HHS Public Access

Author manuscript

ACS Chem Biol. Author manuscript; available in PMC 2015 October 19.

Published in final edited form as:

ACS Chem Biol. 2015 September 18; 10(9): 2034–2047. doi:10.1021/acschembio.5b00342.

Non-enzymatic protein acetylation detected by NAPPA protein arrays*

Adam S. Olia[1,2], **Kristi Barker**[3], **Cheryl E. McCullough**[1,2], **Hsin-Yao Tang**[4], **David W. Speicher**[4], **Ji Qiu**[3,*], **Joshua LaBaer**[3,*], and **Ronen Marmorstein**[1,2,*]

[1]Department of Biochemistry & Biophysics, Abramson Family Cancer Research Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, Pennsylvania

[2]Program in Gene Expression and Regulation, The Wistar Institute, Philadelphia, Pennsylvania, United States of America

[3]Virginia G. Piper Center for Personalized Diagnostics, Biodesign Institute, Arizona State University, Tempe, Arizona, United States of America

[4]Program in Molecular and Cellular Oncogenesis, The Wistar Institute, Philadelphia, Pennsylvania, United States of America

Abstract

Acetylation is a post-translational modification that occurs on thousands of proteins located in many cellular organelles. This process mediates many protein functions and modulates diverse biological processes. In mammalian cells, where acetyl-CoA is the primary acetyl donor, acetylation in the mitochondria is thought to occur by chemical means due to the relatively high concentration of acetyl-CoA located in this organelle. In contrast, acetylation outside of the mitochondria is thought to be mediated predominantly by acetyltransferase enzymes. Here we address the possibility that non-enzymatic chemical acetylation outside of the mitochondria may be more common than previously appreciated. We employed the Nucleic Acid Programmable Protein Array platform to perform an unbiased screen for human proteins that undergo chemical acetylation, which resulted in the identification of a multitude of proteins with diverse functions and cellular localization. Mass spectrometry analysis revealed that basic residues typically precede the acetylated lysine in the −7 to −3 position, and we show by mutagenesis that these basic residues contribute to chemical acetylation capacity. We propose that these basic residues lower the pKa of the substrate lysine for efficient chemical acetylation. Many of the identified proteins reside outside of the mitochondria, and have been previously demonstrated to be acetylated *in vivo*. As such, our studies demonstrate that chemical acetylation occurs more broadly throughout the eukaryotic cell than previously appreciated, and suggests that this post-translational protein modification may have more diverse roles in protein function and pathway regulation.

*Correspondence should be addressed to Ronen Marmorstein: marmor@mail.med.upenn.edu; Joshua LaBaer: joshua.labaer@asu.edu or Ji Qiu: ji.qiu@asu.edu.

The authors declare no conflict of interest.

INTRODUCTION

Post-translational modifications (PTMs) of proteins provide the cell an additional level of regulation. These PTMs include phosphorylation, ubiquitination, glycosylation, and acetylation, as well as a vast assortment of other less studied or less abundant modifications. Along with the varied nature of the modifications, the effects of the PTMs can influence nearly every aspect of protein function, including stability, localization, binding affinities, and enzymatic activity; and biological processes such as transcription and metabolism.

The acetylation of the ε-nitrogen of the lysine side chain has long been studied as primarily a histone modification. In general, the action of any lysine acetyltransferase enzyme is to mediate the transfer of an acetyl group from the acetyl-CoA cofactor onto the lysine residue, while deacetylase enzymes work in opposition to this function. The classical view for this modification was that that the neutralization of the positively charged lysine side chain on the histone tails by acetylation facilitates DNA unwinding from the nucleosome, allowing for transcription and other DNA-templated activities. Currently a much broader role for these modifications has been appreciated since the discovery of histone PTMs. It has been demonstrated that these modifications do not directly contact the DNA, but rather specifically influence the binding and activity of other transcription factors including bromodomain-containing proteins that specifically recognize acetylated lysine residues.(1, 2)

The acetylation of non-histone proteins as a whole is less well understood, with a few exceptions. For example, the acetylation of the extreme C-terminal domain of the tumor suppressor p53 by the acetyltransferase p300 has been well documented, with the effect of increased transcription of p53 target genes.(3) Additionally, the acetylation by Tip60 of p53 on a distinct lysine residue regulates the apoptotic functions of p53.(4) Besides transcription factor acetylation, the microtubule protein α-tubulin has been shown to be acetylated, at least in part by the αTAT acetyltransferase.(5) This PTM increases the longevity of the microtubules, as well as facilitates vesicle trafficking.(6, 7) More recently, proteome wide mass spectrometry studies have shown that acetylation is a pervasive modification, with thousands of acetylated proteins located in many cellular organelles harboring multiple acetylation sites.(8, 9) These proteins are involved in nearly every cellular process, from transcription and translation, to cytoskeleton remodeling and protein transport.(9)

The presence of appropriate acetylation modifications is crucial to maintaining cellular function, and as such, a number of diseases and disorders have been shown to be correlated with aberrant acetylation. In the case of Tip60, misregulation of its target proteins has been implicated in the progression of Alzheimer's disease,(10) while p300 and CBP can act as tumor suppressor genes, with their dysfunction linked to numerous cancers.(11) In addition to their role in disease, acetylation may have bearing on the aging process, as it has been reported that activation of some Sirtuin family deacetylases may extend lifespan in model organisms.(12)

Given the large number of functionally important acetylation sites across various cellular proteins, it becomes imperative to understand the mode of protein acetylation. The mode of

protein acetylation which is most studied is mediated by protein acetyltransferase enzymes, which use the acetyl-coenzyme A cofactor to acetylate protein substrates.(9) In addition, some protein acetyltransferases have been reported to autoacetylate themselves such as the MYST family of HATs: hMOF and Esa1(13), PCAF(14), Rtt109(15)and p300(16). Another mechanism of protein acetylation involves a chemical event, which requires no enzyme. Non-enzymatic acetylation acetylation has been historically reported for both histones and hemoglobin, and this reaction occurs at modest acetyl-coenzyme A concentrations, and under biologically relevant buffer conditions.(17, 18) More recently, evidence has emerged for non-enzymatic acetylation on a number of bacterial proteins (19–21), as well as on eukaryotic proteins in the mitochondria (20, 22, 23). While these studies have detailed a large number of chemically acetylated proteins, bacteria contain millimolar concentrations of acetyl-phosphate which can be used as the acetate donor,(19, 21, 24) and a similar situation exists in the mitochondria of mammalian cells, with up to 2mM acetyl-CoA (20). In contrast, the cytoplasm of mammalian cells contains less than 30μM acetyl-CoA as the sole acetate donor (23), and the mechanism of acetylation in this compartment has not been extensively studied.

Recent advances in printed protein arrays have allowed for near-proteome wide biochemical analyses in a controllable way.(25) While protein arrays have been used previously for the study of protein acetylation, these have focused either on bacterial proteins, or on peptide libraries (21, 26). Here, we employed the Nucleic Acid Programmable Protein Array (NAPPA) platform (20, 22, 27) which circumvents challenges of expression and purification of thousands of proteins in yeast or insect cells and the potential denaturation incurred during purification, printing and storage as required by the conventional production of human protein arrays. Protein integrity is assured by the use of both mammalian expression machinery and chaperone proteins to synthesize and properly fold proteins immediately before assays. Using NAPPA (28), we have screened a large subset of the human proteome for their ability to undergo chemical acetylation, leading to the identification and characterization of a diverse group of chemically acetylated proteins, which contain a signature sequence preceding the chemically acetylated lysine residue. The implications of these findings are discussed.

RESULTS

Adaptation of NAPPA to Screen for Chemically Acetylated Proteins

For initial protocol development, a small set of positive and negative control genes were selected to print onto a test array. The positive control genes (p300, CBP, PCAF, hMOF, and GCN5) were acetyltransferase enzymes, which had been shown to exhibit at least some degree of auto-acetylation activity *in vitro*,(2, 12–14, 16) while the negative control genes (p53 and histone H4) were proteins that were demonstrated substrates of acetyltransferase enzymes.(29, 30) These genes were cloned into a FLAG-tagged *in vitro* translation vector. The plasmids were printed onto the array slide, alongside an α-FLAG antibody to immobilize each translated protein onto the array. Protein expression was initiated by the addition of a Hela cell based *in vitro* translation mixture. The expression mixture was supplemented with T7 polymerase, to transcribe the recombinant proteins, which were under

the control of a T7 promoter. After washing, the final result is multiple pure, immobilized proteins displayed on the array slide. The quality of the printed plasmid DNA templates was assessed using pico-green staining (Figure S1A), and the protein expression and display was gauged using a second α-FLAG antibody (Figure S1B).

To determine the acetylation capability of each protein on the array, the displayed proteins were incubated with ^{14}C labeled acetyl-coenzyme A (acetyl-CoA). Two methods were tested, by acetylating the protein during the translation reaction (Single-Step Method) or after the translation reaction was stopped. In the Single-Step method, ^{14}C-acetyl-CoA was simply added to the transcription/translation reaction. After a one month exposure to film, the ^{14}C incorporation was quantified by integration of the spot intensity on the film. In contrast, in the Multi-Step method, all attempts were made to start the radiolabeled acetylation reaction with a minimum of preexisting acetylation sites. First, the arrays were translated in the presence of the non-specific acetyltransferase enzyme inhibitor *s*-acetonyl-CoA (31), to inhibit any acetyltransferase enzymes which may be present in the translation mixture. Following expression, the arrays were fully deacetylated using high concentrations of the yeast histone deacetylase Hst2. Following this pretreatment, the arrays were then incubated with ^{14}C-acetyl-CoA and processed as described above. For both methods, in addition to arrays processed with full transcription/translation reaction mixtures, arrays were also processed lacking the T7 polymerase in the expression mixture, which should result in no recombinant protein production.

We observed that both methods could differentiate between the positive and negative control proteins (Figure 1). In both cases, hMOF and PCAF showed the highest acetylation signal, as these enzymes are shown to have high intracellular autoacetylation activity, (2, 32) and in the Single-Step method, p300 was also detected with a high signal, in line with its known autoacetylation activity.(16) Both methods did display some background signal in the negative genes as well as on the slides lacking the T7 polymerase, which was attributed to residual bacterial proteins contaminating the plasmid DNA. However, this background was consistent for each DNA template, and the signal to noise was more than sufficient in both methods to allow for distinguishing the positive and negative controls. There was a small distinction between the two methods in that the Single-Step yielded higher signal from the positives, albeit with a slightly higher background from the negative genes (Figure 1B and C); whereas the Multi-Step procedure resulted in a lower background, with lower signal from the positive controls (Figure 1D and E). Based on these results, we concluded that the arrays displayed sufficient protein levels to detect autoacetylation and selected the Single-Step procedure to test the full set of genes, as the increase in signal outweighed the higher background.

Use of NAPPA to Screen for Chemically Acetylated Proteins

Following protocol validation, arrays were produced with approximately 6,000 genes. The plasmid DNA for these genes were taken from the DNASU collection, and encoded a C-terminal FLAG or GST tag in an *in vitro* expression vector. The genes were split into four sets, resulting in about 1,500 genes printed per array slide. The plasmid DNA was co-printed with a corresponding α-GST or α-FLAG antibody for immobilization. The arrays were

processed using the Single-Step method, with the radiolabeled ^{14}C-acetyl-CoA added directly to the protein expression reaction. For each of the four array sets, three slides were produced with the full expression mixture, and two were produced with a translation mixture lacking the T7 polymerase. After processing and exposure, the films were scanned, and the intensity of each spot was integrated (Figure 2A).

By visual inspection of the films, there were often large areas on the film with an overall lower signal near the center of the glass slide. This was attributed to the surface tension of the reaction mixture creating a higher volume around the edge of the slide. To correct for this bias, a local area normalization was carried out on each slide, generating z-scores for each spot. Using this approach, approximately 300 genes had a z-score above 2, and approximately 1,600 genes had z-scores over 1, however there was no clear delineation to mark the cutoff between positive hits and the background, as the genes fell into a continuous distribution. (Figure 2B). Therefore, a secondary confirmation step was performed, and the top ranked 300 genes bases on the z-score were selected and re-arrayed in triplicate to produce a new confirmation array. These candidates were printed onto the array alongside positive and negative controls, as well as BSA and several other non-gene controls.

For the confirmation stage, the arrays were processed using both the Single-Step and Multi-Step procedures, to allow for a cross-validation of the techniques. A total of fourteen slides were used, with four being treated with the full expression mixture, and three lacking the T7 polymerase for each method (Figure 3). Following the routine described earlier, the arrays were processed, and the spot intensities quantified and normalized. Each spot was individually ranked according to the z-score as before, with the replicates being maintained separately. Many of the genes had a high z-score in at least two replicates, however this correlation began to drop off below a z-score of ~1. This allowed for a more rational cutoff for selecting hits to move on to further validation, and as such the final selection criteria was that at least two of the three spots had a z-score greater than 1, with the third spot having a positive value. This narrowed the list of positive hits to 31 proteins from the Single-Step method and 27 proteins from the Multi-Step method, with 11 proteins being common in the two sets, for a total of 47 unique candidate genes. In addition, 11 proteins were included which, by visual inspection, had noticeable and reproducible signal changes in the full array screen, but did not meet the initial selection criteria for confirmation array printing.

In Vitro Validation of Chemically Acetylated Proteins

The mechanism of the acetylation for the 58 candidate genes was investigated next, as the observed signal could have been due to protein acetyltransferase activity (autoacetylation), or direct chemical acetylation. To dissect this, the genes selected for validation were investigated for autoacetylation activity *in vitro* using recombinant purified protein. The genes' functions represented a wide variety of cellular processes. For example, RPS26 and RPS28 are ribosomal subunits, MRPL17 and MRPL40 are mitochondrial ribosomal proteins, UBE2B is an ubiquitinating enzyme, KRT8 is a keratin, IGL@ is an immunoglobulin, and COX5 is a subunit of the cytochrome c complex. While some of the candidate genes were well characterized and unlikely to represent novel acetyltransferase enzymes, all of the candidate genes were recloned, expressed in *E. coli*, and purified using a

GST affinity column. Of the proteins tested, 49 were produced at levels sufficient for acetylation studies *in vitro* (Table 1 and Figure S2). Prior to any acetylation studies, the candidate proteins were treated with high concentrations of the recombinant *S. cerevisiae* Hst2 protein deacetylase to remove any promiscuous acetylation events that may have occurred during the protein overexpression in *E. coli*. The purified proteins were then assayed for their ability to incorporate the radiolabeled acetate from ^{14}C-acetyl-CoA. After incubation with concentrations of ^{14}C-acetyl-CoA from 1μM to 20μM (the upper limit of the acetyl-CoA concentration that is thought to be present in growing eukaryotic cells)(33), the proteins were washed, and the radioactivity counted. (Figure 4A and Table S1). Remarkably, all proteins assayed had incorporation of the radioactive acetyl group. We noted that the level of protein acetylation was significant and showed a dose response in relation to increasing ^{14}C-acetyl-CoA concentrations (Figure 4A, left). However, when compared to the control acetyltransferase enzyme hMOF, the level of ^{14}C-acetyl incorporation in the candidate genes was one to two orders of magnitude lower (Figure 4A, right). In addition, the level of protein acetylation did not plateau at high acetyl-CoA concentrations, even with acetyl-CoA concentrations as high at 300μM (Figure 4B, left and Table S1). This is in contrast to a more defined plateau at high acetyl-CoA concentrations for the hMOF acetyltransferase (Figure 4B, right and Table S1). Since a second-order, enzymatically-driven reaction, at saturating substrate concentration, would be expected to show a plateau of activity, these data were consistent with the chemical acetylation of these protein substrates. To test this possibility, we treated hMOF and the candidate proteins with equimolar concentrations (20μM) of ^{14}C-acetyl-CoA and *s*-acetonyl-CoA, a non-hydrolyzable acetyl-CoA analog in which an extra carbon is inserted between the terminal sulfur of the CoA and the acetyl group, which functions as a non-specific inhibitor of acetyltransferase enzymes, but should have no effect on non-enzymatic acetylation.(31) This analysis revealed that the acetylation of the candidate proteins was not significantly inhibited by *s*-acetonyl-CoA, while in four cases (LARP1, FATE1, SMA4 and STK25) the incorporation was actually substantially increased, for reasons that are unclear. This was in contrast to the acetyltransferase hMOF, which showed more significant inhibition at the concentration of *s*-acetonyl-CoA used (Figure 4C and Table S1). The observation that hMOF did not show more complete inhibition of acetylation upon treatment with *s*-acetonyl-CoA, suggested that hMOF may itself undergo chemical acetylation and this possibility was later confirmed as will be addressed below. Taken together, these data suggested that the candidate proteins were chemically, rather than enzymatically, acetylated. This was further supported by the observation that the candidate proteins all had similar levels of acetylation, along with the identities of some of the proteins. A number of the candidates identified in the screen are well-studied proteins, and are extremely unlikely to be acetyltransferase enzymes; including Birc5 (survivin), KRT8 (keratin), IGL@ (immunoglobulin), and STK25 (kinase). One protein in particular, hemoglobin (HBE1), clearly has no acetyltransferase activity, but has previously been reported to undergo non-enzymatic acetylation.(17)

Identification of a Chemical Acetylation Signature Sequence

To confirm that the signal observed in the *in vitro* acetylation experiments was due to acetylation on lysine side chains, 18 of the candidate proteins were selected for further analysis based on high acetylation signal in the radiolabeling experiments, as well as the

level of expression and quality of the final purified protein. These candidate proteins were chemically acetylated, and subjected to trypsinization and mass spectrometry analysis. All analyzed proteins contained at least one detectable acetylated lysine residue and moreover, the pattern of acetylation was not random; as we found that acetylation occurred at preferred sites within each protein.

The distinct patterns of non-enzymatic acetylation led us to look for a signature sequence for chemical acetylation. The apparent percent acetylation at each site was estimated from the LC-MS/MS data as described in Methods using spectral counts. Each modified site was categorized as high acetylation (apparent acetylation ≥ 20%) or low acetylation (apparent acetylation <20%) and the sequence ranging from −7 to +7 of the lysine in question was submitted to the WebLogo3(34) to generate consensus sequences. Comparing the consensus sequences generated from the high (Figure 5A) and low (Figure 5B) acetylation sites revealed a large charge difference in the seven residues preceding the lysine in question. In particular, acetylated lysines tended to have at least one basic residue (lysine or arginine) in the −7 to −4 position, although many peptides had multiple basic residues from the −7 to −3 position (Figures 5C and 5D). In contrast, lysines with low levels of acetylation more often had acidic residues (glutamate or aspartate) and non-polar residues (typically leucine) flanking the lysine, although the positional requirement seemed to be less stringent. To better quantify this, the pI values for the seven residues preceding the lysine were calculated for each sequence, and the values averaged. Peptides preceding highly acetylated lysines had an average pI value of 9.1 with a SEM of 0.37, while low acetylation peptides had an average pI value of 5.3 with a SEM of 0.39.

To test the requirement of the preceding basic residues for lysine chemical acetylation, we mutated these residues to alanine in three test proteins, BIRC5, DTYMK and DOK4 (Figure 6A) and compared the level of lysine acetylation in the wild-type and mutated sequences. Specifically, the proteins were first fully deacetylated, and then reacted with acetyl-CoA. Replicate aliquots of DTYMK were digested with trypsin or GluC, replicate aliquots of DOK4 were digested with trypsin or AspN and an aliquot of BIRC5 was digested with trypsin. To directly compare the acetylation yields of the wild-type and mutant proteins, the integrated peak area of the corresponding acetylated peptides in the two samples were compared. In the case of Glu-C and AspN this meant the comparison was between the identical peptide from the two samples, while the tryptic peptides between mutant and wild-type protein were similar but not identical as the mutation removed a trypsin cleavage site. However, the calculated decrease in acetylation of the target site in the mutant proteins were similar for both the tryptic and non-tryptic peptides, therefore, the results from the two proteases were averaged. In a striking fashion, in all cases, the mutated proteins had large decreases in acetylation at the lysine following the mutation site (Figure 6B), while the levels of acetylation in the remainder of the proteins remained unchanged (Figure S3). These data strongly suggest that chemical lysine acetylation depends on the presence of preceding basic residues.

Correlation with Whole Cell Acetylomics Studies

Although we have demonstrated that non-enzymatic protein lysine acetylation can occur at physiological acetyl-CoA concentrations *in vitro*, we were curious about whether this phenomena carries over to an *in vivo* environment. To assess this, our final list of 58 validated chemically acetylated proteins *in vitro* was compared to published whole cell proteomics data (Table 1), to determine if the chemically acetylated sites that we identified had been previously observed *in vivo*. In total, 28 of 49 (57%) of the proteins that we validated to be chemically acetylated *in vitro* were also observed to be acetylated in whole cell proteomics studies (Table 1). In addition, several proteins that we analyzed for site-specific acetylation using mass spectrometry identified the same specific modified residues as found in the proteomics studies. For example, the protein survivin, well studied for its role in the cell cycle and as a substrate of CBP, was shown to have a number of lysine residues acetylated *in vivo*. We demonstrated that three of these lysine residues in particular, Lys23, Lys115 and Lys122,(35) were all chemically acetylated in our study. After overnight treatment with acetyl-CoA, residues Lys23 and Lys122 were found to be more than 50% chemically acetylated, while Lys115 was fully acetylated. The protein fascin (FSCN1), involved in actin organization, when treated similarly had apparent complete acetylation on Lys471, which was previously observed to be aceylated in whole cell proteomics studies.(9) The ribosomal subunit RPS26 had two lysine residues with correlation between our study and whole cell proteomic studies, Lys66(9) and Lys82(36), although Lys66 was only acetylated to low levels in our hands. Lys169 of the dTMP kinase has been reported to be acetylated *in vivo*, and was found to be ~10% acetylated by chemical means in our study. Given these extensive correlations between the chemical acetylation sites that we identified *in vitro* and prior demonstration that the same sites were acetylated in cells, we propose that many, if not all, of the sites that we have identified to be chemically acetylated in our confirmation studies *in vitro* may, in fact, become acetylated by chemical means in cells.

hMOF lysine 274 autoacetylation

Upon examination of the sequences of the MYST family of acetyltransferases, a perfectly conserved, positively charged lysine residue (Lys268 in hMOF) was observed in the −6 position relative to the reported autoacetylated lysine residue (Lys274 in hMOF), within a region of generally low sequence similarity (Figure 7A). In the enzyme hMOF, Lys274 is located in a loop region and acetylation of this residue appears to trigger a structural rearrangement of the Lys274 bearing loop from outside to inside the enzyme active site, a rearrangement that appears to be critical for hMOF acetylation of cognate substrates.(13, 19, 24) As the sequence surrounding this lysine appeared to share the signature sequence of a preceding basic residue, a short peptide spanning residues 261–278 of hMOF was synthesized to test the non-enzymatic acetylation of this lysine residue. Additionally, peptides containing an arginine or methionine at the −6 position, rather than the endogenous lysine, were generated, to assess the role of the preceding basic residue. Upon incubation with increasing amounts of radiolabeled acetyl-CoA, only peptides with a positively charged residue in the preceding position were efficiently acetylated, and the level of acetylation was linear with respect to acetyl-CoA concentration (Figure 7B). To confirm the acetylation status of the peptides, peptides were treated with 200μM acetyl-CoA, and subjected to

MALDI mass spectrometry. Consistent with the radioassay, only the lysine and arginine containing peptides showed detectable acetylation, as seen as a peak with a mass shift of 42Da (Figure 7C). While these peptide assays cannot determine the actual mechanism of Lys274 acetylation in the intact enzyme, this data suggests that this modification is at least capable of happening in an non-enzymatic fashion, dependant on the preceding basic residues. This may reflect the full mechanism of the modification *in vivo*, or it may function as a redundancy to ensure that this residue is always acetylated in the cell.

DISCUSSION

We have successfully employed the NAPPA array to identify human proteins that undergo non-enzymatic (chemical) acetylation. Further site-specific analysis of a subset of these proteins revealed that these chemically acetylated lysine residues typically have a pattern in the preceding sequence, where basic residues are found in the −7 to −3 positions. The basis of the increased chemical reactivity of these lysine residues remains unknown; however, it is likely the preceding basic residues function to sufficiently lower the pKa of the substrate lysine so that an enzyme is not required to deprotonate the side chain, thus allowing the acetylation reaction to occur spontaneously. The possibility of positive residues spatially near a lysine residue affecting the pKa of its terminal amines has been recently discussed. (20) The distance between the preceding basic residue and the acetylated lysine in the signature sequence may act to provide enough flexibility for the preceding residue(s) to come in proximity and effect the reactivity of the target lysine. An additional possibility is that the positive residues may provide a docking site for the negatively charged phosphates of the acetyl-CoA, in analogy to the mechanism proposed for acetylation in bacteria by acetyl-phosphate (21). However, in the case of acetyl-phosphate acetylation, a negatively charged residue in the −1 position was required, which we do not observe in the peptides acetylated by acetyl-CoA.

Previous studies of chemically acetylated proteins in bacterial and mitochondrial proteins have demonstrated a correlation between chemically acetylated sites and 3-D protein environment (20). In contrast, our comparison of available 3-D structures containing “low” and “high” chemically acetylated sites (Figure 5A and 5B) reveals no statistically significant difference between the two types of sites for (1) the number of nearby positive residues, (2) the solvent accessibility, or (3) the secondary structure (Table S2). Based on these observations, we conclude that, at least in the case of the mammalian proteins that we analyzed, the propensity of a particular lysine residue to undergo chemical acetylation does not correlate with its 3D environment, but that the sequence features noted in Figures 5A and 5B appear to be the most indicative factor in the propensity for a certain lysine residue to undergo chemical acetylation.

Several of the acetylated lysine residues identified in our analysis also overlap with lysine residues which have been shown to be acetylated in whole cell proteomics studies, implying that these residues may be chemically acetylated sites *in vivo*. However, as the whole cell proteomics studies performed in mammalian systems specifically selected for acetyllysine containing peptides, determining a stoichiometry of acetylation in the cell would not be possible; so it remains unclear if these sites are acetylated to a high degree in the cell, or

rather exist at background levels. Additionally, it is not surprising that some chemically acetylated proteins identified in this study would have been missed in prior acetylome studies because identification of post-translational modifications in low abundance proteins, especially when such modifications are likely sub-stoichiometric, are very challenging and incomplete in global analyses. Further investigation of the remaining chemical acetylation sites that we identified will be necessary to determine whether these remaining modifications also occur *in vivo*.

Interestingly, while chemical acetylation has previously been proposed to occur in the mitochondria (20, 22), we noted that most of our identified targets reside in the cytosol, nucleus and other organelles. The mitochondrial compartment contains millimolar concentrations of acetyl-CoA, approximately 100× more than is found in the cytoplasm (23). These relative concentrations may, in fact, explain our candidates being skewed away from the mitochondria. On an absolute scale, a protein would need to be much more easily acetylated, if that protein were in an environment with lower acetate donor concentrations. And as we exposed all of the candidate proteins to the same concentrations of acetyl-CoA, regardless of their cellular localizations, we would select for proteins most easily acetylated, which would tend towards non-mitochondrial proteins. Overall, this has resulted in the first large-scale identification of mammalian proteins capable of chemical acetylation, which are not localized to the mitochondria.

In retrospect, while many non-histone acetylation sites have been determined, primarily by global proteome analysis after enrichment of acetylated peptides, few have been shown to be the direct result of an enzymatic modification.(9, 20, 22) Inspection of the known non-histone enzymatic sites shows a mixture of sequences with and without preceding basic residues (Figure 8), but in general, there are fewer preceding basic residues in the enzymatic sites overall compared with the chemical acetylation consensus sequence identified here. Specifically, in the case of the enzymes αTAT1 and hMOF, the sequences before the substrate lysine have no positive residues,(5, 26, 37) while in the case of Esco1, PAT and PCAF, some sites contain at least one basic residue in the −7 to −3 position,(38–40) in line with the previously determined signature sequence. Of note, a peptide containing the Lys105 site of the substrate SMC3 can be very efficiently chemically acetylated *in vitro* (*unpublished data*). This may indicate that the site is acetylated by multiple means in the cell, or that the lysine can be simply more efficiently acetylated by the enzyme in the cell. This reasoning may extend to the other enzymatic sites which show preceding basic residues, in that the positive residues can serve to lower the pKa of the substrate lysine, allowing more efficient acetylation, either chemically or enzymatically.

A remaining question is the biological role of the non-enzymatic acetylation sites. Whereas acetylation by enzymes has been shown to affect protein binding affinities, stabilities, and cellular localizations,(35, 41, 42) and is well known to act as a reversible switch,(29, 41, 43, 44) the impact of non-enzymatic acetylation is unknown. Also, it is not clear if the non-enzymatic acetylation sites exist in equilibrium, or are biologically intended to always be acetylated. It is also possible that the acetylation of these sites could act as a sensor to respond to the concentration of acetyl-CoA in the cell. As high acetyl-CoA concentrations are a cellular signal for an upregulation of cellular metabolism,(33) this would linearly

increase the first-order reaction of non-enzymatic acetylation, resulting in modulation of protein functions.

These studies have important implications for the activity of the MYST family of protein acetyltransferases. It has been established that these enzymes must be auto-acetylated on Lys262 of yeast ESA1 and Lys274 of human MOF to be functional.(12) This was originally attributed to a *cis*-enzymatic acetylation, as the acetylated lysine is near the active site of the enzyme and exhibits first order reaction kinetics in regards to protein concentration.(13) However, we have demonstrated that in the context of short peptides, the region surrounding the lysine is sufficient for non-enzymatic acetylation, and moreover, the presence of the preceding basic residue is necessary for this modification to occur. While the mechanism in the intact enzyme cannot be fully described by these experiments, the contribution of chemical acetylation to Lys274 modification is supported by the fact that a catalytically inactive mutant of hMOF is able to undergo Lys274 acetylation,(13) and that *s*-acetonyl-CoA cannot completely inhibit the auto-acetylation of this lysine in hMOF (Figure 4C).

The unrelated acetyltransferase enzyme p300 has also been shown to have a lysine rich activation loop, which requires acetylation to generate the active enzyme.(16) Investigation of the loop sequence reveals that the acetylated lysines responsible for enzymatic activity show a non-enzymatic acetylation consensus sequence, with each acetylated residue being preceded by positive residues in the −5 to −3 positions (Figure 9). While studies to date have demonstrated that p300 is able to acetylate these lysine residues(16), it is possible that some of this acetylation occurs through chemical means.

Taken together, these studies add chemical acetylation to the toolbox of how many proteins in different cellular organelles can be modified. While the major biological roles of non-enzymatic acetylation and its regulation are unknown, it is likely that the process is much more common than previously appreciated, and must be considered when investigating the mode of protein acetylation.

METHODS

Cloning

To generate positive and negative controls for initial establishment of assay protocols on NAPPA, each gene was cloned into the pANT7-cGST in vitro expression vector encoding a C-terminal GST tag using the Gateway recombination cloning system(21). To generate vectors for *E. coli* expression, genes were PCR amplified from the templates in pANT7-cGST (https://DNASU.org) and cloned into the pET-BAPP vector. The pET-BAPP vector was generated from the pET-DUET backbone, which was then edited to generate a single expression cassette containing an N-terminal GST tag, followed by a TEV cleavage site. The multi-cloning site of the vector was re-engineered to contain the BamHI site followed by the 8-cutter restriction sites for AsisI, PacI, and PmeI. All PCR amplified genes were digested with AsisI and PmeI, followed by ligation into a similarly cut vector. All vectors were sequenced to confirm the presence and fidelity of the insert.

Array Printing

Array production and quality control were performed as previously described (28). Template DNAs were obtained from the DNASU collection (https://DNASU.org), a set of plasmid DNAs which have a focus on human ORFs in FLAG or GST tagged *in vitro* translation vectors. Plasmid DNA was prepared in a 96 well format and re-suspended in water. BSA, anti-tag antibody (anti-GST from GE Healthcare or anti-FLAG from Sigma), and BS3 linker (Pierce) were added to a final concentration of 3.7 mg/ml, 375 μg/μl, and 5 mM, respectively. Printing mixtures were incubated overnight at 4°C before being arrayed onto aminosilane (Pierce) coated glass slides by GenetixQArray2 using 350 μm solid pins. Printed plasmid DNA was visualized by picogreen staining (Life Technology) to assess printing quality. We used anti-tag antibodies different from the capture antibodies printed on the arrays (anti-GST from Cell Signaling and anti-FLAG from Sigma) to assess protein display on arrays after cell free expression using 1-Step Human Coupled *in vitro* Expression Kits (Thermo).

Array Processing

NAPPA arrays were expressed using one of two general methods. In both cases, printed arrays were blocked for 1 hour in SuperBlock Blocking Buffer (ThermoScientific) followed by 5 rinses with dH_2O. The slides were further blocked with incubation with 500μM cold acetyl-CoA for 2 hours in 20mM Tris-HCl pH 7.8, 50mM NaCl. Following blocking, the slides were dried under compressed air, and a Hybriwell gasket (Grace Bio-Labs) applied. The slides were then translated using the 1-Step Human Coupled IVT Kit (Pierce) according to the manufacturer's instruction for 90min at 30°C followed by 45min at 15°C. For the Single-Step method, the translation mixture was supplemented with 300μM ^{14}C-acetyl-CoA, and following translation the slides were immediately washed with TBST Buffer. For the Multi-Step method, the translation mixture was supplemented with 400μM *s*-acetonyl-CoA. The slides were then washed extensively with 20mM Tris-HCl pH 7.8, 100mM NaCl, followed by an overnight, room-temperature incubation with 300μM ^{14}C-acetyl-CoA in the same buffer. The slides were then washed extensively with TBST. For both methods, the slides were rinsed with dH_2O and then dried completely under compressed air. The dry slides were then applied to high resolution, single emulsion film and incubated at −80°C for approximately 30 days, after which the film was developed and scanned.

Array Analysis

After processing, spot intensity from scanned film images were quantified using ArrayProImage Analyzer for downstream data analysis. The normalized intensity of each spot on the array was calculated by dividing the intensity of the spot by the average intensity of the spots in an 11 × 11 box surrounding it. The normalized intensities from replicate slides were then averaged, and the average intensity of the (-) T7 Polymerase slides was subtracted from the Full Reaction slides. The change in each spot's intensity was ranked by standard deviations in the local area (local area z-score). The final values for the total collection of genes ranged between −3.5 to 3.5σ, with the negative z-scores generated by the surrounding genes having increased signal, resulting in a lower normalized value.

Protein Expression

Expression plasmids encoding the genes of interest were transformed into BL21 (DE3) cells, and grown in 1L of TB media to an OD_{600} of 0.8. Protein expression was induced with 0.5mM IPTG for 12–16 hours at 20°C. Following expression, the cells were pelleted, and resuspended in PBS supplemented with 300mM NaCl, 2mM β-ME, 0.1mg/ml PMSF. Cells were lysed by sonication, and then centrifuged to remove insoluble material. The cleared lysate was then incubated with 0.5mL of GST resin (Gold Bio) for 2 hr. at 4°C. The beads were washed three times with 50mL of PBS plus 300mM NaCl, 2mM β-ME, and then applied to a gravity flow column. The resin was further washed with PBS plus 300mM NaCl, 2mM β-ME, after which the proteins were deacetylated for ~12hr. at room temperature with 0.5mg/ml HST2 dissolved in PBS plus 1mM NAD^+. The resin was washed extensively with the supplemented PBS buffer described above, followed by washing with un-supplemented PBS buffer. The resin was resuspended with 0.5mL of PBS for assaying. Protein concentration was estimated by denaturing 100μL of resin slurry with 500μL of 6M guanidine-HCl and determining the absorbance at 260nm. Concentration measurements were confirmed by densitometry of coomassie stained SDS-PAGE gels.

In vitro Assay

To assay for acetylation, 25μL of resin slurry was incubated with varying concentration of ^{14}C-acetyl-CoA dissolved in 20mM Tris pH 8.0, 200mM NaCl. The reaction was allowed to proceed for 3.5 hr. at room temperature, and was then applied to a glass-fiber filter plate (Millipore). After pulling through the reaction mixture, the retained resin was washed three times with 200mM Tris pH 8.0, 200mM NaCl, after which 100μL of scintillation fluid was added and the plate sealed. Incorporation was measured by counting on a 96-well TopCount scintillation counter, and all counts were normalized by the mass of protein present in the sample.

Mass Spectrometry & Peptide Analysis

The purified proteins immobilized on resin were acetylated using 300μM acetyl-CoA in 20mM Tris pH 8.0, 200mM NaCl for 3 hr. at room temperature. SDS sample loading buffer was added to stop the reaction, and ~10μg of each sample was applied to an SDS-PAGE gel, which was then stained with coomassie R-250. Bands corresponding to the appropriate molecular weight of each sample were excised and digested with trypsin. In some cases, replicate samples were also digested with either GluC or AspN.

Peptides were loaded into a UPLC Symmetry trap column (180 μm inner diameter × 2 cm packed with 5 μm C18 resin, Waters) and separations were performed on a 15-cm PicoFrit column (75 μm inner diameter, New Objective) packed with Magic 5 μm C18 reversed-phase resin (Michrom Bioresources) using a nanoflow high-pressure LC system (Eksigent Technologies), which was coupled online to a LTQ-Orbitrap XL mass spectrometer (Thermo Fisher Scientific) via a nanoelectrospray ion source. Chromatography was performed with Solvent A (Milli-Q water with 0.1% formic acid) and Solvent B (acetonitrile with 0.1% formic acid). Peptides were eluted at 200 nL/min for 5–30% B over 75 min, 30–80% B over 5 min, 80% B for 10 min before returning to 5% B over 1 min. To minimize sample carryover, a fast blank gradient was run between each sample. The LTQ-Orbitrap

XL was operated in the data-dependent mode to automatically switch between full scan MS (m/z 300–2000) in the Orbitrap analyzer (resolution of 60,000 at m/z 400) and fragmentation of the six most intense ions exceeding a minimum threshold of 1000 by collision-induced dissociation in the ion trap mass analyzer. Monoisotopic precursor selection was enabled, and charge state screening was enabled to reject ions with undefined charge state. Ions subjected to MS/MS were excluded from repeated analysis for 20 s.

MS/MS spectra were extracted and searched using the SEQUEST algorithm in BioWorks (version 3.3.1, Thermo Fisher Scientific) against a combined E. coli proteome database and the recombinant protein sequences. The reversed sequences of the database and a list of common contaminants were also appended. MS/MS spectra were searched using partial specificity for the expected target sites of the protease that was used with up to two missed cleavages, a 100 ppm precursor mass tolerance, 1 amu fragment ion mass tolerance, static modification of cys by carbamidomethylation (+57.0215), and variable modifications for methionine oxidation (+15.9949), asparagine deamidation (+0.9840) and lysine acetylation (42.0106). Consensus protein lists were generated by DTASelect (version 2.0, Scripps Research Institute, La Jolla, CA) using the following data filter: 10 ppm precursor mass accuracy, $\Delta Cn \geq 0.05$, and requiring a minimum of three peptides per protein. The peptide false discovery rate was less than 1%.

Estimation of Acetylation Levels at Specific Lysines

In order to develop a chemical acetylation consensus sequence signature, sites of acetylation were detected by LC-MS/MS analysis of tryptic digests of target proteins excised from SDS gels. The extent of acetylation at each modified lysine was estimated by dividing the number of times a given lysine was identified in a sequence in its acetylated form (acetylated spectral counts) by the total spectral counts for all sequences containing that lysine in either its modified or unmodified state. This is a rough estimate because trypsin does not cleave at acetylated lysines and the MS ionization efficiencies for complete tryptic cleavages of unmodified lysines, incomplete cleavages with internal unmodified lysines, and incomplete cleavages containing a modified lysine will usually be different. As a result, the number of spectral counts of these different forms of a given sequence will not directly correspond to molar yields. However, analysis of tryptic peptides produced more extensive sequence coverage and detection of acetylation sites than other proteases, thereby limiting the value of alternative approaches. In addition, this method will typically under-estimate acetylation levels because in general, complete cleavages at unmodified peptides will yield stronger signals and spectral counts per femtomole than incomplete peptides either with or without modification of the internal lysine. Hence, this is the most practical method to globally estimate the “apparent percent modification” for a relatively large number of proteins of interest. After ranking the peptides by apparent percent acetylation, the signature sequence was determined by supplying the peptides with 20–100% acetylation versus 1%–20% acetylation to the WebLogo3 web server(34). pI prediction calculations were performed on the peptides using an adaptation of the half-interval search algorithm, which excluded the contribution of N- and C-termini.

Comparison of the acetylation level of specific lysine residues in several wild-type and mutant proteins was performed by integrating the acetylated peptide peak area of the extracted ion chromatogram for the mutant protein and comparing it to the peak area of the corresponding peptide from the wild-type protein. To correct for differences in sample load, data were corrected based on the peak areas of two to four non-acetylated peptides for each digest of the mutant and wild-type protein. For comparison, the corrected peak area of the acetylated peptide from the wild-type protein was set to 1.0.

hMOF peptide acetylation assay

Peptides corresponding to residues 261–278 of hMOF were synthesized. The peptides were tested for acetylation by incubation with acetyl-CoA in NaCl, 20mM Tris pH 8.0 for 3 hours. For radioactive experiments, ^{14}C labeled acetyl-CoA was used between 1–200μM, after which the pH was lowered using sodium acetate pH 5.5, after which it was spotted onto P81 filter paper. The paper was washed three times with 20mM sodium acetate pH 5.5 and incorporation was measured by counting in scintillation fluid. For MALDI analysis, the reaction was performed using 200uM cold acetyl-CoA. The reaction was mixed 1:1 with 10 mg/ml alpha-cyanno-4-cinnamic acid (Sigma), spotted onto a 96 well coated plate and analyzed on an ABI/PerSeptive Voyager DE-PRO MALDI-TOF instrument in positive ion mode.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

GM060293 (R.M.) and NIH training grant CA009171 (A.O.); and by the Ovarian Cancer Research Fund (R.M. and D.W.S.). We acknowledge the use of the Wistar Proteomics Core facility for the work reported here, which is supported in part by NIH grant CA010815. We would like to thank J. Domsic for assistance in the peptide analysis and pI calculations.

References

1. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. Genome research. 2004; 14:1188–1190. [PubMed: 15173120]
2. Cosgrove MS, Boeke JD, Wolberger C. Regulated nucleosome mobility and the histone code. Nature structural & molecular biology. 2004; 11:1037–1043.
3. Gu W, Roeder RG. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. Cell. 1997; 90:595–606. [PubMed: 9288740]
4. Tang Y, Luo J, Zhang W, Gu W. Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. Molecular cell. 2006; 24:827–839. [PubMed: 17189186]
5. Shida T, Cueva JG, Xu Z, Goodman MB, Nachury MV. The major alpha-tubulin K40 acetyltransferase alphaTAT1 promotes rapid ciliogenesis and efficient mechanosensation. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:21517–21522. [PubMed: 21068373]
6. Maruta H, Greer K, Rosenbaum JL. The acetylation of alpha-tubulin and its relationship to the assembly and disassembly of microtubules. The Journal of cell biology. 1986; 103:571–579. [PubMed: 3733880]

7. Reed NA, Cai D, Blasius TL, Jih GT, Meyhofer E, Gaertig J, Verhey KJ. Microtubule acetylation promotes kinesin-1 binding and transport. Current biology : CB. 2006; 16:2166–2172. [PubMed: 17084703]
8. Kim SC, Sprung R, Chen Y, Xu Y, Ball H, Pei J, Cheng T, Kho Y, Xiao H, Xiao L, Grishin NV, White M, Yang XJ, Zhao Y. Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. Molecular cell. 2006; 23:607–618. [PubMed: 16916647]
9. Choudhary C, Kumar C, Gnad F, Nielsen ML, Rehman M, Walther TC, Olsen JV, Mann M. Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science (New York, NY). 2009; 325:834–840.
10. Hornbeck PV, Kornhauser JM, Tkachev S, Zhang B, Skrzypek E, Murray B, Latham V, Sullivan M. PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic acids research. 2012; 40:D261–270. [PubMed: 22135298]
11. Wang H, Holloway MP, Ma L, Cooper ZA, Riolo M, Samkari A, Elenitoba-Johnson KS, Chin YE, Altura RA. Acetylation directs survivin nuclear localization to repress STAT3 oncogenic activity. The Journal of biological chemistry. 2010; 285:36129–36137. [PubMed: 20826784]
12. Kaeberlein M, McVey M, Guarente L. The SIR2/3/4 complex and SIR2 alone promote longevity in Saccharomyces cerevisiae by two different mechanisms. Genes & development. 1999; 13:2570–2580. [PubMed: 10521401]
13. Yuan H, Rossetto D, Mellert H, Dang W, Srinivasan M, Johnson J, Hodawadekar S, Ding EC, Speicher K, Abshiru N, Perry R, Wu J, Yang C, Zheng YG, Speicher DW, Thibault P, Verreault A, Johnson FB, Berger SL, Sternglanz R, McMahon SB, Cote J, Marmorstein R. MYST protein acetyltransferase activity requires active site lysine autoacetylation. The EMBO journal. 2012; 31:58–70. [PubMed: 22020126]
14. Thompson PR, Wang D, Wang L, Fulco M, Pediconi N, Zhang D, An W, Ge Q, Roeder RG, Wong J, Levrero M, Sartorelli V, Cotter RJ, Cole PA. Regulation of the p300 HAT domain via a novel activation loop. Nature structural & molecular biology. 2004; 11:308–315.
15. Pirooznia SK, Sarthi J, Johnson AA, Toth MS, Chiu K, Koduri S, Elefant F. Tip60 HAT activity mediates APP induced lethality and apoptotic cell death in the CNS of a Drosophila Alzheimer's disease model. PloS one. 2012; 7:e41776. [PubMed: 22848598]
16. Di Cerbo V, Schneider R. Cancers with wrong HATs: the impact of acetylation. Briefings in functional genomics. 2013; 12:231–243. [PubMed: 23325510]
17. Garbutt GJ, Abraham EC. Non-enzymatic acetylation of human hemoglobins. Biochimica et biophysica acta. 1981; 670:190–194. [PubMed: 6170344]
18. Paik WK, Pearson D, Lee HW, Kim S. Nonenzymatic acetylation of histones with acetyl-CoA. Biochimica et biophysica acta. 1970; 213:513–522. [PubMed: 5534125]
19. Weinert BT, Iesmantavicius V, Wagner SA, Scholz C, Gummesson B, Beli P, Nystrom T, Choudhary C. Acetyl-phosphate is a critical determinant of lysine acetylation in E. coli. Molecular cell. 2013; 51:265–272. [PubMed: 23830618]
20. Baeza J, Smallegan MJ, Denu JM. Site-specific reactivity of nonenzymatic lysine acetylation. ACS chemical biology. 2015; 10:122–128. [PubMed: 25555129]
21. Kuhn ML, Zemaitaitis B, Hu LI, Sahu A, Sorensen D, Minasov G, Lima BP, Scholle M, Mrksich M, Anderson WF, Gibson BW, Schilling B, Wolfe AJ. Structural, kinetic and proteomic characterization of acetyl phosphate-dependent bacterial protein acetylation. PloS one. 2014; 9:e94816. [PubMed: 24756028]
22. Wagner GR, Payne RM. Widespread and enzyme-independent Nepsilon-acetylation and Nepsilon-succinylation of proteins in the chemical conditions of the mitochondrial matrix. The Journal of biological chemistry. 2013; 288:29036–29045. [PubMed: 23946487]
23. Cai L, Sutter BM, Li B, Tu BP. Acetyl-CoA induces cell growth and proliferation by promoting the acetylation of histones at growth genes. Molecular cell. 2011; 42:426–437. [PubMed: 21596309]
24. Klein AH, Shulla A, Reimann SA, Keating DH, Wolfe AJ. The intracellular concentration of acetyl phosphate in Escherichia coli is sufficient for direct phosphorylation of two-component response regulators. Journal of bacteriology. 2007; 189:5574–5581. [PubMed: 17545286]

25. Nand A, Gautam A, Perez JB, Merino A, Zhu J. Emerging technology of in situ cell free expression protein microarrays. Protein & cell. 2012; 3:84–88. [PubMed: 22426976]
26. Thao S, Chen CS, Zhu H, Escalante-Semerena JC. Nepsilon-lysine acetylation of a bacterial transcription factor inhibits Its DNA-binding activity. PloS one. 2010; 5:e15123. [PubMed: 21217812]
27. Weinert BT, Iesmantavicius V, Moustafa T, Scholz C, Wagner SA, Magnes C, Zechner R, Choudhary C. Acetylation dynamics and stoichiometry in Saccharomyces cerevisiae. Molecular systems biology. 2014; 10:716. [PubMed: 24489116]
28. Link AJ, Labaer J. Construction of Nucleic Acid Programmable Protein Arrays (NAPPA) 1: Coating Glass Slides with Amino Silane. CSH protocols. 2008; 2008 pdb prot5056.
29. Sun B, Guo S, Tang Q, Li C, Zeng R, Xiong Z, Zhong C, Ding J. Regulation of the histone acetyltransferase activity of hMOF via autoacetylation of Lys274. Cell research. 2011; 21:1262–1266. [PubMed: 21691301]
30. Mellert HS, McMahon SB. hMOF, a KAT(8) with many lives. Molecular cell. 2009; 36:174–175. [PubMed: 19854127]
31. Rubenstein P, Dryer R. S-acetonyl-CoA. A nonreactive analog of acetyl-CoA. The Journal of biological chemistry. 1980; 255:7858–7862. [PubMed: 6995455]
32. Imhof A, Yang XJ, Ogryzko VV, Nakatani Y, Wolffe AP, Ge H. Acetylation of general transcription factors by histone acetyltransferases. Current biology : CB. 1997; 7:689–692. [PubMed: 9285713]
33. Brooks CL, Gu W. The impact of acetylation and deacetylation on the p53 pathway. Protein & cell. 2011; 2:456–462. [PubMed: 21748595]
34. Turner BM. Histone acetylation and control of gene expression. Journal of cell science. 1991; 99(Pt 1):13–20. [PubMed: 1757496]
35. Davie JR, Hendzel MJ. Multiple functions of dynamic histone acetylation. Journal of cellular biochemistry. 1994; 55:98–105. [PubMed: 8083305]
36. Hassig CA, Schreiber SL. Nuclear histone acetylases and deacetylases and transcriptional regulation: HATs off to HDACs. Current opinion in chemical biology. 1997; 1:300–308. [PubMed: 9667866]
37. Sykes SM, Mellert HS, Holbert MA, Li K, Marmorstein R, Lane WS, McMahon SB. Acetylation of the p53 DNA-binding domain regulates apoptosis induction. Molecular cell. 2006; 24:841–851. [PubMed: 17189187]
38. Zhang J, Shi X, Li Y, Kim BJ, Jia J, Huang Z, Yang T, Fu X, Jung SY, Wang Y, Zhang P, Kim ST, Pan X, Qin J. Acetylation of Smc3 by Eco1 is required for S phase sister chromatid cohesion in both human and yeast. Molecular cell. 2008; 31:143–151. [PubMed: 18614053]
39. You D, Yao LL, Huang D, Escalante-Semerena JC, Ye BC. Acetyl coenzyme A synthetase is acetylated on multiple lysine residues by a protein acetyltransferase with a single Gcn5-type N-acetyltransferase (GNAT) domain in Saccharopolyspora erythraea. Journal of bacteriology. 2014; 196:3169–3178. [PubMed: 24957627]
40. Lin R, Tao R, Gao X, Li T, Zhou X, Guan KL, Xiong Y, Lei QY. Acetylation stabilizes ATP-citrate lyase to promote lipid biosynthesis and tumor growth. Molecular cell. 2013; 51:506–518. [PubMed: 23932781]
41. Nogueira-Ferreira R, Vitorino R, Ferreira-Pinto MJ, Ferreira R, Henriques-Coelho T. Exploring the role of post-translational modifications on protein-protein interactions with survivin. Archives of biochemistry and biophysics. 2013; 538:64–70. [PubMed: 23938875]
42. Icardi L, De Bosscher K, Tavernier J. The HAT/HDAC interplay: multilevel control of STAT signaling. Cytokine & growth factor reviews. 2012; 23:283–291. [PubMed: 22989617]
43. Guan KL, Xiong Y. Regulation of intermediary metabolism by protein acetylation. Trends in biochemical sciences. 2011; 36:108–116. [PubMed: 20934340]
44. Konig AC, Hartl M, Boersema PJ, Mann M, Finkemeier I. The mitochondrial lysine acetylome of Arabidopsis. Mitochondrion. 2014

A

hMOF	Neg.	GCN5	Neg.	CBP	Neg.	PCAF	Neg.
H4	p300	H4	p300	p300	p53	p300	hMOF
GCN5	p53	CBP	hMOF	PCAF	GCN5	Pos.	CBP
PCAF	--	p300	--	p300	--	p53	--

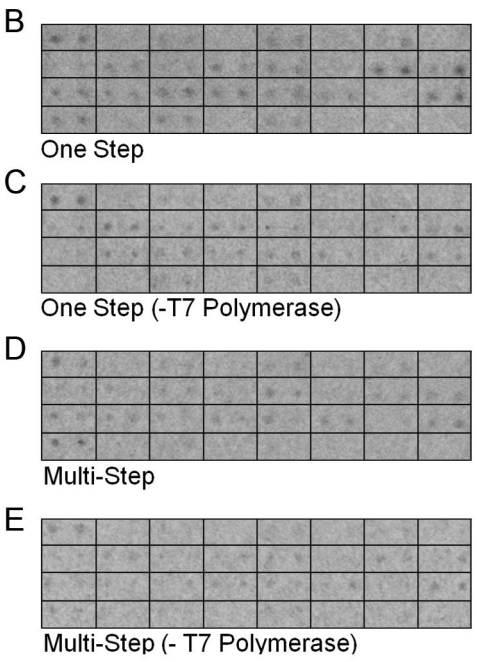

Figure 1. Test array confirms that NAPPA is suitable to screen for chemically acetylated proteins

(A) Layout of printed genes on test array. Positive genes: hMOF, GCN5, CBP, PCAF, p300. Negative genes: H4, p53. Each spot was printed in side-by-side duplicate. (B) 30-day film exposure of test arrays processed using the Single-Step Method, or (C) the Single-Step minus T7 Polymerase. (D) 30-day film exposure of test arrays processed using the Multi-Step Method, or (E) the Multi-Step minus T7 Polymerase.

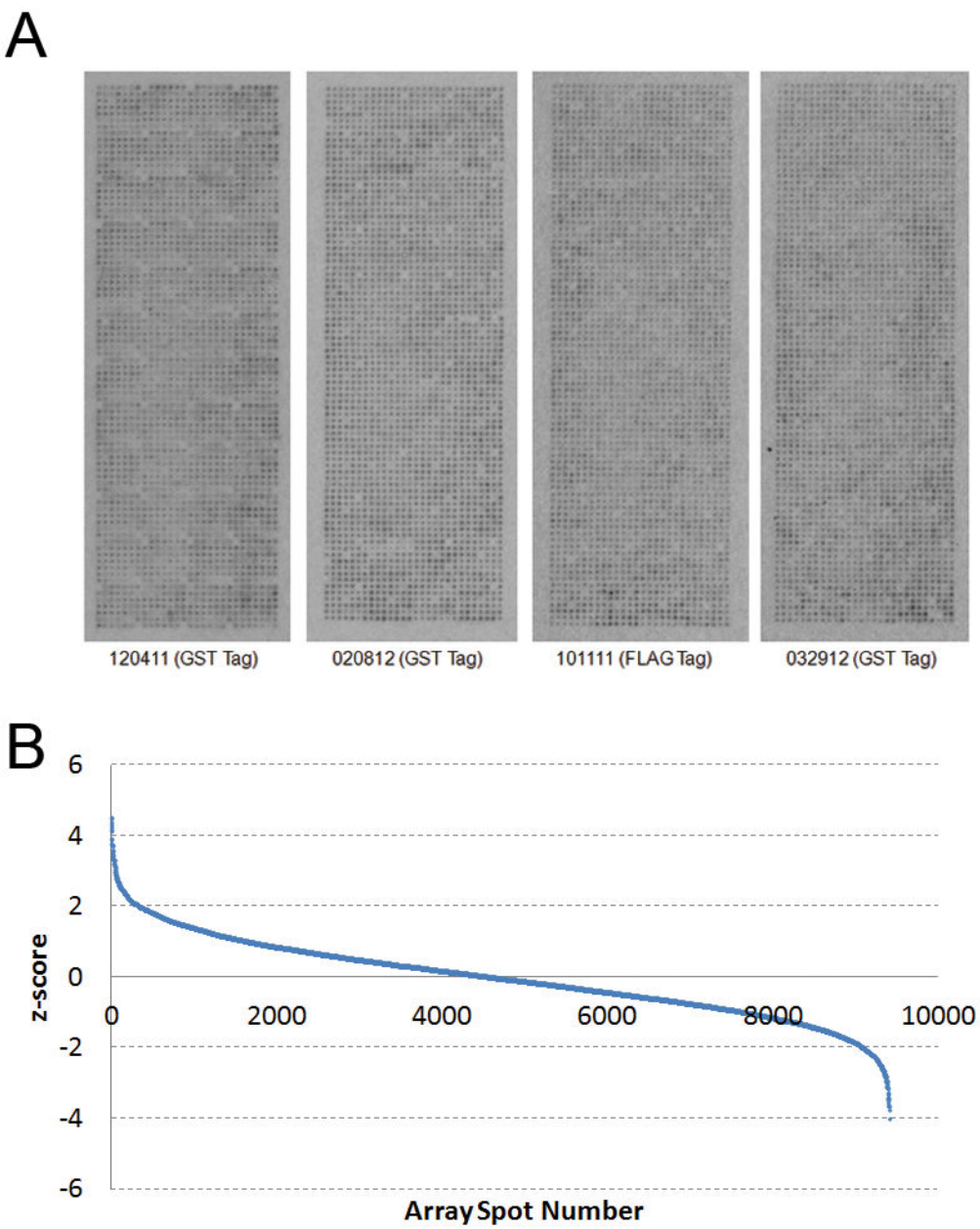

Figure 2. Full NAPPA screen identifies candidate acetylated proteins

(A) Sample films from each of the four chips, which make up the full set of arrays processed using the Single-Step method. Labels indicate the unique identifier for each printed array, along with the affinity tag used for immobilization of the translated protein. (B) z-score of each spot in the array comparing the signal change between control un-translated slides and slides processed using full translation mixtures. For each array, three full translation slides and two untranslated slides were prepared, and the average values from these sets were used to generate z-scores.

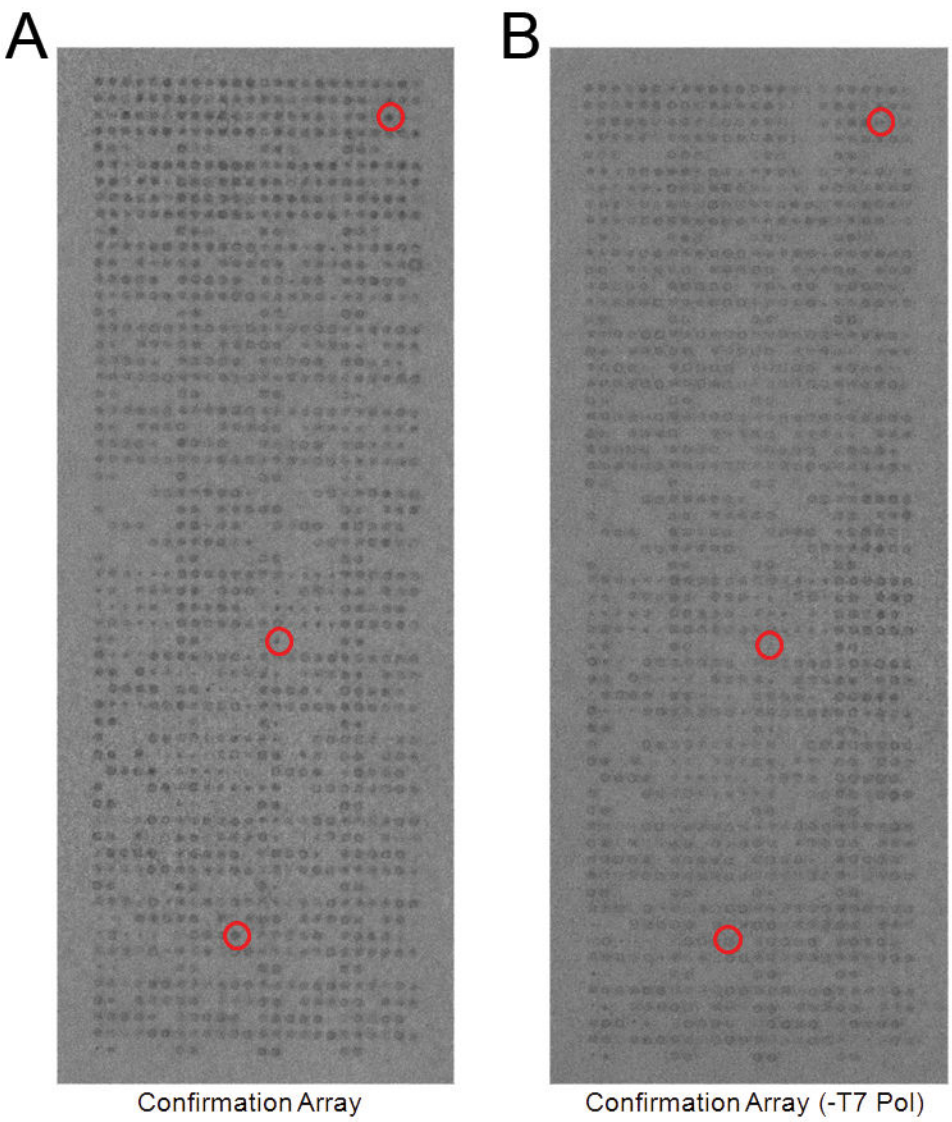

Figure 3. NAPPA screen of acetylated proteins from Figure 3 identifies top candidates

(A) Sample film of confirmation array translated with the full reaction mixture or (B) reaction mixture lacking T7 Polymerase, and processed the Single-Step Methods. Red circles indicate a sample of corresponding spots between the two images showing notable increases in intensity between the control and full reaction arrays.

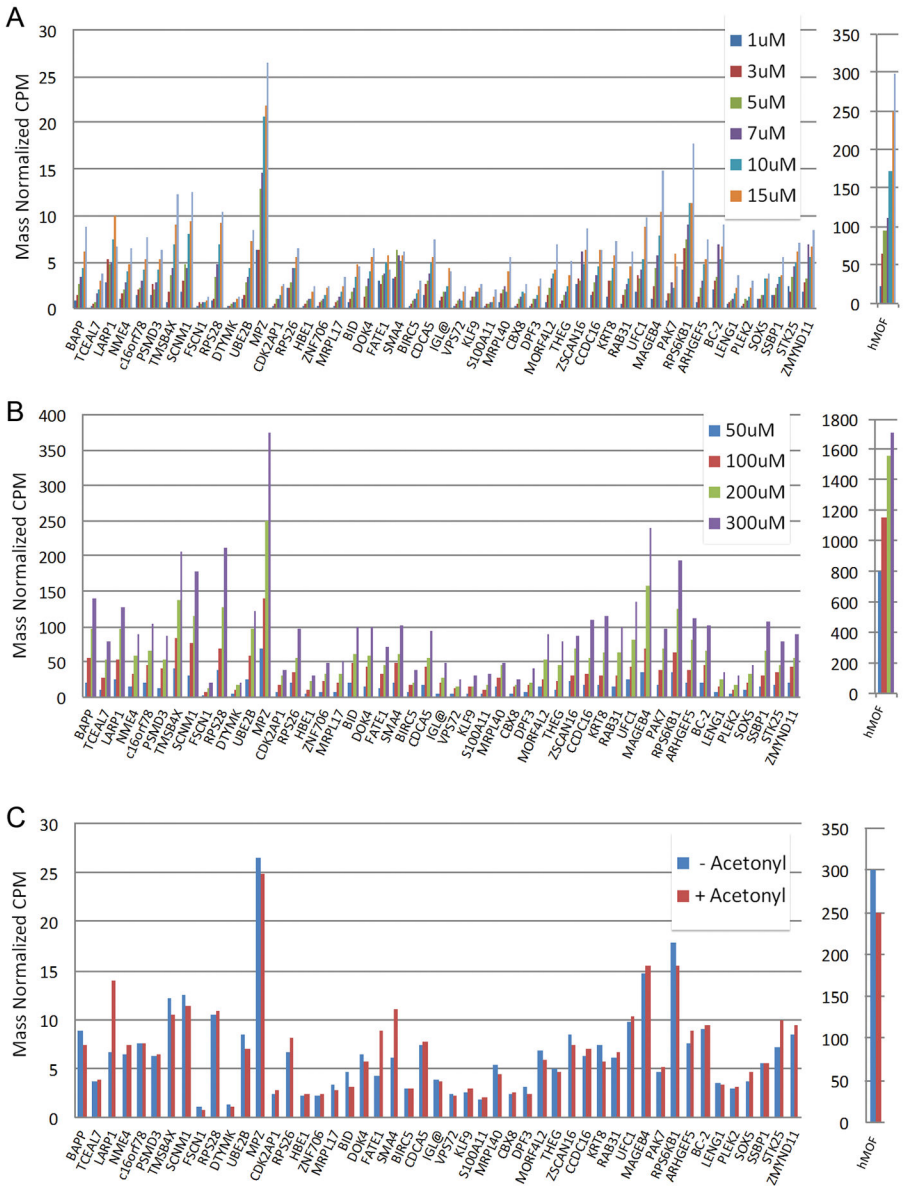

Figure 4. ***In vitro*** **acetylation of candidate proteins shows dose response acetylation indicative of chemical acetylation**

(A) Level of acetylation as measured by ^{14}C incorporation after incubation of purified recombinant protein with varying concentration of ^{14}C-acetyl-CoA. Note different scale between array candidates and hMOF control. (B) Linear increase in acetylation of array candidates upon incubation with high concentrations of ^{14}C-acetyl-CoA, as opposed to the plateau exhibited by hMOF. (C) Lack of effect of general acetyltransferase inhibitor *s*-acetonyl-CoA on acetylation of array candidates, as compared to effect on the acetyltransferase enzyme hMOF.

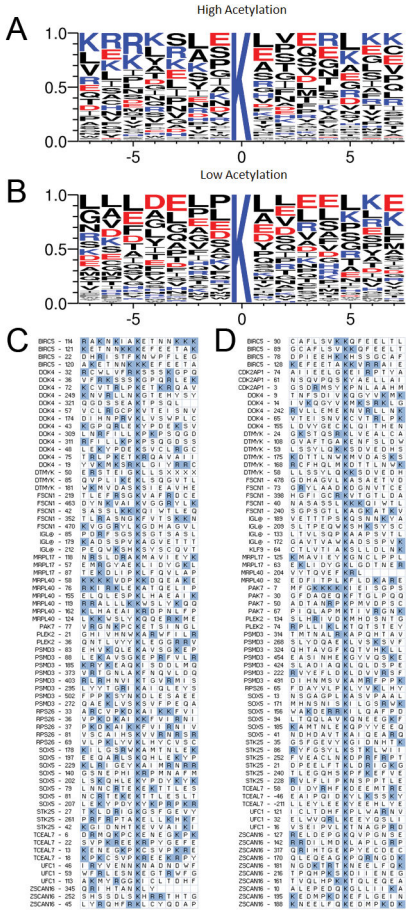

Figure 5. Mass spectrometry reveals an acetylation signature sequence

(A) Signature sequence from alignment of peptides containing (A) highly acetylated lysines (25–100% acetylation) and (B) lysines with low levels of acetylation (1–20%) as determined by trypsinization and LC-MS/MS. (C) List of high acetylation peptides and (D) low acetylation peptides with basic residues highlighted in blue. Annotation denotes the peptide's origin protein, and the amino acid number of the lysine in question.

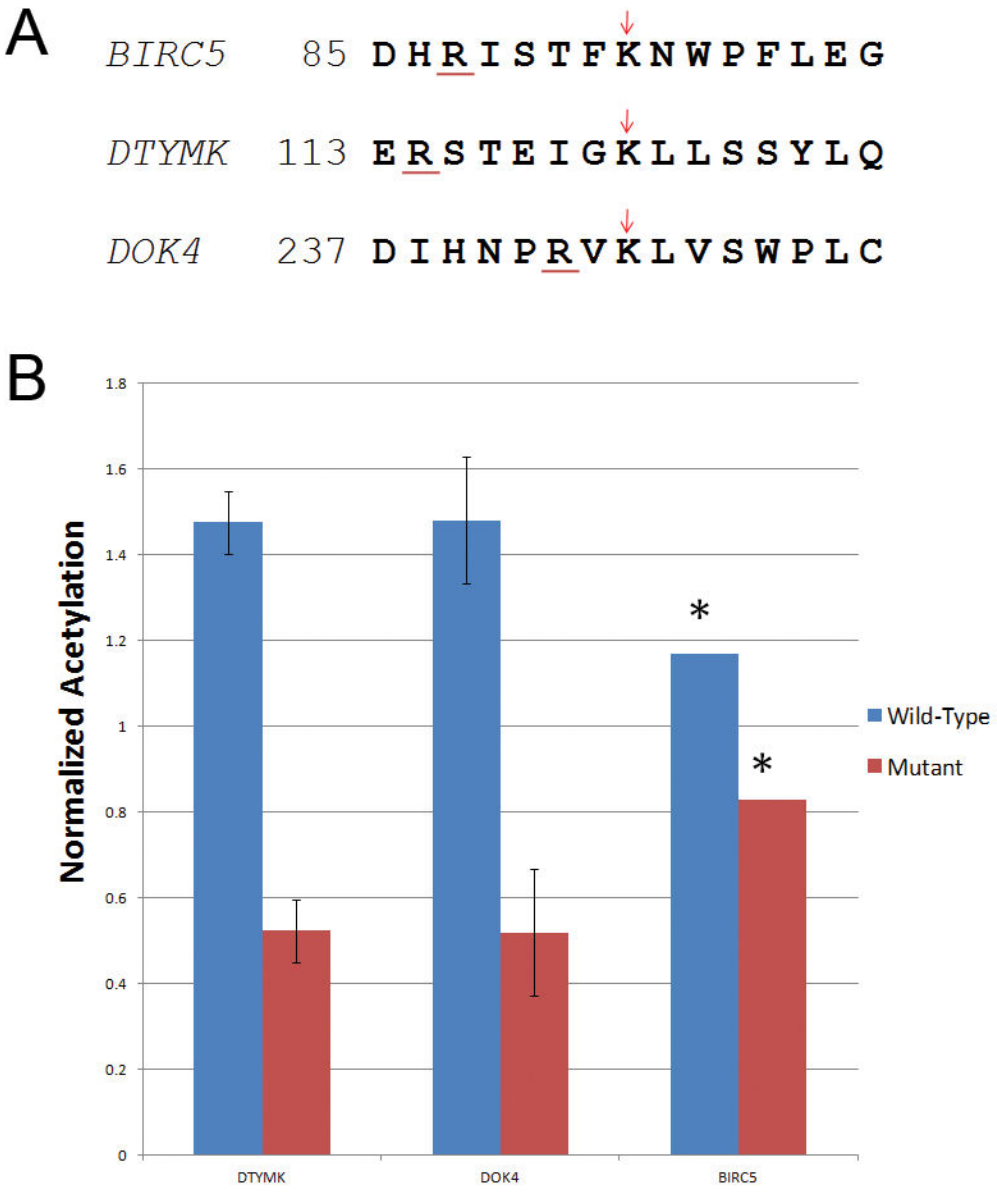

Figure 6. Mutational analysis of basic residues preceding the lysine target of chemical acetylation highlights the importance of these residues for chemical acetylation

(A) Sequence of peptides surrounding highly acetylated lysines (indicated by arrow) and the preceding basic residues mutated to alanine (indicated by underline). (B) Normalized levels of acetylation of the indicated lysine in the wild-type protein and the mutant as determined by digestion and LC-MS/MS. For DTMYK, the protein was individually digested with trypsin and GluC, and the indicated level is the average normalized area of the acetylated peptides from the two protease reactions. DOK4 was digested with trypsin and AspN and similarly analyzed. In the case of BIRC5, only trypsin gave coverage of the peptide in question, so no averaging was possible (indicated by a *).

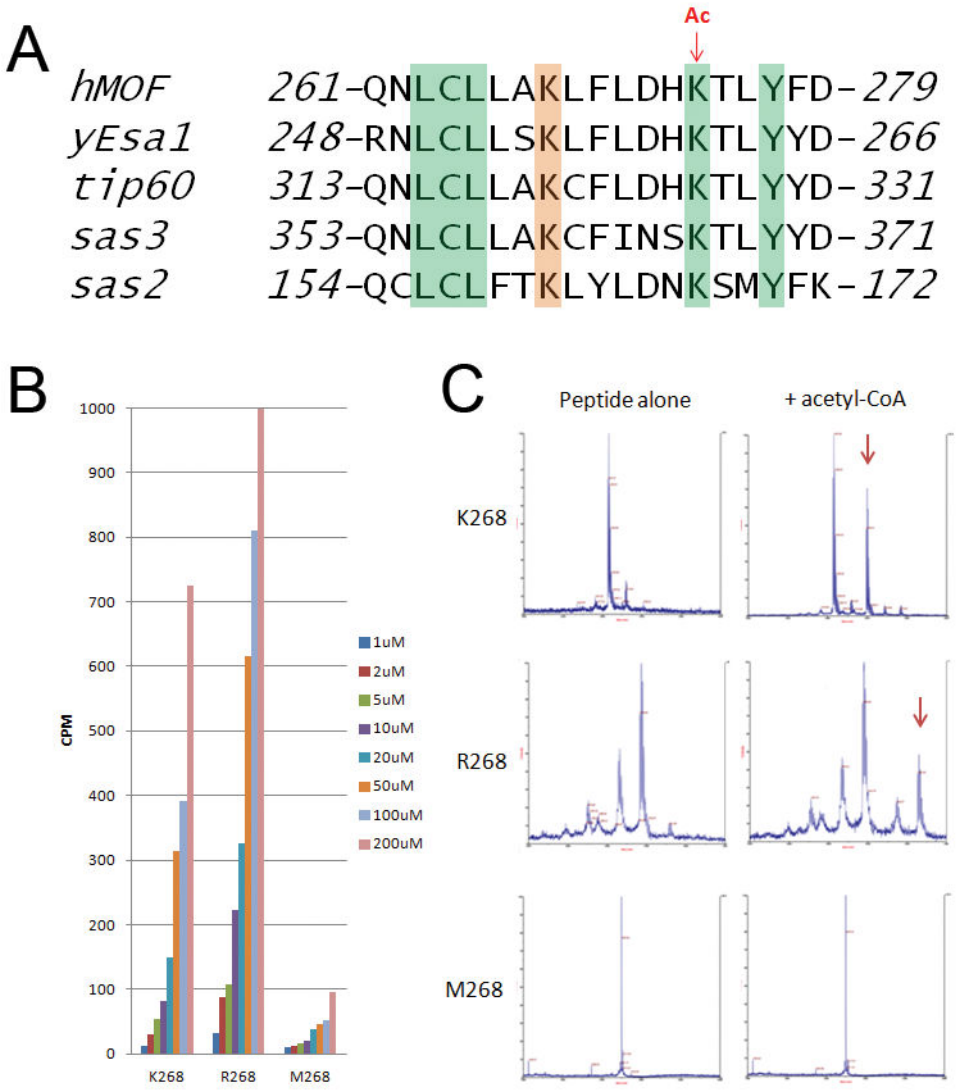

Figure 7. Non-enzymatic modification of acetylated hMOF loop

(A) Sequence alignment of MYST family of acetyltransferase enzymes, with conserved residues highlighted. Auto-acetylation of lysine 274 in hMOF and the corresponding lysine in other family members has been shown to be necessary for catalytic activity. Conserved lysine in the −6 position relative to the acetylated lysine is highlighted in orange. Peptides comprising residues 261–278 of hMOF with different residues at the −6 position were tested for non-enzymatic acetylation. (B) In an ^{14}C incorporation assay, in which the peptides were incubated with increasing amount of labeled acetyl-CoA, peptides with a positive charge at the −6 position showed a linear response with increasing acetyl-CoA concentration, while greatly reduced incorporation was seen with a neutral residue at the −6 position. (C) After incubation with 200μM acetyl-CoA, peptides were analyzed by MALDI mass spectrometry. Only peptides with preceding positive residues showed a mass-shifted peak corresponding to an acetylated peptide (red arrow)

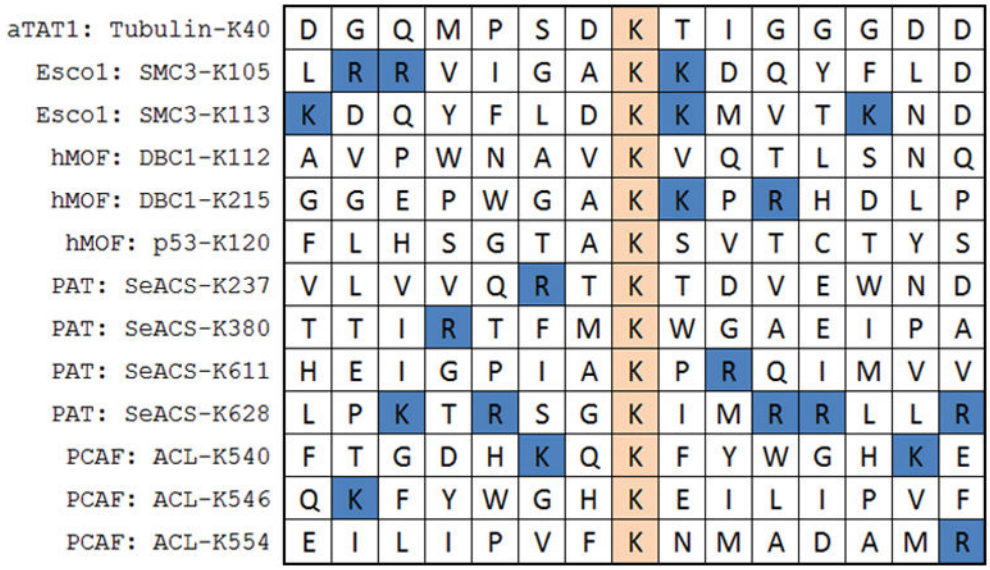

Figure 8. Enzymatic acetylation substrate sites

Sequence alignment of known non-histone sites of enzymatic acetylation. Sites are aligned according to substrate lysine, with basic residues are colored in blue.

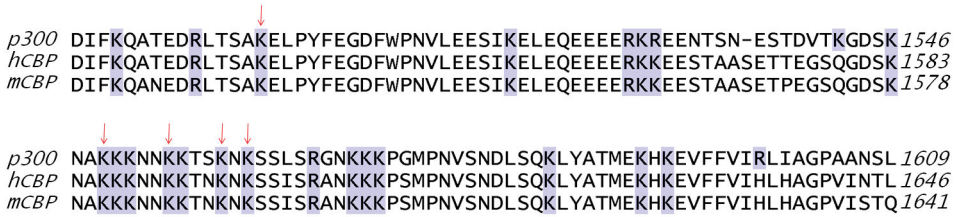

Figure 9. Putative autoacetylation sites of p300/CBP HATs reveals chemical acetylation signature sequences

Sequence alignment of human p300, human CBP, and mouse CBP. Lysine acetylation necessary for catalytic activity indicated by red arrows. For each necessary acetylation site, at least one basic residue is present between the −5 and −3 positions.

Table 1
Validated Chemically Acetylated Proteins

Proteins that have been shown to be acetylated in whole cell acetylomic studies are colored in blue and proteins with identified site specific acetylation sites are colored in purple.

Accession #	Protein	Description	Localization
BC005988	TCEAL7	transcription elongation factor A (SII)-like 7	nucleus
BC001460	LARP1	La ribonucleoprotein domain family, member 1	cytoplasm
BC004880	NME4	non-metastatic cells 4, protein expressed in	mitochondria
BC021181	c16orf78	chromosome 16 open reading frame 78	nucleus
BC020518	PSMD3	proteasome 26S subunit, non-ATPase, 3	cytoplasm, nucleus
BC001631	TMSB4X	thymosin, beta 4, X-linked	cytoplasm, extracellular
BC000264	SCNM1	sodium channel modifier 1	nucleus
BC000521	FSCN1	fascin homolog 1, actin-bundling protein (S. purpuratus)	cytoplasm
BC000354	RPS28	ribosomal protein S28	cytoplasm
BC001827	DTYMK	deoxythymidylate kinase (thymidylate kinase)	cytoplasm, mitochondria
BC005979	UBE2B	ubiquitin-conjugating enzyme E2B	cytoplasm, nucleus
BC006491	MPZ	myelin protein zero	plasma membrane
BC034717	CDK2AP1	CDK2-associated protein 1	cytoplasm, nucleus
BC015832	RPS26	ribosomal protein S26	cytoplasm
BC015537	HBE1	hemoglobin, epsilon 1	cytoplasm
BC015925	ZNF706	zinc finger protein 706	cytoplasm, nucleus
BC012306	MRPL17	Mitochondrial ribosomal protein of the large subunit	mitochondria
BC024240	COX5A	cytochrome c oxidase subunit Va	mitochondria
BC036364	BID	BH3 interacting domain death agonist	mitochondria
BC003541	DOK4	docking protein 4	cytoplasm, mitochondria
BC022064	FATE1	fetal and adult testis expressed 1	ER
BC000908	SMA4	survival of motor neuron 1, telomeric	cytoplasm, nucleus
BC034148	BIRC5	baculoviral IAP repeat-containing 5 (survivin)	nucleus
BC011000	CDCA5	cell division cycle associated 5	cytoplasm, nucleus
BC030984	IGL@	immunoglobulin lambda locus	extracellular
BC003151	VPS72	vacuolar protein sorting 72 (S. cerevisiae)	nucleus
BC069431	KLF9	Kruppel-like factor 9	nucleus
BC001410	S100A11	S100 calcium binding protein A11 (calgizzarin)	cytoplasm, nucleus
BC009707	MRPL40	mitochondrial ribosomal protein L40	mitochondria, nucleus
BC009376	CBX8	chromobox homolog 8	nucleus
BC060801	DPF3	D4, zinc and double PHD fingers, family 3	nucleus
BC093013	MORF4L2	mortality factor 4 like 2	nucleus
BC028574	THEG	testicular haploid expressed gene	nucleus
BC004255	ZSCAN16	zinc finger and SCAN domain containing 16	nucleus
BC011584	CCDC16	coiled-coil domain containing 16	nucleus

Accession #	Protein	Description	Localization
BC000654	KRT8	keratin 8	cytoplasm, nucleus
BC001148	RAB31	RAS oncogene family member	Golgi
BC005187	UFC1	ubiquitin-fold modifier conjugating enzyme 1	cytoplasm, nucleus
BC032852	MAGEB4	melanoma antigen family B, 4	unknown
BC024179	PAK7	p21 protein (Cdc42/Rac)-activated kinase 7	cytoplasm, mitochondria, nucleus
NM_003161	RPS6KB1	ribosomal protein S6 kinase, 70kDa, polypeptide 1	cytoplasm, mitochondria
BC014555	ARHGEF5	Rho guanine nucleotide exchange factor (GEF) 5	cytoplasm
BC002502	BC-2	putative breast adenocarcinoma marker (32kD)	cytoplasm
BC014986	LENG1	leukocyte receptor cluster (LRC) member 1	nucleus
BC001226	PLEK2	pleckstrin 2	cytoplasm
NM_006940	SOX5	SRY (sex determining region Y)-box 5	nucleus
BC000895	SSBP1	single-stranded DNA binding protein 1, mitochondrial	nucleus
BC007852	STK25	serine/threonine kinase 25	cytoplasm, Golgi
BC034784	ZMYND11	zinc finger, MYND domain containing 11	Nucleus