5-T612G treatment induced strong signals, indicating that numerous proteins were biotinylated due to the enzymatic labeling. Now that p300 & GCN5-T612G could make advantageous use of 3AZ-CoA to achieve specific biotinylation of p300 and GCN5 substrates, it would be promising to enrich these target proteins with subsequent pull-down experiment on streptavidin-coated beads. In this regard, the lysates were treated in the same way aforementioned with KAT enzyme and 3AZ-CoA, and then Ak-Azo-BTN was conjugated to target proteins via CuAAC. The biotinylated proteins were enriched through biotin-streptavidin pull-down and the protein eluents were resolved on SDS-PAGE gel and imaged with silver staining **(Figure 2.4b)**. Clearly, very little chemical reactivity of 3AZ-CoA was observed without enzyme, while p300 and GCN5-T612G treatment led to significant protein enrichment.

Therefore, this KAT activity-based labeling strategy and subsequent protein enrichment is highly effective to enrich cellular substrates of p300 and GCN5.

Profiling of enriched proteins with semi-quantitative LC-MS/MS.

After we demonstrated that the activity-based chemical labeling and affinity pull-down efficiently enriched p300/GCN5 substrates, we next conducted proteomic studies to determine the identities of the enriched KAT substrates. LC-MS/MS proteomic analysis in recent years has been used extensively to identify PTM substrates together with the application of chemical biological enrichment.[36,44] Basically, the enriched protein samples were digested and the generated peptide fragments were identified according to their m/z values. These mass-to-charge data were analyzed through database search engines such as Sequest and Mascot to find the corresponding proteins via sequence alignment.[45] Since each protein can yield multiple peptides, a comparison of the peptide counts of individual proteins between control and treated groups has been used as the standard to determine protein hits.[36,44] Although being widely used, this method has low accuracy to quantitate protein abundance. Strong labeling or enrichment of proteins may not always induce increases of peptide fragment ion counts after tryptic digestion; on the contrary, fewer peptide fragments could be yielded if the functional tags that are added to lysines or arginines block tryptic digestion. Herein, we applied a semi-quantitative label-free LC-MS/MS method for protein hit identification. Peak areas of the three most abundant peptide fragment ions are summed up to quantify the individual proteins. Hundreds of proteins were identified in all three groups respectively, but the total abundance of proteins between them were significantly different from each other. As is shown in **Figure 2.4c**, GCN5-T612G and p300 treatment led to 42- and 706-fold enrichment when compared to the control group in which the cell lysate was treated with only 3AZ-CoA but not with enzyme. Proteins that were enriched

more than three times by p300 or GCN5-T612G were chosen as substrate candidates. A total of 568 proteins were identified as KAT substrates; 472 and 475 proteins were identified as GCN5 and p300 substrate candidates, respectively (**Supplementary Table 1** and **Figure 2.5**). Further examination revealed that p300 and GCN5-T612G had 379 shared substrates and ~90 unique ones to each enzyme. Histones are the most enriched proteins in both sets. The abundance of histone proteins in p300 and GCN5-T612G treated samples are hundreds or even thousands of folds higher than the control group. Histone H3 being highly enriched by GCN5-T612G and all the four core histones being highly enriched by p300 is consistent with our understanding that GCN5 primarily acetylates histone H3 while p300 acetylates multiple histones with comparable activities.[46-48] Histone H1 is highly enriched upon treatment of both p300 and GCN5-T612G, indicating that histone H1 may be a good substrate of p300 and GCN5. Other than histones, hundreds of non-histone proteins were highly enriched, which will be annotated later. The high-degree of overlap of GCN5 and p300 substrates suggested that these two enzymes, although not from the same family, can act on some common protein substrates and thus regulate common or related pathways. However, we need to point out that for the same substrates, the modification sites for p300 and GCN5 could be different from each other. To assess which enzyme has higher activity on the shared substrates, we normalized their abundance to the total protein abundance in p300 and GCN5-T612G groups respectively. The ratios of relative abundance of each shared substrate candidates in GCN5-T612G and p300 sets were listed in Figure 4b, showing that the majority of shared substrate candidates are more enriched by GCN5-T612G. As is shown in **Supplementary Table 2**, most histone proteins and a few non-histone proteins such as proteins SET, high mobility group protein B and nucleolin were more enriched by p300. Many of these proteins are involved with either DNA- or RNA- related functions such as transcription

regulation, RNA stabilization and RNA chaperoning.[49] The rest of the hundreds of non-histone proteins are more enriched by GCN5-T612G. This may suggest that p300 is more selective toward histone proteins over non-histone proteins while the preference of GCN5 for histones is at a lesser extent.

Validity of the identified KAT substrates.

We compared the substrate candidates identified from this work with the published data. Yang and coworkers identified 23 proteins as p300 substrates with 4PY-CoA through *in vitro* labeling of Hela nuclear extracts and subsequent CuAAC and proteomic analysis.[36] 17 out of the 23 proteins were found in our list of p300 substrates. The reason that we identified more p300 substrates than their work could be multifold: the binding affinity of 3AZ-CoA probe to p300 is about 10-fold higher than 4PY-CoA (**Table 2.1**); Yang and coworkers used Hela nuclear extracts as substrate source while we used 293T whole cell lysates as the substrate source. Therefore, it is reasonable that we identified more substrates than Ref.[36] Choudhary et al. surveyed an "acetylome" containing 4951 proteins from different human cell lines that undergo lysine-ε-amine acetylation.[50] 67% (378 out of 568) of the substrate candidates identified herein are found in this group **(Figure 2.5a)**. Notably, matching with KAT's canonical function in chromatin biology, all the core histones were identified in the list. These evidence supports that our chemoproteomic strategy is a valid approach to discovering physiologically relevant KAT substrates. We also tested the enzymatic activity of both KATs on a shared substrate, tubulin. While tubulin is a known substrate of GCN5, this study is the first to demonstrate that tubulin is a p300 substrate.[51] The western blotting verified that both p300 and GCN5 serve as *bona fide* tubulin lysine acetyltransferases and the LC-MS/MS proteomic analysis showed that both KATs could modify tubulin on the same lysine residues including Lys60 on tubulin α chain and Lys 59

and 60 on tubulin β chain **(Figure 2.6)**. To further validate our proteomic discovery, we chose two undefined GCN5 substrates identified by our study, Ras GTPase-activating-like protein IQGAP1 and the structural maintenance of chromosomes protein 1 (SMC1) for cellular acetylation test. A plasmid containing either Myc tagged IQGAP1 or SMC1 was overexpressed with or without the FLAG-GCN5 plasmid. The Myc tagged proteins were enriched with immunoprecipitation and their acetylation levels were detected using pan anti-acetyllysine antibody. As shown in **Figure 2.7,** the acetylation level of both IQGAP1 and SMC1 proteins increased significantly in the presence of the overexpression of GCN5, supporting that GCN5 acetylated these two proteins *in vivo*. Overall, the high accordance with previous proteomic work and our validation study highlight the reliability, validity, and biological relevance of our chemoproteomic strategy for KAT substrates discovery. It is worthwhile to point out that the identified GCN5 and p300 sub-acylome may not be exactly the same as their *in vivo* sub-acylome due to the fact that KATs, in many cases, function as acetyltransferase subunit in protein complexes and alteration of binding proteins may change their substrate specificities.[52,53]

Functional annotation of GCN5- and p300- shared substrate candidates.

To understand the broad biological involvement of GCN5 and p300 as well as their functional differences, we annotated physiological and pathological functions of the shared and unique substrates of both enzymes with DAVID Bioinformatics Resources 6.8 (https://david.ncifcrf.gov/) **(Supplementary Table 3, Figure 2.8)**.[54,55]

Canonical pathways related with GCN5 and p300 shared substrates. p300 and GCN5 shared substrates are involved in more than 30 pathways related with gene expression, cellular metabolism, etc. **(Supplementary Table 4, Figure 2.8)**. Not surprisingly, canonical KAT-related processes such as gene expression, DNA repair, cell cycle and apoptosis have been found

in the annotation and involve both known and newly identified substrates of p300 and GCN5.[39,56] For instance, DNA replication and DNA repair regulatory proteins Flap Endonuclease-1 (Fen-1), high mobility group protein B1 (HMGB1), poly (ADP-ribose) polymerase-1 (PARP-1), and proliferating cell nuclear antigen (PCNA) are known p300 substrates.[37,38,57-59] Acetylation of these proteins changes their activities or functions. Furthermore, a few other proteins such as X-ray repair cross-complementing protein and DNA replication licensing factor MCM were identified for the first time as potential GCN5 and p300 shared substrates. Cell cycle and apoptosis are also regulated by KATs. Cyclin-dependent kinase-1 (CDK-1) is an important cell cycle regulatory protein and has an acetylated lysine 40 within its kinase domain.[60] CDK-1 is identified herein as a potential substrate of both p300 and GCN5, indicating a new pathway that p300 and GCN5 regulate the cell cycle through acetylation of CDK protein(s). We have also identified more than 130 proteins related with the metabolism of proteins, carbohydrates, fatty acids as well as amino acids, indicating extensive involvement of GCN5 and p300 in eukaryote metabolism. This is consistent with a recent study by Zhao et al. who demonstrated that lysine acetylation has a dramatic influence on cellular metabolizing enzymes in human liver tissue.[61] As is shown in **Supplementary Table 5**, out of 40 metabolizing enzymes that we have identified using the KEGG pathway database,[62] 21 (black) of them were reported by Zhao et al. as well. The non-overlapped part (red) might be caused by the difference of the cell lines that were used in our and their studies.

GCN5 and p300-shared substrates include many ribosomal and spliceosomal proteins. Spliceosome and ribosome are important RNA-protein complexes that regulate and control post-transcriptional RNA processing and protein translation. Spliceosomes are composed of 5 different small nuclear RNAs (snRNAs) and over a hundred associated proteins. They remove

introns from pre-mRNA and connects flanking exons to form mRNA encoding sequences.[63] As mRNA translation machinery, ribosomes are composed of two subunits containing 50-80 individual proteins depends on the species. These proteins are structurally conserved between different organisms.[64] Structural and functional studies of spliceosomes and ribosomes have been extensively pushed forward in the past decades, while there are still unresolved puzzles such as why a large variety of ribosomal proteins exist and how spliceosomes assembly and RNA splicing are regulated.[64] Our study has identified 44 spliceosomal and 63 ribosomal proteins as GCN5 and p300 shared substrate candidates **(Supplementary Table 4)**, suggesting that lysine acylation may play some roles during ribosome and spliceosome assembly and RNA recognition through regulation of protein-protein or protein-RNA interactions. These findings also indicate that GCN5 and p300 may affect gene expression through regulation of acylation of spliceosomal and ribosomal proteins. This hypothesis is somehow supported by a recent study which suggested that the SAGA (Spt /Ada/ Gcn5/Ada) complex is necessary for co-transcriptional spliceosome assembly.[65] Moreover, 93 proteins are involved in influenza infection and most of them are ribosomal and spliceosomal proteins.[66,67] This is likely because influenza viruses need to replicate and translate its RNA after entering into the host cells, and ribosomes and spliceosomes are essential in these processes. Possibly, p300 and GCN5 may affect progression of influenza through the alteration of ribonucleoprotein acylation. There are also other common substrates and pathways shared by p300 and GCN5 and all these shared substrates and their related pathways are summarized in **Supplementary Table 4.1.**

Functional annotation of GCN5- or p300- unique substrate candidates.

93 and 96 proteins have been identified as GCN5 and p300 unique substrates, respectively **(Figure 2.5)**. Functional annotation of these proteins was carried out with DAVID

Bioinformatics Resources 6.8[54,55] and more than 30 biologically significant proteins from GCN5 and p300 clusters were identified **(Figure 2.8, Supplementary Table 3.2, 3.3**). Some of these proteins share similar functions as GCN5 and p300 shared substrates: GCN5 unique substrates contains 9 spliceosomal proteins and 5 TCA cycle regulatory proteins; p300 unique substrates contain 10 spliceosomal proteins, 10 gene expression, 4 carbohydrate metabolism, and 7 cell cycle regulatory proteins. Importantly, there are also pathways that are regulated by only p300- or GCN5-unique substrates.

p300 unique substrates and related pathways. PELP1 (proline-, glutamic acid-, and leucine-rich protein1) is a coactivator of the estrogen receptor (ER).[68] It binds to ER and p300 at the same time to potentiate histone acetylation and activate gene transcription.[68] p300 is able to potentiate growth and migration of breast cancer cells and the ER has been proven to be a breast cancer initiator.[69-71] Our study suggests that PELP can be acylated by p300 which, together with previous studies, indicates a new possible mechanism that p300 potentiates breast cancer development by acetylating PELP. Signal recognition particles (SRP) are ribonucleoproteins that translocate newly synthesized proteins (polypeptides) from the ribosome to the endoplasmic reticulum.[72] There are 6 different proteins that serve as SRP subunits including SRP 9, 14, 19, 54, 68 and 72. Our study suggests that SRP 68 and SRP 72 can be acylated by p300, indicating a possible role of p300 in regulating the translocation of newly synthesized proteins. Besides, p300 may also regulate arginine and proline metabolism by acylating metabolizing enzymes pyrroline-5-carboxylate reductase, creatine kinase and aspartate aminotransferase 2 and regulate the pentose phosphate pathway by acylating transaldolase and glucose-6-phosphate isomerase.

GCN5 unique substrates and related pathways. Pike et al. demonstrated that the vitamin D receptor (VDR) controls gene expression by recruiting co-regulatory complexes to different

promoters.[73] Herein, we identified two regulatory proteins of this pathway including SNW domain-containing protein 1 and SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily D member 2 (SMARCD2) as GCN5 substrate candidates. This indicates a possible co-regulatory role of GCN5 in gene expression with the vitamin D receptor. Antigen processing and presentation are key immune processes that present antigens to T cells through major histocompatibility complex (MHC) pathways.[74] MHC is a set of cell surface proteins. They can bind with antigenic peptides to form complexes that can be recognized by T cells. There are two classes of MHC (Class I and II) and they mediate the presentation of intracellular and extracellular antigenic peptides respectively.[75] Legumain is a cysteine endopeptidase that digests antigenic proteins such as toxin C fragment (TTCF) to antigenic peptides with other proteolytic enzymes. It has been shown that abolishment of legumain activity results in the resistance of TTCF proteolysis and keeps cells from being presented to T cells.[76] Heat shock proteins (Hsps) also play important roles during antigen processing and presentation. Murshid and coworkers concluded that Hsps bind to and chaperone multiple polypeptides and mediate cross presentation of tumor antigens by antigen present cells (APC).[77] Our study suggests that legumin, poly Hsp90-alpha, beta and Hsp70 protein 4 can be acylated by GCN5, indicating that GCN5 may affect antigen processing and presentation by regulating the acetylation of these proteins. In addition, four RNA helicases including U5 small nuclear ribonucleoprotein 200 kDa helicase, putative pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16, putative ATP-dependent RNA helicase DHX30, regulator of nonsense transcripts 1 and two RNA degrading proteins including 5’-3’ exoribonuclease 2 and exosome complex component RRP45 have been identified as GCN5 substrate candidates. Last but not least, GCN5 is also likely to be involved in the metabolism of the amino acids methionine, serine, threonine, isoleucine, and

valine by acylating cystathionine beta-synthase, cystathionine gamma-lyase and D-3-phosphoglycerate dehydrogenase.

Conclusion

By combining enzyme activity-based chemical labeling using clickable acyl-CoA probes in conjugation with either wild type or mutant KATs, we have demonstrated a bioorthogonal chemoproteomic method to profile sub-acylome of different KATs in cellular contexts. Hundreds of substrates of GCN5 and p300 were identified containing both previously known and first-time identified proteins. We found that histone proteins are the primary substrates of both GCN5 and p300 while GCN5 is less selective to histone proteins in cellular contexts compared to p300. Our functional annotations of the KAT substrates point toward multiple cellular pathways regulated by p300 and GCN5. In addition to the previously studied pathways such as gene expression, cellular metabolism, cell cycle, DNA repair, we found the potential involvement of GCN5 and p300 in other cellular processes such as RNA splicing and translation, antigen processing and presentation, and protein export. We validated two novel substrates of GCN5, i.e. IQGAP1 and SMC1, by co-transfection studies. This work provides a great insight for future work to investigate the functional involvement of GCN5 and p300 in these newly discovered biological pathways.

Methods and Materials

Synthesis of Acyl-CoA Analogs.

4-pentynoyl CoA (4PY-CoA), 5-hexynoyl CoA (5HY-CoA), 6-heptynoyl CoA (6HY-CoA), and 3-azidopropionyl CoA (3AZ-CoA) were synthesized by reacting with carboxylic anhydride with CoA-SH, by following the procedures used in our previous study.[42]

Protein Expression and Purification.

Site-directed mutagenesis was used to generate GCN5-T612G plasmid. Nickel-NTA agarose resin from Bio-Rad was used for the purification of His-tag labeled proteins including GCN5-T612G, GCN5, MOF, Tip60 and PCAF. The expression and purification condition were described in the previous work.[42,78]

p300 protein expression and purification were described in the previous work of our lab.[79] pET28a-LIC-MOZ plasmid was purchased from Addgene Plasmid #25181. MOZ protein was expressed and purified according to the protocol developed by Structural Genomics Consortium (SGC) (PDB code: 2OZU). (http://www.thesgc.org/structures/2ozu)

Maltose binding protein (MBP)-MORF catalytic domain-pMAL plasmid was a gift from Dr. Xiangjiao Yang from McGill University. Plasmids were transformed into BL21 (DE3) cells with heat-shock method followed spreading cells on LB-Agar plate with ampicillin. Colonies were picked up and cultured in 8 mL of growth medium for 12 hours followed by transferring 8 mL medium into 1L medium containing 10g Bacto trypton, 5g yeast extract, 5g sodium chloride, 2g glucose and 100 mg/mL of ampicillin. 1 liter culture medium was incubated at 37^{o}C until the OD_{595nm} reach to 0.7. 0.3 mM of IPTG was added to induce protein expression at 16 ^{o}C overnight. Cells were collected by centrifugation and suspended in lysis/washing buffer containing 20 mM Tris-HCl, pH 8.0, 0.5M KCl, 10% (v/v) glycerol, 5 mM $MgCl_2$, 0.1% (v/v) Nonidet P-40, and 1 mM Phenylmethylsulfonyl fluoride (PMSF). Cells were disrupted with a microfluidizer at 15000 psi followed by collection of supernatant after 12000 rpm centrifugation. Supernatant containing cellular proteins were incubated with Amylose resin (New England Biolabs, Product #E8021S), followed by washing with lysis/washing buffer and eluting with elution buffer containing Tris-HCl, pH 8.0, 0.15M KCl, 10% (v/v) glycerol, 5 mM MgCl2, 0.1%

(v/v) Nonidet P-40, 10 mM Maltose and 1 mM phenylmethylsulfonyl fluoride(PMSF). Eluent was dialyzed with dialysis buffer containing 25 mM Tris pH 8.0, 250 mM of NaCl, 10% (v/v) glycerol, 0.1 mM EDTA and 1 mM DTT. Protein was concentrated and protein purity was checked with sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS–PAGE). Proteins were aliquoted, flash frozen and stored at –80 ^{o}C for future use.

Fluorescence-Based Kinetic Assays.

Kinetic assays were used to determine the binding affinity (K_m) and catalytic efficiency (k_{cat}) of enzyme and analogs. k_{cat}/K_m was used to quantify the activities of enzyme and analog pairs. For GCN5-T612G kinetic characterization, 100 μM of H3-20 peptide (Ac ARTKQTARKSTGG KAPRKQL) were pre-incubated with cofactors at varied concentrations for 5 minutes at 30 ^{o}C in KAT reaction buffer including 50 mM of HEPEs pH 8.0, 0.1 mM of EDTA. 10 nM of enzyme was added to initiate enzymatic reaction, which lasted for 8-15 minutes followed by the addition of excessive amount of CPM in dimethyl sulfoxide (DMSO). CPM DMSO solution not only quenched enzymatic reaction but also reacts with the by-product CoASH and generated fluorescent product CoAS-CPM. The fluorescence intensity was measured with micro-plate reader (FlexStation 3) with fixed excitation and emission wavelength at 392 nm and 482 nm. For p300, H4-20 peptide (Ac-SGRGKGGKGLGKGGAKRHRK) was used as substrates and same conditions were used for the measurement with Ac-CoA and 3AZ-CoA. In order to get detectable and valid fluorescence reading, enzyme and peptide concentrations were increased to 100 nM and 200 μM when reacting with 4PY, 5HY and 6HY-CoA. The kinetic data were summarized in **Table 2.1** and **Figure 2.2**.

In-gel fluorescence and chemi-imaging assay

Extraction of cellular protein. Human embryonic kidney 293 cells (HEK293T cells) were cultured with DMEM medium supplemented with 10% (v/v) FBS at 37 ºC with 5% of CO_2. Cells were scraped after reaching to ~90% confluence and washed with ice-cold PBS for twice. Cells were collected by centrifugation at 12000 rpm at 4 ºC 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, cells were sonicated with 30% amplitude to release whole cellular proteins. Cellular proteins were collected by centrifugation for 20 minutes at 12000 rpm at 4 ºC and protein concentration was measured with Bradford assay. Cell lysates were aliquoted, flash frozen and stored at –80 ºC for future use.

Fluorescent and chemiluminescent gel imaging. 20 μL of mixture containing 20 μg of cell lysate, 25 μM of 3AZ-CoA and 5 μM of GCN5-T612G or 2 μM of p300 in 50 mM of HEPES pH 8.0 and 0.1 mM of EDTA was incubated at 30 ºC for 50 minutes. 10 μL of click cocktail containing 75 μM of alkyne-tetramethylrhodamine (Ak-TAMRA) or alkyne-azo-biotin, 1.5 mM of ligand BTTP, 15 mM of reducing agent sodium ascorbate and 1.5 mM of copper sulfate was added. Reaction mixture was incubated at room temperature for 30 minutes followed by the addition of 8 μL of SDS-PAGE loading dye. Samples were resolved on 4-20% SDS-PAGE gradient gel. For Ak-TAMRA treated samples, the gel was scanned with Typhoon scanner (GE Healthcare Life Sciences) using fixed excitation and emission wavelengths at 532 nm and 580 nm. For Ak-azo-BTN treated samples, proteins were transferred onto nitrocellulose membrane followed by blocking with 5% non-fat milk. The membrane was incubated with streptavidin-HRP and scanned with chemiluminescence scanner.

Streptavidin affinity enrichment, in-gel digestion, and semi-quantitative LC-MS/MS proteomic analysis.

KAT substrates were enriched with biotin-streptavidin pull down. 800 μg of cell lysate was incubated with 2 μM or 5 μM of p300 or GCN5-T612G and 25 μM of 3AZ-CoA at 30 ^{o}C for 50 minutes. Click cocktail containing Ak-Azo-BTN was added to biotinylated target proteins with 30 minutes at room temperature. Excessive Ak-Azo-BTN was removed with spin-dialysis. Samples were incubated with streptavidin affinity resin (ThermoFisher SCIENTIFIC, Product #20347) that was pre-equilibrated with pH 7.4 phosphate buffer saline (PBS). The resins were collected by centrifugation at 1000g for 5 minutes and were washed sequentially with PBS supplemented with 0.2% (w/v) of SDS, PBS supplemented with 0.1% (w/v) of SDS and 6M urea and 50 mM of NH_4HCO_3 supplemented with 0.05% of SDS. Then, the beads were incubated with 25 mM of $Na_2S_2O_4$ in 50 mM NH_4HCO_3 for 1 hour to cleave the protein off the beads. The mixture was centrifuged at 1000g for 5 minutes and the supernatant was carefully collected. The supernatant was dried down by SpeedVac.

Substrates identification with in-gel tryptic digestion and LC-MS/MS proteomic analysis.

A variation of gel assisted proteomics sample preparation was used to generate tryptic peptide digests of the proteins cleaved from the beads.[80] Dried samples were dissolved in 20 μL of water followed by incubation with 20 mM of DTT at 75 ^{o}C for 10 minutes and 150 mM of iodoacetamide at room temperature in dark for 20 minutes. Then, protein sample was fixed in the gel containing 30% acrylamide, ammonium persulfate and TEMED. The gel bands were sliced into small pieces, and then rinsed with 50% (v/v) acetonitrile/20 mM ammonium bicarbonate (~pH7.5–8) twice. The gel pieces were dehydrated by adding 100% of acetonitrile and dried out by a heat block at 60°C. A various amount of Trypsin solution (0.01μg.μL^{-1} 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% (v/v) acetonitrile/0.1% (v/v) formic acid twice. The extracts were dried down with low pressure-centrifugation using a SpeedVac.

The mass spectrometry analysis was carried out on a Thermo-Fisher LTQ Orbitrap Elite Mass Spectrometer coupled with a Proxeon Easy NanoLC system (Waltham, MA) located at Proteomics and Mass Spectrometry Facility, University of Georgia. 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 0.1% (v/v) formic acid as buffer A and 99.9% (v/v) acetonitrile with 0.1% (v/v) formic acid as buffer B starts with 5% (v/v) B, holds at 5% (v/v) B for 2 minutes, then increases to 25% (v/v) B in 60 minutes, to 40% (v/v) B in 10 minutes, and to 95% (v/v) 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 5 ions in the MS scan were selected for following CID and HCD MS/MS analysis. Both MS and MS/MS scans were acquired by Orbitrap at the resolutions of 120,000 and 30,000, respectively.

Data were acquired using Xcalibur software (version 2.2, ThermoFisher SCIENTIFIC). Proteins identification and modification characterization were performed using Thermo Proteome Discoverer (version 1.4) with Mascot (Matrix Science) and Uniprot database. The spectra of possible modified peptides were inspected further to verify the accuracy of the assignments. The semi-quantitative analysis was achieved using a label-free quantification workflow within Proteome Discoverer, which a Precursor Ion Areas Detector node calculates the

average of the top three highest peptide areas in the Extracted Ion Chromatograms (mass precision, 2 ppm).

Identification of GCN5 and p300 acetylation sites on tubulin using LC-MS/MS.

Tubulin extracted from bovine brain (Sigma Aldrich) was incubated with 150 μM deuterium labeled acetyl-CoA (d3-Ac-CoA from Sigma Aldrich) and 0.5 μM GCN5 or p300. Only d3-Ac-CoA treated tubulin was used as negative control. The reaction mixtures were subjected to LC-MS/MS analysis. Tubulin samples was diluted to 0.3 mg. mL^{-1} in 40 mM Tris-HCl (pH 8), 5 mM DTT. The mixture was heated at 95 °C for 20 min. The buffer was exchanged to 50 mM ammonium bicarbonate (pH 8.0) with the desalting column followed by tryptic digestion (protein: enzyme 60:1) at 37 °C for 5 h. Samples were analyzed by ACQUITY UPLC system (Waters, Milford, MA) coupled to a Waters SYNAPT G2 mass spectrometer (Milford, MA). Peptides were separated on the HALO C18 peptide column (2.7 μm, 4.6 x 100mm, Advanced Materials Technology, Wilmington, DE). Mobile phase A is water containing 0.01% (v/v) formic acid and B is acetonitrile (ACN). The injection volume was 20 μL. The tryptic digest was analyzed with a 60 minutes gradient method, 75 min run time, with a 0.3 mL. min^{-1} flow rate, (time. $Minute^{-1}$, % mobile phase B): (0, 5), (60, 50), (60.01, 100), (67.50, 100), (67.51, 5). Key parameter setting for the SYNAPT G2 mass spectrometer were as follows: capillary 2.00 kV, sample cone 35 V, extraction cone 4.0 V, source temperature 120 °C, desolvation temperature 500 °C, desolvation gas 500 L. h^{-1}. All data were acquired by the data-dependent acquisition (DDA) mode. For DDA parameters, 1s MS survey scan in the mass range of 300–1900 were followed by MS/MS scans of up to 3 ions, when intensity rose above 1500 counts per second. MS/MS was acquired over the range of 100-1900, with a 2s scan rate. MS/MS scan was

switched to MS survey scan after 3 scans. Trap collision energy was set using charge state recognition, applying the default files for 1-4 charge states.

Files containing MS/MS spectra were processed with Proteinlynx Global Server 2.4 software (Waters, Milford, MA) to identify peptides. The following parameters were used: (i) peptide tolerance, 500 ppm; (ii) fragment tolerance 0.5 Da; (iii) primary digestion, trypsin; (iv) missed cleavage, 3; (v) Variable Modification, Acetyl K and d3-acetyl K (mass shift 45.01). Results were validated based on the identification of three or more consecutive fragment ions from the same series. We identified three peptides for both p300 and GCN5 treated samples that had d3-acetyl labeled lysines while no d3-acetylated labeled peptide was detected from only d3-acetyl-CoA treated tubulin. **(Figure 2.6)**

Detection of IQGAP1 and SMC1 acetylation by GCN5 *in vivo*.

Human embryonic kidney (HEK) 293T cells were cultured to ~80% confluence at 37 ^{o}C using DMEM medium supplemented with 10% FBS and 1% streptomycin-penicillin. Transient transfection of the plasmids pAdEasy-Flag-GCN5[81] (Addgene Plasmid #14106), pcDNA3-Myc-IQGAP1[82] (Addgene Plasmid #30118), and pCDNA-3 5' cMyc-SMC1-wt[83] (Addgene Plasmid #32363) was conducted using the Lipofectamine 3000 Reagent (ThermoFisher SCIENTIFIC, Product #L3000008). The whole cell lysates were extracted using the same method aforementioned. The GCN5 expression level was detected using anti-FLAG tag antibody. The cell lysates were subjected to immunoprecipitation using the c-Myc-Tag IP/Co-IP Kit (ThermoFisher SCIENTIFIC, Product #23620) for IQGAP1 and SMC1 enrichment. The acetylation level of enriched proteins was detected using western blotting analysis with pan anti-acetyllysine antibody. The results are shown in **Figure 2.7** of the main text.

Figures, Tables and Captions

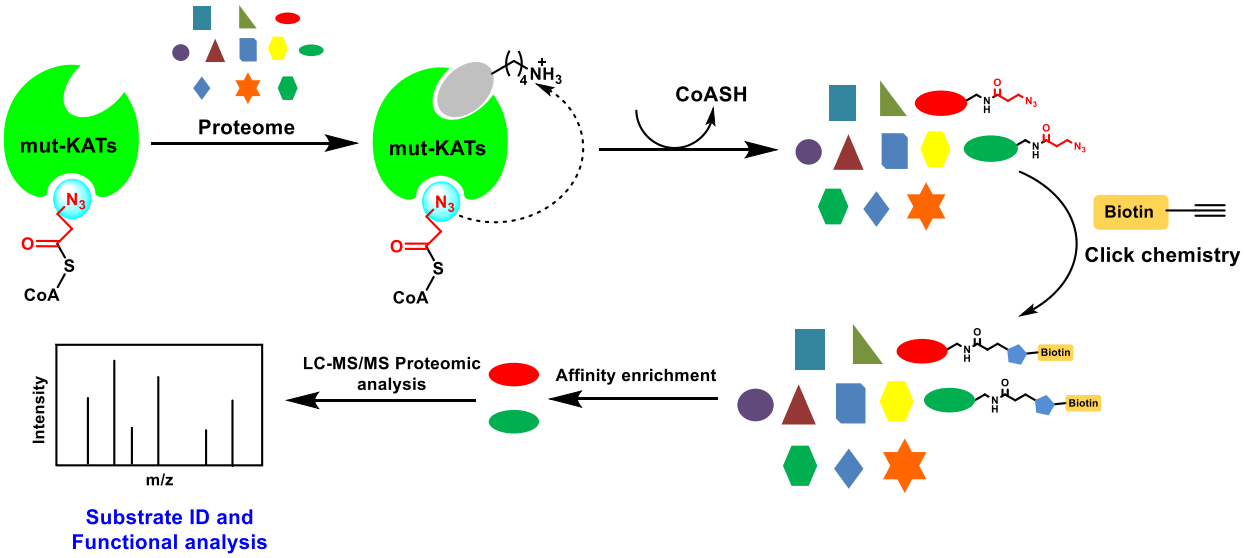


Figure 2.1. Labeling and profiling KAT substrates with wild-type or engineered KAT, functionalized acyl-CoA, click reaction and MS/MS proteomics.

Incubation of cellular lysates with KAT and 3-azidopropionyl-CoA (3AZ-CoA) induces labeling of KAT substrates (oval proteins) with an azide tag. Through a CuAAC reaction, a biotin tag is covalently attached to the KAT substrates, followed by protein enrichment and LC-MS/MS to identify the KAT's sub-acylome from the cellular proteome.

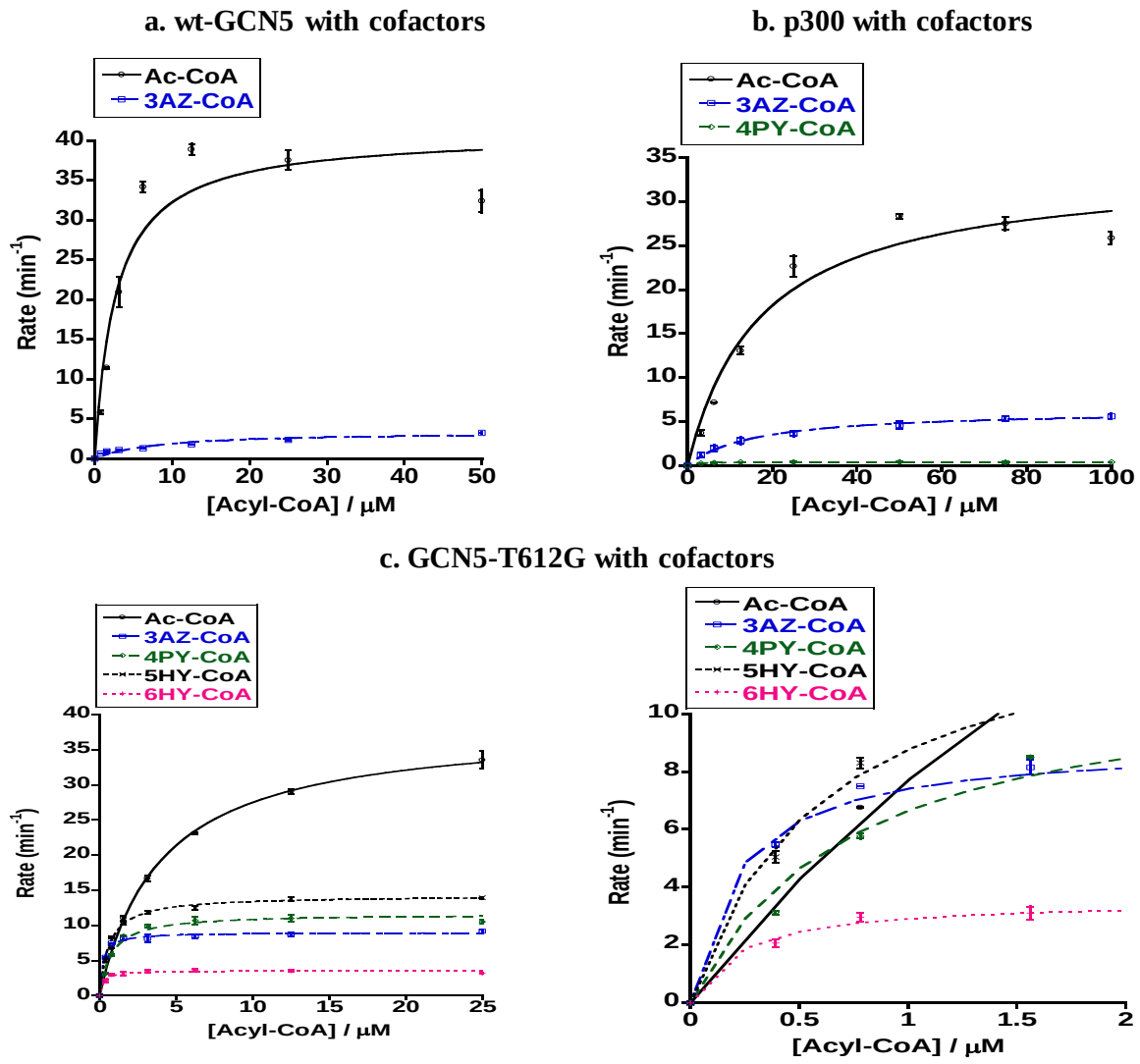


Figure 2.2. Kinetic analysis of KATs with acetyl-CoA and analogs.

The kinetic constants K_m and k_{cat} of each KAT with cofactors were obtained through fluorescence measurement and Michaelis-Menten curve fitting. **a.** wild type GCN5 has much weaker activity for 3AZ-CoA than acetyl-CoA; **b.** p300 with acetyl-CoA, 3AZ-CoA and 4PY-CoA; **c.** GCN5-T612G mutant possesses much stronger activity with Ac-CoA analogs than wild type GCN5. 3AZ-CoA does not lead to the highest V_{max} among all these cofactors, but it has the lowest K_m with GCN5-T612G, therefore, it has the highest k_{cat}/K_m.

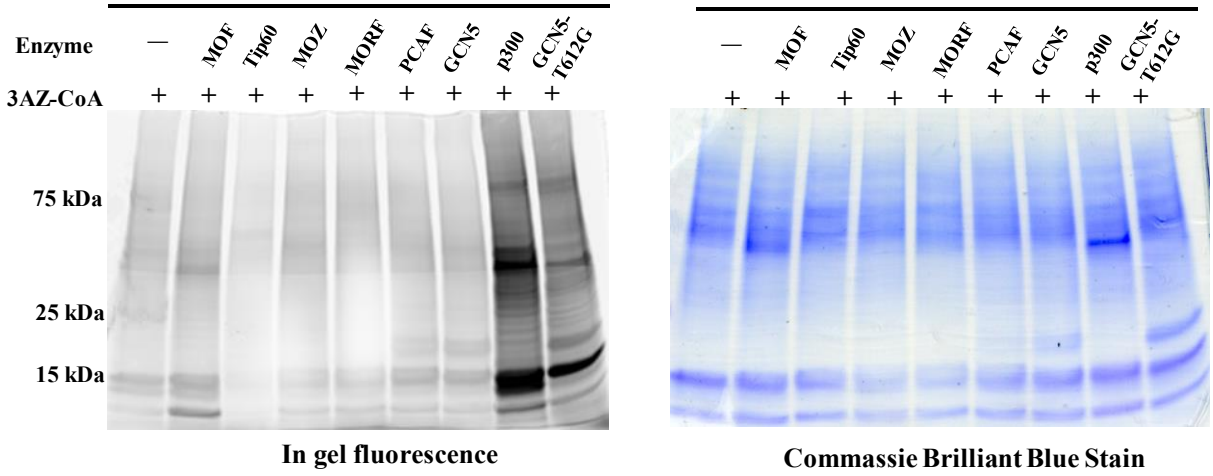


Figure 2.3. Chemoproteomic labeling of cellular substrates using 3AZ-CoA coupled with different KATs.

Cell lysate was treated with 3AZ-CoA and different KATs, followed by conjugation of alkyne-TAMRA with CuAAC. The upper panel shows the in-gel fluorescence image and the lower panel shows protein loading stained with Coomassie Brilliant blue.

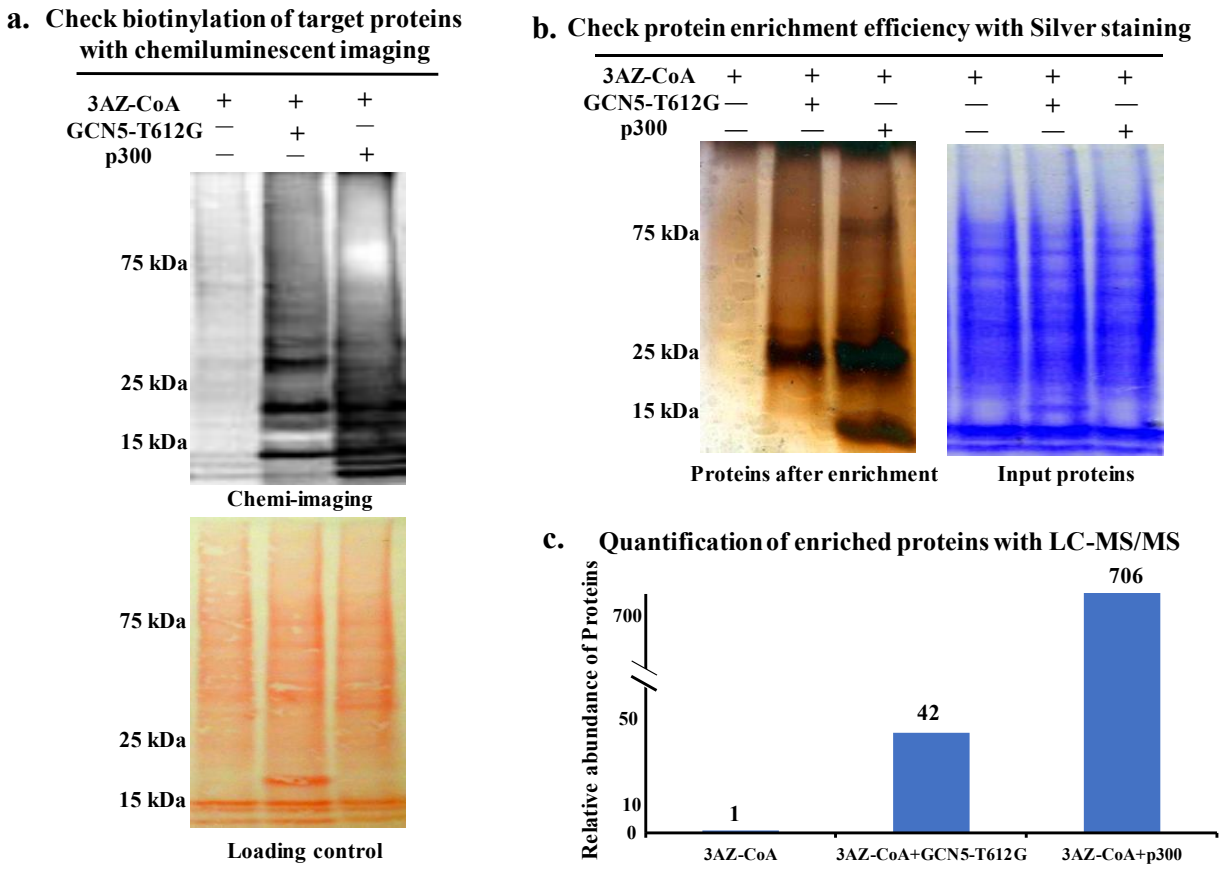


Figure 2.4. KAT substrate enrichment with biotin-streptavidin pulldown and quantification with semi-quantitative LC-MS/MS.

a. KAT substrates were biotinylated with the consecutive labeling strategy. Labeling efficiency and specificity were checked with biotin-streptavidin-HRP imaging. **b.** Biotinylated proteins were enriched with biotin-streptavidin pull-down followed by Na2S2O4 cleavage. Efficiency of KAT substrates enrichment was checked with silver staining. **c.** Enriched proteins were quantified using an LC-MS/MS semi-quantitative technique. p300 and GCN5-T612G treatment enriched proteins by 42 and 706 folds compared with the control group.

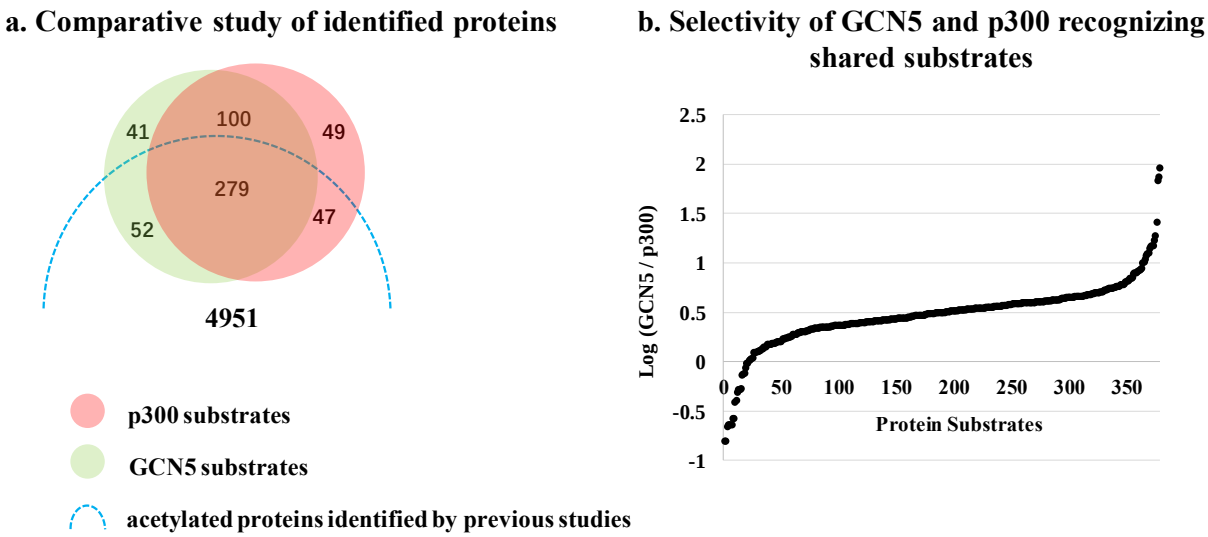
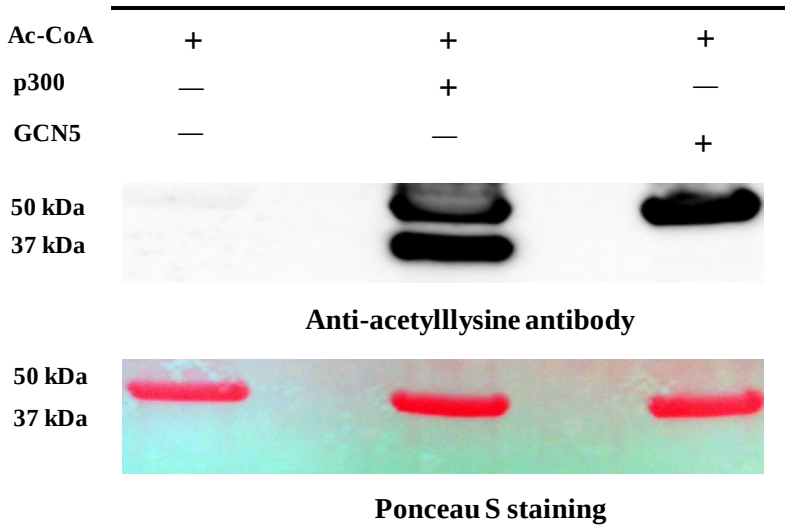


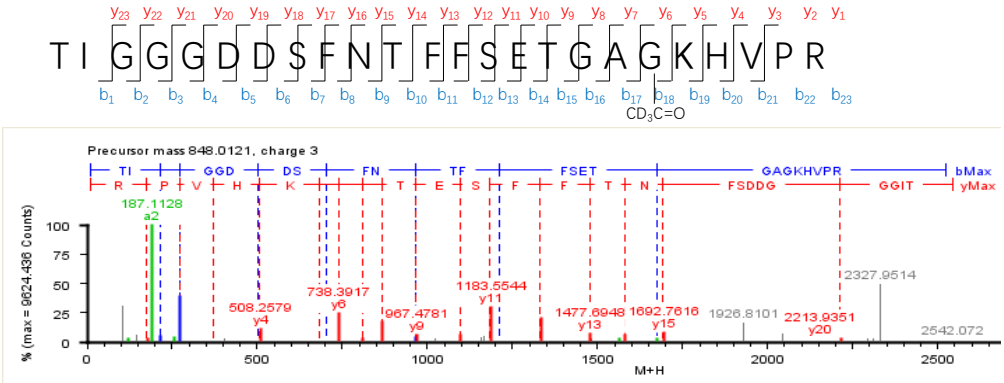
Figure 2.5. Comparative analysis of identified KAT substrates.

a. Identified GCN5 and p300 substrates were compared with each other and previously identified acetylome. *(Ref 34)* **b.** For the shared substrates, the logarithmized ratio of their relative abundances in GCN5 and p300 sets were calculated to compare the selectivity of GCN5 and p300 with the substrates

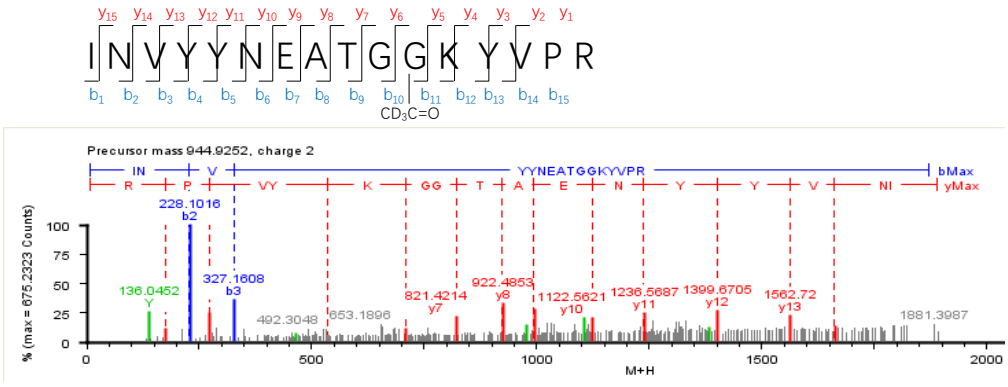
a.



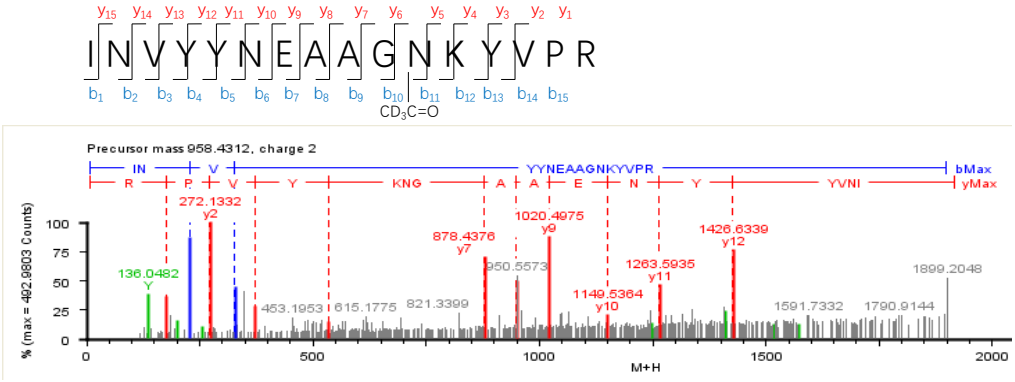
b. Peptide m/z 848.0121



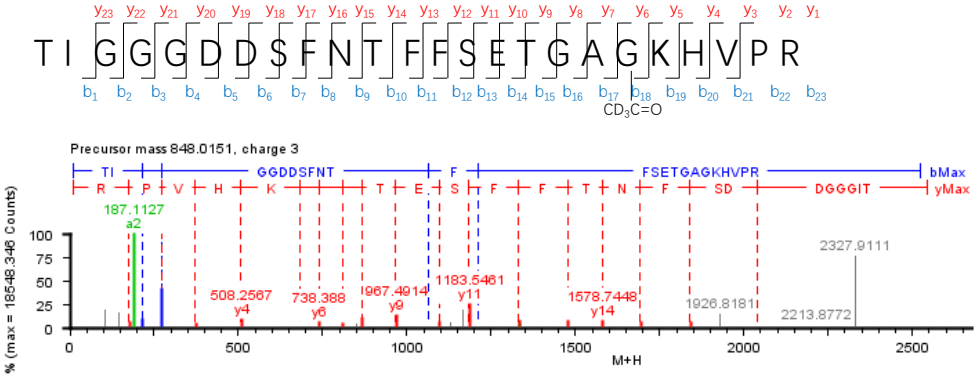
c. Peptide m/z 944.9252



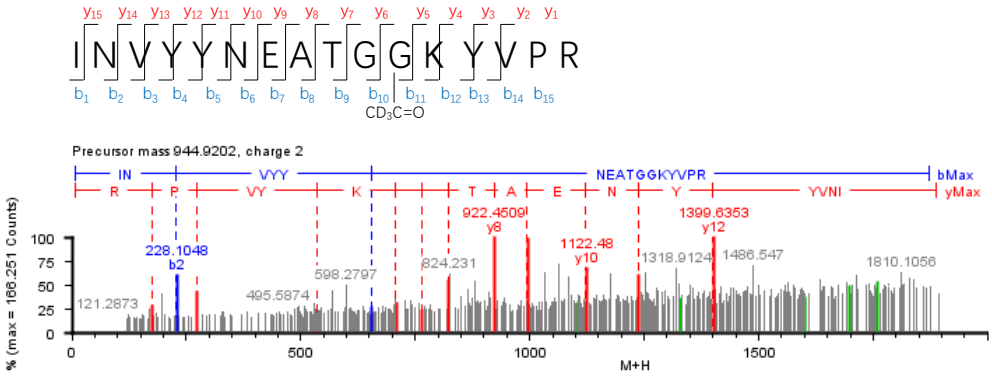
d. Peptide m/z 958.4299



e. Peptide m/z 848.0151



f. Peptide m/z 944.9202



g. Peptide m/z 958.4299

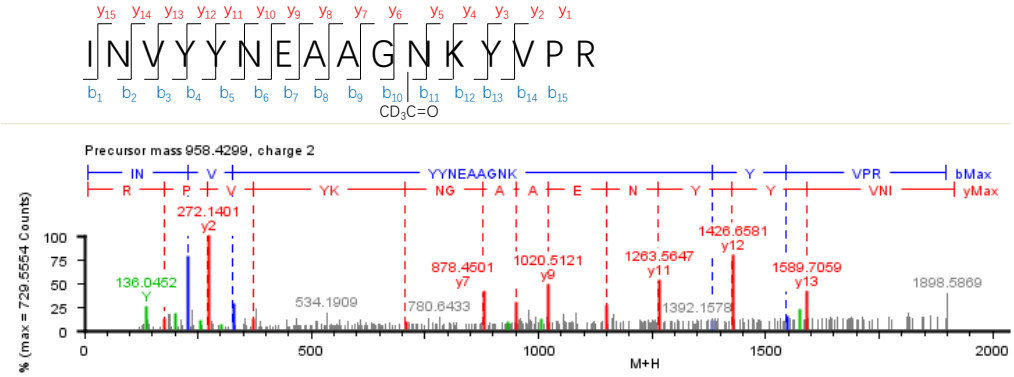


Figure 2.6. Detection of GCN5 and p300 acetyltransferase activities on tubulin.

a. Treatment of p300 and GCN5 resulted in strong acetylation of tubulin. The chemiluminescent protein band right below tubulin is autoacetylated p300. Spectra **b, c, and d** are from p300 treated tubulin and **e, f, and g** are from GCN5 treated tubulin samples. **a&d.** Peptide TIGGGDDSFNTFFSETGAG K60(d3-ac) HVPR from alpha tubulin (tubulin alpha chain, alpha-1D chain and alpha-1B chain). **b&e.** Peptide INVYYNEATGGK59(d3-ac)YVPR from beta tubulin (tubulin beta-4B chain). **c&f.** Peptide INVYYNEAAGNK59 (d3-ac)YVPR from beta tubulin (tubulin beta chain). In panels **b-g**, **b-(blue)** and **y-(red)** and **a-(green)** ion series derived from the precursor ions were marked. Peaks that are not labeled include internal fragments, masses losing a molecule of water or ammonia and baseline noises.

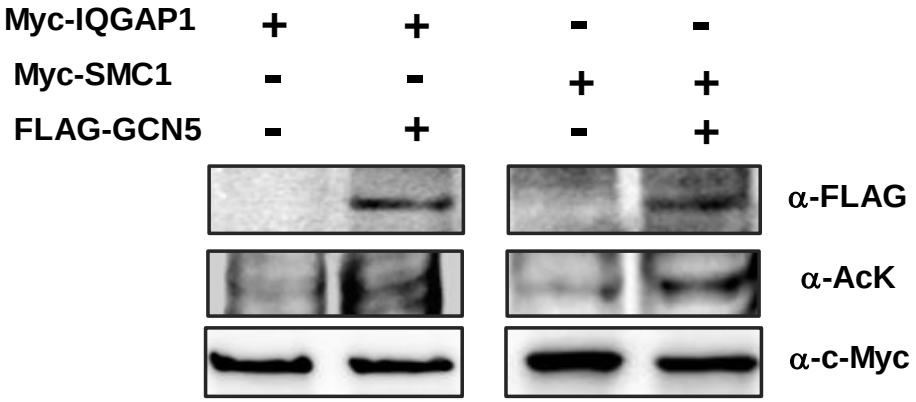


Figure 2.7. Detection of *in vivo* acetylation of IQGAP1 and SMC1 by GCN5.

Expression level and acetylation level of tagged IQGAP1, SMC1, and GCN5 were detected with western blotting. The acetylation level of IQGAP1 and SMC1 was detected using pan anti-acetyllysine antibody.

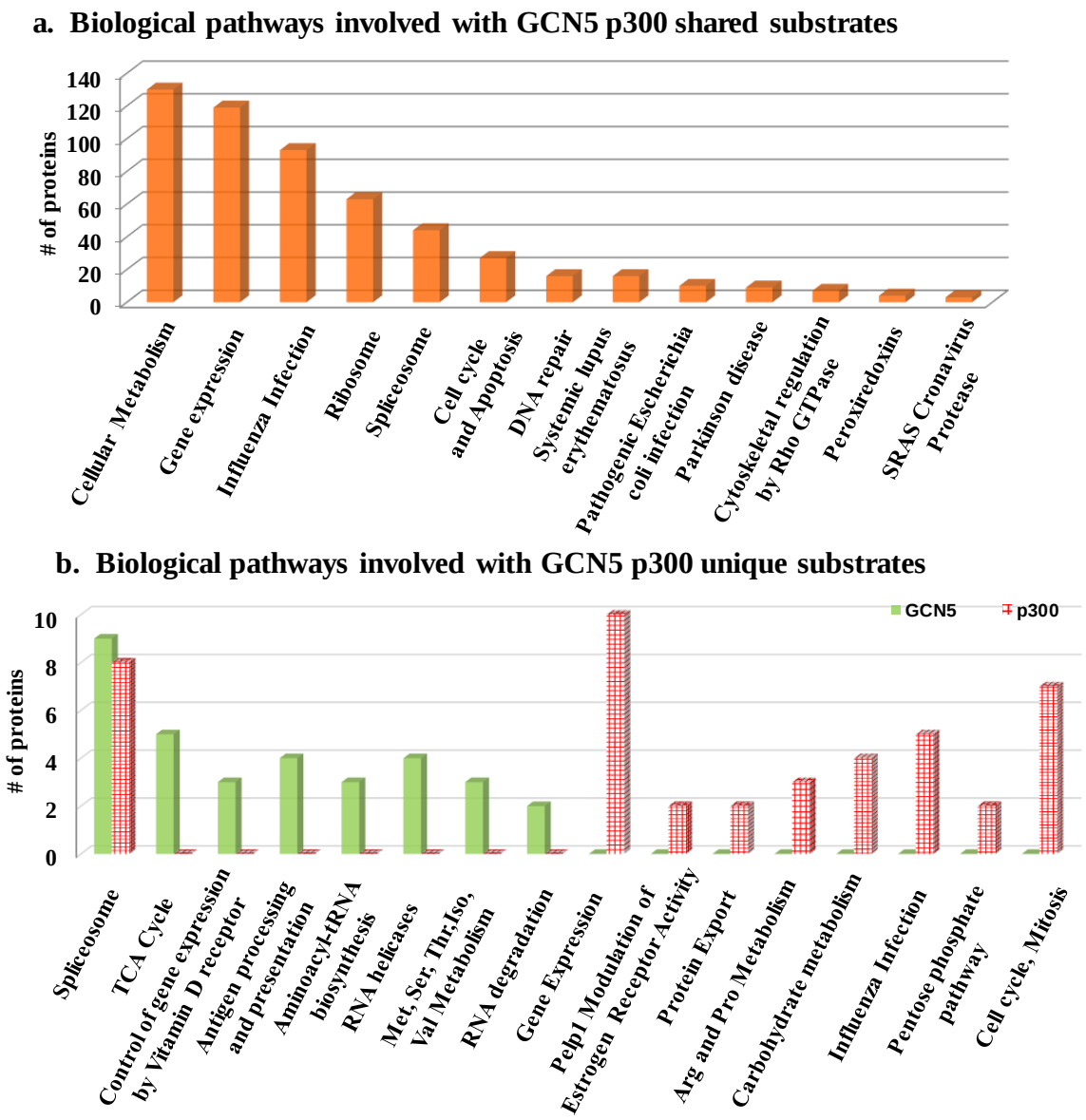


Figure 2.8 Functional annotation of enriched KAT substrates.

a. GCN5 and p300 shared substrates have broad involvement in cellular metabolism, gene expression, etc. **b.** GCN5 and p300 unique substrates show significantly different profiles in pathway involvements.

Table 2.1. Kinetic parameters of acetyl-CoA and synthetic orthogonal acyl-CoA in KAT-mediated histone modifications.

Enzyme	Cofactor	K_m / μM	k_{cat} / min^{-1}	k_{cat}/K_m ($min^{-1}.\mu M^{-1}$)
GCN5	Ac-CoA	3.7±1.0	47.3±4.3	12.8±3.7
	3AZ-CoA	8.2±2.8	0.3±0.04	0.04±0.02
GCN5-T612G	Ac-CoA	4.0±0.1	38.5±0.3	9.6±3.7
	4PY-CoA	0.9±0.1	12.5±0.5	14.7±5.6
	5HY-CoA	0.6±0.03	14.4±0.2	22.5±8.6
	6HY-CoA	0.3±0.04	3.8±0.1	13.1±5.0
	3AZ-CoA	**0.2±0.04**	**8.9±0.3**	**44.5±17.0**
p300	Ac-CoA	17.3±4.8	34±3.0	2.0±0.8
	3AZ-CoA	**8.6±0.4**	**14.0±0.2**	**1.6±0.6**
	4PY-CoA	2.4±0.3	0.4±0.001	0.2±0.07
	5HY-CoA	—[a]	—[a]	—[a]
	6HY-CoA	—[a]	—[a]	—[a]

a. These values could not be determined due to weak activity.

All supplementary tables are available online at:

http://pubs.acs.org/doi/suppl/10.1021/acschembio.7b00114

CHAPTER 3. Integration of Bioorthogonal Probes and Q-FRET for the Detection of Lysine Acetyltransferase Activity

---

Han, Z., Luan, Y., and Zheng, G.Y. 2015. *ChemBioChem*. 16 (18), 2605–2609.

Reprinted here with permission of the publisher.

Abstract

Lysine acetyltransferases (KATs) are key players in the epigenetic regulation of gene function. The recent discovery of diverse KAT substrates implies a broad spectrum of cellular functions of KATs. Many pathological processes are also intimately associated with the dysregulation of KAT levels and activities. However, detecting the enzymatic activity of KATs has been challenging, and this has significantly impeded drug discovery. To advance the field, we developed a convenient one-pot, mix-and-read strategy that is capable of directly detecting the acylated histone product through a fluorescent readout. The strategy integrates three technological platforms-bioorthogonal KAT substrates labeling, alkyne-azide click chemistry, and quenching FRET-into one system for effective probing of HAT enzyme activity.

Introduction

Lysin acetyltransferases (KATs) are important posttranslational modification enzymes that charge acetyl groups to substrate lysine residues; with most well studied in nucleosomal histones. Since their genetic identification in the mid-1990s, KATs have been extensively researched particularly in terms of structural characterization, biochemical properties, and functions in physiology and pathology.[84-88] Based on their conserved domain sequences, KATs are categorized into several major families, including the GCN5/PCAF family, the p300/ CBP family, and the MYST (MOZ, Ybf2/Sas3, Sas2, Tip60) family. Each KAT member transfers the acetyl group from acetyl-coenzyme A (acetyl-CoA) to the epsilon amino group of specific lysine residues in their protein substrates. Owing to their capability of acetylating the nucleosomal histones, KATs are important epigenetic regulators for the control of DNA transcription, replication, and damage repair.[89,90] (In recent years lysine acetylation has

become better appreciated as a common post-translational modification in proteins, far out of the chromatin realm.) Mass spectrometry-based proteomic screening has identified a vast number of non-histone substrates, thus suggesting the general significance of lysine acetylation in biology [50,91-93] As the pathological significance of KATs has become increasingly clear, it is expected that many unknown functions of KATs in the cell will be elucidated. Deregulation of KAT levels and activities have been observed in and are associated with many disease phenotypes such as inflammation,[94] diabetes mellitus,[13,95] obesity,[96] neurological disorder,[97] cancer,[14,86] hematologicdiseases,[85,98,99] cardiovascular disorders,[100,101] and viral infection,[102].

Given the physiological and pharmacological significance of KATs, it is imperative to develop isoform-selective KAT inhibitors as either mechanistic tools or therapeutic lead agents. To move this field forward, reliable and efficient biochemical methods that can be applied to detect KAT activity and screen thousands of library compounds are critically needed. Simplicity, sensitivity, dose-dependent linearity, and signal robustness are often important factors to consider in the design of enzymatic assays. Nevertheless, the detection of KAT activity is technically challenging: the transferred acetyl group is both chemically and spectroscopically inert so that there are no straightforward detectable readouts under routine assay settings. Till now, different strategies had been developed to study KAT activity based on different detection principles, each of which has pros and cons. Radiometric assays relying on isotope-labeled acetyl-CoA represent the gold standard in the characterization of KAT activities because of their relatively high sensitivity.[103-105] Handling radioactive hazards, however, is a concern for many laboratories. Quantitation of CoA, the side product of the KAT reaction, by using either chemical- or enzyme-coupled spectroscopic methods is also widely adopted. Antibody-based immunosorbent assays are sensitive and commonly applied by molecular biologists who study

biological functions of lysine acetylation.[106] However, the high cost of antibody production and sacrifice of lab animals limit their use in large-scale applications. It should be noted that many assay types require products to be separated from the reaction mixture before signal detection; this is a painful technical bottleneck in a high-throughput study. Therefore, homogeneous, spectroscopic, mix-and-read assays for KAT activity detection are the preferred approach.[43,107]

In this study, we demonstrate a fluorescent assay strategy for KAT activity detection by aligning several chemical biology techniques into one system: bioorthogonal protein acetylation labeling, alkyne-azide click chemistry, and quenching-fluorescence resonance energy transfer (Q-FRET). As is shown in **Figure 3.1**, the first step is to create a bioorthogonal enzyme-cofactor matching pair so that the KAT enzyme can take up orthogonal acyl-CoA cofactors for histone modification. We recently identified several active bioorthogonal pairs for the KAT members MOF, GCN5 and p300.[42] Interestingly, 3-azidopropanoyl CoA (3AZ-CoA) is a highly effective surrogate of acetyl-CoA that can be used as a cofactor by several KAT enzyme forms. Interestingly, 3-azidopropanoyl CoA (3AZ-CoA) is a chemically active surrogate of acetyl-CoA that can be used as a cofactor by several KAT enzyme forms. We decided to carry out the enzymatic reaction with a histone substrate containing a fluorescent donor (e.g., fluorescein). Upon KAT-catalyzed acylation, the added 3-azidopropanoyl group in the substrate can be treated with a FRET acceptor chromophore though alkyne–azide cycloaddition click chemistry. In this way, the donor and acceptor will be brought into spatial proximity within the same molecule, leading to FRET signal transduction. The advantage of this new strategy is that the intramolecular FRET mechanism occurs in situ as a function of the progress of the KAT-catalyzed reaction. This will be a straightforward mix-and-read KAT assay approach, without any tedious washing procedure, that possesses highthroughput capacity.

Result and Discussion

As a proof of concept, we chose the MYST enzyme MOF (male absent on the first, MYST1, HAT8/KAT8) to demonstrate the FRET detection strategy. MOF is the key acetyltransferase responsible for the acetylation of lysine-16 of histone H4.[108] We recently determined that active site engineering of MOF by mutating its isoleucine-317 to alanine expanded the cofactor binding site resulting in a consequence that MOF-I317A was able to utilize bulky acetyl-CoA surrogates such as 3AZ-CoA to modify histone substrates.[42] Indeed, kinetic characterization on histone H4 peptide modification showed that MOF-I317A mutant exhibited a strong activity toward 3AZ-CoA, twice that of acetyl-CoA **(Table 3.1)**. In contrast, wt-MOF had good activity for the nascent cofactor acetyl-CoA, but with minimal activity toward 3AZ-CoA. Therefore, MOF-I317A-3AZ-CoA is a superior bioorthogonal pair for MOF substrate labeling. Importantly, the clickable functionality of the 3-azidopropanoyl group offers a great power for downstream detection of modified substrates by using fluorescent reporters or affinity tags through copper-catalyzed azide–alkyne cycloaddition (CuAAC) [109] or strain-promoted azide–alkyne cycloaddition (SPAAC).[110]

To construct the one-pot assay, we used a fluorophore-labeled histone peptide substrate (i.e., H4-FL) containing the 20 N-terminal residues of histone H4 but with Leu10 replaced with a fluorescently labeled lysine. First, 3AZ-CoA and H4-FL were incubated with different concentrations of MOF-I317A for 1 h to allow the acylation to occur. After the reaction, a cocktail of click reagents that contained propargyl-DABCYL, copper sulfate, ligand 3-(4-((bis((1-tert-butyl-1H-1,2,3-triazol-4-yl)methyl)amino) methyl)-1H-1,2,3-triazol-1-yl)propanol (BTTP), and sodium ascorbate in dimethyl sulfoxide (DMSO) was added. After further incubation for 1 h, 0.25 M guanidine hydrochloride was added in order to eliminate any

nonspecific interaction between the cofactor and the histone substrate.[105] In the final step, the fluorescence emission spectra were measured with a fixed excitation wavelength of 491 nm. As expected, emission of fluorescein decreased as more enzymes were added **(Figure 3.2a)**; this illustrates the effectiveness of this detection platform: the enzyme delivered the 3-azidopropanoyl group to Lys16 of the H4 substrate, then propargyl-DABCYL specifically reacted with the azide (CuAAC) to form a covalent triazole linkage; in this way the quencher chromophore was brought into close proximity to the fluorescein donor, and thus an intramolecular quenching-FRET process between DABCYL and fluorescein was introduced, diminishing the fluorescence of the donor. Impressively, the fluorescence intensity changed linearly as a function of enzyme concentration **(Figure 3.2b)**; this indicates that the quenching effect is proportional to the yield of acylated histone product.

The fluorescent changes in response to the enzymatic reaction time were also tested. At different time points of reaction, a DMSO solution of click reagent was added to quench the enzymatic reaction and meanwhile initiate the alkyne-azide cycloaddition reaction. Next, 0.25 M of guanidine hydrochloride was added and the fluorescence intensity was measured. As shown in **Figure 3.3**, the fluorescence intensity decreased linearly as a function of enzymatic reaction time; this further supports the hypothesis that this assay offers a quantitative approach for detecting the product of MOF catalysis. To test whether this method could be used to determine the kinetic parameters of substrates, a range of 3AZ-CoA concentrations were selected for the enzymatic reaction followed by fluorescent measurement. Increasing 3AZ-CoA concentration led to a hyperbolic decrease in fluorescence intensity because more azidoacyl groups were transferred to H4-FL **(Figure 3.4)**. The hyperbolic curve was fitted to a modified Michaelis–Menten equation

to give an apparent K_m value of 1.5±0.1 μM, which was within range of the K_m of 2.7±0.2 μM that was measured by using a regular 7-diethyamino-3-(4'-maleimidylphenyl)-4-methylcoumarin (CPM)-CoA fluorogenic assay **(Table 3.1)**.

The excellent performance of this FRET strategy for detecting KAT activity suggests that it could easily be used for KAT inhibitor screening and optimization. We tested whether our new approach can be used to quantify the potency of anacardic acid—a known inhibitor of the MYST KATs [111] and p300 [112]—in MOF inhibitior. In this experiment, different concentrations of anacardic acid were present in the H4 acylation reaction catalyzed by MOF-I317A. We anticipate that increasing the inhibition of MOF activity by higher concentrations of anacardic acid would reduce the transfer rate of the 3-azidopropanoyl group to H4-FL. Consequently, less propargyl-DABCYL would be recruited to the substrate, and the FRET interaction would be reduced, with concurrent higher fluorescence emission. Indeed, the fluorescence intensity increased as a function of anacardic acid concentration **(Figure 3.5a)**. By fitting the dose response curve to the modified inhibitor concentration-response equation, the IC_{50} value was calculated to be 3.2 ± 0.5 μM. This value is almost the same as the IC_{50} value (3.4 ± 0.5 μM) measured using the standard radiometric acetylation assay in which [^{14}C]Ac-CoA was the cofactor **(Figure 3.5b)**. Therefore, the quenching-FRET assay is a valid approach for studying the KAT inhibitors. Given that the development of selective chemical probes of KATs has lagged in epigenetic drug discovery, biochemical assays that effectively detect KAT activity are of significant value to discovering KAT probes in a high-throughput manner. We expect that the new design will help fill the gap in this important biomedical arena.

Conclusion

In summary, by welding several technologies, that is, bioorthogonal KAT substrate labeling, CuAAC click chemistry, and quenching-FRET, into one platform, we have developed a homogeneous one-pot fluorescent assay approach for detecting KAT activity. This method directly measures the acylated histone product, has advantages over the current CPM fluorescent method that measures the by-product, CoA.[43,113] The latter assay is easily susceptible to background fluorescence interference, and the instability of acetyl-CoA can lead to false positive results. With our new approach, the fluorescent readout is linearly proportional to the yield of the acylated product. The distance-dependent intramolecular FRET mechanism suggests that this assay type is more suitable to organic dyes in inhibitor screenings. Protein acylation is an emerging area of science[114], and the high- performance assay strategy presented here has great potential to be widely applied to other KATs and acyltransferases.

Methods and Materials

Preparation of 3AZ-CoA, MOF mutagenesis, protein expression of MOF-wt and MOF-I317A in BL21(DE3) bacteria cells were done according to previous procedures.[42] Histone peptides H4-20 (Ac-SGRGKGGKGLGKGGAKRHRK) and H4-FL (Ac-SGRGKGGKGK (fluorescein) GKGGAKRHRK) peptides were synthesized according to standard solid-phase peptide synthesis (SPPS) protocols, purified by C18 RP-HPLC and characterized by MALDI-MS, as previously described.[115]

Synthesis and characterization of propargyl-Dabcyl

4-([4-(Dimethylamino)phenyl]azo) benzoic acid succinimidyl ester (DABCYL SE) (50 mg, 136.5 μmol, 1.0eq.), propargyl amine (15mg, 273 μmol, 2.0eq.) and triethylamine (20 μL, 136.5 mol, 1.0 eq.) were mixed in 3.5 mL of anhydrous dimethylformamide (DMF) and stirred

overnight at room temperature. Then, DMF was removed with rotary evaporation and product was dissolved with methanol and loaded onto SiO_2 column. Column was then washed with solvent of hexane: ethyl acetate = 1:2 and orange eluent was collected. Solvent was removed with rotary evaporation and lyophilization. Product structure was confirmed by ESI-MS and NMR.

Enzyme-cofactor kinetic characterization

H4-20 peptides and cofactors at various concentrations were incubated in reaction buffer containing HEPES (50 mm, pH 8.0) and EDTA (0.1 mm) at 30 ^{o}C. Enzyme (0.05 μM) was added to initiate the reaction, which lasted for 10 min. After that, CPM (10 μM) in DMSO was added both to quench the enzymatic reaction and to initiate the CoA CPM reaction. After 20 min, fluorescence intensities were measured by using a microplate reader with fixed λ_{ex}=392 nm and λ_{em}= 482 nm. Fluorescence readouts were plotted against cofactor concentrations with a modified Michaelis–Menten equation $F = F_0 + \frac{(F_f - F_0)*[S]}{K_m+[S]}$ to get the apparent K_m value.

Measurement of IC_{50}

For the filter binding assay, reaction mixture (30 mL), which included various concentrations of anacardic acid, MOF-I317A (0.5 μM), [^{14}C] Ac-CoA (5 μM), and H4-FL (5 μM) was incubated at 30 ^{o}C for 30 min. Then, a sample (20 mL) was spotted onto a p81 filter paper disk. The disks were washed with $NaHCO_3$ (50 mM, pH 9.0) and air dried. Liquid scintillation counting was performed to quantitate the amount of acetylated peptides. The IC_{50} value was calculated by using the inhibitor concentration–response equation:

$$\text{Fration of inhibited enzyme activity} = \frac{1}{1 + (\frac{[I]}{[IC_{50}]})^h}$$

For the IC_{50} measurements using Q-FRET assay, anacardic acid at various concentrations was incubated with MOF-I317A (0.5 μM), 3AZ-CoA (5 μM), and H4-FL (5 μM) at 30 °C for 30 min (20 μL reaction mixture). Click reagent DMSO solution (25 μL) was added, then the mixture was incubated for 1 h at room temperature in the dark. The fluorescence intensity was measured with a CLARIO star microplate reader (BMG LABTECH, Cary, NC, USA) in the fluorescein channel after the addition of guanidine hydrochloride (0.25 M). Fluorescence intensities were plotted against inhibitor concentrations by using a modified inhibitor concentration– response equation: $F_0 - F = \frac{F_0 - F_f}{1+(\frac{[I]}{[IC_{50}]})^h}$.

Figures, Tables and Captions

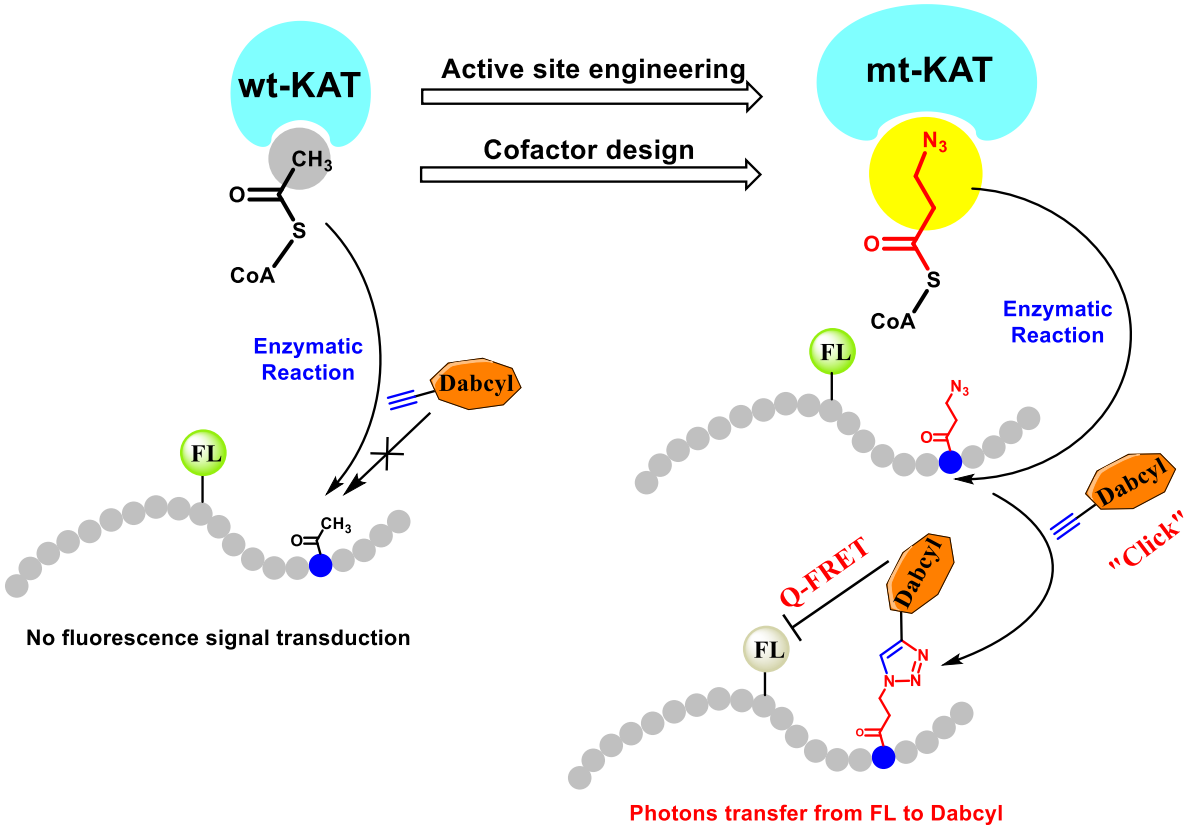


Figure 3.1 Using a quenching fluorescence resonance energy transfer (Q-FRET) assay in conjugation with bioorthogonal substrate labeling to detect KAT activity.

wt-KAT=wild type lysine acetyltransferase, mt-KAT=mutant lysine acetyltransferase, FL=fluorescein.

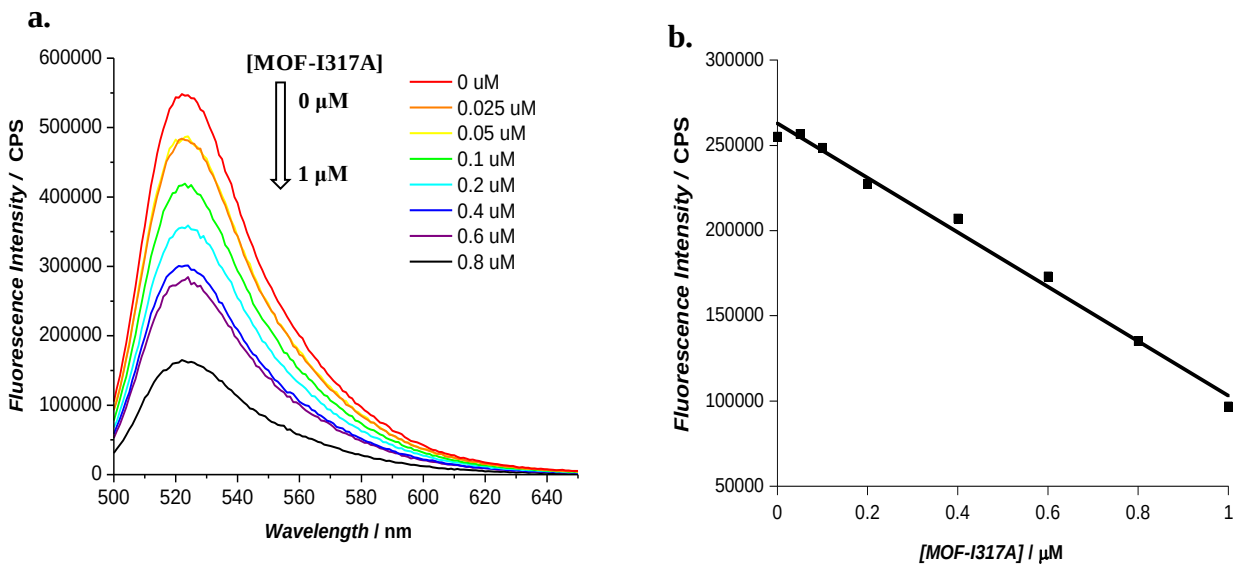


Figure 3.2. The quenching FRET is dependent on the enzyme concentration.

MOF-I317A at various concentrations was incubated with 5 µM of H4-FL and 10 µM of 3AZ-CoA followed by incubation with propargyl-DABCYL. The mixture was excited at 491 nm, then emission spectra were collected at each enzyme concentration. **a.** Fluorescence spectra with λ_{ex}=491 nm. **b.** Linear relationship between fluorescence intensity at 520 nm and enzyme concentration.

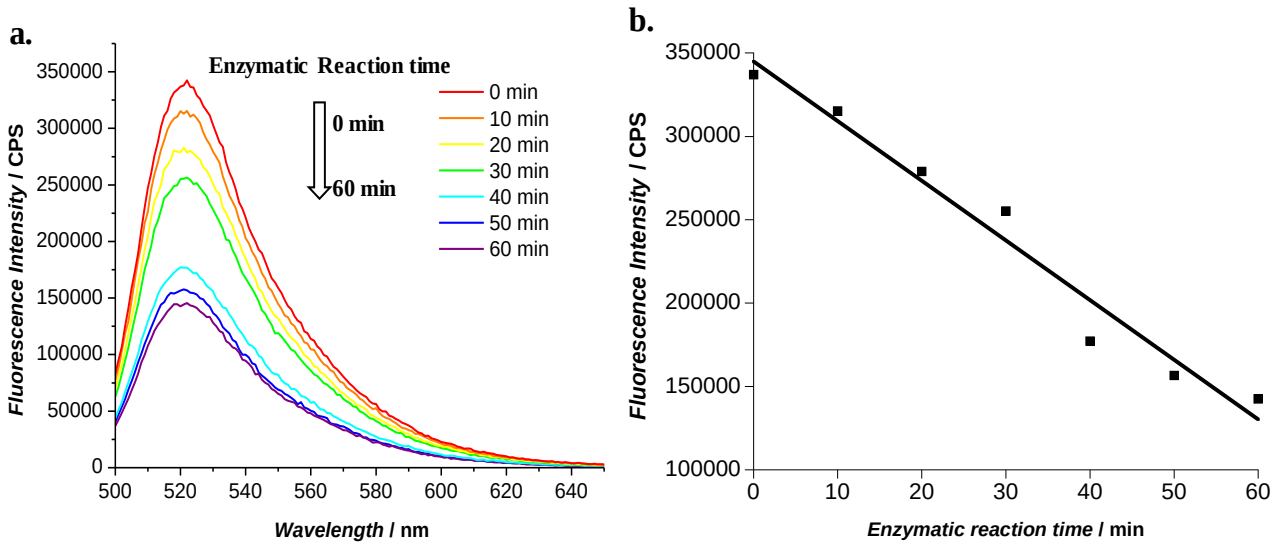


Figure 3.3. Fluorescent changes in response to enzymatic reaction time.

Enzymatic reaction was initiated by mixing MOF-I317A, H4-FL and 3AZ-CoA and was quenched at different time points. Excess alkyne DABCYL was added to the modified peptides to form quenching-FRET pairs with the anchored fluorescein. **a.** Changes in fluorescence spectra in response to reaction time with λ_{ex}=491 nm. **b.** Linear relationship between fluorescence intensity at 520 nm and enzymatic reaction time.

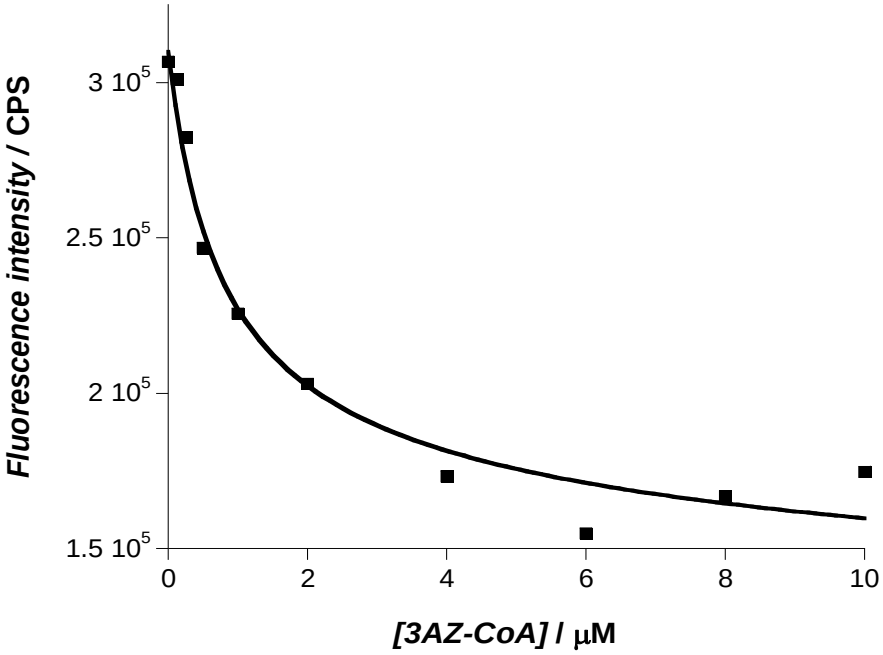


Figure 3.4. Fluorescence intensity as a function of the concentration of 3AZ-CoA.

3AZ-CoA at different concentrations was incubated with 0.7 μM of MOF-I317A and 5 μM of H4-FL to produce various amounts of azide-labeled H4-FL. Propargyl-DABCYL was then conjugated to the modified peptides, and the fluorescence was measured. Fluorescence decreases with increasing concentration of 3AZ-CoA due to the increasing Q-FRET effect. The obtained K_m value of MOF-I317A with 3AZ-CoA is 1.5 ± 0.1 μM.

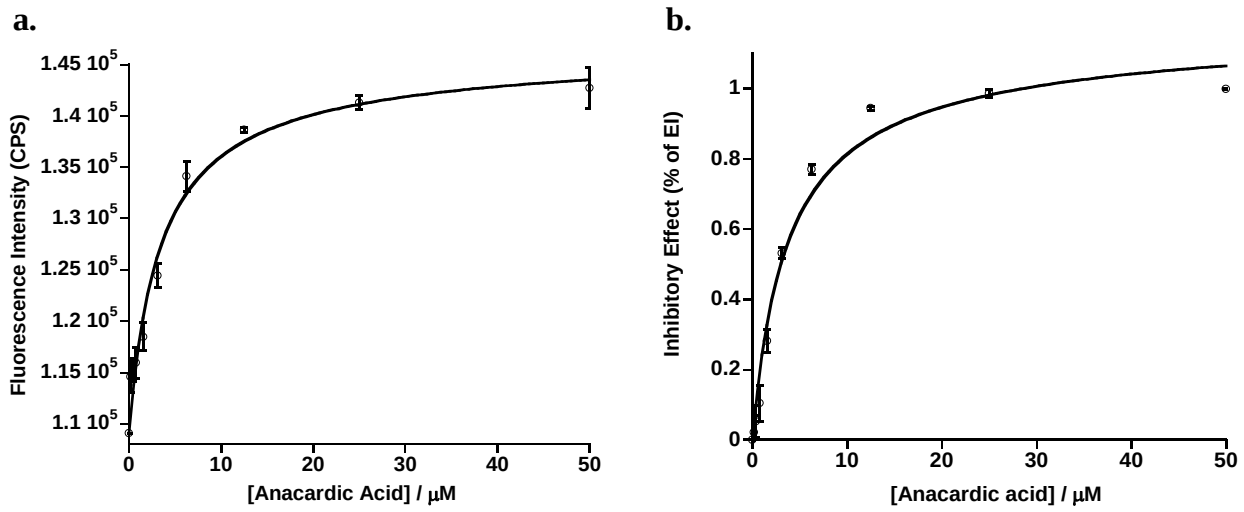


Figure 3.5. Quantitation of anacardic acid potency for HAT inhibition.

a. Using Q-FRET to measure the IC_{50} value of anacardic acid (enzymatic reaction components included MOF-I317A, H4-FL and 3AZ-CoA); the obtained IC_{50} was 3.2±0.5 μM. **b.** Using a filter binding assay to measure IC_{50} value of anacardic acid (enzymatic reaction components included MOF-I317A, H4-FL and ^{14}Ac-CoA); the measured IC_{50} of anacardic acid was 3.4±0.5 μM.

Table 3.1. Kinetic parameters of MOF for the nascent and orthogonal cofactors.

Enzyme	Cofactor	K_m/μM	k_{cat}/min^{-1}	k_{cat}/K_m (μM^{-1}.min^{-1})
MOF-wt	Ac-CoA	1.3 ± 0.1	5.2 ± 0.1	4.0 ± 1.5
	3AZ-CoA	—	—	—
MOF-I317A	Ac-CoA	3.5 ± 0.2	2.2 ± 0.0	0.6 ± 0.2
	3AZ-CoA	2.7 ± 0.2	3.2 ± 0.1	1.2 ± 0.5

"— "means the values were unmeasurable because of the low reading.

CHAPTER 4 Revealing the Lysine Propionylation Activity of the Lysine Acetyltransferase MOF

Han, Z. and Zheng, G.Y., et al. Submitted to *Journal of Biological Chemistry*, 10/20/2017

Abstract

Short-chain acylation of lysine residues recently emerges as a group of reversible posttranslational modifications in mammalian cells. The diversity of acylation further broadens the landscape and complexity of the proteome. Identification of regulatory enzymes and effector proteins for lysine acylation is critical to understand functions of these novel modifications at the molecular level. Here, we report that the MYST family of lysine acetyltransferases (KATs) possesses strong propionyltransferase activity both *in vitro* and *in vivo*. In particular, the propionyltransferase activity of MOF, MOZ, and HBO1 is as strong as their acetyltransferase activity. Overexpression of MOF in human cells induced significantly increased propionylation in multiple histone and non-histone proteins, which manifests that the function of MOF goes far beyond its canonical histone H4 lysine-16 acetylation. We also resolved the X-ray co-crystal structure of MOF bound with propionyl coenzyme A (propionyl-CoA, Pro-CoA) which provides a direct structural basis for the propionyltransferase activity of the MYST KATs. Our data together defines a novel function for the MYST KATs as lysine propionyltransferases and suggests much broader physiological impacts for this family of enzymes.

Introduction

Short-chain acylations of lysine residues in cellular proteins such as acetylation, propionylation, and crotonylation are reversible posttranslational modifications (PTMs) that modulate functions and properties of protein targets [8]. The variety of lysine acylation has been proposed to correlate with divergent biological outputs [6,25]. Etiologically, lysine acylations rely on acyl-coenzyme A (acyl-CoA) molecules to serve as the acyl donors, which are key metabolic intermediates in the Krebs cycle, fatty acid oxidation, and amino acid degradation [116]. The level of acyl-CoA intermediates fluctuates with nutritional status and altered activities of metabolic

enzymes. Toxic accumulation of acyl-CoA molecules is associated with a broad variety of complications [4,117]. A typical example is the propionic acidemia, an autosomal recessive metabolic disorder caused by the deficiency of propionyl-CoA carboxylase (PCC) [118]. Increase of propionyl-CoA and protein lysine propionylation level were reported in PCC deficient patient samples, but the disease relevant targets and the corresponding regulatory enzymes are not defined [4]. It is also increasingly recognized that the gut microbiota-related short-chain fatty acids interacts with and modulates the mammalian epigenetic machinery, especially histone acylation and methylation [119,120]. Nevertheless, it remains poorly understood how the alterations in acyl-CoA metabolites and protein acylation levels are causatively linked with the related diseased phenotypes. Understanding the regulatory mechanisms of short-chain lysine acylation such as lysine propionylation will have profound significance to elucidate the pathophysiological mechanisms of metabolic diseases.

In understanding the mechanistic links of acylation diversity with varied physiological outputs, a significant challenge is to resolve how different acylations are deposited, interpreted, and erased in the cell. In eukaryotes, protein acetylation marks are introduced by lysine acetyltransferases (KATs), which are classified into three major families based on primary amino acid sequence and domain organization including the MYST, the PCAF/GCN5, and the p300/CBP family [10]. These KATs, together with lysine deacetylases (KDACs), orchestrate the dynamics of lysine acetylation in the cell. Numerous studies have shown that aberrant expression and dysfunction of KATs are associated with various kinds of disease phenotypes such as inflammation, neurodegenerative disorders and cancers [40,121,122]. In 2007, Chen et al. detected lysine propionylation in human histone H4 and p53 using high resolution mass spectrometry (MS) and found that p300 and CBP can catalyze protein propionylation [28]. Following that study,

GCN5 and PCAF, members of the GNAT superfamily, were also found to possess propionyltransferase activity [26,27,123]. These findings largely increased mechanistic complexity and diversity of the KAT related physiological activities. The MYST KATs represent a distinct family whose sequence and function are dramatically different from the members of p300/CBP and PCAF/GCN5 families. It remains an open question whether or not the members of the MYST family possess lysine propionyltransferase (KPT) activity **(Figure 4.1)**.[124] In this work, we have conducted a combined suite of biochemical, cellular, and structural experiments and demonstrated that the MYST KATs have genuine KPT activity both *in vitro* and *in vivo*.

Result and Discussion

Propionyl-CoA is an abundant metabolite in human cells.

Several novel acylations other than acetylation on protein lysine residues have been identified, such as propionylation, butyrylation, crotonylation, succinylation, and malonylation [23,28,125]. Thus far, it remains poorly understood as regards to the cellular abundance of the acyl-CoA molecules for corresponding acylations. To understand the regulation of protein propionylation, we first quantified and compared the cellular abundance of endogenous propionyl-CoA and acetyl-CoA in the human embryonic kidney 293T (HEK293T) cells. Following cell culture, acyl-CoA molecules were extracted in 50% methanol and subjected to the LC-MS detection. Deuterated acetyl-CoA and propionyl-CoA were used as internal standards for calibration. Triplicate experiments were performed and showed that the cellular propionyl-CoA concentration is 92 ng/mg cellular proteins and acetyl-CoA is 754 ng/mg cellular proteins. Thus, the abundance of propionyl-CoA is about 12% that of acetyl-CoA **(Figure 4.2)**. Considering that acetyl-CoA is a rich and principal metabolite for cell growth, propionyl-CoA, even 8 times less abundant than acetyl-CoA, has the potential to serve as an ample source for lysine propionylation.

The recombinant MYST KATs have strong lysine propionyltransferase activity.

A few studies have demonstrated that GCN5/PCAF and p300/CBP members were able to carry out lysine propionylation, butyrylation, crotonylation, etc. in addition to their intrinsic acetylation activity [25,26,28]. In contrast, although the MYST enzymes represent the largest KAT family in human cells, their cofactor promiscuity is much less studied. Very recently, MOF was shown to carry out butyrylation and crotonylation on histone substrates [29,126]. These evidence hints for the cofactor promiscuity of the MYST enzymes. Herein, we thoroughly investigated the novel acyltransferase activity of MYST KATs comparatively and from different perspectives. First, we screened and compared the KPT activity of the three families of KATs on their histone H3 or H4 peptide substrates (H3-20 and H4-20, the N-terminal 20-aa sequence of histone H3 and H4). A fluorogenic assay was used to determine the kinetic constants of KATs including K_m and k_{cat} with respect to acetyl-CoA or propionyl-CoA under the initial velocity condition [43]. The k_{cat}/K_m ratio was calculated to quantitatively compare the KPT and KAT activities of individual KATs. In consistence with the previous studies showing that PCAF/GCN5 and p300/CBP members possess KPT activity [26,27,127], our kinetic measurement showed that the KPT activity of GCN5 and HAT1 is almost equally strong compared to their KAT activity while the KPT activity of PCAF and p300 is about 40% and 30% of their acetyltransferase activity **(Table 4.1)**. Interestingly, the MYST KAT members, including MOF, HBO1, and MOZ, show strong KPT activity with ratios of their KPT/KAT activity to be 0.84, 0.99, and 1.04 respectively. The MYST KATs Tip60 and MORF also showed appreciable KPT activities, which are 12% and 17% of their KAT activity. After we demonstrated the propionyltransferase activity of the KAT enzymes on histone peptides, we then tested the KPT activity of the MYST KATs on the proteome of the 293T cell using an anti-propionyllysine antibody **(Figure 4.3)**. Compared with the untreated cell

lysate, sole propionyl-CoA co-incubation did not induce increase of lysine propionylation on cellular proteins while treatment of the cell lysate with propionyl-CoA and individual MYST KATs induced extensive propionylation of multiple histone and non-histone proteins. In particular, MOF showed the strongest KPT activity on cellular proteins. Together, the steady-state kinetic data and the western blot analysis of whole protein propionylation strongly demonstrate that the MYST KATs are *bona fide* KPTs. Next, we focus on MOF for further biological, structural, and biochemical investigation of the novel KPT activity of the MYST enzymes.

MOF acetylates and propionylates largely shared substrates with subtle variation.

After proving the KPT activity of the MYST enzymes on cellular histones and non-histones, we investigated the similarity of MOF-mediated sub-acetylome and sub-propionylome. Peptide proteomic analyses were applied to identify MOF acetylation and propionylation sites on the H4-20 peptide. Mono-acylated H4-20 is the major product from MOF catalyzed acetylation reaction and the di-acylated H4-20 also present with lower yield than mono-acylated product **(Figure 4.4,)**. Notably, the acylation sites were slightly different: K8, K12, and K16 were found acetylated while K8 and K16 were propionylated. In further detail, mono-acetylation of H4-20 occurs on K8 and K16 while di-acetylated peptides have acetylated K8 and K12, K16 and K12, and K8 and K16, respectively. For H4-20 propionylation, K8- and K16- mono-propionylated peptides were detected, and di-propionylated peptides have propionylated K8 and K16 simultaneously **(Table 4.2)**. Thus, both K8 and K16 are prone to undergo both acetylation and propionylation by MOF while K12 may be a weak acetylation site of MOF. Next, we tested MOF substrate specificity on the histone proteins and nucleosomes with western blot analysis. Using free histones as substrates, MOF was able to modify all the four core histones for both

acetylation and propionylation **(Figure 4.5a)** while MOF cannot acylate histone H3 on the nucleosomes **(Figure 4.5b)**. Both MOF catalytic domain and full-length MOF showed the same substrate specificity, suggesting that the chromo domain did not affect the interaction between MOF and nucleosomes. Accumulated studies indicated that the histone proteins are not the only substrate of MOF [6,37,128,129]; therefore, we next explored MOF substrate profile on cellular proteome. Since the extracted cellular proteome has gone through acylations by endogenous KAT enzymes, it is hard to accurately discriminate MOF sub-acylome from the bulk acylome with immunoblot assays. Thus, a radiometric gel assay was performed to compare the substrates acetylation and propionylation driven by MOF **(Figure 4.6)**. In accord with the western blot results shown in **Figure 4.3 and Figure 4.5**, little non-enzymatic labeling on lysine residues driven by the chemical reactivity of acetyl-CoA and propionyl-CoA was observed. Treatment of the cell lysate with acyl-CoA and MOF together induced strong acylation of multiple proteins from the cellular proteome. A highly identical modification pattern was observed between lysine acetylation and propionylation, which is suggestive that MOF acetylome and propionylome may largely overlap. MOF auto-acetylation was previously reported by Yang et al [130]. We observed that MOF underwent auto-propionylation with similar activity compared to auto-acetylation. The fact that multiple non-histone proteins were being acylated by MOF is in agreement with the previous findings that MOF targets both histone and non-histone proteins.[6,128,131-133] Overall, MOF may have highly identical acetylome and propionylome while the modification levels at individual lysine residues could differ to varying degrees.

MOF is a genuine lysine propionyltransferase *in vivo* with a distinct substrate profile.

The above biochemical experiments demonstrated that the recombinant MYST KATs catalyze propionylation of histone and non-histone proteins *in vitro*. We next focused to test

MOF propionyltransferase activity *in vivo* and determine the MOF propionylome. A model of MOF overexpression was created in HEK293T cells by lentivirus transient transfection **(Figure 4.7a)**. Following cell culture and cell lysate extraction, the acylation levels of cell proteome were examined with western blot analyses. Both lysine acetylation and propionylation in extracted histones were drastically upregulated in the presence of MOF overexpression **(Figure 4.7b)**. Particularly, MOF overexpression induced increased levels of acylation on both histone H4 and H2A/2B while histone H3 acylation did not change, which are consistent with the *in vitro* data shown in **Figure 4.5b**. Therefore, H3 is likely not a significant *in vivo* substrate of MOF. Of note, H4K16 propionylation was upregulated with MOF overexpression **(Figure 4.7c)**, in accord with the previous knowledge that H4K16 is a primary substrate of MOF.[134] We also observed that multiple non-histone proteins were acetylated and propionylated at enhanced levels in response to MOF overexpression, and the western blot profiles of substrate propionylation and acetylation are highly identical **(Figure 4.7d)**. This result is in consistence with the radiometric gel assay **(Figure 4.6)** and strengthens our hypothesis that MOF have shared substrates for KAT and KPT activities. In addition, we detected the change of lysine propionylation level on cellular proteome in response to varying propionyl-CoA concentrations **(Figure 4.8)**. The 293T cells were treated with sodium propionate, which can be converted to propionyl-CoA in the cells through the acyl-CoA synthase pathway,[135] and cellular proteins were extracted for western blot detection. Treatment of both the control and MOF-overexpressed cells with sodium propionate induced increase of lysine propionylation on both histone and non-histone proteins. Interestingly, the western blot patterns show appreciable difference between the control cells and MOF-overexpressed cells upon propionate treatment, suggesting that some proteins are only propionylated in the MOF-overexpressed cells. These results indicate that MOF may have a

distinct substrate propionylome and the unique MOF function cannot be compensated by other KPT enzymes. Together, MOF displayed *in vivo* KPT activity on both histone and non-histone proteins. The propionylation substrate profile of MOF is highly similar to that of acetylation, but is partially distinct from other KPTs.

Proteomic profiling of MOF-mediated propionylome.

Tandem MS was carried out to further investigate the propionylation substrates of MOF in the 293T cellular proteome. The MOF-overexpressed and control cells were treated with deuterated sodium propionate (Sodium propionate_D5), followed by cell lysate extraction, affinity enrichment with anti-propionyllysine antibody, and LC-MS/MS analyses. As shown in **Figure 4.9**, the number of d5-propionylated lysine sites (Kpr_D5) in cellular proteins greatly increased, from 36 to 60, upon MOF overexpression. Excitingly, 43 unique propionylation sites were identified in MOF-overexpressed cells which were not present in the control cells. These emerging sites most likely represent *bona fide* cellular substrates of MOF **(Table 4.3)**. These proteins include known MOF substrates such as p53 and MOF [6,130]. H4K16 was found in the propionylation list, which is consistent with the *in vitro* and *in vivo* data (**Table 4.2** and **Figure 4.7c**) and also matches the previous literature showing that MOF can acetylate H4K16 in mammalian cells [133]. Additionally, we identified several new histone propionylation sites at H4K12, K16, and H2AK5, K9 in the MOF-overexpressed cells, which is in agreement with the western blot data that the propionylation levels of histone H2A/H2B and H4 were upregulated upon MOF overexpression **(Figure 4.7b)**. It is worth mentioning that 17 out of the total 60 Kpr_D5 sites in the MOF-overexpressed cells were found in the control cells as well. These shared modification marks could be the common substrates of MOF and other KPTs. Also, 19 Kpr_D5 sites were only identified in the control cells but not in the MOF-overexpressed cells.

These lysine sites might be the substrates of other KPTs such as GCN5 or p300 and their propionylation level are not high enough to be detected in the MOF-overexpressed cells due to the strong competition from the overexpressed MOF. Overall, these proteomic data demonstrate that MOF propionylates a wide range of cellular proteins. The finding that MOF propionylates multiple protein substrates indicates broader involvement of MOF in biological regulation and its versatile functions.

The crystal structure of MOF in complex with propionyl-CoA further supports the KPT activity of the MYST enzymes.

To gain structural insight into the KPT activity of MOF, we solved the crystal structure of MOF catalytic domain in complex with propionyl-CoA at 1.78 Å resolution **(Table 4.4 and Figure 4.10)**. Binding of propionyl-CoA to the MOF catalytic domain did not cause any major structural change in MOF structure, as the MOF-propionyl-CoA binary structure is almost identical to that of the previously solved MOF-acetyl-CoA complex (PDB: 2GIV). The Cα of these two structures can be superimposed with root-mean-square deviation (RMSD) of just 0.17 Å. Similar to the MOF-acetyl-CoA structure, the overall fold of MOF-propionyl-CoA adopts an elongated shape. The structure is comprised of a central core region, flanked on the opposite sides by N- and C-terminal domains **(Figure 4.10a)**. The catalytic site of the MYST family of KATs involves a catalytic cysteine (Cys316 in MOF), and a conserved glutamate (Glu350 in MOF) function as general base. An auto-acetylated lysine at the active site (Lys274 in MOF) is also critical for the KAT activity.[136,137] Our structure shows that the active site of MOF can well accommodate propionyl-CoA and Fo-Fc omit map shows clear density for the terminal methyl group in propionyl-CoA (**Figure 4.10b**). Like acetyl-CoA, propionyl-CoA binds to MOF between the core domain and C-terminal domain in a bent conformation, interacting mostly with

the β10-loop-α4 region of the core domain and loop-α5 region of the C-terminal domain **(Figure 4.10a)**. When the two complex structures are overlaid together, the propionyl-CoA molecule in the MOF-propionyl-CoA structure can superimpose with the corresponding moiety of the acetyl-CoA molecule in the MOF-acetyl-CoA complex structure very well **(Figure 4.10c)**. Most of the interactions between MOF and acetyl-CoA in the MOF-acetyl-CoA structure are still conserved in the MOF-propionyl-CoA structure. However, propionyl–CoA makes two extra hydrophobic interactions through its propionyl group with Val314 and Pro349 in MOF **(Figure 4.10c)**. Sequence alignment reveals that both Pro349 and Val 314 in MOF are conserved amino acid residues through all the five eukaryotic MYST KATs **(Figure 4.11)**. These two interactions could contribute to the proper orientation of the propionyl group. The crystallographic data provide direct structural evidence that the MYST KATs behave as a dual activity enzyme for catalyzing both propionylation and acetylation.

To test the necessity of the two residues Val314 and Pro349 of MOF for its KPT activity, we conducted site-directed mutagenesis to replace MOF-Val314 or Pro349 with Ala and then tested the KAT and KPT activity of wt-MOF, MOF-V314A, and MOF-P349A using the fluorescent CPM assay to obtain the kinetic constants k_{cat} and K_m with respect to acetyl-CoA and propionyl-CoA **(Table 4.5)**. Both MOF-V314A and P349A showed decreased intrinsic enzymatic activity on acetylation, retaining 30% and 20% activity compared with the wt-MOF. The KPT activity of MOF-V314A abolished 55% of wt-MOF while mutation of Pro349 to Ala abolished more than 95% the KPT activity of wt-MOF. These data suggest that the interaction between the propionyl group and MOF involves Pro349 more than Val314; therefore, the conserved Pro residue contributes more to the KPT activity of MYST KATs.

Conclusion

The discovery of chemically diverse acylations in proteins represents an exciting area of research in biology. It embodies the potential and important regulatory roles of cellular acyl-CoA metabolites in the modulation of epigenetics and signal transduction. It would be vitally important to determine the cofactor promiscuity of different KATs in order to elucidate the biochemical etiology of cellular protein acylations. In this study, we found that the MYST family of KATs showed strong *bona fide* KPT activity. Especially, MOF, MOZ, and HBO1 exhibited as strong KPT activity as their classic KAT activity on histone substrates. The analysis of cellular proteins with western blot showed that all members of the MYST KATs promote extensive lysine propionylation, not only on histones but also on non-histone proteins. These findings that the MYST enzymes exhibited a widespread propionylome suggest their much broader physiological roles in the regulation of biological processes. We found that MOF has highly identical acetylome and propionylome profiles, but the modification levels at the individual lysine residues were slightly distinct. That the propionylome of MOF being highly identical to its acetylome strongly advocates for the cofactor promiscuity of the MYST enzymes and predicts that the dynamics of distinct acylation marks in proteins is greatly influenced by the metabolic fluctuation of cellular acyl-CoA variants. We, for the first time, solved the crystal structure of a MYST enzyme bound with propionyl-CoA, which provides direct evidence supporting the KPT activity of the MYST KATs. Moreover, our data suggest that the residue Pro349 in MOF, conserved through the MYST family of KATs, is required for its KPT activity. Future efforts will be needed to clarify how the sub-propionylomes of individual KATs differ from each other and to address how lysine propionylation impacts on the properties of the substrate proteins in various biological pathways.

Methods and Materials

Quantification of cellular acetyl- and propionyl-CoA.

Human embryonic kidney 293T (HEK293T) cells were cultured to 90% confluence with DMEM medium supplemented with 10% fetal bovine serum (FBS) and 1% streptomycin-penicillin. Cells were washed with ice-cold PBS buffer followed by fixing in methanol at −80 °C for 15 minutes. 75 ng deuterated acetyl-CoA and propionyl-CoA were added as internal standards. Cells were collected in 50% of methanol with gentle scrape. Cell suspension was centrifuged at 16000g and supernatant was collected for LC-MS/MS analysis. An Atlantis® T3 (4.6×150 mm, 3 μm) column coupled with a Phenomenex SecurityGuard C-18 guard column (4.0 mm×2.0 mm) was applied for separation. An Agilent 1100 binary pump HPLC system (Santa Clara, CA) coupled to a Waters Micromass Quattro Micro triple quadrupole mass spectrometer with an ESI source (Milford, MA) was applied for LC–MS/MS analysis. The column temperature was controlled at 32 °C. The mobile phase A was 10 mM ammonium acetate, and mobile phase B was acetonitrile. Analytes were separated using a gradient method, with a 0.4 mL/min flow rate, (time/minute, % mobile phase B): (0, 6), (15, 30), (15.01, 100), (22.50, 100), (22.51, 6). The injection volume was 30 μL, and the autosampler injection needle was rinsed with methanol after each injection. Samples were analyzed by the mass spectrometer in positive ion ESI mode. Nitrogen was used as the desolvation gas at a flow rate of 500 L/h. The desolvation temperature was 500 °C and the source temperature was 120 °C. Argon was used as the collision gas, and the collision cell pressure was 3.5×10^{-3} mbar. The capillary voltage was 3.2 kV, the cone voltage was 42 V and the collision energy was 22 eV. Multiple reaction monitoring (MRM) functions were applied for the detection of each acyl-CoA and internal standards. The ion transitions monitored were 810→303 for acetyl-CoA, 813→306 for d3-acetyl-CoA, 824→

317 for propionyl CoA and 829→322 for d5-propionyl-CoA.An Atlantis® T3 (4.6×150 mm, 3 μm) column coupled with a Phenomenex SecurityGuard C-18 guard column (4.0 mm×2.0 mm) was applied for separation. An Agilent 1100 binary pump HPLC system (Santa Clara, CA) coupled to a Waters Micromass Quattro Micro triple quadrupole mass spectrometer with an ESI source (Milford, MA) was applied for LC–MS/MS analysis. The column temperature was controlled at 32 ºC. The mobile phase A was 10 mM ammonium acetate, and mobile phase B was acetonitrile. Analytes were separated using a gradient method, with a 0.4 mL/min flow rate, (time/minute, % mobile phase B): (0, 6), (15, 30), (15.01, 100), (22.50, 100), (22.51, 6). The injection volume was 30 μL, and the autosampler injection needle was rinsed with methanol after each injection. Samples were analyzed by the mass spectrometer in positive ion ESI mode. Nitrogen was used as the desolvation gas at a flow rate of 500 L/h. The desolvation temperature was 500 ºC and the source temperature was 120 ºC. Argon was used as the collision gas, and the collision cell pressure was 3.5×10^{-3} mbar. The capillary voltage was 3.2 kV, the cone voltage was 42 V and the collision energy was 22 eV. Multiple reaction monitoring (MRM) functions were applied for the detection of each acyl-CoA and internal standards. The ion transitions monitored were 810→303 for acetyl-CoA, 813→306 for d3-acetyl-CoA, 824→317 for propionyl CoA and 829→322 for d5-propionyl-CoA.

Kinetic characterization of acetyl-CoA and propionyl-CoA in KAT mediated histone modification.

Synthetic histone peptides H3-20 or H4-20 (20 amino acids from the N-terminal of histone H3 and H4, the sequence of H3-20 is Ac-ARTKQTARKSTGGKAPRKQL, the sequence of H4-20 is Ac-SGRGKGGKGLGKGGAKRHRK) were used as acyl acceptor substrates. Acyl-CoA at varied concentrations was incubated with individual KAT enzymes and peptide at fixed

concentration. KAT enzymes deposit acyl groups on the ε-amine of peptide lysines and release CoASH simultaneously. The fluorogenic probe 7-diethylamino-3-(4'-maleimidylphenyl)- 4-methylcoumarin (CPM) was added to react with by-product CoASH to form the fluorescent CoAS-CPM complex [43]. The fluorescence intensities were measured with a microplate reader (FlexStation® 3). The catalytic rate was determined from the fluorescence intensity. Kinetic constants including binding affinity (K_m) and catalytic efficiency (k_{cat}) were determined by fitting the acyl-CoA concentration-catalytic rate to the Michaelis-Menten equation.

Profiling of MYST KATs *in vitro* propionylome.

The whole lysate of HEK293T cells was extracted using a gentle cell lysis buffer M-PER Mammalian Protein Extraction Reagent (Thermo Scientific) together with sonication at 30% amplitude using a sonicator (Fisher Scientific, Model 120 Sonic Dismembrator). Protein propionylation level was tested with western blot analysis using pan anti-propionyllysine antibody (PTM BIOLABS, product number PTM-210).

Identification of MOF acylation sites on H4-20 peptide using LC-MS/MS analysis.

100 μM H4-20 peptide, 200 μM acyl-CoA were incubated with 1 μM MOF enzyme for 1 hour at 30 ^{o}C followed by desalting with C18 Zip-tip. The desalted samples were analyzed by ACQUITY UPLC system (Waters, Milford, MA) coupled to a Waters SYNAPT G2 mass spectrometer (Milford, MA). Peptides were separated on the HALO C18 peptide column (2.7 μm, 4.6 x 100mm, Advanced Materials Technology, Wilmington, DE). Mobile phase A is water containing 0.01% formic acid and B is acetonitrile (ACN). The injection volume was 10 μL. Peptides were separated with a 0.3 mL/min, isocratic flow of 5 % B. MS tune parameters were as follows: capillary 2.00 kV, sample cone 35 V, extraction cone 4.0 V, source temperature 120 °C, desolvation temperature 500 °C, desolvation gas 500 L/h. Data were first collected in the full

scan mode in the mass range of 300-1900. To identify acetylation/propionylation sites, data were acquired by the data-dependent acquisition (DDA) mode. For DDA parameters, 1 s MS survey scan in the m/z range of 300–1900 were followed by MS/MS scans of up to 3 ions, when intensity rose above 1500 counts per second. MS/MS was acquired in the m/z range of 100-1900, with a 2 s scan rate. MS/MS scan was switched to MS survey scan after 3 scans. Trap collision energy was set using charge state recognition, applying the default files for 1-4 charge states. Files containing MS/MS spectra were processed with Proteinlynx Global Server 2.4 software (Waters, Milford, MA) to identify PTM sites.

Comparative analysis of MOF substrate specificity on both acetylation and propionylation

20 μg cell lysate was incubated with 2 μM MOF and 50 μM carbon-14 labeled acetyl-CoA or propionyl-CoA for 3 hours at 30 ^{o}C. The reaction was quenched with the addition of 5x SDS-PAGE loading dye and the proteins were resolved on a 4-20% gradient SDS-PAGE gel (BioRad). The gel was dried in vacuum and exposed to a phosphorscreen for 72 hours. The autoradiograph was scanned with the GE storm 865 imager (GE Healthcare) and the in gel proteins were imaged with coomassie brilliant blue stain for protein loading control

Study of *in vivo* MOF acyltransferase activity

Full-length MOF encoding sequence was inserted into XbaI and EcoRI sites of lentivirus vector FuCRW to generate the MOF-overexpression plasmid [138]. Plasmid transfections were performed using Lipofectamine 3000 (ThermoFisher). Twenty-four hours before transfection, the HEK 293T cells were seeded at a density of $1x10^{6}$ cells/well for six-well plate. The cells were transfected with 8 μg MOF-vector or vector plasmid as described in the manufacturer's protocol and the transfected cells were maintained for 72 h before harvest. 20 mM sodium propionate was added into the medium 24 hours before the cellular proteins were extracted.

Whole cell lysate and histones were extracted from both MOF-overexpressed and normal cells for the study of cellular protein acetylation and propionylation changing in response to the overexpression of MOF.

Identification of propionylation sites using LC-MS/MS proteomic analysis.

The whole cell lysates and core histone proteins were extracted from the HEK293T cells treated with 20 mM deuterated sodium propionate with or without MOF overexpression. The extracted whole cell lysates or histones were precipitated using trichloroacetic acid. The resulting protein precipitate was washed twice with ice-cold acetone and digested with trypsin using a procedure described previously [139]. To enrich the propionylated peptides, the tryptic digest in NETN buffer (100 mM NaCl, 1mM EDTA, 50 mM Tris-HCl, 0.5% NP40, pH 8.0) was incubated with pan anti-Kpr antibody (PTM-201, PTM Bio, Chicago, IL) that was immobilized to protein A agarose beads at 4 ^{o}C for 6 hours with gentle rotation. After incubation, the beads were washed 3 times with NETN buffer and twice with ddH_2O. The enriched peptides were eluted with 0.1% trifluoroacetic acid (TFA). The eluates were vacuum concentrated and the peptides were suspended in 0.1% TFA followed by desalting using C18-Ziptips. The peptide samples were directly loaded onto a home-made capillary column (10 cm length, 75 μm internal diameter) packed with Reprosil C18 resin (3 μm particle size, 100 Å pore size, Reprosil) on an EASY-nLC™ 1000 System (Thermo Fisher Scientific, Waltham, MA). The binding peptides were eluted with a gradient of 5% to 90% HPLC buffer B (0.1% formic acid in 90% acetonitrile, v/v) in buffer A (0.1% formic acid in water, v/v) over 60 min at a flow rate of 200 nL/min. The eluted peptides were ionized and introduced into a Q-Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA) using a nanospray source. Full MS scans were acquired over the range of m/z 300–1400 with a resolution of 70000, which was followed by data-dependent

MS/MS fragmentation of the 15 most intense peaks with a resolution of 17500 at 27% normalized collision energy. For all the experiments, the dynamic exclusion time was set to 25 s. Peptides identification was performed with MaxQuant (v 1.3.0.5) against the UniProt Human protein database. Oxidation on methionine, protein N-terminal acetylation, lysine acetylation, lysine propionylation, lysine D5-propionylation, lysine mono-/di-/tri-methylation, and arginine mono-/di-methylation were set as variable modifications. Carbamiodomethylation on cysteine was set as fixed modification. False discovery rate (FDR) thresholds for protein, peptide and modification site were specified at 0.01. Peptides with MaxQuant score below 40 or site localization probability below 0.75 were removed. In addition, all the identified peptides were manually verified.

Cloning, expression and purification of recombinant MOF.

The DNA fragment encoding the histone acetyltransferase domain of human MOF (residues 174-449) was sub-cloned into pET28a-LIC vector (GenBank ID: EF442785). Recombinant MOF was over-expressed in *E. coli* BL21 (DE3) codon plus RIL strain (Stratagene) as an N-terminal Hexa-His fusion protein at 15 °C in Terrific Broth (Sigma). The harvested cells were resuspended in 50 mM HEPES buffer, pH 7.4, supplemented with 500 mM NaCl, 5 mM imidazol, 2 mM β-mercaptoethanol, 5% glycerol, 0.1% CHAPS. The cells were lysed by passing through Microfluidizer (Microfluidics Corp.) at 20,000 psi. The lysate was loaded onto 10 ml Chelating Sepharose column (GE Healthcare) charged with Ni^{2+}. After washing the column with 20 mM HEPES buffer, pH 7.4, containing 500 mM NaCl, 50 mM imidazole, 5% glycerol, the protein was eluted with 20 mM HEPES, pH 7.4, 500 mM NaCl, 250 mM imidazole, 5% glycerol. The eluted protein was loaded onto a Superdex 200 column (26x60) (GE Healthcare), equilibrated with 20 mM HEPES buffer, pH 7.4, containing 500 mM NaCl. The fractions

containing recombinant MOF were combined and the protein was further purified to homogeneity by ion-exchange chromatography.

Crystallization of MOF with propionyl-CoA and X-ray crystal structure determination

Purified recombinant MOF protein (5 mg/mL) was mixed with propionyl CoA (Sigma) at 1:5 molar ratio of protein:compound and crystallized using hanging drop vapor diffusion method at 20 °C by mixing 1µl of the protein solution with 1 µl of the reservoir solution. MOF-propionyl CoA was crystallized in buffer containing 20% PEG 3350, 0.2 M sodium malonate, pH 5.0. Crystals were soaked in the corresponding mother liquor supplemented with 20% glycerol as cryoprotectant before freezing in liquid nitrogen.

X-ray diffraction data for MOF + propionyl CoA was collected at 100K on a Rigaku FR-E superbright X-ray generator. Data set was processed using the HKL-3000 suite [140].

The structures of MOF−propionyl CoA complex is isomorphous with PDB entry 2GIV. REFMAC [141] was used for structure refinement. Graphics program COOT [142] was used for model building and visualization. MOLPROBITY [143] was used for structure validation.

Figures, Tables and Captions

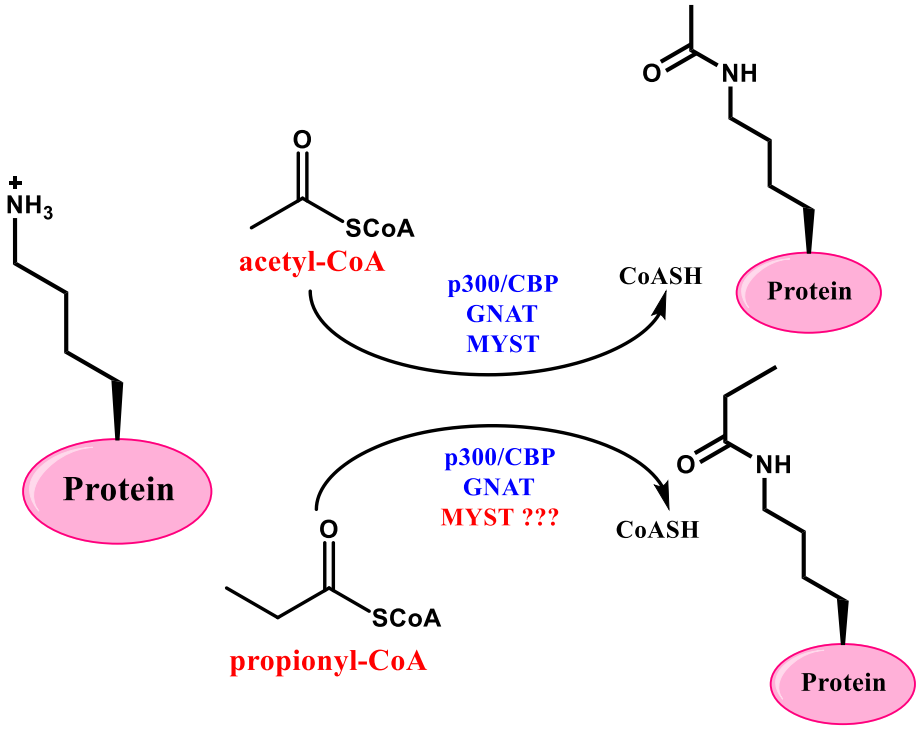
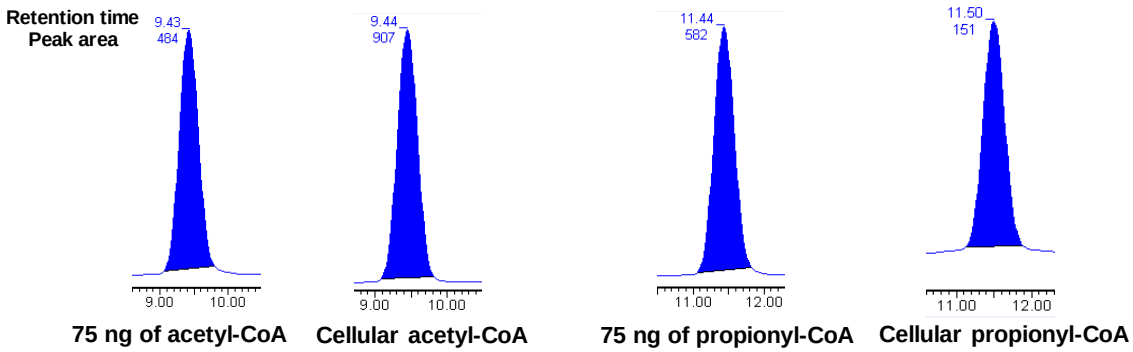


Figure 4.1. Dual enzymatic activity of eukaryotic KAT enzymes.

The three major eukaryotic KAT families catalyze acetylation of lysine residues using acetyl-CoA as the acetyl donor. p300/CBP and GCN5/PCAF KATs have been reported to possess lysine propionyltransferase activity. In this study, we found that all the MYST KATs possess strong KPT activity, providing a holistic view of KATs as lysine propionyltransferase.

a)



b)

	Acyl-CoA Concentration (ng / mg of protein)		Ratio (Pro/Ac)	Average % of Pro-CoA / Ac-CoA
	Ac-CoA	Pro-CoA		
Assay 1	817	113	13.8%	12.2%
Assay 2	731	76	10.4%	
Assay 3	715	88	12.3%	
Average	754	92	12.2%	

Figure 4.2. Quantification of acetyl- and propionyl-CoA abundance in 293T cells using LC-MS/MS.

a. LC-MS chromatograms of acetyl- and propionyl-CoA. The cellular acyl-CoA concentration was calculated based on the deuterated acyl-CoA molecules which serve as internal standards. Acetyl-CoA and propionyl-CoA were sufficiently separated with the retention time at 9.43 and 11.43 minute respectively; **b.** Summary of triplicate experiments. The cellular abundance of propionyl-CoA is about 12% of the abundance of acetyl-CoA.

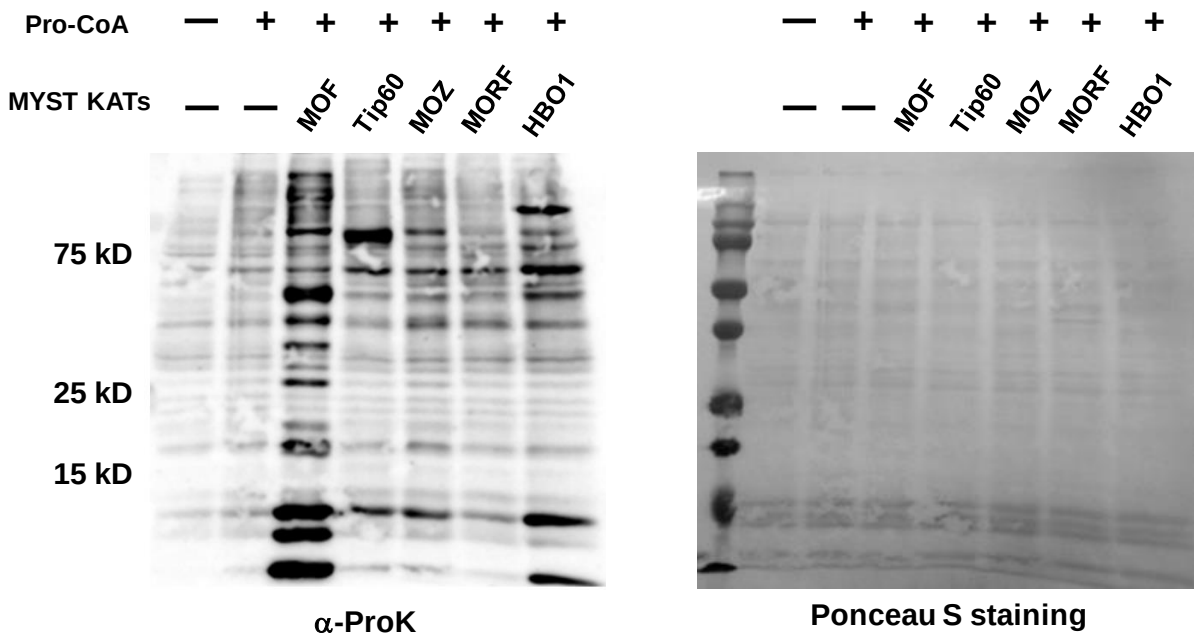


Figure 4.3. Test of lysine propionyltransferase activity of the MYST KATs on cellular proteome with western blot analysis.

HEK 293T cell lysate was incubated with individual KATs in the presence or absence of propionyl-CoA. Treatment of cell lysate with propionyl-CoA and MYST KATs, especially MOF and HBO1, induced strong lysine propionylation on multiple histone and non-histone proteins.

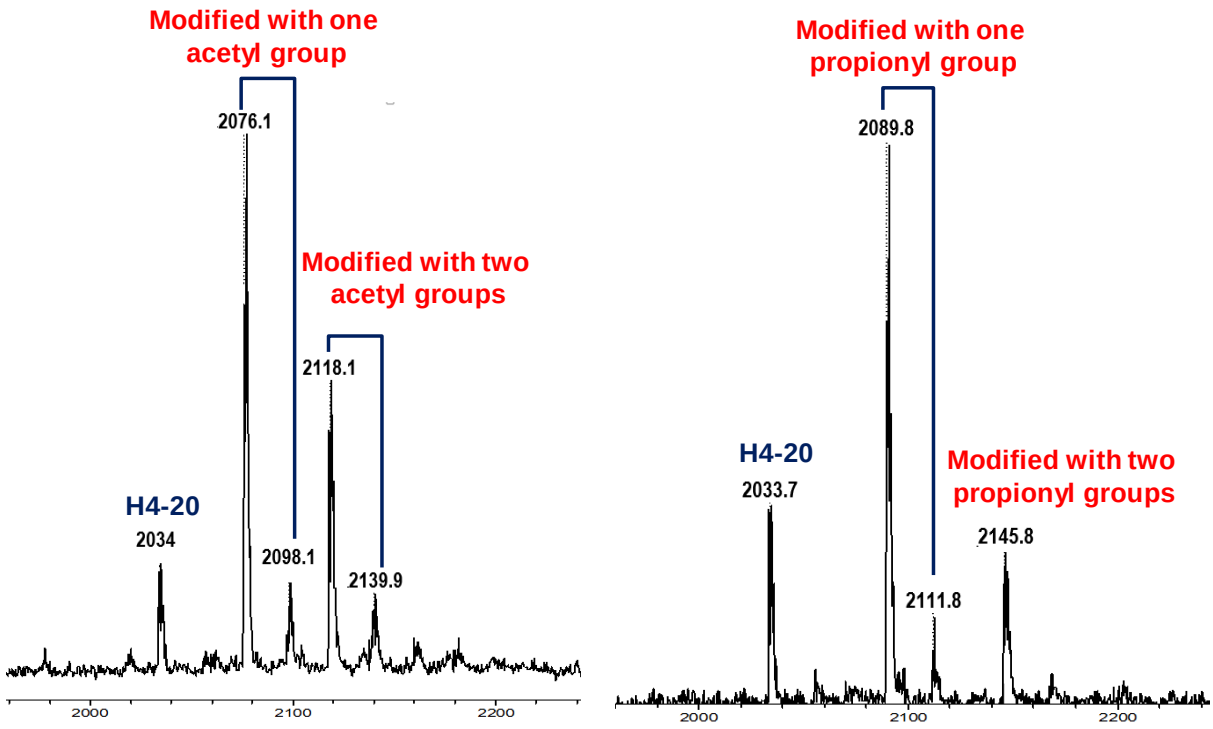


Figure 4.4. Identification of MOF products with H4-20 as a substrate using MALDI-MS analysis.

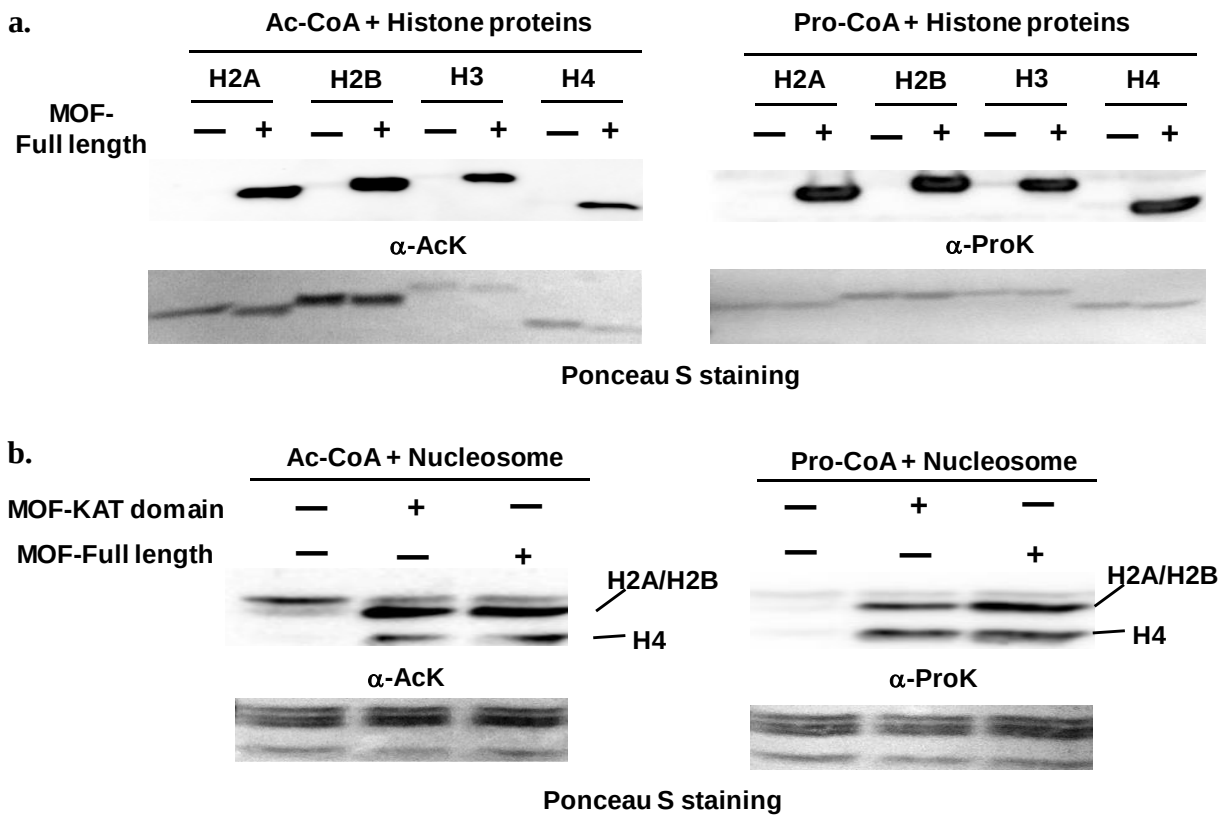


Figure 4.5. Study of MOF acetyl- and propionyl-transferase activity on recombinant histones and nucleosomes.

Both recombinant free histones and reconstituted nucleosomes were used as substrates for MOF acetyl- and propionyl-transferase activity study. Histone proteins with acyl-CoA were incubated with or without MOF; the reaction mixtures were then subjected to western blot analysis with pan anti-acetyllysine and pan anti-propionyllysine antibodies. a. MOF acylates four free core histones; b. MOF acylates histone H2A/H2B and H4 on nucleosome.

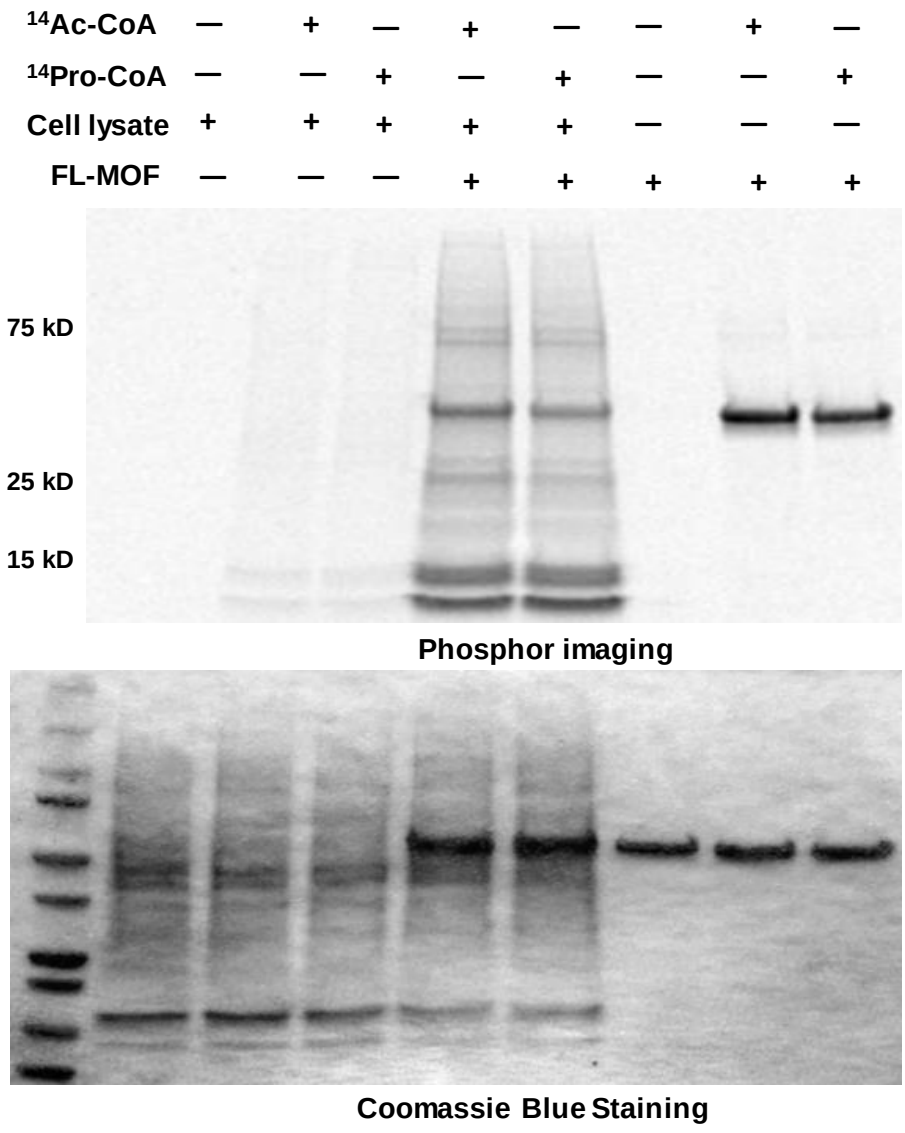


Figure 4.6. Imaging of MOF *in vitro* acetylome and propionylome using radioactive gel assay.

Carbon-14 labeled acetyl-CoA and propionyl-CoA were used for cell lysate acylation by MOF. Proteins were resolved on SDS-PAGE and the MOF acylome was imaged using a phosphor imager.

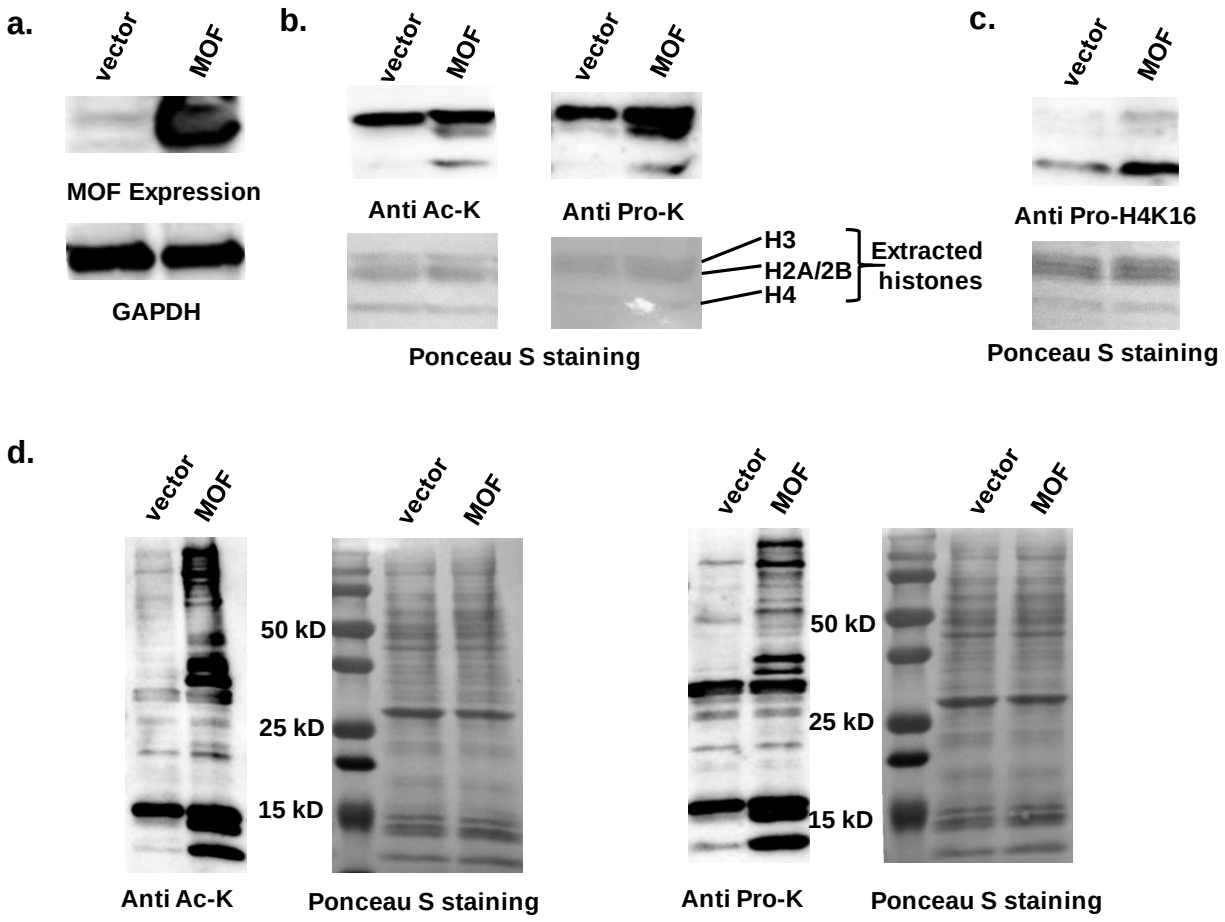


Figure 4.7. Detection of the cellular acetyltransferase and propionyltransferase activities of MOF.

a. MOF was overexpressed in transfected 293T cells compared to the control cells where the vector plasmid was used for transfection; **b.** Histone lysine acylation level was tested with pan anti-acyllysine antibodies. MOF acylated histone H2A/H2B and H4; **c.** MOF overexpression induced increase of H4K16 propionylation; **d.** Acylation of the cellular proteome was tested with western blot. Increase of lysine acetylation and propionylation of multiple proteins were induced by MOF overexpression.

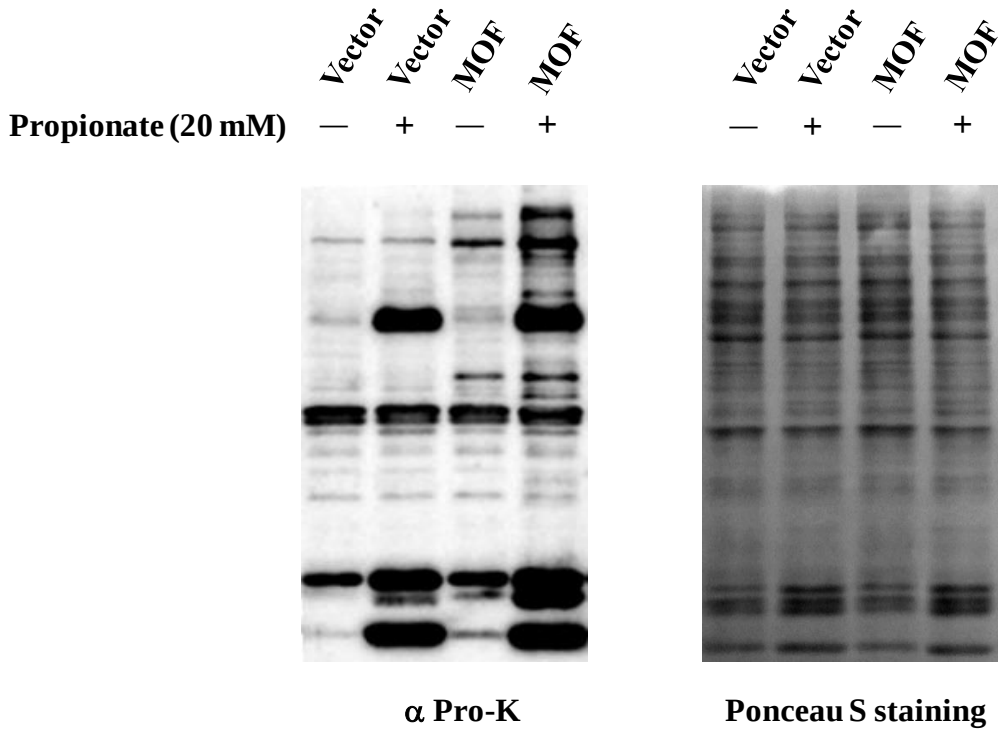


Figure 4.8. Dynamic change of lysine propionylation in response to propionate treatment. MOF-overexpressed and control HEK 293T cells were treated with or without 20 mM propionate. Whole cell lysates were extracted and subjected to western blot analysis. Propionate treatment and MOF overexpression induced increase of lysine propionylation, respectively. Treatment of MOF-overexpressed cells with propionate revealed more propionylated proteins, indicating MOF may have some unique substrates for lysine propionylation compared with other endogenous KATs.

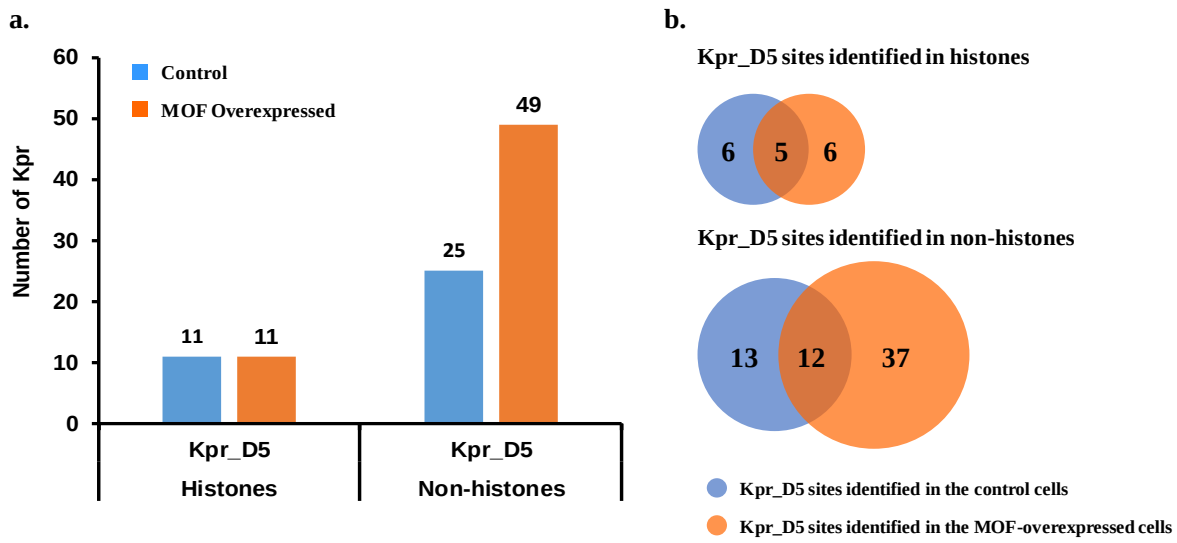


Figure 4.9. A comparison of lysine propionylation sites between the control and MOF overexpressed cells.

a. The numbers of Kpr_D5 modified sites in histones and non-histone proteins were compared between the control and MOF overexpressed cells; **b.** The numbers of overlapped and non-overlapped Kpr_D5 sites between both the control and MOF-overexpressed cells.

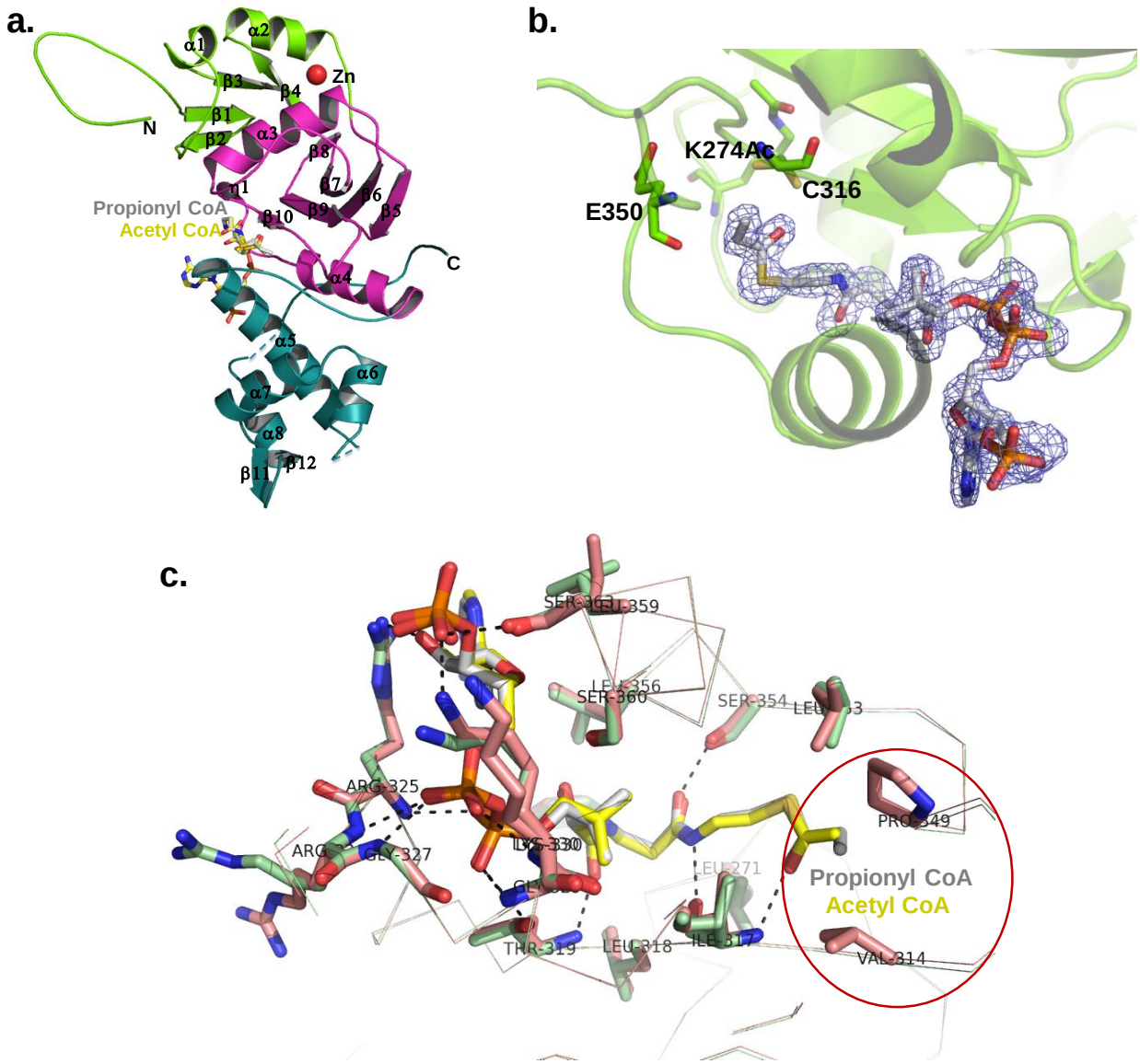


Figure 4.10. X-ray crystal structure of MOF-propionyl CoA complex.

a. Overall structure of MOF-propionyl-CoA. MOF structure is shown in cartoon model, with the N-terminal, central, and C-terminal domains colored in green, magenta and blue, respectively. The bound compounds are shown in sticks, with propionyl CoA colored in grey and Ac-CoA (from 2GIV) in yellow; **b.** Active site of MOF. Catalytic residues C316, E350 and auto-acetylated K274 are shown in stick model. Fo-Fc omit map of propionyl-CoA contoured at 2.5 sigma shows the density for the extra methyl group in propionyl CoA; **c.** Structural comparison

of the binding sites for propionyl CoA (PDB:5WCI) and Ac-CoA (PDB: 2GIV) in MOF. The two structures are superimposed, with MOF-propionyl CoA in pink and MOF-acetyl CoA in green. The MOF residues interacting with the compounds are shown in stick model. Dashed lines represent hydrogen bonds. The two extra interactions in MOF-propionyl CoA are circled.

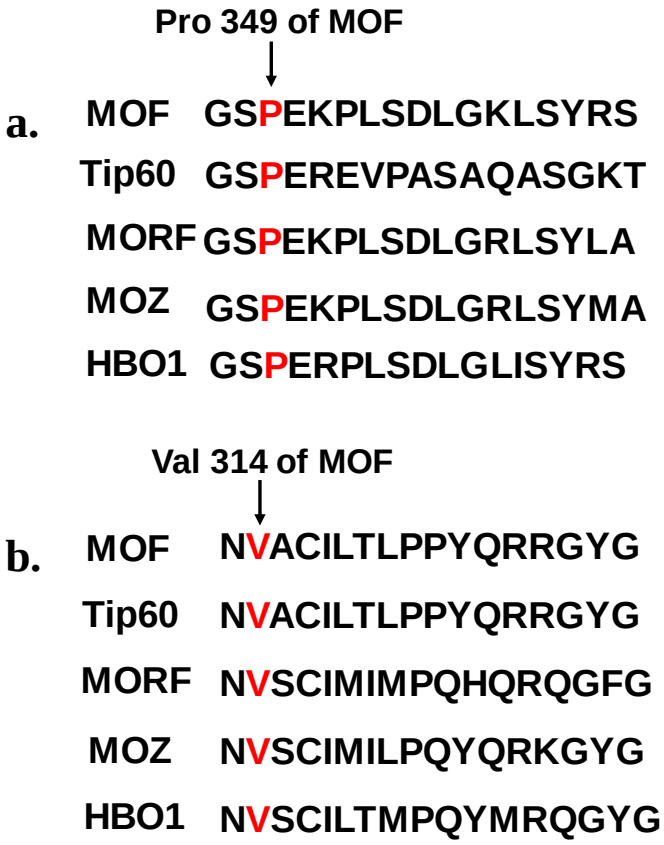


Figure 4.11. Sequence alignment of MYST KATs.

a. MOF Pro349 is a conservative amino acid residue through all the five eukaryotic MYST KATs; **b.** MOF Val314 is a conservative amino acid residue through all the five eukaryotic MYST KATs

Table 4.1. Kinetic parameters of acyl-CoA in KAT mediated histone modification.

KAT Family	Enzyme	Cofactor	K_m (μM)	k_{cat} (min^{-1})	k_{cat}/K_m ($\mu M^{-1}.min^{-1}$)	$[k_{cat}/K_m(Pro)]/[k_{cat}/K_m(Ac)]$
MYST	MOF	Ac-CoA	1.1 ± 0.4	2.8 ± 0.2	2.4±0.9	0.8
		Pro-CoA	0.9 ± 0.2	1.8 ± 0.1	2.0±0.5	
	HBO1	Ac-CoA	6.8 ± 2.6	1.0 ± 0.1	0.14±0.06	1.0
		Pro-CoA	4.8 ± 1.5	0.7 ± 0.1	0.14±0.05	
	Tip60	Ac-CoA	1.5 ± 0.3	3.5 ± 0.1	2.3±0.4	0.13
		Pro-CoA	1.9 ± 0.2	0.6 ± 0.02	0.3±0.03	
	MORF	Ac-CoA	0.8 ± 0.2	1.3 ± 0.1	1.7±0.4	0.18
		Pro-CoA	2.3 ± 1.0	0.7 ± 0.1	0.3±0.1	
	MOZ	Ac-CoA	0.6 ± 0.1	0.6 ± 0.02	0.9±0.2	1.0
		Pro-CoA	0.6 ± 0.2	0.5 ± 0.02	0.9±0.3	
GNAT	GCN5	Ac-CoA	4.4 ± 0.9	58.4 ± 3.2	13.2±2.7	1.1
		Pro-CoA	0.5 ± 0.03	6.5 ± 0.1	14.1±0.9	
	PCAF	Ac-CoA	5.2 ± 1.1	35.0 ± 2.7	6.7±1.5	0.4
		Pro-CoA	1.1 ± 0.1	3.0 ± 0.1	2.8±0.3	
	HAT1	Ac-CoA	5.4 ± 0.8	7.4 ± 0.3	1.4±0.2	0.93
		Pro-CoA	8.5 ± 1.9	10.9 ± 0.8	1.3±0.3	
p300/CBP	p300	Ac-CoA	17.3 ± 4.8	34.0 ± 3.0	2.0±0.6	0.3
		Pro-CoA	16.3 ± 2.8	10.2 ± 0.8	0.6±0.1	

Table 4.2. Identification of MOF acylation sites on H4-20 peptide

	m/z	Charge	Peak MW	Peptide MW	Modification Sites	Retention time
H4-20+acetyl-CoA+MOF	519.8197	4	2075.2476	2075.1938	K16	3.94
	692.7526	3	2075.2344	2075.1938	K8	4.11
	424.4618	5	2117.27	2117.2043	K12, K16	5.92
	530.3186	4	2117.2432	2117.2043	K8, K16	5.11
	706.7573	3	2117.2485	2117.2043	K8, K12	5.48
H4-20+propionyl-CoA+MOF	537.328	4	2145.2808	2145.2031	K8, K16	11.24
	523.3264	4	2089.2744	2089.1931	K16	4.23

Table 4.3. Identification of MOF propionylation sites on cellular proteome from 293T cells

(The lysine residues highlighted in yellow color are identified in both control and MOF-overexpressed cells)

	Gene names	Protein names	Localization prob	Score	Modified sequence	Position in peptide	Position in protein
Kpr_D5 lysine sites in histones from the control cells (11)	HIST1H1C	Histone H1.2	1.0000	80.245	_KPAAATVTK(pr)K_	9	168
	HIST1H1D	Histone H1.3	1.0000	141.08	_KPATAAGTK(pr)K_	9	169
	HIST1H1E	Histone H1.4	1.0000	76.228	_KPAAAAGAK(pr)K_	9	168
	H2AFX	Histone H2A.x	1.0000	61.679	_GK(pr)TGGK(pr)AR_	2	6
	H2AFX	Histone H2A.x	1.0000	61.679	_GK(pr)TGGK(pr)AR_	6	10
	H2AFZ	Histone H2A.Z	1.0000	46.158	_AGK(pr)DSGK(pr)AK_	3	8
	H2AFZ	Histone H2A.Z	1.0000	46.158	_AGK(pr)DSGK(pr)AK_	7	12
	HIST2H2BF	Histone H2B	1.0000	40.103	_SAPAPK(pr)K_	6	12
	HIST2H2BF	Histone H2B type 2-F	1.0000	41.502	_AVTK(pr)VQK_	4	21
	HIST1H3A	Histone H3.1	1.0000	71.342	_STGGK(pr)APR_	5	15
	HIST1H3A	Histone H3.1	1.0000	71.692	_QLATK(pr)AAR_	5	24
Kpr_D5 lysine sites in histones from the MOF-overexpressed cells (11)	HIST1H1C	Histone H1.2	1.0000	57.785	_KPAAATVTK(pr)K_	9	168
	HIST1H1C	Histone H1.2	1.0000	63.673	_KPVGAAK(pr)K_	7	136
	HIST1H1E	Histone H1.4	1.0000	82.029	_KPAAAAGAK(pr)K_	9	168
	HIST1H1E	Histone H1.4	1.0000	100.07	_SAGAAK(pr)R_	6	32
	HIST1H2AA	Histone H2A type 1-A	1.0000	50.878	_GK(pr)QGGK(pr)AR_	2	6
	HIST1H2AA	Histone H2A type 1-A	1.0000	50.878	_GK(pr)QGGK(pr)AR_	6	10
	H2AFX	Histone H2A.x	1.0000	43.083	_GK(pr)TGGK(pr)AR_	2	6
	H2AFX	Histone H2A.x	1.0000	43.083	_GK(pr)TGGK(pr)AR_	6	10
	HIST1H3A	Histone H3.1	1.0000	109.11	_KQLATK(pr)AAR_	6	24
	HIST1H4A	Histone H4	1.0000	58.596	_GLGK(pr)GGAK(pr)R_	4	13
	HIST1H4A	Histone H4	1.0000	58.596	_GLGK(pr)GGAK(pr)R_	8	17

Kpr_D5 lysine sites in non-histones from the control cells (25)	ADNP	Activity-dependent neuroprotector homeobox protein	1.0000	44.816	_IFHAPNASAPSSSLSTFK(pr)DK_	18	144
	BCLAF1	Bcl-2-associated transcription factor 1	1.0000	51.927	_NTEEEGLK(pr)YK_	8	437
	CBX3	Chromobox protein homolog 3	1.0000	63.48	_TTLQK(pr)M(ox)GK_	5	10
	DDX17	Probable ATP-dependent RNA helicase DDX17	1.0000	56.258	_K(pr)FGNPGER_	1	109
	EEF1G	Elongation factor 1-gamma	1.0000	58.981	_AFNQGK(pr)IFK_	6	484
	FUS	RNA-binding protein FUS	1.0000	54.608	_GGFNK(pr)FGGPR_	5	264
	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	1.0000	58.676	_TVDGPSGK(pr)LWR_	8	194
	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	1.0000	74.769	_SSGPYGGGGQYFAK(pr)PR_	14	350
	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	1.0000	67.275	_RPAEDMEEEQAFK(pr)R_	13	34
	HNRNPM	Heterogeneous nuclear ribonucleoprotein M	1.0000	46.408	_GCGVVK(pr)FESPEVAER_	6	698
	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	0.9966	40.589	_VSELKEELK(pr)K_	9	21
	NPM1	Nucleophosmin	1.0000	55.064	_SKGQESFK(pr)K_	8	229
	NPM1	Nucleophosmin	1.0000	107.66	_LLSISGK(pr)R_	7	141
	NPM1	Nucleophosmin	1.0000	109.29	_VEAK(pr)FINYVK_	4	267
	NUDT21	Cleavage and polyadenylation specificity factor subunit 5	1.0000	86.291	_GVTQFGNK(pr)YIQQTK_	8	23
	PABPN1	Polyadenylate-binding protein 2	1.0000	40.025	_VTILCDK(pr)FSGHPK_	7	234
	PSMC1	26S protease regulatory subunit 4	1.0000	52.725	_K(pr)YEPPVPTR_	1	24
	RBMX	RNA-binding motif protein, X chromosome	1.0000	103.71	_GPLPVK(pr)R_	6	150
	RPS27A	Ubiquitin-40S ribosomal protein S27a	1.0000	53.166	_LAVLK(pr)YYK_	5	104
	SRSF10	Serine/arginine-rich splicing factor 10	1.0000	86.67	_THYK(pr)SGSR_	4	218
	SSB	Lupus La protein	1.0000	82.362	_IIEDQQESLNK(pr)WK_	11	328
	STMN1	Stathmin	1.0000	40.589	_EAQM(ox)AAK(pr)LER_	7	119
	THRAP3	Thyroid hormone receptor-associated protein 3	1.0000	76.827	_GFVPEK(pr)NFR_	6	519
	TOP1	DNA topoisomerase 1	0.9890	44.447	_EDIKPLK(pr)R_	7	137
	U2SURP	U2 snRNP-associated SURP motif-containing protein	1.0000	54.023	_NEPIFK(pr)VAPSK_	6	755
	RPL23	60S ribosomal protein L23	1.0000	69.598	_GGSSGAK(pr)FR_	7	13
	RPL28	60S ribosomal protein L28	1.0000	56.547	_TRPTK(pr)SS_	5	135
	RPL34	60S ribosomal protein L34	1.0000	50.878	_IVYLYTK(pr)K_	7	36
	AHSA1	Activator of 90 kDa heat shock protein ATPase homolog 1	1.0000	45.161	_(ac)AK(pr)WGEGDPR_	2	3
	ENO1	Alpha-enolase	1.0000	47.823	_LAK(pr)YNQLLR_	3	406
	DDX54	ATP-dependent RNA helicase DDX54	1.0000	41.502	_DLYQK(pr)WK_	5	770
	BCLAF1	Bcl-2-associated transcription factor 1	1.0000	41.283	_NTEEEGLK(pr)YK_	8	437
	TP53	Cellular tumor antigen p53	1.0000	62.823	_AHSSHLK(pr)SK_	7	370
	NUDT21	Cleavage and polyadenylation specificity factor subunit 5	1.0000	107.37	_GVTQFGNK(pr)YIQQTK_	8	23
	TOP1	DNA topoisomerase 1	1.0000	58.37	_DEDDADYK(pr)PK_	8	148

Kpr_D5 lysine sites in non-histones from the MOF-overexpressed cells (49)	EEF1G	Elongation factor 1-gamma	1.0000	59.8	_AFNQGK(pr)IFK_	6	484
	EIF3G	Eukaryotic translation initiation factor 3 subunit G	1.0000	40.589	_TGK(pr)YVPPSLR_	3	212
	GAPDH	Glyceraldehyde-3-phosphate dehydrogenase	1.0000	40.589	_TVDGPSGK(pr)LWR_	8	194
	HNRNPA1	Heterogeneous nuclear ribonucleoprotein A1	1.0000	51.643	_SSGPYGGGGQYFAK(pr)PR_	14	350
	HNRNPK	Heterogeneous nuclear ribonucleoprotein K	1.0000	51.546	_RPAEDMEEEQAFK(pr)R_	13	34
	HNRNPU	Heterogeneous nuclear ribonucleoprotein U	1.0000	72.643	_QGQQQAGGK(pr)K_	9	213
	HNRNPC	Heterogeneous nuclear ribonucleoproteins C1/C2	1.0000	55.717	_AVVPSK(pr)R_	6	157
	KAT8	Histone acetyltransferase KAT8	1.0000	51.726	_YWK(pr)GQHVICVTPK_	3	410
	KAT8	Histone acetyltransferase KAT8	1.0000	76.151	_NSEK(pr)YLSELAEQPER_	4	125
	KAT8	Histone acetyltransferase KAT8	1.0000	90.15	_EHEAITK(pr)VK_	7	175
	TMPO	Lamina-associated polypeptide 2, isoforms beta/gamma	1.0000	53.968	_TPVTLK(pr)QR_	6	213
	SSB	Lupus La protein	1.0000	71.029	_VQFQGK(pr)K_	6	360
	MECP2	Methyl-CpG-binding protein 2	1.0000	41.283	_SVQETVLPIK(pr)K_	10	316
	NCL	Nucleolin	1.0000	42.789	_GGK(pr)NSTWSGESK_	3	477
	NCL	Nucleolin	1.0000	45.335	_AAVTPGK(pr)K_	7	87
	NCL	Nucleolin	1.0000	49.813	_VVVSPTK(pr)K_	7	70
	NCL	Nucleolin	1.0000	71.223	_VAVATPAK(pr)K_	8	79
	NPM1	Nucleophosmin	1.0000	64.265	_LLSISGK(pr)R_	7	141
	NPM1	Nucleophosmin	1.0000	66.435	_SKGQESFK(pr)K_	8	229
	PHF6	PHD finger protein 6	1.0000	53.967	_TNFK(pr)GLSEDTR_	4	174
	PDCL3	Phosducin-like protein 3	1.0000	58.676	_LAEWK(pr)ATK_	5	83
	PRAMEF20	PRAME family member 20/21	1.0000	41.448	_YISPEQK(pr)K_	7	252
	FAM50A	Protein FAM50A	1.0000	69.598	_ELQMK(pr)LEK_	5	100
	C4orf27	Putative UPF0609 protein C4orf27-like	1.0000	57.804	_SK(pr)FCEADVSSDLRK_	2	28
	RSL1D1	Ribosomal L1 domain-containing protein 1	0.8714	41.996	_KPEAK(pr)FFTTPSK_	5	461
	RRP1B	Ribosomal RNA processing protein 1 homolog B	1.0000	55.452	_GSPTGGAQLLK(pr)R_	11	522
	RBMX	RNA-binding motif protein, X chromosome	1.0000	103.71	_GPLPVK(pr)R_	6	150
	RBM12B	RNA-binding protein 12B	1.0000	49.5	_PEGGK(pr)FDFGK_	5	895
	FUS	RNA-binding protein FUS	1.0000	55.064	_GGFNK(pr)FGGPR_	5	264
	PPP1CB	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	1.0000	80.522	_AK(pr)YQYGGLNSGR_	2	303
	CTTN	Src substrate cortactin	1.0000	44.807	_DYSSGFGGK(pr)YGVQADR_	9	161
	CTTN	Src substrate cortactin	1.0000	48.112	_TGFGGK(pr)FGVQSER_	6	272
	CTTN	Src substrate cortactin	1.0000	48.423	_GFGGK(pr)YGVQK_	5	309
	CTTN	Src substrate cortactin	1.0000	78.548	_GFGGK(pr)YGIDK_	5	198
	STMN1	Stathmin	1.0000	40.589	_EAQMAAK(pr)LER_	7	119
	SMARCAD1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin	1.0000	43.77	_ASISYFK(pr)NQR_	7	84

		subfamily A containing DEAD/H box 1					
	THRAP3	Thyroid hormone receptor-associated protein 3	0.9745	42.789	_AEGK(pr)YKDDPVDLR_	4	709
	THRAP3	Thyroid hormone receptor-associated protein 3	1.0000	80.239	_GFVPEK(pr)NFR_	6	519
	ZNF207	Zinc finger protein 207	1.0000	43.246	_AQLPK(pr)YQR_	5	397

Table 4.4. Crystallography data and refinement statistics.

	MOF-propionyl CoA
PDB Code	5WCI
Data collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	46.3, 58.7, 120.3
α, β, γ (°)	90.0,90.0,90.0
Resolution (Å) (highest resolution shell)	50.00-1.78(1.81-1.78)
Measured reflections	419509
Unique reflections	31768
R_{merge}	0.093(0.916)*
$I/\sigma I$	27.5(1.7)*
Completeness(%)	98.7(97.6)*
Redundancy	13.2(8.1)*
Refinement	
Resolution (Å)	50.00-1.78
No. reflections (test set)	30055(1648)
R_{work}/ R_{free} (%)	18.3/21.9
No. atoms	
Protein	2378
CoA	52
Water	242
B-factors (Å^2)	
Protein	24.7
CoA	25.4
Water	31.2
RMSD	
Bond lengths (Å)	0.010
Bond angles (°)	1.351
Ramachandran plot % residues	
Favored	98.1
Additional allowed	1.9
Generously allowed	0
Disallowed	0

Table 4.5. Kinetic parameters of acyl-CoA in wild type and mutant MOF mediated histone modification

	Cofactor	K_m/μM	k_{cat}/min^{-1}	k_{cat}/K_m (min^{-1}.μM^{-1})
WT-MOF	Ac-CoA	1.1±0.4	2.8±0.2	2.4±0.9
	Pro-CoA	0.9±0.2	1.8±0.1	2.0±0.5
MOF-V314A	Ac-CoA	2.7±0.3	2.0±0.1	0.7±0.1
	Pro-CoA	2.4±0.6	2.1±0.2	0.9±0.2
MOF-P349A	Ac-CoA	1.8±0.5	0.9±0.1	0.5±0.1
	Pro-CoA	11.4±7.2	0.6±0.2	0.05±0.04

CHAPTER 5 Identification of Lysine Isobutyrylation as A Novel Histone Acylation

---

Han, Z. and Zheng, G.Y., et al. To be submitted to *Journal of Biological Chemistry*.

Abstract

Short-chain acylations on ε-amino group of lysine residues have emerged as essential posttranslational modifications (PTMs) that regulate gene transcription, cellular metabolism, and cell signaling, etc. Chemically diverse acylations can lead to a variety of downstream biological outcomes. Lysine butyrylation was initially discovered with a mass spectrometry (MS) study and was recognized as linear n-butyrylation without any bias. Here, we report lysine isobutyrylation as a novel PTM on histone lysines. In particular, we found isobutyryl-CoA as an ample metabolite and the potential acyl donor for lysine isobutyrylation. Isobutyryl-CoA is synthesized from isobutyrate and valine in mammalian cells. Our biochemical data showed that isobutyrylation is a *bona fide* PTM on histone lysines and histone acetyltransferase 1 (HAT1) has strong isobutyryltransferase activity. We also found that cellular level of isobutyryl-CoA is drastically decreased in aged cells, which implied a potential pathway for cell longevity regulation. Together, our data defines isobutyrylation as a novel lysine PTM and suggested the potential regulatory function of lysine isobutyrylation in longevity.

Introduction

Short-chain acylation on the Lysine ε-amino group of protein lysines are an important type of reversible posttranslational modifications (PMTs). Many studies have demonstrated the regulatory functions of lysine acylations on a plethora of physiological events such as gene expression, cellular metabolism, cell cycle, and signal transduction.[20,25,144,145] Since lysine acetylation was first discovered in the middle of 1960s,[5] more than ten lysine acylations such as propionylation, butyrylation, crotonylation, succinylation, malonylation, and glutarylation have been discovered.[21-23,28,145] Lysine acylations are deposited by writer enzymes such as lysine acetyltransferases (KATs) using acyl-CoA molecules as the acyl donor and are interpreted and

removed by reader and eraser proteins respectively. Dysregulation of lysine acylation dynamics due to aberrant expression of the writer, reader, and eraser proteins or unbalanced nutrition status leads to various diseased phenotypes including inflammation, neurodegeneration, and cancers etc.[14,97,146-149] Different acylation on the same locus can lead to different biological outcomes.[30,150] For instance, interchangeable acylations on protein lysines have different impacts on gene transcription: butyrylation competes with acetylation on H4K5K8 and prevents the binding of the Lys-Ac reader protein Brdt on these loci, which causes delayed histone removal and gene expression in spermatogenic cells.[24] Therefore, to discover novel lysine acylations and their regulatory mechanism are important for understanding certain physiological and pathological pathways.

In the past decade, the development of high resolution mass spectrometry has largely facilitated the discovery of novel lysine acylations.[20,23,135] Lysine butyrylation was first discovered by Zhao et al in 2007.[28] At the time of its identification, lysine butyrylation was taken for granted to be linear n-butyrylation (NBT). Later, several studies demonstrated that KAT enzymes, including p300, CBP, and GCN5, possess lysine n-butyryltransferase activity.[26,28] Also, n-butyryl-CoA is an important metabolic intermediate during even-chain fatty acids oxidation[151] and therefore, may serve as an ample source for NBT. These lines of biochemical evidence strongly supported the existence of lysine NBT; nevertheless, butyrylation may also exist as another form, isobutyrylation (IBT) at the meantime **(Figure 5.1)**. This hypothesis is strengthened by the natural existence of isobutyryl-CoA, which is generated during valine degradation.[152] In addition, an enzyme known as butyryl-CoA mutase was found in prokaryotes that interconvert n- and iso-butyryl-CoA.[153] Most importantly, dysregulation of isobutyryl-CoA metabolism caused by genetic deficiency of isobutyryl-CoA dehydrogenase (IBD) leads to

multiple symptoms including speech delay, anaemia, and dilated cardiomyopathy on newborn patients while the mechanism behind such disorders are still unexplored.[154,155] Therefore, to study the catabolism of isobutyryl-CoA and dynamic regulation of lysine isobutyrylation has profound physiological significance. In the present study, we define isobutyrylation as a *bona fide* PTM in histone lysines with a combined suite of analytical and biochemical studies.

Results and Discussion

Isobutyryl-CoA is an abundant metabolite in mammalian cells

Upon its identification with mass spectrometry, lysine butyrylation was unbiasedly defined as lysine n-butyrylation.[28] Nevertheless, two possible structures may have the same molecular weight of butyryl chain: linear n-butyryl and branched isobutyryl group. Etiologically, lysine acylations rely on the acyl-CoA molecules as the acyl donor; therefore, an ample pool of acyl-CoA is the prerequisite for the occurrence of cognate lysine acylation. To elucidate the composition of lysine butyrylation, we first tested the abundance of n- and iso-butyryl-CoA in human embryonic kidney (HEK) 293T cells with a high-performance liquid chromatography tandem mass spectrometry (HPLC-MS/MS) approach. Iso- and n-butyryl-CoA standards were separated on the high-performance liquid chromatography with the retention times to be ~13.85 and 14.05 minutes respectively and were monitored at the same ion transition (838→331) on the mass spectrometer (MS) **(Figure 5.2a)**. The chromatographic peak area of butyryl-CoA molecules was integrated for quantification purpose. Then 293T cell extracts were subjected to the same HPLC-MS/MS analysis and the abundance of n- and iso-butyryl-CoA was compared. Duplicate experiments were conducted, showing that the ratios of n- and iso-butyryl-CoA were 67:27 and 62:21 **(Figure 5.2b)**. Thus, the abundance of isobutyryl-CoA is 2-3 times higher than n-butyryl-CoA. N-butyryl-CoA is a rich metabolic intermediate in fatty acids metabolism and

serves as an ample source for lysine n-butyrylation. Hence, isobutyryl-CoA, which is ~3 times more abundant than n-butyryl-CoA, can also serve as the substantial isobutyryl donor for lysine isobutyrylation.

Isobutyryl-CoA is synthesized through isobutyrate and valine metabolism in mammalian cells

Acyl-CoA molecules are metabolic intermediates mostly in tricarboxylic acid cycle (TCA), fatty acid β-oxidation, and amino acid degradation.[116] Many studies reported that treatment of cells with fatty acids such as propionate induces intracellular synthesis of the cognate acyl-CoA molecules with the catalysis of acyl-CoA synthetase (ACS).[135] Isobutyryl-CoA is a known intermediate during propionyl-CoA production from valine metabolism.[156] In addition, an enzyme called isobutyryl-CoA mutase was found in prokaryotes and is able to interconvert n- and iso-butyryl-CoA **(Figure 5.3a)**.[153]Whether this enzyme exists in eukaryotes remains unclear. These isobutyryl-CoA catabolic pathways further enhance the possibilities that isobutyryl-CoA acts as an essential regulator in physiological processes. Here, we studied the effect of isobutyrate treatment on the changing of cellular isobutyryl-CoA level. In particular, HEK293T cells were treated with varied concentrations of deuterated sodium isobutyrate (d7-isobutyrte). Extracted butyryl-CoA molecules were analyzed with the HPLC-MS/MS method aforementioned. The level of butyryl-CoA molecules was quantified in the unit of peak area/mg of cellular protein **(Figure 5.3b)**. With the increase of d7-isobutyrate concentration, d7-isobutyryl-CoA level increased drastically until a plateau was reached when 5 mM of d7-isobutyrate was used. The level of non-isotopic isobutyryl-CoA remained barely changed until the cells were treated with 10 mM of d7-isobutyrate, which largely competed with non-isotopic isobutyrate for the acyl-CoA synthetase. We also tested how isobutyryl-CoA level change in

response to valine feeding. The data showed that treatment of 293T cells with valine significantly upregulated the level of isobutyryl-CoA until a plateau was reached at the point that 5 mM of valine was used **(Figure 5.3b)**. This observation is consistent with the aforementioned valine metabolic pathway, in which isobutyryl-CoA is an intermediate.[156] Of note, neither d7-isobutyryate nor valine treatment induced the upregulation of n-butyryl-CoA level, possibly suggested that isobutyryl-CoA mutase is not present in 293T cells. These data together demonstrated that isobutyryl-CoA is synthesized from the acyl-CoA synthetase pathway and valine metabolism. These data not only defined the catabolism of isobutyryl-CoA in mammalian cells but also provided us with a nice approach to study the dynamic change of lysine isobutyrylation in cellular proteins.

HAT1 is a potential isobutyryltransferase

Lysine acylations are driven by KATs, which can covalently deposit acyl groups to protein lysines using acyl-CoA as the cofactor. Identification of isobutyryl-CoA as an ample acyl-CoA donor strongly supported the fact that lysine isobutyrylation exists in cellular proteins. We next tested the isobutyrylation activity of nine eukaryotic KAT enzymes. Although were first identified as lysine acetyltransferases, many KAT enzymes function very promiscuously. They carry out lysine propionylation, butyrylation, and crotonylation, etc. other than the classic acetylation.[26-28,123] To test if some of KAT enzymes also act as lysine isobutyryltransferases, we quantitively measured the acylation activities of nine eukaryotic KATs from three major KAT families including the MYST, the GCN5/PCAF, and the p300/CBP family using a fluorescent CPM assay.[43] Histone peptides H3-20 and H4-20 (twenty amino acid residues from the N-terminus of histone H3 or H4) were used as the acyl acceptor to characterize the acetyl-, propionyl-, n-butyryl-, and isobutyrylation activities of individual KAT enzymes **(Figure 5.4)**.

Elongation of acyl chain induces decrease of acylation activity of KAT enzymes, except that HAT1 has stronger propionylation (KPT) activity than acetylation (KAT) activity. All the tested KAT enzymes show appreciable KPT activities, consistent with our recent study, which demonstrates that almost all eukaryotic KATs possess strong KPT activity. **(Chapter 4, Table 4.1)** Nevertheless, further increase of the length of acyl chain to four carbons leads to drastic decrease of butyrylation activity (KBT) of KAT enzymes. Among the tested KAT enzymes, HAT1 showed outstanding isobutyrylation (IBT) activity, ~25% of its KAT activity while the other KAT enzymes show much lower or even barely detectable IBT activity. p300 is known for its great cofactor promiscuity, functions as lysine acetyl-, propionyl-, butyryl-, and crotonyltransferase.[25,28,30] Its unique acyl-CoA binding pocket enables p300 to bind with larger acyl-CoA molecules without significant structural change.[42,157,158] Therefore, we carried out the kinetic characterization of p300 and HAT1 with acyl-CoA molecules to validate their acyltransferase activities using a CPM fluorogenic assay. The kinetic constants K_m and k_{cat} were determined by fitting acyl-CoA concentration-reaction rate to the Michaelis-Menten model. The k_{cat}/K_m ratio represents the activity of KAT enzymes with different acyl-CoA molecules. The k_{cat}/K_m ratio of p300 with n- and isobutyryl-CoA are 0.25 $min^{-1}.\mu M^{-1}$ and 0.13 $min^{-1}.\mu M^{-1}$ respectively, 7% and 13% of p300 with acetyl-CoA **(Figure 5.5a, Table 5.1)**. In comparison, the NBT and IBT activity of HAT1 was about 13% and 31% of its KAT activity respectively **(Figure 5.5b, Table 5.1)**, suggesting that HAT1 may serve as a genuine lysine isobutyryltransferase. We further tested the IBT activity of HAT1 on protein substrate. Recombinant histone H4 was co-incubate with butyryl-CoA and HAT1, followed by testing with anti butyryllysine antibody. Although the antibody was designed for detection of lysine NBT, it also recognizes lysine IBT efficiently. As shown in **Figure 5.6**, sole co-incubation of histone H4

with butyryl-CoA induced little histone labeling while co-incubation of butyryl-CoA, histone H4, and HAT1 drastically increased lysine butyrylation level on histone H4. In the previous studies, anti-butyryllysine antibody was used to enrich or recognize butyrylated lysines in cellular proteins to study the physiological functions of lysine butyrylation.[4,24] Our finding about the low specificity of anti-butyryllysine antibody indicates that many butyrylated lysines identified in previous works could be a mixture of n- and iso-butyrylated lysines. In summary, the kinetic measurement and western blot data validated the IBT activity of HAT1. The anti-lysine NBT antibody provided us with an antibody substitute for further study of lysine IBT in cellular contexts.

Lysine isobutyrylation is a *bona fide* PTM on histone lysines

Identification of isobutyryl-CoA as an ample metabolite and HAT1 as a potential isobutyryltransferase are strong but indirect evidences to demonstrate the existence of lysine isobutyrylation. To gain a direct insight about isobutyrylation distribution on protein lysines, we next focused on the detection of isobutyrylation on both core histones and cellular proteome. In particular, the 293T cells were treated with 10 mM d7-isobutyrate for 16 hours, followed by extraction of core histone proteins and whole cellular proteome. Lysine isobutyrylation level of extracted proteins were tested with anti-butyryllysine antibody. Isobutyrate treatment upregulated the chemiluminescence signal on core histones, especially histone H4 **(Figure 5.7)**. Nevertheless, no appreciable change of chemiluminescence intensity is observed on non-histone proteins upon isobutyrate treatment. Since isobutyrate treatment only induce synthesis of isobutyryl-CoA but not n-butyryl-CoA **(Figure 5.3b)**, the chemiluminescence change is induced by histone isobutyrylation. Therefore, lysine isobutyrylation is a *bona fide* PTM on histone lysines.

Isobutyryl-CoA level decreases in aged cells

Amino acids such as serine, threonine and valine and can sensitize cell to oxidative stress and promote cell aging through different mechanisms.[159] In yeast, threonine and valine can activate the target of rapamycin (TOR) pathway and serine can activate the Pkh1/2 protein. Both processes will converge to activate the serine/threonine-protein kinase SCH9, which can suppress the expression of downstream stress resistance genes and decrease life span extension of yeast.[160-162] Leucine catabolism produces acetic acid, which can subsequently be synthesized to acetyl-CoA. Fluctuation of leucine level in yeast can induce the change of acetylation state of histones, which have major impacts on cellular longevity.[163,164] Therefore, previous studies have well demonstrated how amino acids and their metabolizing products affect cell aging. Since isobutyryl-CoA is an important intermediate during valine catabolism, we measured the abundance of isobutyryl-CoA in 293T cells at different ages. The newly recovered cells, the cells passaged for 4 and 7 times were collected and subjected to butyryl-CoA quantification analyses respectively. The isobutyryl-CoA level is 125, 53, and 13 integrated peak area / mg of cellular protein in the cells passaged for 1, 4, and 7 times respectively while the level of n-butyryl-CoA barely changed in cells with different ages **(Figure 5.8)**. This is possibly because the aged cells are less capable of metabolizing valine and therefore produced less isobutyryl-CoA. In this case, isobutyryl-CoA level can be used as a potential landmark for cell aging characterization. In addition, decrease of isobutyryl-CoA cellular abundance may induce the change of histone PTM states, which may be associated with cell aging pathways. Further study of how histone isobutyrylation change in response to the fluctuation of isobutyryl-CoA level may reveal novel pathways of cellular longevity regulation.

Conclusion

Chemically diverse acylations largely enriched the regulatory mechanisms of lysine PTM towards cellular processes. Lysine butyrylation was recognized as n-butyrylation upon its discovery.[28] Later study revealed that competing butyrylation and acetylation on the same lysine residue would lead to different biological outcomes.[24] In the present study, we identified lysine isobutyrylation as a *bona fide* PTM on histone lysines. While current antibodies cannot distinguish n- and isobutyrylation **(Figure 5.6 and Figure 5.7)**, identification of lysine isobutyrylation indicates that the previously revealed physiological impacts of lysine n-butyrylation can be partially caused by isobutyrylation. We also showed that isobutyryl-CoA is an ample metabolite in mammalian cells and its biosynthetic pathway is independent from n-butyryl-CoA. The enzyme activity assay showed that lysine n-butyrylation and isobutyrylation are regulated by different writer enzymes, p300 and HAT1 respectively. These data strongly support that lysine NBT and IBT may regulate different biological processes and function in different mechanisms. This hypothesis is also supported by the finding that cellular level of isobutyryl-CoA decreased drastically as cell getting aged while n-butyryl-CoA level remain stable during cell aging. Nevertheless, how isobutyryl-CoA level regulates cell longevity remain undefined at this stage. Future work on the study of how lysine isobutyrylation regulate the expression of longevity regulatory genes may reveal novel pathways of cell aging regulation.

Methods and Materials

Quantification of butyryl-CoA level in 293T cells

Human embryonic kidney 293T (HEK293T) cells were cultured to 90% confluence with DMEM medium supplemented with 10% fetal bovine serum (FBS) and 1% streptomycin-penicillin. Cells were also treated with varied concentrations of d7-deuterated isobutyrate

(Product# 632007, from SIGMA-ALDRICH) and valine. Cells were washed with ice-cold PBS buffer followed by fixing in methanol at −80 ºC for 15 minutes. Cells were collected in 50% of methanol with gentle scrape. Cell suspension was centrifuged at 16000g and supernatant was collected for LC-MS/MS analysis.

An Atlantis® T3 (4.6×150 mm, 3 μm) column with a Phenomenex SecurityGuard C-18 guard column (4.0 mm×2.0 mm) was applied to separate analytes. The column temperature was constant at 32 °C. The mobile phase A was 10 mM ammonium acetate, and mobile phase B was acetonitrile. A gradient method was applied for separation, with a 0.4 mL/min flow rate, (time/minute, % mobile phase B): (0, 6), (15, 30), (15.01, 100), (22.50, 100), (22.51, 6). The injection volume was 30 μL, and the autosampler injection needle was washed with methanol after each injection. Nitrogen was used as the desolvation gas at a flow rate of 500 L/h. The desolvation temperature was 500 °C and the source temperature was 120 °C. Argon was used as the collision gas, and the collision cell pressure was 3.5×10^{-3} mbar. Samples were analyzed in the positive ion mode. The capillary voltage was 3.2 kV, the cone voltage was 42 V and the collision energy was 22 eV. A multiple reaction monitoring (MRM) function was applied for the detection of analytes. The ion transition 838→331 was monitored for iso- and n-butyryl-CoA, and 845→338 for d7-isobutyryl CoA.

Measurement of lysine acetyltransferase activity.

Synthetic histone peptides H3-20 or H4-20 (20 amino acids from the N-terminal of histone H3 and H4, the sequence of H3-20 is Ac-ARTKQTARKSTGGKAPRKQL, the sequence of H4-20 is Ac-SGRGKGGKGLGKGGAKRHRK) were used as acyl acceptor substrates. For single-point quantification assays **(Figure 5.4)**, 30 μM acyl-CoA was incubated with individual KAT enzymes and 100 μM histone peptides. KAT enzymes deposit acyl groups on the ε-amine of

peptide lysines and release CoASH simultaneously. The fluorogenic probe 7-diethylamino-3-(4'-maleimidylphenyl)- 4-methylcoumarin (CPM) was added to react with by-product CoASH to form the fluorescent CoAS-CPM complex [43]. The fluorescence intensities were measured with a microplate reader (FlexStation® 3) and the results were shown in a bar graph **(Figure 5.4)**.

For kinetic characterization of p300 and HAT1 with acyl-CoA molecules. Varied concentration of acyl-CoA was incubated with p300 or HAT1 and H4-20 peptide at fixed concentrations. The fluorescence intensity was measured with the same method as the single-point assay and catalytic rate was determined from the fluorescence intensity. Kinetic constants including binding affinity (K_m) and catalytic efficiency (k_{cat}) were determined by fitting the acyl-CoA concentration-catalytic rate to the Michaelis-Menten equation. The results were shown in **Figure 5.5 and Table 5.1**.

Test of histone H4 butyrylation.

1 μg recombinant histone H4 was incubated with 50 μM acyl-CoA and 0.2 μM of HAT1. The enzymatic reaction lasted for 1 hour, followed by lysine acylation detection with western blot analysis using anti-butyryllysine antibody (Product# PTM-301, from PTM Biolabs). The results are shown in **Figure 5.6**.

Test of lysine isobutyrylation changing in response to isobutyrate treatment.

HEK 293T cells were cultured to 90% confluence in DMEM medium supplemented with 10% FBS and 1% streptomycin-penicillin antibiotics. The cells were treated with 10 mM d7-isobutyrate for 16 hours followed by cellular protein extraction. Whole cell lysate was extracted in M-PER™ Mammalian Protein Extraction Reagent (Product# 78501, from ThermoFisher Scientific) with gentle sonication and core histone proteins were extracted with the EpiQuik Total Histone Extraction Kit (Product# OP-0006-100, from Epigentek). The extracted proteins

were tested with western blot analyses using anti-butyryllysine antibody (Product# PTM-301, from PTM Biolabs). The results were shown in **Figure 5.7**.

Figures, Tables and Captions

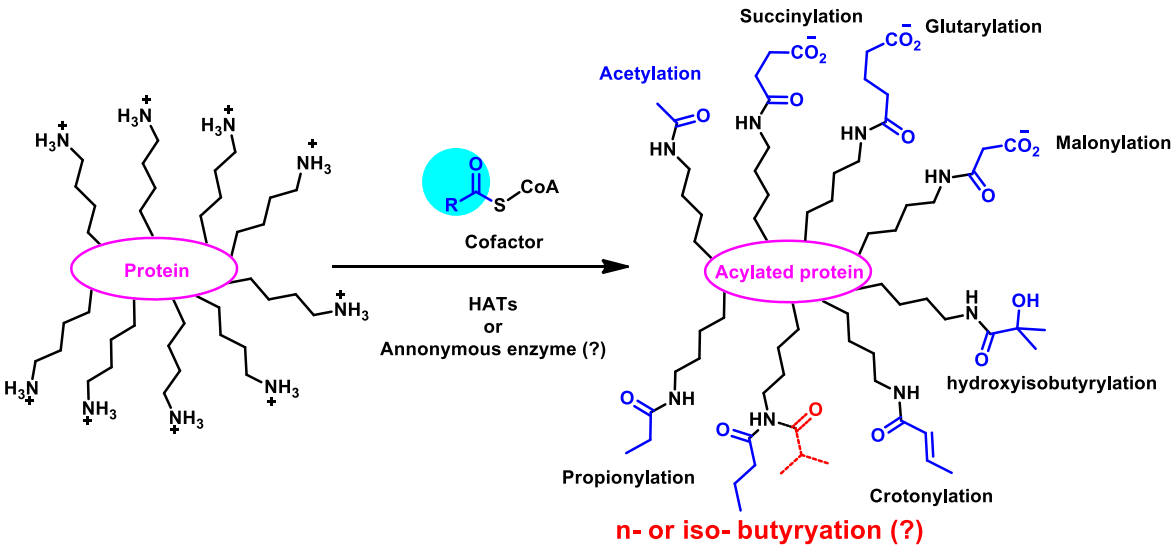


Figure 5.1. Chemically diverse acylations on protein lysines.

Lysine butyrylation was identified in 2007[28] and was recognized as n-butyrylation. Nevertheless, whether butyrylation exists as isobutyrylation remained unrevealed. Identification of butyrylation composition will assist in the mechanistic study of physiological processes that are regulated by lysine butyrylation.

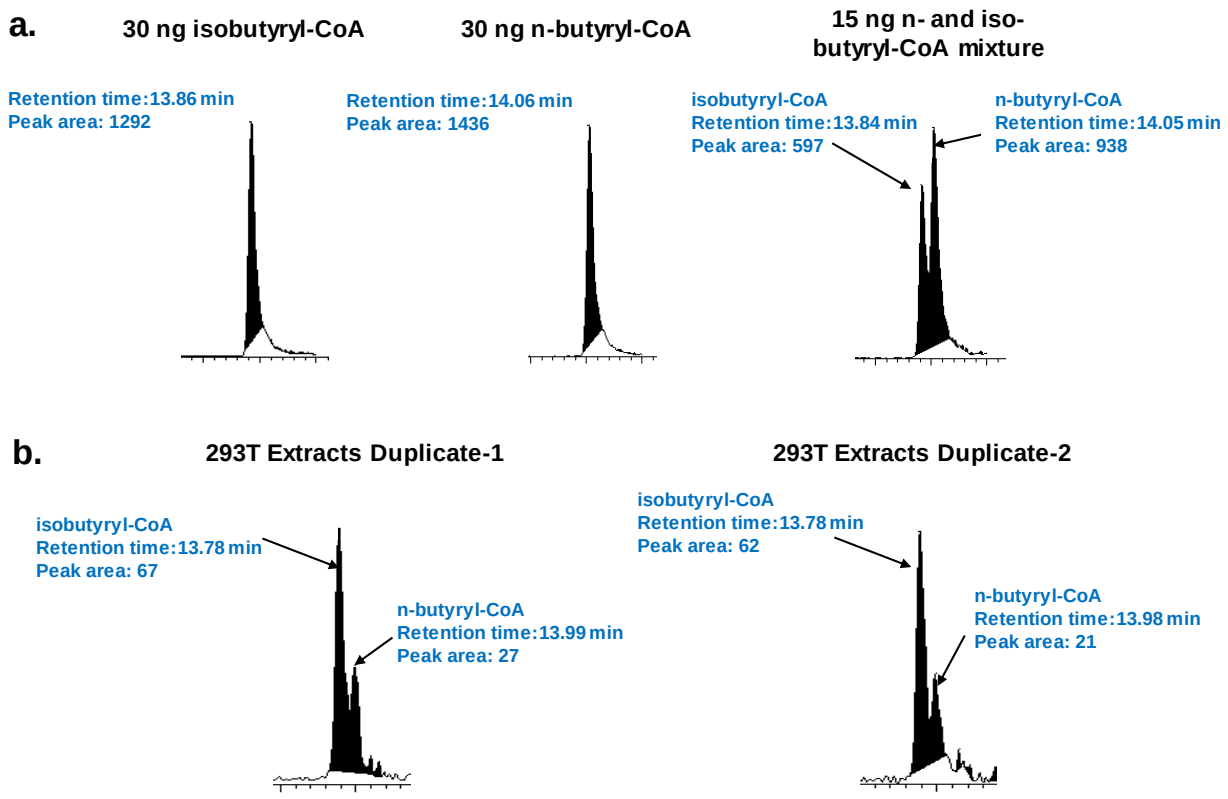


Figure 5.2. Quantification of n- and iso-butyryl-CoA.

a. N- and iso-butyryl-CoA were separated with the LC-MS/MS system with the retention times at 14.06 and 13.86 minutes and were detected the same ion transition 810 → 313. **b.** N- and iso-butyryl-CoA in the cell extracts were detected with the same LC-MS/MS condition. Their levels were compared using the integrated chromatographic peak areas.

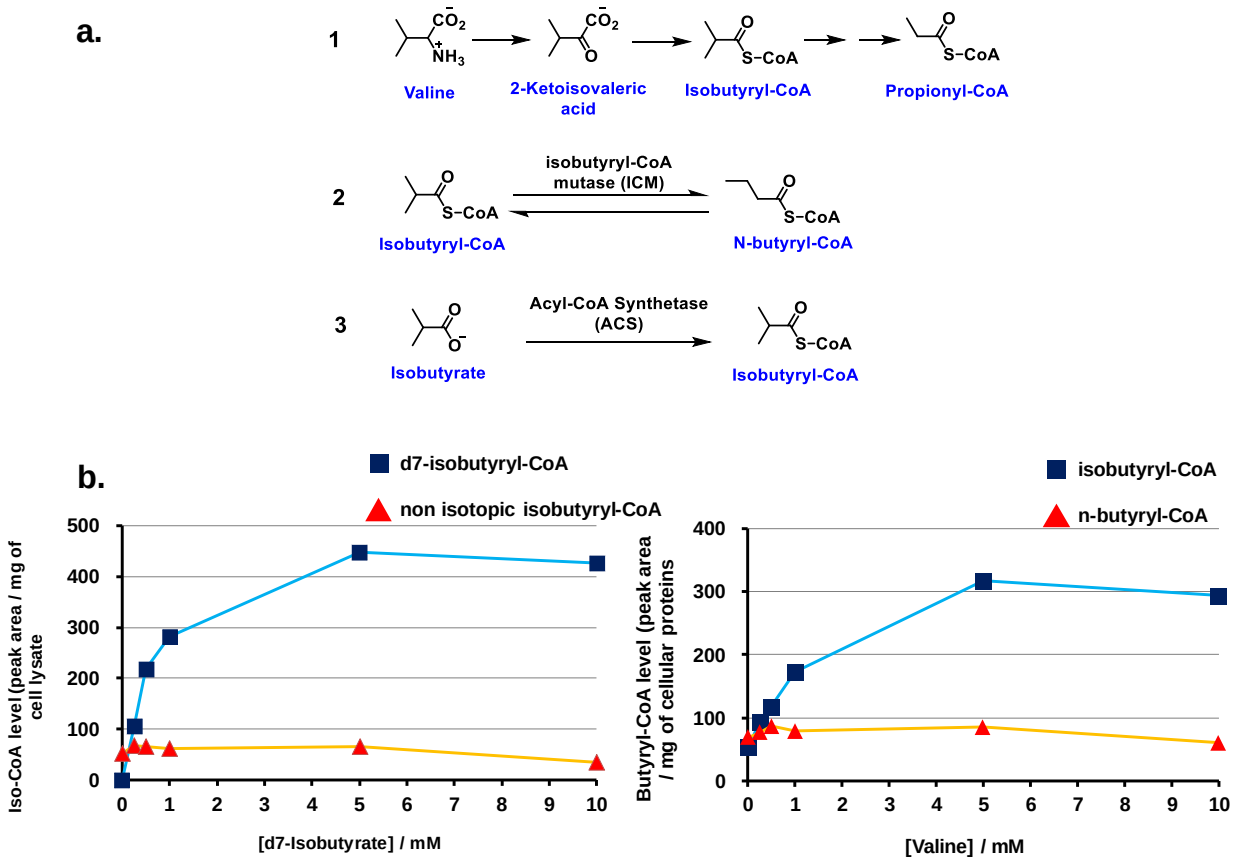


Figure 5.3. Study of isobutyryl-CoA catabolism pathways.

a. Reported and proposed pathways for isobutyryl-CoA catabolism. It is reported that isobutyryl-CoA is an intermediate during valine metabolism (1).[156] A prokaryotic enzyme called isobutyryl-CoA mutase can interconvert iso- and n-butyryl-CoA while whether this enzyme exists in eukaryotes remains unclear (2).[153] Acyl-CoA synthetase is a known enzyme that can produce acyl-CoA molecules from cognate acytates.[165,166] It is highly possible that isobutyryl-CoA can also be synthesized from this pathway (3). **b. Cellular level of isobutyryl-CoA changes in response to isobutyrate and valine feeding.** Isobutyryl-CoA can be produced from acyl-CoA synthetase pathway and valine metabolism in mammalian cells. N-butyryl-CoA level is not affected in response to isobutyrate or valine treatment.

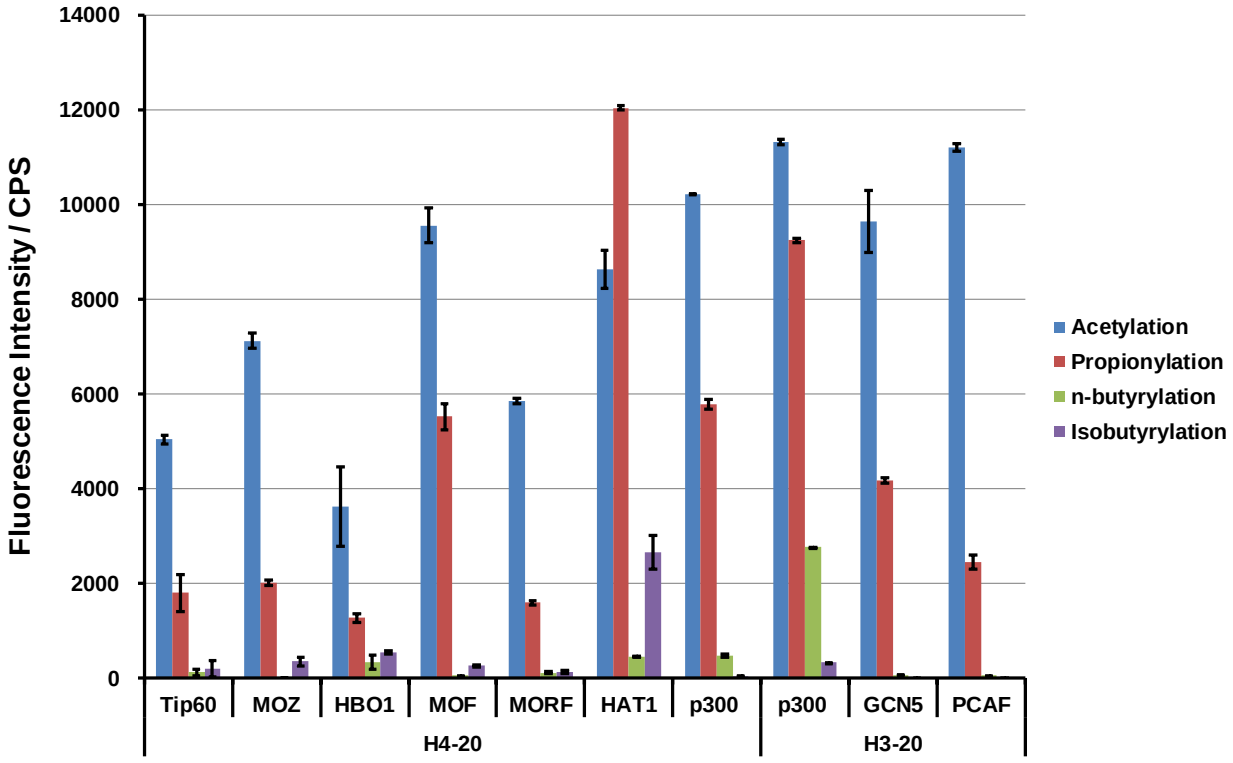


Figure 5.4. Measurement of lysine acylation activity of eukaryotic KAT enzymes.

Lysine acetyl-, propionyl-, n-butyryl-, and isobutyryltransferase activity of each KAT enzymes was tested on H3-20 and H4-20 histone peptide substrates. All tested KATs show strong activity on lysine acetylation and appreciable activity on lysine propionylation. HAT1 shows the strongest activity of carrying out lysine isobutyrylation, about 25% of its acetyltransferase activity.

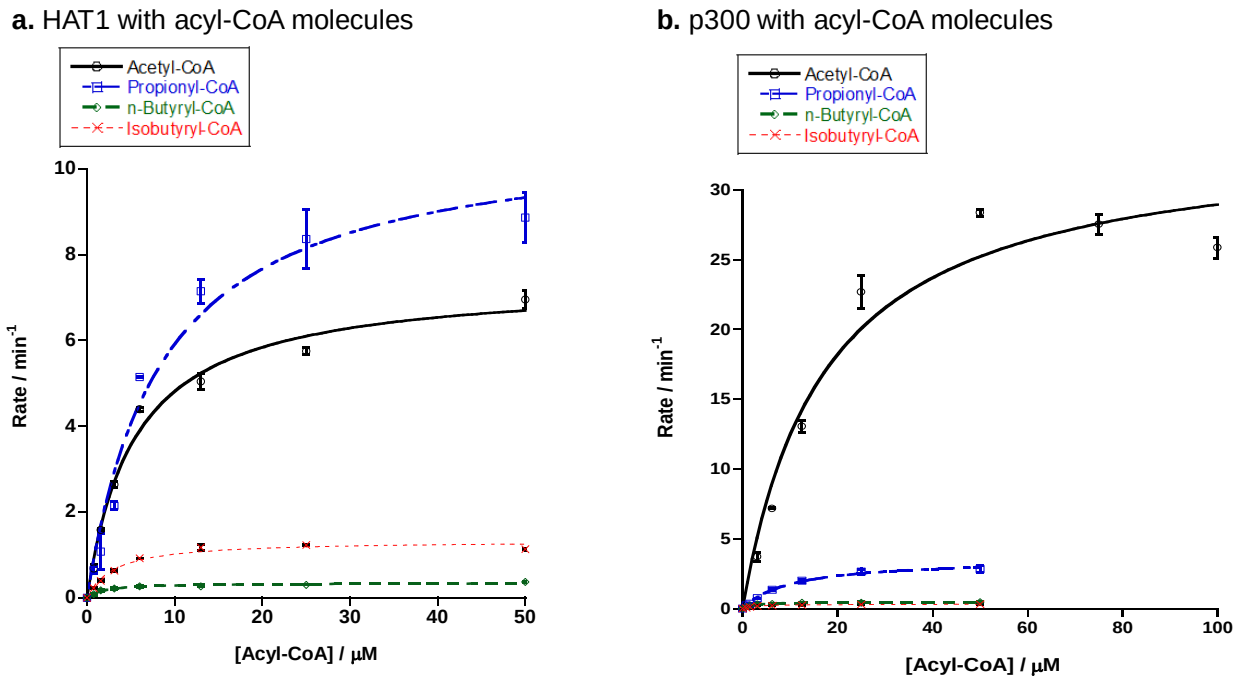


Figure 5.5 Kinetic characterization of HAT1 and p300 with acyl-CoA.

HAT1 or p300 was incubated with individual acyl-CoA molecule at varied concentrations and H4-20 peptide substrates. The by-product CoASH from enzymatic reaction was quantified with a fluorescent assay. The reaction rate-acyl-CoA concentration were plotted with the Michaelis-Menten equation to get the kinetic constants K_m and k_{cat}. The k_{cat}/K_m value were used to evaluate lysine acylation activity. **a.** k_{cat}/K_m value of HAT1 with acetyl-, propionyl-, n-butyryl-, and isobutyryl-CoA are: 1.4 ± 0.2, 1.3 ± 0.3, 0.21 ± 0.06, and 0.42 ± 0.09 min^{-1}. μM^{-1}, **b.** k_{cat}/K_m value of p300 with acetyl-, propionyl-, n-butyryl-, and isobutyryl-CoA are: 2 ± 0.6, 0.63 ± 0.12, 0.25 ± 0.09, and 0.13 ± 0.02 min^{-1}. μM^{-1}.

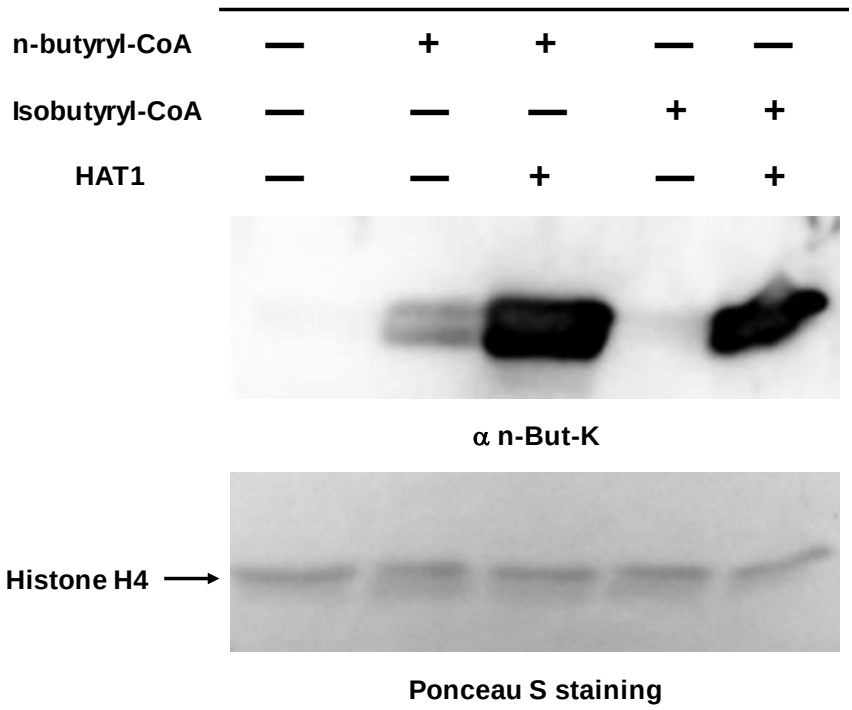


Figure 5.6. Detection of HAT1 mediated lysine isobutyrylation.

Recombinant histone H4 was incubated with HAT1 and butyryl-CoA. Butyrylation level on histone H4 lysines was detected with anti n-butyryllysine antibody (Product# PTM-301 from, PTM Biolabs). The antibody can strongly detect both n- and isobutyrylation on histone lysines.

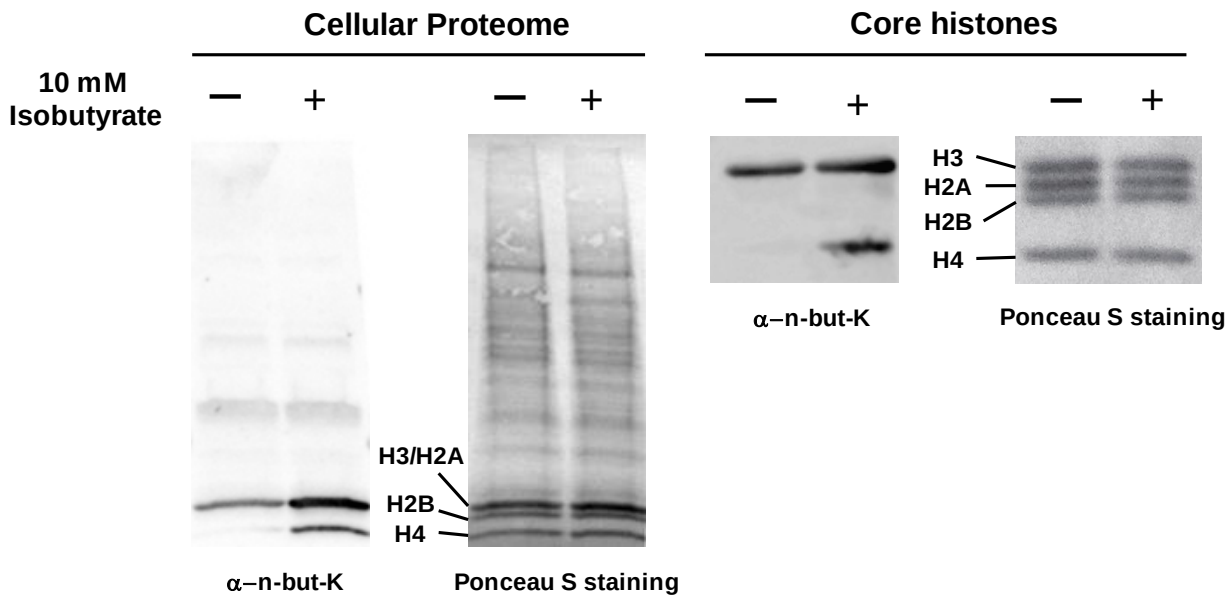


Figure 5.7. Detection of lysine isobutyrylation on protein lysines.

293T cells were treated with isobutyrate to induce the synthesis of isobutyryl-CoA. Cellular proteome and core histone proteins were extracted and tested with anti butyryllysine antibody. Treatment of cells with isobutyrate induced increase of lysine isobutyrylation level on core histone proteins while no appreciable change was observed on non-histone protein isobutyrylation upon isobutyrate treatment.

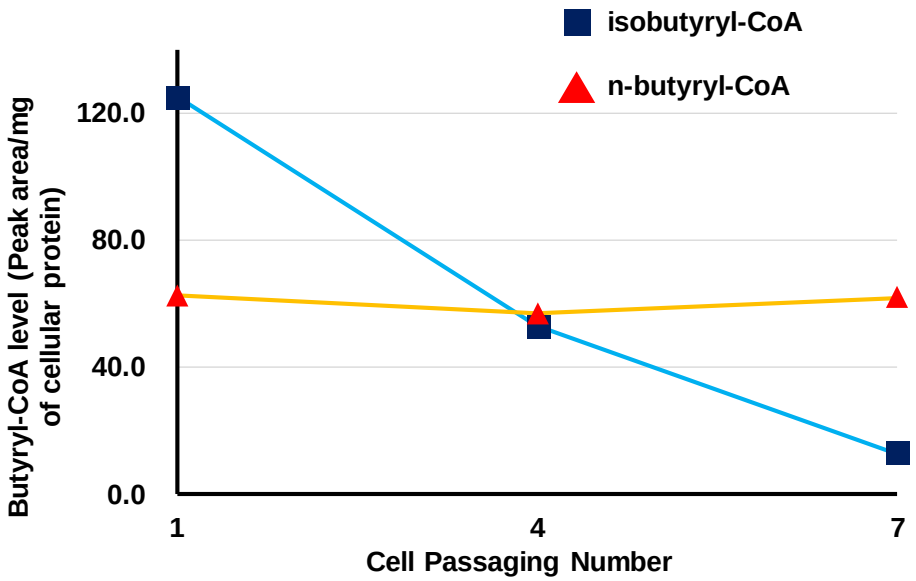


Figure 5.8. Quantification of butyryl-CoA level in response to cell aging.

Frozen 293T cells were recovered in DMEM medium and were passaged when the confluence reaches to 80%. Butyryl-CoA level in the cells being passaged for 1, 4, and 7 times were quantified with the LC-MS/MS method aforementioned. Isobutyryl-CoA level decreases as the cell getting aged while n-butyryl-CoA level remain unchanged.

Table 5.1. Kinetic characterization of HAT1 and p300 to different acyl-CoAs

	Acyl-CoA	K_m / μM	k_{cat} / min^{-1}	k_{cat}/K_m (min^{-1}. μM^{-1})
p300	Acetyl-CoA	17.3 ± 4.8	33.9 ± 2.9	2 ± 0.6
	Propionyl-CoA	16.3 ± 2.8	10.2 ± 0.8	0.63 ± 0.12
	n-butyryl-CoA	1.9 ± 0.2	0.5 ± 0.17	0.25 ± 0.09
	Isobutyryl-CoA	2.4 ± 0.3	0.3 ± 0.01	0.13 ± 0.02
HAT1	Acetyl-CoA	5.4 ± 0.8	7.4 ± 0.3	1.4 ± 0.2
	Propionyl-CoA	8.5 ± 1.9	10.9 ± 0.8	1.3 ± 0.3
	n-butyryl-CoA	1.9 ± 0.5	0.4 ± 0.02	0.21 ± 0.06
	Isobutyryl-CoA	3.1 ± 0.6	1.3 ± 0.1	0.42 ± 0.09

CHAPTER 6 Conclusion

Lysine acylations are reversible posttranslational modifications (PTM) on the ε-amino group of protein lysines. These processes widely regulate biological events such as gene expression, cell cycle, DNA repair, cell metabolism, and cell signaling by altering protein stability, protein-nucleotide interaction, and protein-protein interaction, etc.[6,167,168] The homeostasis of lysine acylation is finely tuned by the writer proteins, lysine acetyltransferases (KATs), and the eraser proteins, lysine deacetylases (KDACs).[146] The effector proteins such as bromodomain can specifically recognize and bind with the acylation marks and induce variety of downstream biological outcomes.[9,169] The acyl donor of lysine acylations, acyl coenzyme A molecules are produced in the metabolism of fatty acids, amino acids, and carbohydrates.[116,166,170] Dysregulation of lysine acylation caused by dysfunction of regulatory proteins or abnormality in cellular metabolism can lead to multiple diseased phenotypes such as diabetes, neurogenerative disorders, and caners.[16,97,146,147,171] Hence, to study the mechanisms of how lysine acylation dynamics are regulated is important to understand KAT functions and to develop effective therapies for KAT-related disorders. In this dissertation, we investigated the novel activities and functions of lysine acetyltransferases (KATs) from different perspectives, which lead to cutting-edge discoveries with respective to methodology development for KAT activity measurement, KAT substrates and function investigation, and discovery of novel KAT activity and novel acylation mark.

We designed and synthesized a clickable acetyl-CoA surrogate 3-azidopropionyl-CoA (3AZ-CoA) that can actively react with wild type and engineered KAT enzymes p300, MOF-

I317A, and GCN5-T612G.[42] Based on the bioorthogonal probe 3AZ-CoA, we developed an activity based protein profiling (ABPP) approach and identified more than four hundred substrates of p300 and GCN5-T612G from cellular contexts. From bioinformatic analyses, we confirmed the canonical regulatory functions of p300 and GCN5 in gene expression, cellular metabolism, cell cycle, and DNA repair. In addition, the pathway investigation shows that GCN5 and p300 may also regulate RNA splicing and translation, antigen processing and presentation, and protein export. This work provides a great insight for future work to investigate the functional involvement of GCN5 and p300 in these newly discovered biological pathways. Other than substrates profiling, 3AZ-CoA was also applied in a KAT activity measurement approach. By welding different technologies including CuAAC click chemistry, enzyme engineering, and quenching fluorescence resonance energy transfer (q-FRET), we developed a mix-and-read method for KAT activity measurement and inhibitor characterization. This method has certain advantages over the current KAT activity assays. In comparison to traditional radiometric assays, this Q-FRET method avoids the complicated procedures of handling radioactive materials. The distance-dependent intramolecular FRET mechanism suggests that this assay type is more suitable to organic dyes in inhibitor screenings. In addition, this method directly measures the production of acylated histone substrates, better than the CPM assays that quantify the by-product CoASH. The instability of acyl-CoA molecules and the thiol containing reagents can easily lead to false positive results in the latter assays system. Therefore, the high-performance assay strategy presented here has great potential to be widely applied for KAT activity measurement and inhibitor screening and characterization.

Chemical biology studies summarized in chapter 2 and 3 lead to valid approaches for KAT sub-acylome profiling and KAT activity measurement. Our biochemistry and molecular biology

investigations are summarized in chapter 4 and 5, which focus on the characterization of novel activity of KAT enzymes and discovery of novel lysine acylation PTM. The discovery of chemically diverse protein lysine acylations including propionylation, butyrylation, crotonylation, and succinylation, etc. represents an exciting area of research in biology.[20,22,23,28,127] It embodies the potential and important regulatory roles of cellular acyl-CoA metabolites in the modulation of epigenetics and signal transduction. It would be vitally important to determine the cofactor promiscuity of different KATs in order to elucidate the biochemical etiology of cellular protein acylations. Among the three major eukaryotic KAT families, the p300/CBP and GCN5/PCAF members are known as promiscuous acyltransferases that can carry out lysine propionylation and butyrylation to different levels other than classic lysine acetylation.[26-28,30] The MYST family, which contains the largest number of KAT enzymes and works differently than the p300/CBP and GCN5/PCAF KATs, has been known as strict acetyltransferases. In this study, we found that the MYST family of KATs showed strong *bona fide* lysine propionyltransferase (KPT) activity on both histone and non-histone substrates. The proteomic analysis revealed that the representative MYST member MOF has highly identical acetylome and propionylome profiles while the modification levels at individual lysine residues were slightly different. Nevertheless, the propionylome of MOF is partially different from other KPTs such as GCN5 and p300. We, for the first time, solved the crystal structure of MOF bound with propionyl-CoA and found that the amino acid residue proline 349 in MOF, conserved through the MYST family of KATs, is required for its KPT activity. These findings suggested that MYST KATs carry out both lysine acetylation and propionylation on a wide range of cellular proteins and therefore, may regulate much broader cellular processes than sole chromatin dynamics. Together with previous studies

showing the KPT activity of GCN5/PCAF and p300/CBP KAT members, the present study defines a holistic view of eukaryotic KAT enzymes as KPTs.

Lysine butyrylation on ε-amino group in protein lysines has two extra carbons on the acyl chain than lysine acetylation. It was first discovered in 2007 with a mass spectrometry study and was recognized as n-butyrylation without any bias.[28] Many studies have suggested that lysine acetylation and butyrylation on the same lysine residue may lead to different biological outcomes, indicated the distinct biological effects of lysine butyrylation.[24] Our study shown in chapter 5 demonstrated that isobutyryl-CoA can be produced from the acyl-CoA synthetase pathways and valine metabolism in mammalian cells. Isobutyryl-CoA in newly recovered 293T cells is three times more abundant than n-butyryl-CoA but the isobutyryl-CoA level is drastically decreased in aged cells. The enzyme activity assay and western blot analyses suggested that HAT1 is a strong lysine isobutyryltransferase and lysine isobutyrylation only exists in histone lysines. These findings together demonstrate that lysine isobutyrylation is a *bona fide* PTM in histone proteins. The previously reported biological effects of lysine butyrylation may be partially induced by lysine isobutyrylation and more biological effects of lysine isobutyrylation may be revealed in future studies.

In summary, this dissertation focuses on the investigation of KAT activities and functions. We developed two chemical biology approaches for KAT sub-acylome profiling and high throughput measurement of KAT activity, respectively. We identified the dual enzymatic activity of MYST KATs as both acetyltransferases and propionyltransferases with a combined suite of biochemistry, mass spectrometry, and structural biology approaches. We also defined lysine isobutyrylation as a novel PTM on histone proteins. These data provide valuable insights on the mechanistic studies of KAT-related physiological and pathological processes.

REFERENCES

1 He, L. *et al.* Metformin and Insulin Suppress Hepatic Gluconeogenesis through Phosphorylation of CREB Binding Protein. *Cell* **137**, 635-646, doi:10.1016/j.cell.2009.03.016 (2009).

2 Popovic, D., Vucic, D. & Dikic, I. Ubiquitination in disease pathogenesis and treatment. *Nat Med* **20**, 1242-1253, doi:10.1038/nm.3739 (2014).

3 Lee, S. Post-translational modification of proteins in toxicological research: focus on lysine acylation. *Toxicol Res* **29**, 81-86, doi:10.5487/TR.2013.29.2.081 (2013).

4 Pougovkina, O., te Brinke, H., Wanders, R. J. A., Houten, S. M. & de Boer, V. C. J. Aberrant protein acylation is a common observation in inborn errors of acyl-CoA metabolism. *Journal of Inherited Metabolic Disease* **37**, 709-714, doi:10.1007/s10545-014-9684-9 (2014).

5 Allfrey, V. G., Faulkner, R. & Mirsky, A. E. Acetylation and Methylation of Histones and their Possible Role in Regulation of Rna Synthesis. *P Natl Acad Sci USA* **51**, 786-+, doi:DOI 10.1073/pnas.51.5.786 (1964).

6 Sykes, S. M. *et al.* Acetylation of the p53 DNA-binding domain regulates apoptosis induction. *Molecular cell* **24**, 841-851, doi:10.1016/j.molcel.2006.11.026 (2006).

7 Kendrick, A. A. *et al.* Fatty liver is associated with reduced SIRT3 activity and mitochondrial protein hyperacetylation. *Biochemical Journal* **433**, 505-514, doi:10.1042/Bj20100791 (2011).

8 Marmorstein, R. & Zhou, M. M. Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harbor perspectives in biology* **6**, a018762, doi:10.1101/cshperspect.a018762 (2014).

9 Zeng, L. & Zhou, M. M. Bromodomain: an acetyl-lysine binding domain. *FEBS Lett* **513**, 124-128 (2002).

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

11 Nile, A. H. & Hannoush, R. N. Fatty acylation of Wnt proteins. *Nature chemical biology* **12**, 60-69, doi:10.1038/nchembio.2005 (2016).

12 Friedmann, D. R. & Marmorstein, R. Structure and mechanism of non-histone protein acetyltransferase enzymes. *FEBS J* **280**, 5570-5581, doi:10.1111/febs.12373 (2013).

13 Vecellio, M. *et al.* The histone acetylase activator pentadecylidenemalonate 1b rescues proliferation and differentiation in the human cardiac mesenchymal cells of type 2 diabetic patients. *Diabetes* **63**, 2132-2147, doi:10.2337/db13-0731 (2014).

14 Di Cerbo, V. & Schneider, R. Cancers with wrong HATs: the impact of acetylation. *Brief Funct Genomics* **12**, 231-243, doi:10.1093/bfgp/els065 (2013).

15 Johnson, C., Warmoes, M. O., Shen, X. & Locasale, J. W. Epigenetics and cancer metabolism. *Cancer Lett* **356**, 309-314, doi:10.1016/j.canlet.2013.09.043 (2015).

16 Holmlund, T., Lindberg, M. J., Grander, D. & Wallberg, A. E. GCN5 acetylates and regulates the stability of the oncoprotein E2A-PBX1 in acute lymphoblastic leukemia. *Leukemia* **27**, 578-585, doi:10.1038/leu.2012.265 (2013).

17 Ionov, Y., Matsui, S. & Cowell, J. K. A role for p300/CREB binding protein genes in promoting cancer progression in colon cancer cell lines with microsatellite instability. *Proc Natl Acad Sci U S A* **101**, 1273-1278, doi:10.1073/pnas.0307276101 (2004).
18 Bienz, M. & Clevers, H. Linking colorectal cancer to Wnt signaling. *Cell* **103**, 311-320 (2000).
19 Brenachot, X. *et al.* The histone acetyltransferase MOF activates hypothalamic polysialylation to prevent diet-induced obesity in mice. *Mol Metab* **3**, 619-629, doi:10.1016/j.molmet.2014.05.006 (2014).
20 Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nature chemical biology* **10**, 365-370, doi:10.1038/nchembio.1497 (2014).
21 Olsen, C. A. Expansion of the lysine acylation landscape. *Angew Chem Int Ed Engl* **51**, 3755-3756, doi:10.1002/anie.201200316 (2012).
22 Zhang, Z. *et al.* Identification of lysine succinylation as a new post-translational modification. *Nature chemical biology* **7**, 58-63, doi:10.1038/nchembio.495 (2011).
23 Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-1028, doi:10.1016/j.cell.2011.08.008 (2011).
24 Goudarzi, A. *et al.* Dynamic Competing Histone H4 K5K8 Acetylation and Butyrylation Are Hallmarks of Highly Active Gene Promoters. *Molecular cell* **62**, 169-180, doi:10.1016/j.molcel.2016.03.014 (2016).
25 Sabari, B. R. *et al.* Intracellular crotonyl-CoA stimulates transcription through p300-catalyzed histone crotonylation. *Molecular cell* **58**, 203-215, doi:10.1016/j.molcel.2015.02.029 (2015).
26 Ringel, A. E. & Wolberger, C. Structural basis for acyl-group discrimination by human Gcn5L2. *Acta Crystallogr D Struct Biol* **72**, 841-848, doi:10.1107/S2059798316007907 (2016).
27 Leemhuis, H., Packman, L. C., Nightingale, K. P. & Hollfelder, F. The human histone acetyltransferase P/CAF is a promiscuous histone propionyltransferase. *ChemBioChem* **9**, 499-503, doi:DOI 10.1002/cbic.200700556 (2008).
28 Chen, Y. *et al.* Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol Cell Proteomics* **6**, 812-819, doi:10.1074/mcp.M700021-MCP200 (2007).
29 Liu, X. *et al.* MOF as an evolutionarily conserved histone crotonyltransferase and transcriptional activation by histone acetyltransferase-deficient and crotonyltransferase-competent CBP/p300. *Cell Discov* **3**, 17016, doi:10.1038/celldisc.2017.16 (2017).
30 Kaczmarska, Z. *et al.* Structure of p300 in complex with acyl-CoA variants. *Nature chemical biology* **13**, 21-29, doi:10.1038/nchembio.2217 (2017).
31 Caron, C., Boyault, C. & Khochbin, S. Regulatory cross-talk between lysine acetylation and ubiquitination: role in the control of protein stability. *BioEssays : news and reviews in molecular, cellular and developmental biology* **27**, 408-415, doi:10.1002/bies.20210 (2005).
32 Kleff, S., Andrulis, E. D., Anderson, C. W. & Sternglanz, R. Identification of a gene encoding a yeast histone H4 acetyltransferase. *J Biol Chem* **270**, 24674-24677 (1995).
33 Eberharter, A. & Becker, P. B. Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep* **3**, 224-229, doi:10.1093/embo-reports/kvf053 (2002).

34 Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567, doi:10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2 (1999).
35 Yang, Y. Y., Ascano, J. M. & Hang, H. C. Bioorthogonal chemical reporters for monitoring protein acetylation. *Journal of the American Chemical Society* **132**, 3640-3641, doi:10.1021/ja908871t (2010).
36 Yang, Y. Y., Grammel, M. & Hang, H. C. Identification of lysine acetyltransferase p300 substrates using 4-pentynoyl-coenzyme A and bioorthogonal proteomics. *Bioorganic & medicinal chemistry letters* **21**, 4976-4979, doi:10.1016/j.bmcl.2011.05.060 (2011).
37 Glozak, M. A., Sengupta, N., Zhang, X. & Seto, E. Acetylation and deacetylation of non-histone proteins. *Gene* **363**, 15-23, doi:10.1016/j.gene.2005.09.010 (2005).
38 Hasan, S. *et al.* Regulation of human flap endonuclease-1 activity by acetylation through the transcriptional coactivator p300. *Mol. Cell* **7**, 1221-1231 (2001).
39 Kikuchi, H., Takami, Y. & Nakayama, T. GCN5: a supervisor in all-inclusive control of vertebrate cell cycle progression through transcription regulation of various cell cycle-related genes. *Gene* **347**, 83-97, doi:10.1016/j.gene.2004.12.007 (2005).
40 Giles, R. H., Peters, D. J. & Breuning, M. H. Conjunction dysfunction: CBP/p300 in human disease. *Trends Genet.* **14**, 178-183 (1998).
41 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. *Journal of the American Chemical Society* **128**, 15356-15357, doi:10.1021/ja066298w (2006).
42 Yang, C. *et al.* Labeling lysine acetyltransferase substrates with engineered enzymes and functionalized cofactor surrogates. *J Am Chem Soc* **135**, 7791-7794, doi:10.1021/ja311636b (2013).
43 Gao, T., Yang, C. & Zheng, Y. G. Comparative studies of thiol-sensitive fluorogenic probes for HAT assays. *Anal Bioanal Chem* **405**, 1361-1371, doi:10.1007/s00216-012-6522-5 (2013).
44 Bao, X., Zhao, Q., Yang, T., Fung, Y. M. & Li, X. D. A chemical probe for lysine malonylation. *Angewandte Chemie* **52**, 4883-4886, doi:10.1002/anie.201300252 (2013).
45 Huang, H., Lin, S., Garcia, B. A. & Zhao, Y. Quantitative proteomic analysis of histone modifications. *Chem Rev* **115**, 2376-2418, doi:10.1021/cr500491u (2015).
46 Kuo, Y. M. & Andrews, A. J. Quantitating the specificity and selectivity of Gcn5-mediated acetylation of histone H3. *PLoS One* **8**, e54896, doi:10.1371/journal.pone.0054896 (2013).
47 Henry, R. A., Kuo, Y. M. & Andrews, A. J. Differences in Specificity and Selectivity Between CBP and p300 Acetylation of Histone H3 and H3/H4. *Biochemistry* **52**, 5746-5759, doi:Doi 10.1021/Bi400684q (2013).
48 Ito, T., Ikehara, T., Nakagawa, T., Kraus, W. L. & Muramatsu, M. p300-mediated acetylation facilitates the transfer of histone H2A-H2B dimers from nucleosomes to a histone chaperone. *Genes Dev.* **14**, 1899-1907 (2000).
49 Parada, C. A. & Roeder, R. G. A novel RNA polymerase II-containing complex potentiates Tat-enhanced HIV-1 transcription. *EMBO J* **18**, 3688-3701, doi:10.1093/emboj/18.13.3688 (1999).

50 Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834-840, doi:1175371 [pii] 10.1126/science.1175371 (2009).
51 Conacci-Sorrell, M., Ngouenet, C. & Eisenman, R. N. Myc-nick: a cytoplasmic cleavage product of Myc that promotes alpha-tubulin acetylation and cell differentiation. *Cell* **142**, 480-493, doi:10.1016/j.cell.2010.06.037 (2010).
52 Grant, P. A. *et al.* Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes & development* **11**, 1640-1650 (1997).
53 Jeffy, B. D. *et al.* An estrogen receptor-alpha/p300 complex activates the BRCA-1 promoter at an AP-1 site that binds Jun/Fos transcription factors: repressive effects of p53 on BRCA-1 transcription. *Neoplasia* **7**, 873-882 (2005).
54 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).
55 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).
56 Dancy, B. M. & Cole, P. A. Protein lysine acetylation by p300/CBP. *Chem Rev* **115**, 2419-2452, doi:10.1021/cr500452k (2015).
57 Hong, R. & Chakravarti, D. The human proliferating Cell nuclear antigen regulates transcriptional coactivator p300 activity and promotes transcriptional repression. *J. Biol. Chem.* **278**, 44505-44513 (2003).
58 Hassa, P. O. *et al.* Acetylation of poly(ADP-ribose) polymerase-1 by p300/CREB-binding protein regulates coactivation of NF-kappaB-dependent transcription. *The Journal of biological chemistry* **280**, 40450-40464, doi:10.1074/jbc.M507553200 (2005).
59 Banerjee, S., Rakshit, T., Sett, S. & Mukhopadhyay, R. Interactions of Histone Acetyltransferase p300 with the Nuclear Proteins Histone and HMGB1, As Revealed by Single Molecule Atomic Force Spectroscopy. *J Phys Chem B* **119**, 13278-13287, doi:10.1021/acs.jpcb.5b07795 (2015).
60 Enserink, J. M. & Kolodner, R. D. An overview of Cdk1-controlled targets and processes. *Cell division* **5**, 11, doi:10.1186/1747-1028-5-11 (2010).
61 Zhao, S. *et al.* Regulation of cellular metabolism by protein lysine acetylation. *Science* **327**, 1000-1004, doi:10.1126/science.1179689 (2010).
62 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).
63 Will, C. L. & Luhrmann, R. Spliceosome structure and function. *Cold Spring Harbor perspectives in biology* **3**, doi:10.1101/cshperspect.a003707 (2011).
64 Korobeinikova, A. V., Garber, M. B. & Gongadze, G. M. Ribosomal proteins: structure, function, and evolution. *Biochemistry (Mosc)* **77**, 562-574, doi:10.1134/S0006297912060028 (2012).
65 Gunderson, F. Q. & Johnson, T. L. Acetylation by the transcriptional coactivator Gcn5 plays a novel role in co-transcriptional spliceosome assembly. *PLoS Genet* **5**, e1000682, doi:10.1371/journal.pgen.1000682 (2009).

66 Garcia-Sastre, A., Muster, T., Barclay, W. S., Percy, N. & Palese, P. Use of a mammalian internal ribosomal entry site element for expression of a foreign protein by a transfectant influenza virus. *J Virol* **68**, 6254-6261 (1994).
67 Chen, W. & Helenius, A. Role of ribosome and translocon complex during folding of influenza hemagglutinin in the endoplasmic reticulum of living cells. *Molecular biology of the cell* **11**, 765-772 (2000).
68 Choi, Y. B., Ko, J. K. & Shin, J. The transcriptional corepressor, PELP1, recruits HDAC2 and masks histones using two separate domains. *The Journal of biological chemistry* **279**, 50930-50941, doi:10.1074/jbc.M406831200 (2004).
69 Fermento, M. E. *et al.* Inhibition of p300 suppresses growth of breast cancer. Role of p300 subcellular localization. *Exp Mol Pathol* **97**, 411-424, doi:10.1016/j.yexmp.2014.09.019 (2014).
70 Horwitz, K. B. & McGuire, W. L. Estrogen control of progesterone receptor in human breast cancer. Correlation with nuclear processing of estrogen receptor. *The Journal of biological chemistry* **253**, 2223-2228 (1978).
71 Gururaj, A. E., Peng, S., Vadlamudi, R. K. & Kumar, R. Estrogen induces expression of BCAS3, a novel estrogen receptor-alpha coactivator, through proline-, glutamic acid-, and leucine-rich protein-1 (PELP1). *Mol Endocrinol* **21**, 1847-1860, doi:10.1210/me.2006-0514 (2007).
72 Walter, P. & Blobel, G. Subcellular distribution of signal recognition particle and 7SL-RNA determined with polypeptide-specific antibodies and complementary DNA probe. *J Cell Biol* **97**, 1693-1699 (1983).
73 Pike, J. W. & Meyer, M. B. The vitamin D receptor: new paradigms for the regulation of gene expression by 1,25-dihydroxyvitamin D3. *Rheumatic diseases clinics of North America* **38**, 13-27, doi:10.1016/j.rdc.2012.03.004 (2012).
74 Germain, R. N. MHC-dependent antigen processing and peptide presentation: providing ligands for T lymphocyte activation. *Cell* **76**, 287-299 (1994).
75 Wijeyesakere, S. J., Gagnon, J. K., Arora, K., Brooks, C. L., 3rd & Raghavan, M. Regulation of calreticulin-major histocompatibility complex (MHC) class I interactions by ATP. *Proceedings of the National Academy of Sciences of the United States of America* **112**, E5608-5617, doi:10.1073/pnas.1510132112 (2015).
76 Antoniou, A. N., Blackwood, S. L., Mazzeo, D. & Watts, C. Control of antigen presentation by a single protease cleavage site. *Immunity* **12**, 391-398 (2000).
77 Murshid, A., Gong, J. & Calderwood, S. K. The role of heat shock proteins in antigen cross presentation. *Frontiers in immunology* **3**, 63, doi:10.3389/fimmu.2012.00063 (2012).
78 Yang, C., Wu, J. & Zheng, Y. G. Function of the active site lysine autoacetylation in Tip60 catalysis. *PloS one* **7**, e32886, doi:10.1371/journal.pone.0032886 (2012).
79 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 and drug development technologies* **13**, 210-220, doi:10.1089/adt.2015.636 (2015).
80 Lu, X. & Zhu, H. Tube-gel digestion: a novel proteomic approach for high throughput analysis of membrane proteins. *Molecular & cellular proteomics : MCP* **4**, 1948-1958, doi:10.1074/mcp.M500138-MCP200 (2005).

81 Lerin, C. *et al.* GCN5 acetyltransferase complex controls glucose metabolism through transcriptional repression of PGC-1alpha. *Cell Metab* **3**, 429-438, doi:10.1016/j.cmet.2006.04.013 (2006).
82 Ho, Y. D., Joyal, J. L., Li, Z. & Sacks, D. B. IQGAP1 integrates Ca2+/calmodulin and Cdc42 signaling. *The Journal of biological chemistry* **274**, 464-470 (1999).
83 Kim, S. T., Xu, B. & Kastan, M. B. Involvement of the cohesin protein, Smc1, in Atm-dependent and independent responses to DNA damage. *Genes & development* **16**, 560-570, doi:10.1101/gad.970602 (2002).
84 Roth, S. Y., Denu, J. M. & Allis, C. D. Histone acetyltransferases. *Annual review of biochemistry* **70**, 81-120 (2001).
85 Yang, X. J. The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. *Nucleic Acids Res* **32**, 959-976 (2004).
86 Avvakumov, N. & Cote, J. The MYST family of histone acetyltransferases and their intimate links to cancer. *Oncogene* **26**, 5395-5407 (2007).
87 Aka, J. A., Kim, G. W. & Yang, X. J. K-acetylation and its enzymes: overview and new developments. *Handb Exp Pharmacol* **206**, 1-12, doi:10.1007/978-3-642-21631-2_1 (2011).
88 Parthun, M. R. Histone acetyltransferase 1: More than just an enzyme? *Biochim Biophys Acta* **1819**, 256-263, doi:10.1016/j.bbagrm.2011.07.006 (2012).
89 Sterner, D. E. & Berger, S. L. Acetylation of histones and transcription-related factors. *Microbiology and molecular biology reviews* **64**, 435-459 (2000).
90 Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
91 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* **41**, 185-198, doi:S1357-2725(08)00347-6 [pii] 10.1016/j.biocel.2008.08.027 (2009).
92 Zhang, J. *et al.* Lysine acetylation is a highly abundant and evolutionarily conserved modification in Escherichia coli. *Mol Cell Proteomics* **8**, 215-225, doi:10.1074/mcp.M800187-MCP200 (2009).
93 Li, T. *et al.* Characterization and prediction of lysine (k)-acetyl-transferase specific acetylation sites. *Mol Cell Proteomics* **11**, M111.011080, doi:10.1074/mcp.M111.011080 (2012).
94 Ghizzoni, M., Haisma, H. J., Maarsingh, H. & Dekker, F. J. Histone acetyltransferases are crucial regulators in NF-kappaB mediated inflammation. *Drug discovery today* **16**, 504-511, doi:10.1016/j.drudis.2011.03.009 (2011).
95 Kadiyala, C. S. *et al.* Acetylation of retinal histones in diabetes increases inflammatory proteins: effects of minocycline and manipulation of histone acetyltransferase (HAT) and histone deacetylase (HDAC). *J Biol Chem* **287**, 25869-25880, doi:10.1074/jbc.M112.375204 (2012).
96 Iyer, A., Fairlie, D. P. & Brown, L. Lysine acetylation in obesity, diabetes and metabolic disease. *Immunology and cell biology* **90**, 39-46, doi:10.1038/icb.2011.99 (2012).
97 Valor, L. M., Viosca, J., Lopez-Atalaya, J. P. & Barco, A. Lysine Acetyltransferases CBP and p300 as Therapeutic Targets in Cognitive and Neurodegenerative Disorders. *Curr Pharm Design* **19**, 5051-5064 (2013).

98 Deguchi, K. *et al.* MOZ-TIF2-induced acute myeloid leukemia requires the MOZ nucleosome binding motif and TIF2-mediated recruitment of CBP. *Cancer Cell* **3**, 259-271 (2003).
99 Perez-Campo, F. M., Costa, G., Lie-a-Ling, M., Kouskoff, V. & Lacaud, G. The MYSTerious MOZ, a histone acetyltransferase with a key role in haematopoiesis. *Immunology* **139**, 161-165, doi:10.1111/imm.12072 (2013).
100 Pons, D. *et al.* Epigenetic histone acetylation modifiers in vascular remodelling: new targets for therapy in cardiovascular disease. *Eur Heart J* **30**, 266-277, doi:ehn603 [pii] 10.1093/eurheartj/ehn603 (2009).
101 Barry, S. P., Davidson, S. M. & Townsend, P. A. Molecular regulation of cardiac hypertrophy. *Int J Biochem Cell Biol* **40**, 2023-2039, doi:S1357-2725(08)00102-7 [pii] 10.1016/j.biocel.2008.02.020 (2008).
102 Col, E. *et al.* HIV-1 Tat targets Tip60 to impair the apoptotic cell response to genotoxic stresses. *Embo J* **24**, 2634-2645 (2005).
103 Aherne, G. W., Rowlands, M. G., Stimson, L. & Workman, P. Assays for the identification and evaluation of histone acetyltransferase inhibitors. *Methods* **26**, 245-253 (2002).
104 Berndsen, C. E. & Denu, J. M. Assays for mechanistic investigations of protein/histone acetyltransferases. *Methods* **36**, 321-331 (2005).
105 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 and drug development technologies* **13**, 210-220, doi:10.1089/adt.2015.636 (2015).
106 Furdas, S. D. *et al.* Synthesis and biological testing of novel pyridoisothiazolones as histone acetyltransferase inhibitors. *Bioorg Med Chem* **19**, 3678-3689, doi:10.1016/j.bmc.2011.01.063 (2011).
107 Ghadiali, J. E., Lowe, S. B. & Stevens, M. M. Quantum-dot-based FRET detection of histone acetyltransferase activity. *Angewandte Chemie* **50**, 3417-3420, doi:10.1002/anie.201008263 (2011).
108 Sharma, G. G. *et al.* MOF and histone H4 acetylation at lysine 16 are critical for DNA damage response and double-strand break repair. *Molecular and cellular biology* **30**, 3582-3595, doi:10.1128/MCB.01476-09 (2010).
109 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* **41**, 2596-2599, doi:10.1002/1521-3773(20020715)41:14<2596::AID-ANIE2596>3.0.CO;2-4 (2002).
110 Agard, N. J., Prescher, J. A. & Bertozzi, C. R. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. *J Am Chem Soc* **126**, 15046-15047, doi:10.1021/ja044996f (2004).
111 Ghizzoni, M. *et al.* 6-alkylsalicylates are selective Tip60 inhibitors and target the acetyl-CoA binding site. *Eur J Med Chem* **47**, 337-344, doi:10.1016/j.ejmech.2011.11.001 (2012).
112 Balasubramanyam, K., Swaminathan, V., Ranganathan, A. & Kundu, T. K. Small molecule modulators of histone acetyltransferase p300. *J Biol Chem* **278**, 19134-19140, doi:10.1074/jbc.M301580200 (2003).

113 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* **287**, 319-328 (2000).
114 Thinon, E. & Hang, H. C. Chemical reporters for exploring protein acylation. *Biochem Soc Trans* **43**, 253-261, doi:10.1042/BST20150004 (2015).
115 Wu, J. & Zheng, Y. G. Fluorescent reporters of the histone acetyltransferase. *Anal Biochem* **380**, 106-110, doi:10.1016/j.ab.2008.05.030 (2008).
116 Grevengoed, T. J., Klett, E. L. & Coleman, R. A. Acyl-CoA metabolism and partitioning. *Annu Rev Nutr* **34**, 1-30, doi:10.1146/annurev-nutr-071813-105541 (2014).
117 Nochi, Z., Olsen, R. K. J. & Gregersen, N. Short-chain acyl-CoA dehydrogenase deficiency: from gene to cell pathology and possible disease mechanisms. *J. Inherit. Metab. Dis.*, doi:10.1007/s10545-017-0047-1 (2017).
118 Scholl-Burgi, S., Sass, J. O., Zschocke, J. & Karall, D. Amino acid metabolism in patients with propionic acidaemia. *J. Inherit. Metab. Dis.* **35**, 65-70, doi:10.1007/s10545-010-9245-9 (2012).
119 Krautkramer, K. A. *et al.* Diet-Microbiota Interactions Mediate Global Epigenetic Programming in Multiple Host Tissues. *Mol. Cell* **64**, 982-992, doi:10.1016/j.molcel.2016.10.025 (2016).
120 Woo, V. & Alenghat, T. Host-microbiota interactions: epigenomic regulation. *Curr. Opin. Immunol.* **44**, 52-60, doi:10.1016/j.coi.2016.12.001 (2017).
121 Avvakumov, N. & Cote, J. Functions of myst family histone acetyltransferases and their link to disease. *Subcell. Biochem.* **41**, 295-317 (2007).
122 Bodai, L., Pallos, J., Thompson, L. M. & Marsh, J. L. Pcaf modulates polyglutamine pathology in a Drosophila model of Huntington's disease. *Neurodegener Dis* **9**, 104-106, doi:10.1159/000330505 (2012).
123 Cheng, Z. *et al.* Molecular characterization of propionyllysines in non-histone proteins. *Mol Cell Proteomics* **8**, 45-52, doi:10.1074/mcp.M800224-MCP200 (2009).
124 Marmorstein, R. & Trievel, R. C. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim. Biophys. Acta* **1789**, 58-68, doi:S1874-9399(08)00157-0 [pii] 10.1016/j.bbagrm.2008.07.009 (2009).
125 Xie, Z. *et al.* Lysine succinylation and lysine malonylation in histones. *Mol. Cell. Proteomics* **11**, 100-107, doi:10.1074/mcp.M111.015875 (2012).
126 Klein, B. J. *et al.* Recognition of Histone H3K14 Acylation by MORF. *Structure* **25**, 650-654 e652, doi:10.1016/j.str.2017.02.003 (2017).
127 Zhang, K., Chen, Y., Zhang, Z. & Zhao, Y. Identification and verification of lysine propionylation and butyrylation in yeast core histones using PTMap software. *J. Proteome Res.* **8**, 900-906, doi:10.1021/pr8005155 (2009).
128 Chen, Z. *et al.* The histone acetylranseferase hMOF acetylates Nrf2 and regulates anti-drug responses in human non-small cell lung cancer. *Br J Pharmacol* **171**, 3196-3211, doi:10.1111/bph.12661 (2014).
129 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*, doi:10.1021/acschembio.7b00114 (2017).
130 Yang, C., Wu, J., Sinha, S. H., Neveu, J. M. & Zheng, Y. G. Autoacetylation of the MYST Lysine Acetyltransferase MOF Protein. *J. Biol. Chem.* **287**, 34917-34926, doi:10.1074/jbc.M112.359356 (2012).

131 Buscaino, A. *et al.* MOF-regulated acetylation of MSL-3 in the Drosophila dosage compensation complex. *Molecular cell* **11**, 1265-1277 (2003).
132 Zhou, Y. *et al.* Reversible acetylation of the chromatin remodelling complex NoRC is required for non-coding RNA-dependent silencing. *Nat Cell Biol* **11**, 1010-1016, doi:10.1038/ncb1914 (2009).
133 Taipale, M. *et al.* hMOF histone acetyltransferase is required for histone H4 lysine 16 acetylation in mammalian cells. *Mol Cell Biol* **25**, 6798-6810, doi:10.1128/MCB.25.15.6798-6810.2005 (2005).
134 Sun, H., Yi Chen, Q. & Costa, M. Structure and function of histone acetyltransferase MOF. *AIMS Biophysics* **2**, 555-569, doi:10.3934/biophy.2015.4.555 (2015).
135 Sun, M. *et al.* Characterization of Protein Lysine Propionylation in Escherichia coli: Global Profiling, Dynamic Change, and Enzymatic Regulation. *J Proteome Res* **15**, 4696-4708, doi:10.1021/acs.jproteome.6b00798 (2016).
136 Cai, Y. *et al.* Subunit composition and substrate specificity of a MOF-containing histone acetyltransferase distinct from the male-specific lethal (MSL) complex. *J Biol Chem* **285**, 4268-4272, doi:10.1074/jbc.C109.087981 (2010).
137 Smith, E. R. *et al.* A human protein complex homologous to the Drosophila MSL complex is responsible for the majority of histone H4 acetylation at lysine 16. *Mol Cell Biol* **25**, 9175-9188, doi:10.1128/MCB.25.21.9175-9188.2005 (2005).
138 Memarzadeh, S. *et al.* Role of autonomous androgen receptor signaling in prostate cancer initiation is dichotomous and depends on the oncogenic signal. *Proc Natl Acad Sci U S A* **108**, 7962-7967, doi:10.1073/pnas.1105243108 (2011).
139 Kim, S. C. *et al.* A clean, more efficient method for in-solution digestion of protein mixtures without detergent or urea. *J Proteome Res* **5**, 3446-3452, doi:10.1021/pr0603396 (2006).
140 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).
141 Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* **53**, 240-255, doi:10.1107/S0907444996012255 (1997).
142 Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126-2132, doi:10.1107/S0907444904019158 (2004).
143 Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66**, 12-21, doi:10.1107/S0907444909042073 (2010).
144 Tan, M. *et al.* Lysine glutarylation is a protein posttranslational modification regulated by SIRT5. *Cell Metab* **19**, 605-617, doi:10.1016/j.cmet.2014.03.014 (2014).
145 Hirschey, M. D. & Zhao, Y. Metabolic Regulation by Lysine Malonylation, Succinylation, and Glutarylation. *Mol Cell Proteomics* **14**, 2308-2315, doi:10.1074/mcp.R114.046664 (2015).
146 Kadiyala, C. S. *et al.* Acetylation of retinal histones in diabetes increases inflammatory proteins: effects of minocycline and manipulation of histone acetyltransferase (HAT) and histone deacetylase (HDAC). *J Biol Chem* **287**, 25869-25880, doi:10.1074/jbc.M112.375204 (2012).

147 Ghizzoni, M., Haisma, H. J., Maarsingh, H. & Dekker, F. J. Histone acetyltransferases are crucial regulators in NF-kappaB mediated inflammation. *Drug Discov Today* **16**, 504-511, doi:10.1016/j.drudis.2011.03.009 (2011).
148 Schneider, A. *et al.* Acetyltransferases (HATs) as Targets for Neurological Therapeutics. *Neurotherapeutics* **10**, 568-588, doi:10.1007/s13311-013-0204-7 (2013).
149 Ott, M. & Verdin, E. HAT trick: p300, small molecule, inhibitor. *Chem Biol* **17**, 417-418, doi:10.1016/j.chembiol.2010.05.002 (2010).
150 Bao, X. C. *et al.* Identification of 'erasers' for lysine crotonylated histone marks using a chemical proteomics approach. *Elife* **3**, doi:ARTN e02999 10.7554/eLife.02999 (2014).
151 Oglesbee, D. *et al.* Development of a newborn screening follow-up algorithm for the diagnosis of isobutyryl-CoA dehydrogenase deficiency. *Genet Med* **9**, 108-116, doi:10.1097GIM.0b013e31802f78d6 (2007).
152 Robinson, W. G., Nagle, R., Bachhawat, B. K., Kupiecki, F. P. & Coon, M. J. Coenzyme-a Thiol Esters of Isobutyric, Methacrylic, and Beta-Hydroxyisobutyric Acids as Intermediates in the Enzymatic Degradation of Valine. *Journal of Biological Chemistry* **224**, 1-11 (1957).
153 Jost, M., Born, D. A., Cracan, V., Banerjee, R. & Drennan, C. L. Structural Basis for Substrate Specificity in Adenosylcobalamin-dependent Isobutyryl-CoA Mutase and Related Acyl-CoA Mutases. *J Biol Chem* **290**, 26882-26898, doi:10.1074/jbc.M115.676890 (2015).
154 Santra, S., Macdonald, A., Preece, M. A., Olsen, R. K. & Andresen, B. S. Long-term outcome of isobutyryl-CoA dehydrogenase deficiency diagnosed following an episode of ketotic hypoglycaemia. *Mol Genet Metab Rep* **10**, 28-30, doi:10.1016/j.ymgmr.2016.11.005 (2017).
155 Yun, J. W. *et al.* A novel ACAD8 mutation in asymptomatic patients with isobutyryl-CoA dehydrogenase deficiency and a review of the ACAD8 mutation spectrum. *Clin Genet* **87**, 196-198, doi:10.1111/cge.12350 (2015).
156 Roe, C. R. *et al.* Isolated isobutyryl-CoA dehydrogenase deficiency: an unrecognized defect in human valine metabolism. *Molecular genetics and metabolism* **65**, 264-271, doi:10.1006/mgme.1998.2758 (1998).
157 Han, Z., Luan, Y. & Zheng, Y. G. Integration of Bioorthogonal Probes and Q-FRET for the Detection of Histone Acetyltransferase Activity. *Chembiochem* **16**, 2605-2609, doi:10.1002/cbic.201500427 (2015).
158 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* **12**, 1547-1555, doi:10.1021/acschembio.7b00114 (2017).
159 Mirzaei, H., Suarez, J. A. & Longo, V. D. Protein and amino acid restriction, aging and disease: from yeast to humans. *Trends Endocrinol Metab* **25**, 558-566, doi:10.1016/j.tem.2014.07.002 (2014).
160 Mirisola, M. G. *et al.* Serine- and threonine/valine-dependent activation of PDK and Tor orthologs converge on Sch9 to promote aging. *PLoS Genet* **10**, e1004113, doi:10.1371/journal.pgen.1004113 (2014).

161 Kingsbury, J. M., Sen, N. D. & Cardenas, M. E. Branched-Chain Aminotransferases Control TORC1 Signaling in Saccharomyces cerevisiae. *PLoS Genet* **11**, e1005714, doi:10.1371/journal.pgen.1005714 (2015).
162 Mansfeld, J. *et al.* Branched-chain amino acid catabolism is a conserved regulator of physiological ageing. *Nat Commun* **6**, doi:ARTN 10043 10.1038/ncomms10043 (2015).
163 Eisenberg, T. *et al.* Nucleocytosolic depletion of the energy metabolite acetyl-coenzyme a stimulates autophagy and prolongs lifespan. *Cell Metab* **19**, 431-444, doi:10.1016/j.cmet.2014.02.010 (2014).
164 Hu, J. *et al.* Tor-Sch9 deficiency activates catabolism of the ketone body-like acetic acid to promote trehalose accumulation and longevity. *Aging Cell* **13**, 457-467, doi:10.1111/acel.12202 (2014).
165 Abdinejad, A., Fisher, A. M. & Kumar, S. Production and utilization of butyryl-CoA by fatty acid synthetase from mammalian tissues. *Arch Biochem Biophys* **208**, 135-145 (1981).
166 Webster, L. T., Jr., Gerowin, L. D. & Rakita, L. Purification and Characteristics of a Butyryl Coenzyme a Synthetase from Bovine Heart Mitochondria. *J Biol Chem* **240**, 29-33 (1965).
167 Faergeman, N. J. & Knudsen, J. Role of long-chain fatty acyl-CoA esters in the regulation of metabolism and in cell signalling. *Biochem J* **323 (Pt 1)**, 1-12 (1997).
168 Sharma, G. G. *et al.* MOF and histone H4 acetylation at lysine 16 are critical for DNA damage response and double-strand break repair. *Mol Cell Biol* **30**, 3582-3595, doi:10.1128/MCB.01476-09 (2010).
169 Flynn, E. M. *et al.* A Subset of Human Bromodomains Recognizes Butyryllysine and Crotonyllysine Histone Peptide Modifications. *Structure* **23**, 1801-1814, doi:10.1016/j.str.2015.08.004 (2015).
170 Han, J. *et al.* Multiple propionyl coenzyme A-supplying pathways for production of the bioplastic poly(3-hydroxybutyrate-co-3-hydroxyvalerate) in Haloferax mediterranei. *Appl Environ Microbiol* **79**, 2922-2931, doi:10.1128/AEM.03915-12 (2013).
171 Su, J., Wang, F., Cai, Y. & Jin, J. The Functional Analysis of Histone Acetyltransferase MOF in Tumorigenesis. *Int J Mol Sci* **17**, doi:10.3390/ijms17010099 (2016).