**Revision**

# A filtered database search algorithm for endogenous serum protein carbonyl modifications in a mouse model of inflammation.

Peter G. Slade[1], Michelle V. Williams[1], Alison Chang[1], Elizabeth Iffrig[2], Steven R. Tannenbaum[1,2], and John S. Wishnok[1]

Department of Biological Engineering[1], and Department of Chemistry[2], Massachusetts Institute of Technology, Cambridge, MA 02139

Address correspondence to JSW
MIT Room 56-747B
Cambridge, MA 02139
617-253-6795
Wishnok@mit.edu

**Abstract**

During inflammation, the resulting oxidative stress can damage surrounding host tissue, forming protein-carbonyls. The SJL mouse is an experimental animal model used to assess *in vivo* toxicological responses to reactive oxygen and nitrogen species from inflammation. The goals of this study were to identify the major serum proteins modified with a carbonyl functionality and to identify the types of carbonyl adducts. To select for carbonyl-modified proteins, serum proteins were reacted with an aldehyde reactive probe (ARP) that biotinylated the carbonyl modification. Modified proteins were enriched by avidin affinity and identified by 2-D LC-MS/MS. To identify the carbonyl modification, tryptic peptides from serum proteins were subjected to avidin affinity and the enriched modified peptides were analyzed by LC-MS/MS. It was noted that the ARP tag created tag-specific fragment ions and neutral losses, and these extra features in the mass spectra inhibited identification of the modified peptides by database searching. To enhance the identification of carbonyl-modified peptides, a program was written which used the tag-specific fragment ions as a fingerprint (i*n silico* filter program) and filtered the mass spectrometry data to highlight only modified peptides. A *de novo*-like database search algorithm was written (biotin peptide identification program, BPI) to identify the carbonyl-modified peptides. Although written specifically for our experiments, this software can be adapted to other modification/enrichment systems. Using these routines, a number of lipid peroxidation-derived protein carbonyls and direct side-chain oxidation proteins carbonyls were identified in SJL mouse serum.

## Introduction

During inflammation, activated phagocytes secrete reactive nitrogen species (RNS) and reactive oxygen species (ROS) that can eliminate infectious agents. If inflammation is chronic, RNS and ROS can also damage surrounding host tissue, resulting in protein modification in the form of protein-carbonyls (1). Total protein carbonylation has been used as a marker of oxidative stress and inflammation (2) and increased levels have been seen in heart disease, lung disease, aging, neurodegenerative disorders, and inflammatory bowel disease (2-7). The carbonylation of proteins can result from the direct oxidation of protein side-chains, forming the glutamate and aminoadipate semialdehydes (Scheme 1) (8,9), but can also occur from the indirect oxidation of polyunsaturated fatty acids (lipid peroxidation) and carbohydrates, leading to a variety of reactive aldehydes (Scheme 2) (10). These aldehydes covalently modify proteins through conjugate addition (often Michael addition) to nucleophilic amino acid side chains, creating protein-bound carbonyls (10,11).

In a previous study, twelve DNA oxidative damage products, from tissues from the SJL mouse model of inflammation, were quantitated (12). Only the lipid peroxidation adducts increased in association with inflammation, which suggested an important role of lipids in inflammatory disease progression and established a direct correlation between inflammation and the increased formation of reactive aldehydes from oxidized lipids. Although DNA modification due to inflammation has been the focus of many animal and human studies, it is proteins that are considered most likely to be ubiquitously affected by disease, response and recovery (13), and the biological consequences include more-rapid protein turnover as well as novel signaling (14-16) . The overall carbonylation of proteins has been well documented in other inflammatory animal models, which have shown significant increases in protein-carbonyls in the mucosal

lining of rat colon (17) and mouse colon(5) while increased levels of protein carbonyls were observed in rat serum, along with a higher turnover of proteins from the inflamed tissue (18,19). Furthermore, increased protein carbonyl modification has been reported in studies of the colon mucosal lining from patients diagnosed with inflammatory bowel disease (20,21). Taken together, these observations suggest that an increase in carbonylated proteins is likely in the SJL mouse and that the extent and type of protein-carbonyls could potentially be a marker for inflammation and disease.

The SJL mouse is an experimental animal model used to assess *in vivo* toxicological responses to nitric oxide (NO) overproduction from inflammation (22). Injections of RscX lymphoma cells into these mice result in rapid tumor growth as well as host T cell proliferation in lymph nodes, spleen and liver, resulting in morbidity within 15 days. The induced macrophages create a 50-fold increase in NO production in spleen and lymph nodes and the post translational modification 3-nitrotyrosine was highly elevated in spleen tissue.

The identification of endogenously formed protein carbonyls in serum is challenging due to their low abundance and the large number of possible modifications (1,2,23), some of which are shown in schemes 1 and 2. We recently identified proteins modified by the carbonyl 9,12-dioxo-10(E)-dodecenoic acid (DODE) in cells treated with the hydroperoxide of linoleic acid (13-HPODE) (24). This work used a technique first demonstrated by Maier and co-workers (25,26). Protein carbonyls were derivatized with an aldehyde reactive probe (ARP), a biotinylated hydroxylamine that reacts preferentially with aldehyde/keto groups (27), allowing for subsequent enrichment of the modified proteins by avidin affinity. DODE-modified proteins were also identified using an anti-DODE antibody and Western blots. Although a number of DODE modified proteins were identified, we were unable to definitively identify the carbonyl

modified peptides by mass spectrometry due both to low abundance and to the interference of ARP-tag-specific fragment ions on database searching.

In this current study, SJL mouse serum was screened for the presence of protein carbonyls endogenously formed during inflammation. Carbonyl-modified proteins were then identified using techniques previously established (24); first anti-DODE Western blotting followed by ARP derivatization/enrichment and 2D-LC-MS/MS. These proteins then formed a database of putative carbonyl-modified proteins from SJL mouse serum. To identify the type of carbonyl modification and the modified peptide, the ARP derivatized peptides were enriched and analyzed by mass spectrometry. To minimize the confounding effect of ARP fragmentation, an algorithm *(in silico* filter) was written that filtered the mass spectrometry data to select only those peptides containing the known ARP pattern of fragmentation. This alone effectively reduced the number of false positives. To further alleviate the interfering effects of ARP fragments on peptide identification by database searching, a *de novo* searching algorithm (Biotin Peptide Identification program, BPI) was written. Peptides were evaluated against the database of proteins that had been previously identified as potentially carbonyl modified. Since modified peptides were searched against a finite list of proteins and all results were manually evaluated, the BPI program did not calculate a statistical peptide score, which allowed the identification of lower abundant modified peptides that would not be considered significant by standard search engines such as Mascot. The BPI program was also written with the flexibility to evaluate a wide range of known carbonyl-adduct masses and could therefore screen for a large number of carbonyl adducts at one time. This should also allow the program to be used with modification/enrichment systems other than the one used here. The program thus selected a finite number of carbonyl modified peptides, resulting in the identification of a number of

proteins that were endogenously carbonylated in serum from the SJL mouse inflammation model.

## Materials and Methods

**Materials.** Aldehyde reactive probe (ARP) was purchased from Invitrogen (Eugene, OR) and biotin-PEO-LC-Amine was purchased from Pierce (Rockford, IL). Cytochrome c (equine heart), acetic acid, and trifluoroacetic acid were purchased from Sigma Chemical Co. (St. Louis, MO). Trypsin was purchased from Promega (Madison, WI). Gases were supplied by AirGas (Salem, NH). DODE was a generous gift from Prof. Ian A. Blair (University of Pennsylvania).

**SJL mouse infection and serum extraction**. RcsX cells (kindly supplied by Prof. Nicolas M. Ponzio, University of New Jersey Medical Center, Newark, NJ) were passaged through SJL mice (Jackson Laboratory, Bar Harbor, ME) and harvested from lymph nodes 14 days after inoculation according to published procedures (28). Cells were manually dissociated from lymph nodes followed by washing in phosphate-buffered saline (PBS; 140 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, pH 7.4) and freezing in aliquots of 5 x $10^7$ cells in 10% dimethylsulfoxide/fetal bovine serum. To initiate NO overproduction, eight SJL mice (5–6 weeks old) were each injected intra-peritoneally with $10^7$ RcsX cells in 200 μL of PBS. Ten mice were injected with 200 μL of PBS as unstimulated controls. Twelve days after injection, five treated and three control mice were anaesthetized with isofluorane, and serum was collected by cardiac puncture. Pooled samples were desalted by filter centrifugation and dried in a SpeedVac. Protein content was determined with a commercial bicinchoninic acid (BCA) protein assay kit (Pierce).

**SDS-PAGE of SJL mouse serum, 1 and 2-D.** 100 µg (Coomassie) or 30 µg (Western) of sample was processed by 2-D gel electrophoresis for protein identification/Western blotting. For anti-biotin Western blotting and Coomassie, proteins were focused on 11 cm 4-7 immobilized pH gradient (IPG) strips (Immobiline™ DryStrip gels, Amersham Biosciences) using an IPGphor focusing apparatus (Amersham Biosciences). For anti-DODE western blotting and Coomassie staining, proteins were focused on an 11 cm 3-10 pH gradient strip. Samples were applied by cup loading. IPG strips were then equilibrated in equilibration buffer (50 mM Tris-HCl, 6M urea, 30% glycerol, 2% SDS) supplemented with 1% dithiothreitol (DTT) to maintain the fully-reduced state of proteins, followed by 2.5% iodoacetamide to prevent re-oxidation of thiol groups during electrophoresis. Proteins were separated on 12.5% Tris/Glycine gels (BioRad) using a Criterion System (BioRad). Proteins were visualized by Coomassie SimplyBlue SafeStain™ (Invitrogen).

**Anti-Biotin and Anti-DODE Western blots.** Proteins were transferred to a polyvinylidine fluoride (PVDF) membrane (BioRad). Precision Plus protein standard (BioRad) was used to estimate molecular weights. Anti-biotin (Cell Signaling) 1:20,000 was used to detect biotinylated proteins. Polyclonal anti-DODE antibody (29) 1:1500 was used to detect DODE-modified proteins.

**In-Gel Digestion.** Protein spots were picked and washed in Milli-Q® water for 15 min, then washed three times in 25 mM $NH_4HCO_3$/50% $CH_3CN$ for 30 minutes. Gel plugs were dehydrated in 100% $CH_3CN$ for 10 min while vortex-mixing. The supernatant was removed, and the plugs were dried in the SpeedVac. Trypsin (1µg/50µL) (Promega) suspended in 25 mM $NH_4HCO_3$ was added, and gel plugs were re-hydrated for 30 min on ice and then digested overnight at 37˚C. The samples were then centrifuged, and the supernatant was removed. The

pellet was re-suspended in $CH_3CN$ with 1% TFA, vortexed, and sonicated for 30 minutes to release hydrophobic peptides. The supernatant was removed and combined with the previous supernatant and stored at -20˚C until ready for MS/MS analysis.

**Neutravidin affinity chromatography of ARP derivatized serum proteins.**

Neutravidin was used to enrich for ARP-carbonyl derivatized proteins from SJL serum. These proteins were identified (described below) to make up the protein database used in the Biotin Peptide Identification program. Immobilized neutravidin (Pierce) was packed into a column with dimensions ID = 9 mm, OD = 7 mm, height = 60 mm. The final column volume was 2 mL. All buffers and samples were brought to room temperature. The column was washed with 4 column volumes of PBS. A protease inhibitor cocktail was added to the protein samples: AEBSF (1 mM), E-64 (10 μM), pepstatin A (1.4 μM), EDTA (1 mM), bestatin (40 μM). Samples (~ 2 mg protein in 1 mL PBS) were added directly to the neutravidin column. To remove non-biotinylated proteins, the column was washed with 10 column volumes of PBS-2% 3-[(3-cholamidopropyl)dimethylammonio]propanesulfonic acid (CHAPS), 5 column volumes of PBS and 5 column volumes of deionized water. Wash fractions were analyzed by UV-Vis spectrometry for baseline absorbance (280 nm). Bound proteins were eluted by the addition of 4 column volumes of 0.4% TFA/80% acetonitrile. Elution fractions were combined, frozen, lyophilized, and stored at -20°C.

**Neutravidin affinity chromatography of ARP derivatized serum peptides.**

Neutravidin columns were prepared as described above. Serum proteins were digested in-solution to peptides with trypsin as follows: to 200 μg of protein 4 μg of sequencing grade trypsin (Promega) was added in ~200 μL 50 mM ammonium bicarbonate; proteins were digested for 4 hours at 37°C. Total serum peptide digests were added directly to the neutravidin column.

Non-biotinylated peptides were removed with 10 column volumes of PBS-2% 3-[(3-cholamidopropyl)dimethylammonio]propanesulfonic acid (CHAPS), 5 column volumes of PBS and 5 column volumes of deionized water. Bound peptides were eluted by the addition of 4 column volumes of 0.4% TFA/80% acetonitrile. Elution fractions were combined, frozen, lyophilized, and stored at -20°C.

**DODE modification and ARP derivatization of mouse serum albumin and cytochrome c.** Mouse serum albumin and cytochrome c (1 mg/mL, 100 µl in pH 7.0 chelex-treated 100 mM HEPES buffer) were reacted with DODE (224 nmols, 10 µL ethanol) in the presence of vitamin C (10 mM) at 37°C for 24 h. Proteins were filtered (regenerated cellulose 3,000 Da MWCO; Amicon,) to remove un-reacted DODE and vitamin C, and brought up in PBS to a concentration of 1 mg/mL. DODE modified protein samples (1 mg/mL) were incubated with Aldehyde Reactive Probe (ARP) (10 mM) at a final volume of 1.0 mL in phosphate buffer, pH = 5-6. The reaction was stirred vigorously for 12 hours at room temperature. The samples were filtered (regenerated cellulose 3,000 Da MWCO; Amicon,) to remove unreacted ARP.

**Strong cation exchange (SCX) separation for SJL serum proteins.** Neutravidin elutions of proteins from SJL mouse serum were digested in-solution with trypsin as follows: to 200 µg of protein 4 µg of sequencing grade trypsin (Promega) was added in ~200 µL 25 mM ammonium bicarbonate; proteins were digested for 4 hours at 37°C. Digested proteins were desalted by SPE (Strata 50um, tri-Func, C18-E), dried and re-dissolved in 100 µL mobile phase A (described below) for SCX separation. Purified peptides were separated by SCX liquid chromatography using an Agilent 1100 LC system with a Polysulfoethyl A 100 x 4.6 mm, 5 µm column (Nest Group) and a flow rate of 0.25 mL/min. Mobile phase A was 10 mM sodium phosphate ($Na_2HPO_4$), 25% acetonitrile, pH = 2.8 and mobile phase B was 10 mM sodium

phosphate ($Na_2HPO_4$), 25% acetonitrile, 0.4 M KCl, pH = 2.8. A linear gradient was as follows: 0% B from 0-10 min, 100 % B at 70 min, 100% B at 75 min and 0% B at 80 min. Peptides were collected in 1-minute fractions to give about 40-60 fractions per sample. These were each desalted by ZipTips® and re-dissolved in 5 μL 0.4% acetic acid for LC-ESI-MS/MS analysis.

**LC-MS/MS (Applied Biosystems QStar Elite).** The electrospray interface for this instrument uses a micro-tee (Upchurch Scientific, Oak Harbor, WA) with a 1-in. piece of platinum rod, inserted into one arm of the micro-tee, to supply the electrical connection. The electrospray voltage was typically 1600-1700 V applied just upstream of the column. Data-dependent MS/MS analysis was performed on the three most intense peaks in each full-scan spectrum, using multiply-charged states (most of the non-peptide background constituents are singly charged). Samples were spiked with 100 fmoles of vasoactive intestinal peptide fragment (amino acids 1-12) as the internal standard for calibration and mass accuracy. Accumulation time and pulsar frequency were maintained at 3 seconds and 6.99 seconds, respectively; the mass tolerance was 50 mmu. MS/MS was performed using nitrogen as the collision gas. Rolling collision energies were calculated as follows: z = 1, (collision energy) = 0.06*($m/z$) + 8; z = 2, (collision energy) = 0.05*($m/z$) + 5.6; z = 3, (collision energy) = 0.04*($m/z$) + 6.7; z > 3, (collision energy) = 0.015*($m/z$) + 25. Chromatography was performed using an Agilent 1100 capillary LC system with fused-silica capillary columns (75 μm i.d. × 360 μm o.d.; 14 cm length, tip 8 μm, New Objective, Woburn, MA) that were packed in-house with 5-μm C18 reverse-phase material (Vydac, Hesperia, CA). The flow rate from the pumps was 3-5 μL/min, and flow was split, before introduction of the sample onto the column, to 100-200 nL/min. Solvent A was 0.5% acetic acid in water and solvent B was 0.5% acetic acid in

acetonitrile. A linear gradient was performed as follows: 2% B at 0 min, 2% B at 3 min, 65% B at 70 min, 80% B at 80 min, 80%B at 90 min, 2%B at 130 min, and 2% B at 170 min.

**Protein and peptide and modification identification and characterization.** Peak lists for the Applied Biosystems QSTAR Elite were generated as mgf files by Applied Biosystems Analyst QS 1.1 software.

Database searches were carried out using Mascot version 2.2 (Matrix Science). For protein identification, either the NCBInr and/or the SwissProt mouse databases were searched. Parameters included: enzyme: trypsin; max missed cleavages = 2; fixed modifications carbamidomethyl (C) for 2-D gel analysis only; variable modifications of methionine oxidation (M); precursor tolerance was set at 0.1 Da; MS/MS fragment tolerance was set at 0.2 Da. Proteins were identified with 3 or more peptides. Significance of a protein match for Mascot was based on an expectation value of $< 0.05$ and a combined peptide score $> 50$.

For ARP-carbonyl modified peptides, the Mascot modification file was edited to add the DODE-ARP modification with elemental composition $C_{24}H_{37}N_5O_7S$ (calculated monoisotopic mass 539.2414 Da) and specificity for lysine. In addition to the elemental composition of the modifying group, the Mascot modification file allows for the inclusion of neutral losses, used in scoring of a fragment ion spectrum and for the peptide neutral loss. The Mascot modification file also allows for ions to be ignored in scoring ("Ignore Masses"). Neutral losses and ignored ions were determined experimentally for ARP-DODE. Neutral-loss fragments included the entire modifying group, $C_{24}H_{37}N_5O_7S$ (calculated monoisotopic mass 539.2414 Da) and the ARP portion of the modifying group, $C_{12}H_{21}N_5O_4S$ (calculated monoisotopic mass 331.12124 Da). "Ignore Mass" ions included $C_{10}H_{15}N_2O_2S$ (calculated monoisotopic mass 227.0854 Da), $C_{10}H_{19}N_4O_2S$ (calculated monoisotopic mass 259.1229 Da), $C_{12}H_{19}N_4O_3S$ (calculated

monoisotopic mass 299.1178 Da) and $C_{12}H_{22}N_5O_4S$ (calculated monoisotopic mass 332.1393 Da).

**Programming language.** All programming for the i*n silico* filter program and the Biotin Identification Program was done in Python Programming Language (http://www.python.org)**.** Python programs were written to be used on the Macintosh OS X operating system. Python text files for the i*n silico* filter program and the Biotin Identification Program are available for view or download at: http://web.mit.edu/toxms/www/filters.htm.

## Results

**Identification of putative carbonyl modified proteins.** Proteins modified by carbonyls were identified by methods previously established (24). SJL mice were injected with RcsX lymphoma cells as described in Materials and Methods while control mice were injected with PBS. After 12 days, mice were bled and serum separated by centrifugation to generate infected and control SJL mouse serum samples. Samples were aliquoted and stored at -80°C.

Twenty μg of ARP-derivatized and non-derivatized serum proteins from control and infected SJL mice were separated by 1-D SDS-PAGE and analyzed by Western blot using either Streptavidin-HRP or an anti-DODE to analyze ARP-derivatized and non-derivatized respectively (Figure 1). No DODE-modified proteins were detected in control mice, establishing the background level of DODE-modified proteins at essentially zero. Streptavidin-HRP Western blot analysis of carbonyl-modified proteins (ARP derivatized) from control (uninfected) SJL mouse serum demonstrated a low abundant background of carbonylation. This low amount of modified protein was below the detection limit of our mass spectrometry methods and therefore background carbonylation was also considered to be zero. Although abundant serum proteins are

generally considered poor markers of disease and inflammation, carbonyl-modification of these proteins may be of interest, therefore the most abundant serum proteins such as albumin or transferrin were not depleted prior to analysis. This decision likely limited the identification of lower abundant carbonyl-modified proteins but allowed the identification of potentially important carbonyl modifications on some of these abundant proteins.

**Identification of DODE-modified proteins.** To identify DODE-modified protein candidates, proteins from the infected mouse serum were separated by 2-D SDS-PAGE and the modified proteins located by anti-DODE Western blot (Supplementary Material Figure 1). Protein spots were cut from a corresponding Coomassie-blue-stained 2-D gel and digested with trypsin. Proteins were identified by nano-LC-MS/MS and Mascot data base searching. A relatively small number of potential DODE carbonyl-modified proteins were identified by this method (these are highlighted in red in Supplemental Table ST1).

**Identification of carbonyl-modified proteins.** To identify serum proteins modified with carbonyls other than DODE, serum from the infected SJL mouse was reacted with the aldehyde reactive probe (ARP) and modified proteins were enriched by avidin affinity as described in Materials and Methods. Enriched modified proteins were digested with trypsin and peptides, then separated by 2D chromatography (strong cation exchange (SCX) and reverse phase). Proteins were identified by tandem mass spectrometry and Mascot database searching. shows the complete list of putative carbonyl-modified proteins identified by both the 2D gel anti-DODE method and the ARP 2D chromatography method. Proteins identified by both methods are highlighted in red and were considered likely candidates for DODE modified serum proteins. This list of proteins was used as a database for the identification of the carbonyl modified peptides, as described below.

**Database searching of ARP-DODE modified peptides**. The ARP tag can generate tag-specific fragments and neutral losses that inhibit successful database searching (24). To further explore this issue a DODE-ARP modified cytochrome c standard was prepared as previously described (24,30). Modified cyt c was enriched by avidin affinity, digested with chymotrypsin and analyzed by nano-LC-MS/MS (Materials and Methods). As previously reported (30), DODE was found to modify lysines on three chymotryptic peptides; $^{98}$LKKATNE$^{104}$, $^{83}$AGIKKKTEREDLIAY$^{94}$ and acyl-$^{1}$GDVEKGKKIF$^{10}$. To evaluate the effectiveness of database searching to identify ARP derivatized carbonyl-modified peptides, mass spectrometry data was searched using Mascot with the DODE-ARP modification ($C_{24}H_{37}N_5O_7S$) added to the Mascot modification file.

The derivatization of protein carbonyls by ARP in effect doubles the size of the modification, increasing the tendency of the modification to undergo fragmentation by collision induced dissociation (CID). Figure 2 shows typical fragmentation patterns of the peptide acyl-$^{1}$GDVEKGKKIF$^{10}$, both unmodified (2A) and ARP-DODE modified (2B). In Figure 2B the ARP-tag-specific fragment ions, as well as peptide fragment ions with a neutral loss from the modification, are annotated. Figure 2C shows the ARP fragments (*m/z* 227, *m/z* 332, *m/z* 299 and *m/z* 159).

Initially it was observed that the ARP-specific fragments (*m/z* 227, *m/z* 332, *m/z* 299 and *m/z* 159) decreased the Mascot statistical score to the extent that even visually-good product ion spectra produced no positive hits. Following adjustment of the Mascot modification file to ignore these fragments, the peptides were identified correctly. Mascot scores of modified peptides, however, remained statistically low (M-score = 15-20), and it was therefore assumed that the ARP-specific fragments were not completely ignored, but were reduced in significance

while still taken into account during scoring. This type of interference has also been observed with the biotin-containing ICAT tag (31).

Another consequence of ARP fragmentation was the introduction of diagnostic neutral losses into the MS/MS spectra. To improve the peptide identification statistical score, the previously-noted neutral losses of 331 Da and 539 Da were added to the Mascot modification file. Neutral losses for y/b fragment ions and peptide neutral losses were also accounted for. Scores then increased to 25, 30 and 22 for the peptides $^{1}$GDVEKGK$^{*}$KIFI$^{10}$, $^{98}$LK$^{*}$KATNE$^{104}$ and $^{83}$AGIK$^{*}$KKTEREDLIAY$^{94}$ respectively. Unmodified peptides nonetheless consistently showed significantly higher scores (> 50); this discrepancy of scores between modified and unmodified peptides is explained as follows. The Mascot peptide score (Mowse score or M-score) is calculated on a probability that the observed match between the experimental data set and sequence database entry is a chance event or M-score = -10*log(P) where P is the probability, *i.e.*, the lower the probability the higher the score. The probability, P, is directly dependent on the number of theoretical fragment ions per given peptide, *i.e.*, the greater the number of fragments the higher the probability of a mismatch. The addition of neutral losses as well as modifications greatly increases the number of possible fragments. For example, Mascot calculated that the unmodified peptide GDVEKGKKIF would have 91 theoretical fragments; 24 ions were identified out of the 91 and the peptide received an M-score of 51. The equivalent DODE-ARP-modified peptide GDVEKGK$^{*}$KIF had a calculated 181 theoretical fragments due to neutral loses and the variable modification; 39 ions of the 181 were identified and the peptide received an M-score of 19. Assuming that both peptides (modified and unmodified) were treated identically, had similar mass spectra, similar number of background ions, similar fragmentation efficiency and similar peak list extractions, then scoring of the biotin modified peptides was

directly limited by Mascot's probability, P, calculation. Achieving the highest score for biotin-modified peptides therefore becomes a balancing act. The addition of neutral losses to the search parameters must not overly increase the theoretical number of fragment ions (and thus increase P) but be sufficient to correctly identify peptide fragments (resulting in lowered P). For the example above, since the peptide neutral loss accounted for only one ion in the MS/MS spectrum, it was removed from the modification file. Although this ion was no longer scored, the theoretical number of possible fragments decreased to 178, which slightly increased the peptide M-score to 20.

**The effect of modification on peptide M-Score.** To better understand the effect of the ARP-DODE modification on peptide identification, the doubly-charged unmodified and modified peptides (acyl-$^{1}$GDVEKGKKIF$^{10}$) were subjected to various collision energies, and Mascot peptide M-scores and fragmentation efficiencies were calculated. Fragmentation efficiency was defined as the ratio of the summed abundances of identified product ions divided by the total ion abundance. Since only the efficiency of peptide fragmentation was being measured, biotin/ARP specific fragment ions (227, 259, 299 and 332) and neutral losses (331 and 539) were excluded from these calculations. The results indicated that the collision energies required to elicit a correct Mascot score (Figure 3A) and for effective fragmentation (Figure 3B) were almost twice as large for ARP-DODE modified peptides than for the corresponding unmodified peptides. In fact, energies sufficient to fragment unmodified peptides produced little fragmentation in modified peptides, which remained unidentified (score = 0). An overall decrease in fragmentation efficiency of the modified peptide was observed. As previously seen, Mascot scores of modified peptides remained low compared to unmodified peptides. Scores for

modified peptides also declined at higher collision energies due to the increasing dominance of biotin/ARP label-specific fragments in the MS/MS spectra.

**Identification of ARP-DODE modified peptides in SJL mouse serum by Mascot.** Serum from RcsX-infected SJL mice serum was reacted with ARP, digested with trypsin and the modified peptides enriched by avidin affinity (Materials and Methods). Enriched peptides were analyzed by nano-LC-MS/MS and modified peptides were identified using Mascot (Materials and Methods). The ARP-derivatized lipid peroxidation adducts 4-hydroxy-2(E)-nonenal (HNE-ARP), 4-oxo-2(E)-nonenal (ONE-ARP) and 9,12-dioxo-10(E)-dodecenoic acid (DODE-ARP) were added to the Mascot modification file and searching parameters were adjusted to account for ARP fragments and neutral losses as previously described. The DODE modified peptide $^{549}$K$^{*}$QTALAELVK$^{558}$ was identified in mouse serum albumin (Figure 4A). The DODE modification was seen on lysine 549, causing a missed cleavage during tryptic digestion. The peptide was identified with a peptide M-score of 34, statistically low but similar to what has previously been seen with the DODE-ARP cyt c peptides. Peptide fragmentation resulted in complete series of y ions and a partial series of b ions, along with the neutral loss of the entire DODE-ARP modification from the peptide (*m/z* = 1100) and the neutral loss of ARP from b2 (*m/z* = 465). To further confirm this peptide modification, mouse serum albumin was reacted *in vitro* with DODE; reacted protein was derivatized with ARP, digested with trypsin and modified peptides enriched by avidin affinity (Materials and Methods). The DODE-ARP modified tryptic peptide K$^{*}$QTALAELVK was also identified, showing an identical mass spectrum fragmentation pattern, (Figure 4B). The expected ARP fragment ions (227, 259 and 299) were identified in both the endogenously modified peptide and the *in vitro* DODE-ARP modified peptide, verifying ARP derivatization of these peptides.

Analysis of the raw mass spectrometry data clearly showed an abundance of ARP derivatized peptides as seen by the identification of the ARP tag specific fragment ions, but further database searching by Mascot did not reveal other peptides that could be verified as modified by the carbonyls HNE, ONE and DODE. As previously described, the ARP fragments and neutral losses were artificially decreasing the peptide M-scores, introducing a significant increase in the number of false positives, thus requiring manual verification of possible hits. Real database hits were easily lost in the hundreds of low scoring spectra that were incorrectly identified as carbonyl-modified. Manual searching of all spectra proved unrealistic, and it was also probable that lower-abundant modified peptides would potentially remain un-scored and therefore unidentified due to the interfering effects of the ARP fragments and neutral losses (previously demonstrated with the DODE-ARP cyt c standards). Therefore a new method for the identification of these ARP derivatized peptides was explored.

***In silico* filter program for the identification of ARP derivatized peptides.** As previously noted, ARP-specific fragments and neutral losses (Figure 2C) can inhibit successful database searching and increase the number of false positives. These same values, however, constitute a potential fingerprint to locate the ARP-carbonyl modified peptides within the MS/MS raw data. Manual extraction of these ions from MS/MS raw data is possible but time-consuming and impractical, therefore a program was written (Materials and Methods) to search for these signals in the MS/MS peak lists generated from data-dependent tandem mass spectrometry experiments: the i*n silico* filter program. A general outline of the program is shown in Figure 5. The raw data is first converted to peak lists using the Mascot script associated with the Analyst software. This built-in script converts the raw mass spectrometry data into text files (mgf files or Mascot Generic Files) containing the parent ions identified by the

mass spectrometer along with their associated CID fragment ions and intensities (see http://www.matrixscience.com/help/data_file_help.html). The *in silico* filter program scans the peak lists within an mgf file to identify lists that contain the ARP fragment ions (227.0854, 332.1387, 299.1127 and 259.1223). The program also searches for parent ions with a neutral loss of 331.1344; corresponding to the ARP (Figure 2). It then creates a set of five new filtered mgf files. The first of these (filter file #1) is made up of peak lists containing the ion at *m/z* 227, corresponding to the most-frequent biotin fragment. The second (filter file #2) contains peak lists with *m/z* 227 plus one other fragment ion (*m/z* 332 or 299 or 259). Peak lists in the third file (filter file #3) contain *m/z* 227 plus two or more fragment ions. The fourth file (filter file #4) required peak lists to contain all four fragment ions. Finally, the fifth filtered file (filter file #5) requires the peak lists to contain all four fragment ions and the parent ion with the neutral loss of 331. These five filtered mgf files allow the user to choose the extent of filtration with the first file corresponding to the least filtered and the fifth file to the most filtered. In general, the presence of three or more fragment ions (filter file #3) was sufficient to mark a peptide as modified with a high degree of confidence.

Fragmentation of the peptide backbone occasionally resulted in the formation of some ARP biotin tag fragment ions (227.0854, 332.1387, 299.1127 and 259.1223). To increase the specificity for these ions, the user could enter an accurate-mass value between ± 1.0 Da and 0.0001 Da. Values in the range of ± 0.01 - 0.001 Da would generally exclude any ions that were associated with peptide fragmentation rather than ARP fragmentation. For these experiments, all tandem mass spectrometry runs contained an internal standard allowing for a mass accuracy within 5 ppm (about ± 0.001 Da).

To prevent the program from selecting apparent ARP fragment ions from the noise of the MS/MS spectra, the program calculates the maximum peak intensity for the ions in each individual peak list. The user then enters the desired percent of this maximum peak intensity to determine a threshold above which ARP-fragments are considered to be present. By inspection, we noted that for ARP derivatized peptides this maximum ion was usually the *m/z* 227 biotin fragment. In general a threshold of 10% of the maximum peak intensity will exclude ions associated with the noise, but allow the program to identify some of the lower-abundance authentic ARP fragments.

The ability of the filter program to highlight only ARP derivatized peptides was tested with an ARP-DODE cytochrome c standard that was digested with chymotrypsin and analyzed by LC-MS/MS. The raw mass spectrometry data was extracted into an mgf peak list file and filtered using the *in silico* filter program with mass accuracy set to ± 0.01 Da and a threshold of 10% of maximum peak intensity for fragment ions. The mgf output files as well as the unfiltered mgf file were then analyzed with Mascot. Table 1A shows the Mascot results from the unfiltered mgf file and the filter file #3 (requiring three or more ARP fragment ions). The unfiltered file resulted in the identification of cyt c and its DODE-ARP modified peptides as well as many unmodified peptides of cyt c, chymotrypsin and keratin. The unfiltered file also included one false-positive modified peptide in chymotrypsin. The search of the filtered data identified only the modified peptides (Table 1B), with the notable removal of the chymotrypsin false positive. None of the DODE-ARP cyt c peptides were lost due to the filtering process. Searching with filter file #5 (requiring all five ARP associated ions) reduced the results to three modified peptides, while searching with filter file #1 (requiring the single ARP fragment at *m/z* 227 ) only slightly increased the number of unmodified peptides. It was concluded that a

minimum of three ARP fragment ions with abundances above 10% of the maximum ion was sufficient to filter modified peptides. A corresponding study using DODE-ARP modified BSA is included in the Supplementary Material; Figure 2.

In summary, the *in silico* filter program highlighted only carbonyl-ARP modified peptides and significantly reduced the complexity of the mgf file submitted for database searching. Three ARP fragment ions were sufficient to filter modified peptides from unmodified peptides and an intensity for ARP fragment ions of 10% of the maximum ion intensity as well as a mass accuracy window of ±0.001 Da gave a high degree of confidence that filtered peptides were modified. Submitting a filtered mgf file for database searching reduced the occurrence of false positive hits since only modified peptides were searched. The *in silico* filter program code is available to view or to download at: http://web.mit.edu/toxms/www/filters.htm.

**Biotin Peptide Identification Program.** The *in silico* filter program output files for the DODE-ARP cyt c standard (described above) highlighted a significant number of modified peptides (or peak lists) that remained unidentified by Mascot, therefore a *de novo*-like sequencing database search algorithm was written: the biotin peptide identification program (BPI). The BPI program was written to search the mgf files generated from the *in silico* filter program, and identify the modified peptides, without prior knowledge of the molecular weight of the carbonyl modification. This enabled searching at one time the many possible variations of protein carbonyls.

To run the BPI program, the user first enters a desired modification mass range (from ±0.0001 Da to large as necessary). The program then selects a peak list from the filtered mgf file. To improve identification specificity, the fragment ion masses in the peak list are calibrated using the known masses of the ARP fragment ions (single or multi point calibration as specified

by user). The program then calculates a background intensity level of the peak list being searched to prevent consideration of ions in the noise of the spectrum (further information on background calculation can be seen at http://web.mit.edu/toxms/www/filters.htm). The BPI program virtually digests (tryptic or chymotryptic) the protein database being searched. The parent ion from the peak list being searched is compared to each peptide mass of the virtually digested peptides from the database. If the differences in mass are within the specified modification mass range, this difference is saved as a potential carbonyl modification mass for that peptide. The peptide is identified by comparing the y and b ions of the virtual peptide with the experimental peak list fragmentation ions as described below.

For a match to be significant, the program first calculates the percent coverage of unmodified y and b fragment ions identified in the experimental peak list. If this is lower than 25-50% (user specified) the peptide is rejected. (Since the peak lists being searched are generated from the *in silico* filter program, they are all considered to be modified with at least one carbonyl, therefore 50% is the maximum possible coverage for unmodified y and b ions.) Fragment ions from the experimental peak lists of un-rejected peptides (percent coverage sufficiently high) are compared a second time to the virtual peak list fragment ions; this second comparison includes the calculated modification mass. A peptide is considered a 'hit' if a minimum of five sequential y or b ions are identified. For example, y3, y4, y5, y6 and y7 ions constitute a hit while y3, y4, y5, y7 and y8 do not. The user can specify between 0 and 5 sequential ions to allow for more or less rigor. This two-fold process (% coverage followed by sequential ions) for peptide identification is necessary to limit false positives, since the modification mass is unknown. Once the peptide is identified, the BPI program will back-calculate the mass of the potential protein carbonyl modification by subtracting the appropriate

mass of the ARP biotin tag and output the peptide, the protein and the calculated carbonyl mass. A general outline for the BPI program can be seen in Figure 5.

The BPI program is written to identify carbonyl-modified peptides that are not being identified by Mascot. To test this, the tandem mass spectrometry mgf files generated from the ARP-DODE cytochrome c standard were filtered using the *in silico* filter program (described above). Filter file #3 was searched by the BPI program and Mascot. To save time, a minimal database, containing only cyt c and serum albumin, was used. The results are shown in Table 2. As described above and shown in Table 3A, Mascot successfully identified the three known ARP-DODE modified peptides $^{1}$GDVEKGK$^{*}$KIFI$^{10}$, $^{98}$LK$^{*}$KATNE$^{104}$ and $^{83}$AGIK$^{*}$KKTEREDLIAY$^{94}$. The BPI program identified these three peptides as well as seven other ARP-DODE modified peptides (Table 2B). These previously unidentified peptides consisted of missed cleavages and secondary cleavages for chymotrypsin as well as the peptide $^{61}$K$^{*}$EETL$^{65}$, which has also been identified as a minor DODE modification site of cyt c *(24)*.

To further test the ability of the BPI program to identify the correct carbonyl mass on a modified peptide, the peak list for the ARP-DODE modified peptide $^{1}$GDVEKGK$^{*}$KIF$^{10}$, generated from the ARP-DODE cyt c standard (described above), was searched against the entire NCBI *Equus caballus* (horse) database. The previously described lysine modification was known to be ARP-DODE, equivalent to an additional molecular weight of +539 Da. The input parameters for the BPI program were as follows: digestion enzyme: chymotrypsin; number of miss-cleavages: 0; N-acetyl protein modification: yes; background level: 2; modification mass range: 300-800 Da; amino acid/s modified: K, H and C; fragment ions tolerance: ±0.01 Da. The program identified the peptide and correctly calculated the DODE modification of lysine (+ 539 Da). The output file generated only two false positive hits (Table 3). A similar result for the

ARP-DODE modified peptide $^{95}$LKK$^{*}$ATNE$^{101}$, generated from the cyt c standard, is summarized in the Supplementary Materials section; Figure 3. Considering that the peptides were searched against the entire NCBI horse database of known proteins and the modification mass range was set at ±500 Da these results demonstrate the effectiveness of the evaluation criteria in the BPI program. The Python program code is available to view or for download at: http://web.mit.edu/toxms/www/filters.htm.

**Identification of carbonyl modified peptides in SJL mouse serum by database search with the BPI program.** To improve the outcome of the BPI program, a searchable database was made using the proteins previously identified as carbonyl modified (Supplemental Table ST1). This database served two purposes, first the BPI program is slow and the use of a smaller database would enable more searches. Second, peptides identified by the program from this database had a higher likelihood of being a true hit since the proteins were previously identified as carbonyl modified. To identify the carbonyl modification, the mgf files from the RcsX infected SJL mouse serum and control mouse serum (described above) were filtered using the *in silico* filter program. The program conditions used a mass accuracy of ± 0.01 Da and 10% maximum peak intensity for fragment ions. As described above, the presence of three major ARP fragment ions in a spectrum was sufficient to select a peptide as modified by the filter program, so the filter file #3 generated by the *in silico* filter was submitted to the BPI program for database searching. The BPI program conditions for database searching were as follows: protease: trypsin; missed cleavages: 1; fragment ion mass accuracy: ± 0.01 Da; N-terminus acetyl modification: no; variable modification oxidized methionine: yes; percent coverage of un-modified fragment ions in spectrum: 35%; number of sequential fragment ions required in spectrum: 5; low mass range: 329; high mass range: 800; background intensity level: 1. The low

mass range was chosen so that the peptides identified would at least contain a modification equivalent to the molecular weight of ARP. The high mass range of 800 was considered sufficient to identify the larger carbonyl modification such as DODE. The program was run to identify the lipid peroxidation carbonyl modifications on amino acids K, H and C; the direct side-chain oxidation of residues P, R and K to the semialdehyde carbonyls, and to identify the theoretical oxidation of O and N linked glycans at residues S, T and N, and was configured to search *in silico* filtered data with a modification mass range of 300-800 Da. Since the oxidation of P, R and K to form the semialdehydes can reduce the molecular weight of the peptide the low mass of 300 was used instead of 331 (mw of ARP).

The search results on the RcsX infected SJL mouse serum peptides are shown in 4. The MS/MS spectra of all peptide hits were verified manually and these are included in the Supplementary Materials section. As previously seen by Western blot (Figure 1), albumin was highly carbonyl-modified. The DODE-modified peptide $^{549}$K$^{*}$QTALAELVK$^{560}$, previously identified by Mascot, was identified, and the albumin peptide $^{546}$QIK$^{*}$K$^{549}$ was also identified as DODE modified. The modification of both lysines 548 and 549 was not surprising since runs of lysines have previously been seen to be modified by DODE in our *in vitro* experiments (30). Albumin was also modified at peptide $^{222}$MK$^{*}$CSSMQK$^{229}$; this modification corresponds to oxidation of the lysine 223 side chain to the aminoadipic semialdehyde (scheme 1). A fourth carbonyl modification was found on the albumin peptide $^{234}$AFK$^{*}$AWAVAR$^{242}$; the calculated carbonyl modification mass was + 112 Da and identified on lysine 236. This carbonyl adduct potentially corresponds to the lipid peroxidation product 2-heptenal. Haptoglobin was modified at peptide $^{112}$GSFPWQAK$^{*}$M$^{*}$ISR$^{123}$. The modification of lysine 119 corresponded to direct oxidation to the aminoadipic semialdehyde and the modification of methionine 120 as the

oxidation product of +16, which had been added as a variable modification during the BPI search. An ONE lipid peroxidation carbonyl modification was identified on lysine 166 of the hemopexin peptide $^{166}$K$^{*}$WFWDFATR$^{174}$. Carbonyl modifications were also identified on hemopexin at peptide $^{103}$GPDSVFLIK$^{*}$EDK$^{114}$ with the oxidation of lysine 111 to the aminoadipic semialdehyde and at peptide $^{78}$GHSGTR$^{*}$ELISAR$^{89}$ with the oxidation of arginine 83 to the glutamic semialdehyde. An HNE lipid peroxidation carbonyl modification was identified at hystidine 335 on the inter-alpha trypsin inhibitor, heavy chain 3 peptide $^{334}$DH$^{*}$LVQATPANLK$^{345}$. An HNE modification was found at histidine 431 of the muranoglobulin-1 peptide $^{431}$H$^{*}$ASAK$^{435}$. A second HNE carbonyl modification was identified at histidine 856 on the muranoglobulin-1 or muranoglobulin-2 peptide $^{856}$H$^{*}$TSSWLVTPK$^{865}$. Pregnancy zone protein (or alpha-2-macroglobulin precursor) was seen to be modified by the oxidation of lysine 306 to the aminoadipic semialdehyde on peptide $^{305}$TK$^{*}$VFQLR$^{311}$ and with a possible 2-heptinal modification at lysine 1205 on peptide $^{1204}$VK$^{*}$ALSFYQPR$^{1213}$. Transferrin was modified by the oxidation of proline 164 to the glutamic semialdehyde carbonyl on the peptide $^{163}$SP*LEK$^{167}$; by the oxidation of lysine 299 to the aminoadipic semialdehyde on the peptide $^{298}$SK$^{*}$DFQLFSSPLGK$^{310}$; by the modification of hystidine 268 with the lipid peroxidation adduct HNE on the peptide $^{265}$IPSH$^{*}$AVVAR$^{273}$; and by the oxidation of either lysine 274 or 278 to the aminoadipic semialdehyde carbonyl on the peptide $^{274}$K$^{?}$NNGK$^{?}$EDLIWEILK$^{287}$ (the exact modification site could not be determined).

As shown in Table 4, carbonyl modified peptides were found on a number of proteins from the database, including alpha-2-HS-glycoprotein, alpha-2-macroglobulin apolipoprotein A1, apolipoprotein E, ceruloplasmin isoform a and b, compliment component 3, esterase 1 and serine (or cysteine) proteinase inhibitor, although the specifically modified peptides could not be

identified. The inability to identify a carbonyl modification on Apo A1 was surprising considering that its modification by HNE has been reported (32).

**Discussion**

Carbonyl modified proteins have proven difficult to analyze, and only a few studies have correctly identified the peptides from endogenously-modified proteins. In this study a number of protein carbonyls were identified in serum of the SJL mouse inflammation model. Modified proteins and peptides were identified using biotin/avidin affinity in conjunction with new software developed in-house that filtered and searched spectra that were not identified during Mascot database searching.

Mascot is a powerful tool for database searching, quantification and identification of protein modifications. Large modifications, however, can fragment during MS/MS CID and these can interfere with Mascot searches (23,33). Regardless of spectral quality, the ARP-biotin modified peptides will generate Mascot scores that are lower than those for the corresponding unmodified peptides. To identify modified peptides the researcher must therefore lower the statistical cut-off point at which spectra are accepted or rejected, consequently leading to an increase in the number of false positives. Furthermore, the low score requires that all identified peptides be manually or visually verified. Even relatively simple systems (such as the ARP-DODE-modified cytochrome c standard used here) become increasingly difficult to analyze and result in many low-scoring identifications, the majority of which are artifacts. For an *in vivo* sample, these problems are compounded by the complexity of the sample and the sample processing, resulting in an inevitable increase in the generation of false positives.

Two programs were written to help overcome some of the problems associated with database searching and identifying biotinylated carbonyl-modified proteins in the SJL mouse serum. The *in silico* filter program takes advantage of the biotin-ARP modification-specific ions that are formed during peptide fragmentation. The same ions that impede effective peptide identification can be used as a fingerprint to highlight modified peptides in the mass spectrometry data. Using this program, modified peptides are filtered *in silico* from the mass spectrometry data, resulting in a new data file containing only carbonyl modified peptides. Mascot database searching of the filtered files result in a reduced the number of false positives, simplifying manual verification of modified peptides.

To improve the identification of the *in silico* filtered peptides, other types of evaluation criteria were explored using an in-house database search algorithm (biotin peptide identification program, BPI) that evaluates the theoretical number of unmodified peptide fragments (y and b ions) and the number of sequentially identified modified and unmodified peptide y and b ion fragments (the longer the sequence the lower the probability of a mismatch). The peptides are searched in this *de novo*-like fragment-dependent manner resulting in modified peptide identification without the necessity of knowing the carbonyl modification mass. The calculated modification mass is part of the BPI software output along with the peptide/protein identification. The program identifies ARP-carbonyl modified peptide standards from a mammalian database, and calculates the correct modification mass with a minimal number of false identifications.

The BPI program was written as a specialized database search engine for identifying biotin-derivatized peptides from data generated by the *in silico* filter program. It was also designed to use a simplified database of known modified proteins. It therefore does not calculate

probability scores for peptides since output is kept to a minimum and can easily be manually verified. Without probability scores, the program does not bias searches with scoring criteria and is able to identify modified peptides not seen during Mascot searching. Although the program allows complete access to the data analysis portion of this research, it is relatively inflexible with respect to its use as a general database search engine, since such searching requires good statistical peptide scoring as well as an ability to search large quantities of data in a relatively short time.

It should be stated that both the *in silico* filter and BPI programs were written to be a semi-empirical approach to identifying spectra from modified peptides. The programs identify spectra that correspond to a few modified peptides, making manual verification possible. These programs use no statistical calculations to limit the number of false positives and manual verification is required. The BPI and the *in silico* filter were written to help with the identification of peptides with large biotinylated modifications and are therefore somewhat specialized. A referee noted that comparisons of the approach with different instrument types, e.g., ion-traps (typically richer in b-ions) *vs* quadrupole time-of-flight, could expand the range and usefulness of the software for identification of other large peptide modifications. Both program codes are available at http://web.mit.edu/toxms/www/filters.htm.

The primary goal of this study was to identify carbonyl protein modifications in serum from the SJL mouse. A number of modifications, including HNE, ONE, DODE and a potential 2-heptenal modification, were identified from the lipid peroxidation pathway. HNE modification of histidine was identified on three proteins (Table 4), while DODE and ONE modifications were identified only on albumin and hemopexin respectively. This result is somewhat contrary to our previous studies in which DODE was the most reactive aldehyde formed from the

hydroperoxide of linoleic acid (30). A more-abundant HNE modification *in vivo,* however, may not be not surprising since a lower reactivity would give the HNE aldehyde a longer half-life, allowing more opportunities for protein modification.

Carbonylation also arose through the direct oxidation of amino acid side chains of proline, lysine or arginine, and the formation of the aminoadipic and glutamic semialdehydes. These modifications were identified on albumin, haptoglobin, pregnancy zone protein and transferrin, and have often been considered a major source of protein carbonyls occurring from the production of free radicals by copper or iron Fenton chemistry. Transferrin and hemopexin were highly modified with semialdehyde carbonyls, with hemopexin containing the only oxidized arginine identified in the study. This is an interesting result considering that the biological function of these proteins involves the sequestering and transport of iron, likely making them more susceptible to Fenton chemistry.

As shown in Table 4, the modified peptides for a number of candidates for protein carbonyl modification remained unidentified. These were marked as "potential DODE" modifications if the protein was identified from an anti-DODE western gel or as "unknown carbonyl" if the protein was identified after biotin/avidin enrichment. The identities and potential carbonyl modification of these proteins will be addressed in future experiments.

**Acknowledgments.**

This work was supported by grants from the National Institutes of Health (NCI Program Project Grant CA26731), and the MIT Center for Environmental Health Sciences (NIEHS grant P30 ES002109). Thanks to Dr. Ioannis Pappayanopoulos for help with the intricacies of Mascot, and Mr. Peter A. Karasev (Georgia Institute of Technology) for contributions to the code.

## References


1. Grimsrud, P.A., Xie, H., Griffin, T. J., Bernlohr, D. A. (2008) Oxidative stress and covalent modification of protein with bioactive aldehydes. *J.Biol.Chem.,* **283,** 21837-21841.

2. Dalle-Donne, I., Giustarini, D., Colombo, R., Rossi, R., Milzani, A. (2003) Protein carbonylation in human diseases. *Trends Mol.Med.,* **9,** 169-176.

3. Butterfield, D.A., Kanski, J. (2001) Brain protein oxidation in age-related neurodegenerative disorders that are associated with aggregated proteins. *Mech.Ageing Dev.,* **122,** 945-962.

4. Buss, I.H., Darlow, B. A., Winterbourn, C. C. (2000) Elevated protein carbonyls and lipid peroxidation products correlating with myeloperoxidase in tracheal aspirates from premature infants. *Pediatr.Res.,* **47,** 640-645.

5. Blackburn, A.C., Doe, W. F., Buffinton, G. D. (1999) Protein carbonyl formation on mucosal proteins in vitro and in dextran sulfate-induced colitis. *Free Radic.Biol.Med.,* **27,** 262-270.

6. Chaudhuri, A.R., de Waal, E. M., Pierce, A., Van Remmen, H., Ward, W. F., Richardson, A. (2006) Detection of protein carbonyls in aging liver tissue: A fluorescence-based proteomic approach. *Mech.Ageing Dev.,* **127,** 849-861.

7. Serdar, Z., Serdar, A., Altin, A., Eryilmaz, U., Albayrak, S. (2007) The relation between oxidant and antioxidant parameters and severity of acute coronary syndromes. *Acta Cardiol.,* **62,** 373-380.

8. Requena, J.R., Chao, C. C., Levine, R. L., Stadtman, E. R. (2001) Glutamic and aminoadipic semialdehydes are the main carbonyl products of metal-catalyzed oxidation of proteins. *Proc.Natl.Acad.Sci.U.S.A,* **98,** 69-74.

9. Yuan, Q., Zhu, X., Sayre, L. M. (2007) Chemical nature of stochastic generation of protein-based carbonyls: metal-catalyzed oxidation versus modification by products of lipid oxidation. *Chem Res.Toxicol.,* **20,** 129-139.

10. Sayre, L.M., Lin, D., Yuan, Q., Zhu, X., Tang, X. (2006) Protein adducts generated from products of lipid oxidation: focus on HNE and ONE. *Drug Metab Rev,* **38,** 651-675.

11. Refsgaard, H.H., Tsai, L., Stadtman, E. R. (2000) Modifications of proteins by polyunsaturated fatty acid peroxidation products. *Proc.Natl.Acad.Sci.U.S.A,* **97,** 611-616.

12. Pang, B., Zhou, X., Yu, H., Dong, M., Taghizadeh, K., Wishnok, J. S., Tannenbaum, S. R., Dedon, P. C. (2007) Lipid peroxidation dominates the chemistry of DNA adduct formation in a mouse model of inflammation. *Carcinogenesis*.

13. Rifai, N., Gillette, M. A., Carr, S. A. (2006) Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat.Biotechnol.,* **24,** 971-983.

14. Rouzer, C.A., Marnett, L. J. (2005) Structural and functional differences between cyclooxygenases: fatty acid oxygenases with a critical role in cell signaling. *Biochem.Biophys.Res.Commun.,* **338,** 34-44.

15. West, J.D., Marnett, L. J. (2005) Alterations in gene expression induced by the lipid peroxidation product, 4-hydroxy-2-nonenal. *Chem Res.Toxicol.,* **18,** 1642-1653.

16. Marnett, L.J., Riggins, J. N., West, J. D. (2003) Endogenous generation of reactive oxidants and electrophiles and their reactions with DNA and protein. *J Clin Invest,* **111,** 583-593.

17. Pelissier, M.A., Marteau, P., Pochart, P. (2007) Antioxidant effects of metronidazole in colonic tissue. *Dig.Dis.Sci.,* **52,** 40-44.

18. Mercier, S., Breuille, D., Mosoni, L., Obled, C., Patureau, M. P. (2002) Chronic inflammation alters protein metabolism in several organs of adult rats. *J Nutr.,* **132,** 1921-1928.

19. El Yousfi, M., Breuille, D., Papet, I., Blum, S., Andre, M., Mosoni, L., Denis, P., Buffiere, C., Obled, C. (2003) Increased tissue protein synthesis during spontaneous inflammatory bowel disease in HLA-B27 rats. *Clin Sci.(Lond),* **105,** 437-446.

20. Keshavarzian, A., Banan, A., Farhadi, A., Komanduri, S., Mutlu, E., Zhang, Y., Fields, J. Z. (2003) Increases in free radicals and cytoskeletal protein oxidation and nitration in the colon of patients with inflammatory bowel disease. *Gut,* **52,** 720-728.

21. Lih-Brody, L., Powell, S. R., Collier, K. P., Reddy, G. M., Cerchia, R., Kahn, E., Weissman, G. S., Katz, S., Floyd, R. A., McKinley, M. J., Fisher, S. E., Mullin, G. E. (1996) Increased oxidative stress and decreased antioxidant defenses in mucosa of inflammatory bowel disease. *Dig.Dis.Sci.,* **41,** 2078-2086.

22. Gal, A., Tamir, S., Tannenbaum, S. R., Wogan, G. N. (1996) Nitric oxide production in SJL mice bearing the RcsX lymphoma: A model for *in vivo* toxicological evaluation of $NO^{\cdot}$ *Proc.Natl.Acad.Sci.U.S.A.,* **93,** 11499-11503.

23. Han, B., Stevens, J. F., Maier, C. S. (2007) Design, synthesis, and application of a hydrazide-functionalized isotope-coded affinity tag for the quantification of oxylipid-protein conjugates. *Anal.Chem,* **79,** 3342-3354.

24. Slade, P.G., Williams, M. V., Brahmbhatt, V., Dash, A., Wishnok, J. S., Tannenbaum, S. R. (2010) Proteins Modified by the Lipid Peroxidation Aldehyde 9,12-Dioxo-10(E)-dodecenoic Acid in MCF7 Breast Cancer Cells. *Chem.Res.Toxicol.,* **23,** 557-567.

25. Chavez, J., Wu, J., Han, B., Chung, W. G., Maier, C. S. (2006) New role for an old probe: affinity labeling of oxylipid protein conjugates by N'-aminooxymethylcarbonylhydrazino d-biotin. *Anal.Chem,* **78,** 6847-6854.

26. Chung, W.G., Miranda, C. L., Maier, C. S. (2008) Detection of carbonyl-modified proteins in interfibrillar rat mitochondria using N'-aminooxymethylcarbonylhydrazino-D-biotin as an aldehyde/keto-reactive probe in combination with Western blot analysis and tandem mass spectrometry. *Electrophoresis,* **29,** 1317-1324.

27. Ide, H., Akamatsu, K., Kimura, Y., Michiue, K., Makino, K., Asaeda, A., Takamori, Y., Kubo, K. (1993) Synthesis and damage specificity of a novel probe for the detection of abasic sites in DNA. *Biochemistry,* **32,** 8276-8283.

28. Gal, A., Tamir, S., Kennedy, L. J., Tannenbaum, S. R., Wogan, G. N. (1997) Nitrotyrosine formation, apoptosis and oxidative damage: relationships to nitric oxide production in SJL mice bearing the RcsX tumor. *Cancer Res.,* **57,** 1823-1828.

29. Slade, P.G., Williams, M. V., Brahmbhatt, V., Dash, A., Wishnok, J. S., Tannenbaum, S. R. (2010) Proteins modified by the lipid peroxidation aldehyde 9,12-dioxo-10(E)-dodecenoic acid in MCF7 breast cancer cells. *Chem.Res.Toxicol.,* **23,** 557-567.

30. Williams, M.V., Wishnok, J. S., Tannenbaum, S. R. (2007) Covalent adducts arising from the decomposition products of lipid hydroperoxides in the presence of cytochrome c. *Chem Res.Toxicol.,* **20,** 767-775.

31. Borisov, O.V., Goshe, M. B., Conrads, T. P., Rakov, V. S., Veenstra, T. D., Smith, R. D. (2002) Low-energy collision-induced dissociation fragmentation analysis of cysteinyl-modified peptides. *Anal.Chem,* **74,** 2284-2292.

32. Szapacs, M.E., Kim, H. Y., Porter, N. A., Liebler, D. C. (2008) Identification of proteins adducted by lipid peroxidation products in plasma and modifications of apolipoprotein A1 with a novel biotinylated phospholipid probe. *J Proteome.Res.,* **7,** 4237-4246.

33. Roe, M.R., McGowan, T. F., Thompson, L. V., Griffin, T. J. (2010) Targeted 18O-labeling for improved proteomic analysis of carbonylated peptides by mass spectrometry. *J Am.Soc.Mass Spectrom.,* **21,** 1190-1203.

**Legends.**

Scheme 1. Direct oxidative carbonylation of proteins to form glutamate and aminoadipate semialdehydes.

Scheme 2. Reactive aldehydes, arising from oxidation of polyunsaturated fatty acids and carbohydrates, can indirectly lead to protein carbonylation

Figure 1. Western blot analysis of ARP-derivatized and non-derivatized serum proteins from control and infected SJL mice.

Figure 2: MS/MS of unmodified (A) and DODE-ARP-modified (B) peptide. (C) Major CID fragments of the ARP tag.

Figure 3**:** Collision induced dissociation analysis of the unmodified (red) and modified (black) peptide, acyl-GDVEKGK$^{*}$KIF. (A) Mascot peptide score with increasing collision energy. (B) Fragmentation efficiency (defined as the ratio of the summed abundances of identified product ions divided by the total ion abundance).

Figure 4: ARP-DODE modified peptide (K$^{*}$QTALAELVK) identified on mouse serum albumin. (A) From infected mouse serum. (B) From mouse serum albumin treated in vitro with DODE.

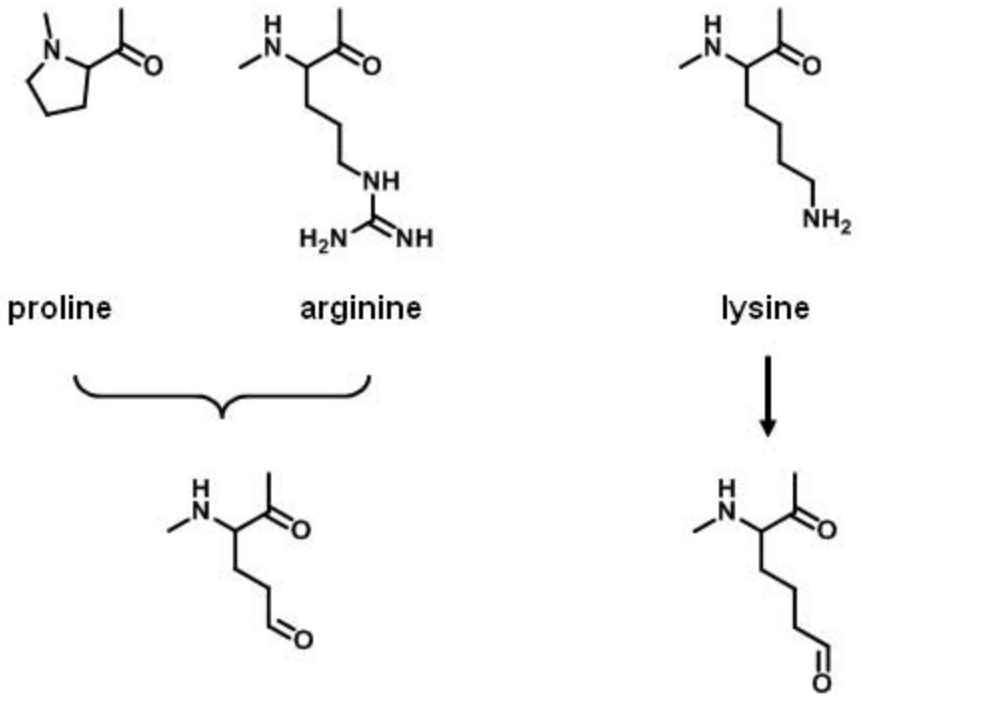
Figure 5: Software workflow for the identification of biotin derivatized protein carbonyls.

Table 1: Mascot search results on DODE-ARP modified BSA and cyt c standards with and without the *in silico* filter. A: Pre-filter; multiple peptides are identified both modified and unmodified. B: Post filter; all un-modified peptides are eliminated and only modified peptides are identified along with the elimination of one false positive.

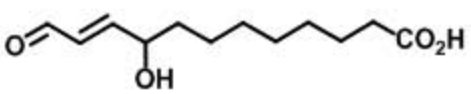
Table 2: Search results on DODE-ARP modified cyt C and BSA, after *in silico* filtration. A: Mascot search results. B: BPI search results. BPI is able to identify a number of low abundant modified peptides that remain un-scored by Mascot. Many of these peptides are secondary cleavage sites of chymotrypsin.

Table 3: Output file from BPI program after searching the ARP-DODE modified peptide GDVEKGK$^{*}$KIF against the entire NCBI *Equus caballus* database. The modification and cytochrome c were correctly identified along with two false positives.

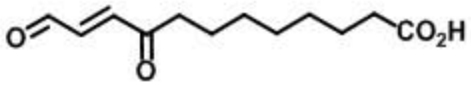
Table 4: Carbonyl modified proteins in SJL mouse serum.


proline
arginine
lysine
glutamate semialdehyde
aminoadipate semialdehyde

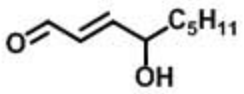
Scheme 1



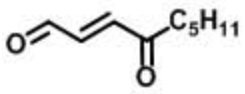

9-hydroxy-12-oxododec-10-enoic acid (HODE)
MW: 228.1362




9,12-dioxododec-10-enoic acid (DODE)
MW: 226.1205

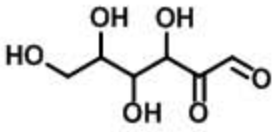



4-hydroxynon-2-enal (HNE)
MW: 156.1150

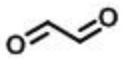



4-oxynon-2-enal (ONE)
MW: 154.0994

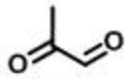

hept-2-enal (ONE)
MW: 112.0888



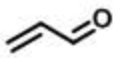

3-deoxyglucosone
MW: 178.0477




glyoxal
MW: 56.0055

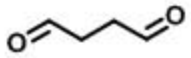



methylglyoxal
MW: 72.0211

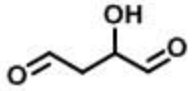



acrolein
MW: 56.0262


malondialdehyde
MW: 72.0211




succinaldehyde
MW: 86.0368




2-hydroxysuccinaldehyde
MW: 102.0317

Scheme 2

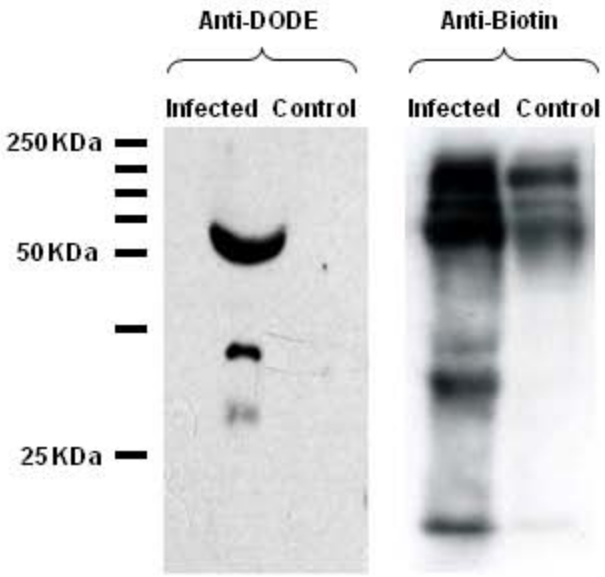
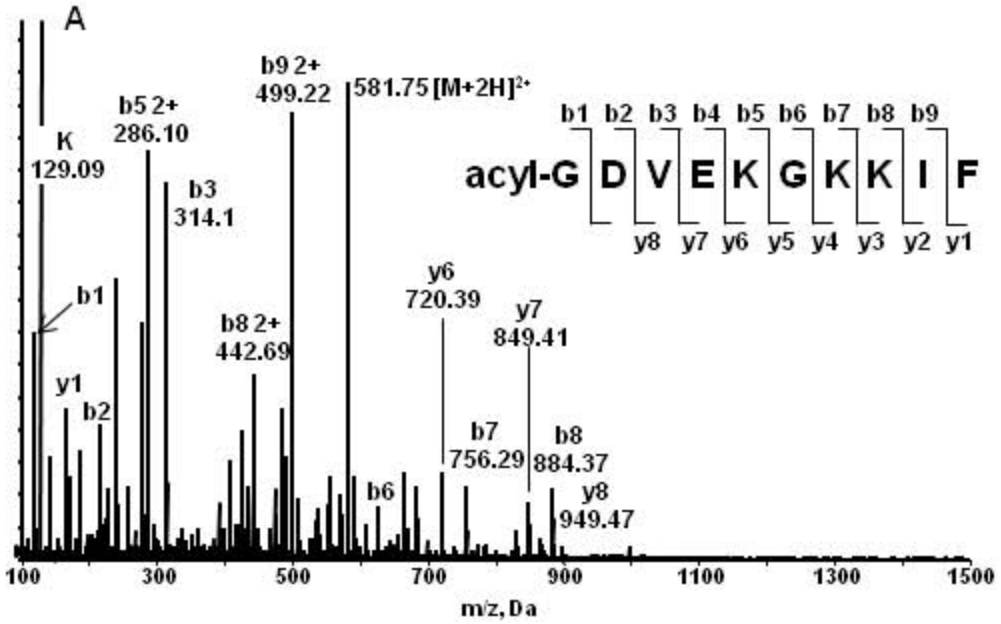
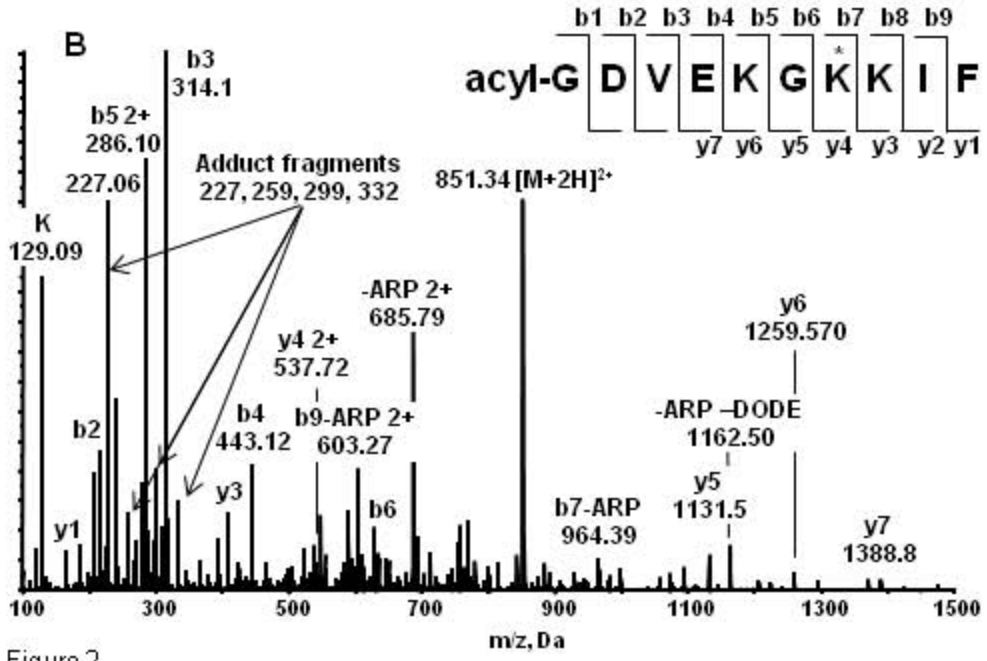

Anti-DODE
Anti-Biotin
Infected
Control
Infected
Control
250KDa
50KDa
25KDa

Figure 1


A
b9 2+
499.22
581.75 [M+2H]2+
b5 2+
286.10
K
129.09
b3
314.1
b1
y6
720.39
y7
849.41
b8 2+
442.69
y1
b2
b7
756.29
b8
884.37
b6
y8
949.47
100
300
500
700
900
1100
1300
1500
m/z, Da
b1 b2 b3 b4 b5 b6 b7 b8 b9
acyl-G D V E K G K K I F
y8 y7 y6 y5 y4 y3 y2 y1


B
b1 b2 b3 b4 b5 b6 b7 b8 b9
acyl-G D V E K G K* K I F
y7 y6 y5 y4 y3 y2 y1
b3
314.1
b5 2+
286.10
227.06
Adduct fragments
227, 259, 299, 332
851.34 [M+2H]2+
K
129.09
-ARP 2+
685.79
y6
1259.570
y4 2+
537.72
b4
443.12
b9-ARP 2+
603.27
-ARP –DODE
1162.50
b2
y5
1131.5
y3
b6
b7-ARP
964.39
y7
1388.8
y1
100
300
500
700
900
1100
1300
1500
m/z, Da

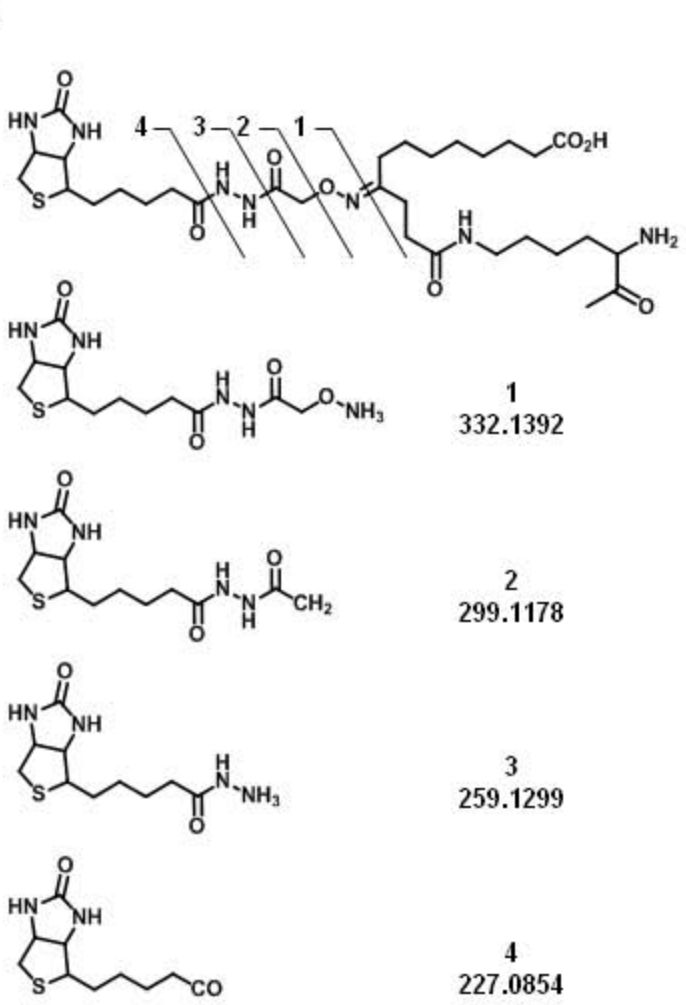

C
4
3
2
1
CO2H
NH2
NH3
CH2
CO
1
332.1392
2
299.1178
3
259.1299
4
227.0854

Figure 2

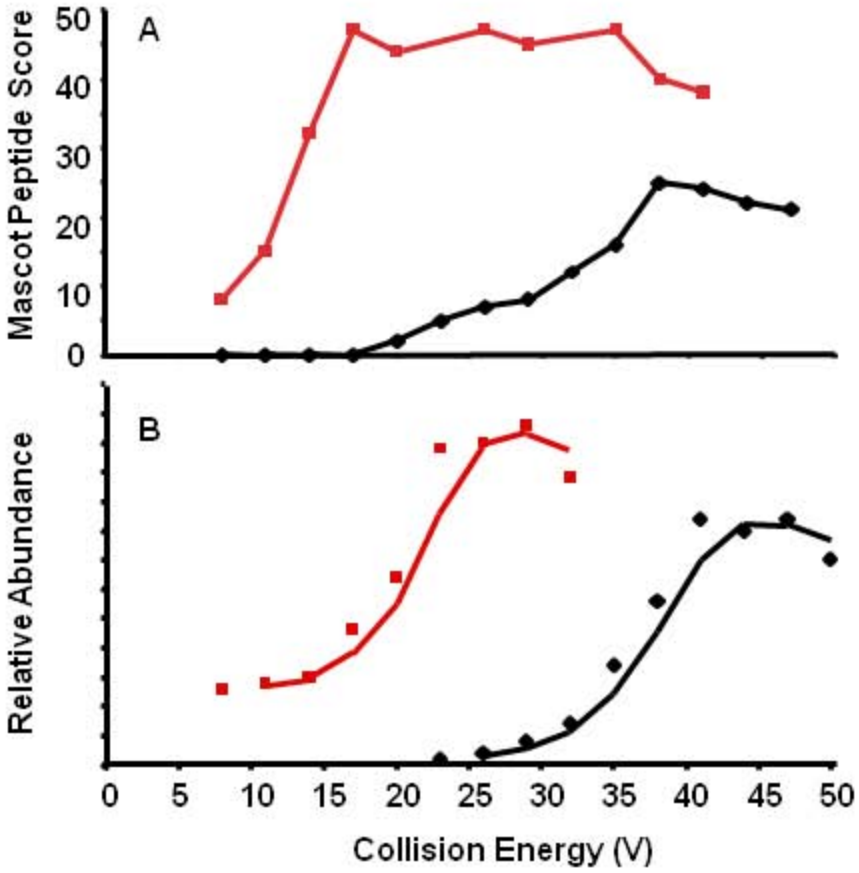

A
B
Mascot Peptide Score
Relative Abundance
0
10
20
30
40
50
0
5
10
15
20
25
30
35
40
45
50
Collision Energy (V)

Figure 3

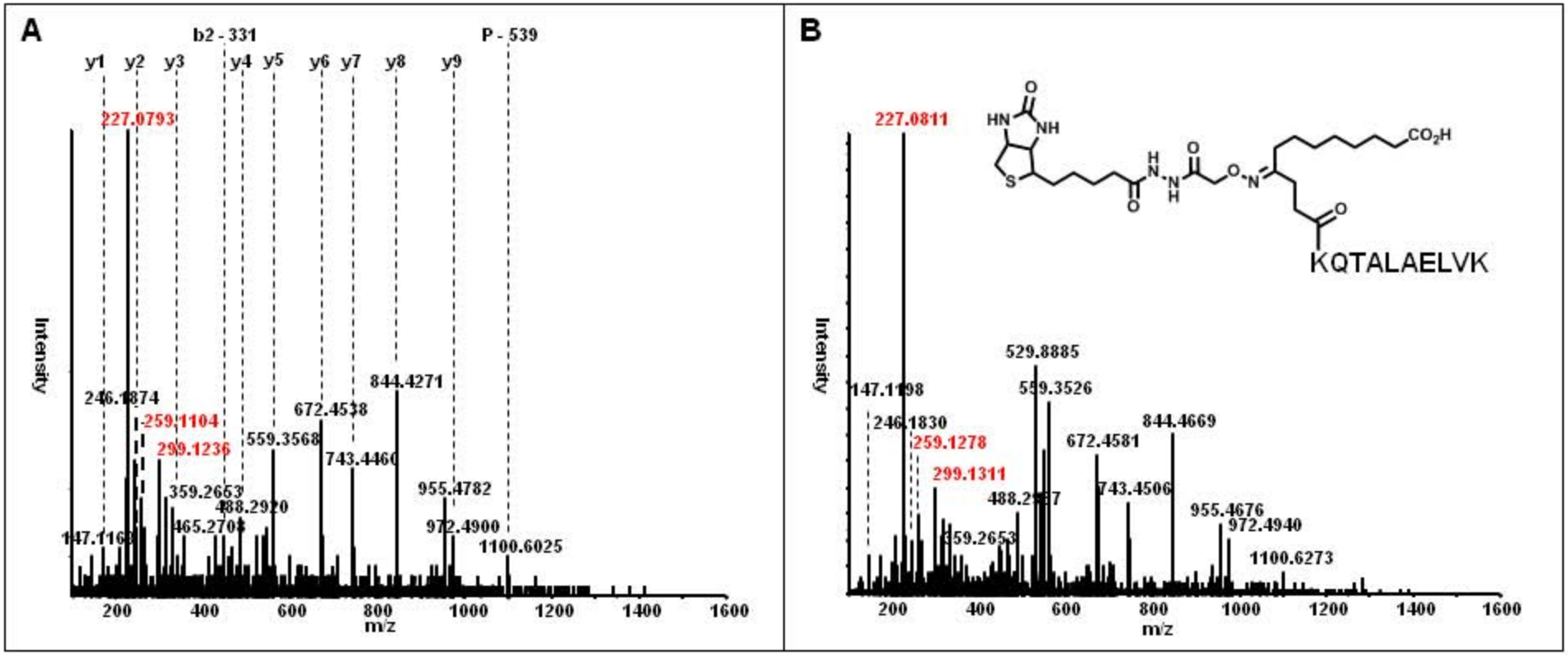

A
b2 - 331
P - 539
y1 y2 y3 y4 y5 y6 y7 y8 y9
227.0793
246.1874
259.1104
299.1236
359.2653
465.2708
488.2920
559.3568
672.4538
743.4460
844.4271
955.4782
972.4900
1100.6025
147.1163
Intensity
200 400 600 800 1000 1200 1400 1600
m/z
B
227.0811
KQTALAELVK
147.1198
246.1830
259.1278
299.1311
359.2653
488.2937
529.8885
559.3526
672.4581
743.4506
844.4669
955.4676
972.4940
1100.6273
Intensity
200 400 600 800 1000 1200 1400 1600
m/z

Figure 4

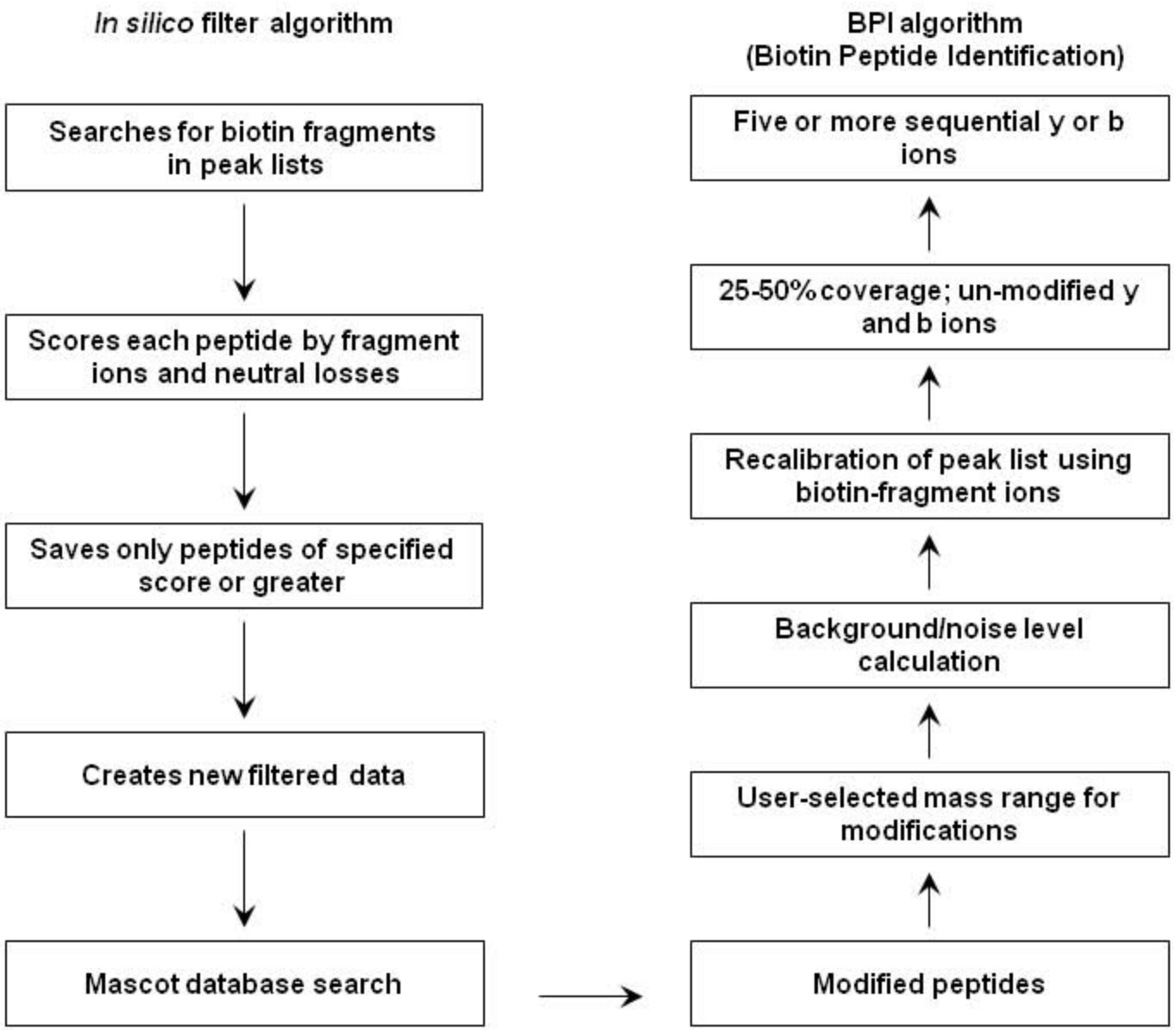

*In silico* filter algorithm
Searches for biotin fragments in peak lists
Scores each peptide by fragment ions and neutral losses
Saves only peptides of specified score or greater
Creates new filtered data
Mascot database search
BPI algorithm
(Biotin Peptide Identification)
Five or more sequential y or b ions
25-50% coverage; un-modified y and b ions
Recalibration of peak list using biotin-fragment ions
Background/noise level calculation
User-selected mass range for modifications
Modified peptides

Figure 5

**Table 1. Search results on DODE-ARP-modified BSA and cytochrome c standards with and without the *in silico* filter.**

## A. Mascot results pre-filter

CTRA_BOVIN Mass: 25650 Score: 95

| Observed | Peptide |
|---|---|
| 412.7117 | L.KLSTAASF.S |
| 412.7117 | L.KLSTAASF.S |
| 437.7442 | W.TLVGIVSW.G |
| 494.25 | L.VNWVQQTL.A |
| 508.7666 | L.TINNDITLL.K |
| 515.2791 | Y.ARVTALVNW.V |
| 561.7727 | W.QVSLQDKTGF.H |
| 561.7727 | W.QVSLQDKTGF.H |
| 562 | W.QVSLQDKTGF.H |
| 665.8442 | Y.NSLTINNDITLL.K |
| 580.2862 | Y.TNANTPDRLQQASLPL.L |
| 836 | L.SNTNCKK...V + ARP-DODE |

K1C19_BOVIN Mass: 43858 Score: 88

| Observed | Peptide |
|---|---|
| 629.3182 | L.QGLEIELQSQL.S |
| 644.8137 | L.ATSDGLLAGNEKL.T |

CYC_HORSE Mass: 11825 Score: 58

| Observed | Peptide |
|---|---|
| 453.7245 | Y.IPGTKMIF.A |
| 453.7435 | Y.IPGTKMIF.A |
| 521.7312 | W.KEETLMEY.L |
| 578.2595 | W.KEETLMEYL.E |
| 388.2056 | M.GDVEKGKKIF.V + Acetyl (N-term) |
| 416.5401 | Y.TDANKNKGITW.K |
| 671.8357 | Y.LKKATNE.- + ARP-DODE |
| 791.337 | W.KEETLMEY.L + ARP-DODE |
| 567.9459 | M.GDVEKGKKIF.V + ARP-DODE |
| 434.4877 | F.AGIKKKTEREDLIAY.L |
| 596.2787 | Y.TDANKNKGITW.K + ARP-DODE |
| 607.9433 | F.GRKTGQAPGFTY.T + ARP-DODE |

K2CA_BOVIN Mass: 18124 Score: 24

| Observed | Peptide |
|---|---|
| 302.6675 | L.MNVKL.A |
| 504.7448 | L.ALDIEIATY.R |

GFAP_BOVIN Mass: 49422 Score: 24

| Observed | Peptide |
|---|---|
| 380.1978 | L.AAELNQL.R |
| 504.7488 | L.ALDIEIATY.R |

## B. Mascot results post-filter

CYC_HORSE Mass: 11825 Score: 0

| Observed | Peptide |
|---|---|
| 671.8357 | Y.LKKATNE.- + ARP-DODE |
| 791.337 | W.KEETLMEY.L + ARP-DODE |
| 567.9459 | M.GDVEKGKKIF.V + ARP-DODE |
| 596.2787 | Y.TDANKNKGITW.K + ARP-DODE |
| 607.9433 | F.GRKTGQAPGFTY.T + ARP-DODE |

**Table 2: Search results on DODE-ARP modified cyt C and BSA, after *in silico* filtration.**
**A: Mascot search results. B: BPI search results**

## A: Mascot

| CYC HORSE | Mass: 11825 |
|---|---|
| **Observed** | **Peptide** |
| 551.3053 | F.AGIKKKTEREDLIAY.L + ARP-ONE |
| 567.9582 | M.GDVEKGKKIF.V + Acetyl: ARP-DODE |
| 671.8545 | Y.LKKATNE.- +ARP-DODE |
| 758.7309 | F.AGIKKKTEREDLIAY.L + ARP-DODE |

## B: Biotin-peptide ID algorithm (BPI)

| CYC HORSE | Mass: 11825 |
|---|---|
| **Observed** | **Peptide** |
| 551.3053 | F.AGIKKKTEREDLIAY.L + ARP-ONE |
| 567.9582 | M.GDVEKGKKIF.V + Acetyl: ARP-DODE |
| 615.2261 | KKATNE + ARP-DODE |
| 635.8450 | Y.LKKATNE.- +ARP-ONE |
| 659.2501 | E NPKKY +ARP-DODE |
| 671.8545 | Y.LKKATNE.- +ARP-DODE |
| 579.7067 | KEETL +ARP-DODE |
| 642.9370 | AGIKKKTE RE DL + ARP-DODE |
| 482.4455 | AGIKKKTEREDL + ARP-DODE |
| 715.7666 | LENPKKY +ARP-DODE |
| 758.7309 | F.AGIKKKTEREDLIAY.L + ARP-DODE |

**Table 3. Output file from the BPI program after searching the ARP-DODE-modified peptide GDVEKGK*KIF against the NCBInr *Equus caballus* database.**

## Correctly identified modified peptide

**GDVEKGkKIF**
**XP_001498872.1 PREDICTED: similar to Cytochrome c [Equus caballus]**
**Mass spec monoisotopic mass: 567.9229**
**Charge: + 3 Elution: 52.22 to 53.3 min**
**Biotin modification mass: 539.1057**
**Carbonyl modificaton mass: 226.2327**

**y ions**
**(1, '+', 166.0950)**
**(2, '+', 279.1613)**
**(3, '+', 407.2481)**
**(4, '+', 1074.4848)**
**(6, '++', 630.3097)**
**(7, '++', 694.8194)**
**(7, '-331 neut. loss ++', 529.2677)**
**(8, '++', 744.3611)**
**(8, '-331 neut. loss ++', 578.8039)**
**(9, '++', 801.8539)**
**(10, '-331 neut. loss ++', 685.8180)**

**b ions**
**(1, '+', 100.05802)**
**(2, '+', 215.0707)**
**(3, '+', 314.1196)**
**(4, '+', 443.1598)**
**(5, '+', 571.2721)**
**(5, '++', 286.1342)**

## Two false positives

**kSALMPAQLF**
**XP_001497818.1 PREDICTED: similar to Phosphoinositide-3-kinase, class 3 [Equus caballus]**
**Mass spec monoisotopic mass: 567.9229**
**Charge: + 3 Elution: 52.22 to 53.3 min**
**Biotin modification mass: 596.14450**
**Carbonyl modificaton mass: 283.2719**

**y ions**
**(1, '+', 166.0950)**
**(2, '+', 279.1613)**
**(3, '+', 407.2481)**
**(4, '+', 478.2573)**
**(5, '+', 575.3043)**
**(6, '+', 706.3402**
**(7, '++', 410.2204)**
**(8, '+', 890.4671)**
**(8, '++', 445.7426)**
**(9, '++', 489.2415)**
**(10, '-331 neut. loss ++', 685.8180)**

**b ions**
**(7, '-331 neut. loss +', 964.4102)**

**VkAELAISTF**
**XP_001500348.1 PREDICTED: similar to protein tyrosine phosphatase, receptor type, T [Equus caballus]**
**Mass spec monoisotopic mass: 567.9229**
**Charge: + 3 Elution: 52.22 to 53.3 min**
**Biotin modification mass: 623.1381**
**Carbonyl modificaton mass: 310.2650**

**y ions**
**(1, '+', 166.0950)**
**(2, '+', 267.1457)**
**(4, '+', 467.2438)**
**(5, '+', 538.2739)**
**(6, '+', 651.3539)**
**(7, '+', 780.4013)**
**(8, '+', 851.4590)**
**(8, '++', 426.2275)**
**(9, '++', 801.8539)**
**(10, '-331 neut. loss ++', 685.8180)**

**b ions**
**(1, '+', 100.05802)**
**(9, '++', 801.8539)**
**(10, '-331 neut. loss ++', 685.8180)**

| Protein Name | Accession Number | Modified Peptide | Modification | Experimental/Calculated Modification Mass | | |
|---|---|---|---|---|---|---|
| | | | | Precursor (*m/z*) | ARP + Carbonyl | Carbonyl |
| Albumin | NP_033784 | K*QTALAELVK | DODE | 547.2894 (+3) | 539.1868 | 226.2337 |
| | | QIK*K | DODE | 528.2976 (+2) | 539.2443 | 226.3712 |
| | | MK*CSSMQK | aminoadipic semialdehyde | 627.7645 (+2) | 312.0997 | -0.7734 |
| | | AFK*AWAVAR | 2-heptenal | 482.2532 (+3) | 425.1982 | 112.3252 |
| Alpha-2-HS-glycoprotein* | NP_038493 | ? | possible DODE | | | |
| Alpha-2-macroglobulin* | NP_783327.1 | | possible DODE | | | |
| Apolipoprotein A1* | NP_659146 | | possible DODE | | | |
| Apolipoprotein E* | NP_033826.1 | | possible DODE | | | |
| Ceruloplasmin isoform a and b** | NP_001036076<br>NP_031778 | ? | unknown carbonyl | | | |
| Compliment Component 3 | NP_033908 | ? | unknown carbonyl | | | |
| Esterase 1 | NP_031980 | ? | unknown carbonyl | | | |
| Haptoglobin | NP_059066 | GSFPWQAK*M*ISR | aminoadipic semialdehyde | 579.2786 (+3) | 312.1156 | -0.7575 |
| Hemopexin | NP_059067 | K*WFWDFATR | ONE | 575.2934 (+3) | 467.2580 | 154.3850 |
| | | GPDSVFLIK*EDK | aminoadipic semialdehyde | 553.948 (+3) | 312.1233 | -0.7497 |
| | | GHSGTR*ELISAR | glutamic semialdehyde | 518.5875 (+3) | 270.0801 | -42.7930 |
| Inter-alpha trypsin inhibitor, heavy chain 3 | NP_032433 | DH*LVQATPANLK | HNE | 592.6391 (+3) | 469.1960 | 156.3230 |
| Kininogen 2 | NP_958763 | ? | unknown carbonyl | | | |
| Murinoglobulin 1 | NP_032671 | H*ASAK | HNE | 491.7722 (+2) | 469.2623 | 156.3893 |
| Murinoglobulin 1 or 2 | NP_032672 | H*TSSWLVTPK | HNE | 542.2829 (+3) | 469.2149 | 156.3418 |
| Pregnancy zone protein | NP_031402 | VK*ALSFYQPR | 2-heptenal | 545.2964 (+3) | 425.2054 | 112.3323 |
| | | TK*VFQLR | aminoadipic semialdehyde | 602.3050 (+2) | 312.0764 | -0.7966 |
| Serine (or cysteine) proteinase inhibitor** | NP_955018<br>NP_659083<br>NP_082273<br>XP_282560<br>NP_035588 | ? | unknown carbonyl | | | |
| Transferrin | NP_598738 | SP*LEK | glutamic semialdehyde | 451.7331 (+2) | 329.1437 | 16.2706 |
| | | KNNGK*EDLIWEILK | aminoadipic semialdehyde | 671.3462 (+3) | 312.0999 | -0.7731 |
| | | K*NNGKEDLIWEILK | aminoadipic semialdehyde | 671.3462 (+3) | 312.0999 | -0.7731 |
| | | SK*DFQLFSSPLGK | aminoadipic semialdehyde | 589.3111 (+3) | 312.1466 | -0.7265 |
| | | IPSH*AVVAR | HNE | 473.6003 (+3) | 469.2462 | 156.3731 |

*** Protein identified as DODE modified by 2-D SDS PAGE and anti-DODE western**

****Protein identified as carbonyl modified after biotin affinity chromatography**