MASSTRPLAN Protocol

Diagnostic ions for small aldehyde modifications of proteins

Grant Agreement number:	675132
Project acronym:	MASSTRPLAN
Project title:	MASS Spectrometry TRaining network for Protein Lipid adduct Analysis
Project coordinator name, title:	Prof. Corinne Spickett
Organization:	Aston University, Birmingham, UK
Project website address:	www.masstrplan.org

Protocol prepared by:	Prof Corinne M. Spickett
Organization:	Department of Bioscience, Aston University, Birmingham, UK
Contact details:	c.m.spickett@aston.ac.uk

Rationale for the Protocol

Small reactive aldehydes derived from lipid peroxidation or glycoxidation can form adducts on protects, a process called lipoxidation. Lipoxidation adducts have been found in several inflammatory diseases, including atherosclerosis and Alzheimer's disease. The detection and analysis of these adducts in biological samples is challenging because of their low abundance and the difficulty of interpreting mass spectrometry data from proteomic analysis. LC-MS is nowadays the technique of choice for the identification of protein-lipid adducts, since it can detect the mass shift caused by the adducts, and when used in tandem mode allows the modified amino acid to be located in the protein sequence. In complex samples thus is still challenging owing to many false positive identifications, as well as the possibility of failing to detect the lower abundance modified peptides in the presence of high abundance unmodified ones. To help overcome this, the information on the fragmentation patterns of aldehyde-protein adducts can be used to identify reporter ions, which are fragment ions characteristic and specific for each modification, rather than for the peptide sequence. These reporter ions can be used to add extra confidence to the identification of modifications in vivo, or search more specifically for the modification by semi-targeted mass spectrometry approaches, e.g. precursor ion scanning (PIS) or neutral loss scanning (NLS), thus simplifying the analysis.

This protocol details the detection of diagnostic ions for several small aldehyde adducts on model proteins lysozyme, human serum albumin and pyruvate kinase, based on our methods reported in Afonso *et al*. [1] and Sousa *et al*. [2].

Protocol

Sample Preparation - Synthesis of aldehyde-protein adducts

Lysozyme should be first reduced with 100 mM DTT to ensure some cysteine thiols are in the reduced form. This step can be omitted for HSA and pyruvate, which can used in their native forms. The model protein are typically prepared at 1 mg/mL.

Acrolein is added to the protein solution at concentrations from 2 µM to 5 mM (use higher concerntrations for good positive controls) and allowed to react for 2 hours at room temperature.

Pentanal is added to the protein solution at 16.6 mM and allowed to react for 24h at 37°C.

Malondialdehyde (MDA) must be prepared via acid hydrolysis of 1,1,3,3-tetramethoxypropane. MDA and 4-hydroxy-2(E)-hexenal (HHE) can be used at 2 µM to 5 mM, but likewise higher concentrations are recommended for clear positive results. 4-hydroxy-2(E)-nonenal (HNE) can be substituted for HHE, taking into account that the mass of HNE is 42 Da higher than that of HHE, so the reporter ions with be correspondingly higher m/z.

Adducts should be stabilized by adding 50 mM NaBH4 to the reaction and leaving for 1 hour at room temperature, with the exception of the HSA-pentanal reaction, which can instead reduced with 50 mM $NaBH_3CN$. For direct infusion mass spectrometry analysis, excess DTT in the reduced lysozyme samples should be removed prior to the reaction with the aldehydes using Microcon Ultracel YM-10 10,000 MWCO centrifugal concentrators (Millipore, Massachusetts, USA), following the manufacturer's protocol.

Protein in-gel digestion

The reaction products from above should be separated by 12.5 or 15 % SDS-polyacrylamide gel electrophoresis followed by staining with Coomassie blue to visualize the bands prior to further processing. Change the acrylamide concentration as required to suit proteins of different molecular weights. Bands present in the gel were then excised and tryptic digestion was performed according to Verrastro et al., 2016 [3]. The peptide extracts were dried for storage, and resuspended in H2O/acetonitrile (98%/2%) with 0.1% formic acid prior to MS analysis.

Alternatively, RapiGest SF (Waters, UK) can be used. To 100 µL of aldehyde-treated protein add 100 µL of RapiGest SF (Waters, UK) and vortex. DTT is then added to a final concentration of 5 mM and the reaction incubated at 60 °C for 30 minutes. Cysteine alkylation is then performed by adding iodoacetamide to a final concentration of 15 mM and incubating in the dark for 30 min at room temperature. Add trypsin (Trypsin Gold, Mass Spectrometry Grade, Promega, Southampton, UK) to a final concentration of 7 µg/mL and incubate the samples overnight at 37 °C. To prepare samples for LC/MS, add trifluoroacetic acid (TFA) to a final concentration of 0.5% and incubate at 37 °C for 45 min, during which time a precipitate forms. Acid-treated samples are centrifuged at 13000 rpm for 10 min then the supernatant carefully transferred to another tube and dried in a centrifugal evaporator. Store them dry at -20 °C. Samples should be resuspended in H2O/acetonitrile (98%/2%), 0.1% formic acid prior to MS analysis.

Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) Analysis

Peptides can be separated and analysed using an Ultimate 3000 system (Thermo Scientific, Hemel Hemstead, UK) coupled to a 5600 TripleTOF (ABSciex, Warrington, UK). The analysis is performed as previously described by Verrastro et al., 2016 [3]. Briefly, load the peptide solution onto a C18 trap column (C18 PepMapTM, 5 µm, 0.5 x 5mm, Thermo Scientific, Hemel Hemstead, UK) at 30 µL/min in 2% acetonitrile 0.5% formic acid followed by a 4 minute wash, before separation on a nano-HPLC column (C18 PepMapTM, 5 µm, 0.075 x 150mm, Thermo Scientific, Hemel Hemstead, UK) at 300 nL/min using a gradient elution running from 2% to 45% aqueous acetonitrile, 0.1% formic acid over 45 minutes.

Ionization of the peptides is achieved with spray voltage set at 2.4 kV, a source temperature of 150□C, declustering potential of 100V, nebulizing gas flow of 15 and a curtain gas setting of 25. Survey scans are collected in positive mode from 350 to 2000 Da using high-sensitivity TOF-MS mode. Information-dependent acquisition (IDA) is used to collect MS/MS data using the following criteria: the 10 most intense with +2 to +5 charge states and a minimum intensity of 200 cps were chosen for analysis, using dynamic exclusion for 12 s and standard rolling collision energy settings.

Data Analysis and Database Searches

The Mascot® probability based search engine (Matrix Science, London, version 2.4.0) is recommended to interrogate the SwissProt 2017-07 primary database, although other search engines can also be used. LC-MS .wiff files of each sample are searched for protein identification and oxidative post-translational modifications (oxPTMs). For protein identification, variable modifications of methionine oxidation and carbamidomethyl cysteine must be used. For the analysis of the lipoxidation products, the initial searches additionally should use a variable modification list including the following:

- pentanal and reduced pentanal adducts at lysine and histidine (mass changes of 68.06 Da and 70.06 Da);
- MDA and reduced MDA adducts at lysine and histidine (mass changes of 54.05 Da, 56.06 Da, 134.13 Da, 36.03 Da and 26.04 Da);
- reduced and unreduced acrolein adducts at cysteine, lysine and histidine (mass changes of 56.06 Da, 58.08 Da, 40.06 Da, 94.11 Da, 56.06 Da, 76.09 Da);
- reduced and unreduced HHE adducts at cysteine, lysine and histidine (mass changes of 114.14 Da, 93.13 Da, 78.11 Da);

The data should then be re-searched using the Mascot error tolerant search function. Other parameters for the searches should be as follows: Enzyme: Trypsin; Peptide tolerance: ±0.6 Da; MS/MS tolerance: ±0.6 Da; Peptide charge state: +2, +3; Max Missed cleavages: 1; #13C: 0; Quantitation: None; Instrument: ESI-QUAD-TOF; Data format: Mascot Generic; Experimental mass values: Monoisotopic; Taxonomy: Chordata. All data identifying modifications must be manually validated.

Example data

The structures and m/z ratios of the proposed reporter ions for acrolein and pentanal are shown in Figure 1.

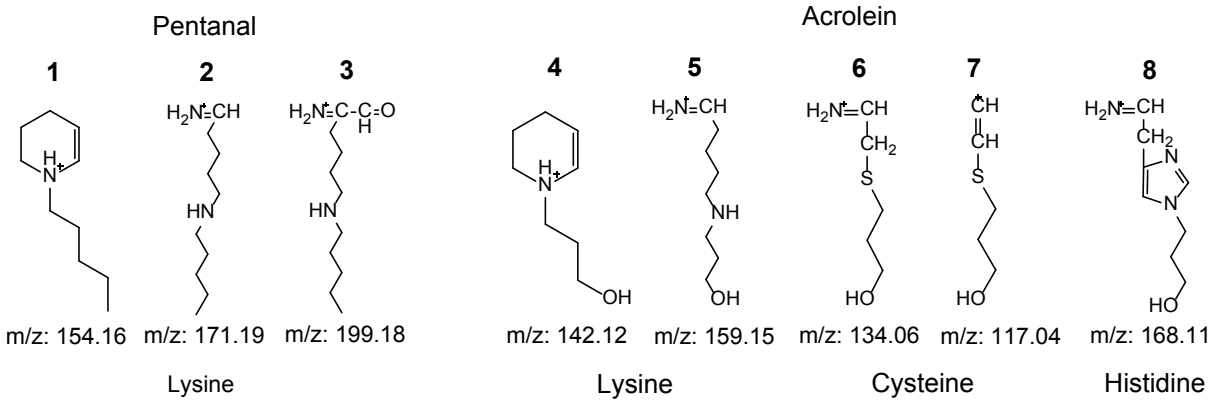

Figure 1. Proposed structures of the diagnostic ions found for the reduced pentanal modifications on the lysine residues (1-3) and for the reduced acrolein modifications on the lysine (4,5), cysteine (6,7) and histidine residues (8). Reproduced from [1]

Examples of MSMS spectra showing the fragmentation of aldehyde-modified peptides and the presence of diagnostic ions can be seen in Figure 2. All modifications were manually validated by de novo sequencing to confirm that the suspected modifications were indeed present. Based on this approach, details of peptides commonly found to be modified in the model proteins tested and their LC-MSMS characteristics are given in Tables 1-3 below.

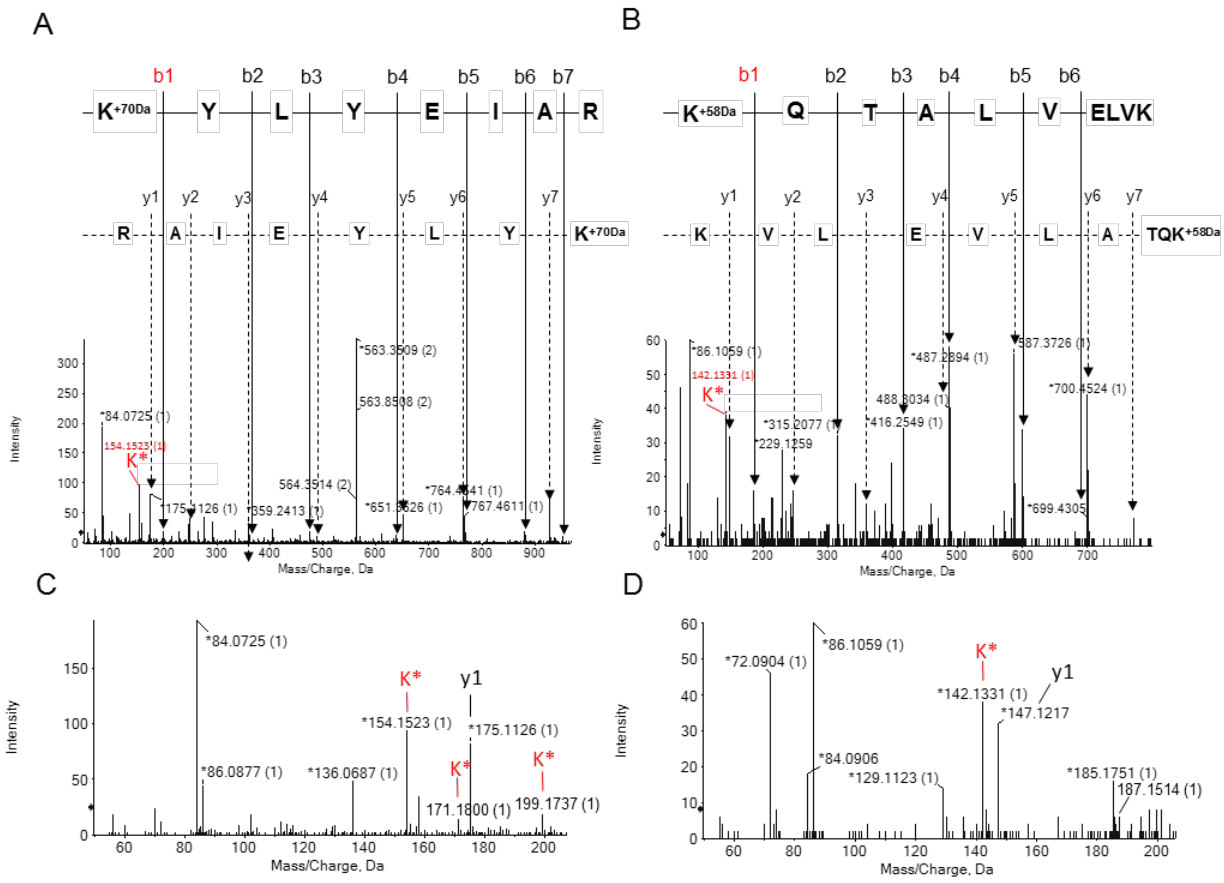

Figure 2. MS/MS spectra of different human serum albumin tryptic peptides (KYLYEIAR and KQTALVELVK) modified on a lysine residue by pentanal (A) and acrolein (B). The y and b ions indicated by the arrows confirm the peptide sequence and the modification on the lysine residue. Additionally, the ions at m/z 154.15, m/z 171.18 and 199.17, diagnostic of pentanal-modified lysine (C) and the ion at m/z 142.13, diagnostic of acrolein-modified lysine (D) are highlighted. Reproduced from [1]

Table 1. Modifications of lysozyme by reactive aldehydes Reproduced from [1]

Lysozyme modified peptide sequence ($_a$ ^{b})	Theoretical mass of modified peptide	Observed mass of modified peptide	m/z (charge state)	Ion score	Rt (min)	Diagnostic Ions ^{c}
Pentanal						
C_6ELAAAM**K**$^{+70}$$R_{14}$	1118.59	1118.62	373.9 (3+)	41	30.95	-
C_{115}**K**$^{+70}$GTDVQAWIR$_{125}$	1402.74	1402.79	368.6 (4+)	37	36.36	-
Acrolein						
C$_6^{+56}$ELAAAMK$_{13}$	891.07	891.42	446.5 (2+)	18	36.19	-
C$_6^{+58}$ELAAAMK$_{13}$	893.12	893.44	447.5 (2+)	79	27.57	6,7

$G_{22}YSLGNWVC^{+58}AAK_{33}$	1325.16	1325.64	663.6 (2+)	80	38.35	-
$W_{62}WC^{+58}NDGR_{68}$	993.06	993.41	497.5 (2+)	35	32.31	6
$N_{74}LC^{+56}NIPCSALLSSDITASVNCAK_{96}$	2506.26	2506.19	836.4 (3+)	77	42.39	-
$N_{74}LC^{+58}NIPCSALLSSDITASVNCAK_{96}$	2508.31	2508.20	837.1 (3+)	50	43.84	6,7
$N_{74}LCNIPC^{+58}SALLSSDITASVNCAK_{96}^{+58}$	2509.33	2509.22	837.4 (3+)	92	44.29	6
$N_{74}LC^{+58}NIPC^{+58}SALLSSDITASVNCAK_{96}^{+58}$	2510.33	2510.24	838.1 (3+)	60	44.84	6
$C_{115}^{+58}KGTDVQAWIR_{125}$	1333.21	1333.68	445.4 (3+)	17	31.19	-
$C_{115}^{+56}K^{+58}GTDVQAWIR_{125}$	1389.19	1389.71	464.1 (3+)	28	30.24	5
$C_{115}K^{+58}GTDVQAWIR_{125}$	1390.21	1390.70	464.4 (3+)	30	31.08	5

a (subscript) – amino acid position in the mature protein for the start and end residues

b (superscript) - mass difference corresponding to the modification on the affected residue (shown in bold red)

c (superscript) – numbers refer to the structures shown in Figure 3.

Table 2. Modifications of human serum albumin (HSA) by reactive aldehydes

Reproduced from [1]

HSA modified peptide sequence ($_a$ ^{b})	Theoretical mass of modified peptide	Observed mass of modified peptide	m/z (charge)	Ion score	Rt (min)	Diagnostic Ions ^{c}
Pentanal						
$F_{11}K^{+70}DLGEENFK_{20}$	1295.69	1295.68	648.9 (2+)	58	35.30	1, 3
$S_{65}LHTLFGDK^{+70}LC^*TVATLR_{81}$	2001.11	2001.12	668.1 (3+)	76	47.16	1, 3
$K_{137}^{+70}YLYEIAR_{144}$	1124.66	1124.69	563.3 (2+)	39	37.19	1, 2, 3
$Y_{161}K^{+70}AAFTEC^*C^*QAADK_{174}$	1731.79	1731.81	578.3 (3+)	58	28.88	1, 3
$L_{198}K^{+70}C^*ASLQK_{205}$	1016.61	1016.62	509.3 (2+)	37	25.15	1, 3
$C_{200}^*ASLQK^{+70}FGER_{209}$	1264.66	1264.68	633.3 (2+)	42	30.91	1, 2, 3
$A_{210}FK^{+70}AWAVAR_{218}$	1088.65	1088.68	363.9 (3+)	51	37.12	1, 2, 3
$Y_{263}ICENQDSISSK^{+70}LK_{276}$	1753.89	1753.91	585.6 (3+)	78	34.16	1, 3
$L_{275}KEC^*C^*EK^{+70}PLLEK_{286}$	1615.87	1615.88	539.6 (3+)	34	28.84	1, 3
$L_{349}AK^{+70}TYETTLEK_{359}$	1365.77	1365.78	456.3 (3+)	35	30.00	1, 3
$Q_{390}NC^*ELFEQLGEYK^{+70}FQNALLVR_{410}$	2668.37	2668.43	890.5 (3+)	130	49.68	1, 2, 3
$K_{414}^{+70}VPQVSTPTLVEVSR_{428}$	1709.01	1709.03	855.5 (2+)	77	36.55	1, 2, 3
$K_{525}^{+70}QTALVELVK_{534}$	1197.76	1197.77	599.9 (2+)	55	37.32	1, 2, 3
$E_{542}QLK^{+70}AVMDDFAAFVEK_{557}$	1909.99	1910.01	637.7 (3+)	65	37.48	1, 3
Acrolein						
$S_{65}LH^{+58}TLFGDK_{73}$	1074.59	1074.57	359.2 (3+)	43	28.51	8

K$_{137}$$^{+58}$YLYEIAR$_{144}$	1112.63	1112.62	557.3 (2+)	16	27.69	4,5
R$_{145}$**H**$^{+58}$PYFYAPELLFFAK$_{159}$	1956.03	1956.03	653.0 (3+)	66	41.35	8
A$_{258}$DLA**K**$^{+58}$YICENQDSISSK$_{274}$	1998.95	1998.96	667.3 (3+)	73	31.29	5
Y$_{263}$ICENQDSISSKL**K**$_{276}$$^{+58}$	1741.86	1741.86	581.6 (3+)	109	26.71	4
S$_{287}$**H**$^{+58}$CIAEVENDEMPADLPSLAADFVESK$_{313}$	3031.34	3031.38	1011.5 (3+)	61	42.41	8
R$_{337}$**H**$^{+58}$PDYSVVLLLR$_{348}$	1524.89	1524.88	509.3 (3+)	83	34.59	8
Q$_{390}$N**C**$^{+56}$ELFEQLGEYK$_{402}$	1655.72	1655.75	829.4 (2+)	42	41.14	6
K$_{414}$$^{+58}$VPQVSTPTLVEVSR$_{428}$	1696.97	1696.97	566.3 (3+)	59	30.89	4
K$_{525}$$^{+58}$QTALVELVK$_{534}$	1185.76	1185.73	396.3 (3+)	52	28.84	4
K$_{574}$$^{+58}$LVAASQAALG$_{584}$	1198.73	1198.73	600.4 (2+)	62	33.67	4

a (subscript) – amino acid position in the mature protein for the start and end residues

b (superscript) - mass difference corresponding to the modification on the affected residue (shown in bold red)

c (superscript) – numbers refer to the structures shown in Figure 3.

Table 3. Pyruvate kinase residues modified after 10 min treatment at high aldehyde concentrations *in vitro*

Modified Residues	Pyruvate kinase modified peptides ($_{a}$ ^{b})	Theoretical mass of modified peptide	Observed mass of modified peptide	m/z (charge)	Ion score	Rt (min)
Acrolein						
Cys49	$_{44}$NTGII**C**$^{+56}$TIGPASR$_{56}$	1357.70	1357.83	679.9 (2+)	52	27.76
Cys49	$_{44}$NTGII**C**$^{+58}$TIGPASR$_{56}$	1359.72	1359.84	680.9 (2+)	54	27.69
Lys66	$_{63}$EMI**K**$^{+58}$SGMNVAR$_{73}$	1292.66	1292.81	431.9 (3+)	16	21.87
Lys166	$_{163}$NIC*__K__$^{+58}$VVDVGSK$_{173}$	1275.68	1275.81	638.9 (2+)	13	21.94
Lys207	$_{207}$**K**$^{+40}$GVNLPGAAVDLPAVSEK$_{224}$	1804.01	1804.19	602.4 (3+)	62	31.00
Cys326	$_{320}$AGKPVI**C**$^{+58}$ATQMLESMIK$_{336}$	1876.98	1876.17	626.7 (3+)	19	52.72
Cys358	$_{343}$AEGSDVANAVLDGAD**C**$^{+58}$IMLSGETAK$_{367}$	2494.16	2494.37	832.5 (3+)	37	49.48
Lys367	$_{343}$AEGSDVANAVLDGADC*IMLSGETA**K**$^{+58}$GDYPLEAVR$_{376}$	3551.67	3551.99	889.0 (4+)	26	47.52
Lys393	$_{393}$**K**$^{+58}$LFELAR$_{400}$	1062.61	1062.72	532.4 (2+)	27	29.34
Lys393	$_{393}$**K**$^{+40}$LFELAR$_{400}$	1060.59	1060.63	531.3 (2+)	9	21.87
Cys423	$_{423}$**C**$^{+58}$LAAALIVLTESGR$_{436}$	1473.82	1473.96	737.9 (2+)	89	51.53
His464	$_{462}$QA**H**$^{+40}$LYR$_{467}$	826.44	826.51	414.3 (2+)	21	21.03
Cys474	$_{468}$GIFPVV**C**$^{+58}$K$_{475}$	975.55	975.63	460.8 (2+)	31	30.97
Lys475	$_{468}$GIFPVVC*__K__$^{+58}$DPVQEAWAEDVDLR$_{489}$	2599.30	2599.51	867.5 (3+)	57	53.72
4-hydroxy-hexenal						
Cys152	$_{152}$**C**$^{+96}$DENILWLDYK$_{162}$	1506.71	1506.86	754.44 (2+)	24	39.35

Cys152	$_{152}$**C^{+114}**DENILWLDYK$_{162}$	1524.72	1524.72	763.43 (2+)	72	33.52
Lys188	$_{187}$Q**K^{+114}**GPDFLVTEVENGGFLGSK$_{206}$	2235.14	2235.34	746.12 (3+)	17	35.84
Lys247	$_{247}$**K^{+96}**AADVHEVR$_{255}$	1119.60	1119.73	374.25 (3+)	27	18.23
Cys326	$_{320}$AGKPVI**C^{+96}**ATQMLESMIKKPRPTR$_{342}$	2650.45	2650.72	531.15 (5+)	14	38.00
Cys326	$_{320}$AGKPVI**C^{+114}**ATQMLESMIKKPRPTR$_{342}$	2668.46	2668.72	445.79 (6+)	13	29.35
Cys358	$_{343}$AEGSDVANAVLDGAD**C^{+114}**IMLSGETAK$_{367}$	2550.18	2550.42	851.15 (3+)	36	36.37
His379	$_{377}$MQ**H^{+114}**LIAR$_{383}$	981.54	981.64	491.83 (2+)	13	19.81
His391	$_{384}$EAEAAMF**H^{+114}**R$_{392}$	1174.54	1174.67	392.56 (3+)	10	21.82
Lys393	$_{393}$**K^{+114}**LFEELAR$_{400}$	1118.63	1118.72	560.37 (2+)	26	29.38
Cys423	$_{423}$**C^{+96}**LAAALIVLTESGR$_{436}$	1529.85	1529.98	765.99 (2+)	76	40.94
Cys474	$_{468}$GIFPVV**C^{+114}**K$_{475}$	975.55	975.62	488.82 (2+)	26	28.22
Lys475	$_{468}$GIFPVVC**K^{+114}**DPVQEAWAEDVDLR$_{489}$	2599.29	2599.49	867.50 (3+)	85	33.83
Lys115	$_{93}$TATESFASDPILYRPVAVALDT**K^{+54}**GPEIR$_{120}$	3070.59	3070.81	768.71 (4+)	42	34.07
Lys135	$_{126}$GSGTAEVEL**K^{+54}**K$_{136}$	1171.61	1171.52	391.51 (3+)	52	18.74
Lys188	$_{187}$Q**K^{+54}**GPDFLVTEVENGGFLGSK$_{206}$	2175.08	2175.21	726.08 (3+)	94	34.69
Lys207	$_{207}$**K^{+54}**GVNLPGAAVDLPAVSEK$_{224}$	1817.99	1818.03	607.02 (3+)	70	28.67
Lys224	$_{208}$GVNLPGAAVDLPAVSE**K^{+54}**DIQDLK$_{230}$	2402.27	2402.45	801.82 (3+)	87	33.97
Lys270	$_{267}$IIS**K^{+54}**IENHEGVR$_{278}$	1447.78	1447.74	483.59 (3+)	82	23.05
Lys305	$_{295}$GDLGIEIPAE**K^{+54}**VFLAQK$_{311}$	1881.03	1881.08	628.03 (3+)	48	34.80
Lys393	$_{393}$**K^{+54}**LFEELAR$_{400}$	1058.58	1058.57	530.29 (2+)	37	27.85
Lys475	$_{468}$GIFPVVC*$^{+54}$DPVQEAWAEDVDLR$_{489}$	2596.26	2596.49	866.50 (3+)	72	36.79

a (subscript) – amino acid position in the mature protein

b (superscript) - mass difference corresponding to the modification

References

1. Afonso CB, Sousa BC, Pitt AR, Spickett CM. (2018) A mass spectrometry approach for the identification and localization of small aldehyde modifications of proteins. Arch Biochem Biophys. 646:38-45. https://doi:10.1016/j.abb.2018.03.026

2. Sousa BC, Ahmed T, Dann WL, Ashman J, Guy A, Durand T, Pitt AR, Spickett CM (2019) Short-chain lipid peroxidation products form covalent adducts with pyruvate kinase and inhibit its activity in vitro and in breast cancer cells. Free Radic Biol Med in press. https://doi.org/10.1016/j.freeradbiomed.2019.05.028

3. Verrastro, K. Tveen-Jensen, R. Woscholski, C.M. Spickett, A.R. Pitt, Reversible oxidation of phosphatase and tensin homolog (PTEN) alters its interactions with signaling and regulatory proteins, Free Radic. Biol. Med. 90 (2016) 24–34. https://doi:10.1016/j.freeradbiomed.2015.11.004