Electronic Supplementary Material (ESI) for Chemical Communications

Supplementary information:

Synthesis of ε-*N*-propionyl-, ε-*N*-butyryl-, and ε-*N*-crotonyl-lysine containing histone H3 using the pyrrolysine system

Michael J. Gattner, Milan Vrabel and Thomas Carell

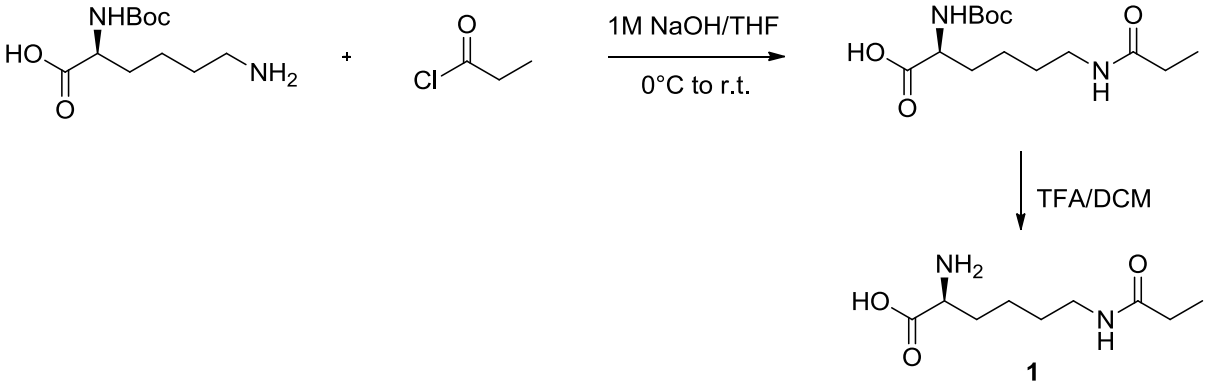
Content

1 General methods

Chemicals were purchased from *Sigma-Aldrich*, *Fluka, ABCR* or *ACROS* and used without further purification. Solvents used were of reagent grade and purified by standard methods. Reactions were monitored on *Merck* Silica 60 F254 TLC plates. Detection was done by irradiation with UV light (254 nm or 366 nm) and/or staining with *p*-anisaldehyde solution in ethanol or CAM staining solution. NMR spectra were recorded on the following spectrometers: *Varian Oxford 200, Bruker AC 300, Varian XL 400* and *Bruker AMX 600*. The chemical shifts (δ) are given in ppm, the coupling constants (*J*) in Hz. The peak assignment were done using 2D spectra. The solutions are referred to residual undeuterated solvents as internal standard. Mass spectra were recorded on the following machines: *Thermo Finnigan LTQ-FT* (ESI-ICR) and *Thermo LTQ-Orbitrap XL*. For analytical HPLC separations of protein and peptide samples with subsequent MS a *Dionex Ultimate 3000 Nano* HPLC was used. Proteins were purified on an *ÄKTA purifier* system.

2 Synthesis of the compounds

2.1 ε -*N*-propionyl lysine 1:



Starting N_α-Boc-Lys-OH (5g, 20.3 mmol) was dissolved in 1M NaOH (50 ml) / THF (50 ml) mixture and the solution was cooled down in ice-bath. Propionyl chlorid (1.95 ml, 22.3 mmol, 1.1 equiv.) was added dropwise to this solution through rubber septum under argon atmosphere. The reaction mixture was stirred at room temperature overnight. Then it was cooled down in an ice-bath again and washed with ice-cold Et_20 (100 ml). The water-phase was acidified to pH=4 with 1N HCl and extracted with ice-cold AcOEt (3x250 ml). The combined organic phases were dried over Na_2SO_4 and the solvent removed under vacuum. Drying on HV (high vacuum) afforded crude N_α-Boc-protected propionyl lysine (5.85 g, 95%) as colorless viscous oil that was used directly in the next step.

HR-ESI MS: $[M+H]^+$ calculated: 303.1920, $[M+H]^+$ found: 303.1917.

To the solution of N_α-*Boc*-protected propionyl lysine in dry DCM (50 ml) was dropwise added TFA (trifluoro acetic acid) (10 ml) and the resulting solution was stirred at room temperature until the disappearance of the starting material on TLC (ca.1.5 hours, TLC in DCM/MeOH = 4/1 + 0.1 % AcOH, CAM staining solution). The volatiles were removed under reduced pressure and the resulting oily residue was taken up in a minimum amount of DCM. The crude product was precipitated from this solution by the addition of Et_20 (ca. 50 ml). The solvents were decanted and the residue dried on HV. The product thus obtained was re-dissolved in water and lyophilized overnight to give 5.9 g (96%) of ε-*N*-propionyl lysine **1** as a

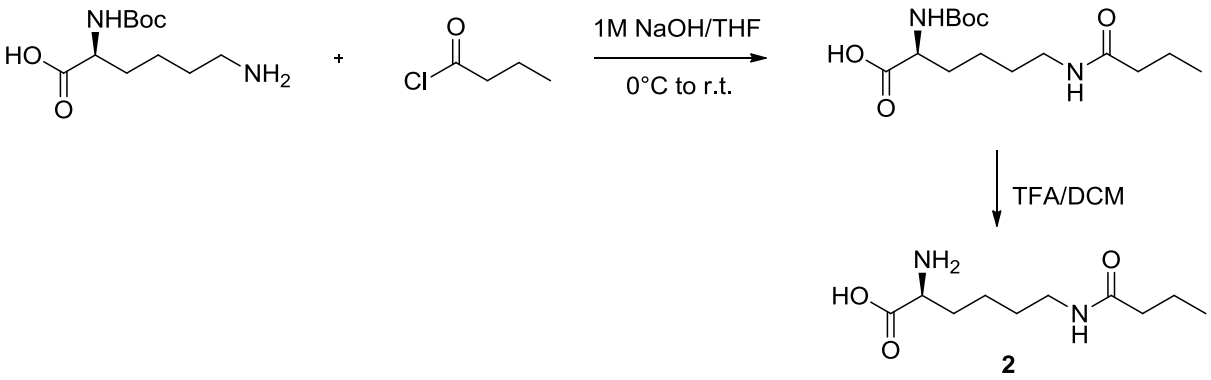
colorless solid. The product was obtained as TFA salt and used as such for the protein expression.

^{1}H-NMR (D_2O) δ = 4.09 (t, *J* = 6.28, 1H, C*H*^{Lys}), 3.20 (t, *J* = 7.03, 2H, C*H*$_2^{Lys}$), 2.24 (q, *J* = 7.56 and 7.38, 2H, C*H*$_2^{Prop}$), 1.97 (m, 2H, C*H*$_2^{Lys}$), 1.37-1.63 (m, 4H, 2xC*H*$_2^{Lys}$), 1.11 (t, 3H, *J* = 7.56, C*H*$_3^{Prop}$) ppm.

^{13}C-NMR (D_2O) δ = 177.9 (NH*C*O^{Prop}), 171.9 (*C*OOH^{Lys}), 52.6 (*C*H^{Lys}), 38.6 (*C*H$_2^{Lys}$), 29.2 (*C*H$_2^{Lys}$), 29.1 (*C*H$_2^{Prop}$), 27.7 (*C*H$_2^{Lys}$), 21.4 (*C*H$_2^{Lys}$), 9.6 (*C*H$_3^{Prop}$) ppm.

HR-ESI MS: $[M+H]^+$ calculated: 203.1396, $[M+H]^+$ found: 203.1391; $[M-H]^-$ calculated: 201.1239, $[M-H]^-$ found: 201.1246.

2.2 ε -*N*-butyryl lysine 2:



Starting *N*$_α$-Boc-Lys-OH (5g, 20.3 mmol) was dissolved in 1M NaOH (50 ml) / THF (50 ml) mixture and the solution was cooled down in ice-bath. Butyryl chlorid (2.1 ml, 20.3 mmol, 1 equiv.) was added dropwise to this solution through rubber septum under argon atmosphere. The reaction mixture was stirred at room temperature overnight. Then it was cooled down in an ice-bath again and washed with ice-cold Et_20 (100 ml). The water-phase was acidified to pH=4 with 1N HCl and extracted with ice-cold AcOEt (3x250 ml). The combined organic phases were dried over Na_2SO_4 and the solvent removed under vacuum. Drying on HV (high vacuum) afforded crude *N*$_α$-Boc-protected butyryl lysine (6 g, 94%) as colorless viscous oil that was used in the next step.

HR-ESI MS: $[M+H]^+$ calculated: 317.2076, $[M+H]^+$ found: 317.2073.

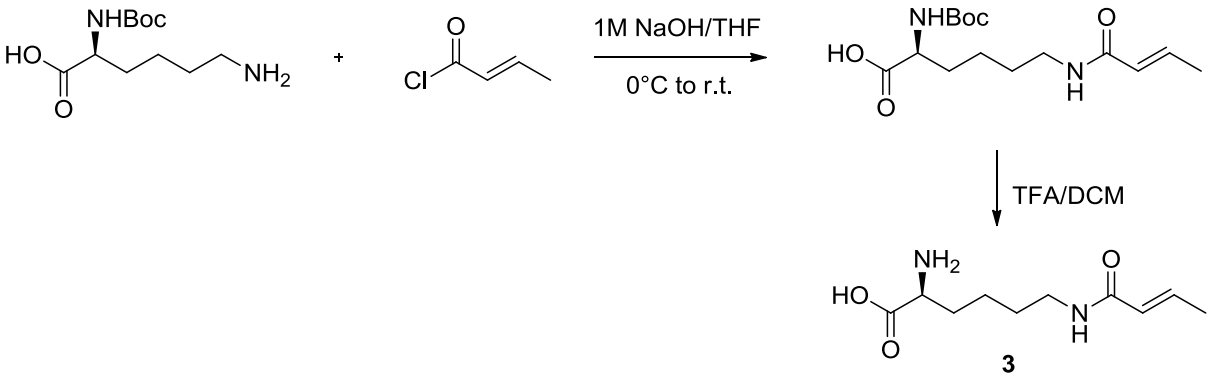
To the solution of N_α-Boc-protected butyryl lysine in dry DCM (50 ml) was added TFA (10 ml) and the resulting solution was stirred at room temperature until the disappearance of the starting material on TLC (ca.1.5 hours, TLC in DCM/MeOH = 4/1 + 0.1 % AcOH, CAM staining solution). All volatiles were removed under reduced pressure and the resulting oily residue was taken up in a minimum amount of DCM. The crude product was precipitated from this solution by the addition of Et_20 (ca. 50 ml). The solvents were decanted and the residue dried on HV. The product thus obtained was re-dissolved in water and lyophilized overnight to give 6.1 g (97%) of ε-*N*-butyryl lysine **2** as a colorless solid. The product was obtained as TFA salt and used as such for the protein expression.

^{1}H-NMR (D_2O) δ = 4.03 (t, *J* = 6.30, 1H, CH^{Lys}), 3.18 (t, *J* = 6.82, 2H, CH_2^{Lys}), 2.18 (t, *J* = 6.72, 2H, CH_2^{But}), 1.85-2.02 (m, 2H, CH_2^{Lys}), 1.33-1.62 (m, 6H, 2xCH_2^{Lys}, CH_2^{But}), 0.87 (t, 3H, *J* = 7.41, CH_3^{But}) ppm.

^{13}C-NMR (D_2O) δ = 176.9 (NHCO^{But}), 172.1 ($COOH^{Lys}$), 52.7 (CH^{Lys}), 38.6 (CH_2^{Lys}), 37.6 (CH_2^{But}) 29.3 (CH_2^{Lys}), 27.7 (CH_2^{But}), 21.4 (CH_2^{Lys}), 18.9 (CH_2^{Lys}), 12.5 (CH_3^{But}) ppm.

HR-ESI MS: $[M+H]^+$ calculated: 217.1552, $[M+H]^+$ found: 217.1547; $[M-H]^-$ calculated: 215.1396, $[M-H]^-$ found: 215.1403.

2.3 ε-N-crotonyl lysine 3:



Starting N_α-Boc-Lys-OH (10g, 40.5 mmol) was dissolved in 1M NaOH (100 ml) / THF (100 ml) mixture and the solution was cooled down in ice-bath. *trans*-crotonyl chlorid (3.9 ml, 40.5

mmol, 1 equiv.) was added dropwise to this solution through rubber septum under argon atmosphere. The reaction mixture was stirred at room temperature overnight. Then it was again cooled down in an ice-bath and washed with ice-cold Et_20 (ca. 100 ml). The water-phase was acidified to pH=4 with 1N HCl and extracted with ice-cold AcOEt (3x250 ml). The combined organic phases were dried over Na_2SO_4 and the solvent removed under vacuum. Drying on HV (high vacuum) afforded crude N_α-Boc-protected crotonyl lysine (12g, 94%) as colorless viscous oil that was used directly in the next step.

HR-ESI MS: $[M+H]^+$ calculated: 315.1920, $[M+H]^+$ found: 315.1915; $[M-H]^-$ calculated: 313.1763, $[M-H]^-$ found: 313.1770.

To the solution of N_α-Boc-protected crotonyl lysine in dry DCM (100 ml) was added TFA (20 ml) and the resulting solution was stirred at room temperature until the disappearance of the starting material on TLC (ca.1.5 hours, TLC in DCM/MeOH = 4/1 + 0.1 % AcOH, CAM staining solution). The volatiles were removed under reduced pressure and the resulting oily residue was taken up in a minimum amount of DCM. The crude product was precipitated from this solution by the addition of Et_20. The solvents were decanted and the residue dried on HV. The product thus obtained was re-dissolved in water and lyophilized overnight to give 12.1 g (97%) of ε-*N*-crotonyl lysine **3** as colorless solid. The product was obtained as TFA salt and used as such for the protein expressions.

^{1}H-NMR (D_2O) δ = 6.80 (m, 1H, CH^{Crot}), 5.99 (m, 1H, CH^{Crot}), 4.09 (t, J = 6.38, 1H, CH^{Lys}), 3.29 (t, J = 6.85, 2H, CH^{Lys}), 1.91-2.09 (m, 2H, CH_2^{Lys}), 1.86-1.91 (m, 3H, CH_3^{Crot}), 1.37-1.67 (m, 4H, 2xCH_2^{Lys}) ppm.

^{13}C-NMR (D_2O) δ = 172.2 (COOH^{Lys}), 168.9 (NHCO^{Crot}), 141.6 (CH^{Crot}), 123.7 (CH^{Crot}), 52.8 (CH^{Lys}), 38.6 (CH$_2^{Lys}$), 33.5 (CH$_2^{Lys}$), 29.3 (CH$_2^{Lys}$), 27.8 (CH$_2^{Lys}$), 21.6 (CH$_2^{Lys}$), 17.0 (CH$_3^{Crot}$) ppm.

HR-ESI MS: $[M+H]^+$ calculated: 215.1396, $[M+H]^+$ found: 215.1388; $[M-H]^-$ calculated: 213.1239, $[M-H]^-$ found: 213.1244.

3 Cloning and mutagenesis

3.1 pETDuet_yfp$_{TAG}$, pylS,T

The cloning procedure of pETDuet_yfp$_{TAG}$ is described in the supporting information of an earlier publication of our group[1]. In brief, *M. mazei pylS* and *pylT* were inserted into the first multiple cloning site using the restriction sites *BamH*I/*Not*I and *Not*I/*Afl*II, respectively. The gene encoding for the C-terminally StrepII-tagged eYFP Lys114Amber was inserted into the second multiple cloning site using the restriction sites *Nde*I/*Kpn*I. In contrast to the previously published construct, the first T7-promoter was not replaced by a *Trp*-promoter and three copies of the *pylT* gene were inserted instead of one.

3.2 pACyc_3xpylT, pylS

The *Methanosarcina mazei pylS* gene was digested using the restriction enzymes *ApaL*I and *Not*I. The 2.2 kb fragment was inserted into pACYC-Duet1 (*Novagen*) using the same sites. The following custom designed DNA-fragment (*GeneArt*) encoding for the *E.coli* pGLN-promoter and a Shine Dalgarno sequence was inserted into this plasmid using the restriction sites *Cla*I/*Nco*I: 5'-TCATCAATCATCCCCATAATCCTTGTTAGATT ATCAATTTTAAAAAACTAACAGTTGTCAGCCTGTCCCGCTTATAATATCATACGCCGTTATACGTTGTT TACGCTTTGAGGAAGCC-3'. The resulting construct pACYC-pGLN was subsequently digested with *Not*I and *Pac*I to insert a custom designed DNA-fragment (*GeneArt*) encoding for three copies of *pylT*, which are flanked by the *E.coli proK* promoter and terminator, respectively. Sequence of this fragment is as follows, with the *proK*-promoter and -terminator highlighted in bold letters, and *pylT* being underlined: 5'-GCGGCCTGCTGACTTTCTCGCCGATCAAAAGGCATTTTGCTATT**AAGGGATTGACGAGGGCGTATCT GCGCAGTAAGATGCGCC**CCGCATTGGAAACCTGATCATGTAGATCGAATGGACTCTAAATCCGTTCA GCCGGGTTAGATTCCCGGGGTTTCCGCCATACATGTTATGGAAACCTGATCATGTAGATCGAATGGA CTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGCCATACATGTTATGGAAACCTGATCAT GTAGATCGAATGGACTCTAAATCCGTTCAGCCGGGTTAGATTCCCGGGGTTTCCGCCA**AATTCGAAA AGCCTGCTCAACGAGCAGGCTTTTTTGC**CTTAAG-3'.

3.3 pPSG IBA3_trxA N65amber

The modified *E. coli* gene *trx*A was cloned by using the *IBA* Stargate System into pPSG IBA3. The gene was amplified from the vector pBAD202 (*Invitrogen*) by using the following primers:

Forward: **AGCGGCTCTTC** AATG GGA TCT GAT AAA ATT ATT CAT CTG ACT GAT G

Reverse: **AGCGGCTCTTC** TCCC CAG GTT AGC GTC GAG GAA CTC TTT C

The PCR-product was inserted into the entry vector IBA51 following the manual of *IBA*. A second cloning step transferred the gene into the destination vector pPSG IBA3. This vector contains an IPTG inducible T7 promotor and a C-terminal Strep-Tag.

The amber codon at position N65 was inserted by blunt-end mutagenesis, using the 5'-phosphorylated primers:

Forward: CCG GGC ACT GCG CCG AAA TAT G

Reverse: CTA GTG ATC GAT GTT CAG TTT TGC AAC GG

3.4 pPSG IBA35_h3 K9amber

The human synthetic gene h3, coding for the histone H3.3 was delivered by *GeneArt*, codon optimized for expression in *E. coli*. The gene was amplified by using the following primers:

Forward: **AGCGGCTCTTC** AATG GCA CGT ACC AAA CAG ACC GCA CGT AAA AG

Forward (+TAG): **AGCGGCTCTTC** AATG GCA CGT ACC AAA CAG ACC GCA CGT TAG AGC ACC G

Reverse: **AGCGGCTCTTC** TCCC TGC ACG TTC ACC ACG AAT ACG ACG TGC

The PCR-product was inserted into the entry vector IBA51 following the manual of *IBA*. A second cloning step transferred the gene into the destination vector pPSG IBA35. This vector contains an IPTG inducible T7 promotor and a N-terminal His_6-Tag.

The amber codon at position K9 was inserted by using a modified Forward primer for the entry vector generation:

Forward (+TAG): **AGCGGCTCTTC** AATG GCA CGT ACC AAA CAG ACC GCA CGT TAG AGC ACC G

4 Testing the incorporation of Kpr, Kbu and Kcr into YFP

Four different *M. mazei* PylRS synthetase variants were tested for their acceptance of the modified lysine amino acids: - wt

- Y306A V401K
- Y306G Y384F I405R
- Y306A Y384F

4.1 Expression of YFP K144Kpr, K144Kbu and K144Kcr

E. coli strain BL21(DE3) (*NEB*) (*fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10*--Tet^S*)2 [dcm] R(zgb-210::Tn10*--Tet^S*) endA1* Δ*(mcrC-mrr)114::IS10)* was used for the production of Kpr, Kbu or Kcr containing YFP. Cells were grown in 50 mL LB-medium containing 5 mM of Kbu or Kcr or 10 mM Kpr (37°C, 220 rpm) and 50 mg/L carbenicillin and 34 mg/L Chloramphenicol. When OD_{600}=0.6 was reached, expression of yfp_{TAG} was induced by addition of 1 mM IPTG. After further 16 h at 30 °C YFP expression was analyzed by fluorescence microscopy.

4.2 Detection of YFP K144Kpr, K144Kbu and K144Kcr

The YFP fluorescence was analyzed by microscopy on a *TCS SPE* spectral confocal microscope (*Leica*) with an inverse stand and the HCX FL APO 63x/140-0.60 oil objective (excitation: 488 nm). As shown in Fig. S4.1 the best YFP-read through results for all three acylated lysine derivatives were obtained with the wt-PylRS. Therefore this variant was used for the incorporation of the amino acids into H3.

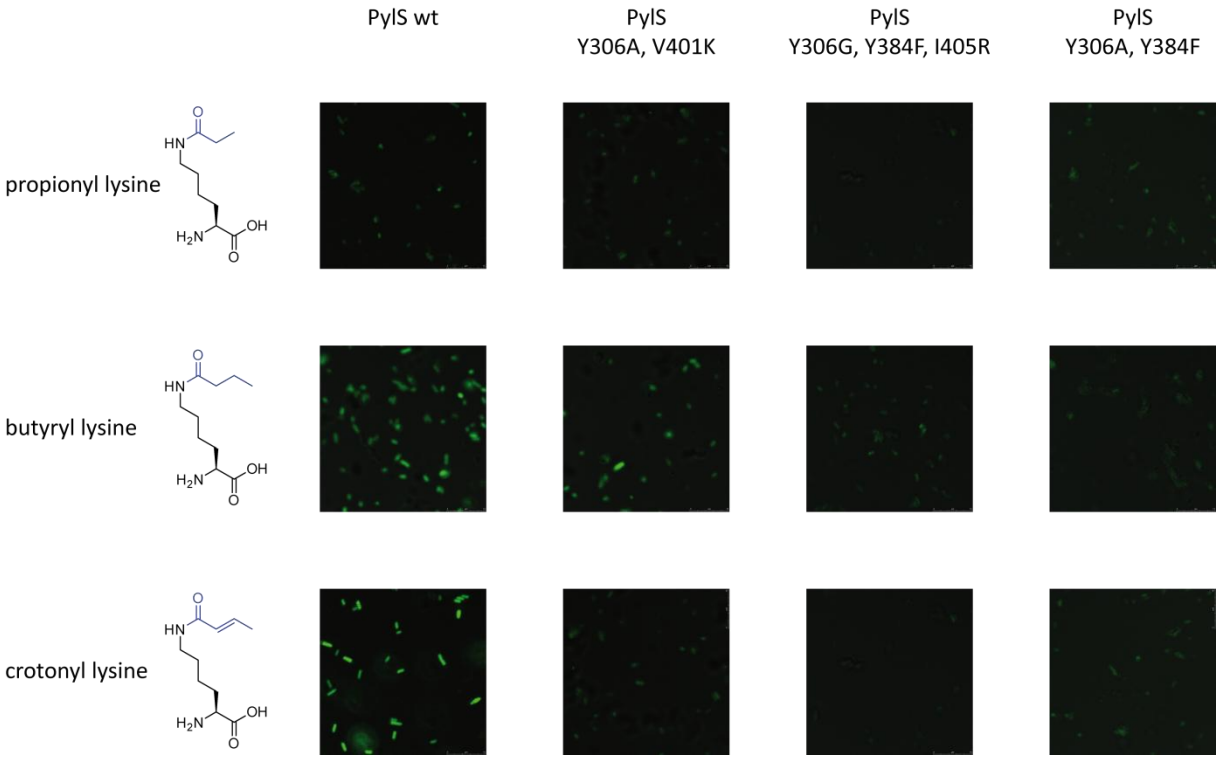


Fig. S4.1. Fluorescence microscopy of E. coli transformed with the pETDuet_YFPK114amber reporter system and addition of the three different lysine derivatives Kpr, Kbu and Kcr

5 Incorporation of Kcr into TrxA

5.1 Expression of TrxA N65Kcr

E. coli strain BL21(DE3) were transformed with pACyc_pylTTT, pylS and pPSG IBA3_trxA N65amber. An overnight culture was used for inoculation (to OD_{600} of 0.1) of 2L LB medium containing 5mM Kcr, 50 mg/L carbenicillin and 34 mg/L chloramphenicol. Cells were grown at 37 °C until an OD_{600} of 0.6. Expression of trxA N65Kcr was induced by addition 1 mM IPTG. The cells were shaked for further 16 h at 30 °C. After centrifugation (10.000 x g, 10 min, 4°C) the cells were stored at -20°C or directly purified.

5.2 Purification of TrxA N65Kcr

All purification steps were carried out at 4°C. Cells were resuspended in StrepA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl) and supplemented with protease inhibitor mix (*Roche*). The cells were lysed using a French press (*Thermo Scientific*). Cell debris was removed by centrifugation (38 000 x g, 30 min, 4 °C) and the supernatant was applied to a 5 mL Strep-Tactin column (*IBA*), equilibrated with StrepA buffer. Proteins were eluted from the column using the same buffer containing 2.5 mM desthiobiotin. The eluted fractions were pooled and concentrated. Typical protein yields of 2 mg per liter expression medium Kcr containing TrxA were achieved. Purified TrxA was equilibrated with StrepA buffer, concentrated and stored at -80°C.

5.3 Intact MS of TrxA N65Kcr

The purified intact protein TrxA N65Kcr was analyzed by HPLC-MS. A self-packed nano C4 column was used for HPLC and mass spectrometry was performed on a LTQ-Orbitrap XL mass spectrometer (*Thermo*). Fig. S5.1 shows the raw mass spectrum. Fig. S5.2 depicts the deconvoluted spectrum. The calculated mass of the protein is 13015.9 Da, the observed mass is 13013.8 Da. Unspecific deacylation of Kcr was not observed at the intact protein level.

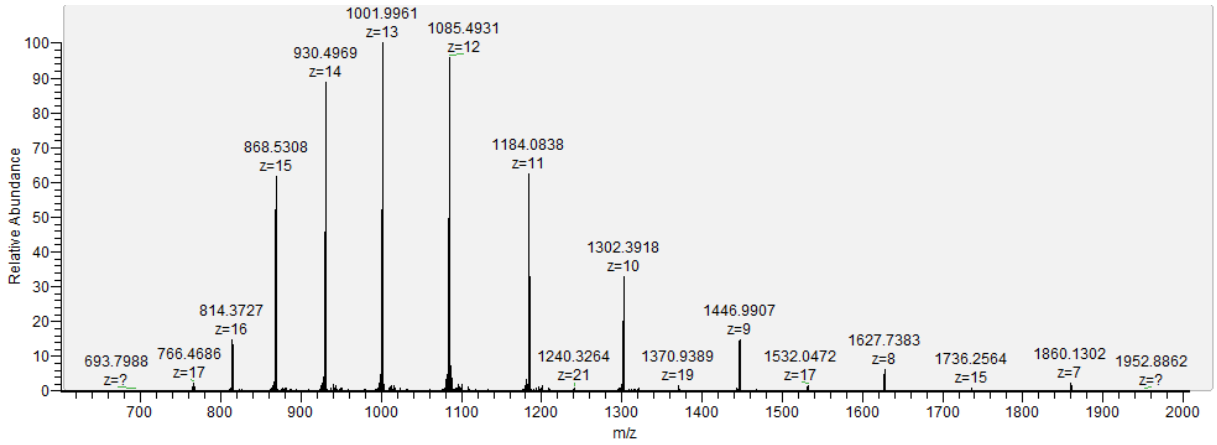


Fig. S5.1. Raw intact mass spectrum of TrxA N65Kcr.

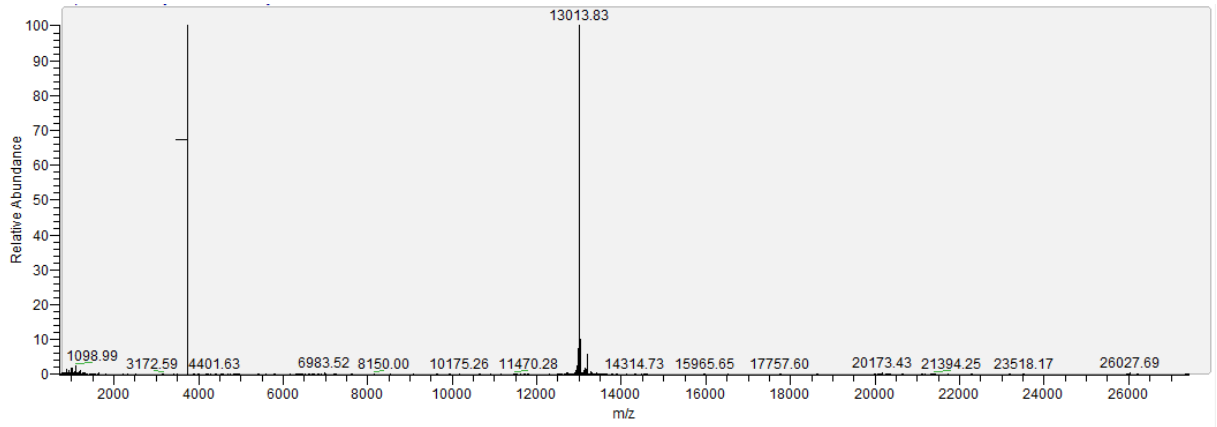


Fig. S5.2. Deconvoluted mass spectrum of TrxA N65Kcr.

5.4 In-Gel chymotryptic digestion of TrxA N65Kcr

The Coomassie Blue stained SDS-gel was washed twice for 10 min with water (bidest.). Protein bands were excised from the gel, cut in small cubes (ca. $1mm^3$) and transferred to *Eppendorf* tubes. The gel slices were washed by incubation in water (100 µL) for 15 min at RT. Then 50 mM NH_4HCO_3/acetonitrile (each 100 µL) was added and incubated for additional 15 min. All liquids were removed and the gel cubes covered with acetonitrile (100 µL). After removal of acetonitrile the gel particles were rehydrated with 50 mM NH_4HCO_3 (100 µL) for 5 min. This was followed by further addition of 100 µL acetonitrile and incubation for 15 min. The liquids were removed and again 100 µL acetonitrile were added to shrink the particles. After 5 min all acetonitrile was removed. The gel cubes were dried at RT for 10 min. Protein denaturation and reduction were carried out by adding DTT (10 mM in 50 mM NH_4HCO_3, 100 µL) and incubation for 45 min at 56 °C. The tubes were cooled to RT, all liquids were removed and alkylation was performed by addition of 100 µL iodoacetamide (55 mM in 50

mM NH_4HCO_3) and incubation for 30 min at RT (in the dark). Further washing steps (15 min at RT with 50 mM NH_4HCO_3/acetonitrile, each 100 µL, twice) and addition of acetonitrile (100 µL) resulted in shrinkage of the gel cubes which was completed by removal of all liquids and drying in a vacuum centrifuge (10 min). In-gel digestion was carried out by rehydration of the gel particles in 100 µL 25 mM NH_4HCO_3 containing 0.5 µg chymotrypsin (*Promega*) and incubation at 37 °C overnight. The supernatant was removed and stored separately. For the extraction of the peptides the gel particles were resuspended in 100 µL 25 mM NH_4HCO_3 and sonicated for 15 min before addition of 100 µL acetonitrile and further sonication for 15 min. The supernatant was again removed and stored separately. For quantitative peptide extraction the gel slices were further sonicated with 5% formic acid and acetonitrile (each 100 µL for 15 min). Finally all peptide solutions were combined and concentrated in a vacuum centrifuge. The samples were used for peptide mass analysis after an extended centrifugation step (10 min, 13000 rpm).

5.5 HPLC-MS/MS of chymotryptic digestion of TrxA N65Kcr

For the nano-HPLC ESI mass spectrometry analysis the tryptic digested peptides were loaded onto a *Dionex* C18 Nano Trap Column (100 µm), subsequently eluted and separated by a self packed C18-AQ, 3 µm, 120 Å column. The peptide fragments were analyzed by tandem MS followed by HR-MS using a coupled *Dionex* Ultimate 3000 LTQ-Orbitrap XL MS system (*Thermo*).

The obtained peptide fragment data were analyzed with the SEQUEST algorithm implemented in the software "*Proteome discoverer 1.10.263*" (*Thermo*) against a peptide library of *E. coli* containing the TrxA variant. The search was limited to chymotryptic peptides, two missed cleavage sites, +2 charged monoisotopic precursor ions and a peptide tolerance of <10 ppm. Sequence coverage and probability values of the chymotryptic TrxA N65Kcr digestion sample is shown in Table S5.1.

Table S5.1. Sequence coverage and probability values of the chymotryptic TrxA N65Kcr digestion.

Sample	**Sequence coverage**	**score**
TrxA N65Kcr	69.2%	362.3

Table S5.2 sums up the MS/MS evaluation data of the chymotryptic peptide NIDHXPGTAPKY containing either Kcr or K at position X. The amount of deacylation was calculated from the

integrals of the mass signals based on the fact that the ionization properties of the peptides are almost similar.

Table S5.2. MS/MS evaluation data of the chymotryptic peptide NIDHXPGTAPKY of the TrxA N65Kcr digestion.

Sample	$[M+2H]^{2+}$	ΔM	Probability	Ions found
TrxA N65Kcr	704.8649	+ 1.24 ppm	25.78	14/22
TrxA N65K (deacylated)	670.8529	+ 2.06 ppm	32.52	16/22

The MS/MS spectrum of the described $[M+2H]^{2+}$ ions is depicted in Figure S5.3. Sequencing data (MS/MS data) of the relevant peptide fragment NIDHXPGTAPKY of Kcr containing TrxA (X=Kcr, Table S5.3) proof the insertion of the amino acid Kcr at the correct position (N65Kcr). Fig. S5.4 and table S5.4 show the relevant peptide fragments of deacylated Trx N65Kcr (Trx N65K). As fragmentation of the peptide can occur starting either with the N-terminus (b ion series) or the C-terminus (y ion series) both ion series can always be found.

TrxA N65Kcr

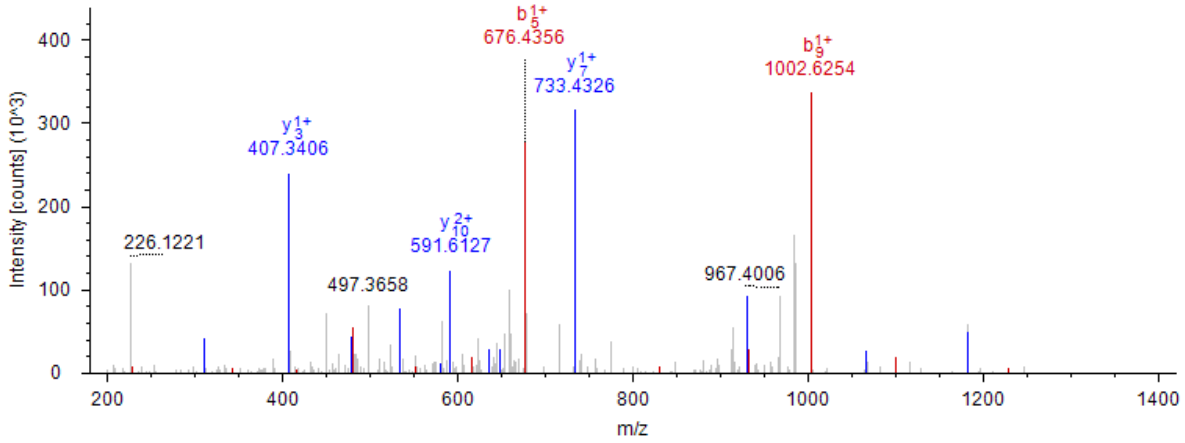


Fig. S5.3. MS/MS spectrum of the chymotryptic peptide NIDHXPGTAPKY (X = Kcr) of the TrxA N65Kcr digestion.

Table S5.3. Assigned MS/MS fragments resulting from fragmentation of the chymotryptic peptide NIDHXPGTAPKY (X = Kcr) of the TrxA N65Kcr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	115.05021	58.02874	N			12
2	228.13428	114.57078	I	1294.67900	647.84314	11
3	343.16123	172.08425	D	1181.59493	591.30110	10
4	480.22014	240.61371	H	1066.56798	533.78763	9
5	676.34127	338.67427	N-KCr	929.50907	465.25817	8
6	773.39404	387.20066	P	733.38794	367.19761	7
7	830.41551	415.71139	G	636.33517	318.67122	6
8	931.46319	466.23523	T	579.31370	290.16049	5
9	1002.50031	501.75379	A	478.26602	239.63665	4
10	1099.55308	550.28018	P	407.22890	204.11809	3
11	1227.64805	614.32766	K	310.17613	155.59170	2
12			Y	182.08116	91.54422	1

TrxA N65K (deacylated Kcr)

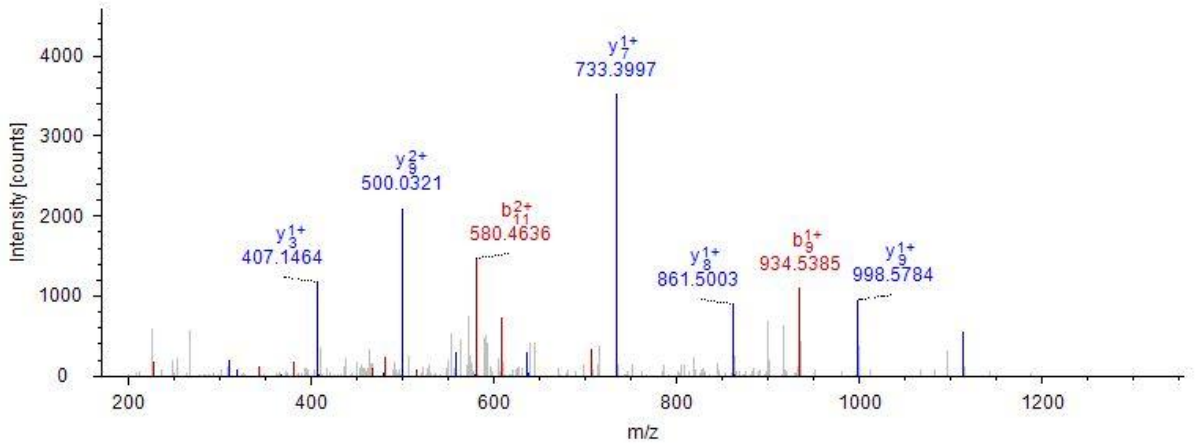


Fig. S5.4. MS/MS spectrum of the deacylated chymotryptic peptide NIDHXPGTAPKY (X = K) of the TrxA N65Kcr digestion.

Table S5.4. Assigned MS/MS fragments resulting from fragmentation of the deacylated chymotryptic peptide NIDHXPGTAPKY (X = K) of the TrxA N65Kcr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	115.05021	58.02874	N			12
2	228.13428	114.57078	I	1226.65284	613.83006	11
3	343.16123	172.08425	D	1113.56877	557.28802	10
4	480.22014	240.61371	H	998.54182	499.77455	9
5	608.31510	304.66119	N-Asn->Lys	861.48291	431.24509	8
6	705.36787	353.18757	P	733.38794	367.19761	7
7	762.38934	381.69831	G	636.33517	318.67122	6
8	863.43702	432.22215	T	579.31370	290.16049	5
9	934.47414	467.74071	A	478.26602	239.63665	4
10	1031.52691	516.26709	P	407.22890	204.11809	3
11	1159.62188	580.31458	K	310.17613	155.59170	2
12			Y	182.08116	91.54422	1

6 Incorporation of Kpr, Kbu and Kcr into H3

6.1 Expression of H3 wt, K9Kpr, K9Kbu, K9Kcr

E. coli strain BL21(DE3) was transformed with pACyc_pylTTT, pylS and pPSG IBA35_h3 K9amber. An overnight culture was used for inoculation (to OD_{600} of 0.1) of 0.5 L LB medium containing 5 mM Kcr or Kbu or 10 mM Kpr, 50 mg/L carbenicillin and 34 mg/L chloramphenicol. Cells were grown at 37 °C until an OD_{600} of 0.6. Expression of h3 K9Kmod was induced by addition 1 mM IPTG. Optionally 20 mM nicotinamide (NAM) were added at this point to inhibit a possible deacylation of the incorporated modified amino acids. The cells were shaked for further 16 h at 30 °C. After centrifugation (10.000 x g, 10 min, 4°C) the cells were stored at -20°C or directly purified.

6.2 Purification of H3 wt, K9Kpr, K9Kbu, K9Kcr

All purification steps were carried out at room temperature. Cells were resuspended in HisA buffer (50 mM Tris-HCl, 500 mM NaCl, 20 mM imidazol, pH 7.5) supplemented with 7 M urea. The cells were lysed using a French press (*Thermo Scientific*). Cell debris was removed by centrifugation (38 000 x g, 30 min, rt) and the supernatant was applied to a 1 mL HisTrap FF column (*GE*), equilibrated with HisA buffer. Proteins were eluted from the column using the same buffer containing 250 mM imidazol. The eluted fractions were pooled and concentrated. Typical protein yields of 1 mg per liter expression medium Kcr or Kbu containing H3 were achieved. The yield of Kpr containing H3 was typically lower at 0.5 mg/L expression medium. H3 wt protein yield was at 4-5 mg/L. Purified H3 was optionally dialyzed twice (2 h, than over night) against 2 L StrepA buffer (10 mM Tris-HCl pH 8, 1 mM EDTA, 150 mM NaCl) supplemented with 5 mM DTT at 4 °C. concentrated and stored at -80°C.

6.3 In-Gel Propionylation and butyrylation and tryptic digestion of H3 variants H3 wt, K9Kpr, K9Kbu, K9Kcr

The first washing steps of this procedure were performed in the same way as for the chymotryptic digestion described in 5.5.

Instead of the reduction and alkylation, the capping of unmodified lysine residues was carried out[2]. After this procedure trypsin cleaves exclusively after arginine residues. Consequently longer peptides are generated from the digestion which can be analyzed.

Another advantage is the possible comparison of originally modified with deacylated peptides.

The dried gel pieces were incubated with 20 µL of propionic or butyric anhydride and 40 µL of 2 M ammonium hydroxide for 30 min at 51 °C. This step was repeated once. Afterwards the gel pieces were washed as described in 5.5. The tryptic digestion was performed by adding 100 µL of a 5 ng/µL trypsin (*Promega*) solution in 25 mM ammonium hydrogencarbonate. The digestion solution was incubated over night at 37 °C. Afterwards the capping procedure was repeated twice as described after the digestion to also cap the N-termini generated during digestion.

The peptides were extracted from the gel pieces and analyzed by HPLC-MS/MS in the same way as described for the chymotryptic peptides in 5.5.

6.4 HPLC-MS/MS of propionylated or butyrylated tryptic peptides of H3 variants H3 wt, K9Kpr, K9Kbu, K9Kcr

For the nano-HPLC ESI mass spectrometry analysis the tryptic digested peptides were loaded onto a *Dionex* C18 Nano Trap Column (100 µm), subsequently eluted and separated by a self packed C18-AQ, 3 µm, 120 Å column. The peptide fragments were analyzed by HR-MS followed by tandem MS using a coupled *Dionex* Ultimate 3000 LTQ-Orbitrap XL MS system (*Thermo*).

The obtained peptide fragment data were analyzed with the SEQUEST algorithm implemented in the software "*Proteome discoverer 1.1.0.263*" (*Thermo*) against a peptide library of *E. coli* containing the protein sequence of human histone H3.3. The search was limited to tryptic peptides, two missed cleavage sites, +2 charged monoisotopic precursor ions and a peptide tolerance of <10 ppm. The sequence coverage and probability values of the capped tryptic H3 wt, H3 K9Kpr, Kbu and Kcr digestion samples are shown in Table S6.1.

Table S6.1. Sequence coverage and probability values of the capped tryptic H3 K9Kpr, Kbu and Kcr digestion samples.

Sample	capping	Sequence coverage	score
H3 wt	propionylation	62.05%	73.51
H3 wt	butyrylation	61.03%	52.72
H3 K9Kpr	butyrylation	54.41%	105.58
H3 K9Kbu	propionylation	60.29%	47.37
H3 K9Kcr	propionylation	48.53%	21.44

Table S6.2 sums up the MS/MS evaluation data of the tryptic peptide XSTGGKAPR containing either K (capped), Kpr, Kbu or Kcr at position X. The amount of deacylation mentioned in the main paper was calculated from the integrals of the mass signals based on the fact that the ionization properties of the peptides are almost similar.

Table S6.2. MS/MS evaluation data of the tryptic peptide XSTGGKAPR (X = K, Kpr, Kbu or Kcr) of the capped H3 digestion.

Sample	capping	$[M+2H]^{2+}$	ΔM	Probability	Ions found
H3 wt	propionylation	535.3057	3.86 ppm	45.19	14/16
H3 wt	butyrylation	556.3284	2.09 ppm	24.71	14/16
H3 K9Kpr	butyrylation	549.3199	1.07 ppm	10.75	13/16
H3 K9Kbu	propionylation	542.3140	4.65 ppm	17.23	13/16
H3 K9Kcr	propionylation	541.3046	1.79 ppm	21.44	12/16

MS Spectra of the described $[M+2H]^{2+}$ ions are depicted in Figures S6.1-S6.5. Sequencing data (MS/MS data) of the relevant peptide fragment XSTGGKAPR of modified lysine containing H3 (X = Kpr, Kbu or Kcr, Table S6.3-S6.7) proof the insertion of the modified amino acids at the correct position in H3 (K9). As fragmentation of the peptide can occur starting either with the N-terminus (b ion series) or the C-terminus (y ion series) both ion series can always be found.

H3 wt propionylated

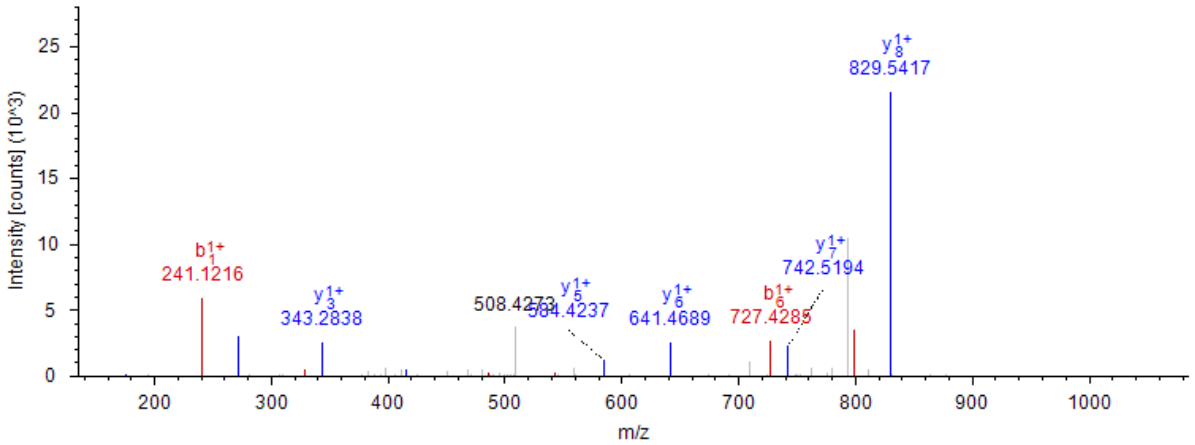


Fig. S6.1. MS/MS spectrum of the propionyl capped tryptic peptide XSTGGKAPR (X = K, capped by propionylation) of the H3 wt digestion.

Table S6.3. Assigned MS/MS fragments resulting from fragmentation of the propionyl capped tryptic peptide XSTGGKAPR (X = K, capped by propionylation) of the H3 wt digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	241.15465	121.08096	K-Propionyl N-term-Propionyl			9
2	328.18668	164.59698	S	829.45267	415.22997	8
3	429.23436	215.12082	T	742.42064	371.71396	7
4	486.25583	243.63155	G	641.37296	321.19012	6
5	543.27730	272.14229	G	584.35149	292.67938	5
6	727.39847	364.20287	K-Propionyl	527.33002	264.16865	4
7	798.43559	399.72143	A	343.20885	172.10806	3
8	895.48836	448.24782	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

H3 wt butyrylated

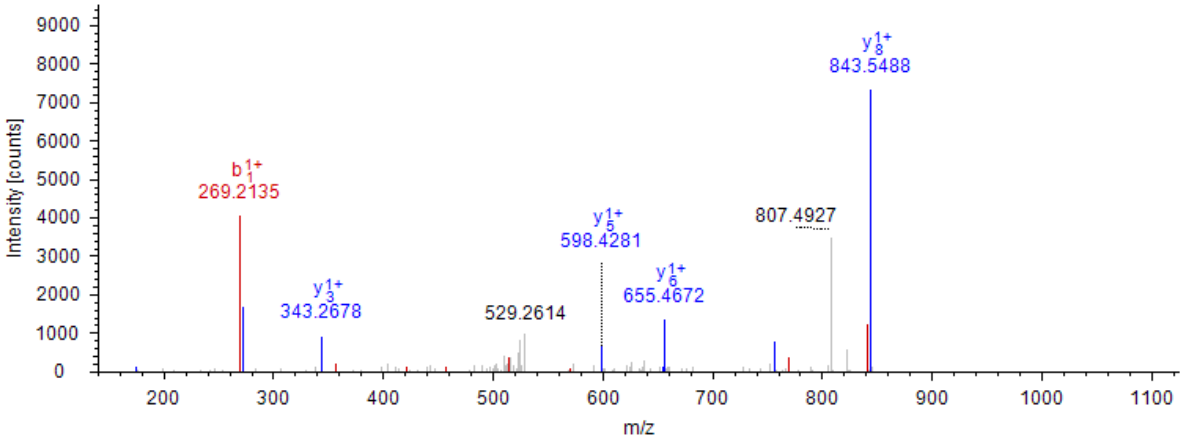


Fig. S6.2. MS/MS spectrum of the butyryl capped tryptic peptide XSTGGKAPR (X = K, capped by butyrylation) of the H3 wt digestion.

Table S6.4. Assigned MS/MS fragments resulting from fragmentation of the butyryl capped tryptic peptide XSTGGKAPR (X = K, capped by butyrylation) of the H3 wt digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	269.18605	135.09666	K-Butyryl N-term-Butyryl			9
2	356.21808	178.61268	S	843.46837	422.23782	8
3	457.26576	229.13652	T	756.43634	378.72181	7
4	514.28723	257.64725	G	655.38866	328.19797	6
5	571.30870	286.15799	G	598.36719	299.68723	5
6	769.44557	385.22642	K-Butyryl	541.34572	271.17650	4
7	840.48269	420.74498	A	343.20885	172.10806	3
8	937.53546	469.27137	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

H3 K9Kpr butyrylated

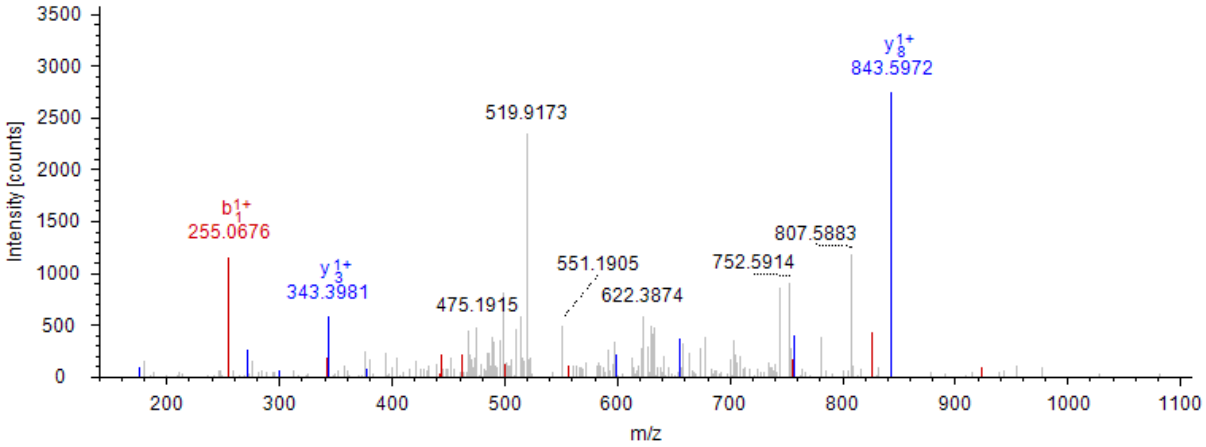


Fig. S6.3. MS/MS spectrum of the butyryl capped tryptic peptide XSTGGKAPR (X = Kpr) of the H3 K9Kpr digestion.

Table S6.5. Assigned MS/MS fragments resulting from fragmentation of the butyryl capped tryptic peptide XSTGGKAPR (X = Kpr) of the H3 K9Kpr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	255.17035	128.08881	K-Butyryl N-term-Propionyl			9
2	342.20238	171.60483	S	843.46837	422.23782	8
3	443.25006	222.12867	T	756.43634	378.72181	7
4	500.27153	250.63940	G	655.38866	328.19797	6
5	557.29300	279.15014	G	598.36719	299.68723	5
6	755.42987	378.21857	K-Butyryl	541.34572	271.17650	4
7	826.46699	413.73713	A	343.20885	172.10806	3
8	923.51976	462.26352	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

H3 K9Kbu propionylated

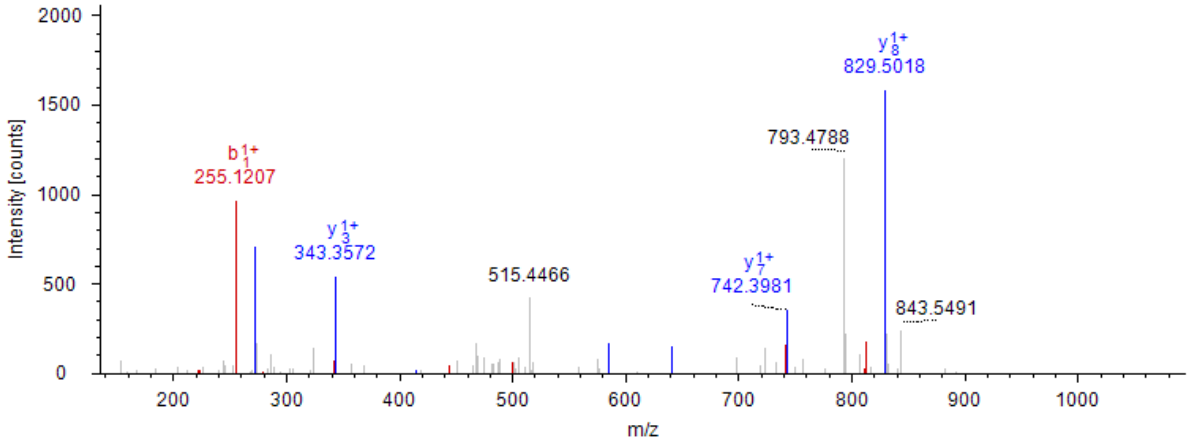


Fig. S6.4. MS/MS spectrum of the propionyl capped tryptic peptide XSTGGKAPR (X = Kbu) of the H3 K9Kbu digestion.

Table S6.6. Assigned MS/MS fragments resulting from fragmentation of the propionyl capped tryptic peptide XSTGGKAPR (X = Kbu) of the H3 K9Kbu digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	255.17035	128.08881	K- Butyryl N-term- Propionyl			9
2	342.20238	171.60483	S	829.45267	415.22997	8
3	443.25006	222.12867	T	742.42064	371.71396	7
4	500.27153	250.63940	G	641.37296	321.19012	6
5	557.29300	279.15014	G	584.35149	292.67938	5
6	741.41417	371.21072	K-Propionyl	527.33002	264.16865	4
7	812.45129	406.72928	A	343.20885	172.10806	3
8	909.50406	455.25567	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

H3 K9Kcr propionylated

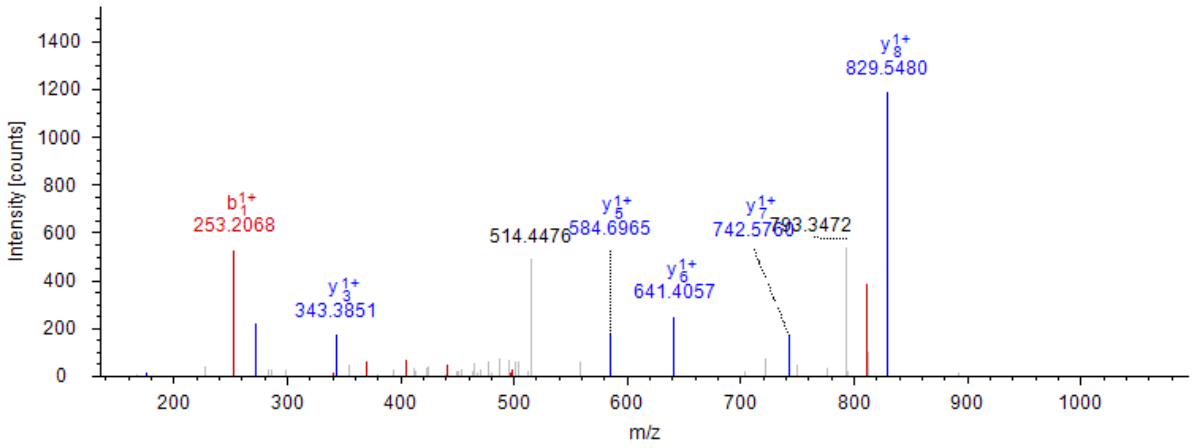


Fig. S6.5. MS/MS spectrum of the propionyl capped tryptic peptide XSTGGKAPR (X = Kcr) of the H3 K9Kcr digestion.

Table S6.7. Assigned MS/MS fragments resulting from fragmentation of the propionyl capped tryptic peptide XSTGGKAPR (X = Kcr) of the H3 K9Kcr digestion.

#1	b(1+)	b(2+)	Seq.	y(1+)	y(2+)	#2
1	253.15465	127.08096	K-Crotonyl N-term-Propionyl			9
2	340.18668	170.59698	S	829.45267	415.22997	8
3	441.23436	221.12082	T	742.42064	371.71396	7
4	498.25583	249.63155	G	641.37296	321.19012	6
5	555.27730	278.14229	G	584.35149	292.67938	5
6	739.39847	370.20287	K-Propionyl	527.33002	264.16865	4
7	810.43559	405.72143	A	343.20885	172.10806	3
8	907.48836	454.24782	P	272.17173	136.58950	2
9			R	175.11896	88.06312	1

7 Western Blot

The antibodies against Kpr, Kbu and Kcr were purchased from *PTM biolabs*. The western blot was performed following the procedures recommended by *PTM biolabs*. As shown in Figure S7.1 the H3 K9Kpr, H3 K9Kbu and H3 K9Kcr histone modifications can be selectively labeled.using the corresponding antibodies when compared to the wild type histone (H3 K9). In a cross experiment every antibody was tested against every histone modification (H3 K9Kpr, H3 K9Kbu and H3 K9Kcr). This experiment revealed that the antibodies also recognize the related modifications (e.g. H3 K9Kpr is recognized also by anti Kbu antibody). This cross labeling activity could be a result of the high concentrations of purified histones used for the western blot. The results are summed up in Fig. S7.2.

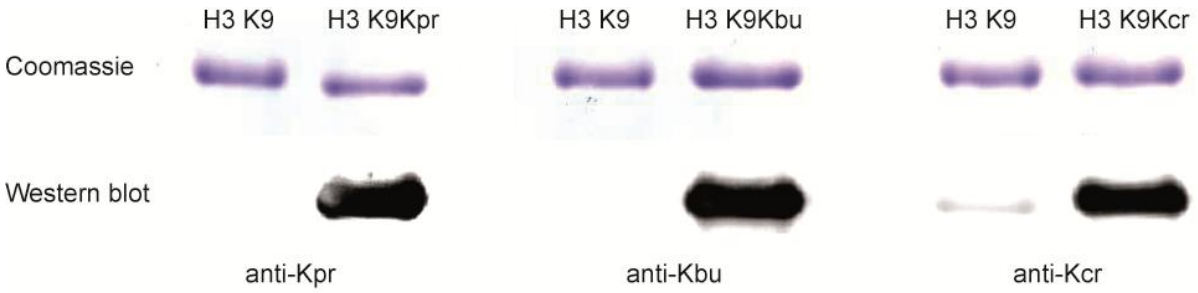


Fig. S7.1. Western Blot of the three different antibodies against different modifications

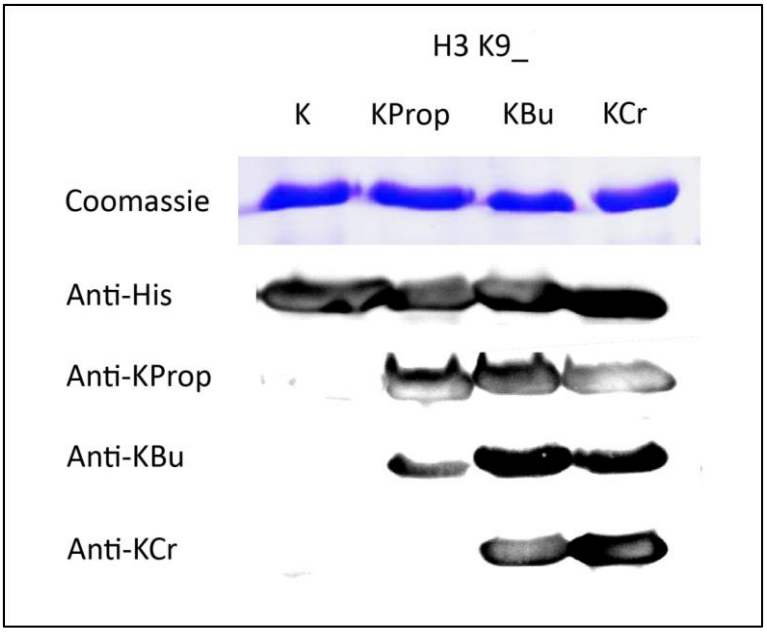
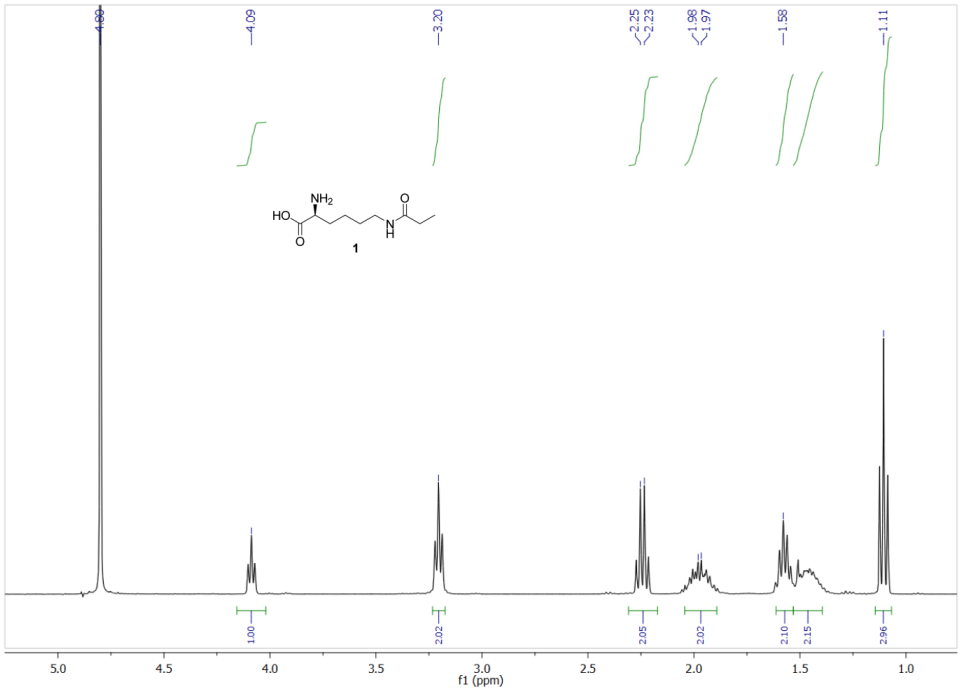


Fig. S7.2. Western Blot of the three different antibodies against different modifications

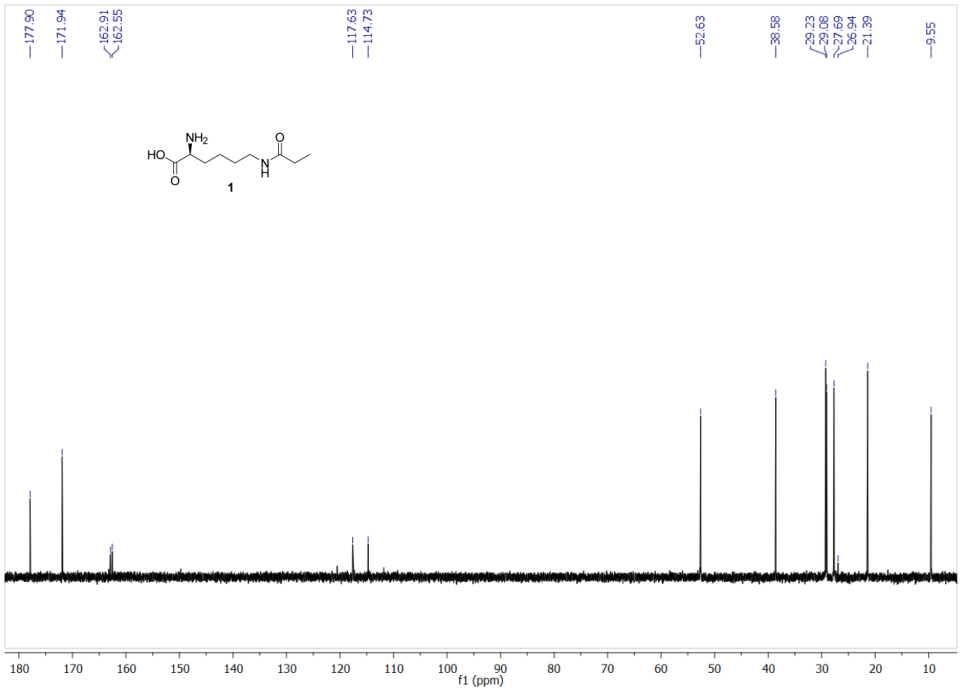
8 Appendix

8.1 NMR spectra

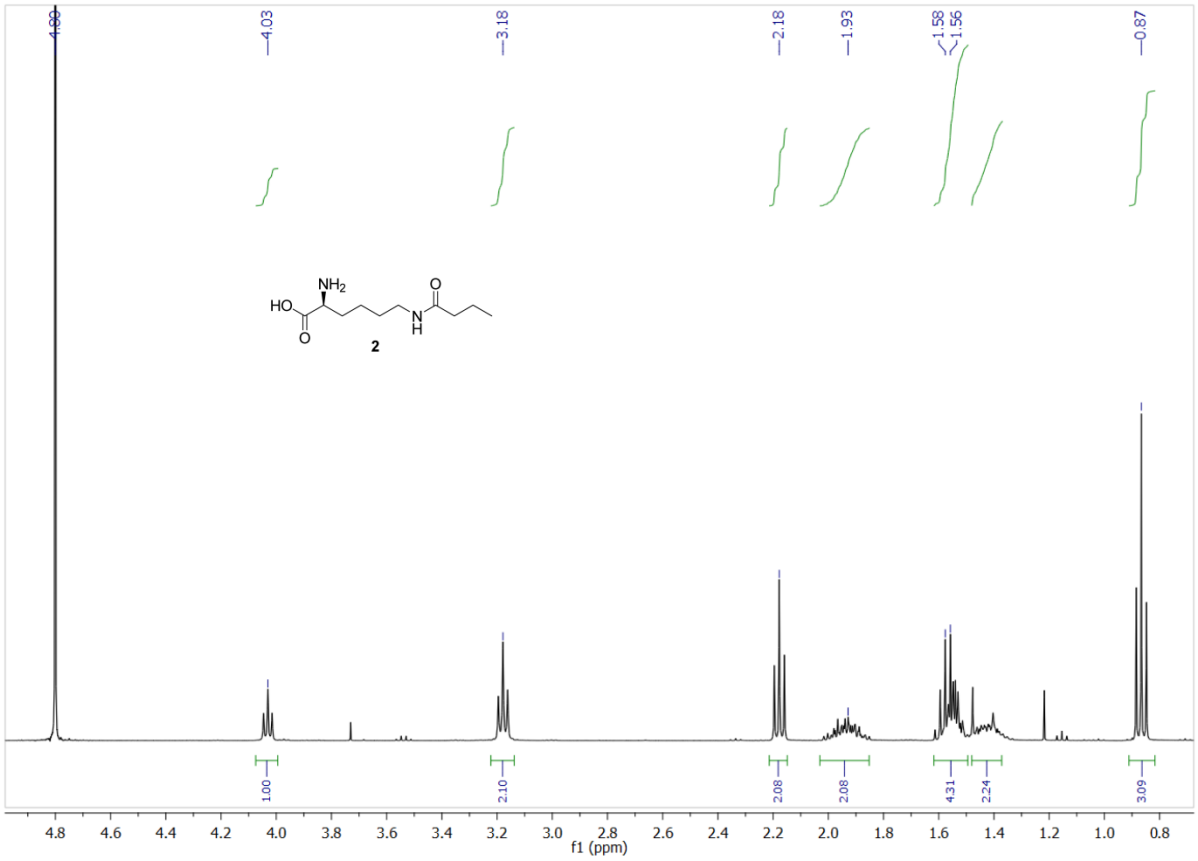
^{1}H NMR of compound **1**



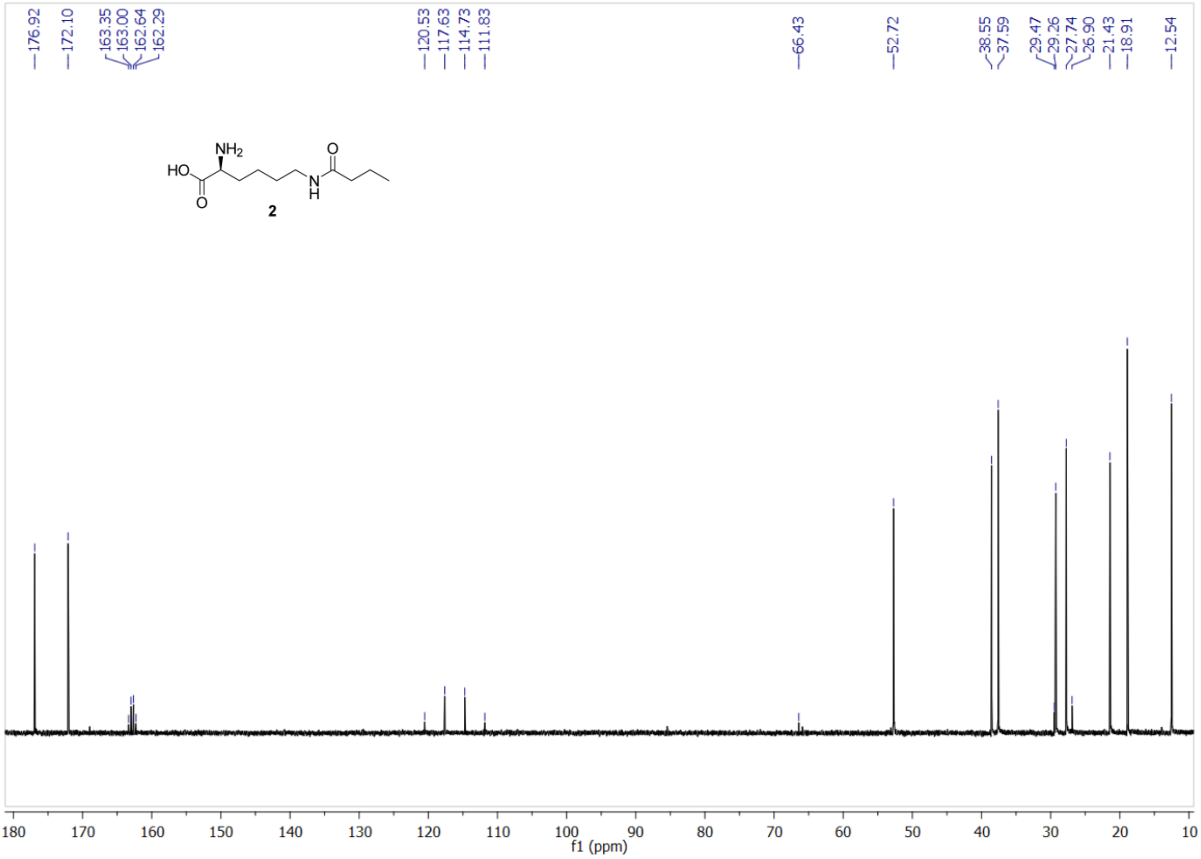
^{13}C NMR of compound **1**



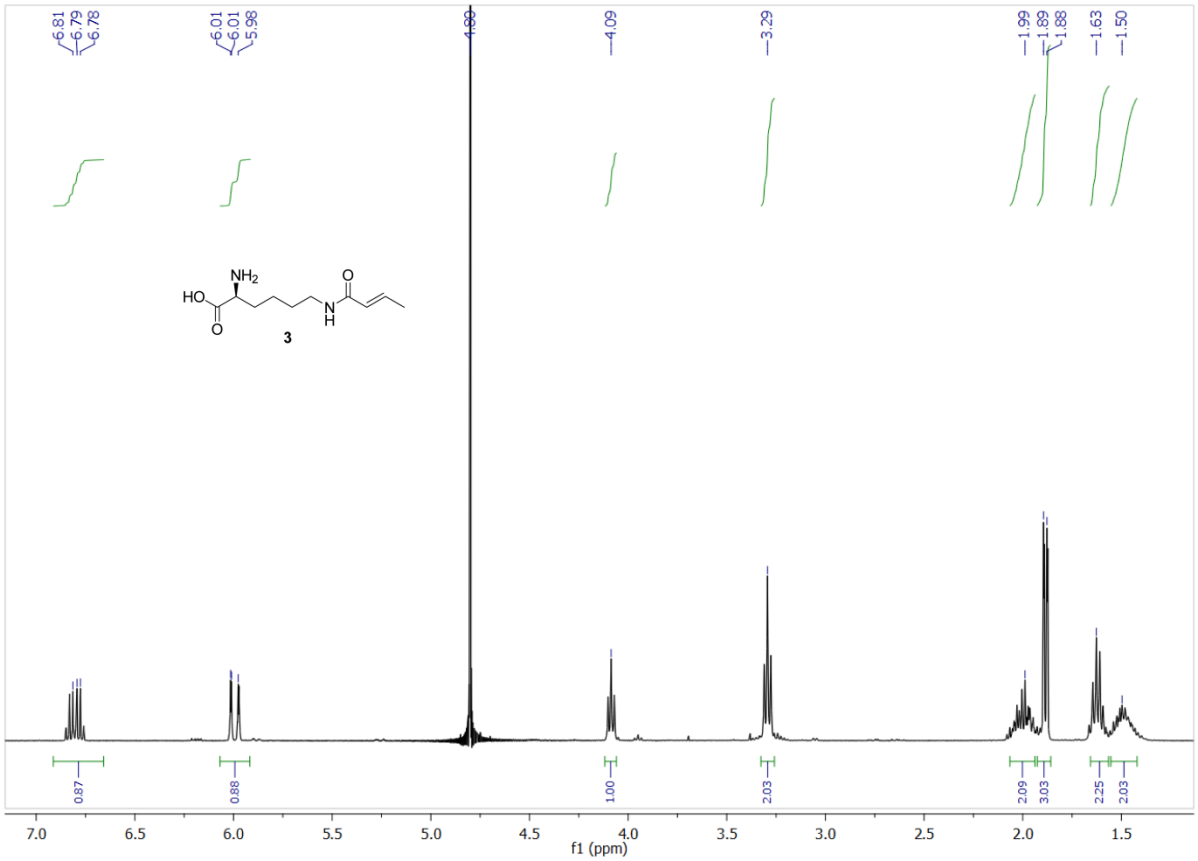
^{1}H NMR of compound **2**



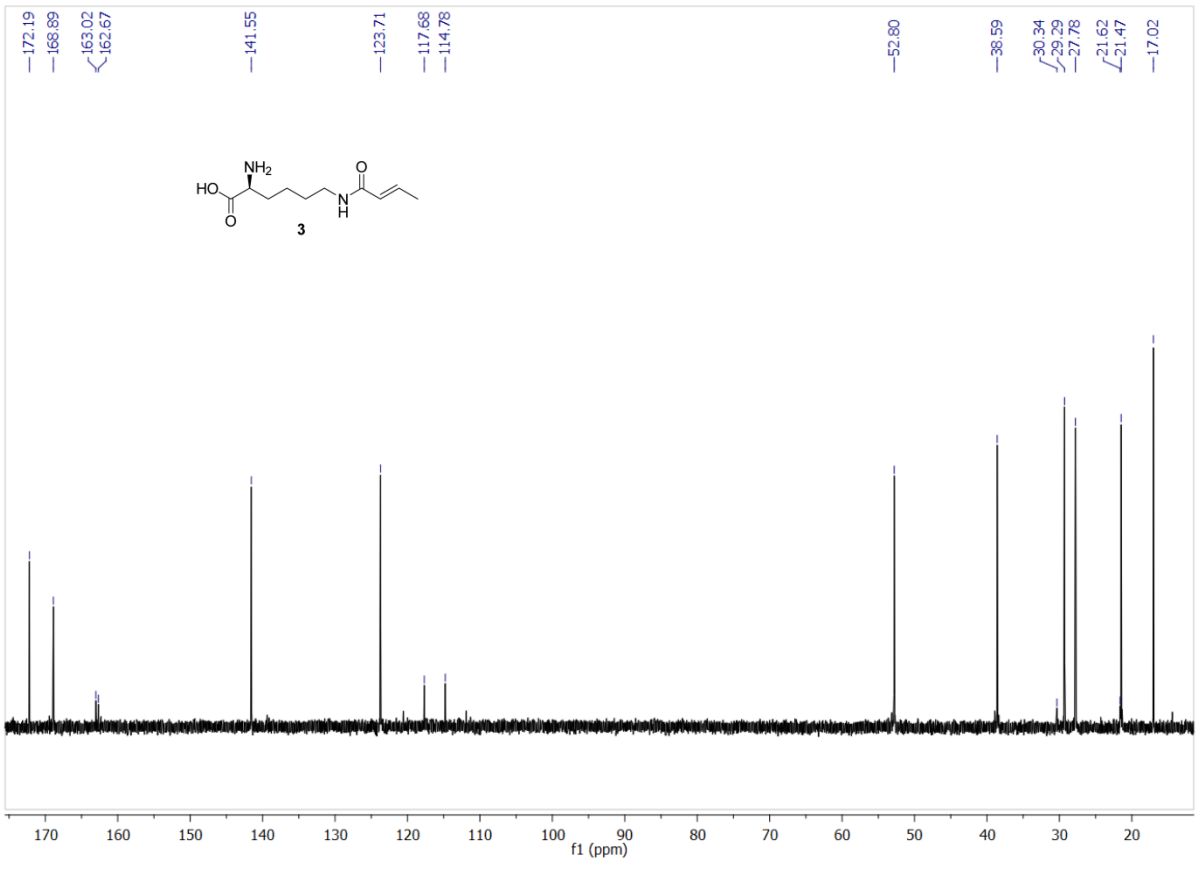
^{13}C NMR of compound **2**



^{1}H NMR of compound **3**



^{13}C NMR of compound **3**



9 Literature

[1] E. Kaya, K. Gutsmiedl, M. Vrabel, M. Muller, P. Thumbs, T. Carell, *ChemBioChem* **2009**, *10*, 2858-2861.

[2] B. A. Garcia, S. Mollah, B. M. Ueberheide, S. A. Busby, T. L. Muratore, J. Shabanowitz, D. F. Hunt, *Nat. Protoc.* **2007**, *2*, 933-938.