Supplementary Material

A facile method to synthesize histones with posttranslational modification mimics

Zhiyong Wang, Yane-Shih Wang, Pei-Jing Pai, William K. Russell, David H. Russell and Wenshe R. Liu*

Department of Chemistry, Texas A&M University, College Station, TX 77843
E-mail: wliu@chem.tamu.edu

Table of Contents

1. General Experimental

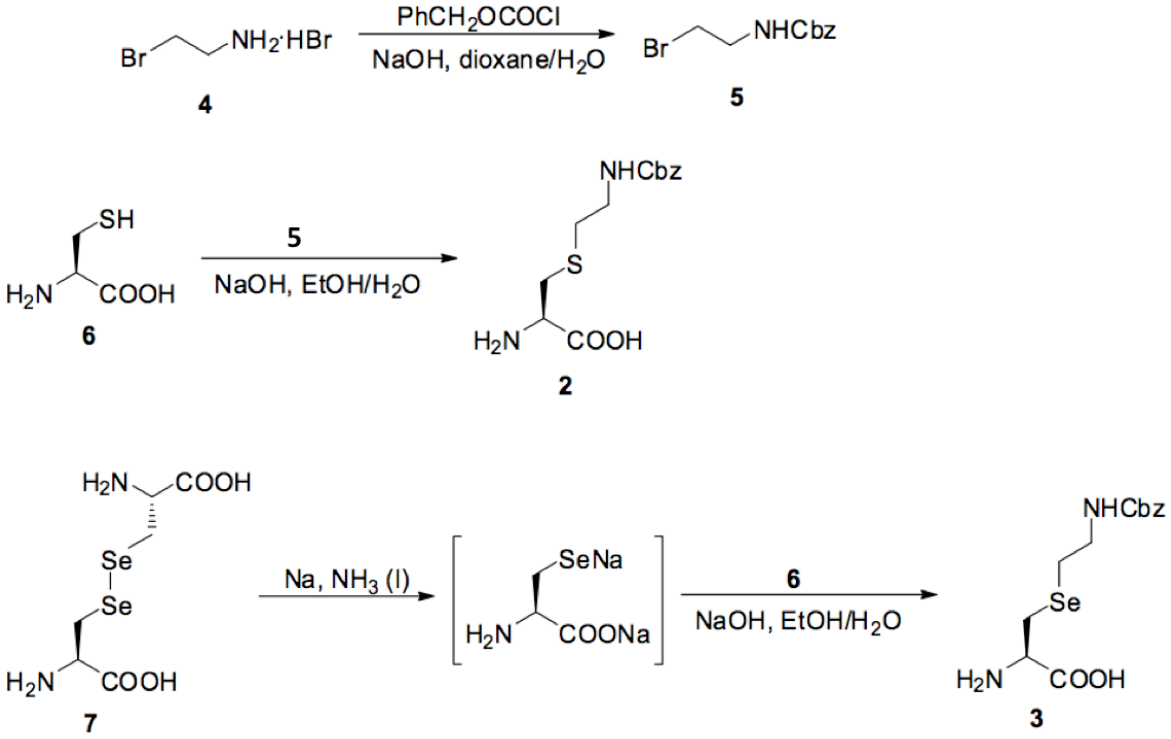
All reactions involving moisture sensitive reagents were conducted in oven-dried glassware under an argon atmosphere. Anhydrous solvents were obtained through standard laboratory protocols. Analytical thin-layer chromatography (TLC) was performed on Whatman SiO_2 60 F-254 plates. Visualization was accomplished by UV irradiation at 254 nm or by staining with ninhydrin (0.3% w/v in glacial acetic acid/n-butyl alcohol 3:97). Flash column chromatography was performed with flash silica gel (particle size 32-63 μm) from Dynamic Adsorbents Inc (Atlanta, GA).

Specific rotations of chiral compounds were obtained at the designated concentration and temperature on a Rudolph Research Analytical Autopol II polarimeter using a 0.5 dm cell. Proton and carbon NMR spectra were obtained on Varian 300 and 500 MHz NMR spectrometers. Chemical shifts are reported as δ values in parts per million (ppm) as referenced to the residual solvents: chloroform (7.27 ppm for ^{1}H and 77.23 ppm for ^{13}C) or water (4.80 ppm for ^{1}H). A minimal amount of 1,4-dioxane was added as the reference standard (67.19 ppm for ^{13}C) for carbon NMR spectra in deuterium oxide, and a minimal amount of sodium hydroxide pellet or concentrated hydrochloric acid was added to the NMR sample to aid in the solvation of amino acids which have low solubility in deuterium oxide under neutral conditions. ^{1}H NMR spectra are tabulated as follows: chemical shift, multiplicity (s = singlet, bs = broad singlet, d = doublet, t = triplet, q = quartet, m = multiplet), number of protons, and coupling constant(s). Mass spectra were obtained at the Laboratory for Biological Mass Spectrometry at the Department of Chemistry, Texas A&M University.

H-Lys(Z)-OH (**1**) was obtained from Chem-Impex International, Inc. (Wood Dale, IL). *O*-Mesitylsulfonylhydroxylamine (MSH), 2-(methylamino)ethanethiol hydrochloride (monomethyllysine mimic precursor), 2-(dimethylamino)ethanethiol hydrochloride (dimethyllysine mimic precursor), and 2-(mercaptoethyl)trimethyl-ammonium chloride (trimethyllysine mimic precursor) were prepared according to procedures by Bernardes and coworkers.[1] All other reagents were obtained from commercial suppliers and used as received.

2. Chemical Synthesis

Compounds **2** and **3** were synthesized by nucleophilic substitution of **5** with appropriate precursors (***Scheme 1***). Reductive cleavage of the diselenide bond in **7** with sodium borohydride[2] turned out to be incomplete, and was best effected with sodium in liquid ammonia.



Scheme 1. Synthesis of **2** and **3**.

Benzyl (2-bromoethyl)carbamate (5).[3,4] To a solution of **5** (13.5 g, 64.6 mmol) in a mixture of sodium hydroxide (2.0 *N*, 70 mL, 0.14 mol) and 1,4-dioxane (50 mL) cooled in an ice bath was added a solution of benzyl chloroformate (11.6 g, 64.6 mmol) in 1,4-dioxane (20 mL) dropwise over 15 min. The mixture was stirred at room temperature overnight, and most of the dioxane was evaporated. The residue was adjusted to pH 5 with hydrochloric acid (2 *N*) and extracted with ethyl acetate (100 mL x 2). The combined organic layers were washed with brine, dried (Na_2SO_4), evaporated, chromatographed (EtOAc/hexanes, 1:9), and crystallized in hexanes to give **5** (8.9 g, 53%) as a white solid. ^{1}H NMR ($CDCl_3$, 500 MHz) δ 7.40-7.33 (m, 5 H), 5.29 (bs, 1 H), 5.12 (s, 2 H), 3.60 (appar. q, 2 H, J = 6.0 Hz), 3.47 (t, 2 H, J = 5.7 Hz); ^{13}C NMR ($CDCl_3$, 75 MHz) δ 156.3, 136.4, 128.8, 128.4, 128.3, 67.2, 43.0, 32.7.

(*R*)-2-Amino-3-((2-(((benzyloxy)carbonyl)amino)ethyl)thio)propanoic acid (2). To a degassed solution of L-cysteine (**6**, 1.25 g, 10.0 mmol) in sodium hydroxide (2 *N*, 15.0 mL, 30.0 mmol) cooled in an ice bath was added **5** (2.85 g, 11.0 mmol) in degassed ethanol (10 mL) dropwise over 5 min. The mixture was stirred at room temperature overnight, and hydrochloric acid (3 *N*, 8.0 mL, 24.0 mmol) was added to give a white suspension. Filtration followed by washing with excessive water, ethanol, and dichloromethane and drying under vacuum afforded **2** (2.4 g, 81%) as a white solid. $[\alpha]_D^{21}$ +2.4 (*c* 1.20, 1 *N* NaOH); ^{1}H NMR (D_2O, 500 MHz, pH > 10) δ 7.45-7.39 (m, 5 H), 5.11 (s, 2 H), 3.38 (t, 1 H, J = 5.7 Hz), 3.36-

3.28 (m, 2 H), 2.84 (dd, 1 H, J = 13.5, 4.5 Hz), 2.76 (dd, 1 H, J = 13.2, 6.7 Hz), 2.68 (t, 2 H, J = 6.5 Hz); ^{13}C NMR (D_2O, 125 MHz, pH > 10) 181.6, 159.0, 137.1, 129.4, 128.3, 67.5, 55.8, 40.5, 37.5, 32.2; HRMS (ESI) calcd for $C_{13}H_{19}N_2O_4S$ ($[M+H]^+$) 299.1066, found 299.1074.

(*R*)-2-Amino-3-((2-(((benzyloxy)carbonyl)amino)ethyl)selanyl)propanoic acid (3). To an argon-protected solution of seleno-L-cystine (**7**, 3.75 g, 11.0 mmol) in liquid ammonia (~ 80 mL) cooled in a dry ice/acetone bath was added sodium metal (1.3 g, 56.5 mmol) in small pieces (*CAUTION!*) over 2 h, and a yellow suspension resulted in the end. The bath temperature was gradually raised up to room temperature and excessive ammonia was blown away with a gentle stream of argon *inside a well-ventilated fume hood*. Residual ammonia was removed on a rotorvap *inside a well-ventilated fume hood*, and the solid was cooled in an ice bath and carefully dissolved with degassed ice-cold water (25 mL). Compound **5** (5.85 g, 22.7 mmol) in degassed ethanol (15 mL) was added dropwise over 5 min, and the mixture was stirred at room temperature for 20 h and then filtered to give a red solution. Hydrochloric acid (3 *N*, 38.0 mL, 0.11 mol) was added to give a pink suspension, which was filtered, washed with plenty of water, ethanol, and dichloromethane, and dried under vacuum to afford **3** (6.6 g, 87%) as a slightly pink solid. $[\alpha]_D^{21}$ +14.8 (*c* 1.42, 1 *N* NaOH); ^{1}H NMR (D_2O, 500 MHz, pH < 1) δ 6.94-6.91 (m, 5 H), 4.62 (s, 2 H), 3.84 (m, 1 H), 2.91 (dt, 2 H, J = 6.7, 1.5 Hz), 2.67-2.65 (m, 1 H), 2.57 (m, 1 H), 2.28 (t, 2 H, J = 6.5 Hz); ^{13}C NMR (D_2O, 125 MHz, pH ~ 14) δ 181.4, 158.6, 136.8, 129.1, 128.7, 128.0, 66.9, 55.9, 41.0, 29.7, 24.1; HRMS (ESI) calcd for $C_{13}H_{19}N_2O_4Se$ ($[M+H]^+$) 349.0512/347.0510/345.0518 (major Se isotopes), found 349.0521/347.0501/345.0506.

3. DNA and Protein Sequences

3.1 DNA Sequences

GFP_{UV}:

atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtcagtgga
gagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacctgttccatggccaacacttgtca
ctactttctcttatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccgaaggttatgt
acaggaacgcactatatctttcaaagatgacgggaactacaagacgcgtgctgaagtcaagtttgaaggtgataccttgttaatcgtat
cgagttaaaaggtattgattttaaagaagatggaaacattctcggacacaaactcgaatacaactataactcacacaatgtatacatcacg
gcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacattgaagatggatccgttcaactagcagaccattatca
acaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacatagtctgccctttcgaaagatcccaacgaa
aagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgaactctacaaagagctccatcaccatc
accatcactaa

sfGFP:

atgtagaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaattttctgtccgtgga
gagggtgaaggtgatgctacaaacggaaaactcacccttaaatttatttgcactactggaaaactacctgttccgtggccaacacttgtc
actactctgacctatggtgttcaatgcttttcccgttatccggatcacatgaaacggcatgacttttcaagagtgccatgcccgaaggttat
gtacaggaacgcactatatctttcaaagatgacgggacctacaagacgcgtgctgaagtcaagtttgaaggtgataccttgttaatcgt
atcgagttaaagggtattgattttaaagaagatggaaacattcttggacacaaactcgagtacaactttaactcacacaatgtatacatcac
ggcagacaaacaaaagaatggaatcaaagctaacttcaaaattcgccacaacgttgaagatggttccgttcaactagcagaccattatc

aacaaaatactccaattggcgatggccctgtccttttaccagacaaccattacctgtcgacacaatctgtcctttcgaaagatcccaacga
aaagcgtgaccacatggtccttcttgagtttgtaactgctgctgggattacacatggcatggatgagctctacaaaggatcccatcaccat
caccatcactaa

pylT:

ggaaacctgatcatgtagatcgaatggactctaaatccgttcagccgggttagattcccggggtttccgcca

Methanosarcina mazei PylRS:

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacaccacg
aagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggactgcaagagc
gctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcacaaaggcaaacga
agaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatccgttgcgagagccccga
aacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgataccggtttccacccaagagtcagtttc
tgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactggtaaaagggaatacgaaccccattac
atccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactgacaggcttgaagtcctgttaaacccaaaagat
gagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctctctcgcagaaaaaaagacctgcagcagatctacgc
ggaagaaagggagaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacaggggttttctggaaataaaatccccga
tcctgatccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaagaacttctgc
ctgagacccatgcttgctccaaaccttacaactacctgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaataggc
ccatgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttctgccagatgggatcgggatgcaca
cgggaaaatcttgaaagcataattacggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctatggg
gataccttgatgtaatgcacggagacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataa
accctggataggggcaggtttcgggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgag
tcttactataacgggatttctaccaacctgtaa

mkRS1

atggataaaaaaccactaaacactctgatatctgcaaccgggctctggatgtccaggaccggaacaattcataaaataaaacaccacg
aagtctctcgaagcaaaatctatattgaaatggcatgcggagaccaccttgttgtaaacaactccaggagcagcaggactgcaagagc
gctcaggcaccacaaatacaggaagacctgcaaacgctgcagggtttcggatgaggatctcaataagttcctcacaaaggcaaacga
agaccagacaagcgtaaaagtcaaggtcgtttctgcccctaccagaacgaaaaaggcaatgccaaaatccgttgcgagagccccga
aacctcttgagaatacagaagcggcacaggctcaaccttctggatctaaattttcacctgcgataccggtttccacccaagagtcagtttc
tgtcccggcatctgtttcaacatcaatatcaagcatttctacaggagcaactgcatccgcactggtaaaagggaatacgaaccccattac
atccatgtctgcccctgttcaggcaagtgcccccgcacttacgaagagccagactgacaggcttgaagtcctgttaaacccaaaagat
gagatttccctgaattccggcaagcctttcagggagcttgagtccgaattgctctctcgcagaaaaaaagacctgcagcagatctacgc
ggaagaaagggagaattatctggggaaactcgagcgtgaaattaccaggttctttgtggacaggggttttctggaaataaaatccccga
tcctgatccctcttgagtatatcgaaaggatgggcattgataatgataccgaactttcaaaacagatcttcagggttgacaagaacttctgc
ctgagacccatgcttgctccaaaccttatgaactacgcgcgcaagcttgacagggccctgcctgatccaataaaaatttttgaaataggc
ccatgctacagaaaagagtccgacggcaaagaacacctcgaagagtttaccatgctgaacttcacgcagatgggatcgggatgcaca
cgggaaaatcttgaaagcataattaaggacttcctgaaccacctgggaattgatttcaagatcgtaggcgattcctgcatggtctttgggg
ataccttgatgtaatgcacggagacctggaactttcctctgcagtagtcggacccataccgcttgaccgggaatggggtattgataaac
cctggataggggcaggtttcgggctcgaacgccttctaaaggttaaacacgactttaaaaatatcaagagagctgcaaggtccgagtct
tactataacgggatttctaccaacctgtaa

Xenopus laevis H3/C110A

atggctcgtactaagcagaccgcccgttagtccaccggagggaaggctccccgcaaacagctggccaccaaggcagccaggaag

agcgctccggccacaggcggagtcaagaaacctcaccgttaccggcccggcacagtcgctctccgcgagatccgccgctaccaga
aatccaccgagctgctcatccgcaaactgcctttccagcgcctggtccgggagatcgctcaggacttcaagaccgacctgcgcttcca
gagctcggccgtcatggctctgcaggaggccagcgaggcttatctggtcggtttgtttgaggacaccaacctggccgccatccacgc
caagagggtcaccatcatgcccaaggacatccagctggcccgcaggatccggggcgagagggctgagctccatcaccatcaccat
cactaa

3.2 Proteins Sequences

GFP_{UV}:
MSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPW
PTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGNYKTRAEVKF
EGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYITADKQKNGIKANFKIRHNIE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTXSALSKDPNEKRDHMVLLEFVTAAGI
THGMDELYKELHHHHHH

X represents a noncanonical amino acid.

sfGFP:
MXKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFI**C**TTGKLPVPW
PTLVTTLTYGVQ**C**FSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKF
EGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYITADKQKNGIKANFKIRHNVE
DGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAAGI
THGMDELYKGSHHHHHH

X represents a noncanonical amino acid.

Methanosarcina mazei PylRS:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTA
RALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSV
ARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKG
NTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKD
LQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRV
DKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFCQ
MGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGDTLDVMHGDLELSSAVVGPIPL
DREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

mkRS1:

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSSRTA
RALRHHKYRKTCKRCRVSDEDLNKFLTKANEDQTSVKVKVVSAPTRTKKAMPKSV
ARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSTSISSISTGATASALVKG
NTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKKD
LQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSKQIFRV
DKNFCLRPMLAPNLMNYARKLDRALPDPIKIFEIGPCYRKESDGKEHLEEFTMLNFT
QMGSGCTRENLESIIKDFLNHLGIDFKIVGDSCMVFGDTLDVMHGDLELSSAVVGPIP
LDREWGIDKPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGISTNL

Xenopus laevis H3/C110A

MARTKQTARXSTGGKAPRKQLATKAARKSAPATGGVKKPHRYRPGTVALREIRRY

QKSTELLIRKLPFQRLVREIAQDFKTDLRFQSSAV**M**ALQEASEAYLVGLFEDTNLAAI
HAKRVTI**M**PKDIQLARRIRGERAELHHHHHH

4. Construction of plasmids

4.1 Construction of pET-pylT-GFP

The plasmid pET-pylT-GFP was derived from the plasmid pAcKRS-pylT-GFP1Amber[5] in which GFP_{UV} has an amber mutation at Q204. The restriction enzyme *BglII* was used to cut off the ACKRS gene. The digested pAcKRS-pylT-GFP1Amber plasmid was ligated to form pET-pylT-GFP.

4.2 Construction of pET-sfGFP2TAG

The plasmid pET-sfGFP2TAG was derived from the plasmid pET-pylT-Z[6] in which sfGFP has an amber mutation at S2. This gene was amplified from the Superfolder GFP (sfGFP) plasmid (Theranostech®). Two restriction sites, *NdeI* at the 5' end and *SacI* at the 3' end, were introduced in the PCR product which was subsequently digested and used to replace Z domain gene in pET-sfGFP2TAG.

4.3 Construction of pBK-mKRS1 with the Y384F mutation

The pBK-mKRS1 plasmid in which mKRS1 has the Y384 mutation was derived from a pBK plasmid containing *p*-iodo-L-phenylalanyl-tRNA synthetase that was initially evolved from *M. jannaschii* tyrosyl-tRNA synthetase.[7] The PylRS gene is under the control of *E. coli glnS* promoter and terminator. This gene was amplified from the pBK-mKRS1 (Y306M/L309A/C348T/T364K) plasmid by flanking primers, pBK-mmPylRS-NdeI-F and pBK-mmPylRS-PstI~NsiI-R. To construct the pBK-mkRS1F plasmid Y384F mutation was introduced by overlap extension PCR. The following pairs of primers were used to generate an mKRS1 gene with the Y384F mutation: (1) pBK-mmPylRS-NdeI-F (5'-gaatcccatatggataaaaaaccactaaacactctg-3') and mmPylRS-Y384F-R (5'-tacatcaagggtatcccc<u>aaa</u>gaccatgcaggaatcgcctacg-3'); (2) mmPylRS-Y384F-F (5'-cgtaggcgattcctgcatggtc<u>ttt</u>ggggataccttgatgta-3') and pBK-mmPylRS-PstI~NsiI-R (5'-gtttgaaaatgcatttacaggttggtagaaatccc-3'). The amplified gene was digested with the restriction enzymes *NdeI* and *NsiI*, gel-purified, and ligated back into the pBK vector digested by *NdeI* and *PstI* to afford plasmid pBK-mKRS1.

4.4 Construction of pEVOL-pylT

The pEVOL-pylT plasmid was derived from a pEVOL plasmid.[8] The pylT gene with *prok* promoter and terminator was synthesized by overlap PCR with eight primers ((1) pEVOL-PylT-ApaLI-F-1 (5'-gatatgatcagtgcacggctaactaagcggcctgctgactttctcg-3'); (2) pEVOL-PylT-R-2 (5'-caatcccttaatagcaaaatgccttttgatcggcgagaaagtcagcag-3'); (3) pEVOL-pylT-F-3 (5'-gctattaagggattgacgagggcgtatctgcgcagtaagatgcgcccc-3'); (4) pEVOL-pylT-R-4 (5'-agtccattcgatctacatgatcaggtttccaatgcggggcgcatcttac-3'); (5) pEVOL-pylT-F-5 (5'-gtagatcgaatggactctaaatccgttcagccgggttagattcccgggg-3'); (6) pEVOL-pylT-R-6 (5'-ggcttttcgaatttggcggaaaccccgggaatctaac-3'); (7) pEVOL-pylT-F-7 (5'-caaattcgaaaagcctgctcaacgagcaggctttttg-3'); (8) pEVOL-pylT-Xho1-R-8 (5'-

ctgagctgctcgagcatgcaaaaaagcctgctc-3') and introduce by two restriction sites, ApaLI at the 5' end and XhoI at the 3' end. Two pairs of restriction sites (SpeI and SalI sites between the *pBAD* promoter and terminator; NdeI and NotI between the *glnS* promoter and terminator) were introduced using two pairs of primers ((1) pEVOL-SpeI-R (5'-ttactagtaattcctcctgttagccc-3') and pEVOL-SalI-F (5'-ccgtcgaccatcatcatcatc-3'); (2) 5'-pEVOL-NdeI-R (5'-atcatatgggattcctcaaagcgtaaac-3') and pEVOL-NotI-F (5'-acgcggccgctttcaaacgctaaattgc-3'). These restriction sites in pEVOL-pylT were constructed for further installations of two copies of mKRS1.

4.5 Construction of pEVOL-mKRS1-pylT

The pEVOL-mKRS1-pylT plasmid was derived from the pET-sfGFP2TAG plasmid with sequential insertion of two copies of the mKRS1 gene. The first copy of mKRS1 gene was amplified from the pBK-mKRS1 plasmid by flanking primers, pEVOL-PylRS-SpeI-F and pEVOL-PylRS-SalI-R, digested by SpeI and SalI restriction enzymes, and ligated to a precut pEVOL-pylT plasmid. The resulted plasmid was digested by NdeI and NotI enzymes and used to insert the second copy of the mKRS1 gene that was amplified using primers pEVOL-PylRS-NdeI-F and pEVOL-PylRS-NotI-R and digested by NdeI and NotI restriction enzymes. The resulted plasmid is pEVOL-mKRS1-pylT.

4.6 Construction of pET-H3K9TAG

The plasmid pET-H3K9TAG was derived from the plasmid pET-sfGFP2TAG in which H3 has an amber mutation at K9. This gene was amplified from the plasmid pET22b-xlH3.[9] Two restriction sites, *NdeI* at the 5' end and *SacI* at the 3' end, were introduced in the PCR product which was subsequently digested and used to replace sfGFP gene in pET-H3K9TAG.

4.7 Construction of pET-H3

The plasmid pET-H3 with wild type H3 gene was derived from the commercial plasmid pET-Duet1. This gene was amplified from the plasmid pET22b-xlH3/C110A. Two restriction sites, *NdeI* at the 5' end and *KpnI* at the 3' end, were introduced in the PCR product which was subsequently digested and used to install H3 gene in pET-Duet1 to afford pET-H3.

5. Screening Procedures for mKRS and mKRS1 Strains

The initial mKRS1 strain (designated as mKRS) and the mKRS1 strain with the Y384F mutation (designated as mKRS1) in pBK plasmid were cotransformed with pY+ to *E. Coli* TOP10 cells and tested for their ability to grow on plates with 102 μg/mL chloramphenicol (Cm), 25 μg/mL kanamycin (Kan), 12 μg/mL tetracycline (Tet), and 1 mM of **1**, **2**, or **3**. A plate without any noncanonical amino acid (NAA) supplement was used as the negative control. *E. Coli* cells containing the two plasmids were separately placed on LB agar plates with decreasing cell number by serial dilution from 3×10^6 to 1. The original cell solution was prepared with OD_{600} = 1.0 (~1×10^9 cells/mL). The cells on the plate were grown at 37 ^oC for 48 h. Images of colonies growing on different plates were shown in **Supplementary Figure 1**.

6. Protein Expression and Purification

6.1 GFP_{UV} and sfGFP Expression and Purification

To express GFP_{UV} or sfGFP incorporated with a NAA, *E. Coli* BL21(DE3) cells were cotransformed with pEVOL-mKRS1-pylT and pET-pylT-GFP or pET-sfGFP2TAG. Cells were recovered in 1 mL of LB medium for 1 h at 37 ºC before being plated on LB agar plate containing chloramphenicol (Cm) (34 μg/mL) and ampicillin (Amp) (100 μg/mL). A single colony was then selected and grown overnight in a 10 mL culture. This overnight culture was used to inoculate 500 mL of LB medium supplemented with 34 μg/mL Cm and 100 μg/mL Amp. Cells were grown at 37 ºC in an incubator (300 r.p.m.) and protein expression was induced when OD_{600} reached 0.7 by adding 1 mM IPTG, 0.2% arabinose and 2 mM **1**. After 6 h induction, cells were harvested, resuspended in a lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0), and sonicated. The cell lysate was clarified by centrifugation (60 min, 11,000 r.p.m., 4 ºC). The supernatant clear solution was incubated with 3 mL Ni^{2+}-NTA resin (Qiagen) (2 h, 4 ºC) and washed with 30 mL of the lysis buffer. Protein was finally eluted out by the lysis buffer with 250 mM imidazole. Eluted fractions were collected and concentrated. The collected fractions were further purified on an anion exchange column. The buffer was later changed to 10 mM ammonium bicarbonate using an Amicon Ultra-15 Centrifugal Filter Devices (10,000 MWCO cut, Millopore). The purified proteins were analyzed by 12% SDS-PAGE. GFP_{UV} or sfGFP proteins incorporated with **2** and **3** were expressed and purified similarly. The stock solution of 100 mM **3** was freshly prepared in 200 mM aqueous NaOH containing 20% dimethyl sulfoxide (DMSO) on ice-cold water bath.

6.2 H3 Expression and Purification

To express H3 incorporated with a NAA, *E. Coli* BL21(DE3) cells were cotransformed with pEVOL-mKRS1-pylT and pET-H3K9TAG. Cells were recovered in 1 mL of LB medium for 1 h at 37 ºC before being plated on LB agar plate containing Cm (34 μg/mL) and Amp (100 μg/mL). A single colony was then selected and grown overnight in a 10 mL culture. This overnight culture was then used to inoculate 500 mL of 2YT media supplemented with 34 μg/mL Cm and 100 μg/mL Amp. Cells were grown at 37 ºC in an incubator (300 r.p.m.) and protein expression was induced when OD_{600} reached 0.7 - 1.0 by adding 1 mM IPTG, 0.2% arabinose and 2 mM **1**. After 6-8 h induction, cells were harvested, resuspended in a lysis buffer (50 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, 0.1% NaN_3, 0.5% Triton-X100, 0.1 mM PMSF and 1mM DTT, pH 8.0) and sonicated. $MgSO_4$ (final concentration 10 mM) was added to chelate EDTA, and 0.01 mg/ml DNase and 0.1 mg/ml lysozyme were then added to the solution. The mixture was incubated at room temperature for 20 min. The cell lysate was clarified by centrifugation (20 min, 6,000 r.p.m., 4 ºC). The centrifuged pellet was crushed with a spatula, then resuspended by sonication in the lysing buffer. Another portion of DNase and lysozyme was added at this point to improve the purity of the pellet. After centrifugation, the inclusion body was washed twice with the washing buffer (50 mM Tris-HCl, 100 mM NaCl, 5 mM EDTA, 0.1% NaN_3, pH 8.0).

The semi-purified H3 protein inclusion body was dissolved by a dissolving buffer

(100 mM NaH_2PO_4, 10 mM Tris HCl, 8 M urea, pH 8.0) and incubated for 1 h at 37 ^oC. The solution was then centrifuged (20 min, 10,000 r.p.m, 4 ºC). The supernatant was incubated with 3 mL Ni^{2+}-NTA resin (Qiagen) (2 h, 4 ºC) and washed with 30 mL of washing buffer (100 mM NaH_2PO_4, 10 mM Tris HCl, 8 M urea, pH 6.2). The protein was finally eluted out by an elution buffer (100 mM NaH_2PO_4, 10 mM Tris HCl, 8 M urea, pH = 4.5). The pure eluted fractions were collected and concentrated. The buffer was later changed to 10 mM ammonium bicarbonate and 8 M urea using an Amicon Ultra -15 Centrifugal Filter Devices (10,000 MWCO cut, Millopore). The purified proteins were analyzed by 15% SDS-PAGE. H3 proteins incorporated with **2** and **3** were expressed in the presence of 2 mM **2** and 1.5 mM **3**, respectively, and similarly purified.

7. Synthesis of H3 Mimics with Posttranslational Modifications

H3K9Dha was synthesized from H3K9-**3**. Aqueous H_2O_2 solution (100 mM, 5 μL, 100 eq, 500 nmol) was added to the protein H3K9-**3** (1.6 mg/mL, 100 μM, 50 μL, 5 nmole) in a dissolving buffer (100 mM NaH_2PO_4, 10 mM Tris HCl, 8 M urea, pH 8.0), and the mixture was periodically agitated at room temperature for 1 h. The mixture was then dialyzed by Amicon Ultra -15 Centrifugal Filter Devices (10,000 MWCO cut, Millopore) against the dissolving buffer to terminated the reaction.

To synthesize H3 mimics H3K9AcsK, H3K9mesK, H3K9m^{2}sK, H3K9m^{3}sK and H3K9pC, solutions of the corresponding thiol nucleophiles (400 mM, 12.5 μL, 5 ・mole) in the dissolving buffer (100 mM NaH_2PO_4, 10 mM Tris HCl, 8 M urea, pH 8.0) were add into protein samples in the same buffer. The Michael addition reaction was performed at room temperature for 1 h and then terminated by dialysis against the dissolving buffer.

8. Immunoprecipitation of wt-H3, H3K9mesK, H3K9m^{2}sK and H3K9m^{3}sK by HP1 ・ and Western Blotting Assay of H3K9AcsK

HP1β (1 μM) was incubated with wt-H3, H3K9mesK, H3K9m^{2}sK and H3K9m^{3}sK, in 500 μl of binding buffer (0.5 M NaCl, 1% NP40, 0.5% sodium deoxycholate, 0.1% SDS, 50 mM Tris HCl, pH 8.0). A portion of this sample (20 μl) was removed to check the total protein level (input). The remaining supernatant was incubated for 4 h at 4 ºC with 1 μg of a goat polyclonal antibody to CBX1/HP1 beta (Abcam, ab40828). After 1 h of incubation, 30 μl of protein A-agarose (Sigma) was added. The beads were pelleted, washed 5 times with 700 μl of RIPA buffer, and the bound protein was released by boiling in SDS-sample buffer. A Rabbit polyclonal antibody to C-terminus of H3 (9715, Cell Signaling Technology) was used to detect H3 proteins immunoprecipitated by HP1β.

A Rabbit polyclonal antibody for histone H3 (acetyl K9) (Abcam, ab10812) was used to detect wtH3 or H3K9-AcsK. Further ECL test was carried out by treatment with a donkey polyclonal secondary antibody to rabbit IgG-H&L (HRP) (Abcam, ab16284).

9. ESI-MS Analysis of Intact Proteins

Protein samples were prepared by desalting the protein using C18 ZipTip™ (Millipore) following the manufacturer's protocol and eluted with 60% acetonitrile containing 0.1% formic acid. The resulting solution was diluted to 1 μM with 50% methanol containing 0.1% formic acid, and then used for electrospray mass spectrometry (ESI-MS) analysis. ESI ion-mobility (IM) MS experiments were performed on a SYNAPT G2 HDMS mass spectrometer (Waters Corp., Milford, MA) equipped with a nano-ESI source. The IM-MS data were acquired in positive ion mode (400-2500 Da) using spray voltage of +1800 V. Data analysis and protein signal extraction were performed using the MassLynx™ and DriftScope™ software packages (Waters Corp., Milford, MA). For the histone H3 series, a mass range of m/z 500-1200 was used for spectral deconvolution and the output range was 15000 to 19000 Da using a resolution of 0.1 Da per channel. For the green fluorescent proteins, a mass range of m/z 700-1300 was used for spectral deconvolution and the output range was 27000 to 29000 Da using a resolution of 0.1 Da per channel. For top-down analysis of H3K9-AcsK by CID-IM-MS, the precursor ion of m/z 655.54, corresponding to the $^{25+}$ charged state of H3K9-AcsK, was selected to perform collision-induced dissociation (CID) experiment using argon as collision gas and 25 V of collision energy in the Trap region. High-resolution MS experiments were performed on a SolariX 9.4 T: hybrid quadrupole-FTICR mass spectrometer (Burker Daltonik GmbH, Bremen, Germany) equipped with a nano-ESI source and acquired in positive ion mode (m/z 300-3000) using electrospray voltage of +1600 V. All MS spectra were obtained by quadrupole mass selection of m/z 800 to 1000 and accumulation of 100 spectra. Data analysis and protein signal extraction were performed using the DataAnalysis™ software packages (Burker Daltonik GmbH, Bremen, Germany). For spectral deconvolution to a singly charged spectrum, the output range was 5000 to 100000 m/z using an abundance cutoff of 0 %.

10. Tandem Mass Spectrometry analysis

GFP-**4** from the SDS-PAGE gel was excised and digested with endoproteinase Asp-N (Roche Diagnostics Co., Indianapolis, IN) or trypsin (Promega, Madison, WI) at 37 ^{0}C overnight using the following protocol: the gel slice was washed with 25 mM ammonium bicarbonate (ABC, pH 8) and dehydrated with a solution mixture of acetonitrile (ACN) and 50mM ABC (v/v, 2/1). The washing and dehydrating steps were repeated for another two times. Supernatant was removed and the gel slice was dried in a vacuum centrifuge (SpeedVac Concentrator, Savant, Farmingdale, NY). 10 μL of 20 ng/μL Asp-N or trypsin in 25 mM ABC was added to the dried gel slice. After the gel slice was completely rehydrated, 20 μL of 25mM ABC was added to cover gel slice and incubated at 37 °C overnight. Peptides resulting from the Asp-N or trypsin digestion were mixed 1:1 (v/v) with matrix (5 mg/mL α -cyano-4-hydroxycinnamic acid, 50% (v/v) acetonitrile, 10 mM ammonium dihydrogen phosphate, 0.1% TFA) and 1 μL of the resulting mixture was spotted onto a stainless steel target plate. Mass spectra and tandem MS spectra were collected using an Applied Biosystems 4800 TOF/TOF™ Analyzer (Framingham, MA). Collision induced dissociation tandem MS spectra were acquired using air at the medium pressure setting and at 2 kV of collision energy. Tandem MS data was manually interpreted using the Data Explorer™ software package (Applied Biosystems, Framingham, MA).

11. References:

[1] G. J. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, *J. Am. Chem. Soc.* **2008**, *130*, 5052.

[2] S. J. Foster, R. J. Kraus, H. E. Ganther, *J. Labelled Comp. Rad.* **1985**, *22*, 301.

[3] A. J. Brouwer, R. M. J. Liskamp, *Eur. J. Org. Chem.* **2005**, 487.

[4] F. E. Michael, P. A. Sibbald, B. M. Cochran, *Org. Lett.* **2008**, *10*, 793.

[5] Y. Huang, W. K. Russell, W. Wan, P. J. Pai, D. H. Russell, W. Liu, *Mol. Biosyst.* **2010**, *6*, 683

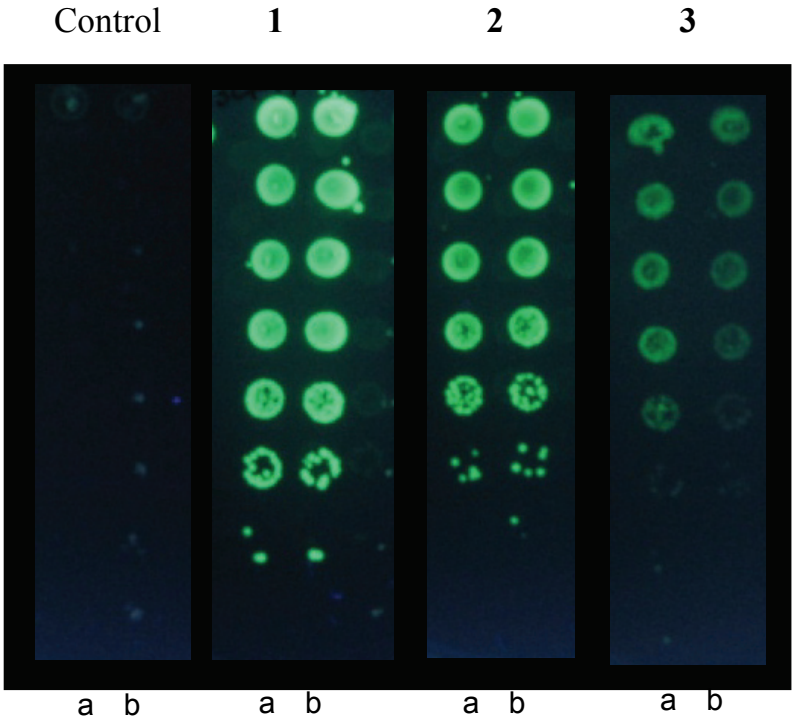
[6] Y. S. Wang, B. Wu, Z. Wang, Y. Huang, W. Wan, W. K. Russell, P. J. Pai, Y. N. Moe, D. H. Russell, W. R. Liu, *Mol. Biosyst.* **2010**, *6*, 1557.

[7] J. Xie, L. Wang, N. Wu, A. Brock, G. Spraggon, P. G. Schultz, *Nat. Biotechnol.* **2004**, *22*, 1297.

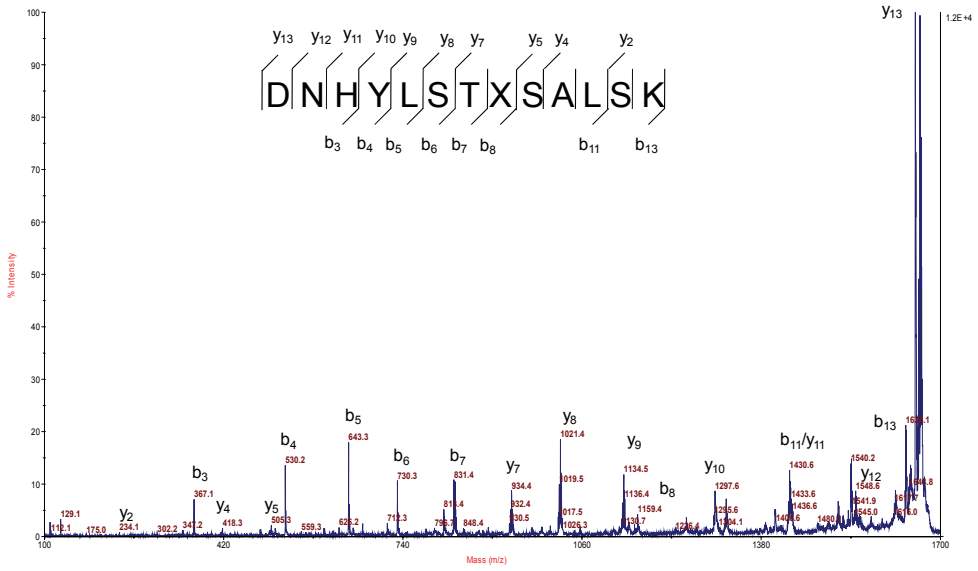
[8] T. S. Young, I. Ahmad, J. A. Yin, P. G. Schultz, *J. Mol. Biol.* **2010**, *395*, 361.

[9] J. Guo, J. Wang, J. S. Lee, P. G. Schultz, *Angew. Chem. Int. Ed.* **2008**, *47,* 6399.

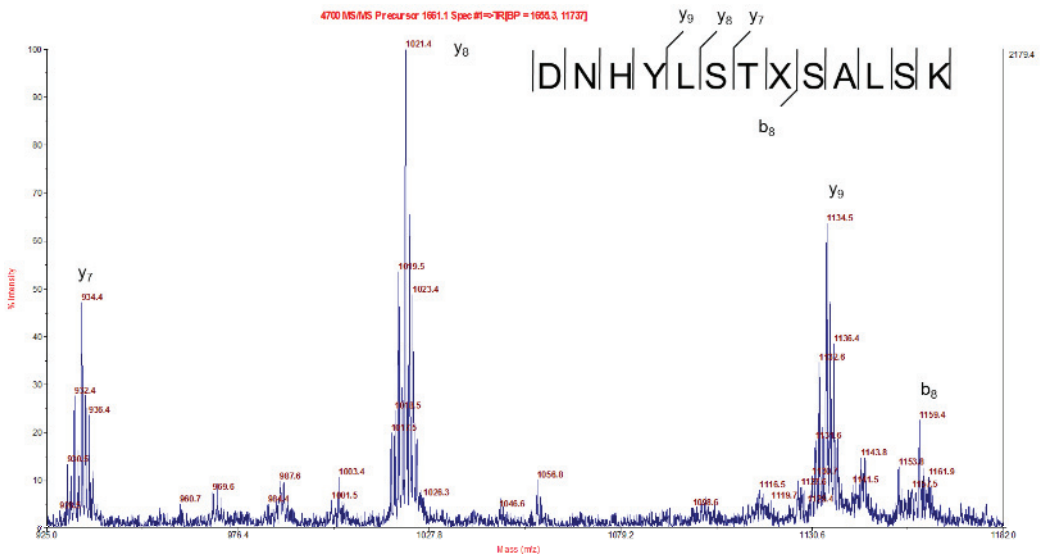
12. Supplemtary Figures



Supplementary Figure 1. Growth of two strains, mKRS1 (a) and mKRS (b), on LB plates with different supplements. The LB agar plates contain 102 μg/mL Cm, 25 μg/mL Kan and 12 μg/mL Tet. The TOP10 cells with pBK-mKRS1 (a) (Y306M/L309A/C348T/T364K/Y384F) and pY+ or pBK-mKRS (b) (Y306M/L309A/C348T/T364K) and pY+ were plated with serial dilution from 3×10^6 (1) cells to 0.3 (8) (10 fold for each dilution). The pY+ plasmid has a GFP_{UV} gene under the control of a T7 promoter. The expression is promoted by the suppression of two amber mutations at positions 1 and 107 of a T7 RNA polymerase gene in pREP. The fluorescent intensity of the expression of GFP_{UV} roughly represents the suppression efficiency at amber codons.

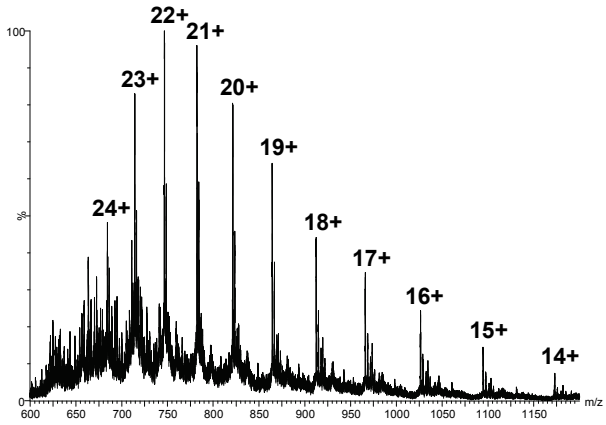


Supplementary Figure 2. Tandem MS analysis of a tryptic fragment of GFP_{UV} with **3** incorporated at Q204. X represents **3**.

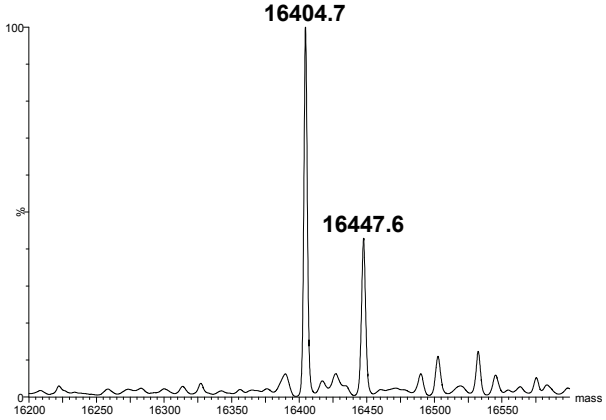


Supplementary Figure 3. Additional support of **3** in the tryptic peptide is provided by the observation of the selenium isotopes in tandem mass spectra. Shown here are the y_7, y_8, y_9 and b_8 fragment ions.

(A) ESI-IMMS spectrum

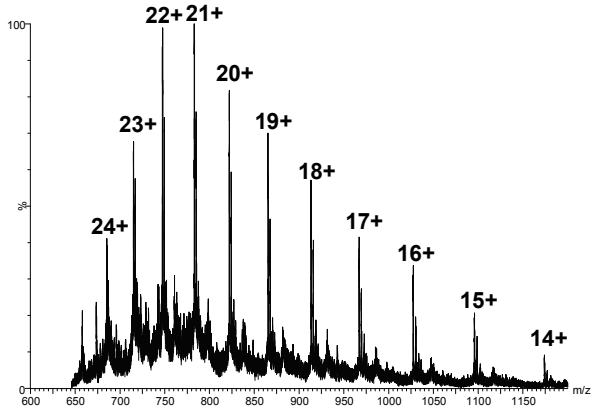


(B) The deconvoluted result

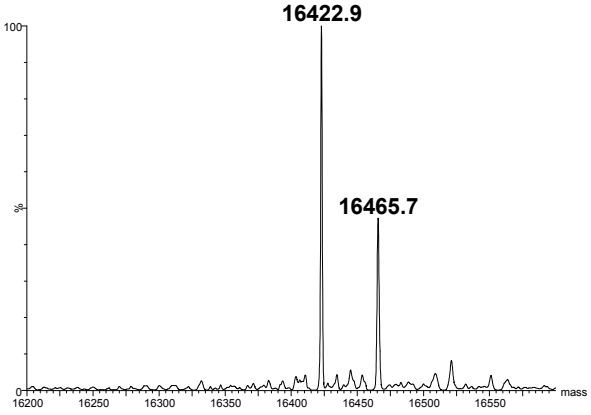


Supplementary Figure 4. (**A**) ESI-IM-MS and (**B**) the deconvoluted spectra of H3K9-**1**.

(A) ESI-IMMS spectrum

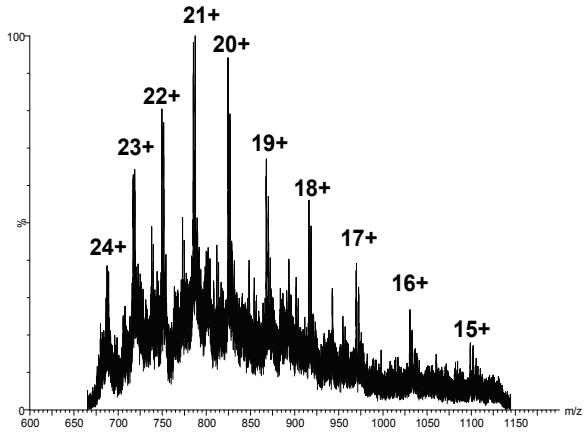


(B) The deconvoluted result

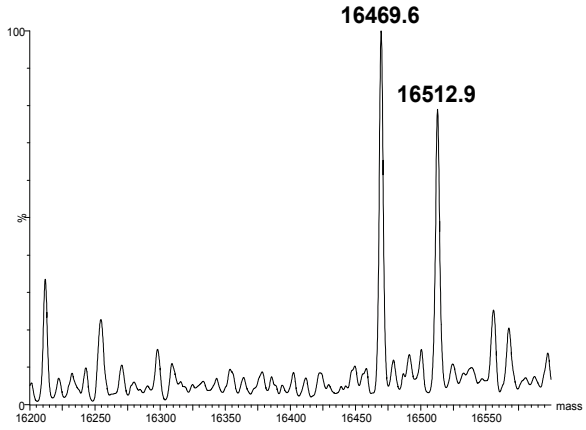


Supplementary Figure 5. (**A**) ESI-IM-MS and (B) the deconvoluted spectra of H3K9-**2**.

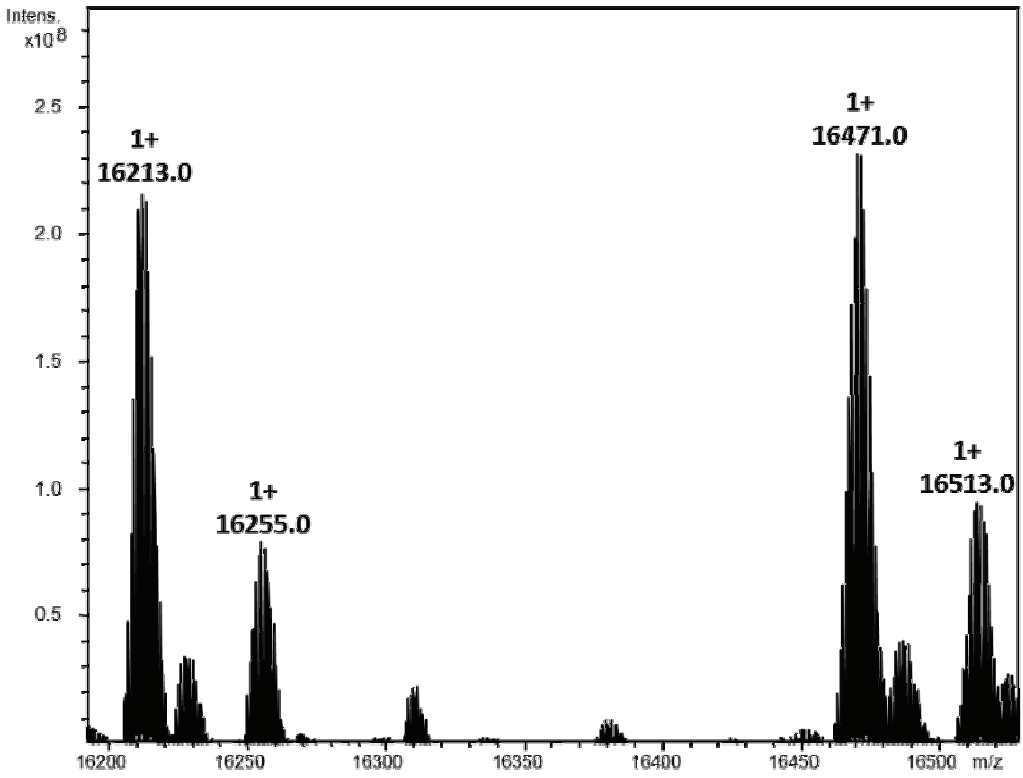
(A) ESI-IMMS spectrum



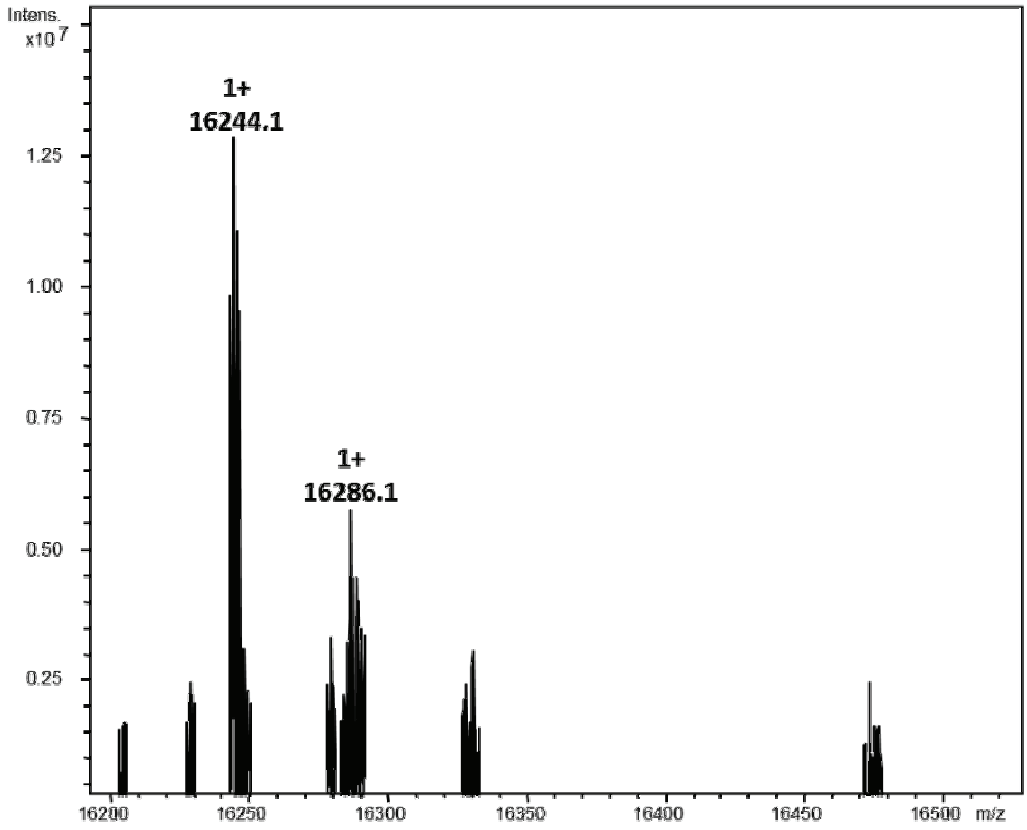
(B) The deconvoluted result



Supplementary Figure 6. (**A**) ESI-IM-MS and (B) the deconvoluted spectra of H3K9-**3**.

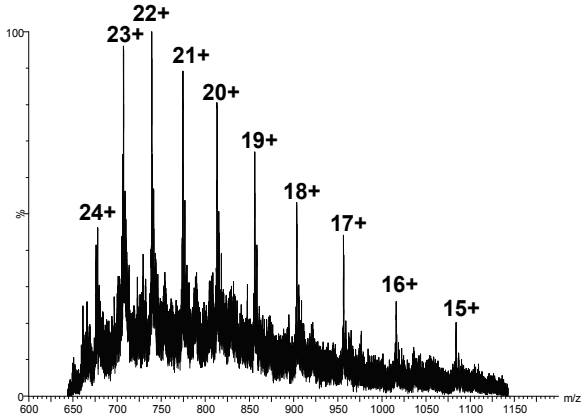


Supplementary Figure 7. Mass determination of H3K9-**3** after its treatment with 5eq H_2O_2. The deconvoluted singly charged ESI-MS spectrum by FT-ICR MS. Both H3K9-**3** and H3K9Dha were found.

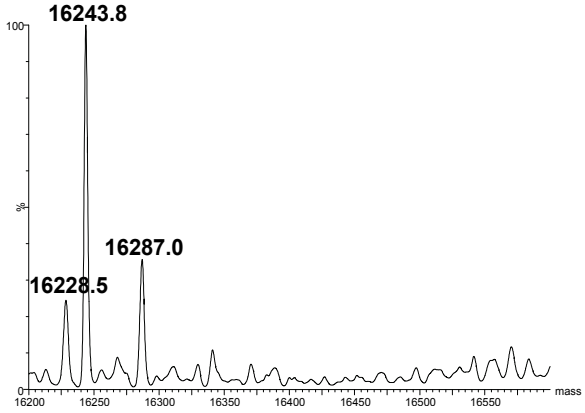


Supplementary Figure 8. Massdetermination of H3K9Dha. The deconvoluted singly charged ESI-MS spectrum of H3K9Dha by FT-ICR MS. H3K9Dha was synthesized from H3K9-**3** by treatment with 100 eq. H_2O_2 at room temperature for 1 h.

(A) ESI-IMMS spectrum

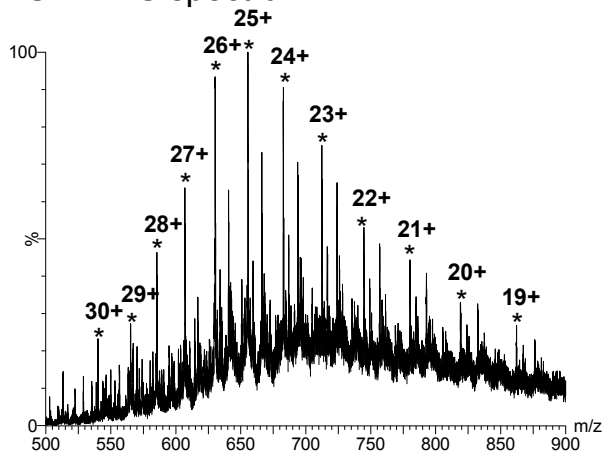


(B) The deconvoluted result

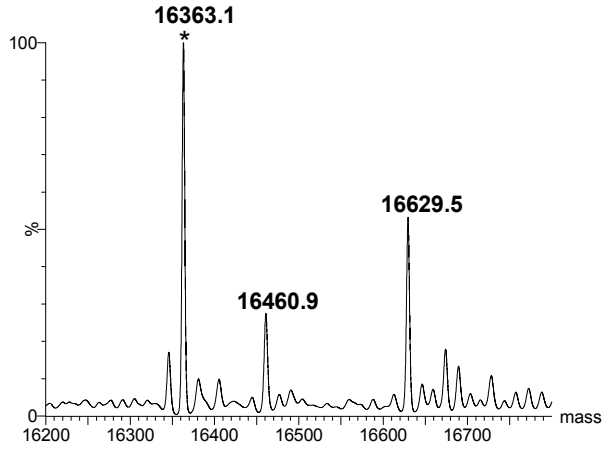


Supplementary Figure 9. Mass determination of H3K9Dha: (A) ESI-IM-MS and (B) the deconvoluted spectra of H3K9Dha. H3K9Dha was synthesized from H3K9-**3** by treatment with 100 eq. H_2O_2 at room temperature for 1 h.

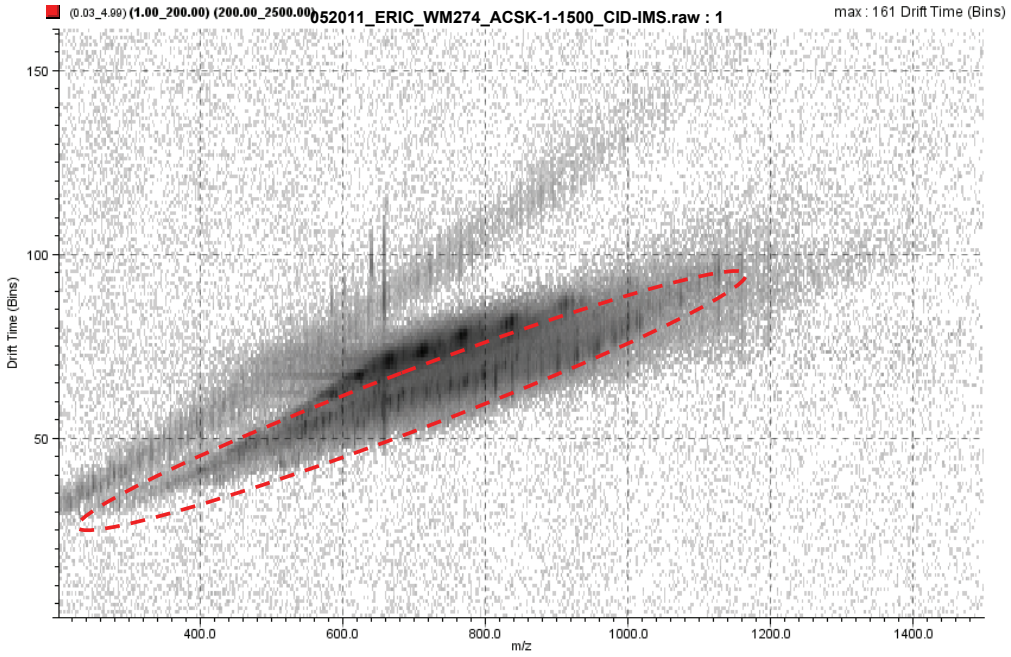
(A) ESI-IMMS spectrum



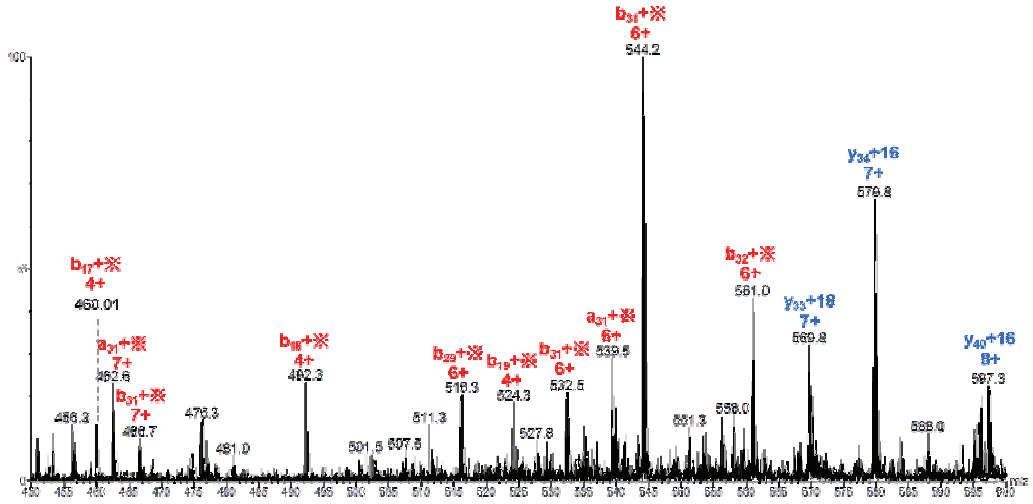
(B) The deconvoluted result



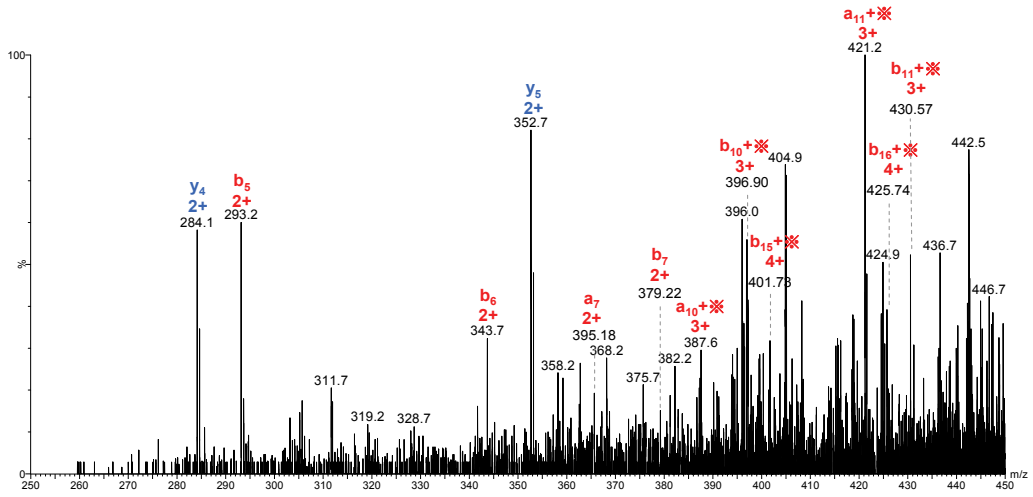
Supplementary Figure 10. Mass determination of the protein H3K9AcsK: (A) ESI-IM-MS and (B) the deconvoluted spectra of H3K9AcsK.



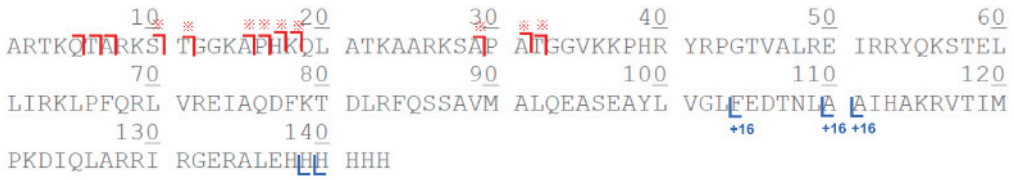
Supplementary Figure 11. Top-down analysis of H3K9AcsK by CID-IM-MS.



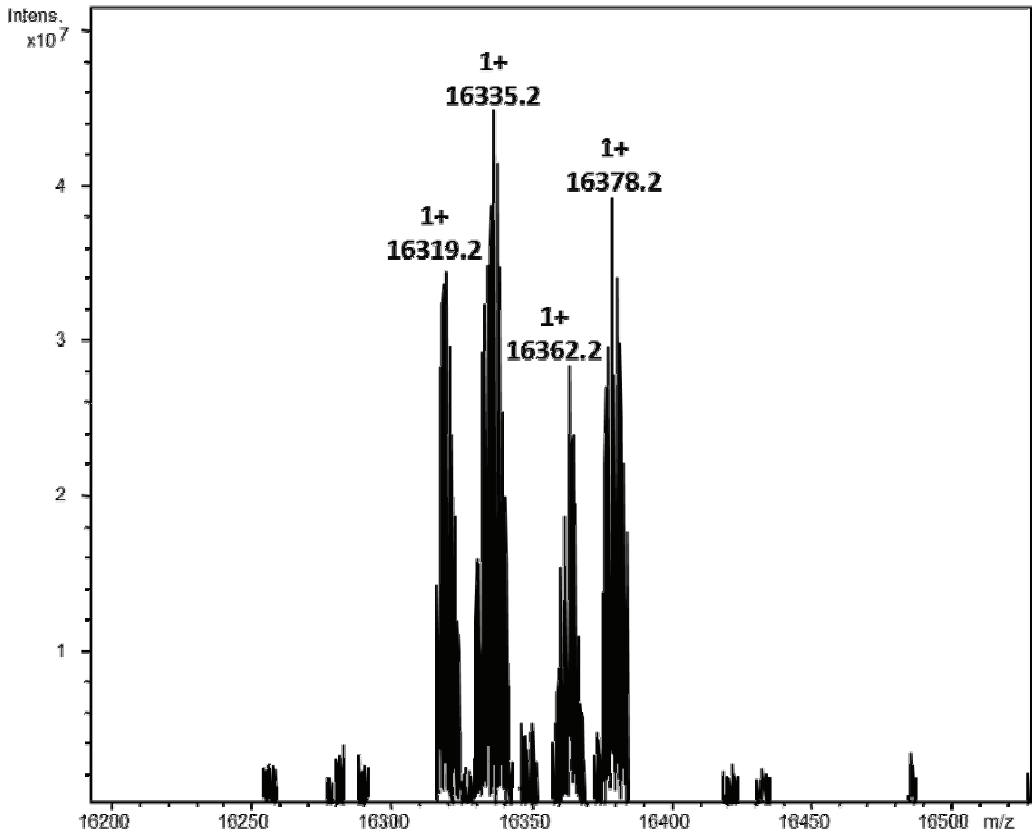
Supplementary Figure 12. The extracted MS/MS spectrum for top-down analysis of H3K9AcsK by CID-IM-MS. The extracted MS/MS region was represented as red dotted line. The spectrum was shown in zoom-in region of m/z 450-600.



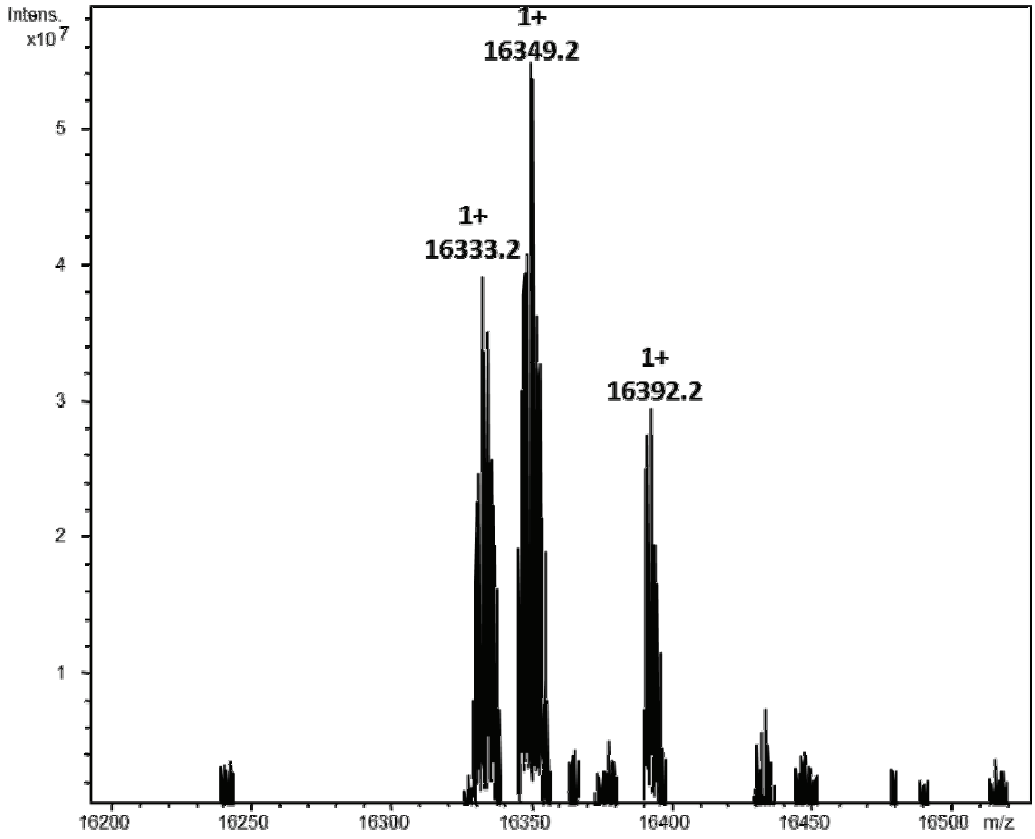
Supplementary Figure 13. The extracted MS/MS spectrum for top-down analysis of H3K9AcsK by CID-IM-MS. The spectrum was shown in zoom-in region of m/z 250-450.



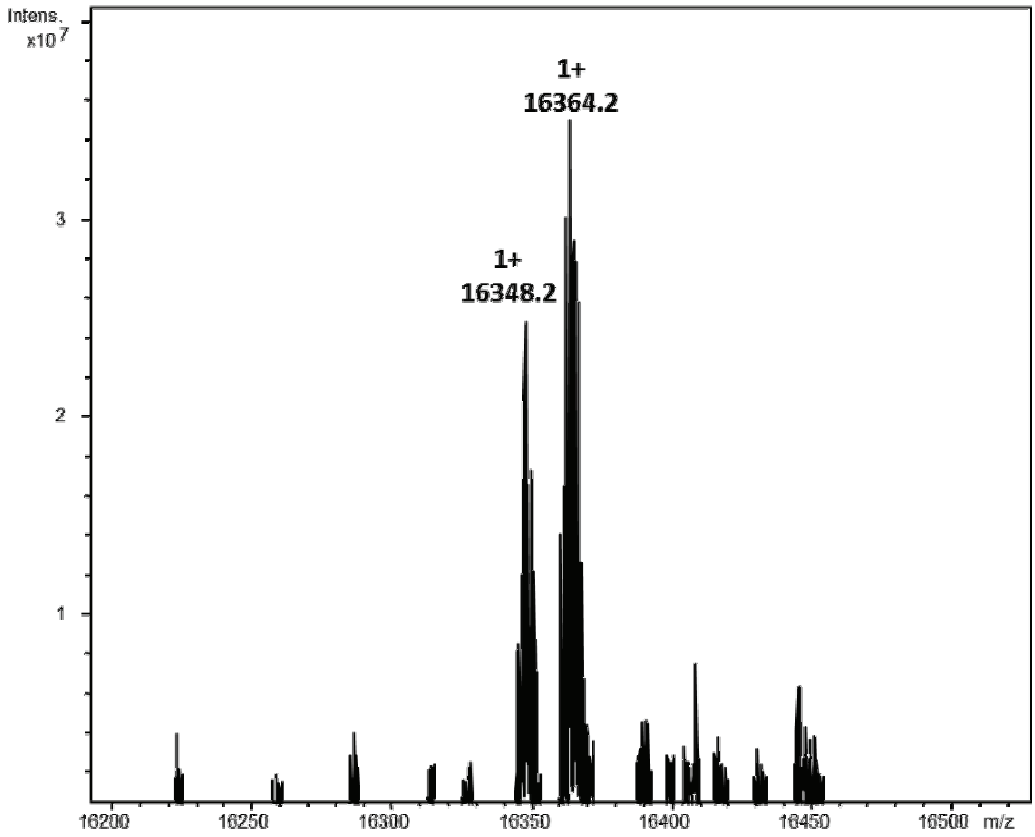
Supplementary Figure 14. The summary of Top-down analysis for H3K9AcsK by CID-IM-MS. This result indicates the AcsK modification site is between the protein sequences of R8 to S10. ※ denotes the AcSK modification (M+60); +16 denotes methionine oxidation (M+16). The b fragment ions were represented in red and y in blue.



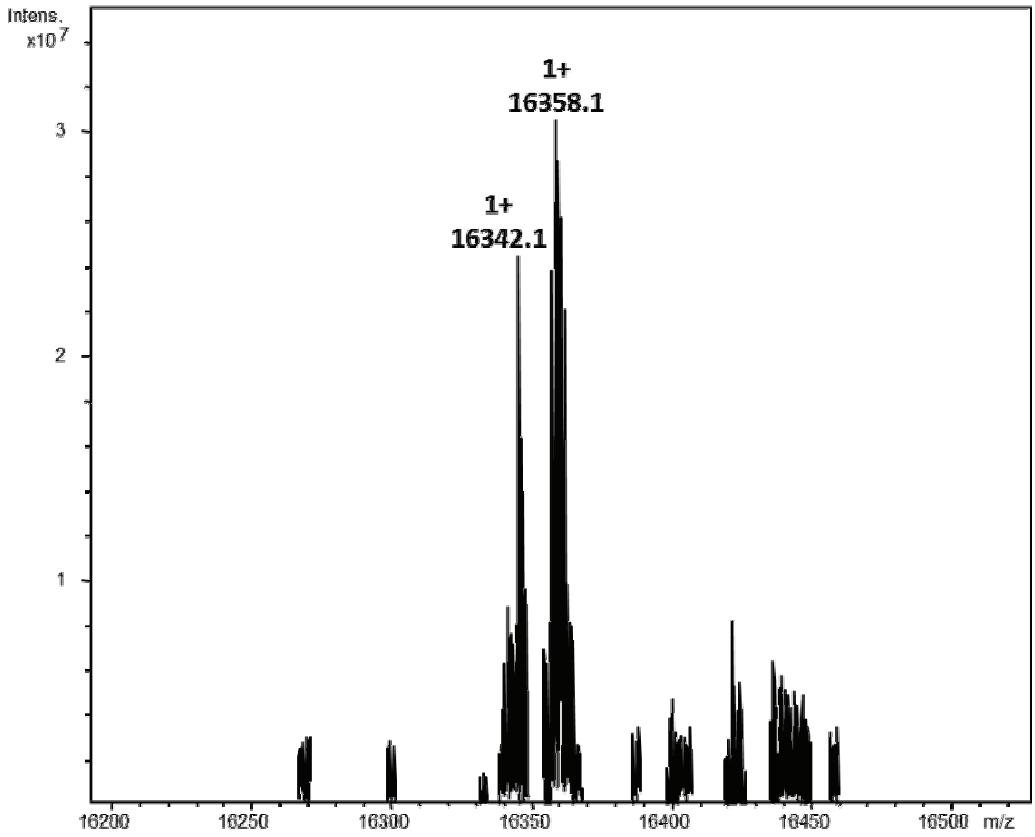
Supplementary Figure 15. Mass determination of Protein H3K9mesK. The deconvoluted singly charged ESI-MS spectrum of H3K9 H3K9mesK by FT-ICR MS.



Supplementary Figure 16. Mass determination of Protein H3K9m^{2}sK. The deconvoluted singly charged ESI-MS spectrum of H3K9 H3K9m^{2}sK by FT-ICR MS.

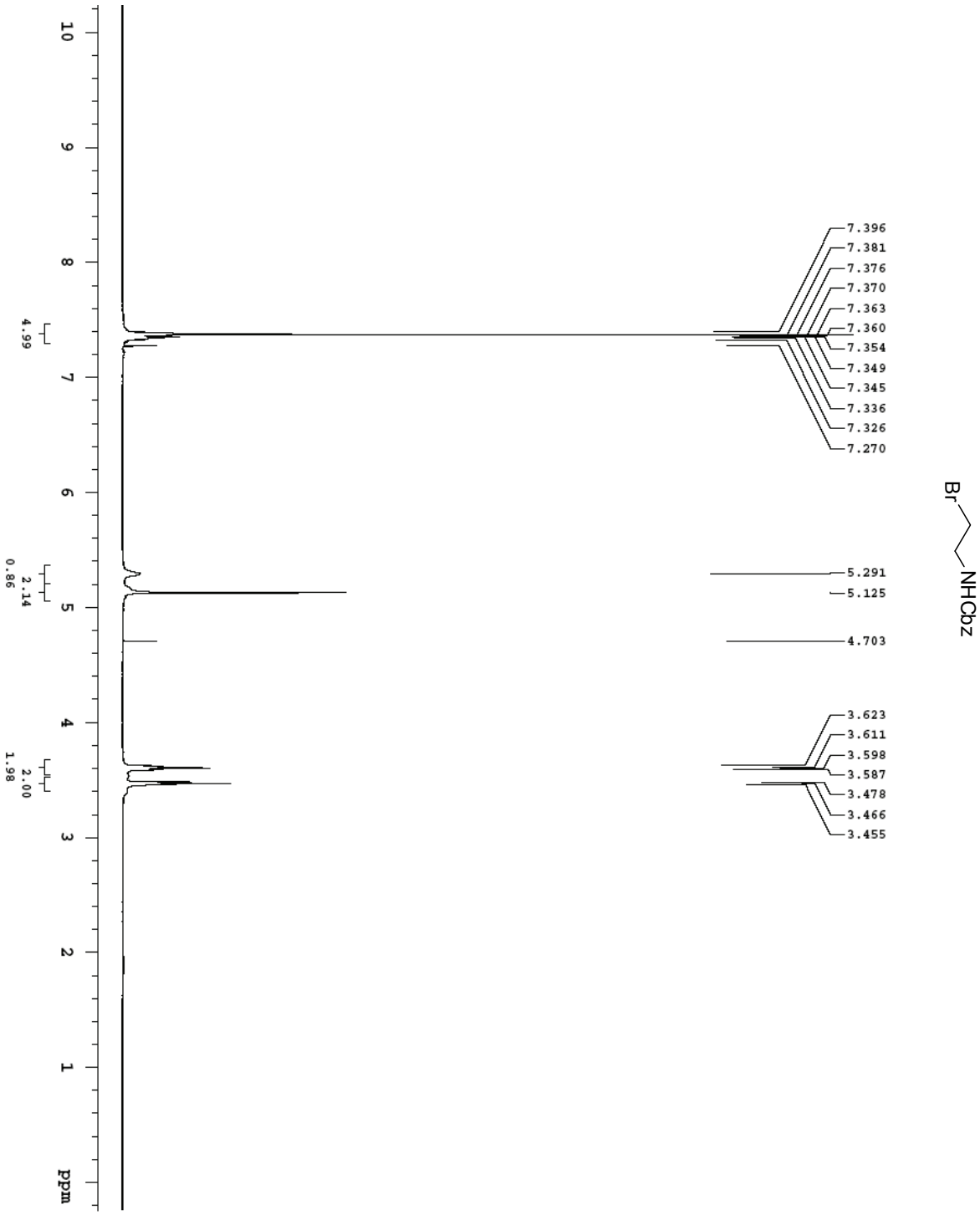


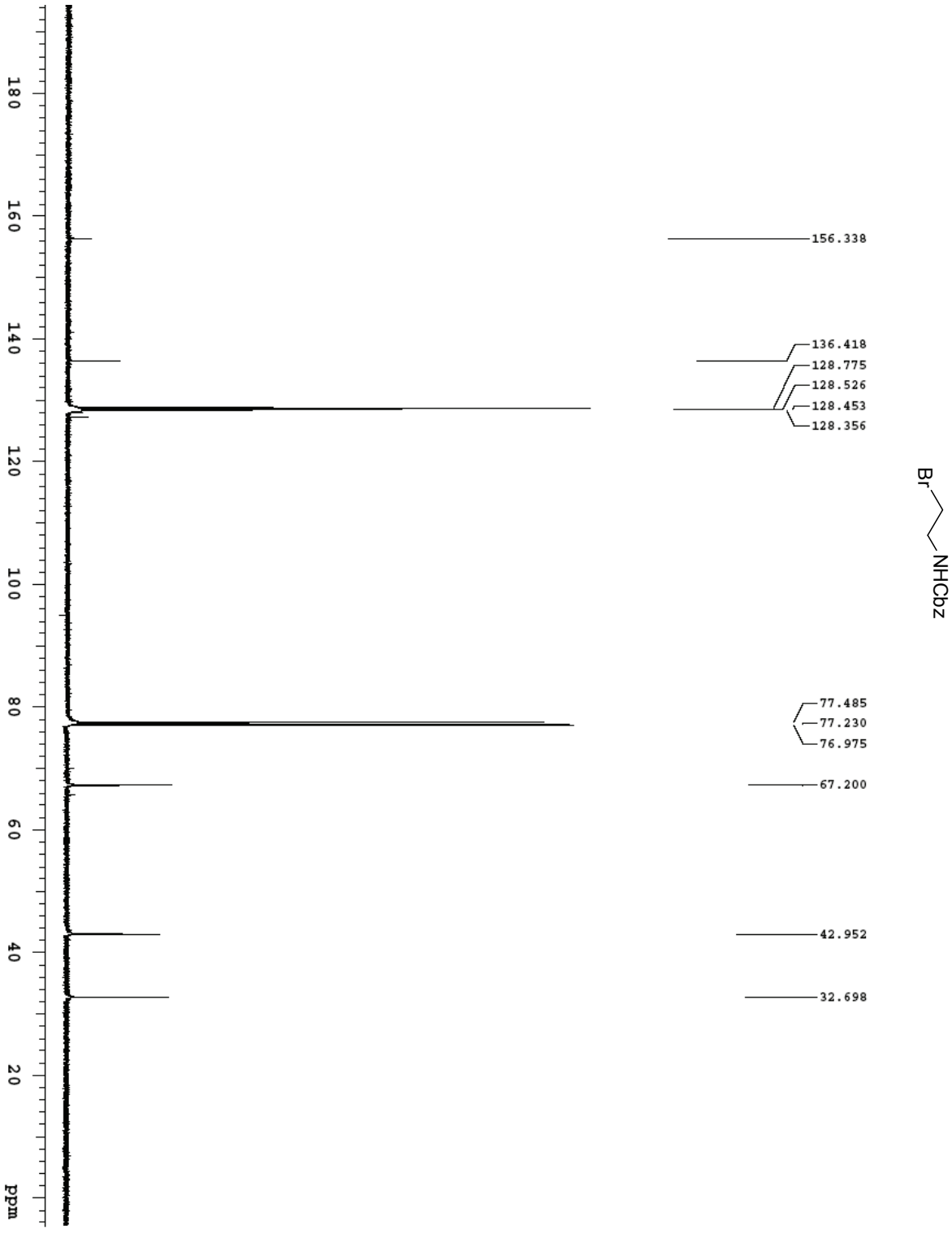
Supplementary Figure 17. Mass determination of Protein H3K9m^{3}sK. The deconvoluted singly charged ESI-MS spectrum of H3K9 H3K9m^{3}sK by FT-ICR MS.

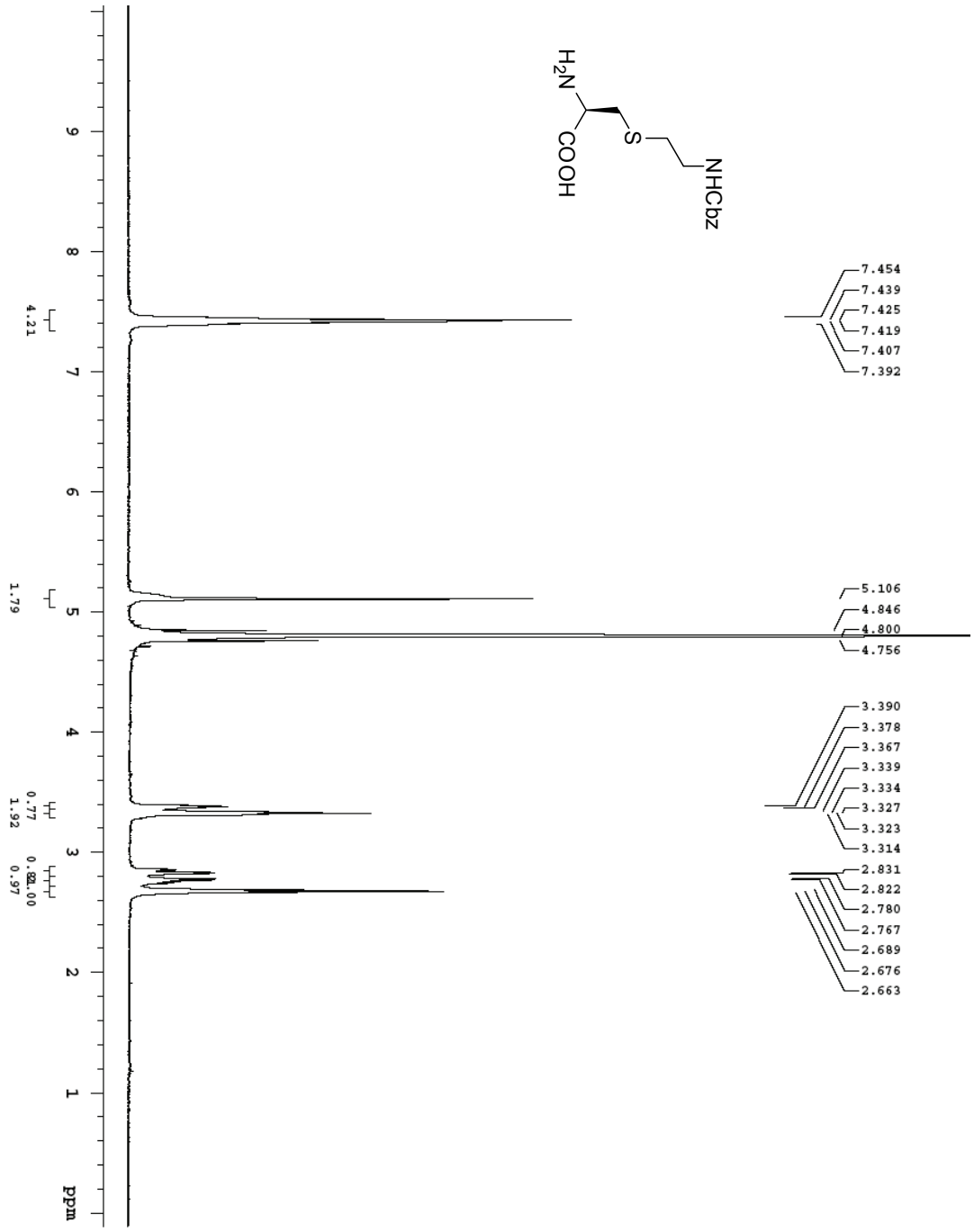


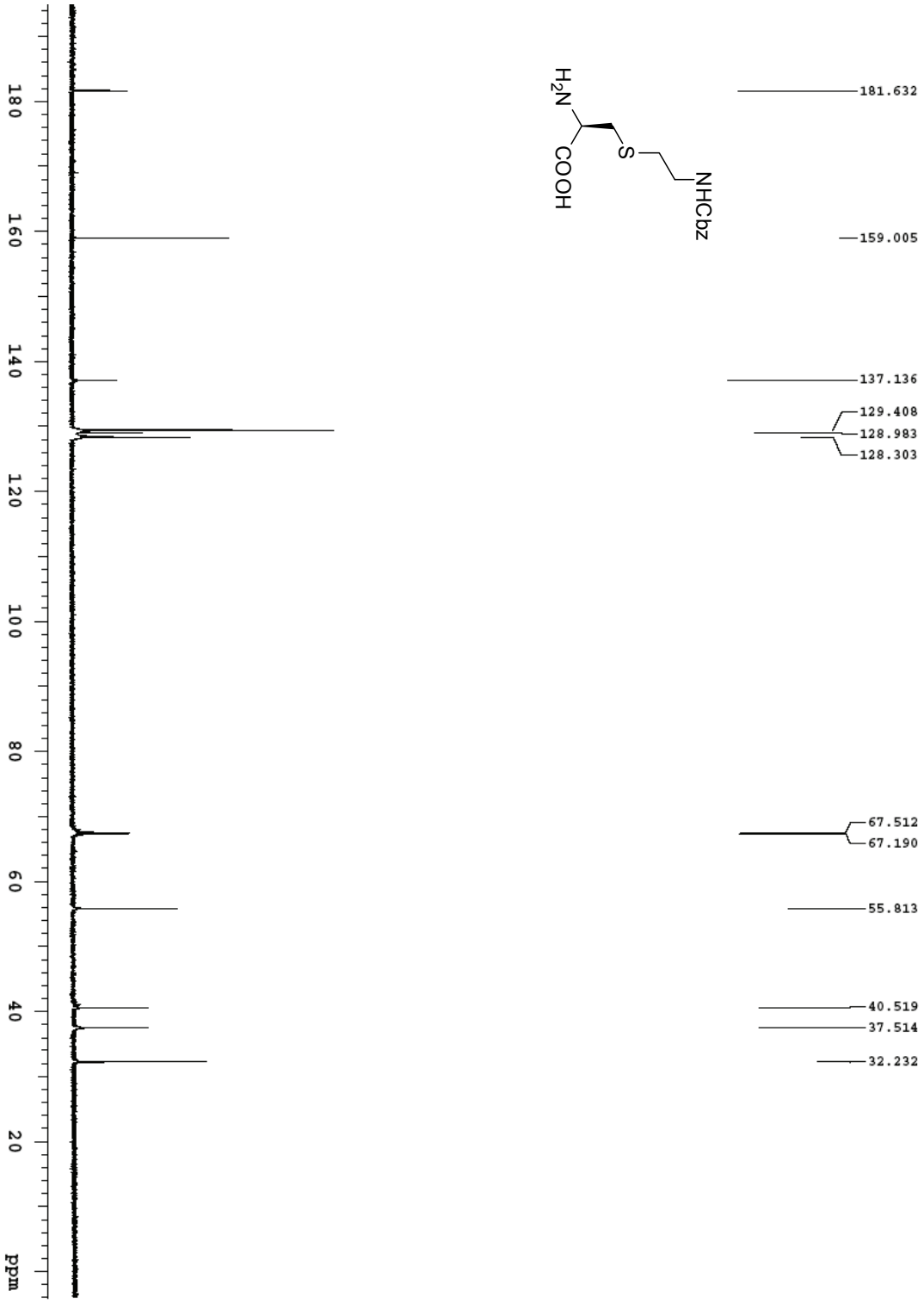
Supplementary Figure 18. Mass determination of Protein H3K9pC. The deconvoluted singly charged ESI-MS spectrum of H3K9 H3K9pC by FT-ICR MS.

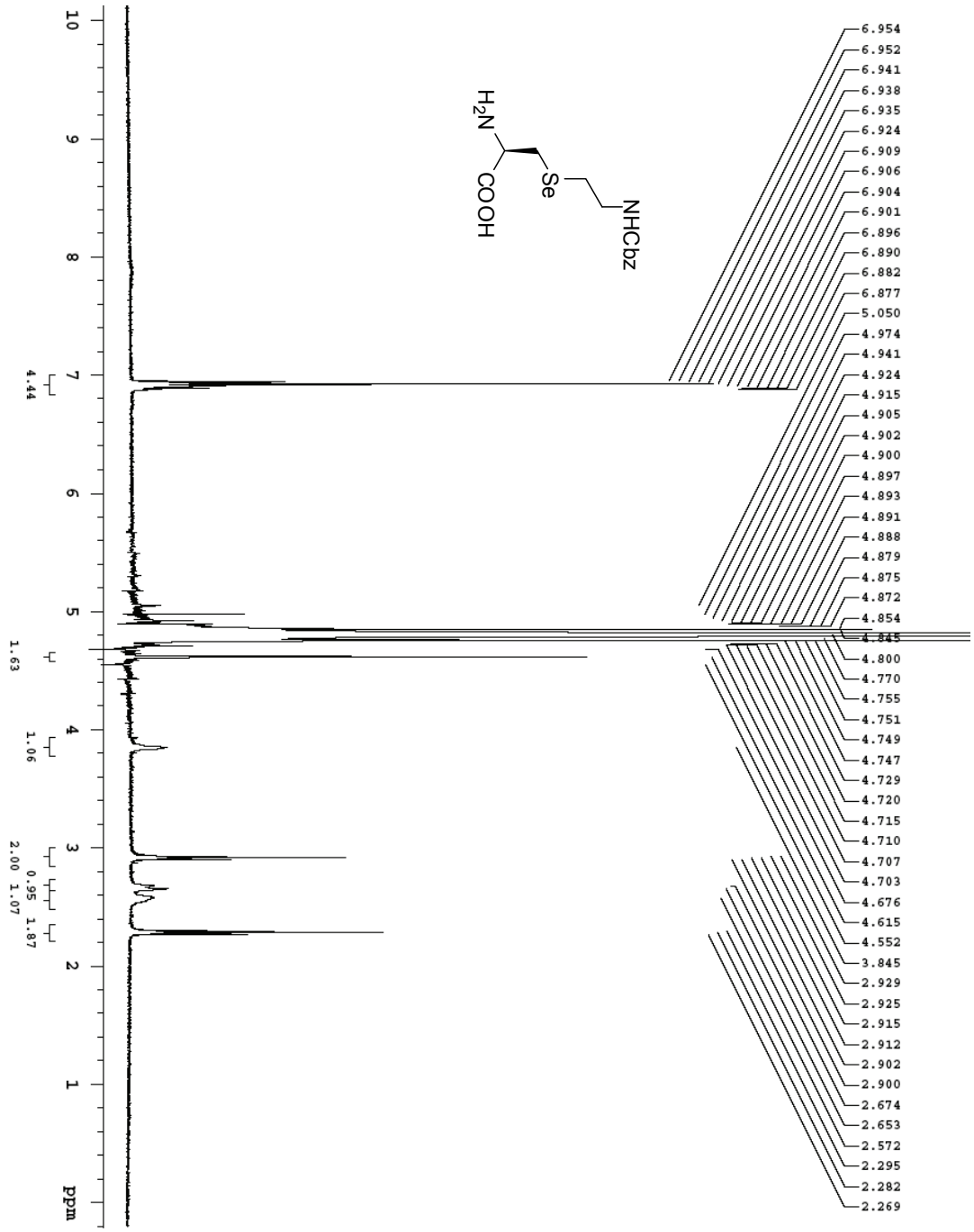
13. NMR Spectra of Synthesized Compounds

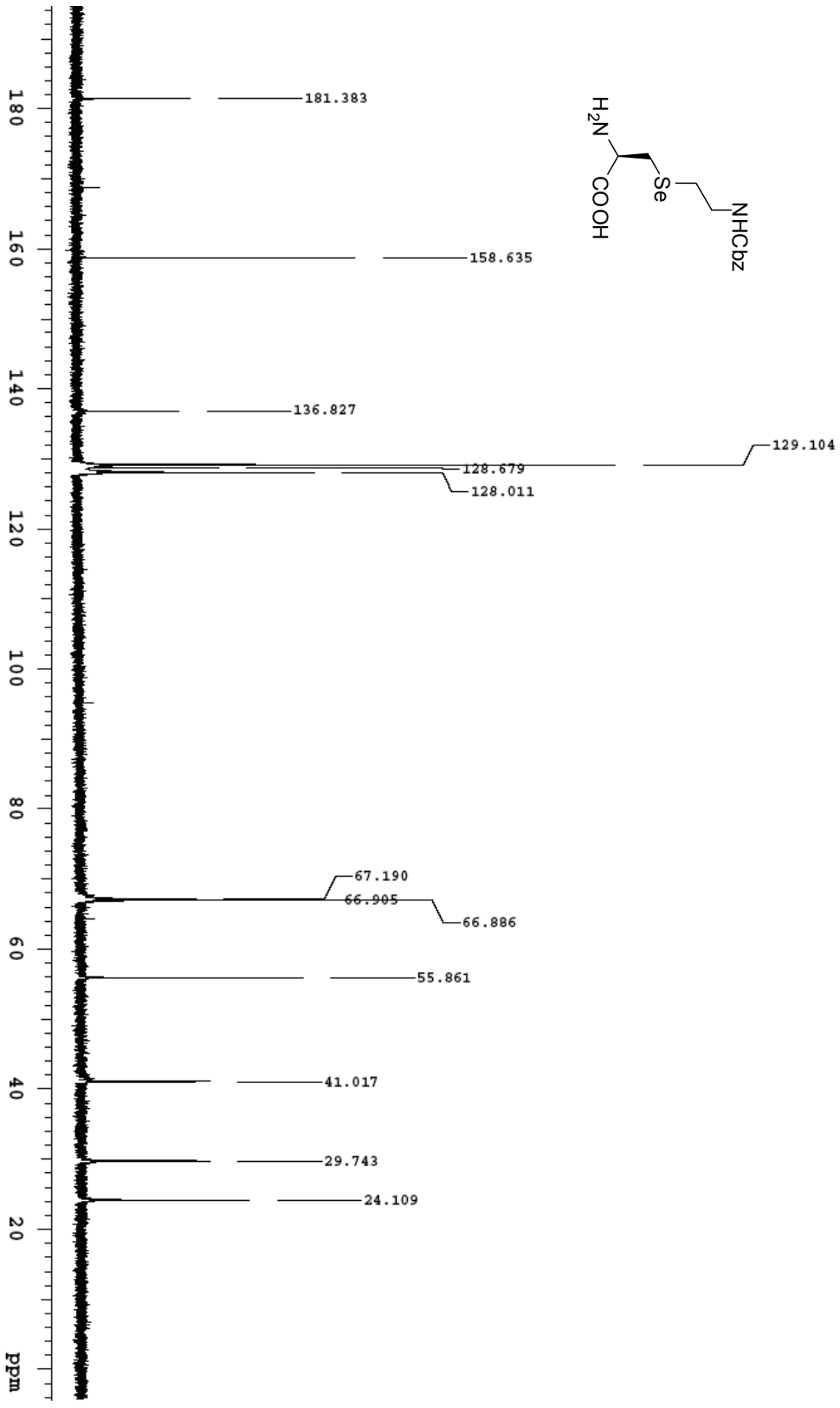



156.338
136.418
128.775
128.526
128.453
128.356
77.485
77.230
76.975
67.200
42.952
32.698
Br
NHCbz
180
160
140
120
100
80
60
40
20
ppm


H2N
COOH
S
NHCbz
7.454
7.439
7.425
7.419
7.407
7.392
5.106
4.846
4.800
4.756
3.390
3.378
3.367
3.339
3.334
3.327
3.323
3.314
2.831
2.822
2.780
2.767
2.689
2.676
2.663
4.21
1.79
0.77
1.92
0.97
ppm


H2N
COOH
S
NHCqz
181.632
159.005
137.136
129.408
128.983
128.303
67.512
67.190
55.813
40.519
37.514
32.232
180
160
140
120
100
80
60
40
20
ppm


H2N
COOH
Se
NHCbz
6.954
6.952
6.941
6.938
6.935
6.924
6.909
6.906
6.904
6.901
6.896
6.890
6.882
6.877
5.050
4.974
4.941
4.924
4.915
4.905
4.902
4.900
4.897
4.893
4.891
4.888
4.879
4.875
4.872
4.854
4.845
4.800
4.770
4.755
4.751
4.749
4.747
4.729
4.720
4.715
4.710
4.707
4.703
4.676
4.615
4.552
3.845
2.929
2.925
2.915
2.912
2.902
2.900
2.674
2.653
2.572
2.295
2.282
2.269
4.44
1.63
1.06
2.00
1.07
0.95
1.87
ppm


181.383
158.635
136.827
129.104
128.679
128.011
67.190
66.905
66.886
55.861
41.017
29.743
24.109
180
160
140
120
100
80
60
40
20
ppm
H2N
COOH
Se
NHCbz