



PCT

(51) International Patent Classification:

C07K 1/107 (2006.01) *A61K 39/395* (2006.01)
C07K 16/18 (2006.01) *A61P 39/02* (2006.01)
C12P 21/08 (2006.01)

(21) International Application Number:

PCT/US2010/053523

(22) International Filing Date:

21 October 2010 (21.10.2010)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

61/253,688 21 October 2009 (21.10.2009) US

(71) Applicant (for all designated States except US): **THE JOHNS HOPKINS UNIVERSITY** [US/US]; 3400 North Charles Street, Baltimore, Maryland 21218 (US).

(72) Inventor; and

(71) Applicant : **COLE, Philip A.** [US/US]; 16 Hambleton Ct., Pikesville, Maryland 21208 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **HUANG, Rong** [CN/US]; 304 Lord Byron Lane, Apt. T3, Cockeysville, Maryland 21030 (US). **HWANG, You-sang** [KR/US]; 2124 Bexley Drive, Woodstock, Maryland 21205 (US).

(74) Agent: **CHILDERS, Jeffrey W.**; 100 N. Charles Street, 5th Floor, Baltimore, Maryland 21201 (US).

(81) Designated States (unless otherwise indicated, for every kind of national protection available): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report (Rule 48.2(g))

(54) Title: SITE-SPECIFIC MODIFICATION OF PROTEINS

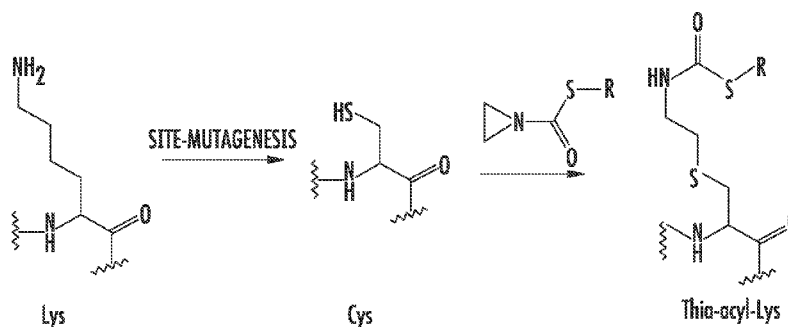


FIG. 1

(57) Abstract: Methods and compositions for site-specific modification of amino acids, peptides, polypeptides and proteins are disclosed.

SITE-SPECIFIC MODIFICATION OF PROTEINS

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No.
5 61/253,688, filed October 21, 2009; which is incorporated herein by reference in its entirety.

FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

This invention was made in part with United States government support under
10 U54 RR020839 awarded by the National Institutes of Health (NIH). The government has certain rights in the invention.

BACKGROUND

Post-translation modifications to proteins regulate protein function and
15 activity, and therefore play a role in cellular processes and disease. To date, more than 200 post-translational modifications have been described on 15 amino acid residues. Walsh, C. T., et al., *Angew. Chem. Int. Ed. Engl.* 44:7342-7372 (2005). The electron-rich and nucleophilic nature of the lysine side chain makes it suitable for undergoing covalent post-translational modification reactions with diverse substrates,
20 including methylation, acetylation, biotinylation, propionylation, butyrylation, ubiquitination, and sumoylation. Chen, Y., et al., *Mol. & Cell. Proteomics* 6: 812-819 (2007).

Reversible lysine acetylation of histones catalyzed by histone acetyltransferases (HATs) and histone deacetylases (HDACs) is recognized as a
25 central mechanism of chromatin regulation. Beyond histones, Lys acetylation has been observed in more than 2,500 mammalian proteins on more than 4,000 sites and has been shown to govern many biological processes. Wang, L., et al., *Curr. Opin. Struct. Biol.* 18:741 (2008); Choudhary, C., et al., *Science* 325: 834 (2009); Wang, Q., et al., *Science* 327:1004 (2010); and Cole, P. A., *Nat. Chem. Biol.* 4:590 (2008).

30 Several approaches are available to explore the structural and functional consequences of protein acetylation. Traditional site-directed mutagenesis is used to replace Lys with Arg or Gln to provide indirect insights into Lys acetylation. Expressed protein ligation and unnatural amino acid mutagenesis via nonsense suppression can install acetyl-Lys (AcK) at desired protein sites, but these approaches

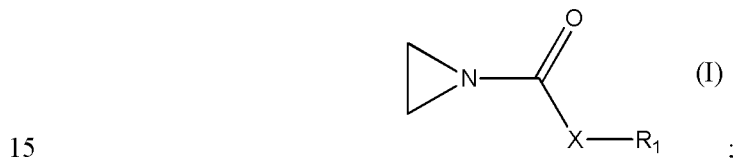
have various technical limitations. Muir, T. W., et al., *Proc. Natl. Acad. Sci. U.S.A.* 95:6705 (1998); Shogren-Knaak, M., et al., *Science* 311:844 (2006); and Neumann, H., et al., *Nat. Chem. Biol.* 4:232 (2008).

5

SUMMARY

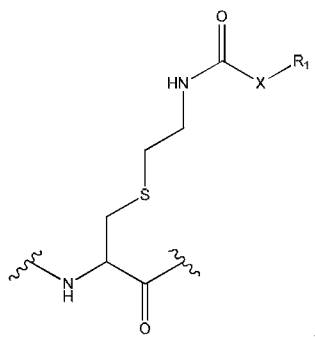
The presently disclosed subject matter provides methods and compositions for site-specific modification of amino acids, peptides, polypeptides and proteins.

In one aspect, the presently disclosed subject matter provides a method for modifying one or more cysteine residues of an amino acid, peptide, polypeptide or protein, the method comprising: providing an amino acid, peptide, polypeptide or
 10 protein comprising one or more cysteine residues; and contacting the amino acid, peptide, polypeptide or protein with a compound of formula (I) for a period of time to modify one or more cysteine residues of the amino acid, peptide, polypeptide or protein:



wherein: X is S or O; R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{O}-\text{R}_2$, and $-\text{O}-(\text{C}=\text{O})-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{OH}$; and wherein: n is an integer from 0 to 20; and R₂ is selected from
 20 the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl; or a pharmaceutically acceptable salt thereof.

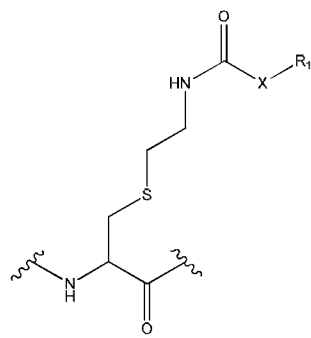
25 In a second aspect, the presently disclosed subject matter provides an isolated amino acid, peptide, polypeptide or protein comprising one or more modified cysteine residues having the following structure:



wherein X and R₁ are as defined hereinabove.

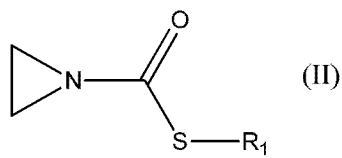
In other aspects, the presently disclosed subject matter provides a method for assaying the effect of modifying at least one lysine residue of a native amino acid, peptide, polypeptide or protein, the method comprising: mutating one or more selected cysteine residues in the native amino acid, peptide, polypeptide or protein to a different amino acid residue to provide a mutated amino acid, peptide, polypeptide or protein; replacing one or more lysine residues with one or more cysteine residues in the mutated amino acid, peptide, polypeptide or protein to provide a mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues; contacting the mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues with a compound of formula (I) for a period of time to modify one or more cysteine residues of the mutated amino acid, peptide, polypeptide or protein; and comparing an activity of the mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues modified with a compound of formula (I) to an activity of the native amino acid, peptide, polypeptide or protein, thereby assaying the effect of modifying at least one lysine residue of the native amino acid, peptide, polypeptide or protein.

In yet other aspects, the presently disclosed subject matter provides a method for producing an antibody to a naturally occurring antigen, the method comprising immunizing a mammal with a polypeptide comprising one or more modified cysteine residues having the following structure:



wherein X and R₁ are as defined hereinabove and wherein the polypeptide is substantially identical to the naturally occurring antigen; and isolating an antibody from the mammal that binds the naturally occurring antigen.

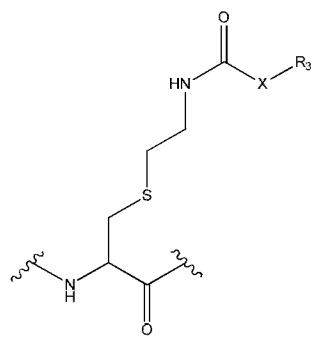
- 5 In further aspects, the presently disclosed subject matter provides a compound of formula II:



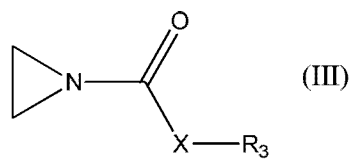
wherein R₁ is as defined hereinabove, or a pharmaceutically acceptable salt thereof.

- 10 In yet further aspects, the presently disclosed subject matter provides a kit comprising a compound of formula (II).

In other aspects, the presently disclosed subject matter provides a method for preparing an antibody-conjugate comprising one or more cysteine residues having the following structure:



- 15 wherein X is S or O; and R₃ is a therapeutic agent or an imaging agent; the method comprising: introducing one or more cysteine residues in an antibody to provide a mutated antibody; and reacting the one or more cysteine residues of the mutated antibody with a compound of formula (III) to prepare an antibody-conjugate:



wherein X is S or O; and R₃ is a therapeutic agent or an imaging agent. In certain aspects, the therapeutic agent is a cytotoxic agent. In certain other aspects, the imaging agent is a fluorescent dye or a radionucleotide.

5 Certain aspects of the presently disclosed subject matter having been stated hereinabove, which are addressed in whole or in part by the presently disclosed subject matter, other aspects will become evident as the description proceeds when taken in connection with the accompanying Examples and Figures as best described herein below.

10

BRIEF DESCRIPTION OF THE FIGURES

Having thus described the presently disclosed subject matter in general terms, reference will now be made to the accompanying Figures, which are not necessarily drawn to scale, and wherein:

15 FIG. 1 is a scheme illustrating an embodiment of the presently disclosed methods for modifying a lysine residue of an amino acid, peptide, polypeptide or protein;

FIGS. 2a-2d show (a) acetyl-Lys (AcK) and acetyl-thiaLys analogue; (b) strategies attempted to install acetyl-Lys analogues, including a prior art strategy (top
20 scheme) and presently disclosed strategy (bottom scheme); (c) competition assay of methylthiocarbonyl-aziridine (MTCA)-modified H4 derived peptides by tetracetylated N-terminal H4 tail. Extracts from GFP-Brdt expressing cells (GFP-Brdt is a recombinant fusion protein composed of green fluorescent protein linked to the bromodomain from testis) were preincubated with peptides and then used in a pull-
25 down assay with immobilized tetracetylated H4 tail or the same nonmodified peptide (- lane). The GFP-Brdt retained on the beads was then revealed using an anti-GFP antibody. C* corresponds to methylthiocarbonyl-thiaLys (MTCTK); and (d) Western blots with site-specific anti-AcK9 Ab on MTCA-treated mutant H3;

FIGS. 3a and 3b are mass spectra (MALDI-TOF) for reaction of N-acetyl-
30 aziridine with peptide Ala-Thr-Arg-Cys-Gln-Thr-Ala (ATRCQTA) (SEQ. ID. NO. 1) derived from the N-terminus of histone H3 with Cys at residue 4 instead of Lys: (a)

MALDI-TOF mass spectrum of ATRCQTA (SEQ. ID. NO. 1) after addition of acetonitrile without alkylating reagent; and (b) MALDI-TOF mass spectrum taken after reaction of ATRCQTA (SEQ. ID. NO. 1) with 40 mM of N-acetyl-aziridine (0.5 M stock in acetonitrile);

5 FIGS. 4a-4c are mass spectra (MALDI-TOF) of p53 peptide variants with or without MTCA: (a) MALDI-TOF mass spectrum of p53 peptide KKGQSTSRHKCLMFKTEG (SEQ. ID. NO. 2) derived from the C-terminus (amino acid residue 372-389) of the human p53 tumor suppressor protein after addition of acetonitrile. Calculated p53 peptide m/z ($M+H^+$) = 2066; (b) mass spectrum of
10 reaction of p53 peptide KKGQSTSRHKCLMFKTEG (SEQ. ID. NO. 2) with methylthiocarbonyl-aziridine. The peptide was dissolved in 100 mM ammonium bicarbonate buffer (pH 8.0) and reacted with 100 mM methylthiocarbonyl-aziridine (MTCA, from 1 M acetonitrile stock) for 3 hrs at room temperature. The crude reaction mixture was lyophilized and subjected to MS analysis. Calculated p53
15 peptide m/z ($M+H^+$) = 2183. No evidence of disulfide is observed; and (c) mass spectrum of reaction of methylthiocarbonyl-aziridine with p53 peptide KKGQSTSRHKCLMFKTEG (SEQ. ID. NO. 3) following the condition described in (b). Calculated p53 peptide m/z ($M+H^+$) = 2093;

 FIG. 5a-5c show specific reactivity of methylthiocarbonyl-aziridine with a Cys
20 residue within different peptides: (a) mass spectra (MALDI-TOF) taken of reactions of peptide ATRCQTA (SEQ. ID. NO. 1) with or without 100 mM methylthiocarbonyl-aziridine in 100 mM ammonium bicarbonate buffer (pH 8.0); (b) mass spectra (MALDI-TOF) for reactions of peptide Ahx-IRRFQCSTELL-NH₂ peptide (SEQ. ID. NO. 4) (derived from histone H3 residues 52-62) with or without
25 100 mM methylthiocarbonyl-aziridine in 100 mM ammonium bicarbonate buffer (pH 8.0). Ahx-aminohexanoic acid; and (c) mass spectra (MALDI-TOF) taken of reactions of commercial peptide PHCKRM (SEQ. ID. NO. 5) with or without 100 mM methylthiocarbonylaziridine in 100 mM ammonium bicarbonate buffer (pH 8.0);

 FIGS. 6a-6d are mass spectra (MALDI-TOF) of peptides derived from the N-
30 terminus of histone H4 with Cys at residue 5 and 8 instead of Lys: (a) MALDI-TOF mass spectrum of SGRGAcKGGCGLGK-NH₂ (SEQ. ID. NO. 6); (b) MALDI-TOF mass spectrum of SGRGCGGAcKGLGK-NH₂ (SEQ. ID. NO. 7); and (c) MALDI-TOF mass spectrum taken after reaction of SGRGAcKGGCGLGK-NH₂ (SEQ. ID. NO. 6) with 100 mM MTCA in 100 mM ammonium bicarbonate buffer (pH 8.0); and

(d) MALDI-TOF mass spectrum taken after reaction of MTCA with SGRGCGGAcKGLGK-NH₂ (SEQ. ID. NO. 7) as shown in (c);

FIGS. 7a and 7b show a competition pulldown assay of H4 derived peptides by tetracetylated N-terminal H4 tail: (a) competition pulldown assay of H4C5*K8ac. C5* refers to an MTCTK at the 5 position. Extracts from GFP-Brdt-expressing cells (GFP-Brdt is a recombinant fusion protein composed of green fluorescent protein linked to the bromodomain from testis) were pre-incubated with increasing amounts of the indicated peptides and then used in a pull-down assay with immobilized (via a biotin group) tetracetylated N-terminal H4 (K5ac/K8ac/K12ac/K16ac). H4 K5acK8ac is a doubly acetylated 12-mer N-terminal H4 tail peptide with AcK at the 5 and 8 positions; H4K5C*K8ac is the identical peptide with MTCTK at the 5-position in place of AcK (see FIG. 6); H4 is an unmodified H4 tail peptide. The GFP-Brdt retained on the streptavidin beads was then revealed using an anti-GFP antibody. "Input" corresponds to 10% of the material used for each pull-down; (b) competition pulldown assay of H4 peptides K5ac or K8ac (monoacetylated with AcK at 5 or 8 position, sequences as in FIG. 6). "-" corresponds to a pull-down assay with non-modified peptide;

FIGS. 8a-8c show analytical RP-HPLC analysis of HDAC8 assay: (a) HPLC chromatogram of co-injection of p53 peptide containing methylthiocarbonyl-thiaLys (MTCTK) at residue 382 and its theoretical deacetylated peptide (10:1 by weight); (b) HPLC chromatogram of p53 peptide containing MTCTK at 382 position from a HDAC8 assay. Peptide (0.3 mM) was incubated with 3 mM HDAC8 for 20 min at 37°C. The reactions were quenched with stop buffer (1 M HCl, 0.16 M acetic acid), followed by addition of 8 M urea (10% v/v). The mixtures were centrifuged and the supernatant was analyzed by RP-HPLC. It should be noted that the lack of precise alignment between the MTCTK-peptide peaks in (a) and (b) likely represent shifts in the column retention due to different sample handling and different states of the C18 column used at various times; and (c) HPLC chromatogram of p53 peptide containing acetyl-Lys at the 382 position from an HDAC8 assay using same conditions as described in (b);

FIGS. 9a-9c show an analytical RP-HPLC analysis of Sir2Tm assay: (a) HPLC chromatogram of co-injection of p53 peptide containing MTCTK at position 382 and its theoretical deacetylated peptide (10:1 by weight); (b) HPLC chromatogram of p53 peptide containing MTCTK at 382 position from a Sir2Tm

(from *Thermatoga maritima*) assay. Peptide (0.3 mM) was incubated with 0.5 mM Sir2Tm for 20 min at 37°C. The reactions were quenched with stop buffer (0.1 M HCl, 0.16 M acetic acid) followed by addition of 8 M urea (10% v/v). The mixtures were centrifuged and the supernatant was analyzed by RP-HPLC. It should be noted
 5 that the lack of precise alignment between the MTCTK-peptide peaks in (a) and (b) likely represent shifts in the column retention due to different sample handling and different states of the C18 column used at time of assay; and (c) HPLC chromatogram of p53 peptide containing acetyl-Lys at the 382 position from a Sir2Tm assay. The slightly different retention times compared with FIG. 8 likely
 10 represent different states of the C18 column used at these different times;

FIGS. 10a-10d show installation of the presently disclosed acetyl-Lysine mimic into histones: (a) intact mass spectra (ESI) of histone H3 C110A/K9C with (right panel) or without (left panel) reacting with MTCA. Calculated unmodified histone H3 m/z (M+H) = 15,214; calculated MTCA-modified histone H3 m/z (M+NH₄⁺) = 15,348 (estimated standard error +/- 3 mass units); (b) Western blot
 15 with site-specific anti-AcK histone H3 antibodies. Both wild type full-length recombinant *Xenopus* histone H3 (1 mg/mL) and mutant H3 (1 mg/mL; mutant H3 also contain C110A) were dissolved in 100 mM NH₄HCO₃ (pH 8.0) and treated with or without methylthiocarbonyl-aziridine (MTCA, 100 mM). Excess alkylating
 20 reagent was quenched by addition of 2-mercaptoethanol after reacting for 3 hrs at room temperature. The reaction mixtures were lyophilized and subjected to western blot. The coomassie stain of histones appears below; (c) Western blot of p300-acetylated histone with antibodies for AcK9, AcK18, and AcK27. 8 µg of purified recombinant H3 protein was incubated with 1 µM of p300 and 500 µM acetyl-CoA in
 25 50 µL reaction buffer at 30°C for 30 min. The reaction mixture was quenched by addition of loading dye and subjected to Western blot. The Coomassie stain of histones appears in the lower panel. This panel serves as a positive control for the anti-AcK Abs used here; and (d) absence of crossreactivity between antibodies for AcK9 and AcK27 for MTCA-modified H3K9C and H3K27C. This panel shows that,
 30 despite the related sequences surrounding the H3 K9 and K27 positions, selectivity could be achieved in antibody recognition;

FIG. 11 show acetylation of histone H3 by Rtt109/Vps75, Rtt109 K290C/Vps75 and Rtt109 K290R/Vps75. The wild type and mutant Rtt109/Vps75 heterodimer proteins were co-expressed and purified from *E. coli* and assayed using

full length recombinant histone H3 as substrate and ^{14}C acetyl-CoA. The Rtt109/Vps75 heterodimer is known to acetylate K9 in histone H3, as well as other Lys sites including K56. Acetylation rates were measured using phosphorimage analysis standardized to ^{14}C -BSA. Enzyme activity was measured at acetyl-CoA [21 μM] and 8 μM of full-length recombinant *Xenopus* histone H3;

FIGS. 12a and 12b show (a) HAT activity of Rtt109 K290C and RTT109 K290R before and after MTCA treatment (autoradiograph below); and (b) kinase activity of CK2 α (carrying C147A and C220A mutations) containing either Cys102 or Lys102 after MTCA treatment; and

FIG. 13 shows kinase activity of CK2 α and mutants with no MTCA treatment. CK2 α is the common catalytic subunit of CK2 protein kinase which phosphorylates many protein and peptide substrates on Ser and Thr residues. The kinases were assayed with 5 nM enzyme, 100 μM ATP, and 40 μM peptide substrate (biotinylated $\text{NH}_2\text{-RRRADDSDDDDK-NH}_2$) (SEQ. ID. NO. 8) for 6 min at 30°C. Assays employed a $\gamma^{32}\text{P}$ -ATP substrate and the radioactivity transferred to peptide was quantified by binding to avidin which was separated from low molecular weight compounds by ultrafiltration. CK2 WT contains no point mutations. CK2 K102* contains point mutations to remove natural cysteines (C147A, C220A). CK2 C102* contains point mutations to remove natural cysteines and to install the acetyl lysine mimic (C147A, C220A, K102C).

DETAILED DESCRIPTION

The presently disclosed subject matter now will be described more fully hereinafter with reference to the accompanying Figures, in which some, but not all embodiments of the presently disclosed subject matter are shown. Like numbers refer to like elements throughout. The presently disclosed subject matter may be embodied in many different forms and should not be construed as limited to the embodiments set forth herein; rather, these embodiments are provided so that this disclosure will satisfy applicable legal requirements. Indeed, many modifications and other embodiments of the presently disclosed subject matter set forth herein will come to mind to one skilled in the art to which the presently disclosed subject matter pertains having the benefit of the teachings presented in the foregoing descriptions and the associated Figures. Therefore, it is to be understood that the presently disclosed

subject matter is not to be limited to the specific embodiments disclosed and that modifications and other embodiments are intended to be included within the scope of the appended claims.

5 I. SITE-SPECIFIC INTRODUCTION OF AN ACETYL-LYSINE MIMIC
INTO PEPTIDES AND PROTEINS BY CYSTEINE MODIFICATION

A. Overview

Protein acetylation on Lys residues is recognized as a significant post-translational modification in cells, but it is often difficult to discern the direct
10 structural and functional effects of individual acetylation events. In some
embodiments, the presently disclosed subject matter describes the use of
methylthiocarbonyl-aziridine (MTCA) to install acetyl-Lys mimics site-specifically
into peptides and proteins by alkylation of cysteine residues (see FIG. 1). Further, the
presently disclosed subject matter demonstrates that the resultant thiocarbamate
15 modification can be recognized by the Brdt bromodomain and site-specific antiacetyl-Lys
antibodies; is resistant to histone deacetylase cleavage; and can confer activation
of the histone acetyltransferase Rtt109 by simulating autoacetylation. This approach
also can be used to obtain functional evidence that acetylation of CK2 protein kinase
on Lys102 can stimulate its catalytic activity.

20 As an analogue of methyl-Lys, methyl-thiaLys can be introduced at targeted
protein locations via a strategy involving Cys alkylation with N-methyl-
aminoethylbromide derivatives. Simon, M. D., et al., *Cell* 128:1003 (2007). To
create an acetyl-thiaLys analogue (FIG. 2a), use of the corresponding acetamide
reagent with Cys-containing peptide was attempted; however, no significant
25 conversion was observed. Without wishing to be bound to any one particular theory,
it is thought that the reduced ability of the amide compounds to form three-membered
ring intermediates and/or intramolecular imidate formation led to diminished activity
of *N*-acetyl-aminoethylbromide. The reaction of *N*-acetyl-aziridine and a Cys-
containing peptide was then explored (FIG. 2b). The predominant product observed
30 by mass spectrometry was *M* +42, which suggests that acetyl-aziridine treatment led
to an undesired acetyl transfer by nucleophilic attack at the carbonyl rather than the
aziridine methylene (FIG. 3).

To reduce reactivity at the carbonyl, the use of methylthiocarbonyl-aziridine
(MTCA) was explored (FIG. 2b). Cys alkylation with this reagent was investigated to

provide methylthiocarbonyl-thiaLys (MTCTK), a thiocarbamate analogue of AcK (FIG. 2b). It was thought that this thiocarbamate could preserve key functional features of AcK, but be resistant to hydrolysis by deacetylases. This property could be beneficial in cell extracts. In the presence of aqueous buffer (pH 8.0) with 100 mM MTCA, mass spectrometric analysis revealed essentially complete conversion of six different Cys-containing peptides with a wide variety of residues to the desired products (FIGS. 4-6). In a Cys-free peptide control, no modification was detected, indicating high selectivity of MTCA for the Cys thiol (FIG. 4).

Some AcK residues mediate bromodomain interaction. To explore the potential of MTCTK to mimic this aspect of AcK function, a pulldown assay was used to assess the binding efficiency of the Brdt bromodomain with H4 acetylated tails containing MTCTK groups in place of AcK at the K5 and K8 positions. Moriniere, J., et al., *Nature* 461:664 (2009). The presently disclosed subject matter confirmed that unacetylated or monoacetylated H4 tail peptides do not compete with the Brdt binding to a tetracetylated H4 peptide, whereas an excess of untagged diacetylated AcK5/AcK8 H4 tail peptide can efficiently block the Brdt pull-down (FIG. 2c and FIG. 7). Further, the presently disclosed subject matter shows that peptides containing substitution of either AcK5 or AcK8 with the presently disclosed MTCTK modification also can compete with the pulldown of Brdt, although approximately 2-to 4-fold less efficiently than the natural AcK-containing peptides (FIG. 2c and FIG. 7). These data show that Brdt's first AcK bromodomain binding pocket can accommodate MTCTK interaction at two sites.

To explore the susceptibility of the methylthiocarbonyl linkage to deacetylase activity, the reactivity of MTCTK was compared to that of an AcK-containing substrate in the context of a p53 peptide with the class I histone deacetylase HDAC8 and the class III NAD-dependent deacetylase Sir2 using an HPLC readout (FIGS. 8-9). Fatkins, D. G., et al., *Bioorg. Med. Chem. Lett.* 16:3651 (2006). In contrast to the AcK p53 peptide, no thiocarbamate removal was detected under similar reaction conditions. These data suggest that the presently disclosed AcK mimic, i.e., MTCTK, could be useful in the presence of deacetylase-rich cell extracts where the consequences of a "constitutive" acetylation state could be explored.

The reactivity of recombinant histone H3 mutants with MTCA was further investigated. Natural *Xenopus* histone H3 has a single Cys (C110), and so three histone mutant proteins were prepared that replaced this site with Ala and introduced

a Cys at the known AcK sites K9, K18, and K27. Alkylation reactions were monitored with site-specific anti-AcK antibodies (Abs) by Western blot. MTCA treatment of C110A/K9C histone H3 led to efficient recognition by anti-AcK9 Ab whereas wt H3 displayed background detection under similar reaction conditions (FIG. 2d). These results suggest that modification of Cys9 in C110A/K9C H3 was efficient, as confirmed by MS (FIG. 10a), and further demonstrate that MTCTK is a reasonable surrogate for AcK in mediating Ab interaction. In a similar fashion, it was shown that MTCA treatment of C110A/K18C and C110A/K27C histone H3 gave rise to efficient recognition by anti-AcK18 Ab and anti-AcK27 Ab, respectively (FIG. 10).

Several HATs have been suggested to undergo autoacetylation as a potential mode of regulation. Wang, L., et al., *Opin. Struct. Biol.* 18:741 (2008); Thompson, P. R., et al., *Nat. Struct. Mol. Biol.* 11:308 (2004); Tang, Y., et al., *Nat. Struct. Mol. Biol.* 15:738 (2008); Stavropoulos, P., et al., *Proc. Natl. Acad. Sci. U.S.A.* 105:12236 (2008). In particular, the yeast HAT Rtt109 has been shown recently to be self-acetylated on active site Lys290. *Id.* Mutation of this residue to Arg has been demonstrated to reduce catalytic activity, Thompson, P. R., et al., *Nat. Struct. Mol. Biol.* 11:308 (2004); Tang, Y., et al., *Nat. Struct. Mol. Biol.* 15:738 (2008); Stavropoulos, P., et al., *Proc. Natl. Acad. Sci. U.S.A.* 105:12236 (2008). This observation also was confirmed by the presently disclosed subject matter (FIG. 11).

In other embodiments, to examine whether introduction of the AcK mimic at this site would stimulate catalysis, a Cys was introduced at position 290. The Rtt109 K290C mutant showed essentially identical HAT activity to the K290R mutant (FIGS. 12a and 11). Since the sequences of Rtt109 and its heterodimeric partner Vps75 have several natural Cys residues, multiple alkylations of these proteins are likely, which could complicate interpretation. To account for this possibility, the HAT activities of Rtt109 K290R and K290C mutants treated with MTCA were directly compared under identical conditions. A reproducible 4-fold relative increase in HAT activity was observed after MTCA treatment of the Rtt109 K290C enzyme compared with the treated Rtt109 K290R enzyme (FIG. 12a). It should be noted that the 4-fold activation by alkylation of Rtt109 K290C does not fully recapitulate the 20-fold rate differential between wt and K290R Rtt109 (FIG. 12), and this observation could be related to multiple Cys alkylation in the proteins or imperfect mimicry of AcK by the MTCTK group. Nevertheless, this partial recovery of activity by introduction of the

AcK analogue can be interpreted as evidence that autoacetylation of Rtt109 on K290 stimulates its HAT activity.

In yet further embodiments, MCTA alkylation was used to explore the effect of the recently discovered acetylation of protein kinase CK2 α on Lys102. Choudhary, C., et al., *Science* 325:834 (2009); Pinna, L. A. and Allende, J. E. *Cell. Mol. Life Sci.* 66:1795 (2009). CK2 α has two natural Cys residues (147, 220), which were mutated to Ala, and Lys102 was further replaced with Cys. Kinase assays showed that each of these mutants had essentially identical kinase activity to wild-type CK2 α (FIG. 13). MTCA alkylation and kinase assays revealed that MTCA-treated Cys102 showed about 2-fold higher catalytic activity compared to MTCA-treated Lys102 (FIG. 12b), providing functional evidence for the acetylation-dependent activation of this important protein kinase. Pinna, L. A. and Allende, J. E. *Cell. Mol. Life Sci.* 66:1795 (2009).

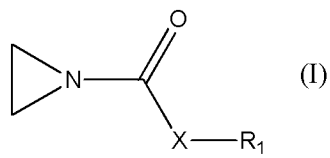
In summary, the presently disclosed subject matter describes an approach to install AcK mimics into recombinant proteins. While the presently disclosed MTCTK is not identical to the AcK structure, it appears to maintain aspects of the molecular recognition by binding proteins and in enzymatic regulation. The thiocarbamate moiety appears resistant to HDAC cleavage, which may provide utility in the complex environments found in transcriptional and chromatin assays. Lu, W., et al., *Mol. Cell* 8:759 (2001).

Accordingly, along with other approaches, Wang, L., et al., *Curr. Opin. Struct. Biol.* 18:741 (2008); Choudhary, C., et al., *Science* 325:834 (2009); Wang, Q., et al., *Science* 327:1004 (2010); Cole, P. A. *Nat. Chem. Biol.* 4:590 (2008); Muir, T. W., et al., *Proc. Natl. Acad. Sci. U.S.A.* 95:6705 (1998); Shogren-Knaak, M., et al., *Science* 311:844 (2006); Neumann, H., et al., *Nat. Chem. Biol.* 4:232 (2008), the presently disclosed Cys modification strategy provides methods and tools for dissecting the diverse and important functions of acetylation in the proteome.

B. Methods of Modifying One or More Cysteine Residues of an Amino Acid, Peptide, Polypeptide or Protein

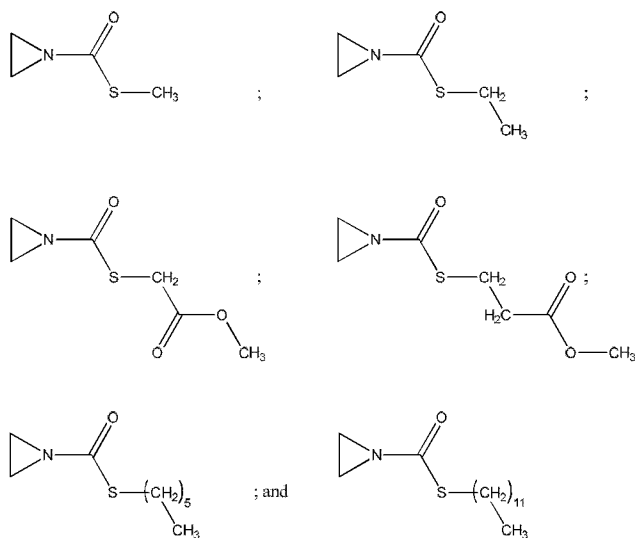
In some embodiments, the presently disclosed subject matter provides a method for modifying one or more cysteine residues of an amino acid, peptide, polypeptide, or protein, the method comprising: providing an amino acid, peptide, polypeptide, or protein comprising one or more cysteine residues and contacting the amino acid, peptide, polypeptide, or protein with a compound of formula (I) for a

period of time to modify one or more cysteine residues of the amino acid, peptide, polypeptide, or protein:



wherein: X is S or O; R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{O}-\text{R}_2$, and $-\text{O}-(\text{C}=\text{O})-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{OH}$; and wherein: n is an integer from 0 to 20; and R₂ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl; or a pharmaceutically acceptable salt thereof.

In some embodiments of the presently disclosed method, X is S and R₁ is selected from the group consisting of unsubstituted, straight-chain C₁ to C₁₂ alkyl, $-(\text{CH}_2)-(\text{C}=\text{O})-\text{O}-\text{CH}_3$, $-(\text{CH}_2)_2-(\text{C}=\text{O})-\text{O}-\text{CH}_3$, and $-\text{O}-(\text{C}=\text{O})-(\text{CH}_2)_2-(\text{C}=\text{O})-\text{OH}$. In particular embodiments, the compound of formula (I) is selected from the group consisting of:



In some embodiments, the presently disclosed method further comprises replacing one or more selected lysine residues of an amino acid, peptide, polypeptide or protein with one or more cysteine residues to provide an amino acid, peptide,

polypeptide or protein comprising one or more selected cysteine residues. In yet further embodiments, the presently disclosed method further comprises mutating one or more cysteine residues of a native amino acid, peptide, polypeptide or protein to a different amino acid residue prior to replacing the one or more selected lysine
5 residues with one or more selected cysteine residues.

Accordingly, in some embodiments, the presently disclosed subject matter provides a method for modifying at least one lysine residue of a native amino acid, peptide, polypeptide or protein the method comprising: (a) mutating one or more selected cysteine residues in the native amino acid, peptide, polypeptide or protein to
10 a different amino acid residue to provide a mutated amino acid, peptide, polypeptide or protein; (b) replacing one or more lysine residues with one or more cysteine residues in the mutated amino acid, peptide, polypeptide or protein to provide a mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues; and (c) contacting the mutated amino acid, peptide, polypeptide or protein
15 comprising one or more cysteine residues with a compound of formula (I) for a period of time to modify one or more cysteine residues of the mutated amino acid, peptide, polypeptide or protein, thereby modifying at least one lysine residue of the native amino acid, peptide, polypeptide or protein.

While the following terms are believed to be well understood by one of
20 ordinary skill in the art, the following definitions are set forth to facilitate explanation of the presently disclosed subject matter. The term "polypeptide" as used herein, refers to two or more amino acids joined to each other by peptide bonds or by modified peptide bonds, i.e., peptide isosteres, and can contain modified amino acids other than the 20 gene-encoded amino acids. The terms "polypeptide," "peptide" and
25 "protein" are used interchangeably herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers.

30 The terms "amino acid" or "amino acid sequence" as used herein refer to an oligopeptide, peptide, polypeptide, or protein sequence, or to a fragment, portion, or subunit of any of these, and to naturally occurring or synthetic molecules. The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the

naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, e.g., hydroxyproline, γ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refer to compounds that have the same basic chemical structure as a naturally occurring amino acid, i.e., an α -carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, e.g., homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (e.g., norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. A representative example of such an analog is aminoethylated cysteine, which functions as a modified lysine residue. Amino acid mimetics refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid.

As to amino acid sequences, one of ordinary skill in the art will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a “conservatively modified variant” where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologues, and alleles of the invention.

Accordingly, a “substantially identical” amino acid sequence is a sequence that differs from a reference sequence by one or more conservative or non-conservative amino acid substitutions, deletions, or insertions, particularly when such a substitution occurs at a site that is not the active site of the molecule, and provided that the polypeptide essentially retains its functional properties. A conservative amino acid substitution, for example, substitutes one amino acid for another of the same class (e.g., substitution of one hydrophobic amino acid, such as isoleucine, valine, leucine, or methionine, for another, or substitution of one polar amino acid for another, such as substitution of arginine for lysine, glutamic acid for aspartic acid or glutamine for asparagine). One or more amino acids can be deleted from a particular polypeptide resulting in modification of the structure of the polypeptide without

significantly altering its biological activity. For example, amino- or carboxyl-terminal amino acids that are not required for a given biological activity can be removed.

The following eight groups each contain amino acids that are conservative substitutions for one another: (1) Alanine (A), Glycine (G); (2) Aspartic acid (D),
5 Glutamic acid (E); (3) Asparagine (N), Glutamine (Q); (4) Arginine (R), Lysine (K);
(5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); (6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W); (7) Serine (S), Threonine (T); and (8) Cysteine (C), Methionine (M) (see, e.g., Creighton, *Proteins* (1984)).

A polypeptide can be modified by natural processes, such as post-translational
10 processing, or by chemical modification techniques, which are well known in the art. Modifications can occur anywhere in the polypeptide, including the peptide backbone, the amino acid side-chains, and the amino or carboxyl termini. It will be appreciated that the same type of modification can be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide can have many types
15 of modifications. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of a phosphatidylinositol, cross-linking cyclization, disulfide bond formation, demethylation, formation of covalent
20 crosslinks, formation of cysteine, formation of pyroglutamate, formylation, propionylation, butyrylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, and transfer-RNA mediated addition of amino acids to
25 protein such as arginylation. (See, for example, Creighton, T. E., *Proteins--Structure and Molecular Properties* 2nd Ed., W.H. Freeman and Company, New York (1993); *Posttranslational Covalent Modification of Proteins*, B. C. Johnson, Ed., Academic Press, New York, pp. 1-12 (1983)).

As used herein, the term "isolated polypeptide" means that the polypeptide is
30 removed from its original environment (e.g., the natural environment if it is naturally occurring). For example, a naturally-occurring polypeptide present in a living animal is not isolated, but the same polypeptide separated from some or all of the coexisting materials in the natural system is isolated. Such polypeptides could be part of a composition and still be isolated in that the composition is not part of its natural

environment. Further, as used herein, the term “purified” does not require absolute purity; rather, it is intended as a relative term.

The terms “recombinant polypeptides” or “recombinant proteins” refer to polypeptides or proteins produced by recombinant DNA techniques; i.e., produced from cells transformed by an exogenous DNA construct encoding the desired polypeptide or protein. “Synthetic polypeptides” or “synthetic proteins” are those prepared by chemical synthesis. Solid-phase chemical peptide synthesis methods also can be used to synthesize such polypeptides. Such methods have been known in the art since the early 1960’s (see e.g., Merrifield, R. B., J. Am. Chem. Soc., 85:2149-2154 (1963); see also Stewart, J. M. and Young, J. D., Solid Phase Peptide Synthesis, 2nd Ed., Pierce Chemical Co., Rockford, Ill., pp. 11-12)) and recently have been employed in commercially available laboratory peptide design and synthesis kits (for example, from Cambridge Research Biochemicals). Such commercially available laboratory kits generally have adopted the teachings of H. M. Geysen et al., Proc. Natl. Acad. Sci., USA, 81:3998 (1984) and provide for synthesizing peptides on the tips of multiple “rods” or “pins,” all of which are connected to a single plate. When such a system is used, a plate of rods or pins is inverted and inserted into a second plate of corresponding wells or reservoirs, which contain solutions for attaching or anchoring an appropriate amino acid to the tips of the pins or rods. By repeating such a process step, i.e., inverting and inserting the rod’s and pin’s tips into appropriate solutions, amino acids are built into desired peptides. In addition, a number of available Fmoc peptide synthesis systems are available. For example, assembly of a polypeptide or fragment can be carried out on a solid support using an Applied Biosystems, Inc. Model 431A automated peptide synthesizer. Such equipment provides ready access to polypeptides, either by direct synthesis or by synthesis of a series of fragments that can be coupled using other known techniques.

The phrase “substantially identical” in the context of two polypeptides refers to two or more sequences that have at least 50%, 60%, 70%, 80%, and in some aspects 90-95% amino acid residue identity when compared and aligned for maximum correspondence, as measured using one of the known sequence comparison algorithms or by visual inspection. Typically, the substantial identity exists over a region of at least about 100 residues, and most commonly the sequences are substantially identical over at least about 150-200 residues. In some embodiments, the sequences are substantially identical over the entire length of the coding regions.

In representative embodiments, the polypeptide modified by the presently disclosed subject methods is selected from the group consisting of a histone, a transcription factor, a nuclear receptor, a cytoskeletal protein; an adaptor protein involved in nuclear transport; a multifunctional protein; a molecular chaperone; and a virally encoded protein.

More particularly, in some embodiments, the amino acid, peptide, polypeptide or protein modified by the presently disclosed methods is a histone. Generally, histones are proteins responsible for packaging and ordering DNA into structural units referred to as nucleosomes and play an important role in gene regulation. Histones generally can be grouped into five major classes or families: H1/H5, H2A, H2B, H3, and H4. These five families of histones can be further organized into two super families as follows: core histones (H2A, H2B, H3, and H4) and linker histones (H1 and H5). The families of histones can be further organized into subfamilies and representative, non-limiting members as follows: H1 (subfamily H1F): H1F0, H1FNT, H1FOO, and H1FX; H1 (subfamily H1H1): HIST1H1A, HIST1H1B, HIST1H1C, HIST1H1D, HIST1H1E, and HIST1H1T; H2A (subfamily H2AF): H2AFB1, H2AFB2, H2AFB3, H2AFJ, H2AFV, H2AFX, H2AFY, H2AFY2, and H2AFZ; H2A (subfamily H2A1): HIST1H2AA, HIST1H2AB, HIST1H2AC, HIST1H2AD, HIST1H2AE, HIST1H2AG, HIST1H2AI, HIST1H2AJ, HIST1H2AK, HIST1H2AL, and HIST1H2AM; H2A (subfamily H2A2): HIST2H2AA3 and HIST2H2AC; H2B (subfamily H2BF): H2BFM, H2BFO, H2BFS, H2BFWT; H2B (subfamily H2B1): HIST1H2BA, HIST1H2BB, HIST1H2BC, HIST1H2BD, HIST1H2BE, HIST1H2BF, HIST1H2BG, HIST1H2BH, HIST1H2BI, HIST1H2BJ, HIST1H2BK, HIST1H2BL, HIST1H2BM, HIST1H2BN, HIST1H2BO; H2B (subfamily H2B2): HIST2H2BE; H3 (subfamily H3A1): HIST1H3A, HIST1H3B, HIST1H3C, HIST1H3D, HIST1H3E, HIST1H3F, HIST1H3G, HIST1H3H, HIST1H3I, HIST1H3J; H3 (subfamily H3A2): HIST2H3C; H3 (subfamily H3A3): HIST3H3; H4 (subfamily H41): HIST1H4A, HIST1H4B, HIST1H4C, HIST1H4D, HIST1H4E, HIST1H4F, HIST1H4G, HIST1H4H, HIST1H4I, HIST1H4J, HIST1H4K, HIST1H4L; and H4 (subfamily H44): HIST4H4.

In other embodiments, non-histone proteins can undergo modification by the presently disclosed methods. See, Glozak, M. A., et al., Acetylation and Deacetylation of Non-Histone Proteins, *Gene* 363:15-23 (2005). Representative examples of such non-histone proteins include, but are not limited to, transcription

factors, such as p53, Yin Yang 1 (YY1), HMG proteins, including the subfamilies HMG-A1/A2, HMG-B1/B2, HMG-N1/N2, STAT3, and c-MYC; and nuclear receptors, such as androgen receptor (AR), estrogen receptor α (ER α), SHP, GATA factors, including GATA1, GATA2, and GATA3, erythroid Kruppel like factor
 5 (EKLF), MyoD, E2F/Rb, NF- κ B, HIF-1 α , and Smad7.

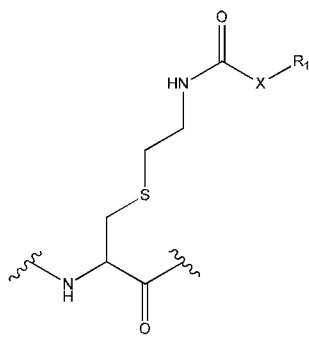
Other classes of non-histone proteins also are regulated by acetylation and deacetylation including, but not limited to, the cytoskeletal protein α -tubulin; importin- α , an adaptor protein involved in nuclear transport; Ku70, a multifunctional protein; and Hsp90, a molecular chaperone.

10 In addition to the cellular proteins described immediately hereinabove, some virally encoded proteins also are substrates of acetylation and deacetylation including, but not limited to, E1A, an adenovirus protein; and HDAG (including L-HDAG and S-HDAG).

In some embodiments, the presently disclosed subject matter provides a
 15 modified amino acid, peptide, polypeptide or protein prepared by the method disclosed immediately hereinabove.

C. Amino Acid, Peptides, Polypeptides or Proteins Comprising One or More Modified Cysteine Residues

In some embodiments, the presently disclosed subject matter provides an
 20 isolated amino acid, peptide, polypeptide or protein comprising one or more modified cysteine residues having the following structure:



wherein: X is S or O; R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl,
 25 substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{O}-\text{R}_2$, and $-\text{O}-(\text{C}=\text{O})-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{OH}$; and wherein n is an integer from 0 to 20; and R₂ is selected from

the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl.

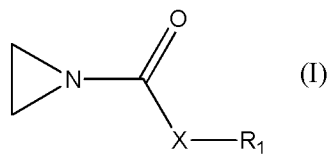
In representative, non-limiting embodiments, amino acids, peptides, polypeptides or proteins modified by the presently disclosed methods include, but are not limited to: p53 peptide KKGQSTSRHKCLMFKTEG (SEQ. ID. NO. 2) derived from the C-terminus (amino acid residue 372-389) of the human p53 tumor suppressor protein; peptide Ala-Thr-Arg-Cys-Gln-Thr-Ala (ATRCQTA) (SEQ. ID. NO. 1) derived from the N-terminus of histone H3 with Cys at residue 4 instead of Lys; peptide Ahx-IRRFQCSTELL-NH₂ peptide (SEQ. ID. NO. 4) (derived from histone H3 residues 52-62) (Ahx=aminohexanoic acid); commercial peptide PHCKRM (SEQ. ID. NO. 5); peptides derived from the N-terminus of histone H4 with Cys at residue 5 and 8 instead of Lys, including, but not limited to, SGRGAcKGGCGLGK-NH₂ (SEQ. ID. NO. 6) and SGRGCGGAcKGLGK-NH₂ (SEQ. ID. NO. 7); H4C5*K8ac, H4K5C*K8ac, where C5* refers to an MTCTK at the 5 position; histone H3 C110A/K9C; Rtt109 K290C and RTT109 K290R, CK2α, H3 derived peptide with biotin tag: QTARKSTGGC*APRKQLATKAK(biotin)-OH (C*=acetyl-Lys mimic) (SEQ. ID. NO. 9); fluorescent labeled NS1 peptide: FAM-RRRRRRRRRGGGKMARTARSC*V-OH (SEQ. ID. NO. 10); and H4 derived peptides: SGRGKGGC^GLGK-OH (SEQ. ID. 11) (C^=succinic modifications with the presently disclosed compounds of formula (I)).

One of ordinary skill in the art upon review of the presently disclosed subject matter would appreciate that the presently disclosed methods are suitable for modifying any cysteine-containing amino acid, peptide, polypeptide, or protein. Thus, the amino acids, peptides, polypeptides or proteins comprising one or more modified cysteine residues are not limited to the examples disclosed herein.

D. Assaying the Effect of Modifying at Least One Lysine Residue of an Amino Acid, Peptide, Polypeptide, or Protein

In some embodiments, the presently disclosed subject matter provides a method for assaying the effect of modifying at least one lysine residue of a native amino acid, peptide, polypeptide or protein, the method comprising: (a) mutating one or more selected cysteine residues in the native amino acid, peptide, polypeptide or protein to a different amino acid residue to provide a mutated amino acid, peptide, polypeptide or protein; (b) replacing one or more lysine residues with one or more

cysteine residues in the mutated amino acid, peptide, polypeptide or protein to provide a mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues; and (c) contacting the mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues with a compound of formula (I) for
 5 a period of time to modify one or more cysteine residues of the mutated amino acid, peptide, polypeptide or protein:



wherein: X is S or O; R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl,
 10 substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{O}-\text{R}_2$, and $-\text{O}-(\text{C}=\text{O})-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{OH}$; and wherein: n is an integer from 0 to 20; and R₂ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or
 15 cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl; or a pharmaceutically acceptable salt thereof; and (d) comparing an activity of the mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues modified with a compound of formula (I) to an activity of the native amino acid, peptide, polypeptide or protein, thereby assaying the effect of modifying at least one lysine
 20 residue of the native amino acid, peptide, polypeptide or protein.

Such assays could include, but are not limited to, investigating protein-protein interactions, protein-nucleic interactions, protein-lipid interactions, protein carbohydrate-interactions, protein-small molecule interactions, enzyme activities, transcription assays, cell-based assays, and the like.

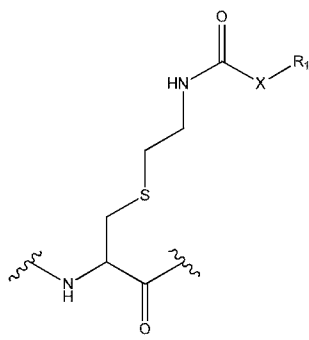
25 *E. Preparation of Antibodies Using Amino Acids, Peptides, Polypeptides, or Proteins Comprising One or More Modified Cysteine Residues*

As used herein, the term “antibodies” refers to proteins that exhibit binding specificity to a specific antigen. Native antibodies are usually heterotetrameric glycoproteins of about 150,000 daltons, composed of two identical light (L) chains
 30 and two identical heavy (H) chains. Each light chain is linked to a heavy chain by one covalent disulfide bond, while the number of disulfide linkages varies between the

heavy chains of different immunoglobulin isotypes. Each heavy chain has at one end a variable domain followed by a number of constant domains. Each light chain has a variable domain at one end and a constant domain at its other end; the constant domain of the light chain is aligned with the first constant domain of the heavy chain,
 5 and the light chain variable domain is aligned with the variable domain of the heavy chain.

Further, as used herein, the term “antibody” is used in the broadest sense and encompasses single monoclonal antibodies, antibody compositions with polyepitopic specificity (i.e., polyclonal antibody compositions), chimeric antibodies, humanized
 10 antibodies, single-chained antibodies, and antibody fragments (e.g., Fab, F(ab'), Fv).

In some embodiments, the presently disclosed subject matter provides a method for producing an antibody to a naturally occurring antigen, the method comprising immunizing a mammal with a polypeptide comprising one or more modified cysteine residues having the following structure:



15

wherein: X is S or O; R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{O}-\text{R}_2$, and $-\text{O}-(\text{C}=\text{O})-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{OH}$; and wherein n is an integer from 0 to 20; and R₂ is selected from
 20 the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl; and wherein the polypeptide is substantially identical to the naturally occurring antigen and isolating an antibody
 25 from the mammal that binds the naturally occurring antigen.

Methods for producing antibodies are well-known in the art. For example, polyclonal antibodies to a particular antigen of interest can be raised in animals.

Accordingly, in some embodiments, polyclonal antibodies can be generated by multiple subcutaneous (sc) or intraperitoneal (ip) injections of a polypeptide comprising one or more acetylated cysteine residues, or a fragment thereof, and an adjuvant. Under certain conditions, it may be useful to conjugate the polypeptide, or
5 fragment thereof, to a protein that is immunogenic in the species to be immunized, e.g., keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, or soybean trypsin inhibitor using a bifunctional or derivatizing agent, for example maleimidobenzoyl sulfosuccinimide ester (conjugation through cysteine residues), N-hydroxysuccinimide (through lysine residues), and glutaraldehyde. Aggregating
10 agents such as alum may also be used to enhance the immune response.

In other embodiments, monoclonal antibodies can be made using the hybridoma method first described by Kohler *et al.*, (1975), *Nature*, 256:495), or can be made by recombinant DNA methods (see e.g., U.S. Patent No. 4,816,567).

In the hybridoma method, a mouse or other appropriate host animal, such as
15 hamster, is immunized as described hereinabove to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the protein used for immunization. Alternatively, lymphocytes can be immunized *in vitro*. Lymphocytes then are fused with myeloma cells using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma (Goding, *Monoclonal Antibodies: Principles and*
20 *Practice*, pp.59-103 (Academic Press, 1986)).

The hybridoma cells thus prepared are seeded and grown in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, parental myeloma cells. Culture medium in which hybridoma cells are growing can be assayed for production of monoclonal antibodies that bind
25 the immunogen. Preferably, the binding specificity of monoclonal antibodies produced by hybridoma cells is determined by immunoprecipitation or by an *in vitro* binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). After hybridoma cells are identified that produce antibodies of the desired specificity, the clones may be subcloned by limiting dilution procedures and
30 grown by standard methods. Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium or RPMI-1640 medium. In addition, the hybridoma cells may be grown *in vivo* as ascites tumors in an animal.

Preferably, the monoclonal antibodies secreted by the subclones are separated from the culture medium, ascites fluid, or serum by conventional immunoglobulin

purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

DNA encoding the monoclonal antibodies can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells serve as a preferred source of such DNA. Once isolated, the DNA may be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA also may be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences, (see Morrison *et al.*, (1984), *Proc. Nat. Acad. Sci.*, 81:6851), or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide.

Further, methods for humanizing non-human antibodies are well known in the art. Generally, a humanized antibody has one or more amino acid residues introduced into it from a non-human antibody. Humanization can be essentially performed following the method of Winter and co-workers (Jones *et al.*, (1986), *Nature*, 321:522-525; Riechmann *et al.*, (1988), *Nature*, 332:323-327; Verhoeven *et al.*, (1988), *Science*, 239:1534-1536), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody.

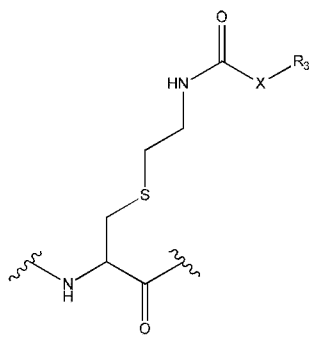
The choice of human variable domains, both light and heavy, to be used in making the humanized antibodies is an important consideration in reducing antibody antigenicity. In the "bestfit" method, the sequence of the variable domain of a non-human antibody is screened against the entire library of known human variable domain sequences. The human sequence which is closest to that of the non-human sequence is then accepted as the human framework (FR) for the humanized antibody (Sims *et al.*, (1993), *J. Immunol.*, 151:2296; Chothia *et al.*, (1987), *J. Mol. Biol.*, 196:901). In another method, a particular framework derived from the consensus sequence of all human antibodies of a particular subgroup of light or heavy chains may be used. The same framework may be used for several different humanized antibodies (Carter *et al.*, (1992), *Proc. Natl. Acad. Sci. USA*, 89:4285; Presta *et al.*, (1993), *J. Immunol.* 151:2623).

F. Antibody-Conjugates Comprising Cysteine Residues Modified by Substituted Thiocarbonyl-Aziridine Compounds

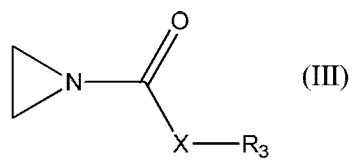
Antibodies have been conjugated to a variety of cytotoxic agents, including small molecule drugs that alkylate DNA, e.g., duocarmycin and calicheamicin; disrupt
 5 microtubules, e.g., maytansinoids and auristatins; or bind DNA, e.g., anthracyclins. Such antibody-drug conjugates have displayed potent and selective killing of target tumor cells *in vitro* and in mouse tumor xenograft studies. Junutula, J. R., et al., Site-specific Conjugation of a Cytotoxic Drug to an Antibody Improves the Therapeutic Index, *Nature Biotechnology* 26 (8):925-932 (2008).

10 Cytotoxic drugs generally can be conjugated to antibodies either through lysine side-chain amines or through cysteine sulphydryl groups activated by reducing interchain disulfide bonds. Both of these techniques, however, yield heterogeneous products, which comprise a mixture of species having different molar ratios of drug to antibody linked at different sites and each having distinct *in vivo* pharmacokinetic,
 15 efficacy, and safety profiles. The presently disclosed subject matter provides a conjugation method for modifying cysteine residues at specific sites in antibodies to allow therapeutic agents or imaging agents to be conjugated with defined stoichiometry without disruption of interchain disulfide bonds.

Accordingly, in some embodiments, the presently disclosed subject matter
 20 provides an antibody-conjugate comprising one or more cysteine residues having the following structure:



wherein: X is S or O; and R₃ is a therapeutic agent or an imaging agent. Such antibody-conjugates can be prepared by introducing a cysteine residue to an antibody
 25 to provide a mutated antibody and then reacting the cysteine residue of the mutated antibody with a compound of formula (III) to prepare an antibody-conjugate:



wherein: X is S or O; and R₃ is a therapeutic agent or an imaging agent.

In some embodiments, R₃ is a therapeutic agent. Therapeutic agents suitable for use with the presently disclosed antibody-conjugate include, but are not limited to, those pharmaceutical compounds developed for use in the treatment of cancer, arthritis, septicemia, myocardial arrhythmia's and infarctions, viral and bacterial infections, autoimmune disease and prion diseases. In particular embodiments, the therapeutic agent comprises a cytotoxic agent (i.e., an agent that impairs the viability and/or the functions of a cell). In such embodiments, the antibody-conjugate can be administered to a subject (e.g., a mammal, such as a human) having or at risk of having a disease or condition, so as to treat (e.g., alleviating, mitigating, reducing, preventing, postponing the onset of) the disease.

When the antibody-conjugate of interest is to be delivered to treat a particular disease, disorder, or condition, the therapeutic agent can be selected for the particular disorder. For example, where the antibodies are targeted to a unique tumor antigen found on a tumor cell at a specific tumor site, the antibodies can be conjugated to an anti-tumor agent for specific delivery to that site and to minimize or eliminate collateral pathology to normal tissue. The agent can be delivered to a specific target ligand recognized by the surrogate molecule and found specifically at the tumor site.

The therapeutic agents can be any type of anti-tumor or anti-angiogenic compound (i.e., an agent that disrupts the vasculature supplying a tumor) that can be attached to the antibody, and can include, for purpose of example, synthetic or natural compounds such as a cytotoxin, an interleukin, a chemotactic factor, a therapeutic radionucleotide, a DNA alkylating agent (e.g., duocarmycin and calicheamicin), an agent that disrupts microtubules (e.g., maytansinoids and auristatins), an agent that binds DNA (e.g., anthracyclins), methotrexate, cis-platin, anastrozole/Arimidex and tamoxifen.

Non-limiting examples of cytotoxic agents include diphtheria toxin, Pseudomonas exotoxin, ribosome inactivating proteins, ricin A, deglycosylated ricin A chain, abrin, alpha sarcin, aspergillin, restrictocin, ribonucleases, bacterial endotoxin, the lipid A moiety of bacterial endotoxin, and cholera toxin. Additional

biological toxins include, but are not limited to, fungal-derived calicheamicins and maytansinoids. Other examples of cytotoxic agents include, but are not limited to, peptides derived from proteins involved in apoptosis, such as Bcl-x, Bax, or Bad.

Representative antibody-conjugates that can be prepared by the presently disclosed methods include, but are not limited to, gemtuzumab ozogamicin (Mylotarg), for example, for treating acute myelogenous leukemia; Trastuzumab-DM1 (T-DM1), for example, for treating breast cancer; Inotuzumab ozogamicin (CMC-544), for example, for treating non-Hodgkin's lymphoma; glembatumumab vedotin (CDX-011, CR011-vcMMAE), for example, for treating melanoma and metastatic breast cancer; SGN-35; IMGN242, for example, for treating CanAg expressing gastric cancer; lorvotuzumab mertansine (IMGN901), for example, for treating CD56 cancers; AN-152, LMB2, TP-38, VB4-845; Cantuzumab mertansine (huC242-DM1); AVE9633; SAR3419; CAT-8015 (anti-CD22); Anti-CD22-MCC-DM1, for example, for treating leukemia; and others antibody-conjugates known in the art.

Additional antibody-conjugates contemplated by the presently disclosed methods are disclosed, for example, in Junutula, J. R., et al., Site-specific Conjugation of a Cytotoxic Drug to an Antibody Improves the Therapeutic Index, *Nature Biotechnology* 26 (8):925-932 (2008); and Wakankar, A. A., et al., *Bioconjugate Chem.* 21:1588-1595 (2010), each of which is included herein by reference in its entirety.

In further embodiments, R₃ is an imaging agent, for example, for use in *in vivo* imaging. The imaging agent can include a fluorescent dye moiety or a radionucleotide. In some embodiments, the imaging agent comprises a fluorescent dye moiety that emits light in the visible or near infrared spectrum. The fluorescent dye moiety can include any additional atoms or linkers necessary to attach the fluorescent dye moiety to the rest of the compound. For instance linking groups having alkyl, aryl, combination of alkyl and aryl, or alkyl and aryl groups having heteroatoms can be present in the fluorescent dye moiety, so long as the linker does not interfere with the fluorescence of the dye. In some embodiments, the fluorescent dye moiety includes a poly(ethyleneglycol) linker.

Numerous fluorescent dye moieties are known in the art, and will be readily apparent to one of ordinary skill in the art. Representative, non-limiting examples of fluorescent compounds suitable for use as an imaging agent comprising the presently

disclosed antibody-conjugates include, but are not limited to, carbocyanine, indocarbocyanine, oxacarbocyanine, thiocarbocyanine, merocyanine, polymethine, coumarine, rhodamine, xanthene, fluorescein, and boron-dipyrromethane (BODIPY) compounds.

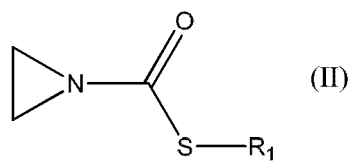
5 Additional examples of fluorescent dye moieties include those described in WO 20089/109832, which is incorporated herein by reference in its entirety.

Further, particular dyes that emit in the near infrared spectrum include commercially available compounds Cy5, Cy5.5, and Cy7, available from GE Healthcare; VivoTag-680, VivoTag-S680, and VivoTag-S750, available from VisEn
 10 Medical; AlexaFluor660, AlexaFluor680, AlexaFluor700, AlexaFluor750, and AlexaFluor790, available from Invitrogen; Dy677, Dy676, Dy682, Dy752, and Dy780, available from Dyonics; DyLight547, and Dylight647, available from Pierce; HiLyte Fluor 647, HiLyte Fluor 680, and HiLyte Fluor 750, available from AnaSpec; IRDye 800CW, IRDye 800RS, and IRDye 700DX, available from Li-Cor; and
 15 ADS780WS, ADS830WS, and ADS832WS, available from American Dye Source.

In some embodiments, the imaging agent is a radionucleotide. In particular embodiments, the radionucleotide is selected from the group consisting of C-11, F-18, I-123, I-124, I-125, I-131, Tc-99m, In-111, Ga-67, Ga-68, Y-86, Y-90, Lu-177, Re-186, Re-188, Cu-64, Cu-67, Co-55, Co-57, Sc-47, Ac-225, Bi-213, Bi-212, Pb-212,
 20 Sm-153, Ho-166, and Dy-166.

G. Compounds of Formula (II)

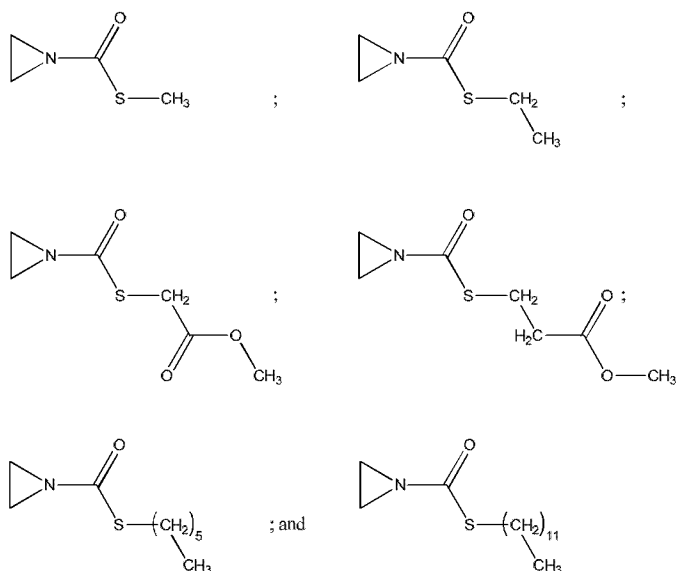
In some embodiments, the presently disclosed subject matter provides a compound of formula (II):



25 wherein: R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, -(CH₂)_n-(C=O)-O-R₂, and -O-(C=O)-(CH₂)_n-(C=O)-OH; and wherein n is an integer from 0 to 20; and R₂ is selected from the group consisting of
 30 substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl,

substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl; or a pharmaceutically acceptable salt thereof.

In some embodiments of the compound of formula (II), X is S and R₁ is selected from the group consisting of unsubstituted, straight-chain C₁ to C₁₂ alkyl, –
 5 (CH₂)–(C=O)–O–CH₃, –(CH₂)₂–(C=O)–O–CH₃, and –O–(C=O)–(CH₂)₂–(C=O)–OH. In particular embodiments, the compound of formula (II) is selected from the group consisting of:



While the following terms in relation to compounds of formula (I) are believed
 10 to be well understood by one of ordinary skill in the art, the following definitions are set forth to facilitate explanation of the presently disclosed subject matter. These definitions are intended to supplement and illustrate, not preclude, the definitions that would be apparent to one of ordinary skill in the art upon review of the present disclosure.

15 The terms substituted, whether preceded by the term “optionally” or not, and substituent, as used herein, refer to the ability, as appreciated by one skilled in this art, to change one functional group for another functional group provided that the valency of all atoms is maintained. When more than one position in any given structure may be substituted with more than one substituent selected from a specified group, the
 20 substituent may be either the same or different at every position. The substituents also may be further substituted (e.g., an aryl group substituent may have another substituent off it, such as another aryl group, which is further substituted, for example,

with fluorine at one or more positions).

Where substituent groups or linking groups are specified by their conventional chemical formulae, written from left to right, they equally encompass the chemically identical substituents that would result from writing the structure from right to left,

- 5 e.g., $-\text{CH}_2\text{O}-$ is equivalent to $-\text{OCH}_2-$; $-\text{C}(=\text{O})\text{O}-$ is equivalent to $-\text{OC}(=\text{O})-$; $-\text{OC}(=\text{O})\text{NR}-$ is equivalent to $-\text{NRC}(=\text{O})\text{O}-$, and the like.

- When the term “independently selected” is used, the substituents being referred to (e.g., R groups, such as groups R_1 , R_2 , and the like, or variables, such as “m” and “n”), can be identical or different. For example, both R_1 and R_2 can be
10 substituted alkyls, or R_1 can be hydrogen and R_2 can be a substituted alkyl, and the like.

- The terms “a,” “an,” or “a(n),” when used in reference to a group of substituents herein, mean at least one. For example, where a compound is substituted with “an” alkyl or aryl, the compound is optionally substituted with at least one alkyl
15 and/or at least one aryl. Moreover, where a moiety is substituted with an R substituent, the group may be referred to as “R-substituted.” Where a moiety is R-substituted, the moiety is substituted with at least one R substituent and each R substituent is optionally different.

- A named “R” or group will generally have the structure that is recognized in
20 the art as corresponding to a group having that name, unless specified otherwise herein. For the purposes of illustration, certain representative “R” groups as set forth above are defined below.

- Descriptions of compounds of the present disclosure are limited by principles of chemical bonding known to those skilled in the art. Accordingly, where a group
25 may be substituted by one or more of a number of substituents, such substitutions are selected so as to comply with principles of chemical bonding and to give compounds which are not inherently unstable and/or would be known to one of ordinary skill in the art as likely to be unstable under ambient conditions, such as aqueous, neutral, and several known physiological conditions. For example, a heterocycloalkyl or
30 heteroaryl is attached to the remainder of the molecule via a ring heteroatom in compliance with principles of chemical bonding known to those skilled in the art thereby avoiding inherently unstable compounds.

The term hydrocarbon, as used herein, refers to any chemical group comprising hydrogen and carbon. The hydrocarbon may be substituted or

unsubstituted. As would be known to one skilled in this art, all valencies must be satisfied in making any substitutions. The hydrocarbon may be unsaturated, saturated, branched, unbranched, cyclic, polycyclic, or heterocyclic. Illustrative hydrocarbons are further defined herein below and include, for example, methyl, ethyl, n-propyl, 5 iso-propyl, cyclopropyl, allyl, vinyl, n-butyl, tert-butyl, ethynyl, cyclohexyl, methoxyl, diethylamino, and the like.

The term "alkyl," by itself or as part of another substituent, means, unless otherwise stated, a straight (i.e., unbranched) or branched chain, acyclic or cyclic hydrocarbon group, or combination thereof, which may be fully saturated, mono- or 10 polyunsaturated and can include di- and multivalent groups, having the number of carbon atoms designated (i.e., C₁-C₁₀ means one to ten carbons). In particular embodiments, the term "alkyl" refers to C₁₋₂₀ inclusive, linear (i.e., "straight-chain"), branched, or cyclic, saturated or at least partially and in some cases fully unsaturated (i.e., alkenyl and alkynyl) hydrocarbon radicals derived from a hydrocarbon moiety 15 containing between one and twenty carbon atoms by removal of a single hydrogen atom.

Representative saturated hydrocarbon groups include, but are not limited to, methyl, ethyl, n-propyl, isopropyl, n-butyl, isobutyl, sec-butyl, tert-butyl, n-pentyl, sec-pentyl, iso-pentyl, neopentyl, n-hexyl, sec-hexyl, n-heptyl, n-octyl, n-decyl, n- 20 undecyl, dodecyl, cyclohexyl, (cyclohexyl)methyl, cyclopropylmethyl, and homologs and isomers thereof.

"Branched" refers to an alkyl group in which a lower alkyl group, such as methyl, ethyl or propyl, is attached to a linear alkyl chain. "Lower alkyl" refers to an alkyl group having 1 to about 8 carbon atoms (i.e., a C₁₋₈ alkyl), e.g., 1, 2, 3, 4, 5, 6, 7, 25 or 8 carbon atoms. "Higher alkyl" refers to an alkyl group having about 10 to about 20 carbon atoms, e.g., 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbon atoms. In certain embodiments, "alkyl" refers, in particular, to C₁₋₈ straight-chain alkyls. In other embodiments, "alkyl" refers, in particular, to C₁₋₈ branched-chain alkyls.

Alkyl groups can optionally be substituted (a "substituted alkyl") with one or 30 more alkyl group substituents, which can be the same or different. The term "alkyl group substituent" includes but is not limited to alkyl, substituted alkyl, halo, arylamino, acyl, hydroxyl, aryloxy, alkoxy, alkylthio, arylthio, aralkyloxy, aralkylthio, carboxyl, alkoxycarbonyl, oxo, and cycloalkyl. There can be optionally inserted along the alkyl chain one or more oxygen, sulfur or substituted or

unsubstituted nitrogen atoms, wherein the nitrogen substituent is hydrogen, lower alkyl (also referred to herein as “alkylaminoalkyl”), or aryl.

Thus, as used herein, the term “substituted alkyl” includes alkyl groups, as defined herein, in which one or more atoms or functional groups of the alkyl group
5 are replaced with another atom or functional group, including for example, alkyl, substituted alkyl, halogen, aryl, substituted aryl, alkoxy, hydroxyl, nitro, amino, alkylamino, dialkylamino, sulfate, and mercapto.

The term “heteroalkyl,” by itself or in combination with another term, means, unless otherwise stated, a stable straight or branched chain, or cyclic hydrocarbon
10 group, or combinations thereof, consisting of at least one carbon atoms and at least one heteroatom selected from the group consisting of O, N, P, Si and S, and wherein the nitrogen, phosphorus, and sulfur atoms may optionally be oxidized and the nitrogen heteroatom may optionally be quaternized. The heteroatom(s) O, N, P and S and Si may be placed at any interior position of the heteroalkyl group or at the
15 position at which alkyl group is attached to the remainder of the molecule. Examples include, but are not limited to, -CH₂-CH₂-O-CH₃, -CH₂-CH₂-NH-CH₃, -CH₂-CH₂-N(CH₃)-CH₃, -CH₂-S-CH₂-CH₃, -CH₂-CH₂-S(O)-CH₃, -CH₂-CH₂-S(O)₂-CH₃, -CH=CH-O-CH₃, -Si(CH₃)₃, -CH₂-CH=N-OCH₃, -CH=CH-N(CH₃)-CH₃, -O-CH₃, -O-CH₂-CH₃, and -CN. Up to two or three
20 heteroatoms may be consecutive, such as, for example, -CH₂-NH-OCH₃ and -CH₂-O-Si(CH₃)₃.

As described above, heteroalkyl groups, as used herein, include those groups that are attached to the remainder of the molecule through a heteroatom, such as -C(O)R', -C(O)NR', -NR'R'', -OR', -SR, and/or -SO₂R'. Where “heteroalkyl” is
25 recited, followed by recitations of specific heteroalkyl groups, such as -NR'R'' or the like, it will be understood that the terms heteroalkyl and -NR'R'' are not redundant or mutually exclusive. Rather, the specific heteroalkyl groups are recited to add clarity. Thus, the term “heteroalkyl” should not be interpreted herein as excluding specific heteroalkyl groups, such as -NR'R'' or the like.

30 “Cyclic” and “cycloalkyl” refer to a non-aromatic mono- or multicyclic ring system of about 3 to about 10 carbon atoms, e.g., 3, 4, 5, 6, 7, 8, 9, or 10 carbon atoms. The cycloalkyl group can be optionally partially unsaturated. The cycloalkyl group also can be optionally substituted with an alkyl group substituent as defined herein, oxo, and/or alkylene. There can be optionally inserted along the cyclic alkyl

chain one or more oxygen, sulfur or substituted or unsubstituted nitrogen atoms, wherein the nitrogen substituent is hydrogen, alkyl, substituted alkyl, aryl, or substituted aryl, thus providing a heterocyclic group. Representative monocyclic cycloalkyl rings include cyclopentyl, cyclohexyl, and cycloheptyl. Multicyclic cycloalkyl rings include adamantyl, octahydronaphthyl, decalin, camphor, camphane, and noradamantyl, and fused ring systems, such as dihydro- and tetrahydronaphthalene, and the like.

The term "cycloalkylalkyl," as used herein, refers to a cycloalkyl group as defined hereinabove, which is attached to the parent molecular moiety through an alkyl group, also as defined above. Examples of cycloalkylalkyl groups include cyclopropylmethyl and cyclopentylethyl.

The terms "cycloheteroalkyl" or "heterocycloalkyl" refer to a non-aromatic ring system, unsaturated or partially unsaturated ring system, such as a 3- to 10-member substituted or unsubstituted cycloalkyl ring system, including one or more heteroatoms, which can be the same or different, and are selected from the group consisting of nitrogen (N), oxygen (O), sulfur (S), phosphorus (P), and silicon (Si), and optionally can include one or more double bonds.

The cycloheteroalkyl ring can be optionally fused to or otherwise attached to other cycloheteroalkyl rings and/or non-aromatic hydrocarbon rings. Heterocyclic rings include those having from one to three heteroatoms independently selected from oxygen, sulfur, and nitrogen, in which the nitrogen and sulfur heteroatoms may optionally be oxidized and the nitrogen heteroatom may optionally be quaternized. In certain embodiments, the term heterocyclic refers to a non-aromatic 5-, 6-, or 7-membered ring or a polycyclic group wherein at least one ring atom is a heteroatom selected from O, S, and N (wherein the nitrogen and sulfur heteroatoms may be optionally oxidized), including, but not limited to, a bi- or tri-cyclic group, comprising fused six-membered rings having between one and three heteroatoms independently selected from the oxygen, sulfur, and nitrogen, wherein (i) each 5-membered ring has 0 to 2 double bonds, each 6-membered ring has 0 to 2 double bonds, and each 7-membered ring has 0 to 3 double bonds, (ii) the nitrogen and sulfur heteroatoms may be optionally oxidized, (iii) the nitrogen heteroatom may optionally be quaternized, and (iv) any of the above heterocyclic rings may be fused to an aryl or heteroaryl ring. Representative cycloheteroalkyl ring systems include, but are not limited to pyrrolidinyl, pyrrolinyl, imidazolidinyl, imidazolinyl, pyrazolidinyl, pyrazolinyl,

piperidyl, piperaziny, indolyl, quinuclidinyl, morpholinyl, thiomorpholinyl, thiadiazinyl, tetrahydrofuran, and the like.

The terms “cycloalkyl” and “heterocycloalkyl”, by themselves or in combination with other terms, represent, unless otherwise stated, cyclic versions of “alkyl” and “heteroalkyl”, respectively. Additionally, for heterocycloalkyl, a heteroatom can occupy the position at which the heterocycle is attached to the remainder of the molecule. Examples of cycloalkyl include, but are not limited to, cyclopentyl, cyclohexyl, 1-cyclohexenyl, 3-cyclohexenyl, cycloheptyl, and the like. Examples of heterocycloalkyl include, but are not limited to, 1-(1,2,5,6-tetrahydropyridyl), 1-piperidinyl, 2-piperidinyl, 3-piperidinyl, 4-morpholinyl, 3-morpholinyl, tetrahydrofuran-2-yl, tetrahydrofuran-3-yl, tetrahydrothien-2-yl, tetrahydrothien-3-yl, 1-piperaziny, 2-piperaziny, and the like. The terms “cycloalkylene” and “heterocycloalkylene” refer to the divalent derivatives of cycloalkyl and heterocycloalkyl, respectively.

An unsaturated alkyl group is one having one or more double bonds or triple bonds. Examples of unsaturated alkyl groups include, but are not limited to, vinyl, 2-propenyl, crotyl, 2-isopentenyl, 2-(butadienyl), 2,4-pentadienyl, 3-(1,4-pentadienyl), ethynyl, 1- and 3-propynyl, 3-butenyl, and the higher homologs and isomers. Alkyl groups which are limited to hydrocarbon groups are termed “homoalkyl.”

More particularly, the term “alkenyl” as used herein refers to a monovalent group derived from a C₁₋₂₀ inclusive straight or branched hydrocarbon moiety having at least one carbon-carbon double bond by the removal of a single hydrogen atom. Alkenyl groups include, for example, ethenyl (i.e., vinyl), propenyl, butenyl, 1-methyl-2-buten-1-yl, pentenyl, hexenyl, octenyl, and butadienyl.

The term “cycloalkenyl” as used herein refers to a cyclic hydrocarbon containing at least one carbon-carbon double bond. Examples of cycloalkenyl groups include cyclopropenyl, cyclobutenyl, cyclopentenyl, cyclopentadiene, cyclohexenyl, 1,3-cyclohexadiene, cycloheptenyl, cycloheptatrienyl, and cyclooctenyl.

The term “alkynyl” as used herein refers to a monovalent group derived from a straight or branched C₁₋₂₀ hydrocarbon of a designed number of carbon atoms containing at least one carbon-carbon triple bond. Examples of “alkynyl” include ethynyl, 2-propynyl (propargyl), 1-propynyl, pentynyl, hexynyl, heptynyl, and allenyl groups, and the like.

The term “alkylene” by itself or a part of another substituent refers to a

straight or branched bivalent aliphatic hydrocarbon group derived from an alkyl group having from 1 to about 20 carbon atoms, e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20 carbon atoms. The alkylene group can be straight, branched or cyclic. The alkylene group also can be optionally unsaturated and/or substituted with one or more “alkyl group substituents.” There can be optionally inserted along the alkylene group one or more oxygen, sulfur or substituted or unsubstituted nitrogen atoms (also referred to herein as “alkylaminoalkyl”), wherein the nitrogen substituent is alkyl as previously described. Exemplary alkylene groups include methylene ($-\text{CH}_2-$); ethylene ($-\text{CH}_2-\text{CH}_2-$); propylene ($-(\text{CH}_2)_3-$); cyclohexylene ($-\text{C}_6\text{H}_{10}-$); $-\text{CH}=\text{CH}-\text{CH}=\text{CH}-$; $-\text{CH}=\text{CH}-\text{CH}_2-$; $-\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_2-$; $-\text{CH}_2\text{CH}=\text{CHCH}_2-$; $-\text{CH}_2\text{C}(\text{CH}_3)\text{CH}_2-$; $-\text{CH}_2\text{CH}_2\text{CH}(\text{CH}_2\text{CH}_2\text{CH}_3)\text{CH}_2-$; $-(\text{CH}_2)_q-\text{N}(\text{R})-(\text{CH}_2)_r-$, wherein each of q and r is independently an integer from 0 to about 20, e.g., 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20, and R is hydrogen or lower alkyl; methylenedioxy ($-\text{O}-\text{CH}_2-\text{O}-$); and ethylenedioxy ($-\text{O}-(\text{CH}_2)_2-\text{O}-$). An alkylene group can have about 2 to about 3 carbon atoms and can further have 6-20 carbons. Typically, an alkyl (or alkylene) group will have from 1 to 24 carbon atoms, with those groups having 10 or fewer carbon atoms being some embodiments of the present disclosure. A “lower alkyl” or “lower alkylene” is a shorter chain alkyl or alkylene group, generally having eight or fewer carbon atoms.

The term “heteroalkylene” by itself or as part of another substituent means a divalent group derived from heteroalkyl, as exemplified, but not limited by, $-\text{CH}_2-\text{CH}_2-\text{S}-\text{CH}_2-\text{CH}_2-$ and $-\text{CH}_2-\text{S}-\text{CH}_2-\text{CH}_2-\text{NH}-\text{CH}_2-$. For heteroalkylene groups, heteroatoms can also occupy either or both of the chain termini (e.g., alkyleneoxo, alkylenedioxo, alkyleneamino, alkylenediamino, and the like). Still further, for alkylene and heteroalkylene linking groups, no orientation of the linking group is implied by the direction in which the formula of the linking group is written. For example, the formula $-\text{C}(\text{O})\text{OR}'-$ represents both $-\text{C}(\text{O})\text{OR}'-$ and $-\text{R}'\text{OC}(\text{O})-$.

The term “aryl” means, unless otherwise stated, an aromatic hydrocarbon substituent that can be a single ring or multiple rings (such as from 1 to 3 rings), which are fused together or linked covalently. The term “heteroaryl” refers to aryl groups (or rings) that contain from one to four heteroatoms (in each separate ring in the case of multiple rings) selected from N, O, and S, wherein the nitrogen and sulfur atoms are optionally oxidized, and the nitrogen atom(s) are optionally quaternized. A heteroaryl group can be attached to the remainder of the molecule through a carbon or

heteroatom. Non-limiting examples of aryl and heteroaryl groups include phenyl, 1-naphthyl, 2-naphthyl, 4-biphenyl, 1-pyrrolyl, 2-pyrrolyl, 3-pyrrolyl, 3-pyrazolyl, 2-imidazolyl, 4-imidazolyl, pyrazinyl, 2-oxazolyl, 4-oxazolyl, 2-phenyl-4-oxazolyl, 5-oxazolyl, 3-isoxazolyl, 4-isoxazolyl, 5-isoxazolyl, 2-thiazolyl, 4-thiazolyl, 5-thiazolyl, 2-furyl, 3-furyl, 2-thienyl, 3-thienyl, 2-pyridyl, 3-pyridyl, 4-pyridyl, 2-pyrimidyl, 4-pyrimidyl, 5-benzothiazolyl, purinyl, 2-benzimidazolyl, 5-indolyl, 1-isoquinolyl, 5-isoquinolyl, 2-quinoxaliny, 5-quinoxaliny, 3-quinolyl, and 6-quinolyl. Substituents for each of above noted aryl and heteroaryl ring systems are selected from the group of acceptable substituents described below. The terms

10 “arylene” and “heteroarylene” refer to the divalent forms of aryl and heteroaryl, respectively.

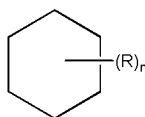
For brevity, the term “aryl” when used in combination with other terms (e.g., aryloxo, arylthioxo, arylalkyl) includes both aryl and heteroaryl rings as defined above. Thus, the terms “arylalkyl” and “heteroarylalkyl” are meant to include those

15 groups in which an aryl or heteroaryl group is attached to an alkyl group (e.g., benzyl, phenethyl, pyridylmethyl, furylmethyl, and the like) including those alkyl groups in which a carbon atom (e.g., a methylene group) has been replaced by, for example, an oxygen atom (e.g., phenoxymethyl, 2-pyridyloxymethyl, 3-(1-naphthyloxy)propyl, and the like). The term “haloaryl,” however, as used herein, is meant to cover only aryls

20 substituted with one or more halogens.

Where a heteroalkyl, heterocycloalkyl, or heteroaryl includes a specific number of members (e.g. “3 to 7 membered”), the term “member” refers to a carbon or heteroatom.

Further, a structure represented generally by the formula:

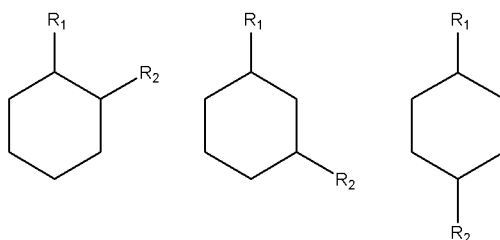


25 as used herein refers to a ring structure, for example, but not limited to a 3-carbon, a 4-carbon, a 5-carbon, a 6-carbon, a 7-carbon, and the like, aliphatic and/or aromatic cyclic compound, including a saturated ring structure, a partially saturated ring structure, and an unsaturated ring structure, comprising a substituent R group, wherein

30 the R group can be present or absent, and when present, one or more R groups can each be substituted on one or more available carbon atoms of the ring structure. The presence or absence of the R group and number of R groups is determined by the

value of the variable “n,” which is an integer generally having a value ranging from 0 to the number of carbon atoms on the ring available for substitution. Each R group, if more than one, is substituted on an available carbon of the ring structure rather than on another R group. For example, the structure above where n is 0 to 2 would

5 comprise compound groups including, but not limited to:



and the like.

A dashed line representing a bond in a cyclic ring structure indicates that the bond can be either present or absent in the ring. That is, a dashed line representing a bond in a cyclic ring structure indicates that the ring structure is selected from the group consisting of a saturated ring structure, a partially saturated ring structure, and an unsaturated ring structure. The symbol () denotes the point of attachment of a moiety to the remainder of the molecule.

When a named atom of an aromatic ring or a heterocyclic aromatic ring is defined as being “absent,” the named atom is replaced by a direct bond. Each of above terms (e.g., “alkyl,” “heteroalkyl,” “cycloalkyl,” and “heterocycloalkyl,” “aryl,” “heteroaryl,” “phosphonate,” and “sulfonate” as well as their divalent derivatives) are meant to include both substituted and unsubstituted forms of the indicated group. Optional substituents for each type of group are provided below.

Substituents for alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl monovalent and divalent derivative groups (including those groups often referred to as alkylene, alkenyl, heteroalkylene, heteroalkenyl, alkynyl, cycloalkyl, heterocycloalkyl, cycloalkenyl, and heterocycloalkenyl) can be one or more of a variety of groups selected from, but not limited to: -OR', =O, =NR', =N-OR', -NR'R'', -SR', -halogen, -SiR'R''R''', -OC(O)R', -C(O)R', -CO₂R', -C(O)NR'R'', -OC(O)NR'R'', -NR''C(O)R', -NR'-C(O)NR''R''', -NR''C(O)OR', -NR-C(NR'R'')=NR''', -S(O)R', -S(O)₂R', -S(O)₂NR'R'', -NRSO₂R', -CN and -NO₂ in a number ranging from zero to (2m'+1), where m' is the total number of carbon atoms in such groups. R', R'', R'''

and R^{'''} each may independently refer to hydrogen, substituted or unsubstituted heteroalkyl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl (e.g., aryl substituted with 1-3 halogens), substituted or unsubstituted alkyl, alkoxy or thioalkoxy groups, or arylalkyl groups. As used herein, an "alkoxy" group is an alkyl attached to the remainder of the molecule through a divalent oxygen. When a compound of the disclosure includes more than one R group, for example, each of the R groups is independently selected as are each R', R'', R''' and R^{'''} groups when more than one of these groups is present. When R' and R'' are attached to the same nitrogen atom, they can be combined with the nitrogen atom to form a 4-, 5-, 6-, or 7- membered ring. For example, -NR'R'' is meant to include, but not be limited to, 1-pyrrolidinyl and 4-morpholinyl. From the above discussion of substituents, one of skill in the art will understand that the term "alkyl" is meant to include groups including carbon atoms bound to groups other than hydrogen groups, such as haloalkyl (e.g., -CF₃ and -CH₂CF₃) and acyl (e.g., -C(O)CH₃, -C(O)CF₃, -C(O)CH₂OCH₃, and the like).

Similar to the substituents described for alkyl groups above, exemplary substituents for aryl and heteroaryl groups (as well as their divalent derivatives) are varied and are selected from, for example: halogen, -OR', -NR'R'', -SR', -halogen, -SiR'R''R''', -OC(O)R', -C(O)R', -CO₂R', -C(O)NR'R'', -OC(O)NR'R'', -NR'C(O)R', -NR'-C(O)NR''R''', -NR'C(O)OR', -NR-C(NR'R''R''')=NR^{'''}, -NR-C(NR'R'')=NR^{'''} -S(O)R', -S(O)₂R', -S(O)₂NR'R'', -NRSO₂R', -CN and -NO₂, -R', -N₃, -CH(Ph)₂, fluoro(C₁-C₄)alkoxo, and fluoro(C₁-C₄)alkyl, in a number ranging from zero to the total number of open valences on aromatic ring system; and where R', R'', R''' and R^{'''} may be independently selected from hydrogen, substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl and substituted or unsubstituted heteroaryl. When a compound of the disclosure includes more than one R group, for example, each of the R groups is independently selected as are each R', R'', R''' and R^{'''} groups when more than one of these groups is present.

Two of the substituents on adjacent atoms of aryl or heteroaryl ring may optionally form a ring of the formula -T-C(O)-(CRR')_q-U-, wherein T and U are independently -NR-, -O-, -CRR'- or a single bond, and q is an integer of from 0 to 3. Alternatively, two of the substituents on adjacent atoms of aryl or heteroaryl ring may

optionally be replaced with a substituent of the formula $-A-(CH_2)_r-B-$, wherein A and B are independently $-CRR'$ -, $-O-$, $-NR'$ -, $-S-$, $-S(O)-$, $-S(O)_2-$, $-S(O)_2NR'$ - or a single bond, and r is an integer of from 1 to 4.

One of the single bonds of the new ring so formed may optionally be replaced
5 with a double bond. Alternatively, two of the substituents on adjacent atoms of aryl or heteroaryl ring may optionally be replaced with a substituent of the formula $-(CRR')_s-X'-(C''R'')_d-$, where s and d are independently integers of from 0 to 3, and X' is $-O-$, $-NR'$ -, $-S-$, $-S(O)-$, $-S(O)_2-$, or $-S(O)_2NR'$ -. The substituents R, R', R'' and R''' may be independently selected from hydrogen, substituted or unsubstituted alkyl,
10 substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl, and substituted or unsubstituted heteroaryl.

As used herein, the term "acyl" refers to an organic acid group wherein the $-OH$ of the carboxyl group has been replaced with another substituent and has the general formula $RC(=O)-$, wherein R is an alkyl, alkenyl, alkynyl, aryl, carbocyclic,
15 heterocyclic, or aromatic heterocyclic group as defined herein). As such, the term "acyl" specifically includes arylacyl groups, such as an acetylfuran and a phenacyl group. Specific examples of acyl groups include acetyl and benzoyl.

The terms "alkoxyl" or "alkoxy" are used interchangeably herein and refer to a saturated (i.e., $alkyl-O-$) or unsaturated (i.e., $alkenyl-O-$ and $alkynyl-O-$) group
20 attached to the parent molecular moiety through an oxygen atom, wherein the terms "alkyl," "alkenyl," and "alkynyl" are as previously described and can include C_{1-20} inclusive, linear, branched, or cyclic, saturated or unsaturated oxo-hydrocarbon chains, including, for example, methoxyl, ethoxyl, propoxyl, isopropoxyl, n-butoxyl, sec-butoxyl, t-butoxyl, and n-pentoxyl, neopentoxyl, n-hexoxyl, and the like.

25 The term "alkoxyalkyl" as used herein refers to an $alkyl-O-alkyl$ ether, for example, a methoxyethyl or an ethoxymethyl group.

"Aryloxy" refers to an $aryl-O-$ group wherein the aryl group is as previously described, including a substituted aryl. The term "aryloxy" as used herein can refer to phenyloxy or hexyloxy, and alkyl, substituted alkyl, halo, or alkoxyl substituted
30 phenyloxy or hexyloxy.

"Aralkyl" refers to an $aryl-alkyl$ -group wherein aryl and alkyl are as previously described, and included substituted aryl and substituted alkyl. Exemplary aralkyl groups include benzyl, phenylethyl, and naphthylmethyl.

"Aralkyloxy" refers to an $aralkyl-O-$ group wherein the aralkyl group is as

previously described. An exemplary aralkyloxy group is benzyloxy.

“Alkoxy carbonyl” refers to an alkyl-O-CO- group. Exemplary alkoxy carbonyl groups include methoxy carbonyl, ethoxy carbonyl, butyloxy carbonyl, and t-butyloxy carbonyl.

5 “Aryloxy carbonyl” refers to an aryl-O-CO- group. Exemplary aryloxy carbonyl groups include phenoxy- and naphthoxy-carbonyl.

“Aralkoxy carbonyl” refers to an aralkyl-O-CO- group. An exemplary aralkoxy carbonyl group is benzyloxy carbonyl.

“Carbamoyl” refers to an amide group of the formula $-\text{CONH}_2$.

10 “Alkyl carbamoyl” refers to a $\text{R}'\text{RN}-\text{CO}-$ group wherein one of R and R' is hydrogen and the other of R and R' is alkyl and/or substituted alkyl as previously described.

“Dialkyl carbamoyl” refers to a $\text{R}'\text{RN}-\text{CO}-$ group wherein each of R and R' is independently alkyl and/or substituted alkyl as previously described.

The term carbonyldioxy, as used herein, refers to a carbonate group of the
15 formula $-\text{O}-\text{CO}-\text{OR}$.

“Acyloxy” refers to an acyl-O- group wherein acyl is as previously described.

The term “amino” refers to the $-\text{NH}_2$ group and also refers to a nitrogen containing group as is known in the art derived from ammonia by the replacement of one or more hydrogen radicals by organic radicals. For example, the terms
20 “acylamino” and “alkylamino” refer to specific N-substituted organic radicals with acyl and alkyl substituent groups respectively.

An “aminoalkyl” as used herein refers to an amino group covalently bound to an alkylene linker. More particularly, the terms alkylamino, dialkylamino, and trialkylamino as used herein refer to one, two, or three, respectively, alkyl groups, as
25 previously defined, attached to the parent molecular moiety through a nitrogen atom. The term alkylamino refers to a group having the structure $-\text{NHR}'$ wherein R' is an alkyl group, as previously defined; whereas the term dialkylamino refers to a group having the structure $-\text{NR}'\text{R}''$, wherein R' and R'' are each independently selected from the group consisting of alkyl groups. The term trialkylamino refers to a group
30 having the structure $-\text{NR}'\text{R}''\text{R}'''$, wherein R', R'', and R''' are each independently selected from the group consisting of alkyl groups. Additionally, R', R'', and/or R''' taken together may optionally be $-(\text{CH}_2)_k-$ where k is an integer from 2 to 6. Examples include, but are not limited to, methylamino, dimethylamino, ethylamino, diethylamino, diethylaminocarbonyl, methylethylamino, iso-propylamino, piperidino,

trimethylamino, and propylamino.

The amino group is -NR'R'', wherein R' and R'' are typically selected from hydrogen, substituted or unsubstituted alkyl, substituted or unsubstituted heteroalkyl, substituted or unsubstituted cycloalkyl, substituted or unsubstituted heterocycloalkyl, substituted or unsubstituted aryl, or substituted or unsubstituted heteroaryl.

The terms alkylthioether and thioalkoxyl refer to a saturated (i.e., alkyl-S-) or unsaturated (i.e., alkenyl-S- and alkynyl-S-) group attached to the parent molecular moiety through a sulfur atom. Examples of thioalkoxyl moieties include, but are not limited to, methylthio, ethylthio, propylthio, isopropylthio, n-butylthio, and the like.

“Acylamino” refers to an acyl-NH- group wherein acyl is as previously described. “Aroylamino” refers to an aroyl-NH- group wherein aroyl is as previously described.

The term “carbonyl” refers to the -(C=O)- group.

The term “carboxyl” refers to the -COOH group. Such groups also are referred to herein as a “carboxylic acid” moiety.

The terms “halo,” “halide,” or “halogen” as used herein refer to fluoro, chloro, bromo, and iodo groups. Additionally, terms such as “haloalkyl,” are meant to include monohaloalkyl and polyhaloalkyl. For example, the term “halo(C₁-C₄)alkyl” is meant to include, but not be limited to, trifluoromethyl, 2,2,2-trifluoroethyl, 4-chlorobutyl, 3-bromopropyl, and the like.

The term “hydroxyl” refers to the -OH group.

The term “hydroxyalkyl” refers to an alkyl group substituted with an -OH group.

The term “mercapto” refers to the -SH group.

The term “oxo” as used herein means an oxygen atom that is double bonded to a carbon atom or to another element.

The term “nitro” refers to the -NO₂ group.

The term “thio” refers to a compound described previously herein wherein a carbon or oxygen atom is replaced by a sulfur atom.

The term “sulfate” refers to the -SO₄ group.

The term thiohydroxyl or thiol, as used herein, refers to a group of the formula -SH.

The term ureido refers to a urea group of the formula -NH-CO-NH₂.

Unless otherwise explicitly defined, a “substituent group,” as used herein,

includes a functional group selected from one or more of the following moieties, which are defined herein:

(A) -OH, -NH₂, -SH, -CN, -CF₃, -NO₂, oxo, halogen, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, unsubstituted heteroaryl, and

(B) alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl, substituted with at least one substituent selected from:

(i) oxo, -OH, -NH₂, -SH, -CN, -CF₃, -NO₂, halogen, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, unsubstituted heteroaryl, and

(ii) alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl, aryl, and heteroaryl, substituted with at least one substituent selected from:

(a) oxo, -OH, -NH₂, -SH, -CN, -CF₃, -NO₂, halogen, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, unsubstituted heteroaryl, and

(b) alkyl, heteroalkyl, cycloalkyl, heterocycloalkyl, aryl, or heteroaryl, substituted with at least one substituent selected from oxo, -OH, -NH₂, -SH, -CN, -CF₃, -NO₂, halogen, unsubstituted alkyl, unsubstituted heteroalkyl, unsubstituted cycloalkyl, unsubstituted heterocycloalkyl, unsubstituted aryl, and unsubstituted heteroaryl.

A “lower substituent” or “lower substituent group,” as used herein means a group selected from all of the substituents described hereinabove for a “substituent group,” wherein each substituted or unsubstituted alkyl is a substituted or unsubstituted C₁-C₈ alkyl, each substituted or unsubstituted heteroalkyl is a substituted or unsubstituted 2 to 8 membered heteroalkyl, each substituted or unsubstituted cycloalkyl is a substituted or unsubstituted C₅-C₇ cycloalkyl, and each substituted or unsubstituted heterocycloalkyl is a substituted or unsubstituted 5 to 7 membered heterocycloalkyl.

A “size-limited substituent” or “size-limited substituent group,” as used herein means a group selected from all of the substituents described above for a “substituent group,” wherein each substituted or unsubstituted alkyl is a substituted or unsubstituted C₁-C₂₀ alkyl, each substituted or unsubstituted heteroalkyl is a substituted or unsubstituted 2 to 20 membered heteroalkyl, each substituted or unsubstituted cycloalkyl is a substituted or unsubstituted C₄-C₈ cycloalkyl, and each

substituted or unsubstituted heterocycloalkyl is a substituted or unsubstituted 4 to 8 membered heterocycloalkyl.

Throughout the specification and claims, a given chemical formula or name shall encompass all tautomers, congeners, and optical- and stereoisomers, as well as
5 racemic mixtures where such isomers and mixtures exist.

Certain compounds of the present disclosure possess asymmetric carbon atoms (optical or chiral centers) or double bonds; the enantiomers, racemates, diastereomers, tautomers, geometric isomers, stereoisomeric forms that may be defined, in terms of absolute stereochemistry, as (R)- or (S)- or, as (D)- or (L)- for amino acids, and
10 individual isomers are encompassed within the scope of the present disclosure. The compounds of the present disclosure do not include those which are known in art to be too unstable to synthesize and/or isolate. The present disclosure is meant to include compounds in racemic and optically pure forms. Optically active (R)- and (S)-, or (D)- and (L)-isomers may be prepared using chiral synthons or chiral
15 reagents, or resolved using conventional techniques. When the compounds described herein contain olefinic bonds or other centers of geometric asymmetry, and unless specified otherwise, it is intended that the compounds include both E and Z geometric isomers.

Unless otherwise stated, structures depicted herein are also meant to include
20 all stereochemical forms of the structure; i.e., the R and S configurations for each asymmetric center. Therefore, single stereochemical isomers as well as enantiomeric and diastereomeric mixtures of the present compounds are within the scope of the disclosure.

It will be apparent to one skilled in the art that certain compounds of this
25 disclosure may exist in tautomeric forms, all such tautomeric forms of the compounds being within the scope of the disclosure. The term "tautomer," as used herein, refers to one of two or more structural isomers which exist in equilibrium and which are readily converted from one isomeric form to another.

Unless otherwise stated, structures depicted herein are also meant to include
30 compounds which differ only in the presence of one or more isotopically enriched atoms. For example, compounds having the present structures except for the replacement of a hydrogen by a deuterium or tritium, or the replacement of a carbon by ^{13}C - or ^{14}C -enriched carbon are within the scope of this disclosure.

The compounds of the present disclosure may also contain unnatural

proportions of atomic isotopes at one or more of atoms that constitute such compounds. For example, the compounds may be radiolabeled with radioactive isotopes, such as for example tritium (^3H), iodine-125 (^{125}I) or carbon-14 (^{14}C). All isotopic variations of the compounds of the present disclosure, whether radioactive or
5 not, are encompassed within the scope of the present disclosure.

The compounds of the present disclosure may exist as pharmaceutically acceptable salts. The term "pharmaceutically acceptable salts" is meant to include salts of active compounds which are prepared with relatively nontoxic acids or bases, depending on the particular substituent moieties found on the compounds described
10 herein. Pharmaceutically acceptable salts are generally well known to those of ordinary skill in the art, and may include, by way of example but not limitation, acetate, benzenesulfonate, besylate, benzoate, bicarbonate, bitartrate, bromide, calcium edetate, carnsylate, carbonate, citrate, edetate, edisylate, estolate, esylate, fumarate, gluceptate, gluconate, glutamate, glycolylarsanilate, hexylresorcinate,
15 hydrabamine, hydrobromide, hydrochloride, hydroxynaphthoate, iodide, isethionate, lactate, lactobionate, malate, maleate, mandelate, mesylate, mucate, napsylate, nitrate, pamoate (embonate), pantothenate, phosphate/diphosphate, polygalacturonate, salicylate, stearate, subacetate, succinate, sulfate, tannate, tartrates, (e.g. (+)-tartrates, (-)-tartrates or mixtures thereof including racemic mixtures), or teclate. These salts
20 may be prepared by methods known to those skilled in art. Other pharmaceutically acceptable salts may be found in, for example, Remington: The Science and Practice of Pharmacy (20th ed.) Lippincott, Williams & Wilkins (2000).

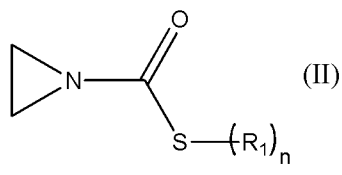
Also included are base addition salts such as sodium, potassium, calcium, ammonium, organic amino, or magnesium salt, or a similar salt. When compounds of
25 the present disclosure contain relatively basic functionalities, acid addition salts can be obtained by contacting the neutral form of such compounds with a sufficient amount of the desired acid, either neat or in a suitable inert solvent. Examples of acceptable acid addition salts include those derived from inorganic acids like hydrochloric, hydrobromic, nitric, carbonic, monohydrogencarbonic, phosphoric,
30 monohydrogenphosphoric, dihydrogenphosphoric, sulfuric, monohydrogensulfuric, hydriodic, or phosphorous acids and the like, as well as the salts derived organic acids like acetic, propionic, isobutyric, maleic, malonic, benzoic, succinic, suberic, fumaric, lactic, mandelic, phthalic, benzenesulfonic, p-tolylsulfonic, citric, tartaric, methanesulfonic, and the like.

Also included are salts of amino acids such as arginate and the like, and salts of organic acids like glucuronic or galactunoric acids and the like, see, for example, Berge et al, "Pharmaceutical Salts", Journal of Pharmaceutical Science, 1977, 66, 1-19). Certain specific compounds of the present disclosure contain both basic and acidic functionalities that allow the compounds to be converted into either base or acid addition salts. The neutral forms of the compounds may be regenerated by contacting the salt with a base or acid and isolating the parent compound in the conventional manner. The parent form of the compound differs from the various salt forms in certain physical properties, such as solubility in polar solvents.

Certain compounds of the present disclosure can exist in unsolvated forms as well as solvated forms, including hydrated forms. In general, the solvated forms are equivalent to unsolvated forms and are encompassed within the scope of the present disclosure. Certain compounds of the present disclosure may exist in multiple crystalline or amorphous forms. In general, all physical forms are equivalent for the uses contemplated by the present disclosure and are intended to be within the scope of the present disclosure.

H. Kits Comprising Compounds of Formula (II)

In some embodiments, the presently disclosed subject matter provides a kit comprising a compound of formula (II):



wherein: n is 1, 3, or 5; R_1 is selected from the group consisting of alkyl, substituted alkyl, heteroalkyl, substituted heteroalkyl, cycloalkyl, substituted cycloalkyl, cycloheteroalkyl, substituted cycloheteroalkyl, aryl, substituted aryl, aralkyl, hydroxyl, alkoxy, halogen, cyano, nitro, CF_3 , hydroxyalkyl, hydroxycycloalkyl, alkoxycycloalkyl, aminoalkyl, acyloxy, alkylaminoalkyl, and alkoxy carbonyl; or a pharmaceutically acceptable salt thereof. The kit also can include suitable carriers, buffers, and the like. The kit, in some embodiments, also includes instructions for uses of the compound of formula (II) including the several embodiments disclosed herein.

III. General Definitions

Although specific terms are employed herein, they are used in a generic and descriptive sense only and not for purposes of limitation. Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly
5 understood by one of ordinary skill in the art to which this presently described subject matter belongs.

Following long-standing patent law convention, the terms “a,” “an,” and “the” refer to “one or more” when used in this application, including the claims. Thus, for example, reference to “a subject” includes a plurality of subjects, unless the context
10 clearly is to the contrary (e.g., a plurality of subjects), and so forth.

Throughout this specification and the claims, the terms “comprise,” “comprises,” and “comprising” are used in a non-exclusive sense, except where the context requires otherwise. Likewise, the term “include” and its grammatical variants are intended to be non-limiting, such that recitation of items in a list is not to the
15 exclusion of other like items that can be substituted or added to the listed items.

For the purposes of this specification and appended claims, unless otherwise indicated, all numbers expressing amounts, sizes, dimensions, proportions, shapes, formulations, parameters, percentages, parameters, quantities, characteristics, and other numerical values used in the specification and claims, are to be understood as
20 being modified in all instances by the term “about” even though the term “about” may not expressly appear with the value, amount or range. Accordingly, unless indicated to the contrary, the numerical parameters set forth in the following specification and attached claims are not and need not be exact, but may be approximate and/or larger or smaller as desired, reflecting tolerances, conversion factors, rounding off,
25 measurement error and the like, and other factors known to those of skill in the art depending on the desired properties sought to be obtained by the presently disclosed subject matter. For example, the term “about,” when referring to a value can be meant to encompass variations of, in some embodiments, $\pm 100\%$ in some
embodiments $\pm 50\%$, in some embodiments $\pm 20\%$, in some embodiments $\pm 10\%$, in
30 some embodiments $\pm 5\%$, in some embodiments $\pm 1\%$, in some embodiments $\pm 0.5\%$, and in some embodiments $\pm 0.1\%$ from the specified amount, as such variations are appropriate to perform the disclosed methods or employ the disclosed compositions.

Further, the term “about” when used in connection with one or more numbers or numerical ranges, should be understood to refer to all such numbers, including all

numbers in a range and modifies that range by extending the boundaries above and below the numerical values set forth. The recitation of numerical ranges by endpoints includes all numbers, e.g., whole integers, including fractions thereof, subsumed within that range (for example, the recitation of 1 to 5 includes 1, 2, 3, 4, and 5, as well as fractions thereof, e.g., 1.5, 2.25, 3.75, 4.1, and the like) and any range within that range.

EXAMPLES

The following Examples have been included to provide guidance to one of ordinary skill in the art for practicing representative embodiments of the presently disclosed subject matter. In light of the present disclosure and the general level of skill in the art, those of skill can appreciate that the following Examples are intended to be exemplary only and that numerous changes, modifications, and alterations can be employed without departing from the scope of the presently disclosed subject matter. The synthetic descriptions and specific examples that follow are only intended for the purposes of illustration, and are not to be construed as limiting in any manner to make compounds of the disclosure by other methods.

EXAMPLE 1

METHODS AND MATERIALS

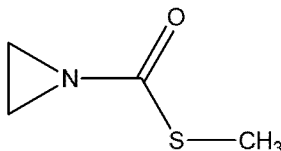
All commercially available reagents were used as purchased without further purification. Most chemicals and reagents were purchased from Sigma-Aldrich (St. Louis, Missouri) and Fisher (Fairlawn, New Jersey) unless noted. Aziridine was purchased from ChemService (West Chester, Pennsylvania). Histone H3 acetyl-K9 antibody was purchased from Abcam (Cambridge, Massachusetts). Histone H3 acetyl-K18 antibody was purchased from Cell Signaling (Danvers, Massachusetts). Histone H3 acetyl-K27 antibody was purchased from Active Motif (Carlsbad, California). Peptides H4K5ac and H4K8ac were bought from AnaSpec (Fremont, California). NMR spectra were collected on a Varian Mercury 400 MHz spectrometer. Chemical shifts are reported as δ in units of parts per million (ppm). Peptide samples were spotted with MALDI matrix (α -cyano-4-hydroxycinnamic acid saturated in 50% aq. acetonitrile with 0.1% TFA). Spectra were acquired on a Perseptive Biosystems Voyager-DETM PRO. Amino acid analysis was carried out to confirm the concentration of peptides used in the bromodomain pulldowns, HDAC

and kinase assays (W.M. Keck Foundation Biotechnology Resource Laboratory, Yale University).

EXAMPLE 2

5 SYNTHESIS OF SUBSTITUTED THIOCARBONYL-AZIRIDINES

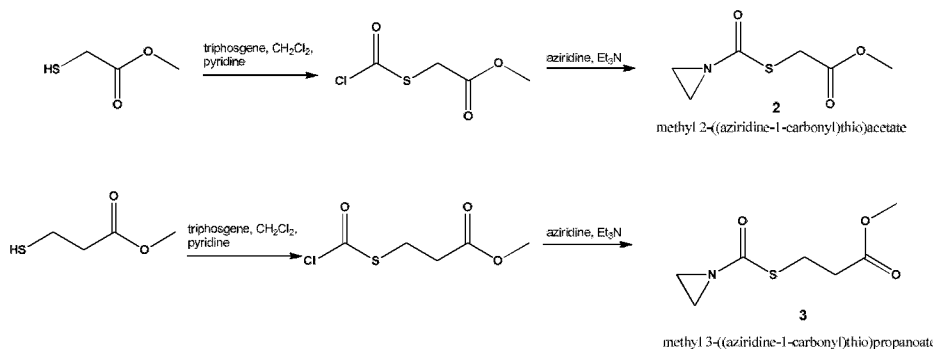
a. Methylthiocarbonyl-aziridine (1-aziridinecarbothioic acid, S-methyl ester) (1)



To a stirred solution of aziridine (100 μ L, 1.93 mmol) in Et₂O (30 mL) at -80°C was added triethylamine (0.27 mL, 1.93 mmol), followed by addition of methyl chlorothioformate (0.16 mL, 1.89 mmol). Fishbein, P.L., Kohn, H. Synthesis and antineoplastic activity of 1a-Formyl and 1a-Thioformyl derivatives of mitomycin C and 2-methyaziridine. *J. Med. Chem.* 30:1767-1773 (1987). Other alkylchlorothioformate reagents can be used to prepare analogues of methylthioformate having substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl substituent groups on the sulfur atom.

After 3 h, the reaction was diluted with Et₂O:ddH₂O (15 mL:10 mL). The organic layer was separated and washed with 0.01 M HCl (5 mL $\times$ 2), ddH₂O (5 mL $\times$ 2), brine (10 mL $\times$ 2). The organic layer was dried over MgSO₄ and concentrated to afford a colorless oil (0.22 g, quantitative) without further purification. ¹H NMR (CDCl₃): δ 2.33 (s, 3H), 2.32 (br, 4H) ¹³C NMR (CDCl₃): δ 182.6, 27.1, 13.1 HRMS (CI/EI) m/z: calculated for 118.0313 [M+H]⁺ found 118.0317.

b. Methyl 2-((aziridine-1-carbonyl)thio)acetate (2) and methyl 3-((aziridine-1-carbonyl)thio)propanoate (3)



25

To a solution of methyl-3-mercapto propionate (216 μ L, 2.0 mmol) in dichloromethane (3 mL) was added pyridine (160 μ L, 2.0 mmol) and triphosgene (0.236 g, 0.8 mmol) at 0°C (Jiao, L., et al. *Synthesis* 6, 0659-0665 (2006)). The reaction was continued stirring for 2 h at 0°C. Then 2 mL ddH₂O was added to the reaction mixture. The aqueous layer was extracted with dichloromethane (5 mL $\times$ 3). The combined organic layer was dried over MgSO₄ and concentrated to yield a colorless oil which was dissolved in 2 mL dichloromethane and directly used without purification. The dichloromethane solution was added dropwise to a stirring solution of ethyleneimine (100 μ L, 1.93 mmol) in dichloromethane and diethylisopropylamine (0.36 mL, 2.0 mmol) at -78 °C. After 2 h, the volatile was removed and redissolved in ethyl ether. The organic layer was washed with ddH₂O, 0.01 M HCl, ddH₂O, brine. The it was dried over MgSO₄ and concentrated to yield a faint yellow oil, which was purified via flash chromatography with hexanes:ethyl acetate (2:1) to provide a colorless oil (yield 43%-60%). ¹H-NMR for compound **3** in CDCl₃: δ = 2.15 (s, 4H), 2.65 (t, 2H), 3.08 (t, 2H), 3.72 (s, 3H). ¹H-NMR for compound **2** in CDCl₃: δ = 2.38 (s, 4H), 3.70 (s, 2H), 3.78 (s, 3H).

EXAMPLE 3

PEPTIDE SYNTHESIS AND ALKYLATION

Standard Fmoc (9H-fluoren-9-ylmethoxycarbonyl) solid phase peptide synthesis techniques were used to prepare peptides. Cysteine was used at the corresponding position to replace lysine. Universal deprotection and cleavage of the peptides from the Wang resin were accomplished with 95:5 TFA:H₂O in the presence of phenol, ethanedithiol, and thioanisole for 4 hrs at 25 °C. Precipitation of the peptides with diethyl ether followed by lyophilization yielded crude peptide as an off-white solid. The crude peptides were purified by preparative RP-HPLC (A = ddH₂O with 0.1% TFA; B = MeCN with 0.1% TFA). For peptide alkylation, HPLC-purified peptides were dissolved in 100 mM ammonium bicarbonate (pH 8). Methylthiocarbonyl-aziridine (MTCA) was added (50-200 mM from a 1 M acetonitrile stock) and the reaction was allowed to proceed for 3 hrs at room temperature. The reactions were lyophilized and analyzed by MALDI-TOF MS. For protein deacetylase HPLC assay, p53 peptide containing MTCTK was purified from the reaction mixture by preparative RP-HPLC and its structure was confirmed by

MALDI-TOF MS. Modified H4 peptides for bromodomain pulldown assay were prepared in the same way.

EXAMPLE 4

5 HISTONE EXPRESSION AND PURIFICATION

Two general methods were used to prepare histone proteins. Mutant *Xenopus* histone 3 (H3 C110A/K9C, aa 1-135, leaving an amino-terminal Ala) was subcloned into a pET SUMO vector and overexpressed in *E. coli* BL21-CodonPlus (DE3)-RIPL cells (Stratagene). Cells were grown to OD600 of 1.0 in LB media at 37°C then
10 induced with a final concentration of 1 mM IPTG and grown for 3 hrs at 37 °C. Cell pellets were harvested by centrifugation at 5000 rpm for 20 min and resuspended in ice cold lysis buffer (50 mM Tris-HCl, 500 mM NaCl, 0.1 mM PMSF, 1 mM TCEP, 5% glycerol, 1 mM EDTA, pH 7.5). The cell suspensions were then lysed via double passage through a French press (16,000 psi), and the lysates clarified by
15 centrifugation at 20,000 rpm for 30 min. Pellets were washed twice with wash buffer (50 mM Tris-HCl, 500 mM NaCl, 0.1 mM PMSF, 5% glycerol, 1 M urea, 1 mM TCEP, pH 7.5) + 1% v/v Triton X-100. Pellets were then treated twice with wash buffer without Triton X-100. Pellets were then suspended in extraction buffer (50 mM Tris-HCl, 500 mM NaCl, 5% glycerol, 6 M guanidine HCl, 1 mM TCEP, 4%
20 DMSO, pH 7.5) twice. The clarified supernatants were loaded on to a 2 mL Ni Sepharose 6 Fast Flow column (GE Healthcare) that was pre-equilibrated with wash buffer (50 mM Tris-HCl, 500 mM NaCl, 5% glycerol, 6 M urea, 1 mM TCEP, pH 7.5) containing 10 mM imidazole. The column was then eluted with 5 × 2 mL fractions of wash buffer containing 100 mM imidazole. SUMO-H3 containing
25 fractions were pooled and dialyzed (10 kDa cutoff, 3 × 1 L) against dialysis buffer (50 mM Tris-HCl, 100 mM NaCl, 5% glycerol, 1 mM TCEP, pH 7.5) containing 3 M urea, 1 M urea, 1 M urea respectively. The dialyzed protein was aliquoted and stored at -80°C. Final protein concentration was determined by Bradford assay using BSA as the standard. Purification of SUMO-H3 by this procedure yielded approximately
30 25 mg of protein/L. To SUMO-H3 (25 mL, 1.1 mg/mL) in dialysis buffer containing 1 M urea was gently added 50 mL SUMO protease (1U/μL, Invitrogen), then the mixture was allowed to stand at 4°C overnight. The SUMO protease cleavage reaction then underwent purification via a Mono-S column (high salt buffer: 50 mM Tris-HCl, 1 M NaCl, 5% glycerol, 1 mM TCEP, pH 7.5, 1 M urea; low salt buffer:

50 mM Tris-HCl, 100 mM NaCl, 5% glycerol, 1 mM TCEP, pH 7.5, 1 M urea). The fractions containing H3 were pooled and dialyzed in 0.1% TFA aqueous solution (3.5 kDa cutoff, 3 times 1 L). The dialyzed protein was lyophilized to a powder and stored at -80°C. WT and Mutant Xenopus histone 3 (H3 C110A/K18C, aa1-135) was
5 subcloned into a pET vector and transformed in E. coli BL21(DE3) pLysS cells (Stratagene). Cells were grown to OD₆₀₀ of 1.0 in 2XYT media (16 g B. Tryptan, 10 g Yeast Extract, 5g NaCl in 1 L ddH₂O) at 37°C then induced with a final concentration of 0.5 mM IPTG and grown for 3 hrs at 37°C. Cell pellets were harvested by centrifugation at 5000 rpm for 20 min and resuspended in ice cold lysis
10 buffer (50 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, 1 mM PMSF, pH7.5). Cells were then lysed via double pass on a French press (16000 psi), and the lysates clarified by centrifugation at 20,000 rpm for 30 min. Pellets were washed twice with lysis buffer plus 1% v/v Triton X-100 and again with lysis buffer twice to remove excess Triton-X. The pellets were then suspended and incubated in extraction buffer
15 (20 mM Tris-HCl, 7 M guanidine HCl, 10 mM DTT, 4% DMSO, pH 7.5) twice for 30 min. The pellets were spun down at 20,000 rpm for 15 min at 4°C. The combined supernatants were dialyzed against 1 mM 2-mercaptoethanol in ddH₂O three times before lyophilization.

20

EXAMPLE 5

HDAC8 HPLC ASSAY

Enzymatic reactions were performed in plastic tubes with 0.3 mM peptide, mixed with BIOMOL assay buffer on ice and initiated with addition of 3 mM human HDAC8 (BIOMOL Research Laboratories, Inc., Plymouth Meeting, Pennsylvania) at
25 37°C. The reactions were quenched with the stop solution (1.0 M HCl and 0.16 M acetic acid) after 20 min. After addition of 10% v/v 8 M urea to the quenched assay mixture, the supernatant was analyzed by analytical C₁₈ RP-HPLC, eluting with linear increase from 5% B to 25% B over 45 min (A: ddH₂O containing 0.05% TFA; B: acetonitrile containing 0.05% TFA) and UV monitoring at 214 nm. Fatkins, D.G., et
30 al., N-Thioacetyl-lysine: a multi-facet functional probe for enzymatic protein lysine N-deacetylation. *Bioorg. Med. Chem. Lett.* 16:3651-3656 (2006).

EXAMPLE 6

Sir2 HPLC ASSAY

A typical assay solution contains the following components: Metal buffer (25 mM Tris (pH 8.0), 137 mM NaCl, 2.7 mM KCl, 1 mM MgCl₂), 0.5 mM β-NAD⁺,
5 0.3 mM peptide, and 0.5 mM Sir2Tm (from *Thermatoga maritima*) enzyme (prepared as described in Smith et al.). Smith, J. S., et al., Sir2 family of NAD dependent protein deacetylases. *Meth. Enz.* 353:282-300 (2002).

The reaction was initiated with addition of Sir2Tm at 37 °C and quenched with the stop solution (0.1 M HCl and 0.16 M acetic acid) after 20 min. After addition of
10 10% v/v 8 M urea to the quenched assay mixture, the reaction mixture was centrifuged at 13,000 rpm for 10 min and the supernatant was analyzed by analytical C18 RP-HPLC, eluting with linear increase from 5% B to 25% B over 45 min (A: ddH₂O containing 0.05% TFA; B: acetonitrile containing 0.05% TFA) and UV monitoring at 214 nm. Smith, J. S., et al., Sir2 family of NAD dependent protein
15 deacetylases. *Meth. Enz.* 353:282-300 (2002).

EXAMPLE 7

BROMODOMAIN PULLDOWN ASSAY

Cos7 cells grown in standard conditions in 10 cm diameter petri dishes (in this
20 case 12 dishes were used), were transfected with 5 mg of a GFP-Brdt construct described previously, Pivot-Pajot, C., et al., Acetylation dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol. Cell Biol.* 23:5354-5365 (2003); Morinière, J. et al. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 461:664-668 (2009), using
25 the lipofectamine reagent according to the supplier's instructions (Invitrogen). Twenty-four hrs posttransfection, the cells from each petri dish were recovered, pooled and disrupted in 1.2 mL of LSDB500 lysis buffer (20% glycerol, 3 mM MgCl₂, 50 mM HEPES [pH 7.9], 500 mM KCl, 0.1% NP-40, containing one Complete Protease Inhibitor Cocktail tablet [Roche]) for 15 min on ice. After
30 centrifugation, the supernatant was recovered and diluted with one volume (1.2 mL) of the lysis buffer without KCl (final KCl concentration becomes 250 mM, LSDB250). 100 μL of this extract were used for each pulldown as follows.

Extracts were incubated with 15-60 nmol of the competing peptides (H4 tail peptides aa 1-12 containing acetyl-Lys or MTCTK at positions 5 and 8 as well as

control peptides: unmodified H4 tail aa 1-15 and monoacetylated (at K5 or K8) H4 tail peptides aa 1-21) at 4°C for 30 min under agitation after which time 20 µL of streptavidin beads containing 3 nmol of biotin-tagged, tetra-acetylated histone H4 peptides (or non-modified H4 peptides) were added. Incubation was carried out for two additional hrs under agitation. Beads were washed twice with LSDB250 and once with PBS and finally re-suspended in 20 µL of SDS-PAGE loading buffer from which 5 µL were used for western analysis. GFP-Brdt retained on the beads was revealed with an anti-GFP antibody.

10

EXAMPLE 8

ACETYLATION OF HISTONE H3 BY P300

For p300-mediated acetylation of histone H3, 8 µg of purified recombinant H3 protein was incubated with 1 µM of p300 and 500 µM acetyl-CoA in 50 µL reaction volume in HAT buffer (50 mM HEPES, pH 8.0, 150 mM NaCl, 5% glycerol, 0.1 mM EDTA, 1 mM DTT, 0.05 mg/mL BSA) at 30°C for 30 min. Thompson, P. R. et al. Regulation of the p300 HAT domain via a novel activation loop. *Nature Struct. Mol. Biol.* 11:308-315 (2004). The reaction mixture was quenched by addition of loading dye. 1 µg of protein from the reaction was subjected to SDS-PAGE and analyzed by western blot.

20

EXAMPLE 9

WESTERN BLOTS

Histone H3 proteins were run on SDS-PAGE and transferred onto polyvinylidene difluoride membranes. For experiments with histone AcK9, AcK18, and AcK27 antibodies, 0.4 µg, 1 µg, and 0.6 µg of protein from each reaction was loaded per lane, respectively. Blots were washed in 10-20 mL Milli-Q water for 2-3 min with mild agitation, then blocked with 20 mL of TBS (50 mM Tris HCl, 150 mM NaCl, pH 7.4) with 5% nonfat dry milk for 15 min at room temperature with agitation. The blocking agents were then removed and blots were rinsed with 10-20 mL TBS. Membranes were incubated with anti-H3X antibody (1:1000 dilution) in TBS with 3% nonfat dry milk overnight at 4 °C.

30

The antibody solutions were removed and membranes were washed three times with TBS. The blots were then incubated with anti-rabbit IgG peroxidase conjugate with 3% nonfat dry milk (1:5000 dilution) in TBST for 1 hr at room

temperature. After washing with TBST (5 min $\times$ 5), proteins were visualized by chemiluminescence.

EXAMPLE 10

5 ELECTROSPRAY MASS SPECTROMETRY ANALYSIS OF MODIFIED HISTONE H3

Samples (concentration approximately 75 pmol μL^{-1}) were diluted to 250 fmol μL^{-1} using 1.8% acetonitrile/0.1 % trifluoroacetic acid (TFA). The samples (40 μL , 10 pmol) were separated using a 50 mm $\times$ 75 μm selfpacked C4 XBridge BEH 3 μm / 10 300 Å (Waters, Milford, MA) column following a 15 minute desalting wash using 20 mm $\times$ 75 μm Delta pack 15 μm (Waters) self-packed trap and 95 % mobile phase A (0.1 % formic acid). Proteins were eluted on a 20 minute gradient of 5 – 100 % B (90 % acetonitrile/0.1 % formic acid). Eluting proteins were sprayed at 2.2 kV through a 10 μm fused silica nanospray needle (New Objectives, Inc.) into a QSTAR 15 Pulsar quadrupole orthogonal TOF tandem mass spectrometer (Applied Biosystems, Foster City, California). Full scan MS data were acquired from 400 m/z to 2,000 m/z for 46 minutes with an accumulation time of one second. Chromatograms were manually inspected for spectra containing characteristic protein charge envelopes. Protein masses were reconstructed from raw spectra using the Bayesian Protein 20 Reconstruct tool in Analyst QS 1.1 (Applied Biosystems) with: a specified mass range of 10,000 – 20,000 Da; step mass of 0.1; signal-to noise threshold of 20%; hydrogen adduct; and, 20 iterations. Standard errors on these measurements are estimated at $\pm$ 3 mass units.

25 EXAMPLE 11

PROTEIN PREPARATION FOR RTT109-VPS75 COMPLEXES

Production of co-expressed wild-type and mutant Rtt109-Vps75 complexes was carried out analogously to methods described previously. Tang, Y., et al., Fungal Rtt109/KAT11 histone acetyltransferase: an unexpected structural homolog of 30 metazoan p300/CBP. *Nature Struct. Mol. Biol.* 15:738-745 (2008).

Briefly, DNAs encoding full-length Vps75 and full length Rtt109 (wild-type or bearing particular mutations) were subcloned into the BamHI/Sall and NdeI/XhoI sites, respectively, of a modified 6His-TEV-pCDF-Duet1 vector. Plasmids were transformed into *E. coli* BL21(DE3) Gold cells. Transformed cells were grown at

37°C in LB media and induced at 20°C with 0.8 mM IPTG for complex overnight overexpression when cell density reached an OD₆₀₀=0.8. Cell pellets were harvested by centrifugation and lysed by sonication in 20 mM Hepes pH 7.5, 500 mM NaCl, 10 mM imidazole (pH 7.0), and 5 mM 2-mercaptoethanol, supplemented with
5 protease cocktail. The tightly associated complex was purified to homogeneity using a combination of Ni-affinity, overnight cleavage with TEV protease, followed by MonoQ anion exchange (in 20 mM Hepes pH7.0 buffer with a NaCl gradient from 50 mM to 1000 mM and 5 mM 2-mercaptoethanol) and Superdex 200 gel filtration chromatography (in 20 mM Hepes pH7.0, 150 mM NaCl and 5 mM 2-
10 mercaptoethanol). The protein complexes were concentrated to 10 mg/mL in the gel filtration buffer and then dialyzed against 25 mM Hepes pH7.5, 300 mM NaCl and 0.1 mM TCEP before flash freezing them in liquid nitrogen until further use.

EXAMPLE 12

15 HAT ENZYME ASSAY

The radioactive HAT assay was adopted from previously published methods. Tang, Y., et al., Fungal Rtt109/KAT11 histone acetyltransferase: an unexpected structural homolog of metazoan p300/CBP. *Nature Struct. Mol. Biol.* 15:738-745 (2008); Lau, O.D. et al. p300/CBP-associated factor histone acetyltransferase
20 processing of a peptide substrate. Kinetic analysis of the catalytic mechanism. *J. Biol. Chem.* 275:21953-21959 (2000); Thompson, P. R., et al., Transcriptional coactivator protein p300. Kinetic characterization of its histone acetyltransferase activity. *J. Biol. Chem.* 276:33721-33729 (2001).

Enzymatic activity of the Rtt109/Vps75 complex and mutants was determined
25 using this gel based assay that measures incorporation of a ¹⁴C-labeled group from acetyl-CoA into the *X. laevis* histone H3 substrate. Reactions were carried out at 30 °C for times varying from 1 to 5 minutes under the following reaction conditions: 0.1 M Hepes, pH 7.9, 50 mM NaCl, 0.1 mg/mL BSA, 1 mM DTT, 0.1 mM EDTA. The mixture was allowed to equilibrate at 30 °C for 10 minutes before the reaction
30 was initiated with enzyme. After the appropriate time the reaction was quenched with 6 X Tris-Tricine gel loading buffer which contained: 0.2 M Tris-Cl pH 6.8, 40% v/v glycerol, 14% w/v SDS, 0.3M DTT, and 0.06% w/v coomassie blue. The ¹⁴C-labeled histone substrates were separated from reactants by running the reaction out on a 16.5% Tris-Tricine SDS-PAGE gel. The rate of ¹⁴C-incorporation into histone H3

was quantified by autoradiography relative to a ^{14}C -BSA standard. Enzyme activity for both WT and mutant Rtt109 complexes was linear with respect to time and enzyme concentration in the ranges used. All assays were performed in duplicate, and the duplicate assays generally agreed within 20%. Alkylation of the Rtt109 mutant
5 complexes was undertaken under the following reaction conditions. First the mutant complexes were dialyzed overnight at 4 °C into 100 mM NH_4HCO_3 , pH 8.0. After dialysis, methylthiocarbonyl-aziridine (1M in acetonitrile solution) was added to a final concentration of 50 mM and the reaction was allowed to proceed for 200 minutes at room temperature. The reaction was then dialyzed twice, for one hour
10 each time at 4 °C, into 10 mM Hepes, pH 7.5, 350 mM NaCl, and 10 mM DTT. The reacted mutant complexes were then assayed for histone acetyltransferase activity as detailed above. It should be noted that because of the much greater activity of the wt Rtt109 complex, effects of alkylation conditions on its HAT activity could not be directly compared because it was diluted to a much greater extent.

15

EXAMPLE 13

PROTEIN PREPARATION FOR CK2 α

DNA encoding full-length CK2 α was subcloned into the pTYB2 vector (Impact-CN system, New England Biolabs). Point mutations for C147A, C220A, and
20 K102C were made using standard Quikchange protocols and mutations were confirmed by DNA sequencing of the complete open reading frames. Plasmids corresponding to CK2 α containing all three point mutations (K102C, C147A, C220A) and CK2 α containing only cysteine to alanine mutations (C147A, C220A) were transformed into E. coli BL21 (DE3) codon plus cells. Transformed cells were grown
25 at 37°C in LB media for 2 h until cell density reached $\text{OD}_{600}=0.6$, then induced at 16°C with 350 μM IPTG for overexpression for 20 hrs. Cell pellets were harvested by centrifugation and lysed via double-pass on French press in 25 mM Hepes pH 8.0, 150 mM NaCl, 1 mM MgSO_4 , 5% glycerol, 5% ethylene glycol, 1 mM EDTA, and 100 μM PMSF. Cell lysates were clarified by centrifugation at 15,000 rpm,
30 supernatant was loaded onto chitin beads (New England BioLabs) for affinity purification and washed with 25 mM Hepes pH 7.5, 250 mM NaCl, 1 mM EDTA, 0.1% Triton X-100 followed by a wash with 50 mM Hepes pH 7.5, 250 mM NaCl, 1 mM EDTA. Full-length CK2 α was eluted from the chitin beads with 50 mM DTT in 50 mM Hepes pH 7.5, 250 mM NaCl, 1 mM EDTA at room temperature for 16 hrs.

The purified proteins were concentrated to >1 mg/mL, dialyzed into 100 mM NH_4CO_3 pH 8.0, 1 mM 2-mercaptoethanol, then treated with methylthiocarbonyl-aziridine (2 M in acetonitrile solution) for a final concentration of 100 mM for 3 h at room temperature. The reaction was dialyzed extensively at 4°C into 20 mM Hepes pH 7.5, 150 mM NaCl, 5% glycerol, 10 mM 2-mercaptoethanol.

Following dialysis, the concentrations of the reacted mutant proteins were determined via Bradford assay using BSA (bovine serum albumin) as a standard. Subsequently, they were assayed for kinase activity as described below.

10

EXAMPLE 14

KINASE ENZYME ASSAY

To determine the kinase activity for the MCTA reacted mutant CK2 α proteins and those not treated with MTCA, radiometric kinase assays were carried out in 50 mM Hepes (pH 7.5), 75 mM NaCl, 10 mM MgCl_2 , 1 mM DTT, 5% glycerol, 250 $\mu\text{g/mL}$ BSA, and 40 μM biotinylated $\text{NH}_2\text{-RRRADDSDDDDK-NH}_2$ (SEQ. ID. NO. 8) as the peptide substrate and 100 μM $\gamma\text{-}^{32}\text{P-ATP}$ in a 25 μL reaction volume with 5 nM kinase. The peptide substrate was prepared using standard Fmoc solid-phase peptide synthesis including a biotinylated lysine residue (NovaBiochem) and purified by reverse phase HPLC. Reactions were initiated by the addition of kinase (5 nM final concentration), carried out at 30°C for 6 min, and stopped by the addition of 10 μL of 100 mM EDTA. To each sample, 10 μL of 10 mg/mL avidin (Thermo Scientific) was added, and all samples were transferred to centrifugal filtration units with 30,000 nominal molecular weight limit (NMWL) membranes (Millipore) and washed three times with 100 μL wash solution (0.5 M Phosphate, 0.5 M NaCl, pH 8.5). Radioactivity of the washed filters was quantified using scintillation counting. The limiting substrate turnover was less than 10% for all rate measurements. Experiments were done in duplicate and error bars correspond to the span of values measured.

30

REFERENCES

All publications, patent applications, patents, and other references mentioned in the specification are indicative of the level of those skilled in the art to which the presently disclosed subject matter pertains. All publications, patent applications, patents, and other references are herein incorporated by reference to the same extent

as if each individual publication, patent application, patent, and other reference was specifically and individually indicated to be incorporated by reference. It will be understood that, although a number of patent applications, patents, and other references are referred to herein, such reference does not constitute an admission that
5 any of these documents forms part of the common general knowledge in the art.

Wang, L., Tang, Y., Cole, P. A., and Marmorstein, R. *Curr. Opin. Struct. Biol.* 18:741 (2008).

Choudhary, C., Kumar, C., Gnad, F., Nielsen, M. L., Rehman, M., Wlather, T. C., Olsen, J. V., and Mann, M. *Science* 325:834 (2009).

10 Wang, Q., Zhang, Y., Yang, C., Xiong, H., Lin, Y., Yao, J., Li, H., Xie, L., Zhao, W., Yao, Y., Ning, Z., Zeng, R., Xiong, Y., Guang, K., Zhao, S., and Zhao, G. *Science* 327:1004 (2010).

Cole, P. A. *Nat. Chem. Biol.* 4:590 (2008).

Muir, T. W., Sondhi, D., and Cole, P. A. *Proc. Natl. Acad. Sci. U.S.A.* 95:6705
15 (1998).

Shogren-Knaak, M., Ishii, H., Sun, J. M., Pazin, M. J., Davie, J. R., and Peterson, C. L. *Science* 311:844 (2006).

Neumann, H., Peak-Chew, S. Y., and Chin, J. W. *Nat. Chem. Biol.* 4:232 (2008).

20 Simon, M. D., Chu, F., Racki, L. R., de la Cruz, C. C., Burlingame, A. L., Panning, B., Narlikar, G. J., and Shokat, K. M. *Cell* 128:1003 (2007).

Fishbein, P.L., Kohn, H. Synthesis and antineoplastic activity of 1a-Formayl and 1a-Thioformyl derivatives of mitomycin C and 2-methyaziridine. *J. Med. Chem.* 30:1767-1773 (1987).

25 Morinière, J. et al. Cooperative binding of two acetylation marks on a histone tail by a single bromodomain. *Nature* 461:664-668 (2009).

Fatkins, D.G., Monnot, A.D., Zheng, W. N-Thioacetyl-lysine: a multi-facet functional probe for enzymatic protein lysine N-deacetylation. *Bioorg. Med. Chem. Lett.* 16:3651-3656 (2006).

30 Thompson, P. R. et al. Regulation of the p300 HAT domain via a novel activation loop. *Nature Struct. Mol. Biol.* 11:308-315 (2004).

Tang, Y., Holbert, M. A., Wurtele, H., Meeth, K., Rocha, W., Gharib, M., Jiang, E., Thibault, P., Verrault, A., Cole, P. A., Marmorstein, R. Fungal

- Rtt109/KAT11 histone acetyltransferase: an unexpected structural homolog of metazoan p300/CBP. *Nature Struct. Mol. Biol.* 15:738-745 (2008).
- Stavropoulos, P., Nagy, V., Blobel, G., and Hoelz, A. *Proc. Natl. Acad. Sci. U.S.A.* 105:12236 (2008).
- 5 Pinna, L. A. and Allende, J. E. *Cell. Mol. Life Sci.* 66:1795 (2009).
- Lu, W., Gong, D., Bar-Sagi, D., and Cole, P. A. *Mol. Cell* 8:759 (2001).
- Smith, J.S., Avalos, J., Celic, I., Muhammad, S., Wolberger, C., Boeke, J.D. Sir2 family of NAD dependent protein deacetylases. *Meth. Enz.* 353:282-300 (2002).
- Pivot-Pajot, C., Caron, C., Govin, J., Vion, A., Rousseaux, S., Khochbin, S.
- 10 Acetylation dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol. Cell Biol.* 23:5354-5365 (2003).
- Lau, O.D. et al. p300/CBP-associated factor histone acetyltransferase processing of a peptide substrate. Kinetic analysis of the catalytic mechanism. *J. Biol. Chem.* 275:21953-21959 (2000).
- 15 Thompson, P.R., Kurooka, H., Nakatani, Y., and Cole, P.A. Transcriptional coactivator protein p300. Kinetic characterization of its histone acetyltransferase activity. *J. Biol. Chem.* 276:33721-33729 (2001).
- Glozak, M. A., et al., Acetylation and Deacetylation of Non-Histone Proteins, *Gene* 363:15-23 (2005).
- 20 Shokat et al., Site-Specific Installation of Methyl-Lysine Analogues Into Recombinant Histones, U.S. Patent Application Publication No. 2008/0199964 A1, published Aug. 21, 2008.
- Junutula, J. R., et al., Site-specific Conjugation of a Cytotoxic Drug to an Antibody Improves the Therapeutic Index, *Nature Biotechnology* 26 (8):925-932
- 25 (2008).
- Wakankar, A. A., et al., *Bioconjugate Chem.* 21:1588-1595 (2010).

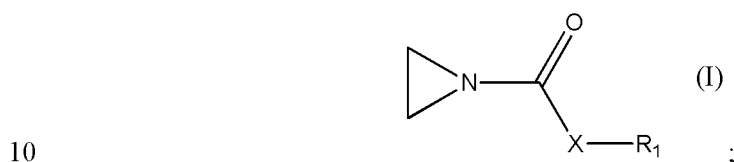
Although the foregoing subject matter has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be understood by those skilled in the art that certain changes and modifications can be

30 practiced within the scope of the appended claims.

THAT WHICH IS CLAIMED:

1. A method for modifying one or more cysteine residues of an amino acid, peptide, polypeptide or protein, the method comprising:

- 5 (a) providing an amino acid, peptide, polypeptide or protein comprising one or more cysteine residues; and
- (b) contacting the amino acid, peptide, polypeptide or protein with a compound of formula (I) for a period of time to modify one or more cysteine residues of the amino acid, peptide, polypeptide or protein:



wherein:

X is S or O;

15 R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(CH_2)_n-(C=O)-O-R_2$, and $-O-(C=O)-(CH_2)_n-(C=O)-OH$; and wherein:

n is an integer from 0 to 20; and

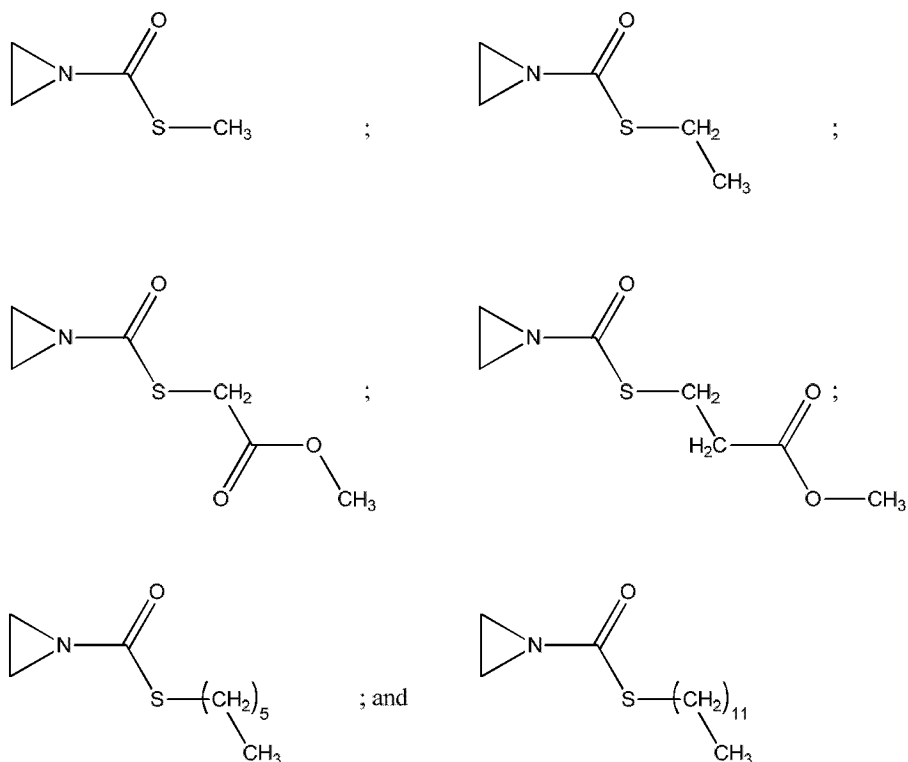
20 R₂ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl;

or a pharmaceutically acceptable salt thereof.

25

2. The method of claim 1, wherein X is S and R₁ is selected from the group consisting of unsubstituted, straight-chain C₁ to C₁₂ alkyl, $-(CH_2)-(C=O)-O-CH_3$, $-(CH_2)_2-(C=O)-O-CH_3$, and $-O-(C=O)-(CH_2)_2-(C=O)-OH$.

30 3. The method of claim 2, wherein the compound of formula (I) is selected from the group consisting of:



4. The method of claim 1, wherein the polypeptide is selected from the
 5 group consisting of a histone, a transcription factor, a nuclear receptor, a cytoskeletal
 protein; an adaptor protein involved in nuclear transport; a multifunctional protein; a
 molecular chaperone; and a virally encoded protein.

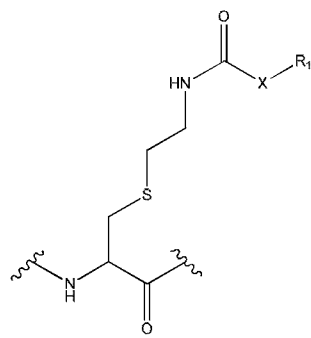
5. The method of claim 4, wherein the histone is selected from the group
 10 consisting of H1, H2A, H2B, H3, H4, and H5.

6. The method of claim 4, wherein the transcription factor is selected
 from the group consisting of p53, Yin Yang 1 (YY1), a HMG protein, STAT3, and c-
 MYC.

15

7. The method of claim 4, wherein the nuclear receptor is selected from
 the group consisting of androgen receptor (AR), estrogen receptor α (ER α), SHP, a
 GATA factor, erythroid Kruppel like factor (EKLF), MyoD, E2F/Rb, NF- κ B, HIF-1 α ,
 and Smad7.

8. The method of claim 4, wherein the cytoskeletal protein comprises α -tubulin.
- 5 9. The method of claim 4, wherein the adaptor protein involved in nuclear transport comprises importin- α .
- 10 10. The method of claim 4, wherein the multifunctional protein comprises Ku70.
11. The method of claim 4, wherein the molecular chaperone comprises Hsp90.
12. The method of claim 4, wherein the virally encoded protein is selected from the group consisting of E1A, an adenovirus protein; and HDAG.
13. The method of claim 1, further comprising replacing one or more selected lysine residues of an amino acid, peptide, polypeptide or protein with one or more cysteine residues to provide an amino acid, peptide, polypeptide or protein comprising one or more selected cysteine residues.
14. The method of claim 13, further comprising mutating one or more cysteine residues of a native amino acid, peptide, polypeptide or protein to a different amino acid residue prior to replacing the one or more selected lysine residues with one or more selected cysteine residues.
15. A modified amino acid, peptide, polypeptide or protein prepared by the method of claim 1.
16. An isolated amino acid, peptide, polypeptide or protein comprising one or more modified cysteine residues having the following structure:



wherein:

X is S or O;

5 R_1 is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C_1 to C_{20} alkyl or C_1 to C_{20} heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(CH_2)_n-(C=O)-O-R_2$, and $-O-(C=O)-(CH_2)_n-(C=O)-OH$; and wherein

10 n is an integer from 0 to 20; and

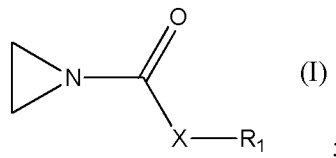
R_2 is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C_1 to C_{20} alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl.

15

17. A method for assaying the effect of modifying at least one lysine residue of a native amino acid, peptide, polypeptide or protein, the method comprising:

- 20 (a) mutating one or more selected cysteine residues in the native amino acid, peptide, polypeptide or protein to a different amino acid residue to provide a mutated amino acid, peptide, polypeptide or protein;
- (b) replacing one or more lysine residues with one or more cysteine residues in the mutated amino acid, peptide, polypeptide or protein to provide a mutated amino acid, peptide, polypeptide or protein
- 25 comprising one or more cysteine residues;
- (c) contacting the mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues with a compound of formula

(I) for a period of time to modify one or more cysteine residues of the mutated amino acid, peptide, polypeptide or protein:



wherein:

5

X is S or O;

10

R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{O}-\text{R}_2$, and $-\text{O}-(\text{C}=\text{O})-(\text{CH}_2)_n-(\text{C}=\text{O})-\text{OH}$; and wherein:

n is an integer from 0 to 20; and

15

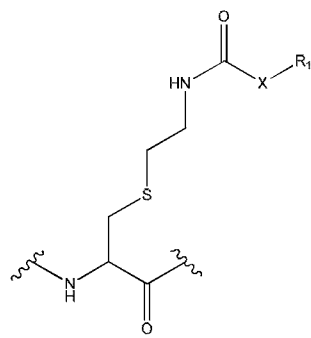
R₂ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl; or a pharmaceutically acceptable salt thereof; and

20

- (d) comparing an activity of the mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues modified with a compound of formula (I) to an activity of the native amino acid, peptide, polypeptide or protein, thereby assaying the effect of modifying at least one lysine residue of the native amino acid, peptide, polypeptide or protein.

25

18. A method for producing an antibody to a naturally occurring antigen, the method comprising immunizing a mammal with a polypeptide comprising one or more modified cysteine residues having the following structure:



wherein:

X is S or O;

5 R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(CH_2)_n-(C=O)-O-R_2$, and $-O-(C=O)-(CH_2)_n-(C=O)-OH$; and wherein

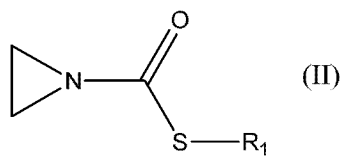
10 n is an integer from 0 to 20; and

R₂ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl.

15 wherein the polypeptide is substantially identical to the naturally occurring antigen and isolating an antibody from the mammal that binds the naturally occurring antigen.

19. A compound of formula (II):

20



wherein:

25 R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or

cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl,
alkoxy, $-(CH_2)_n-(C=O)-O-R_2$, and $-O-(C=O)-(CH_2)_n-(C=O)-OH$;
and wherein

n is an integer from 0 to 20; and

5

R_2 is selected from the group consisting of substituted or
unsubstituted, straight-chain or branched C_1 to C_{20} alkyl or heteroalkyl,
substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted
or unsubstituted aryl or heteroaryl;

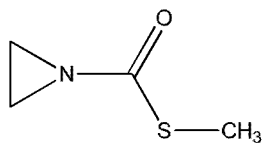
or a pharmaceutically acceptable salt thereof.

10

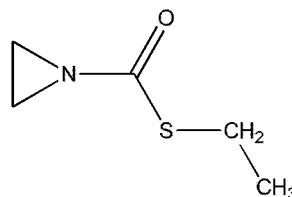
20. The compound of claim 17, wherein X is S and R_1 is selected from the
group consisting of unsubstituted, straight-chain C_1 to C_{12} alkyl, $-(CH_2)-(C=O)-O-$
 CH_3 , $-(CH_2)_2-(C=O)-O-CH_3$, and $-O-(C=O)-(CH_2)_2-(C=O)-OH$.

15

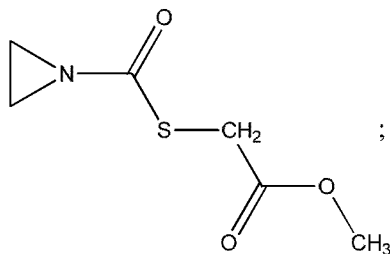
21. The compound of claim 18, wherein the compound of formula (II) is
selected from the group consisting of:



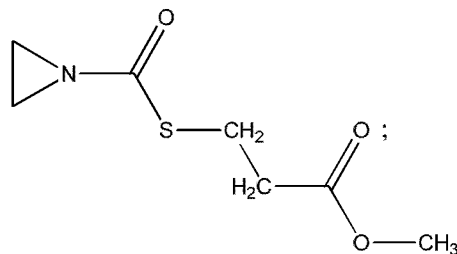
;



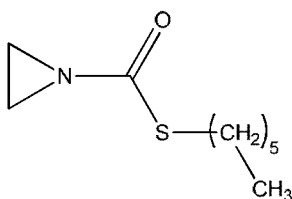
;



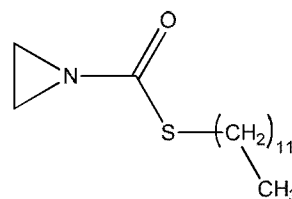
;



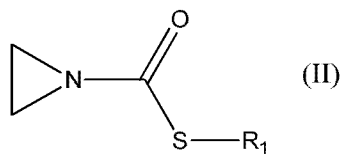
;



; and



22. A kit comprising a compound of formula (II):

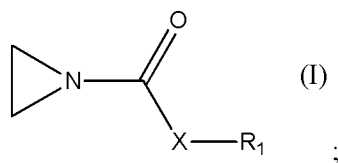


wherein:

- 5 R_1 is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C_1 to C_{20} alkyl or C_1 to C_{20} heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(CH_2)_n-(C=O)-O-R_2$, and $-O-(C=O)-(CH_2)_n-(C=O)-OH$;
 10 and wherein:
 n is an integer from 0 to 20; and
 R_2 is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C_1 to C_{20} alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted
 15 or unsubstituted aryl or heteroaryl;
 or a pharmaceutically acceptable salt thereof.

23. A method for modifying at least one lysine residue of a native amino acid, peptide, polypeptide or protein the method comprising:

- 20 (a) mutating one or more selected cysteine residues in the native amino acid, peptide, polypeptide or protein to a different amino acid residue to provide a mutated amino acid, peptide, polypeptide or protein;
 (b) replacing one or more lysine residues with one or more cysteine residues in the mutated amino acid, peptide, polypeptide or protein to
 25 provide a mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues; and
 (c) contacting the mutated amino acid, peptide, polypeptide or protein comprising one or more cysteine residues with a compound of formula (I) for a period of time to modify one or more cysteine residues of the
 30 mutated amino acid, peptide, polypeptide or protein:



wherein:

X is S or O;

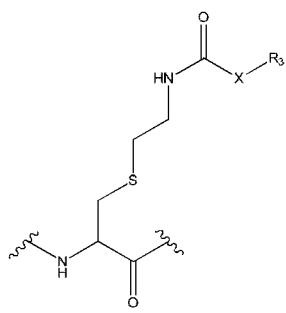
5 R₁ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or C₁ to C₂₀ heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl, alkoxy, $-(CH_2)_n-(C=O)-O-R_2$, and $-O-(C=O)-(CH_2)_n-(C=O)-OH$; and wherein:

10 n is an integer from 0 to 20; and

R₂ is selected from the group consisting of substituted or unsubstituted, straight-chain or branched C₁ to C₂₀ alkyl or heteroalkyl, substituted or unsubstituted cycloalkyl or cycloheteroalkyl, substituted or unsubstituted aryl or heteroaryl; or a pharmaceutically acceptable salt thereof;

15 thereby modifying at least one lysine residue of the native amino acid, peptide, polypeptide or protein.

24. A method for preparing an antibody-conjugate comprising one or more
20 cysteine residues having the following structure:

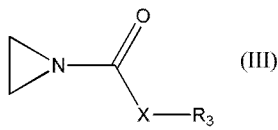


wherein:

X is S or O; and

R₃ is a therapeutic agent or an imaging agent; the method comprising:

- (a) introducing one or more cysteine residues in an antibody to provide a mutated antibody; and
- (b) reacting the cysteine residue of the mutated antibody with a compound of formula (III) to prepare an antibody-conjugate:



5

wherein: X is S or O; and
R₃ is a therapeutic agent or an imaging agent.

25. The method of claim 24, wherein the therapeutic agent is a cytotoxic agent.

10

26. The method of claim 25, wherein the cytotoxic agent is selected from the group consisting of diphtheria toxin, Pseudomonas exotoxin, a ribosome inactivating protein, ricin A, deglycosylated ricin A chain, abrin, alpha sarcin, aspergillin, restrictocin, a ribonuclease, bacterial endotoxin, the lipid A moiety of bacterial endotoxin, cholera toxin, a DNA alkylating agent, an agent that disrupts microtubules, an agent that binds DNA, a calicheamicin, and a maytansinoid.

15

27. The method of claim 24, wherein the imaging agent is selected from the group consisting of a fluorescent dye and a radionucleotide.

20

28. The method of claim 27, wherein the fluorescent dye is selected from the group consisting of carbocyanine, indocarbocyanine, oxacarbocyanine, thiocarbocyanine, merocyanine, polymethine, coumarine, rhodamine, xanthene, fluorescein, and a boron-dipyrromethane (BODIPY) compound.

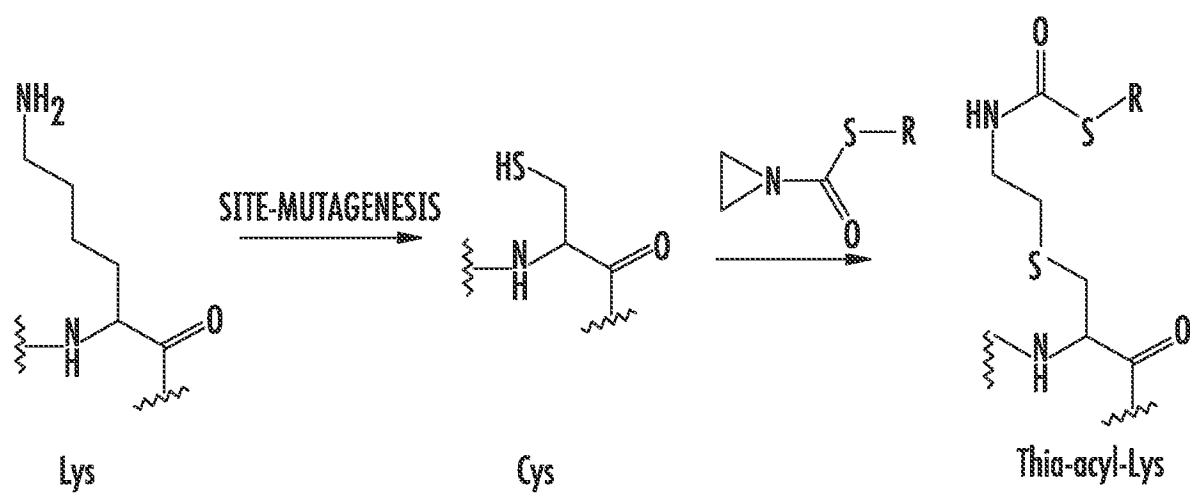
25

29. The method of claim 27, wherein the radionucleotide is selected from the group consisting of C-11, F-18, I-123, I-124, I-125, I-131, Tc-99m, In-111, Ga-67, Ga-68, Y-86, Y-90, Lu-177, Re-186, Re-188, Cu-64, Cu-67, Co-55, Co-57, Sc-47, Ac-225, Bi-213, Bi-212, Pb-212, Sm-153, Ho-166, and Dy-166.

30

30. An antibody-conjugate prepared by the method of claim 24.

1/13

**FIG. 1**

2/13

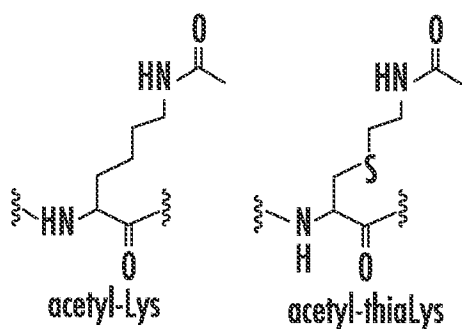


FIG. 2a

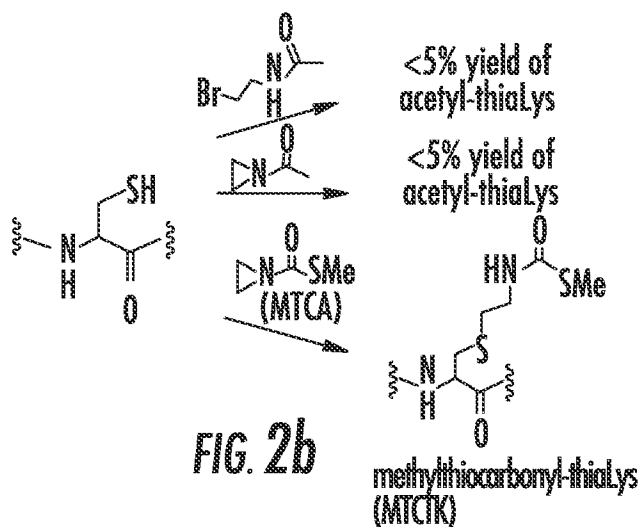


FIG. 2b

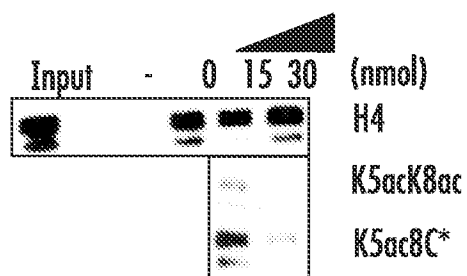


FIG. 2c

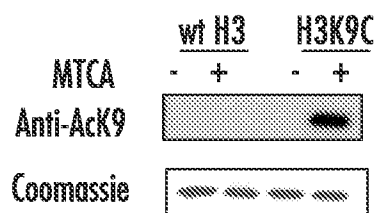


FIG. 2d

3/13

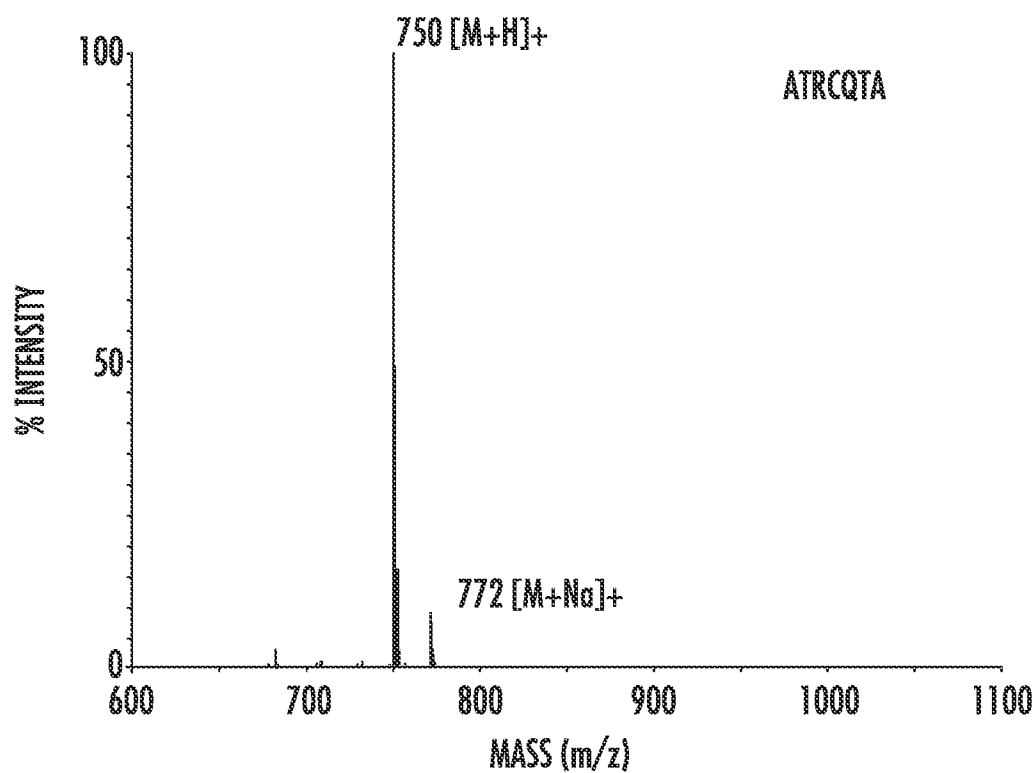


FIG. 3a

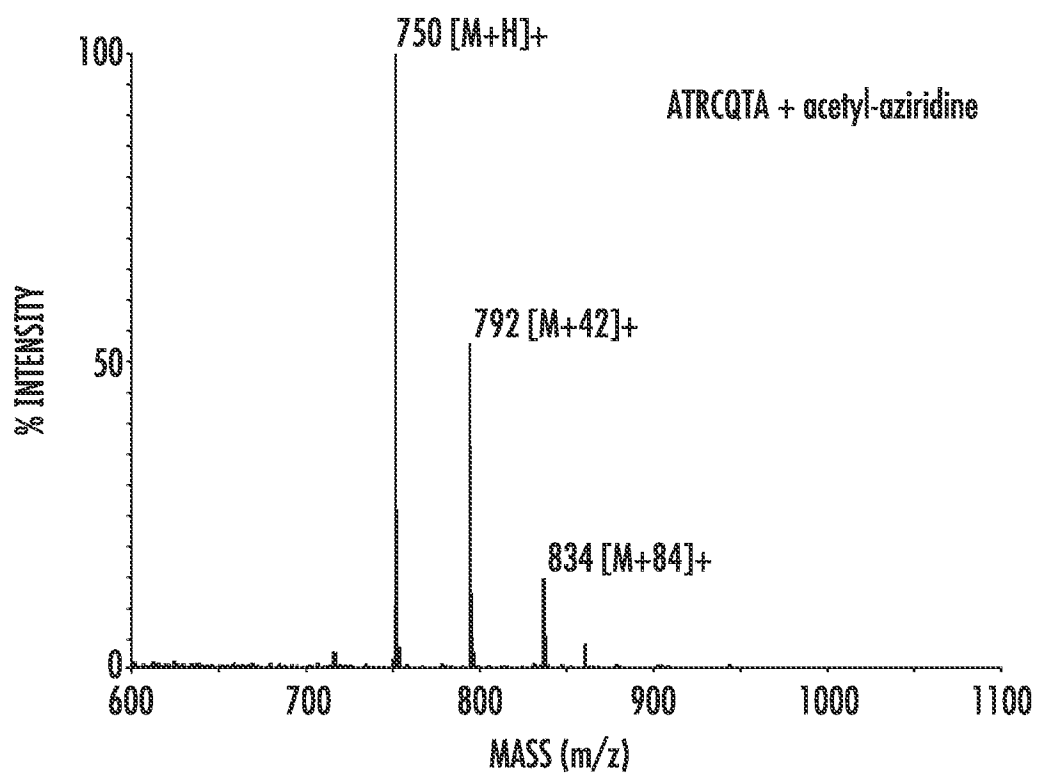


FIG. 3b

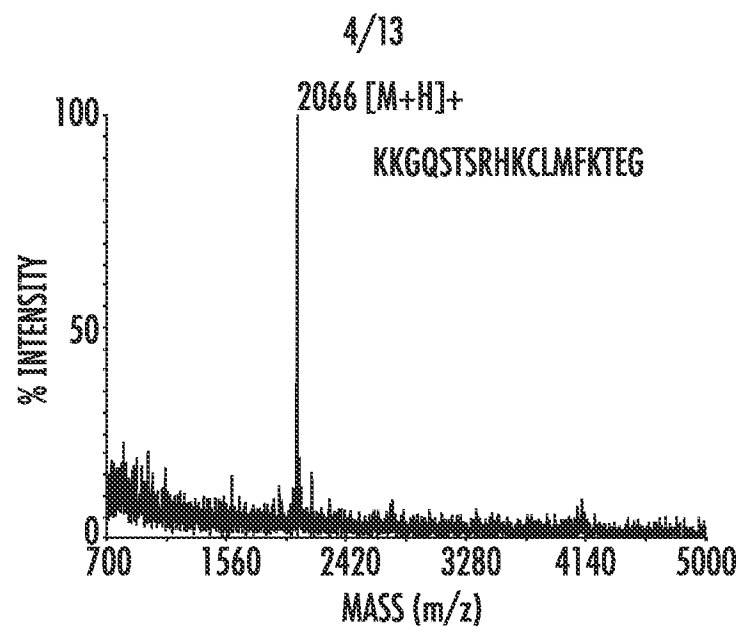


FIG. 4a

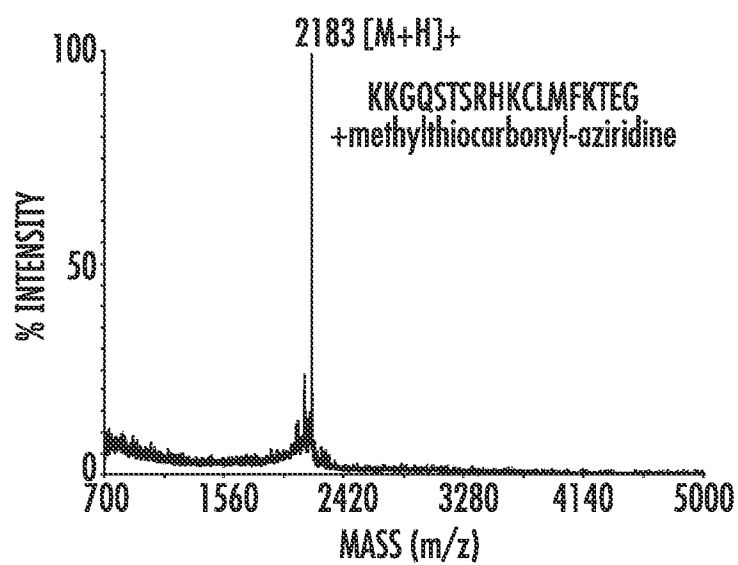


FIG. 4b

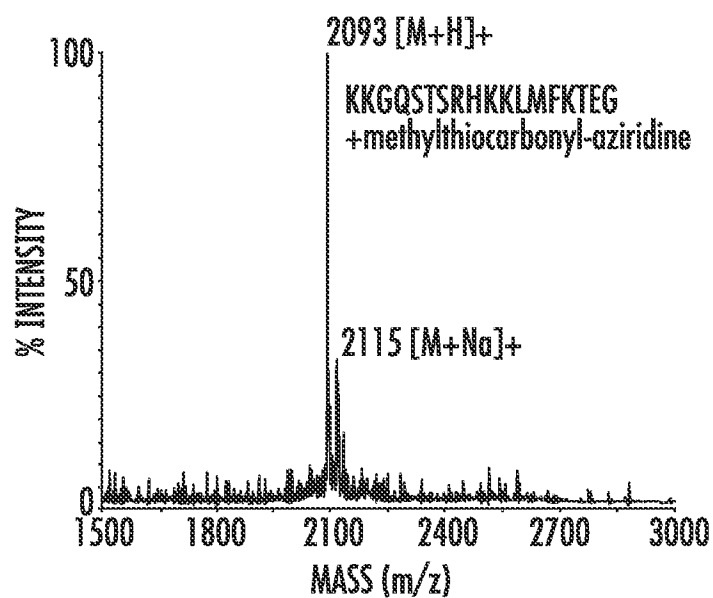


FIG. 4c

5/13

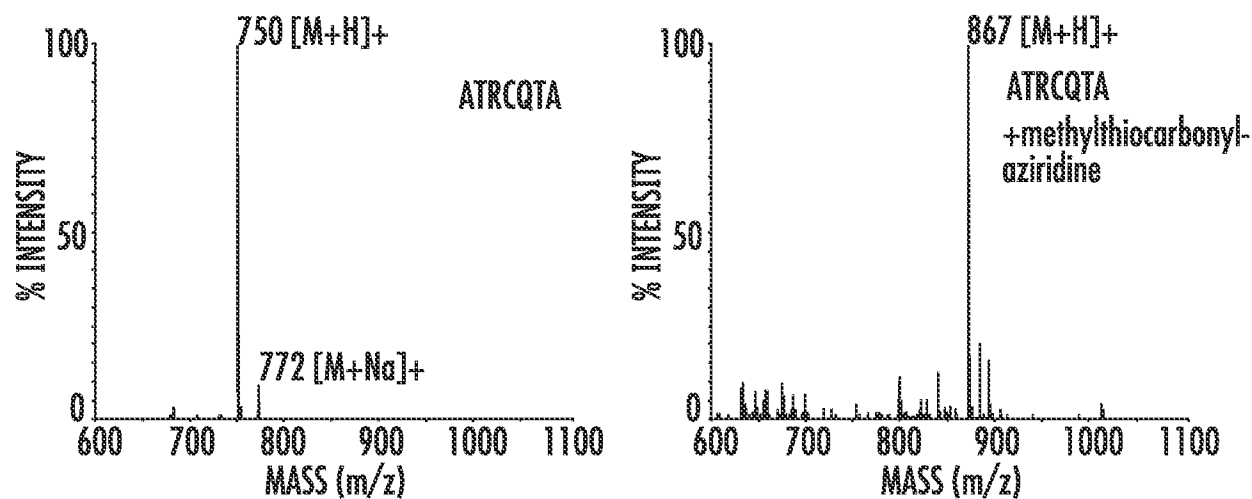


FIG. 5a

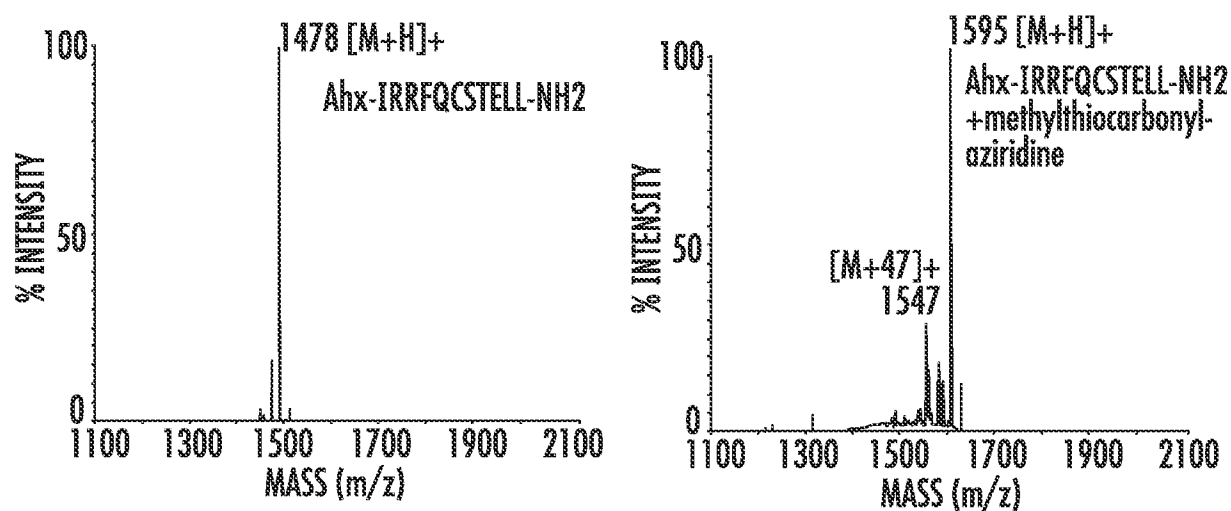


FIG. 5b

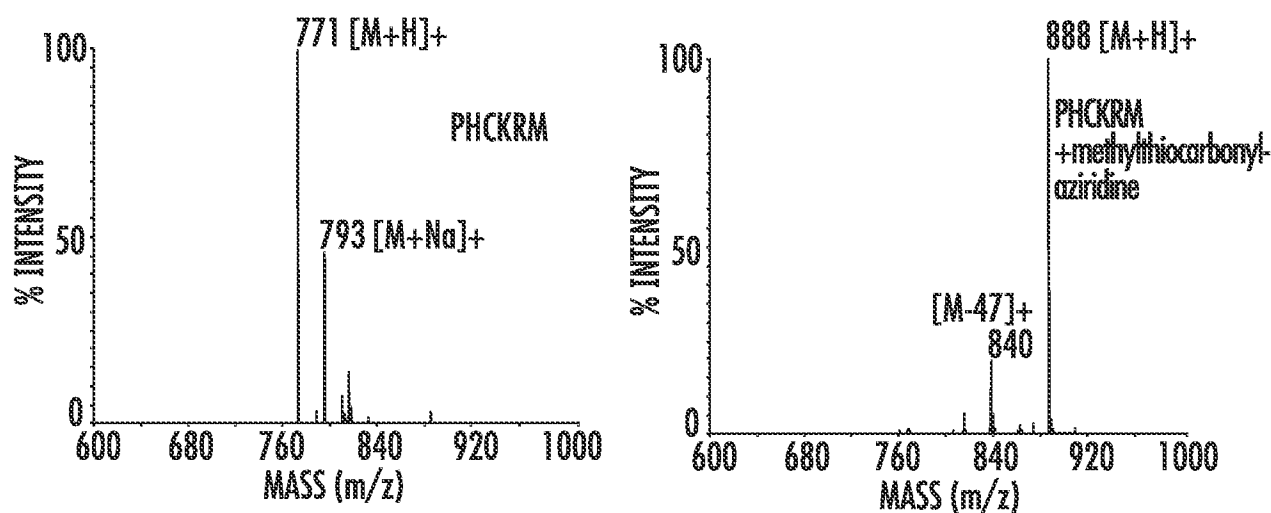


FIG. 5c

6/13

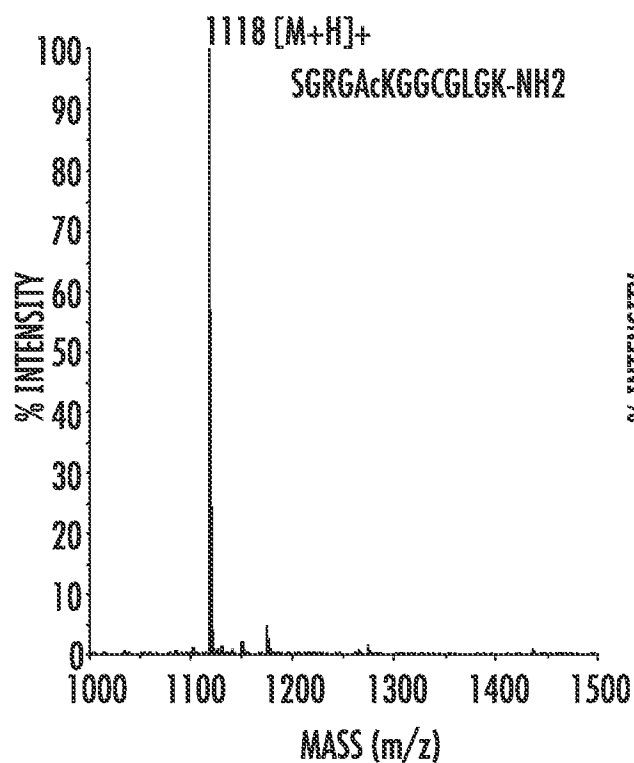


FIG. 6a

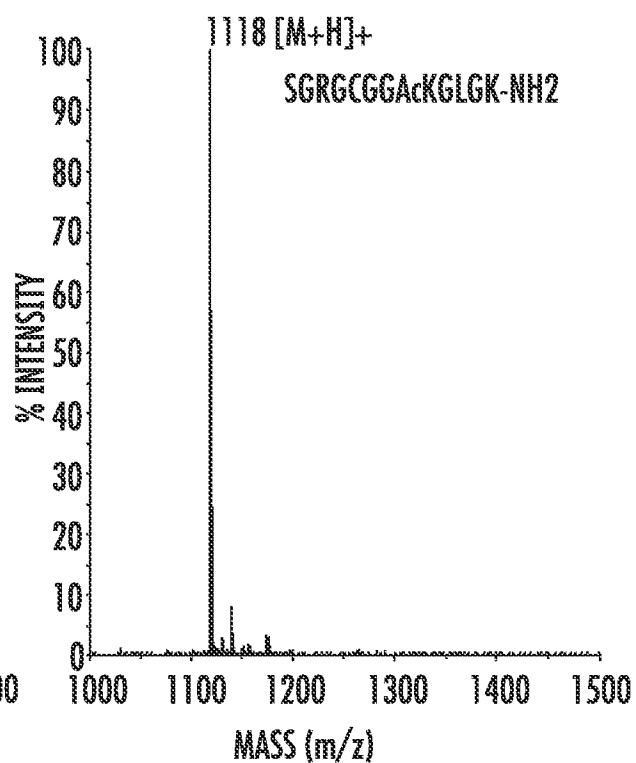


FIG. 6b

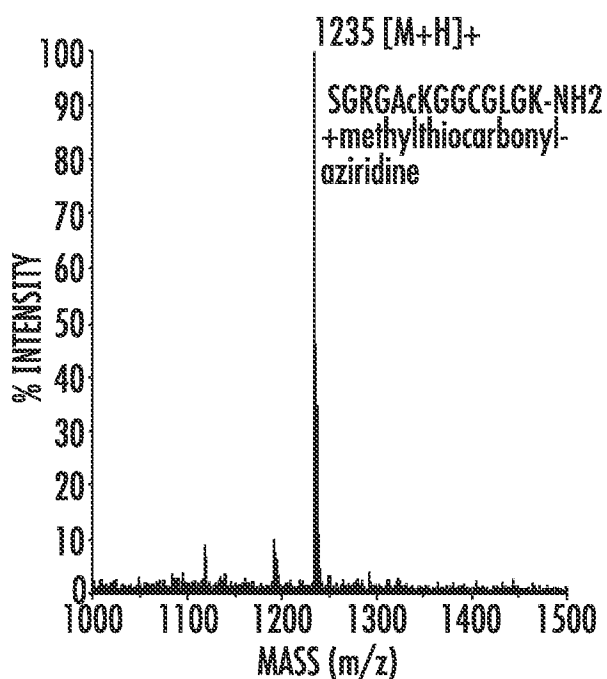


FIG. 6c

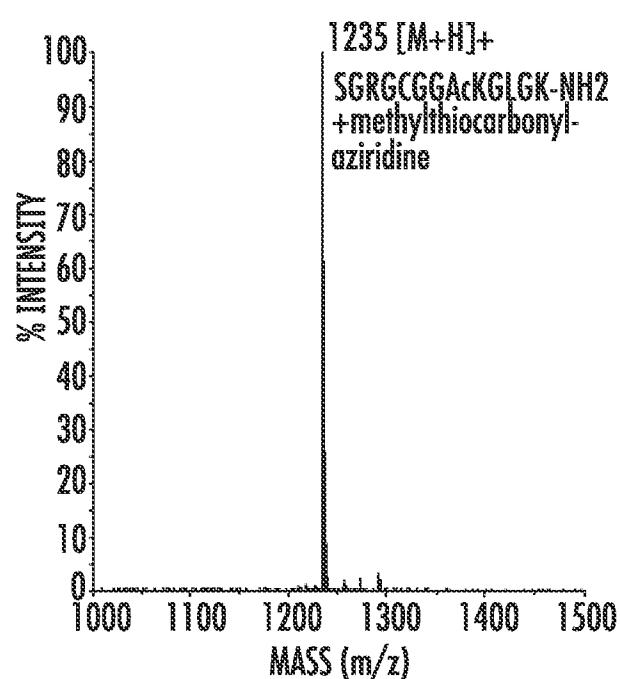


FIG. 6d

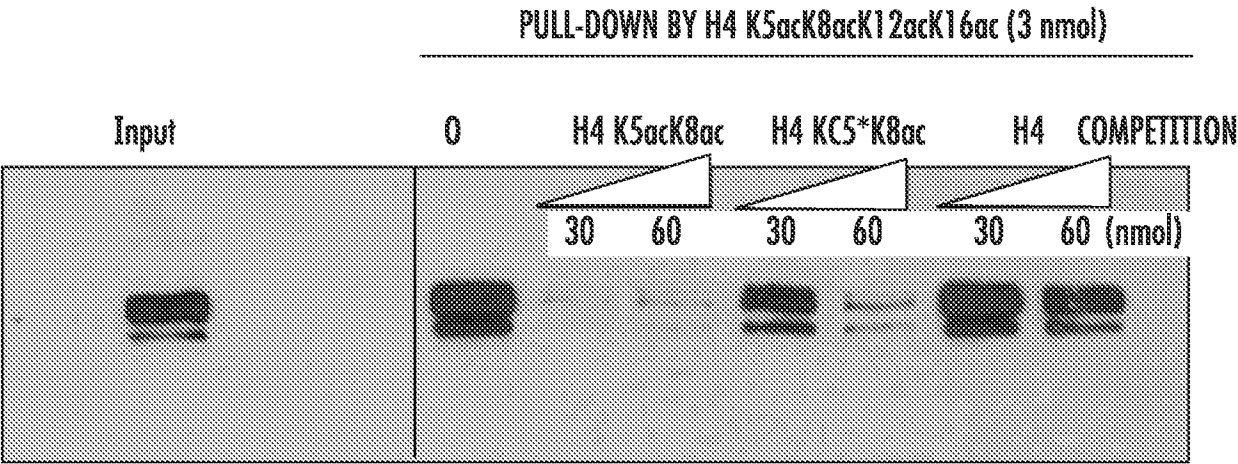


FIG. 7a

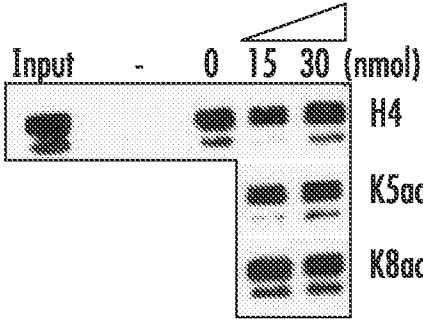


FIG. 7b

8/13

STANDARDS

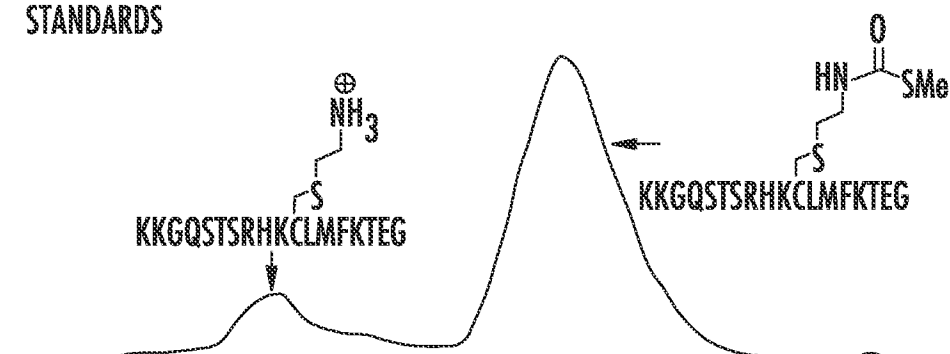


FIG. 8a

HDAC8 TREATMENT

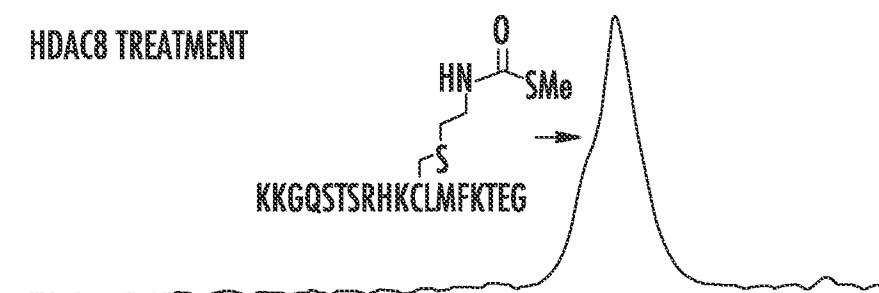


FIG. 8b

HDAC8 TREATMENT

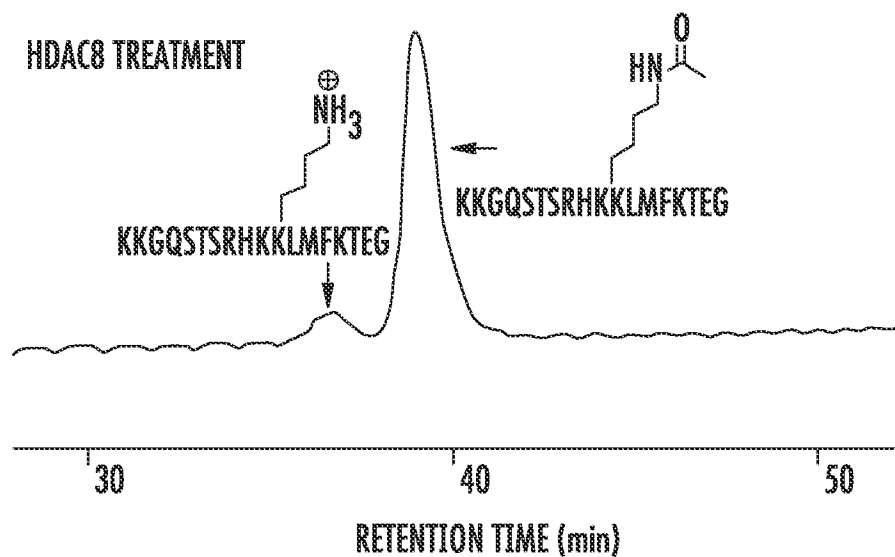


FIG. 8c

9/13

STANDARDS

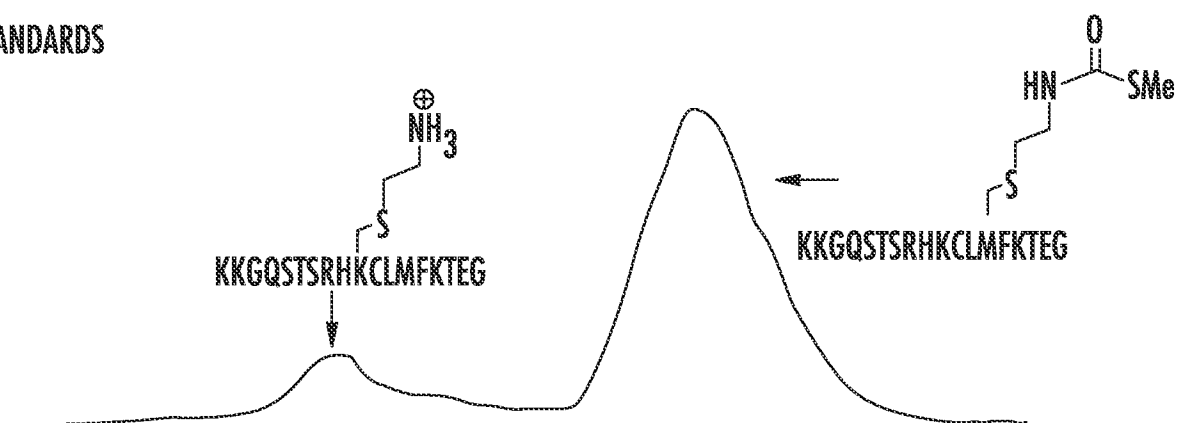


FIG. 9a

SIR2 TREATMENT

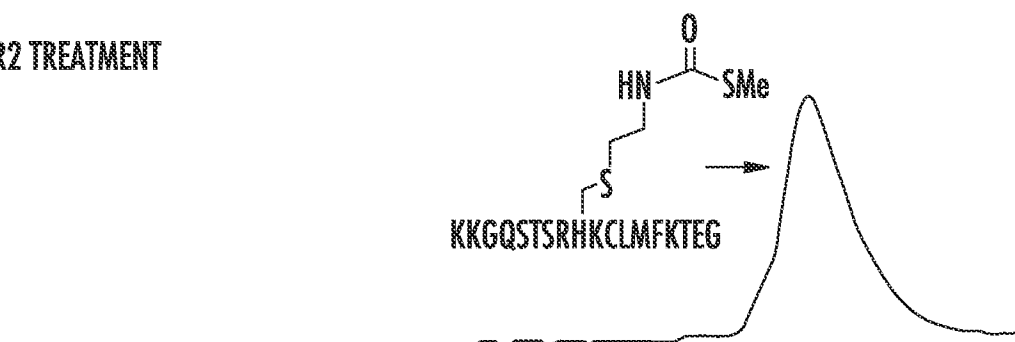


FIG. 9b

SIR2 TREATMENT

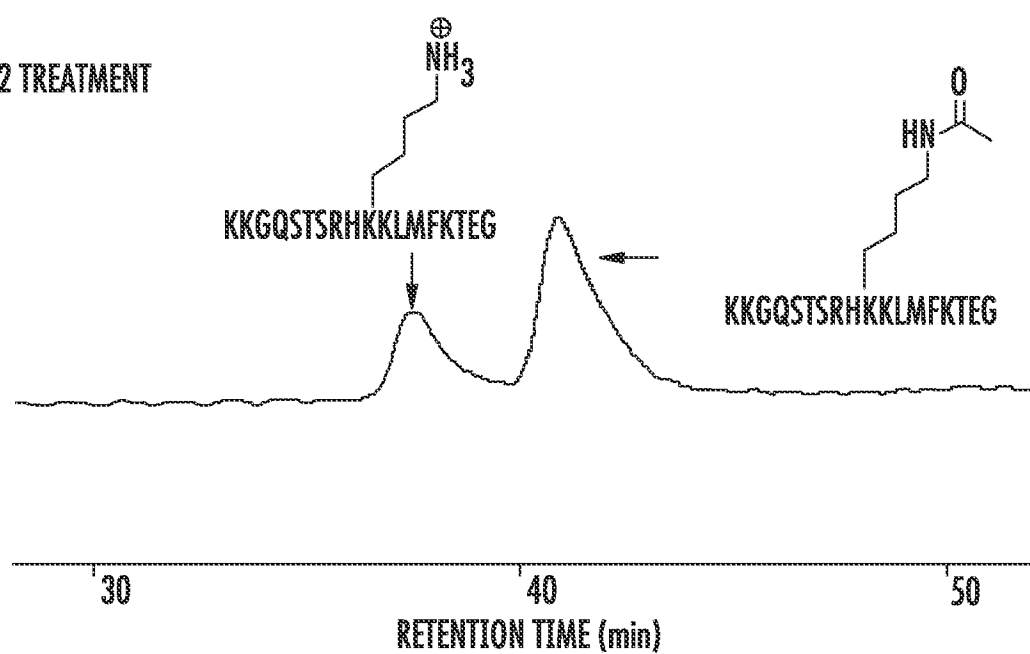


FIG. 9c

10/13

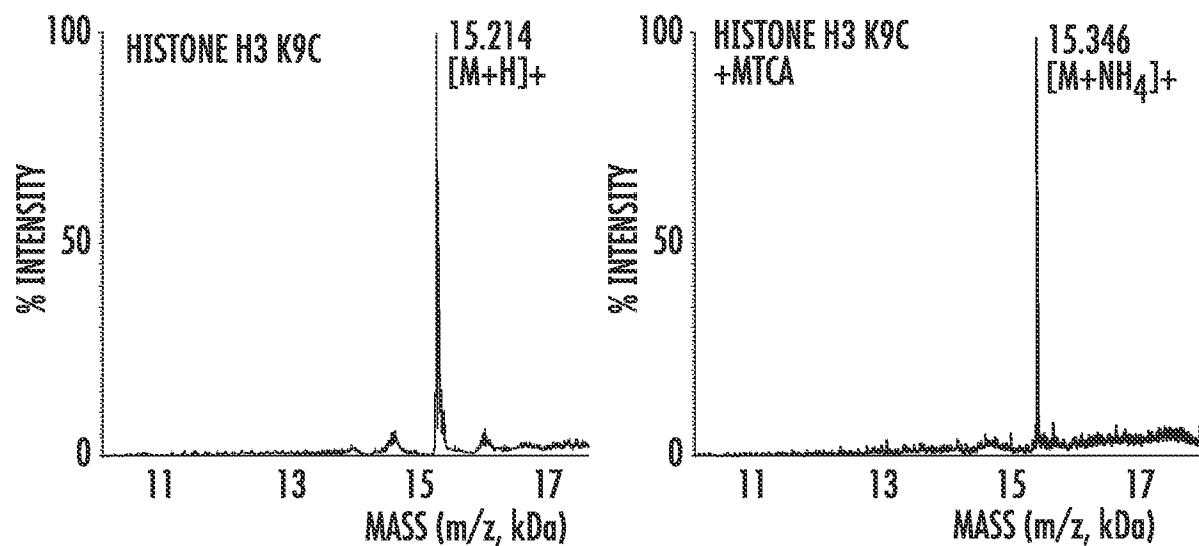


FIG. 10a

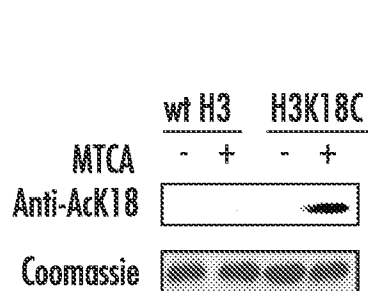


FIG. 10b

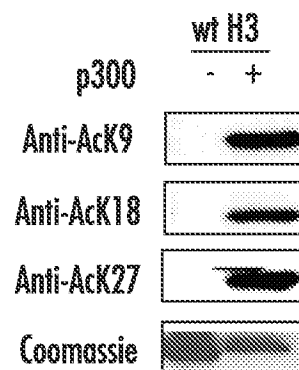
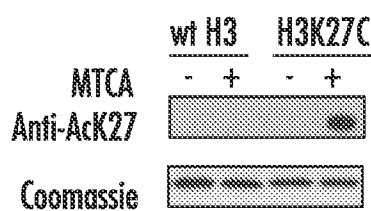


FIG. 10c

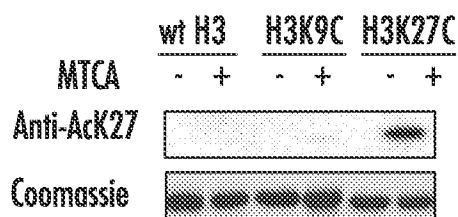
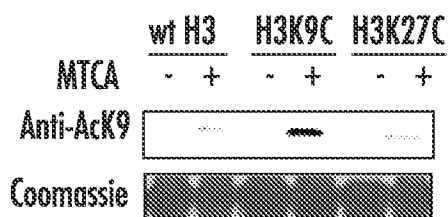
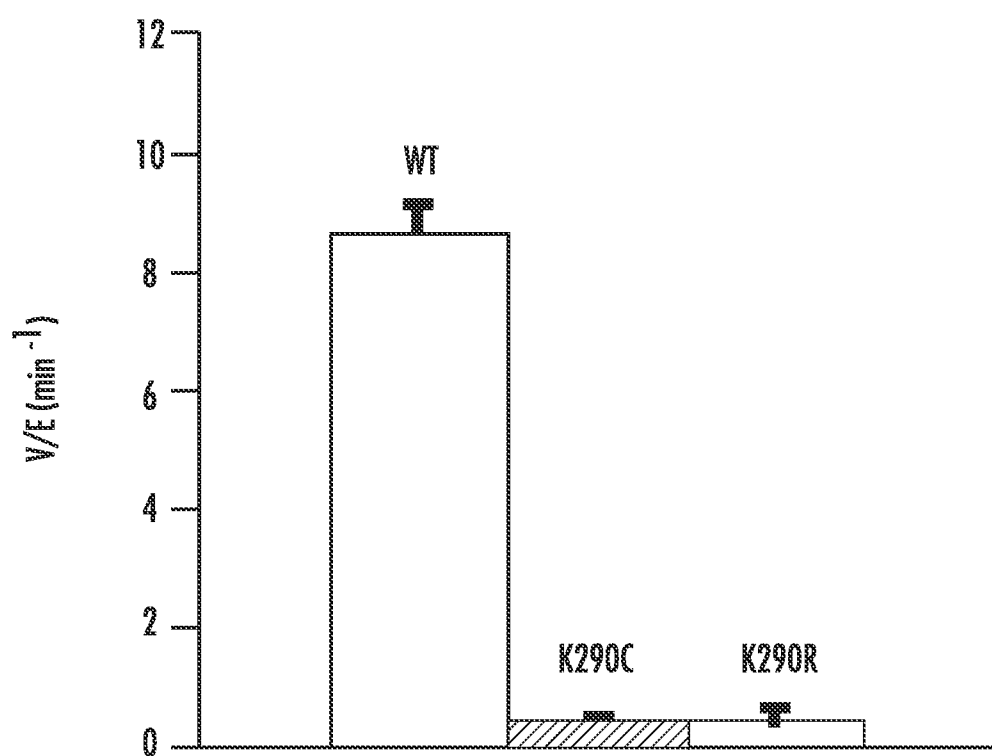


FIG. 10d

11/13

**FIG. 11**

12/13

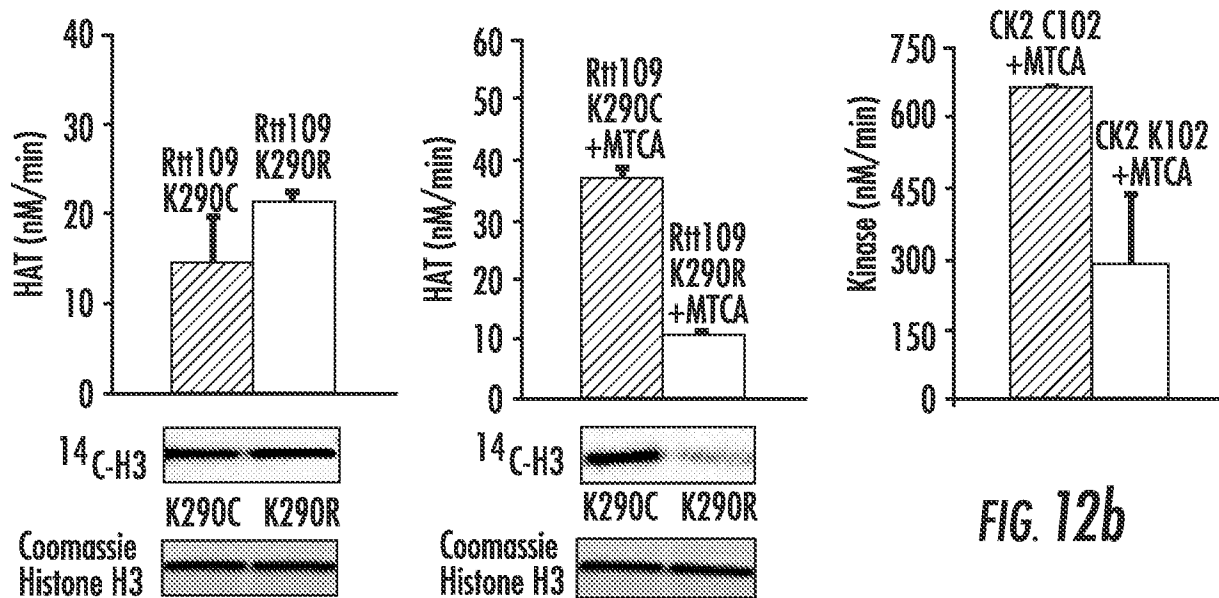
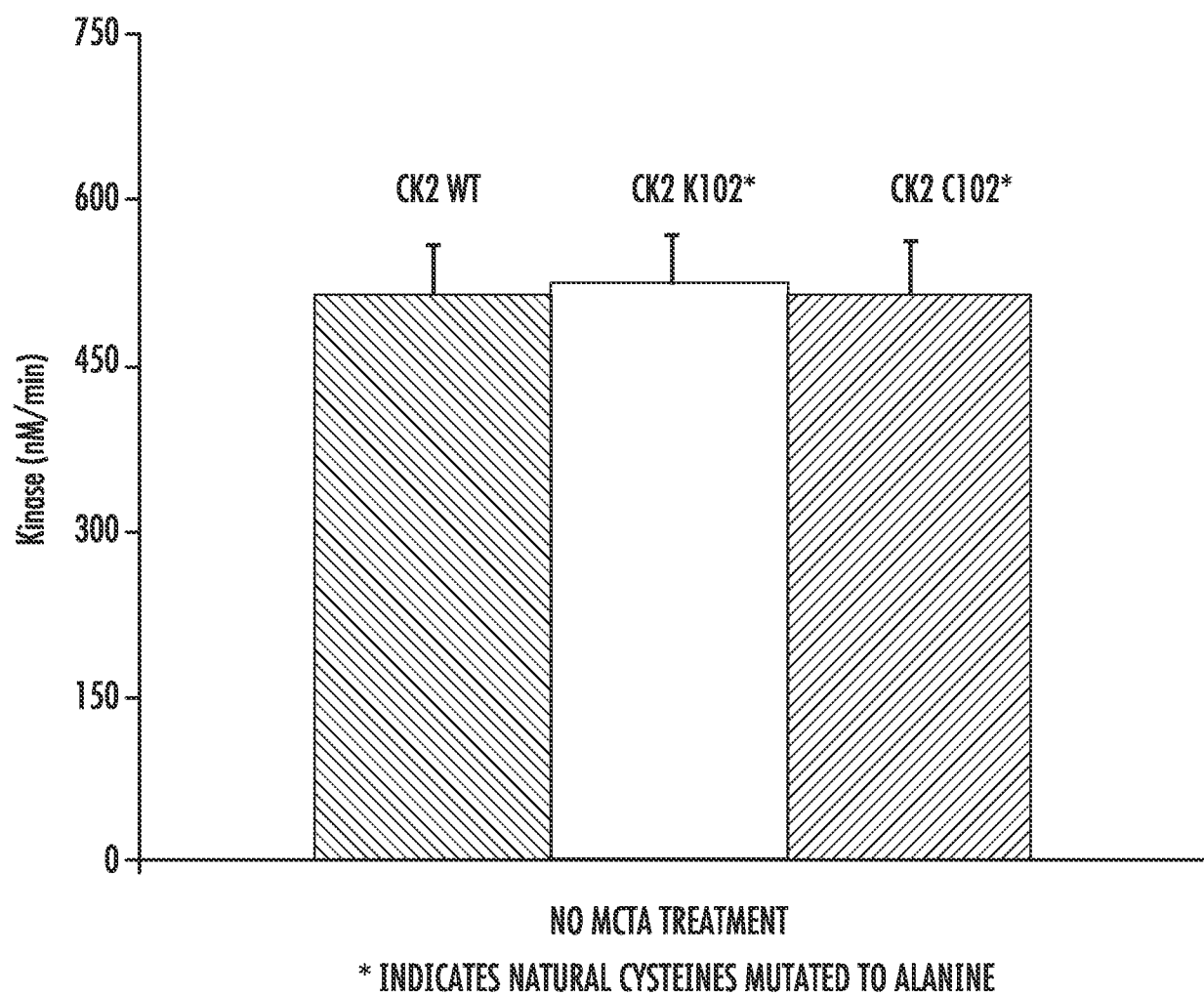


FIG. 12b

FIG. 12a

13/13

**FIG. 13**