The Thesis Committee for William Ryan Parker
Certifies that this is the approved version of the following thesis:

Characterization of Proteins and Peptides via Enhanced 266 nm Ultraviolet Photodissociation Mass Spectrometry utilizing a Selenium Based Chromophore

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Jennifer S. Brodbelt

Livia Schiavinato Eberlin

Characterization of Proteins and Peptides via Enhanced 266 nm Ultraviolet Photodissociation Mass Spectrometry utilizing a Selenium Based Chromophore

by

William Ryan Parker, B.S.

Thesis

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May 2016

Acknowledgements

The guidance and mentorship provided by the thesis supervisor, Jennifer Brodbelt, throughout the duration of this work is acknowledged. Funding for this research was provided by the NSF and Welch Foundation.

Abstract

Characterization of Proteins and Peptides via Enhanced 266 nm Ultraviolet Photodissociation Mass Spectrometry utilizing a Selenium Based Chromophore

William Ryan Parker, M.A.

The University of Texas at Austin, 2016

Supervisor: Jennifer S. Brodbelt

Mass spectrometry and chemical derivatization have been used as tools for the identification of proteins in both top-down and bottom-up studies. Cysteine is the rarest and most nucleophilic amino acid thus making it a popular target for chemical tagging strategies. Ultraviolet photodissociation (UVPD) is a versatile activation technique for fragmentation of peptides and proteins. For successful photodissociation, ions of interest must contain a suitable chromophore that matches the wavelength of irradiation. *N*-(Phenylseleno)phthalimide (NPSP) is a fast reacting reagent which attaches a selenium based chromophore that absorbs at 266 nm light to free thiols. In the studies presented in this thesis, NPSP was used to derivatize free cysteine residues in both intact proteins and tryptic peptides. Activation with 266 nm photons causes a dominant neutral loss of the benzeneselenol groups on the tagged protein or peptide ions. This diagnostic neutral loss allows the determination of the number of free versus bound cysteine residues in intact proteins. Additionally, tagging peptides with benzeneselenol provides a means to target

only the cysteine-containing peptides in bottom-up proteomics experiments. Both of these methods provide a significantly reduced search space for identification of cysteine-containing proteins. Counting the number of cysteine residues also provides an effective way to restrict the number of protein candidates for database searches. Moreover, cysteine peptides are inherently more unique than other peptides created upon enzymatic digestion of proteins due to the low frequency of cysteine in the proteome, thus allowing these peptides to be used as surrogates for protein identification.

Table of Contents

List of Tables

List of Figures

Chapter One: Characterization of the Cysteine Content in Proteins utilizing Cysteine Selenylation with 266 nm Ultraviolet Photodissociation (UVPD)[1]

OUTLINE 1.1

Characterization of the cysteine content of proteins is a key aspect of proteomics. By defining both the total number of cysteines and their bound/unbound state, the number of candidate proteins considered in database searches is significantly constrained. Herein we present a methodology that utilizes 266 nm UVPD to count the number of free and bound cysteines in intact proteins. In order to attain this goal, proteins were derivatized with N-(phenylseleno)phthalimide (NPSP) to install a selectively cleavable Se-S bond upon 266 UVPD. The number of Se-S bonds cleaved upon UVPD, a process that releases SePh moieties, corresponds to the number of cysteine residues per protein.

INTRODUCTION 1.2

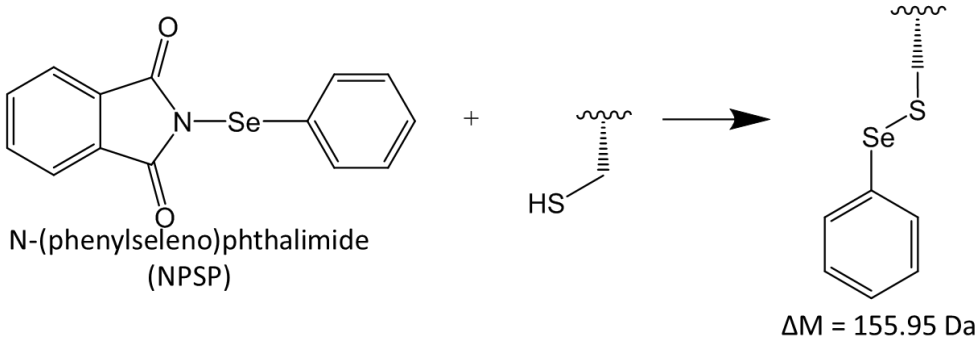
Cysteine is involved in both stabilizing tertiary protein structure through disulfide bonding and modulation of protein redox activity [1]. There have been numerous studies aimed at characterization of the redox states of cysteines in proteins owing to the inherent importance of cysteine-mediated chemistry in countless biological processes. Methods such as UV absorption spectroscopy [2], fluorescent labeling [3], and X-ray absorption spectroscopy [4] have been used to quantify or characterize the cysteine content of proteins. Characterizing cysteine content by mass spectrometry (MS) has also become a popular option as a consequence of the development of well-established bottom-up proteomics approaches [5–7] often in combination with various clever cysteine-selective derivatization

[1] Parker, W.R., Brodbelt, J.S.: Characterization of the Cysteine Content in Proteins Utilizing Cysteine Selenylation with 266 nm Ultraviolet Photodissociation (UVPD). J. Am. Soc. Mass Spectrom. ASAP (2016). Co-author Jennifer Brodbelt provided research funding.

methods [8]. MS strategies for determination of cysteines have been based on the utilization of mass tags [9] or isotopic labels [10, 11], differential monitoring of ESI mass spectra after cysteine-selective reactions [12], proteolysis in isotopically heavy solvents [13], characterization by high resolution top down MS/MS [14], or utilization of selective ion/ion reactions [15, 16], all of which have proven to be versatile methods for either qualitative or quantitative characterization of cysteine content of proteins and peptides. Electrochemical tagging reactions of cysteines have also been implemented via coupling an electrochemical cell in an on-line fashion to a mass spectrometer [17–21]. The latter methodologies have even shown promise for mapping reactivities of cysteine residues in different locations of proteins [22]. Chemical derivatization of cysteine residues in peptides and proteins by reactions with quinone has been shown to differentiate free thiols versus disulfide bonds [23]. For this prior study, 266 nm UVPD was used to promote homolytic cleavage of C-S bonds of quinone-derivatized proteins and peptides, generating neutral loss products which were subjected to CID to achieve radical-directed dissociation (RDD) [23]. Selective cleavage of disulfide bonds to elucidate disulfide-linked peptide pairs has also been shown to be effective using 266 nm photodissociation [24].

N-(phenylseleno)phthalmide (NPSP) is a well-known selenylating reagent [25, 26]. Fast, efficient, and selective selenylation of cysteine-containing species has been achieved previously using NPSP (**Scheme 1.1**) [27–29]**.** Additionally, on-line electrolytic cleavage of disulfides via an electrochemical cell prior to reactions with NPSP and MS analysis has been reported as a means to identify disulfide-containing peptides in digests [30]. Herein we present a strategy for counting cysteine residues in proteins based on NPSP-based

selenylation followed by 266 nm UVPD to selectively and exclusively cleave the tags. The cysteine content information can be utilized in conjunction with informatics engines such as UniProt to generate a list of candidate proteins from a proteome that match the cysteine content observed experimentally.



Scheme 1.1: Scheme showing reaction of N-(phenylseleno)phthalimide (NPSP) with a free thiol

METHODS 1.3

Sample Preparation

Proteins containing up to eight cysteines were analyzed in three ways: (i) as intact proteins prior to NPSP derivatization, (ii) after derivatization with NPSP without reduction of intrinsic disulfide bonds, and (iii) after derivatization with NPSP and reduction of disulfide bonds. For both (i) and (ii) proteins were suspended at a concentration of 10 μM in 50:50 H2O:acetonitrile and 1% formic acid. For (ii) the proteins were reacted with 10 mM NPSP dissolved in dry acetonitrile (typically using 200 uL of protein solution and 2 uL of NPSP solution) for 30 seconds prior to analysis by mass spectrometry (without clean-up). For (iii) proteins were reduced by incubation with 10 mM DTT in 150 mM NH_4HCO_3 for three hours at 55°C (typically using 200 ul of protein solution and 2 uL of DTT solution). Following reduction, proteins were buffer-

exchanged three times into 1% formic acid using Amicon Ultra Centrifugal Filters (Merck Millipore). Proteins were then diluted by addition of one volume 1% formic acid in acetonitrile and incubated with 10 mM NPSP for 30 seconds. The final protein solutions were analyzed without further clean-up.

Mass Spectrometry

Mass spectrometry experiments were undertaken in positive ion mode on a Thermo Scientific Velos Pro dual-pressure linear ion trap mass spectrometer (San Jose, CA) equipped with CID, HCD, ETD, and 266 nm UVPD capabilities. Each spectrum consisted of 3 μscans averaged. The applied ESI spray voltage was 3 kV. Tandem mass spectrometry was carried out in the high pressure trap for all activation strategies using the most abundant ion of the charge state envelope. For CID and HCD a normalized collision energy of 35 NCE was used (during an activation period of 10 ms for CID and 2 ms for HCD), while for ETD a reaction time from 40-120 ms was utilized. UVPD was implemented in a manner described previously [31] and was performed using the fourth harmonic of a Continuum Minilite Nd:YAG laser (San Jose, CA) with an energy output of approximately 6 mJ per pulse. Ions were subjected to an increasing number of 266 nm laser pulses until either the precursor was completely eliminated or no additional neutral losses were observed. Post-acquisition, data analysis was assisted by charge state deconvolution software, MagTran [32]. Bovine β-lactoglobulin, bovine α-lactalbumin, chicken lysozyme, bovine ribonuclease A., bovine aprotinin, horse cytochrome C, and bovine serum albumin (BSA) were used in this study.

RESULTS AND DISCUSSION 1.4

ESI-MS Analysis

Derivatization of each cysteine residue with NPSP results in a mass shift of 156 Da by the addition of a SeC_6H_5 moiety through formation of a S-Se bond as depicted in **Scheme 1.1**. Cysteine sites can be counted in several ways using this reactive tag. First, the net mass shift of a protein is readily observed in the MS1 mass spectrum. In **Figure 1.1** this is observed for the β-lactoglobulin variants A and B. Upon deconvolution of the charge state envelope, an average molecular weight was determined. Compared to the underivatized proteins in **Figure 1.1a**, the ones in **Figure 1.1b** showed a mass shift of approximately 155 Da per protein, indicative of one reactive site (one free cysteine) for both A and B variants of β-lactoglobulin. When reduction was performed prior to derivatization of the proteins (**Figure 1.1c**), a more prominent mass addition was observed along with a shift to higher charge states. The shift to higher charge states was indicative of the cleavage of disulfide bonds that facilitated more extensive protonation of the protein. The net mass shift observed in **Figure 1.1c** for reduced β-lactoglobulin was 781 Da, corresponding to addition of five SePh tags. From these observations, it is deduced that there are five cysteine residues of which four are inaccessible until after reduction.

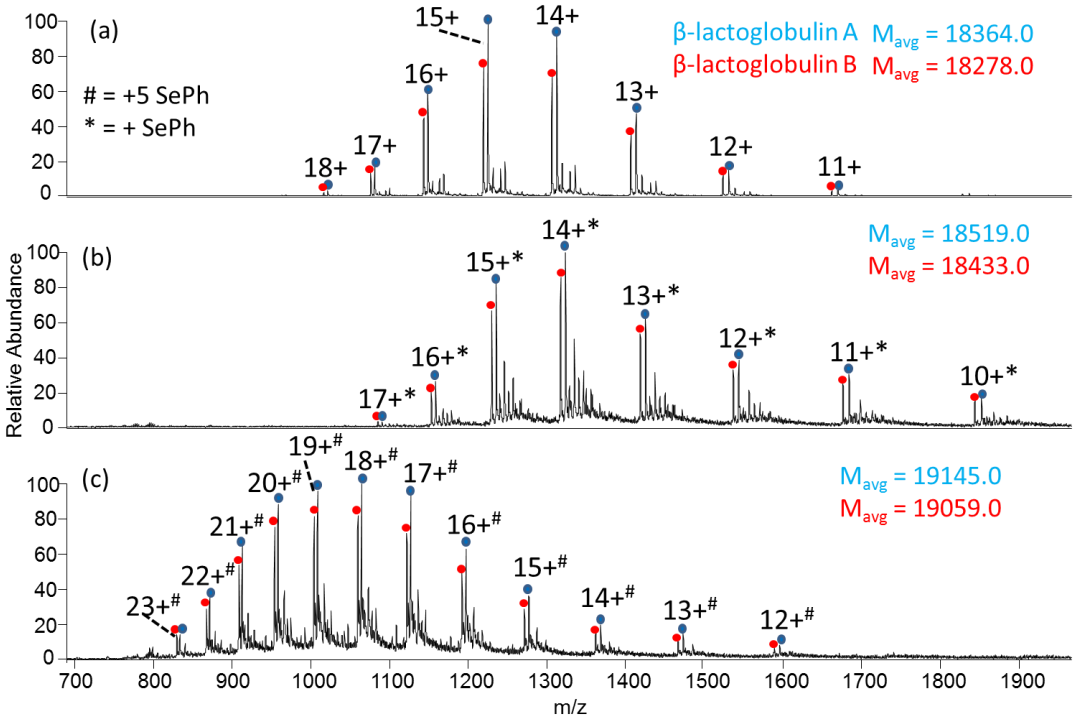


Figure 1.1. β-lactoglobulin A (blue) and B (red) for (a) unmodified, (b) NPSP-modified, and (c) NPSP-modified after reduction. (# = + 5 SePh, * = + SePh)

The same strategy was used for lysozyme (**Figure 1.2)**, ribonuclease A **(Figure 1.3)**, aprotinin (**Figure 1.4)**, α-lactalbumin (**Figure 1.5**), cytochrome C (**Figure 1.6**) and BSA (**Figure 1.7)**, and the findings are summarized in **Table 1.1.** In each case the number of free cysteines and disulfides was correctly deduced based on monitoring the mass shifts upon reaction with NPSP prior to or after reduction of the protein. Ribonuclease A and aprotinin exhibited no reactions with NPSP prior to reduction, indicating the absence of free cysteines. After reduction and reaction with NPSP, ribonuclease A displayed a mass shift consistent with addition of eight SePh tags; for aprotinin it was addition of six SePh tags. For both ribonuclease A (**Figures 1.3a, 1.3b)** and aprotinin (**Figures 1.4a, 1.4b**), a +98 Da mass addition to the intact protein was observed in the mass spectra prior to NPSP reactions.

This adduct was attributed to non-covalently bound sulfuric or phosphoric acid which is known to interact strongly with surface accessible basic sites of the protein [31]. This adduct was easily dislodged in the cleanup step performed on each protein after reduction and prior to selenylation (**Figures 1.3c, 1.4c**) and did not impede the NPSP tagging strategy. For aprotonin, the smallest protein used in this study, a distinct broadening of the peaks in the charge envelope was observed in **Figure 1.4c**. This was due to the broad isotopic distribution of the selenium component of the tag and was not especially noticeable for the larger proteins due to their significantly higher charge states (and thus more compressed peak widths).

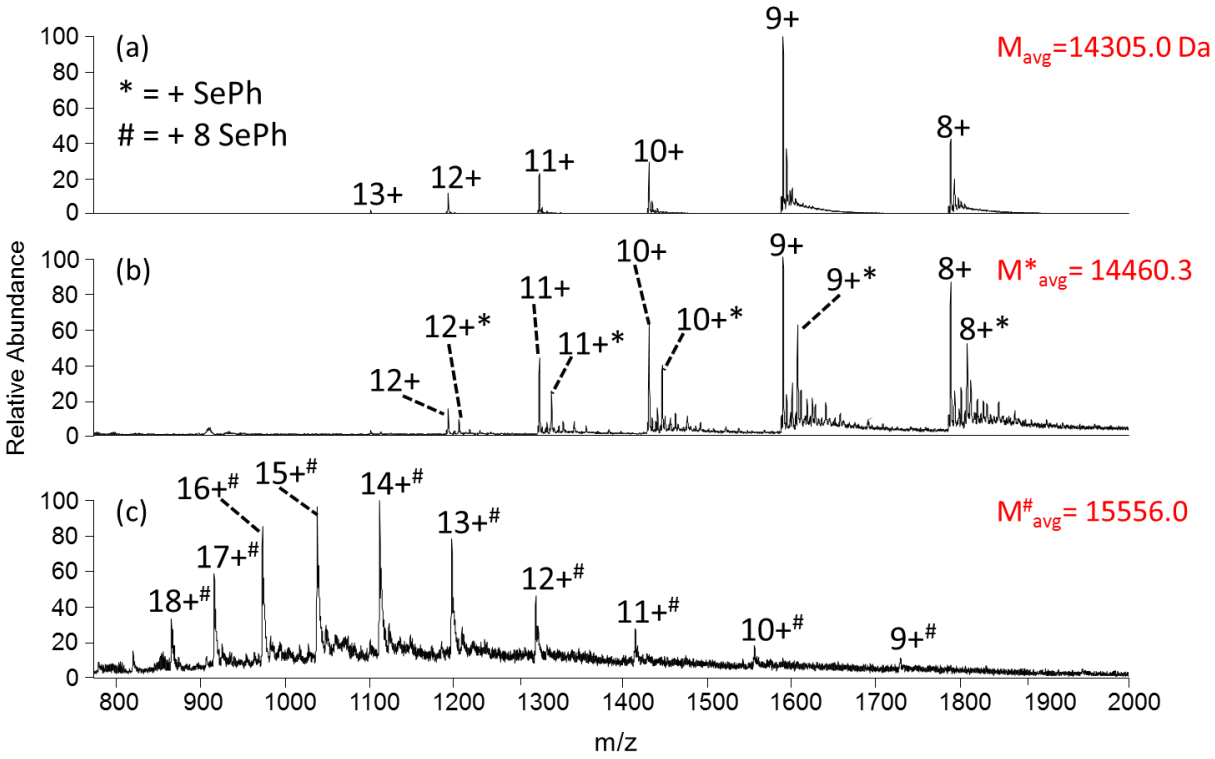


Figure 1.2. ESI mass spectra of lysozyme: (a) unmodified, (b) NPSP-modified prior to reduction, and (c) NPSP-modified after reduction

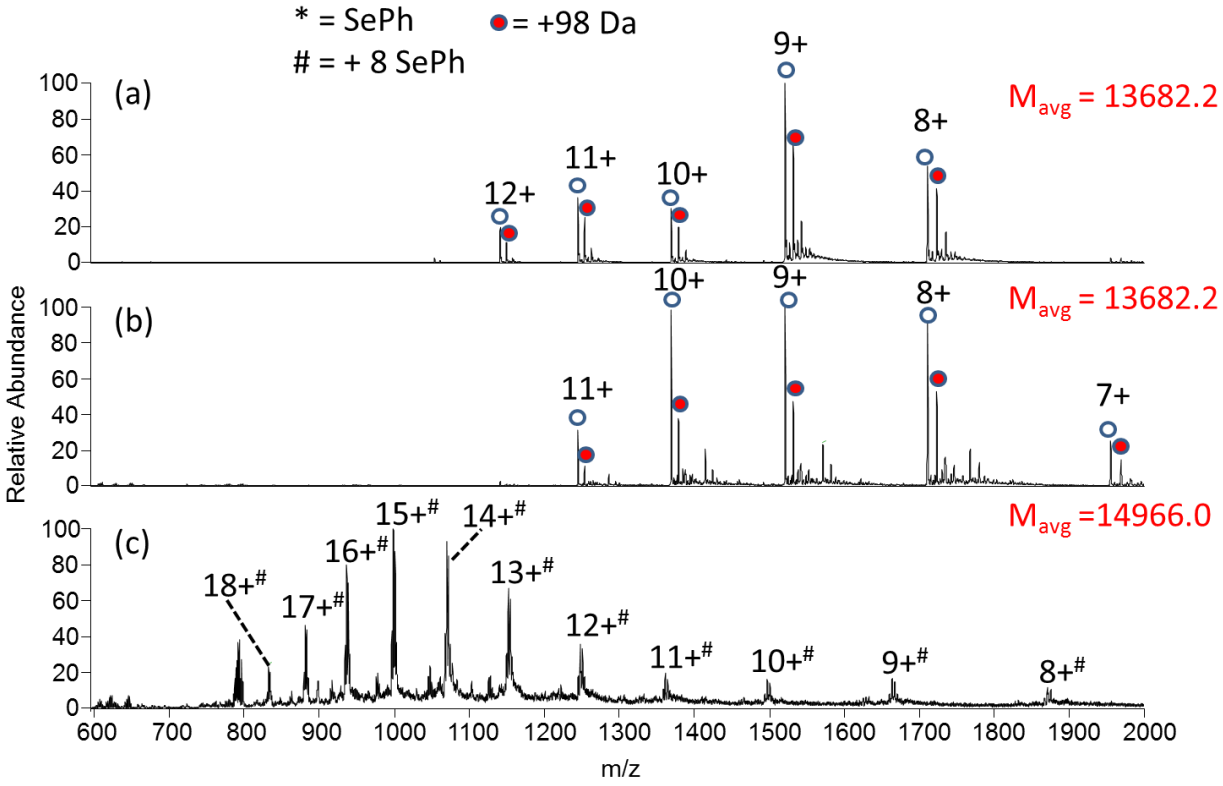


Figure 1.3: ESI mass spectra of ribonuclease A: (a) unmodified, (b) NPSP-modified prior to reduction, and (c) NPSP-modified after reduction. Ions containing contaminating adduct (~98 Da) for ribonuclease A are denoted by a solid red circle. (* = SePh, # = 8 SePh)

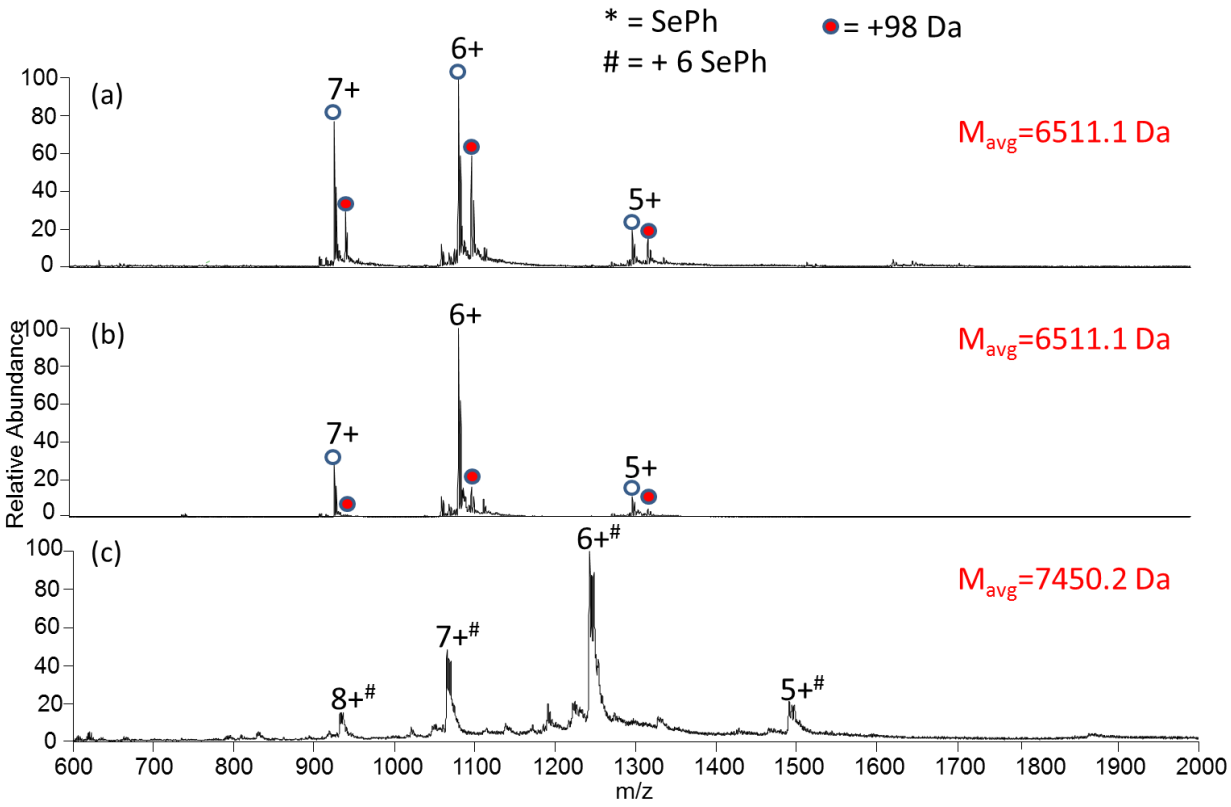


Figure 1.4. ESI mass spectra of aprotinin: (a) unmodified, (b) NPSP-modified prior to reduction, and (c) NPSP-modified after reduction. Ions containing contaminating adduct (~+98 Da) aprotinin are denoted by a solid red circle.

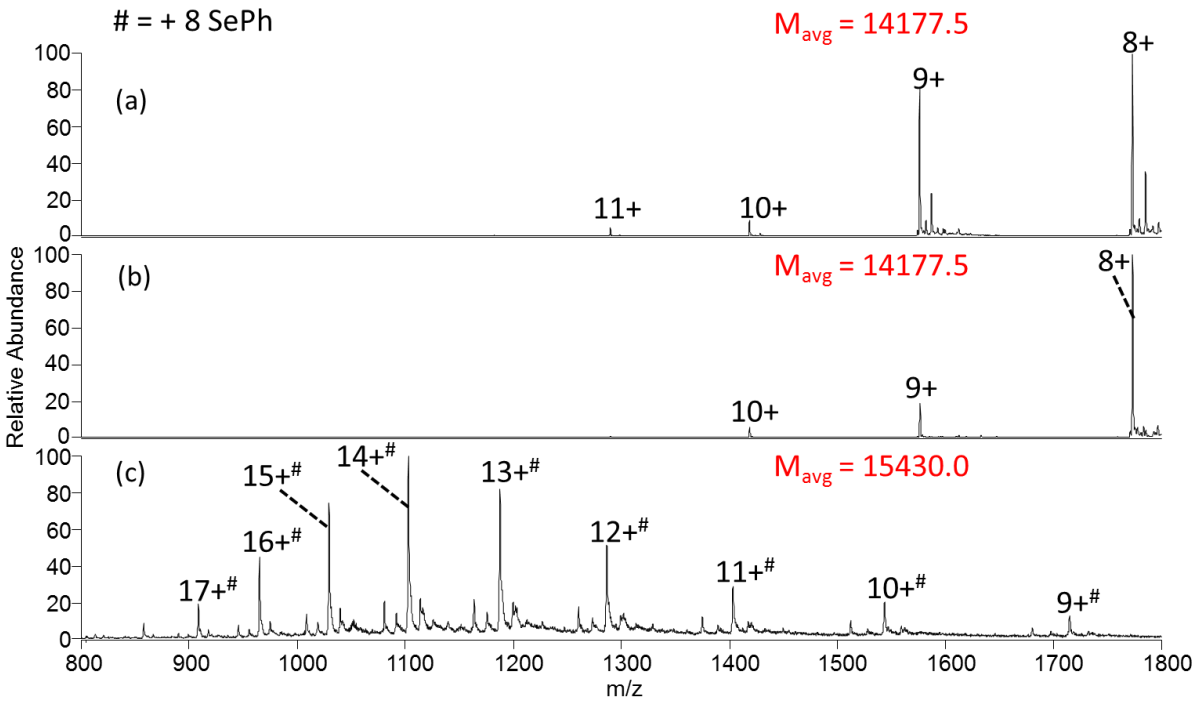


Figure 1.5. ESI mass spectra of α-lactalbumin: (a) unmodified, (b) NPSP-modified prior to reduction, and (c) NPSP-modified after reduction.

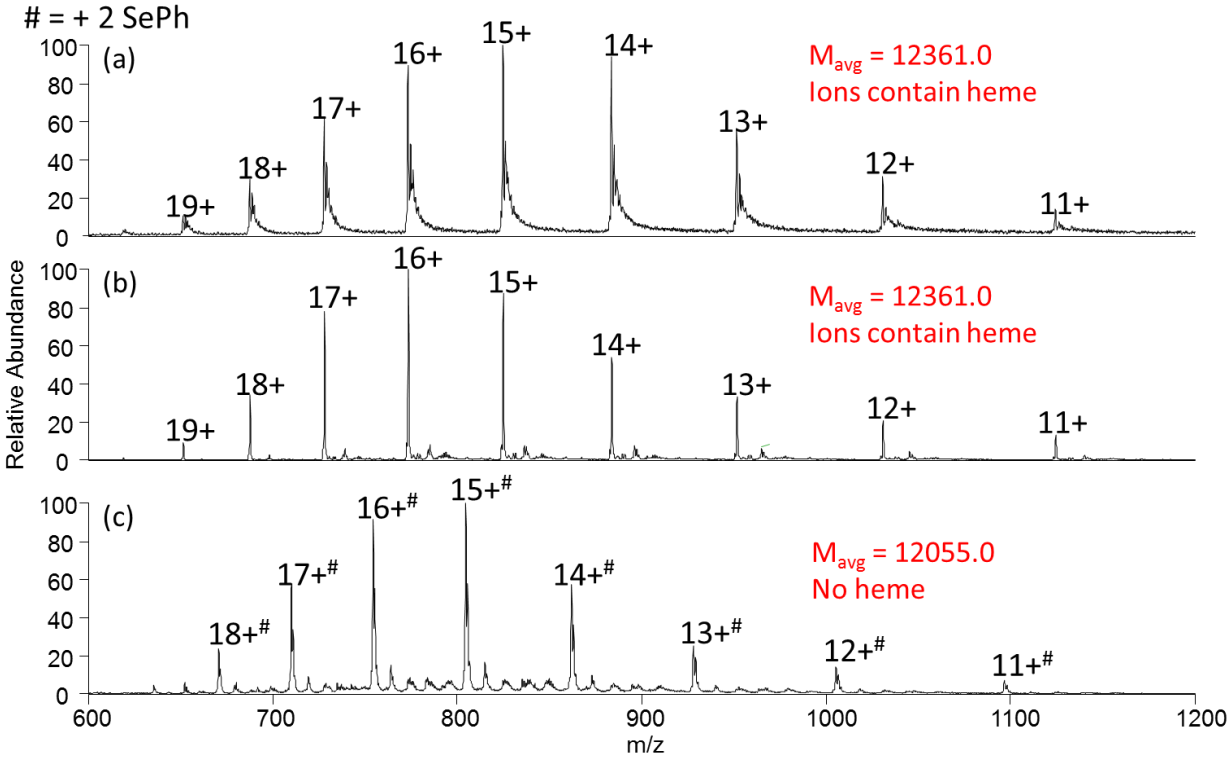


Figure 1.6. ESI mass spectra of cytochrome c: (a) unmodified, (b) NPSP-modified prior to reduction, and (c) NPSP-modified after reduction

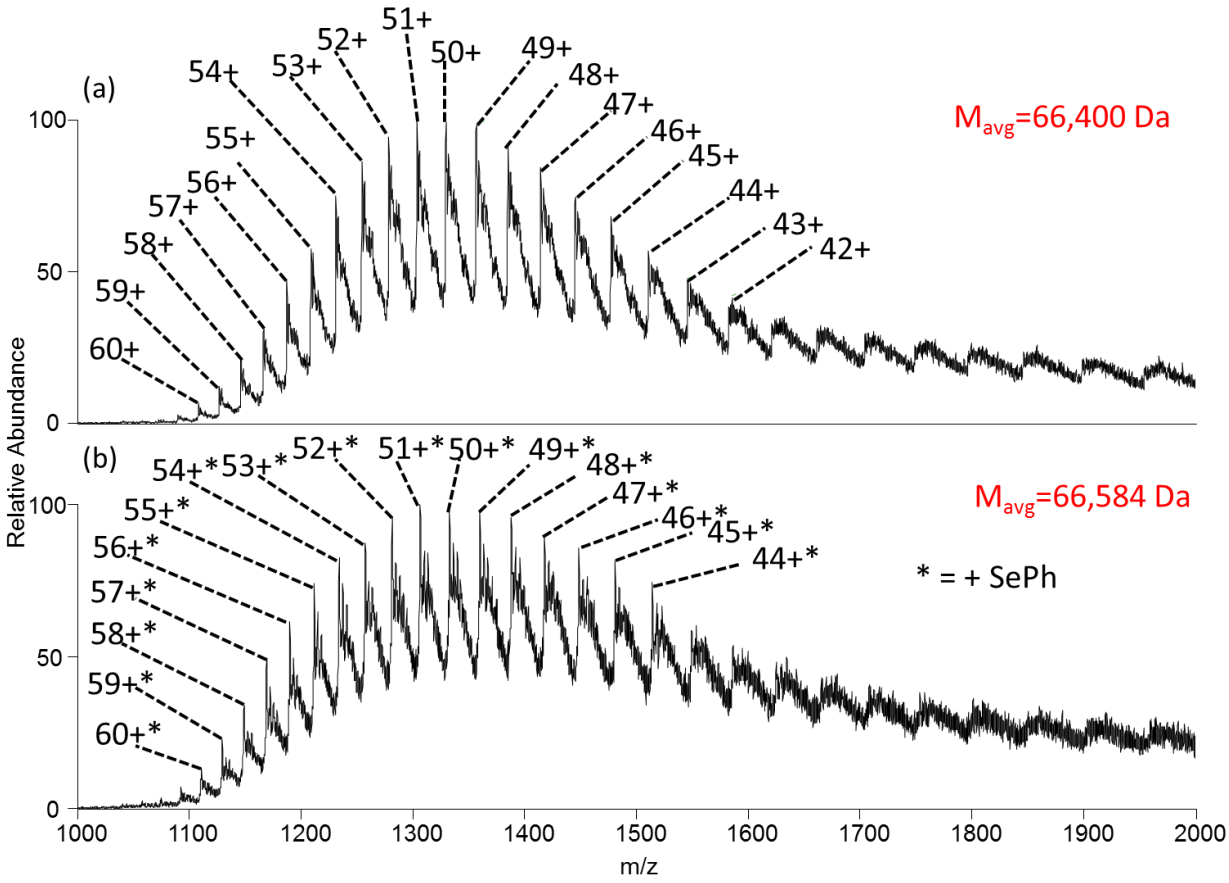


Figure 1.7. ESI mass spectra of BSA: (a) unmodified and (b) NPSP-modified prior to reduction

The NPSP tagging method indicated that alpha-lactalbumin had four disulfide-bound cysteines and no free cysteines and that cytochrome c had two disulfide-bound cysteines and no free cysteines (**Table 1.1**). The NPSP strategy was unsuccessful for characterization of BSA due to incomplete tagging, an outcome not unexpected owing to the fact that BSA has 35 cysteines. The method showed mixed success for lysozyme. Lysozyme has eight cysteines, all engaged in disulfide bonds. Upon reaction with NPSP, the resulting mass spectrum displayed a mass shift corresponding to attachment of one SePh tag which suggested the presence of one free cysteine. This odd result may correspond to partial

degradation of lysozyme, thus leading to cleavage of one disulfide bond in the intact protein. Although addition of two SePh tags might be expected for this scenario, it is possible that one cysteine remains inaccessible owing to the six other disulfide bonds in the protein. Upon reduction of lysozyme and reaction with NPSP, the resulting mass shift was consistent with eight cysteines as expected.

Table 1.1: Comparison of theoretical and experimental cysteine counting results. All calculations are done based on the chain region of the protein as defined by UniProt. The percent proteome match signifies the percentages of proteins in each proteome which have the same cysteine content (free vs bound cysteines) as the protein listed in that row. The percent proteome match was calculated using a custom python script.

Protein (AC) [Species]	Free cys (Uniprot)	Free cys (exp)	Total cys (Uniprot)	Total cys (exp)	Figures	% Proteome Match
β-lactoglobulin (P02754) [*Bos taurus*]	1	1	5	5	1.1, 1.8	0.11% (28 out of 24113)
Lysozyme (P00698) [*Gallus gallus*]	0	1	8	8	1.2, 1.9	0.03% (6 out of 17691)
ribonuclease A (P61823) [*Bos taurus*]	0	0	8	8 by mass 7 by UVPD	1.3, 1.10	0.15% (36 out of 24113)
Aprotinin (P00974) [*Bos taurus*]	0	0	6	6	1.4, 1.11, 1.15	0.31% (74 out of 24113)
α-lactalbumin (P00711) [*Bos taurus*]	0	0	8	8	1.5, 1.12	0.15% (36 out of 24113)
cytochrome C (P00004) [*Equus caballus*]	0	0	2	2	1.6, 1.13, 1.16	0.08% (19 out of 22718)
BSA[1] (P02769) [*Bos taurus*]	1	1	35	N/A	1.7, 1.14	< 0.01% (2 out of 24113)

[1]Only the oxidized form of BSA was evaluated because it has only one free cysteine out of 35 total cysteines; the products formed after reduction of BSA would not be resolvable.

Analysis by 266 nm UVPD

Sulfur-selenium bonds have been shown previously to be photolytically cleavable in solution, producing radical products [32]. Implementation of this type of photoreaction for ions in the gas phase using 266 nm photons in the present study resulted in exclusive cleavage of the S-Se bonds and loss of the SePh tags. Monitoring the loss of SePh tags was used as a second facile means to count the number of cysteine residues per protein (i) prior to reduction and (ii) post-reduction.

Examples of the mass spectra obtained upon UVPD of the SePh-tagged proteins are shown in **Figure 1.8** for beta-lactoglobulin A and B. Irradiation of unmodified β-lactoglobulin with 266 nm photons (2 laser pulses, 6 mJ) resulted in no significant dissociation, as shown in **Figures 1.8a** and **1.8b** for the A and B forms of the protein (15+ charge state), respectively. However, the SePh-modified proteins (**Figure 1.8c-f**), dissociated by loss of one SePh tag for the non-reduced proteins or by loss of five SePh tags for the reduced proteins upon 266 nm UVPD (2 laser pulses, 6 mJ). Cleavage of the S-Se bonds is both extremely efficient and highly selective upon absorption of 266 nm photons, yielding a neutral loss of 156 Da per cleaved tag. For both variants of β-lactoglobulin, 266 nm UVPD indicated one free and four bound cysteine residues. This UVPD method was extended to lysozyme (**Figure 1.9**), ribonuclease A (**Figure 1.10**), aprotinin (**Figure 1.11**), α-lactalbumin (**Figure 1.12**), cytochrome C (**Figure 1.13**), and BSA (**Figure 1.14**). Using multiple laser pulses facilitated removal of the SePh tags as shown for α-lactalbumin in **Figure 1.12**. The results of the UVPD strategy for counting free and bound cysteines are summarized in **Table 1.1**.

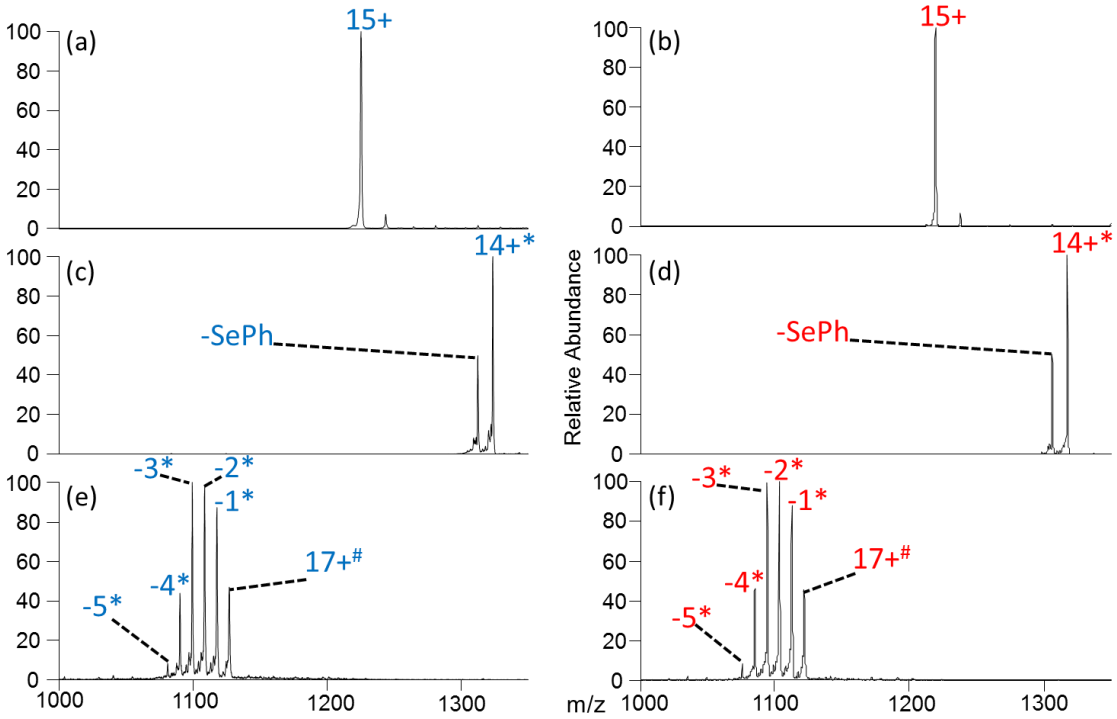


Figure 1.8. β-lactoglobulin A (blue) and B (red) activated by 2 laser pulses (266 nm) as follows (a,b) unmodified, (c,d) modified with NPSP without reduction, and (e,f) modified with NPSP after reduction. (* denotes 1 SePh group, # denotes 5 SePh groups)

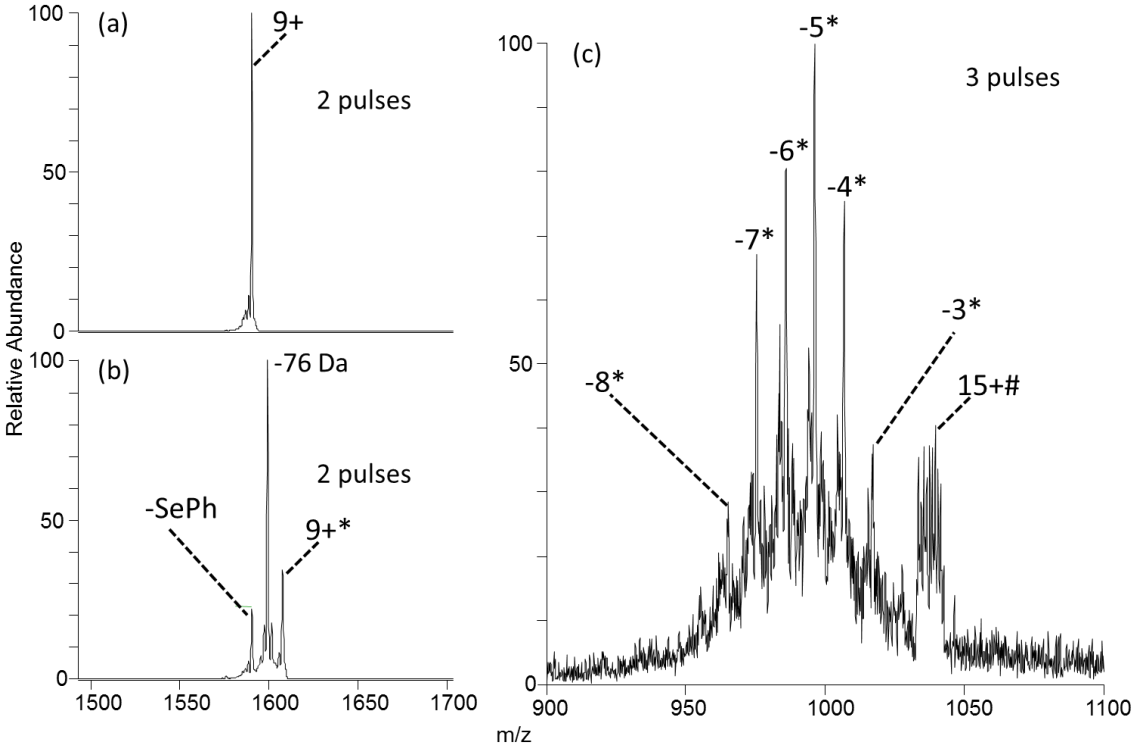


Figure 1.9. 266 nm UVPD mass spectra of lysozyme for (a) unmodified (9+), (b) NPSP-modified before reduction (9+), and (c) NPSP-modified after reduction (15+). (* = SePh, # = 8 SePh)

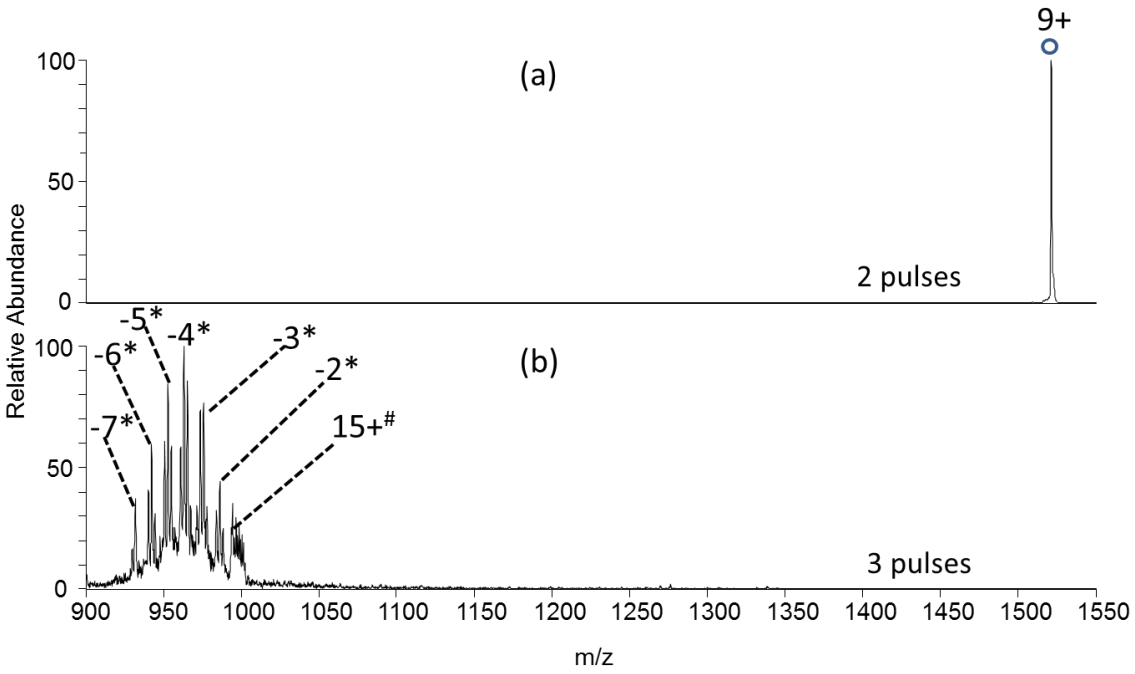


Figure 1.10. 266 nm UVPD mass spectra of ribonuclease A for (a) unmodified (9+) and (b) NPSP-modified after reduction (15+)

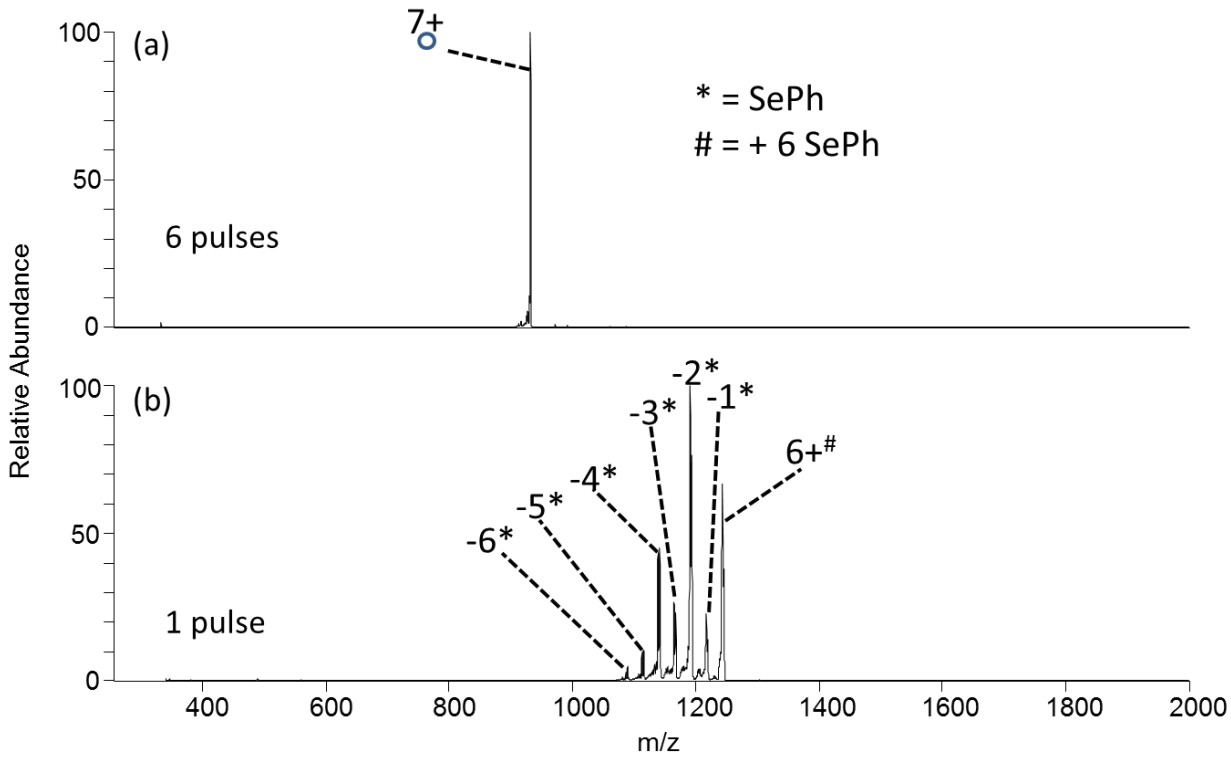


Figure 1.11. 266 nm UVPD mass spectra of aprotinin for (a) unmodified (7+) and (b) NPSP-modified after reduction (6+). (* = 1 SePh, # = 6 SePh)

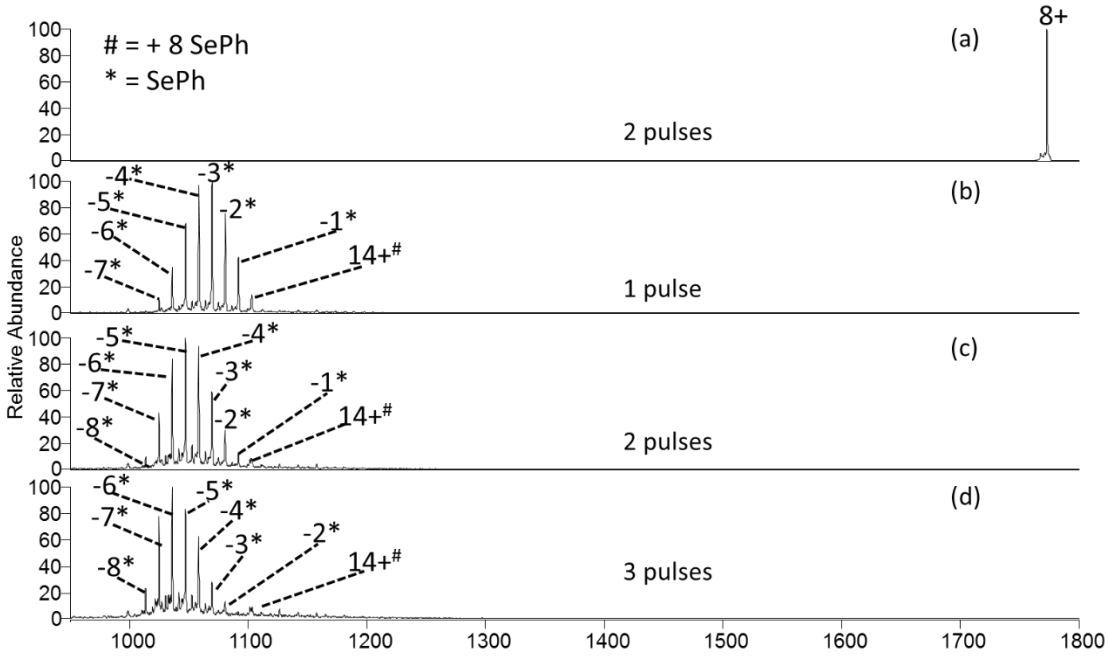


Figure 1.12. UVPD mass spectra of α-lactalbumin for (a) unmodified (8+), and for NPSP-modified after reduction (14+) using (b) 1 pulse, (c) 2 pulses, and (d) 3 pulses.

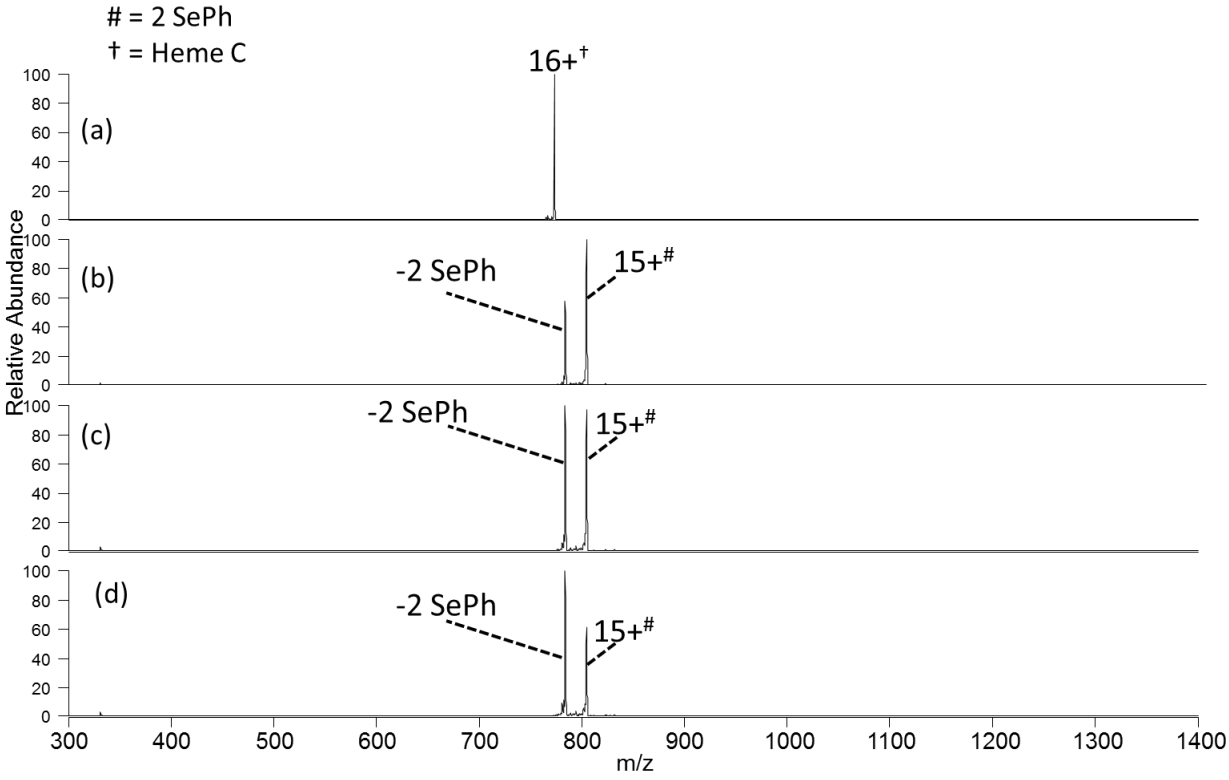


Figure 1.13. UVPD mass spectra of cytochrome c for (a) unmodified (16+) using 2 pulses, and for NPSP-modified cytochrome c after reduction (15+) using (b) 1 pulse, (c) 2 pulses, and (d) 3 pulses.

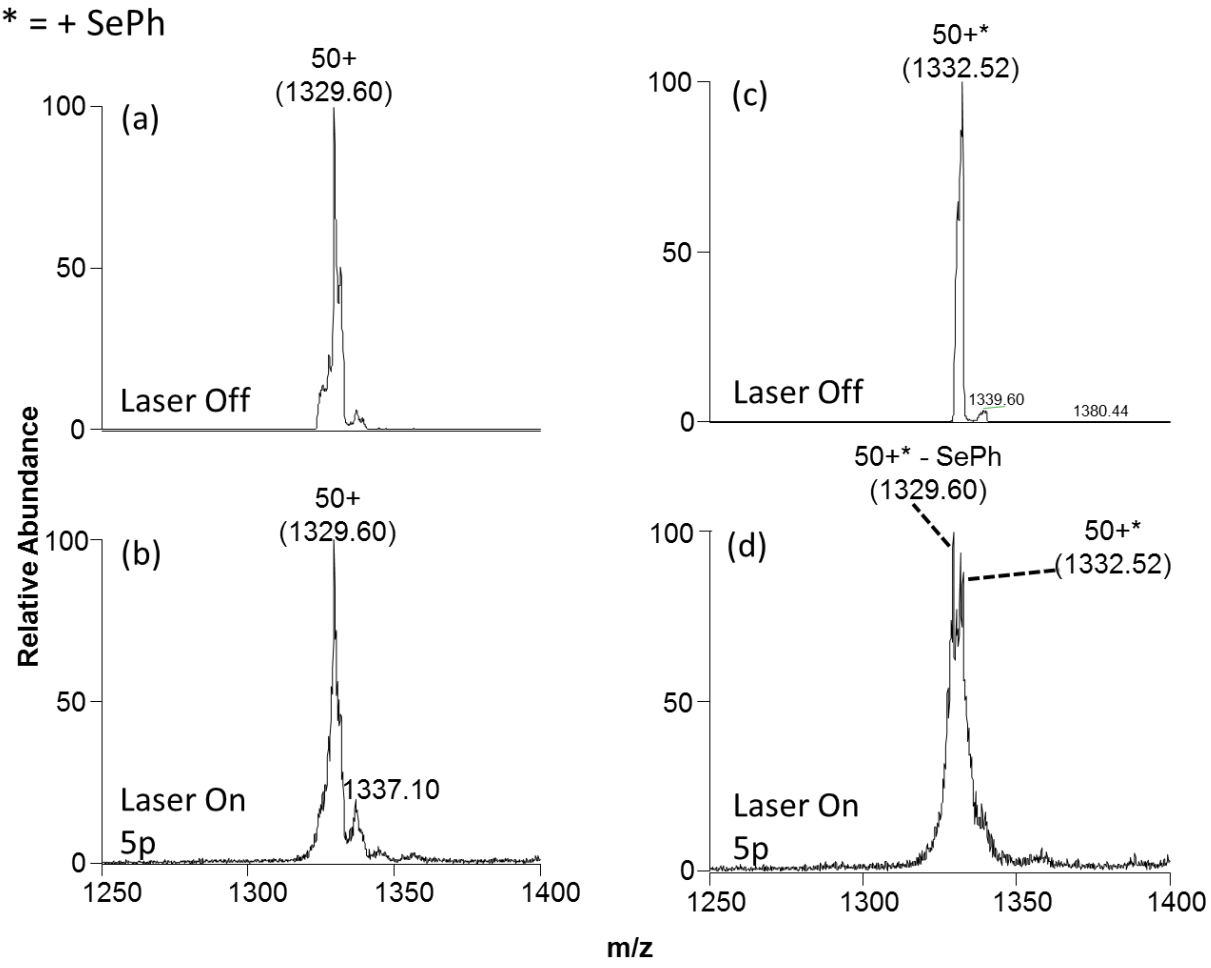


Figure 1.14. Isolation of the (a) 50+ and (c) 50+* charge states of BSA and 266 nm UVPD (5 laser pulses) of the (b) 50+ and (d) 50+* species.

Among the proteins examined, ribonuclease A yielded discrepancies in the characterization of cysteine content via the UVPD method compared to the result obtained from the mass shift observed in the mass spectrum. For ribonuclease A, only seven neutral losses were observed with high confidence upon UVPD (**Figure 1.10b**), whereas the NPSP-modified protein contains eight tags and is thus expected to lose eight tags. One hypothesis to explain the fact that only seven out of eight Se-S bonds were cleaved in ribonuclease A is that the presence of the aromatic amino acids tyrosine,

tryptophan and phenylalanine, all absorb 266 nm photons (with tryptophan and tyrosine having significantly larger photoabsorption cross-sections than phenylalanine and thus absorbing better). Consider the comparison of 266 nm UVPD spectra for NPSP-tagged β-lactoglobulin, lysozyme, ribonuclease A, and α-lactalbumin (**Figures 1.9, 1.10, 1.12**). Each of these proteins has a number of aromatic residues (tryptophan/tyrosine/phenylalanine) that may absorb 266 nm photons: 2/4/4 (β-lactoglobulin), 6/3/3 (lysozyme), 0/6/3 (ribonuclease A), and 4/4/4 (α-lactalbumin) for tryptophan/tyrosine/phenylalanine residues. Based on absorbance profiles of amino acids in solution, it is anticipated that the photoabsorption cross-section for tryptophan in the gas phase is likely greater than that of tyrosine at 266 nm, and the photoabsorption cross-section for phenylalanine is expected to be rather low at 266 nm (these remarks are derived from solution profiles, not the gas phase [35]). Excitation energy transfer has been shown to occur between tryptophan or tyrosine and disulfide bonds, ultimately resulting in homolytic cleavage of the disulfide bond via an excited state [36]. We speculate that a similar phenomenon may occur for the NPSP-tagged proteins. For example, lysozyme and α-lactalbumin contain multiple tryptophan residues which may enhance S-Se bond cleavage from an excited state induced by absorption of 266 nm photons (similar to that shown for disulfide bonds [33]). The lack of tryptophan residues in ribonuclease A may explain the inhibition of tag loss for the protein. Additionally, lysozyme appears to have a less efficient SePh tag loss series than α-lactalbumin, yet both proteins contain the same number of cysteine residues. Lysozyme has six tryptophans in its primary sequence compared to α-lactalbumin which has only four tryptophans, and this may contribute to a greater absorption cross-section for lysozyme and may lead to fragmentation by other pathways. While our results indicate that the photoabsorption cross-section of the benzeneselenol group is significantly greater than that of the aromatic

side-chains at 266 nm, the availability of other absorbing moieties may inhibit S-Se cleavage by affording access to other fragmentation pathways, specifically pathways caused upon photoabsorption by the aromatic side-chains. The availability of other fragmentation pathways for lysozyme was further supported by activation of non-reduced SePh-tagged lysozyme (**Figure 1.9b**). For this protein, in addition to the characteristic SePh loss, an unexpected loss of 76 Da was also observed suggesting an alternative fragmentation route.

Other Activation Methods (HCD, ETD, CID)

It was previously reported that NPSP-derivatized peptides undergo Se-S cleavage upon ETD or CID [28]. Thus, for comparative purposes collision-based (CID and HCD) and electron-based (ETD) methods were also used to activate the SePh-tagged proteins in the present study. Examples of the resulting MS/MS spectra are shown in **Figures 1.15** for aprotinin (with six SePh tags) and **Figure 1.16** for cytochrome C (with two SePh tags). Neither HCD nor CID nor ETD resulted in efficient Se-S cleavage. ETD promoted cleavage of up to two S-Se cleavages in conjunction with charge reduction for tagged aprotinin; the analogous MS/MS spectra for cytochrome c were not readily interpretable. Based on this comparison, 266 nm UVPD showed remarkably high efficiency and selectivity for Se-S cleavage relative to the other activation methods.

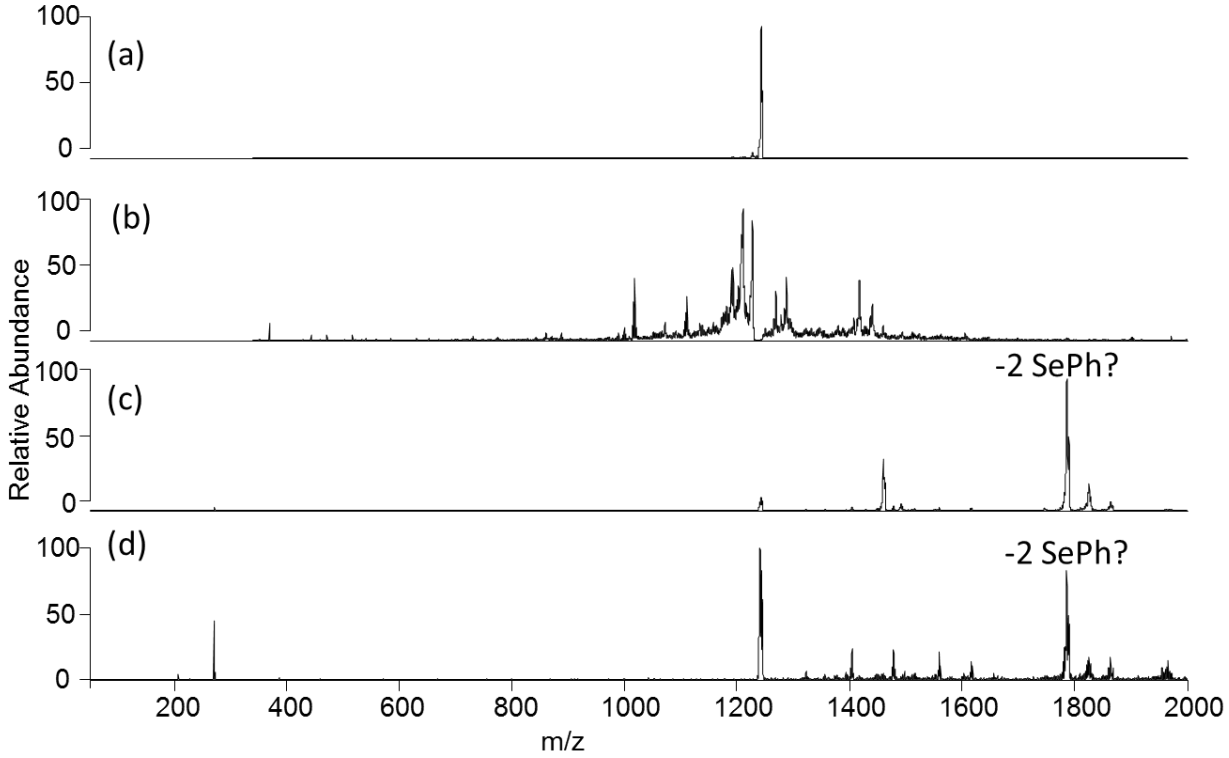


Figure 1.15. Activation of NPSP-modified aprotinin (6+) via (a) HCD (NCE 35), (b) CID (NCE 35), and ETD with an activation time of (c) 40 ms and (d) 100 ms.

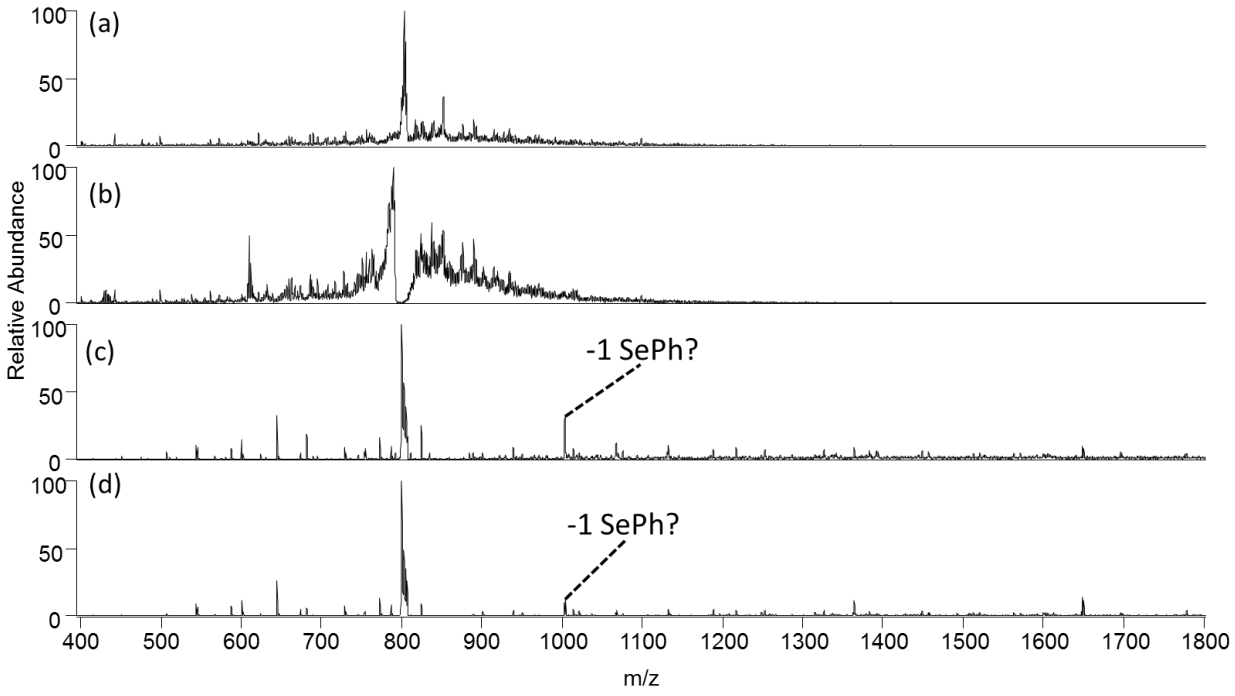


Figure 1.16: Activation of NPSP-modified cytochrome c (15+) via (a) HCD (NCE 35), (b) CID (NCE 35), and ETD with an activation time of (c) 40 ms and (d) 100 ms.

CONCLUSION 1.5

The NPSP-derivatization strategy and 266 nm UVPD proved successful as a new means to count free and bound cysteines in proteins. Proteins containing up to eight cysteine residues were successfully characterized. The SePh tag served as an excellent chromophore for absorption of 266 nm photons, and the selective cleavage of the Se-S bond was striking. Tracking free and bound cysteines has numerous applications in proteomics and offers opportunities for incorporation in informatics engines. For example, the last column of **Table 1.1** shows the percentages of proteins in each proteome which have the same cysteine content (free versus bound cysteines) as each protein included in this study. On average

about 0.1% of all possible proteins match each combination of free and bound cysteines, thus illustrating that characterizing cysteine content offers a significant way to constrain protein identification in database searches.

Chapter Two: Cysteine-Selective Peptide Identification: Selenium-Based Chromophore for Selective S-Se Bond Cleavage with 266 nm Ultraviolet Photodissociation (UVPD)

OUTLINE 2.1

The tremendous number of peptides identified in current bottom-up mass spectrometric workflows, although impressive for high throughput proteomics, results in little selectivity for more targeted applications. We describe a strategy for cysteine-selective proteomics based on a tagging method that installs a S-Se bond in peptides that is cleavable upon 266 nm ultraviolet photodissociation (UVPD). The alkylating reagent, *N*-(phenylseleno)phthalimide (NPSP), reacts with free thiols in cysteine residues and attaches a chromogenic benzeneselenol (SePh) group. Upon irradiation of tagged peptides with 266 nm photons, the S-Se bond is selectively cleaved, releasing a benzeneselenol moiety corresponding to a neutral loss of 156 Da per cysteine. Herein we showcase a new MS/MS scan mode, UVPDnLossCID, which facilitates selective screening of cysteine-containing peptides. A "prescreening" event occurs by activation of the top N peptide ions by 266 nm UVPD. Peptides exhibiting a neutral loss corresponding one or more SePh groups are reactivated and sequenced by CID. Because of the low frequency of cysteine in the proteome, unique cysteine-containing peptides may serve as surrogates for entire proteins. UVPDnLossCID does not generate as many PSMs as conventional bottom-up methods, however UVPDnLossCID provides far greater selectivity.

INTRODUCTION 2.2

Recent advances in bottom-up proteomics strategies have led to unsurpassed numbers of peptide and protein identifications from complex cell lysates [34, 35]. In a typical bottom-up proteomic workflow, proteins are subjected to proteolytic digestion, and the resulting peptides are separated chromatographically prior to ionization and identification by tandem mass spectrometry (MS/MS). Collision induced dissociation (CID) [36], including the higher-energy beam type method termed HCD [37], and electron-transfer dissociation (ETD) [38, 39] are the most popular methods for activation and fragmentation of peptides, and numerous database search methods have been adapted for the resulting MS/MS spectra to facilitate identification of the peptides and assignment of their parent proteins. Currently it is possible to identify over ten thousand peptides in a standard LC-MS/MS run routinely, corresponding to thousands of proteins [40]. Ultraviolet photodissociation (UVPD) is a more recent activation technique that has gained traction and has been applied to a broad range of proteomics applications [41–48]. In particular, UVPD using 193 nm photons generates very rich fragmentation patterns [43, 48], providing high sequence coverage of peptides.

Despite the impressive level of performance obtained by bottom-up proteomics approaches, there remain a number of challenges. Several obstacles are related to the concept of selectivity, either in terms of streamlining the massive LC-MS/MS datasets to focus on specific proteins or features of interest, or with respect to redundancies or ambiguities in the thousands of spectra obtained in each run. For example, one issue encountered in bottom-up proteomics approaches arises from the presence of

confounding peptides, meaning peptides that are characteristic of more than one protein [49]. These ambiguous peptides occur when two or more proteins yield the same proteolytic product. A related problem occurs upon MS/MS analysis of co-eluting peptides, particularly ones that generate ions with overlapping *m/z* values [50, 51]. The resulting fragmentation patterns which contain product ions from more than one peptide are referred to as chimera spectra and increase false negative identifications by up to twofold by diluting search scores for true positive assignments [52]. Strategies to resolve these issues have employed database filtering prior to the search to incorporate only unique peptides [53–55]. This workaround has shown to lower statistical q-values (not to be confused with ion trap Mathieu stability) by reducing the search space, thus increasing peptide confidence [55–58]. Building selectivity into a proteomic workflow to allow meaningful database filtering to include only those peptides that contain a more specific type of amino acid composition have shown similar types of improvement [58, 59].

Another rationale for building selectivity into bottom-up workflows is to allow enhanced sampling for peptides of interest. Typical data-dependent acquisitions are set up to fragment the top N most abundant ions as the peptides elute from the column [60]. Dynamic exclusion settings prevent re-activation of the most abundant ions in each subsequent spectrum, thus facilitating the analysis of less abundant ions [61]. However, as sample complexity increases it becomes more probable that many peptides will elute simultaneously. Thus, normal dynamic exclusion parameters may overlook less abundant ions, masking features of special interest. Employing selective enrichment methods using an affinity tag or a residue selective resin offers one solution to this problem [62, 63].

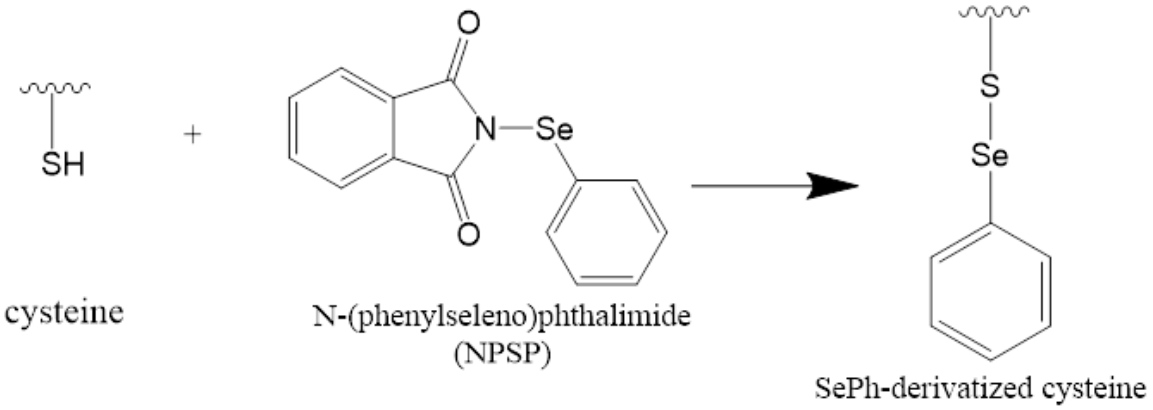
Typically the enrichment methods remove a large portion of peptides from a sample, thus enhancing the detection of the remaining peptides while simultaneously alleviating the production of MS/MS spectra that contain fragment ions from multiple peptides [64]. Enrichment procedures can be somewhat time-consuming and thus are not always feasible or efficient depending on the aim of the proteomics application. Therefore, a similar outcome can be accomplished alternatively by using a method that essentially ignores ions not exhibiting a unique characteristic, such as a neutral loss or isotope pattern [65–67].

Another strategy for incorporating a higher degree of selectivity into a proteomics workflow exploits the use of specific chromophores and UVPD [59, 68–70]. In this case, only peptides that are tagged with the chromophore absorb photons, are activated, and undergo dissociation. This concept can yield a highly selective and targeted approach depending on how the chromophore tags are incorporated, for example via derivatization of a particular amino acid side-chain. The combination of derivatization and UVPD affords a tunable way to streamline the search methods used for peptide identification and can increase the probability of analyzing low abundance peptides. For example, standard tryptic peptides do not absorb photons in the range of 350 nm. Derivatization of the peptides to attach suitable chromophores has been used to convert non-absorbing peptides to highly chromogenic peptides that undergo efficient photodissociation when exposed to 351 nm photons (from an excimer laser) [58, 68, 70]. This concept has been used to target peptides containing histidine or tyrosine via a diazonium reaction [58], as well as cysteine-containing peptides to allow selective analysis of the antigen-binding

regions of antibodies [70]. A similar UVPD method was used to map solvent accessibility of proteins based on analysis of peptides tagged with a chromogenic chemical probe [68]. Strategies using photodissociation at 473 nm have also shown great promise for selective fragmentation of cysteine-containing through cysteinyl alkylation with a Dabcyl group [69]. This strategy has been employed as a photo-selected reaction monitoring (photo-SRM, employing 473 nm photodissociation) experiment and has shown to be as sensitive as conventional SRM experiments for complex human serum samples [71]. There have also been a number of reports showing selective bond cleavages of peptides derivatized with quinone or napththalenethiol tags (among others) upon UVPD at 266 nm, the latter via a type of radical directed dissociation [23, 24, 72–75].

In the present study, we report a strategy that uses 266 nm UVPD to target cysteine-containing peptides via tagging with *N*-(phenylseleno)phthalimide (NPSP) [25, 26] (see **Scheme 2.1**). NPSP is a class of benzeneselenyl radical precursors which have shown to be photo-labile under ultraviolet radiation [32]. Numerous chemical probe strategies have been developed that target specific amino acids in order to map topologies of proteins, to evaluate protein interactions, and to streamline the identification of primary sequences [8, 76–78]. Development of methods to focus on cysteine residues offers a particularly compelling opportunity for enhanced selectivity for bottom-up proteomics. Cysteine (Cys) is the only amino acid that contains a sulfhydryl moiety, thus making it one of the most nucleophilic and reactive amino acids [78]. Cysteine is also one of the rarest amino acids in the proteome; however, due to its importance in

stabilizing protein structure through disulfide bonds and modulating redox-based reactions *in-vivo*, the residue is present in almost all proteins thus making it an ideal candidate for targeted applications. In fact, it is estimated that 97% of the proteins in the human proteome contain at least one cysteine [8]. Furthermore, since peptides containing cysteine are likely to be unique, building a targeted cysteine method allows the potential of whole-protein identification based on surrogate peptides at higher levels of confidence. NPSP has been shown previously to yield fast and efficient derivatization of free cysteine residues [27–30]. Other methods for detecting cysteine-containing residues based on mass shifts due to the loss of different hydroquinone derivatives have also been explored and shown to increase protein scores when incorporated into MASCOT database searches, demonstrating the advantages of selectivity in HPLC-MS/MS workflows [20, 22].



Scheme 2.1. Reaction of free thiols with *N*-(phenylseleno)phthalimide.

METHODS 2.3

Materials and Reagents

N-(Phenylseleno)phthalimide and all proteins were obtained from the Sigma-Aldrich company. High purity water, acetonitrile, and formic acid were obtained from ThermoFisher Scientific. Molecular weight cutoff filters were obtained from Pierce.

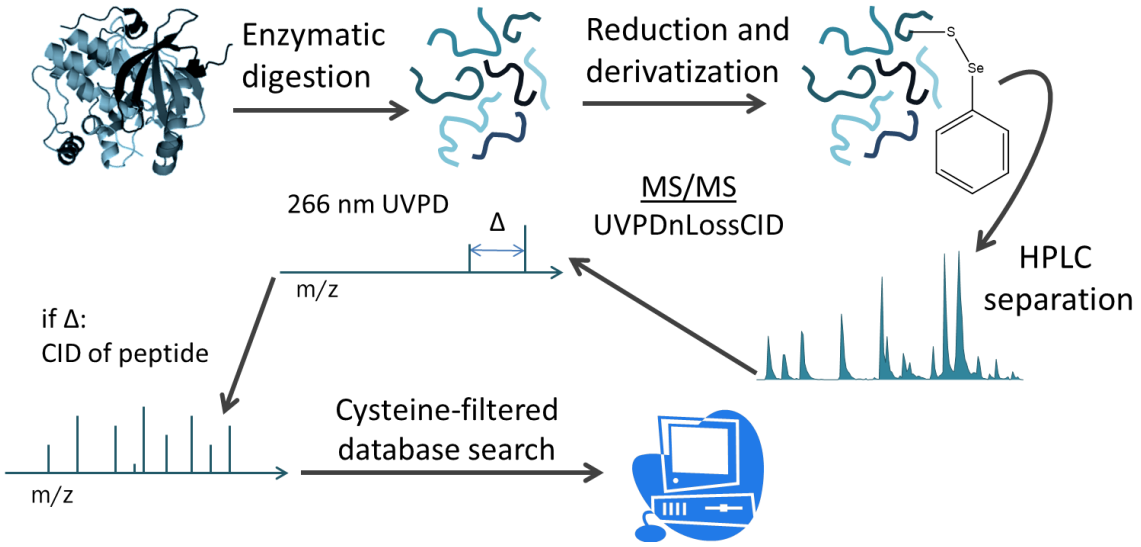
Chemical Modification and Preparation

Single proteins were digested by trypsin in a ratio of 1:50 protease:protein for 6 hours at 37ºC. Protein mixtures of horse cytochrome C and myoglobin and bovine α-lactalbumin, bovine β-lactoglobulin A/B, bovine carbonic anhydrase II, bovine fetuin B, and bovine serum albumin (BSA) were digested in a 1:20 protease:protein ratio for 18 hours at 37ºC. Following digestion, samples were subjected to reduction with 5 mM DTT at 55ºC for 1 hour. Samples were desalted using Pierce C18 spin column and then reconstituted in 50/50/1 acetonitrile/water/formic acid and reacted for 30 seconds with a 10 molar excess of *N*-(phenylseleno)phthalimide (NPSP) per cysteine residue (**Scheme 2.1**). An addition of 1% formic acid was added to all solvents used in the C18 cleanup to prevent disulfide bonds from reformation. Because selenylsulfide (Se-S) and disulfide bonds (S-S) have similar reduction potentials, the NPSP alkylation reaction could not be undertaken in the presence of reducing reagents such as dithiothreitol (DTT) or *tris*(2-carboxyethyl)phosphine (TCEP). In order to achieve successful derivatization of disulfide-containing peptides, a buffer exchange step into formic acid following reduction was implemented.

Chromatography and Mass Spectrometry

CID, HCD, and UVPDnLossCID were carried out on a Velos Pro dual linear ion trap mass spectrometer (Thermo Scientific, San Jose, CA) equipped with a Continuum frequency quadrupled Nd:YAG laser (266 nm) in a manner similar to that described previously [79]. ETD experiments were performed on a LTQ XL linear ion trap mass spectrometer (Thermo Scientific, San Jose, CA). LCMS was performed using a Dionex 3500RSC nanoLC system. The trap and analytical columns were packed in-house with Michrom Magic C_{18} packing material with inner diameters of 150 and 75 μm, respectively. Solvent A was composed of 0.1% formic acid in water, and solvent B had a composition of 0.1% formic acid in acetonitrile**.** Approximately 250 ng (1 μL of a 250 ng/μL digestion) of single protein digests and 500 ng of protein mixture digestions were loaded and separated using a gradient from 2% to 35% B over 75 minutes at a flow rate of 300 nL/min. Dynamic exclusion parameters had a repeat and exclusion duration of 30 seconds with an exclusion list size of 500 and repeat count of 1. For CID, HCD and ETD experiments, the ten most abundant ions in each MS1 scan were activated using an NCE of 35 for HCD and CID. The activation times of ETD and CID were 100 ms and 10 ms, respectively. A new scan mode termed UVPDnLossCID was developed in the data-dependent acquisition code. For UVPDnLossCID the top five most abundant peaks were activated by 266 nm UVPD using one laser pulse at approximately 6 mJ of energy, followed by CID for those peptides which exhibited neutral losses characteristic of the SePh tag. Protein mixtures were prepared and analyzed in quadruplicate, and single protein digests were analyzed in triplicate. The overall experimental workflow is presented in **Scheme 2.2**. In brief, proteins are digested, reduced,

derivatized with NPSP, analyzed by LC-MS/MS using the UVPDnLossCID scan mode, then subjected to a cysteine-selective database search.



Scheme 2.2. Overall experimental workflow for selective characterization of cysteine-containing peptides

Database Searching

Data files were converted to the mzXML format using the software MM File Conversion and submitted to MassMatrix for analysis [80–83]. For UVPDnLossCID data, a filtering step was performed in SEQUEST which removed all UVPD scans leaving only CID MS/MS events in an exported MGF file which was submitted to MassMatrix for analysis. Fragment ion databases were built upon FASTA files obtained from Uniprot, and the resulting theoretical fragment ions were compared to the experimental results. A custom version of the algorithm was implemented which allowed solely cysteine-containing peptides to be searched. A maximum of three modifications per peptide was searched with variable modifications of oxidation of methionine and alkylation of cysteine by NPSP. Peptides with a statistical p value of < 0.05 (as

calculated based on the pp_{tag}, pp2, pp scores, all which are scoring parameters used to estimate the quality of peptide-spectrum matches) [81] were considered significant. Decoy hits were excluded from the reported results.

Safety Considerations

N-(Phenylseleno)phthalimide (NPSP) is a hazardous chemical which is classified as a health hazard having the potential to cause acute toxicity and is also dangerous for the environment. Precautions for working with this chemical should be taken by wearing a dust mask and using a fume hood.

RESULTS 2.4

NPSP Reaction efficiency and MS/MS for Model Peptides

Since this strategy depends on efficient tagging of cysteines, initial efforts focused on optimization of the NPSP derivatization reaction and measurement of reaction efficiency. **Figure 2.1** shows a comparison of the ESI mass spectra obtained for a model peptide, CDPGYIGSR, prior to and after reaction with NPSP. The addition of the benzeneselenol group (ΔM = +155.9 Da) to the free cysteine thiol of CDPGYIGSR is highly efficient with little unreacted peptide observed in **Figure 2.1b**. The relative abundances of the 1+ and 2+ charge states for both the unmodified and modified peptides remain similar, thus indicating that the benzenselenol tag does not have a significant effect on charging of the peptide.

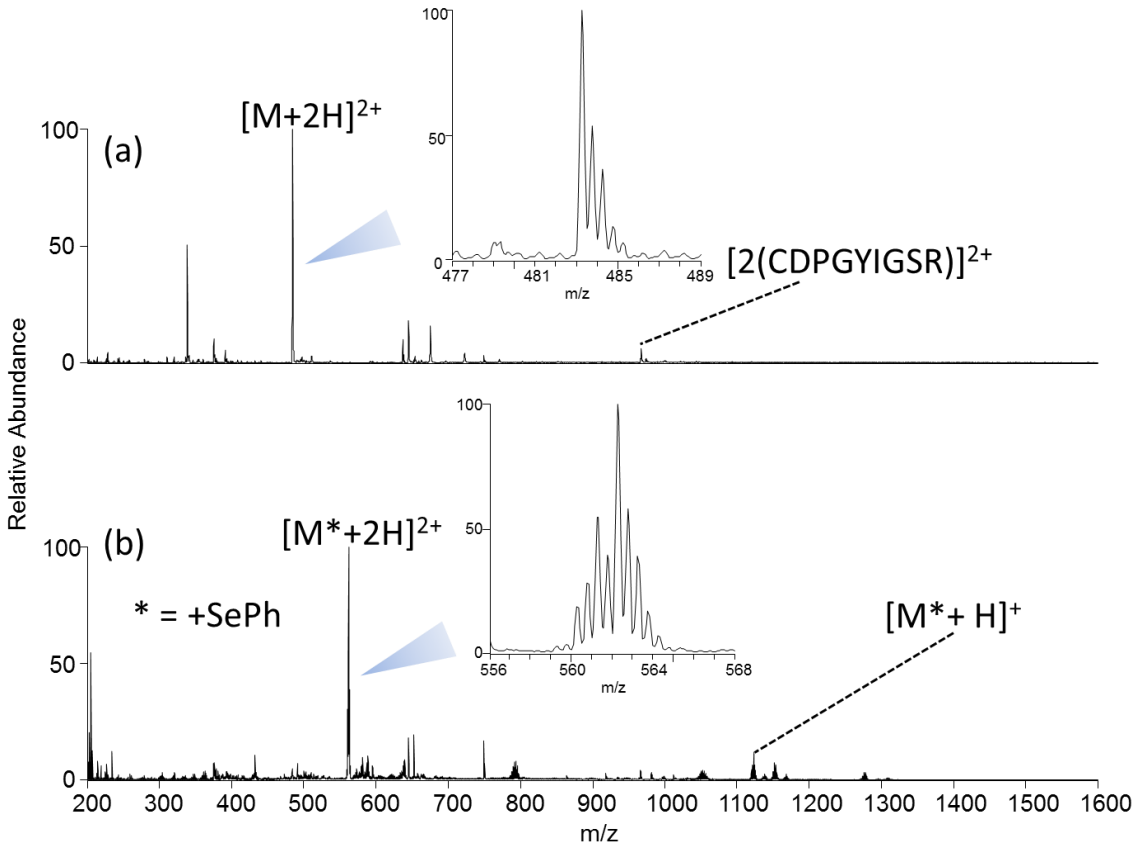


Figure 2.1. ESI mas spectra of (a) unmodified and (b) SePh-modified peptide CDPGYIGSR with insets showing characteristic Se isotope pattern of the 2+ charge state

CID was used to characterize the fragmentation patterns of the unmodified and modified CDPGYIGSR peptide (2+) as shown in **Figure 2.2**. While there are some modest differences in the types and relative abundances of the diagnostic *b*/*y* sequence ions, the overall sequence coverage is comparable for both the unmodified and NPSP-tagged peptides. Owing to the fact that the cysteine is located on the N-terminus, all characteristic *b*-type ions observed in **Figure 2.2b** retain the benzeneselenol group after CID. Thus, the

benzeneselenol group is not labile, and the peptide can be readily identified based on the conventional *b/y* product ions.

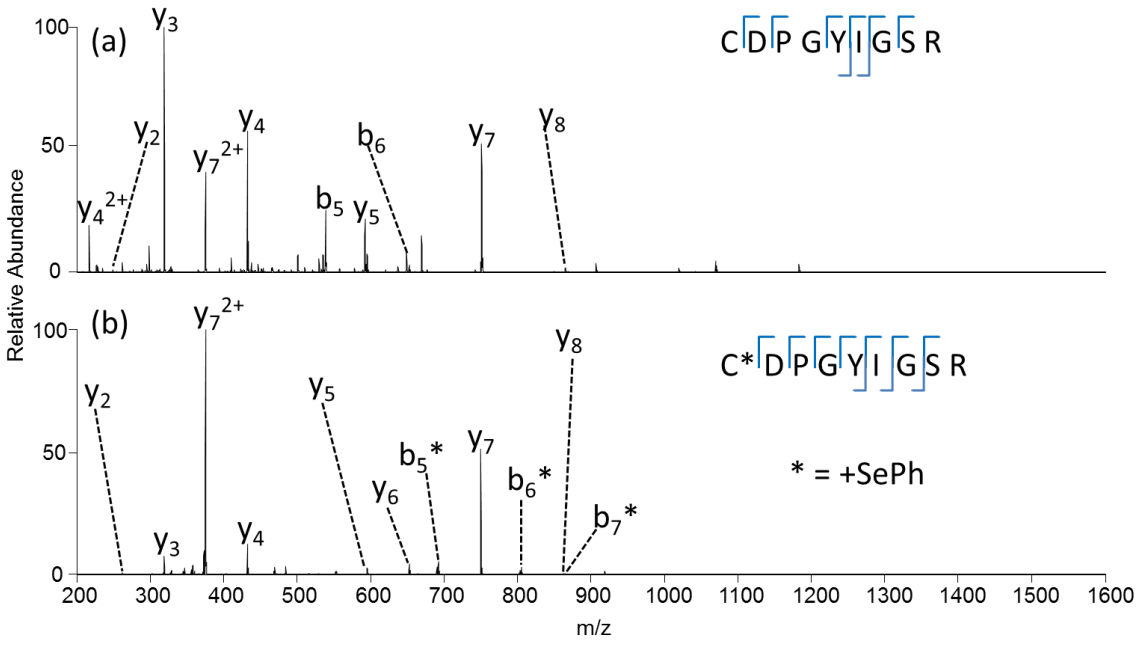


Figure 2.2. CID mass spectra of (a) CDPGYIGSR (2+) and (b) SePh-modified CDPGYIGSR (2+).

Unmodified peptides, as exemplified by CDPGYIGSR (2+) (**Figure 2.3a**), yielded no product ions upon 266 nm UVPD. The peptide backbone does not efficiently absorb at 266 nm, and the sole aromatic chromophore of Tyr (Y) does not result in significant absorption and dissociation upon exposure to a single laser pulse. When the 2+ charge state of the NPSP-modified CDPGYIGSR peptide was subjected to 266 nm UV photoactivation (one laser pulse), the sole prominent product arose from Se-S cleavage corresponding to loss of the benzeneselenol group (**Figure 2.3b**). The loss of the benzeneselenol group results in a product that is 1 Da less than the unmodified peptide, suggesting that 266 nm UVPD occurs

via homolytic cleavage of the Se-S bond to yield a thiol radical which presumably may undergo subsequent electron migration. CID of the resulting Se-S cleavage product (the presumed radical thiol peptide species) was performed, resulting in low abundance ions which were difficult to assign as traditional *a, b, c, x, y, z* ions (spectrum not shown). Nonetheless, irradiation of the NPSP-modified peptides by 266 nm photons resulted in highly selective preferential cleavage of the Se-S bond and loss of the benzeneselenol group, thus affording a highly characteristic neutral loss pathway for NPSP-labelled cysteine-containing peptides.

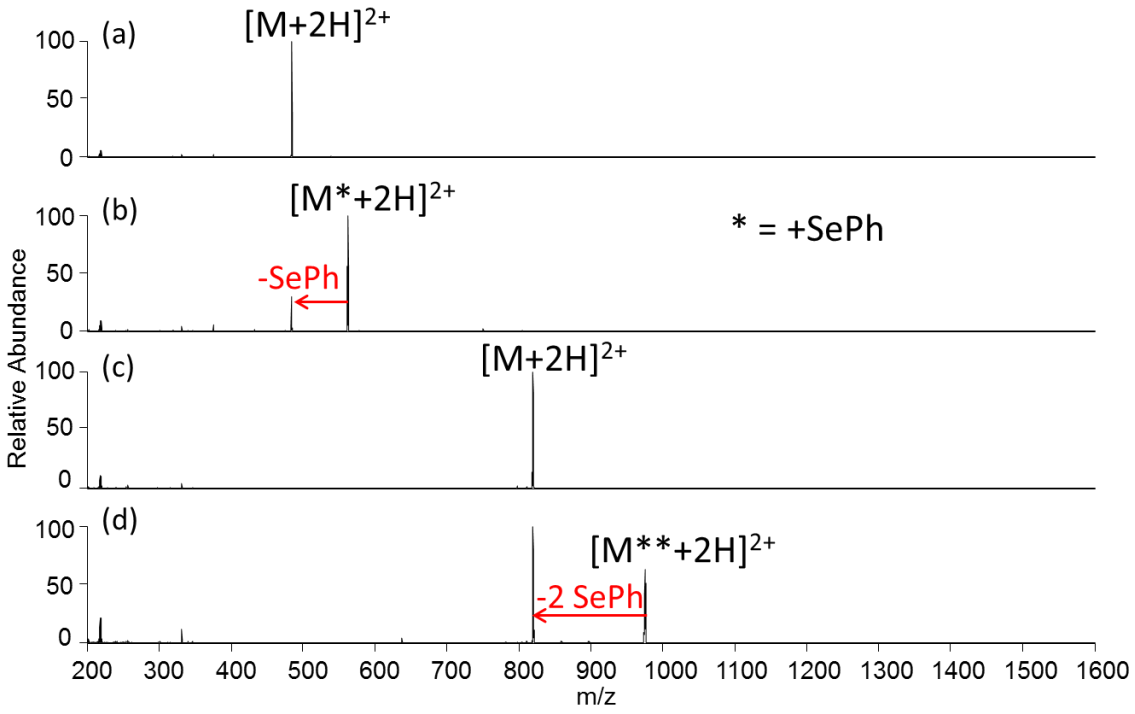


Figure 2.3. 266 nm UVPD mass spectra of (a) unmodified and (b) modified CDPGYIGSR (2+) as well as (c) unmodified and (d) modified somatostatin AGCKNFFWKTFTSC (2+) exhibiting the characteristic neutral loss of the SePh group from the modified peptides.

As another example, somatostatin (AGCKNFFWKTFTSC) contains a disulfide bond between Cys3 and Cys14 and was used to optimize the reaction efficiency for disulfide-bound cysteine residues. Upon reaction with NPSP, two benzeneselenol (Δm = 311.9 Da) groups were appended to somatostatin (**Figure 2.4**), and the bis-labelled peptide was the dominant product. The doubly NPSP-modified AGCKNFFWKTFTSC peptide was subjected to 266 nm UVPD, and the resulting mass spectrum is shown in **Figure 2.3d** (and 266 nm UVPD of unmodified somatostatin is shown in **Figure 2.3c**). The exclusive product arose from the loss of both benzeneselenol tags via cleavage of both Se-S bonds. Interestingly, a product corresponding to cleavage of a single Se-S bond was not observed. The clean loss of both tags from the doubly-modified peptide was an unexpected and striking result that attests to the lability of the Se-S bonds upon UVPD. Comparison of the UVPD mass spectra in **Figure 2.3d** for the doubly-modified peptide (AGC*KNFFWKTFTSC*) to that observed for the singly-modified peptide (C*DPGYIGSR) in **Figure 2.3b** demonstrates that the UVPD efficiency was significantly greater for the doubly NPSP-modified peptide. We rationalize this outcome owing to greater molar absorptivity of the doubly-modified peptide due to the addition of the second SePh group. These proof-of-principle experiments with model peptides demonstrated that 266 nm UVPD promoted highly efficient Se-S cleavage and thus proved to be very selective for cysteine-containing peptides that contained the benzeneselenol tag. Using UVPD as the first step of a data-dependent MS/MS strategy allows ready pinpointing of the cysteine-containing peptides based on the characteristic loss of the SePh groups, then CID is subsequently

triggered for those peptides to characterize the sequences. The net method is termed UVPDnLossCID.

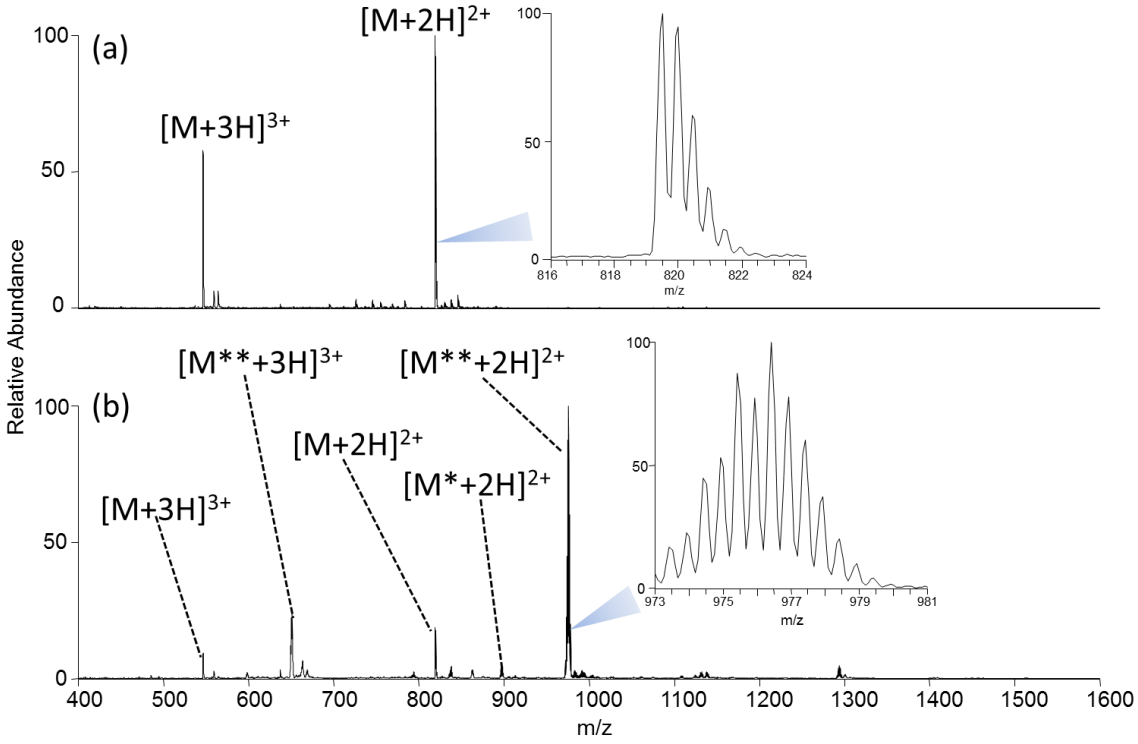


Figure 2.4. ESI mass spectra of (a) unmodified and (b) SePh-modified peptide somatostatin (AGCKNFFWKTFTSC) with insets showing unique Se isotope pattern of the 2+ charge state

Previous reports have indicated that ETD also provides a way to cleave S-Se bonds [28], and thus the performance of ETD was compared to UVPD for both singly modified CDPGYIGSR (**Figure 2.5a**) and bis-modified AGCKNFFWKTFTSC (somatostatin) (**Figure 2.5b**). For both singly NPSP-modified CDPGYIGSR and doubly NPSP-modified somatostatin, the dominant pathway upon ETD was charge reduction and S-Se cleavage resulting in the neutral loss of the benzeneselenol group. A few sequence-type ions were observed, as well as notable ETnoD products. The loss of both tags from the bis-labeled

peptide is a minor pathway (less than 1% abundance relative to the precursor. The efficiency of the S-Se cleavage is lower for ETD than for UVPD. UVPD offered a more efficient means to cleave all S-Se bonds simultaneously in a faster activation period and without any charge-state dependence, so it was used as the activation method of choice for the rest of the study.

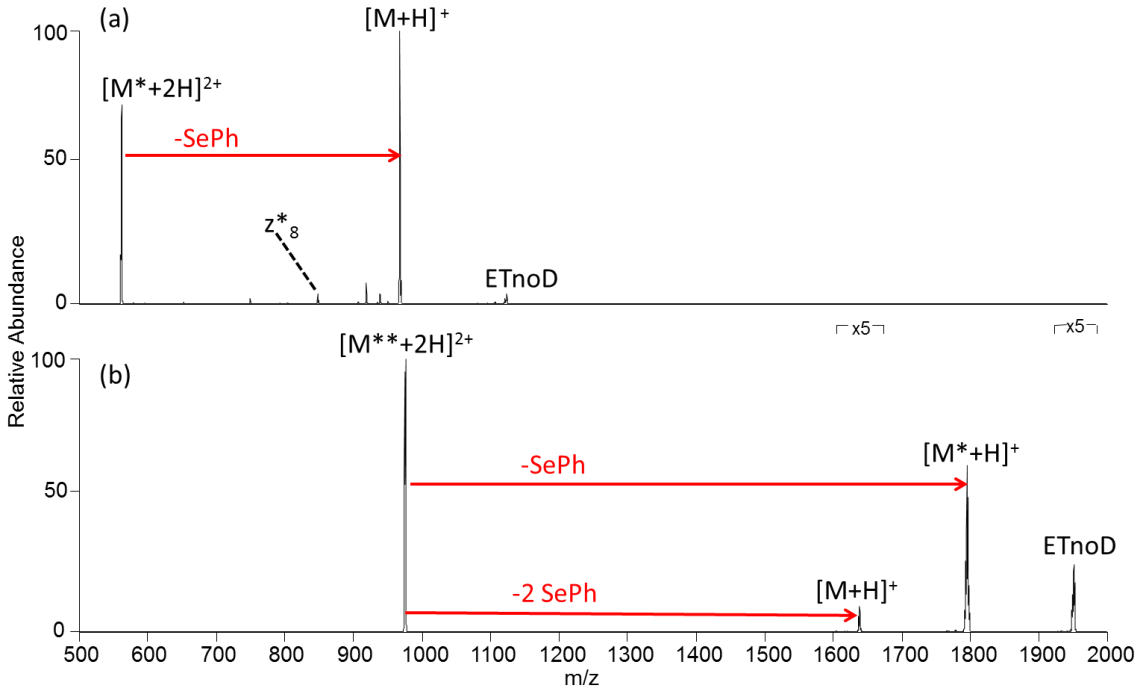
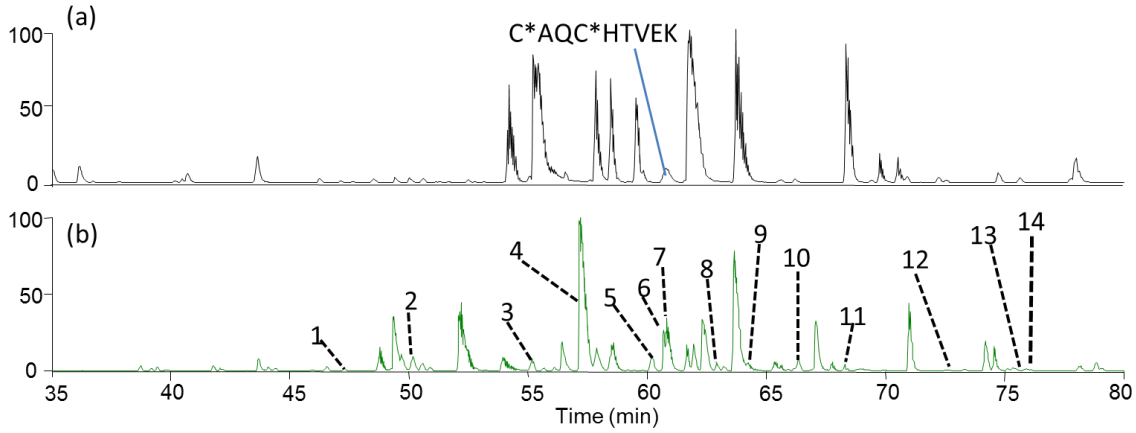


Figure 2.5. ETD mass spectra of (a) SePh-modified CDPGYIGSR (2+, 100 ms) and (b) SePh-modified AGCLNFFWKTFTSC (somatostatin) (2+, 100 ms) showing charge reduction and neutral losses of SePh.

Derivatization of Protein Digests

The NPSP reaction and UVPDnLossCID method was optimized for identification of cysteine-containing peptides based on examination of tryptic digests of cytochrome C and bovine serum albumin (BSA). Using the new UVPDnLoss CID scan mode, any ion

displaying the targeted neutral loss corresponding to the cleavage of one or more SePh tags was selected for characterization in a second CID event (NCE 35). Neutral losses for peptide ions containing up to two cysteine modifications of NPSP were monitored as follows: -156 or -312 Da for 1+ peptides, -78 or -156 Da for 2+ peptides, and -52 or -104 for 3+ peptides. In brief, typical experiments involved 266 nm UVPD, followed by isolation and CID of any ion which exhibited one of the specified neutral losses upon UVPD. The base peak chromatogram of the cyt C digest is shown in **Figure 2.6a**, and another is shown in **Figure 2.6b** for a BSA digest with all cysteine-containing peptides labelled. As a simple example, cytochrome c contains two cysteine residues, and tryptic digestion results in a single peptide containing both cysteine residues (CAQCHTVEK). After reduction of the peptide, alkylation of CAQCHTVEK with NPSP yielded a mass shift of 311.9 Da. The bis-labeled tryptic peptide (C*AQC*HTVEK) is shown in **Figure 2.7a.** The UVPD spectrum is shown in **Figure 2.7b** and exhibits a single product due to cleavage of both SePh tags, and the subsequent neutral-loss triggered CID spectrum of the original tagged peptide is displayed in **Figure 2.7c**. The protein BSA, containing 35 cysteine residues and 17 disulfide bonds, provided a larger number of cysteine-containing peptides (as labeled in **Figure 2.6b**). Fourteen of the cysteine-containing tryptic peptides generated from BSA were successfully characterized using the UVPDnLossCID method. The complete lists of all peptides identified in the tryptic digests of cyt C and BSA are provided in **Tables 2.1** and **2.2**.



1	C*ASIQK
2	C*C*AADDK
3	C*C*TESLVNR
4	LC*VLHEK
5	YIC*DNQDTISSK
6	DDPHAC*YSTVFDK
7	YNGVFQEC*CQAEDKGACLLPK

8	C*CAADDKEACFAVEGPK
9	LKEC*C*DKPLLEK
10	GAC*LLPK
11	RPC*FSALTPDETYVPK
12	LKPDPNTLC*DEFK
13	SLHTLFGDELC*K
14	DAIPENLPPLTADFAEDKDVC*K

Figure 2.6. LC traces of tryptic digests of (a) cytochrome C (horse) and (b) bovine serum albumin. The cysteine-containing SePh-modified peptides are labelled, and the ones for BSA are listed in the table below the LC trace.

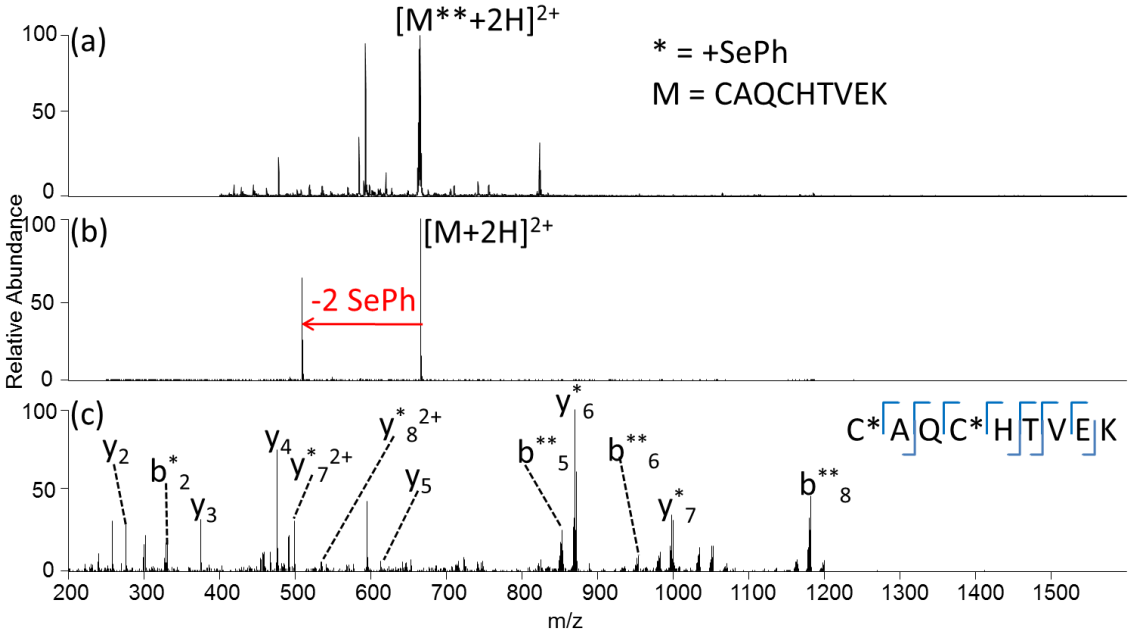


Figure 2.7. (a) ESI mass spectrum of the tryptic peptide CAQCHTVEK eluting at 61.13 minutes in the LC trace shown in Figure 2a, and (b) UVPD mass spectrum of C*AQC*HTVEK (2+), and (c) UVPDnLossCID spectrum.

Table 2.1. All identified peptides for Cytochrome C (**Figure 2.6a**) from MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Min)	Charge States
CAQCHTVEK	41.95	2+
CAQCHTVEK seph(1)	42.95	2+
CAQCHTVEK seph(1) seph(4)	70.35	2+ 1+
CAQCHTVEK seph(4)	47.73	2+
CAQCHTVEKGGKHK	81.81	3+
EDLIAYLK	80.46	2+ 1+
EDLIAYLKK	63.61	1+ 2+
EETLMEYLENPK	74.34	2+ 3+ 1+
EETLMEYLENPK OxiM(5)	60.67	2+ 3+
EETLMEYLENPKK	70.07	3+ 2+ 1+
EETLMEYLENPKK OxiM(5)	73.4	3+ 2+
EETLMEYLENPKKYIPGTK OxiM(5)	66.68	3+
GITWKEETLMEYLENPK	76.59	3+ 2+
GITWKEETLMEYLENPK OxiM(10)	70.9	3+ 2+
GITWKEETLMEYLENPKK	73.16	3+ 2+
GITWKEETLMEYLENPKK OxiM(10)	67.22	3+ 2+
IFVQKCAQCHTVEK	49.26	3+
IFVQKCAQCHTVEK seph(6)	59.08	3+
IFVQKCAQCHTVEK seph(9)	55.3	3+
KIFVQK	45.04	1+
KTEREDLIAYLK	61.06	3+ 2+ 1+
KTGQAPGFTYTDANK	53.84	3+ 2+
KYIPGTK	41.63	2+ 1+
MGDVEKGK	46.29	2+
MGDVEKGKK	81.97	2+
MIFAGIK	62.92	1+
MIFAGIKK	55.72	2+
MIFAGIKK OxiM(1)	49.98	2+
MIFAGIKKK	47.69	2+
NKGITWK	65.13	2+
NKGITWKEETLMEYLENPK OxiM(12)	69.85	3+
TEREDLIAYLK	65.68	2+ 3+
TEREDLIAYLKK	60.3	3+ 2+
TGPNLHGLFGR	67.83	2+ 1+

Table 2.1 (continued)

TGQAPGFTYTDANK	44.87	2+ 1+ 3+
YIPGTK	43.23	1+

Table 2.2. All identified peptides for BSA (**Figure 2.6b**) from MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Min)	Charge States
ADLAKYICDNQDTISSK seph(8)	69.57	3+
AEFVEVTK	43.61	2+ 1+
AEFVEVTKLVTDLTK	74.44	2+ 3+
ALKAWSVAR	45	2+
ATEEQLK	30.08	2+ 1+
ATEEQLKTVMENFVAFVDK	82.73	3+ 2+
ATEEQLKTVMENFVAFVDK OxiM(10)	70.42	3+ 2+
AWSVAR	40.07	1+
CASIQK seph(1)	47.06	1+
CASIQKFGER seph(1)	32.97	3+
CASIQKFGERALK	57.64	3+
CCAADDK seph(1) seph(2)	43.42	2+
CCAADDKEACFAVEGPK seph(1)	62.15	3+
CCAADDKEACFAVEGPK seph(1) seph(2) seph(10)	82.1	3+
CCAADDKEACFAVEGPK seph(10)	63.51	3+
CCAADDKEACFAVEGPK seph(2)	60.14	3+
CCTESLVNR	42.64	2+
CCTESLVNR seph(1) seph(2)	77.09	2+
CCTESLVNRR seph(1)	37.32	3+
CCTESLVNRR seph(2)	41.28	3+
CCTKPESER	31.38	2+
CCTKPESERMPCTEDYLSLILNR	76.65	3+
DAFLGSFLYEYSR	92.88	1+ 3+ 2+
DAIPENLPPLTADFAEDK	66.92	3+ 2+
DAIPENLPPLTADFAEDKDVCK seph(21)	77.2	3+ 2+
DAIPENLPPLTADFAEDKDVCKNYQEAK seph(21)	73.08	3+
DDPHACYSTVFDK seph(6)	66.02	3+ 2+
DDPHACYSTVFDKLK seph(6)	69.66	3+
DDSPDLPK	37.87	2+ 1+
DLGEEHFK	42.15	2+ 1+
EACFAVEGPK seph(3)	66.84	2+
ECCDKPLLEK	42.64	2+
ECCHGDLLECADDR	91.62	3+
ECCHGDLLECADDR seph(10)	67.32	3+

Table 2.2 (continued)

ECCHGDLLECADDR seph(2)	78.86	3+
ECCHGDLLECADDR seph(3)	80.31	3+
EKVLASSAR	78.71	2+
ETYGDMADCCEK OxiM(6)	36.03	2+
ETYGDMADCCEKQEPER OxiM(6)	37.62	3+
EYEATLEECCAK	50.4	2+
EYEATLEECCAK seph(9) seph(10)	42.53	3+
EYEATLEECCAKDDPHACYSTVFDK seph(18)	65.64	3+
EYEATLEECCAKDDPHACYSTVFDK seph(9)	65.94	3+
FGERALK	59.81	2+
FKDLGEEHFK	45.87	2+ 3+
FPKAEFVEVTK	51.58	3+ 2+
GACLLPK seph(3)	29.67	2+
GACLLPKIETMREK	74.62	3+
GACLLPKIETMREK OxiM(11)	69.57	3+
GLVLIAFSQYLQQCPFDEHVK seph(14)	92.37	3+
HLVDEPQNLIK	50.68	2+ 1+ 3+
HLVDEPQNLIKQNCDQFEK seph(14)	69.9	3+ 2+
HPEYAVSVLLR	58.16	2+ 1+ 3+
HPYFYAPELLYYANK	69.99	3+ 2+
IETMREK	29.9	2+
IETMREK OxiM(4)	28.13	2+
KQTALVELLK	56.62	2+ 1+
KVPQVSTPTLVEVSR	57.22	3+ 2+
KVPQVSTPTLVEVSRSLGK	58.74	3+
LCVLHEK	39.25	2+
LCVLHEK seph(2)	57.43	2+
LCVLHEKTPVSEK seph(2)	55.35	3+
LCVLHEKTPVSEKVTK	39.79	3+
LCVLHEKTPVSEKVTK seph(2)	53.81	3+
LFTFHADICTLPDTEK seph(9)	80.13	2+ 3+
LGEYGFQNALIVR	63.94	2+ 1+ 3+
LKECCDK seph(5)	53.77	2+
LKECCDKPLLEK	43.98	3+ 2+
LKECCDKPLLEK seph(4) seph(5)	70.61	3+
LKHLVDEPQNLIK	49.38	2+ 3+
LKPDPNTLCDEFK seph(9)	74.05	2+ 3+
LKPDPNTLCDEFKADEK	49.64	2+
LKPDPNTLCDEFKADEK seph(9)	70.29	3+ 2+

Table 2.2 (continued)

LSQKFPK	35.87	2+
LSQKFPKAEFVEVTK	50.6	2+ 3+
LVNELTEFAK	57.53	2+ 1+
LVTDLTK	41.82	1+
LVVSTQTALA	53.01	1+ 2+
MPCTEDYLSLILNR	81	2+ 3+
MPCTEDYLSLILNR OxiM(1)	78.57	3+ 2+
MPCTEDYLSLILNR OxiM(1) seph(3)	82.29	2+ 3+
MPCTEDYLSLILNR seph(3)	77.55	2+
NECFLSHKDDSPDLPK seph(3)	60.61	3+ 2+
NYQEAKDAFLGSFLYEYSR	71.96	3+ 2+
PCFSALTPDETYVPK	87.73	2+
PCFSALTPDETYVPK seph(2)	74.14	3+ 2+
PDPNTLCDEFK	50.45	2+
PDPNTLCDEFK seph(7)	72.99	3+ 2+
PDPNTLCDEFKADEK seph(7)	70.46	2+ 3+
PLLEKSHCIAEVEK	35.55	3+
QEPERNECFLSHKDDSPDLPK seph(8)	58.25	3+
QIKKQTALVELLK	54.6	3+
QNCDQFEK	40.02	2+
QNCDQFEK seph(3)	55.18	2+
QNCDQFEKLGEYGFQNALIVR seph(3)	77.64	3+
QRLRCASIQK	47.15	3+
QTALVELLK	62.8	2+ 1+
QTALVELLKHKPK	51.46	3+
RHPEYAVSVLLR	78.92	3+ 2+
RHPEYAVSVLLRLAK	52.67	3+
RHPYFYAPELLYYANK	64.89	3+ 2+
RPCFSALTPDETYVPK	74.69	3+ 2+
RPCFSALTPDETYVPK seph(3)	78.74	3+ 2+
SHCIAEVEK seph(3)	52.43	2+
SLGKVGTRCCTK	44.13	3+
SLGKVGTRCCTK seph(10)	40.53	3+
SLHTLFGDELCK	91.84	3+ 2+
SLHTLFGDELCK seph(11)	75.76	3+ 2+ 1+
TCVADESHAGCEK	31.69	3+
TCVADESHAGCEK seph(11)	35.88	3+
TCVADESHAGCEK seph(2) seph(11)	60.61	3+ 2+
TPVSEK	28.15	1+

Table 2.2 (continued)

TPVSEKVTK	30.78	2+
TPVSEKVTKCCTESLVNR	76.67	3+
TVMENFVAFVDK	71.65	2+ 1+ 3+
TVMENFVAFVDK OxiM(3)	61.97	2+ 3+
TVMENFVAFVDKCCAADDK OxiM(3)	64.02	3+
TVMENFVAFVDKCCAADDKEACFAVEGPK + seph(13, 14, 22)	92.55	3+
VPQVSTPTLVEVSR	55.26	2+ 3+
VTKCCTESLVNR seph(4)	54.83	3+
VTKCCTESLVNR seph(4) seph(5)	40.67	3+
YICDNQDTISSK	38.26	2+
YICDNQDTISSK seph(3)	61.9	2+
YLYEIAR	49.79	2+ 1+
YLYEIARR	45.55	2+
YNGVFQECCQAEDK	51.27	2+
YNGVFQECCQAEDKGACLLPK seph(17)	72.4	3+
YNGVFQECCQAEDKGACLLPK seph(8)	72.76	3+

A more complex mixture comprised of eight protein inclusing cytochrome C, myoglobin, α-lactalbumin, β-lactoglobulin A/B, carbonic anhydrase II, fetuin B, and bovine serum albumin (BSA) was digested, derivatized, and analyzed. The resulting analysis of the number of unique cysteine-containing PSMs and the total number of unique PSMs (i.e. both cysteine-containing and non-cysteine-containing peptides) for this mixture is shown in **Figure 2.8**. As illustrated, conventional CID and HCD returned large numbers of total PSMs (**Figure 2.8a**) and unique PSMs (**Figure 2.8b**), and there is no selectivity for cysteine-containing peptides (as expected for any type of non-targeted search). UVPDnLossCID yielded a much smaller number of total PSMs and unique PSMs, an outcome arising from the significant selectivity of the UVPDnLossCID method for targeting cysteine-containing peptides. Additionally, all seven of the cysteine-containing proteins were identified from a unique peptide surrogate in at least three out of

four of the experimental replicates using the UVPDnLossCID method (the replicate LC traces are depicted in **Figure 2.9**). A filtered database search in a regular proteomics experiment can be thought of as an "incomplete database" when used to search a tryptic digest, a factor known to inhibit peptide identification by causing a higher false discovery rate [84]. When the filtered database search is restricted to only MS/MS events that exclusively sequence peptides contained in a database, then this problem is resolved and the chance of a false positive is decreased.

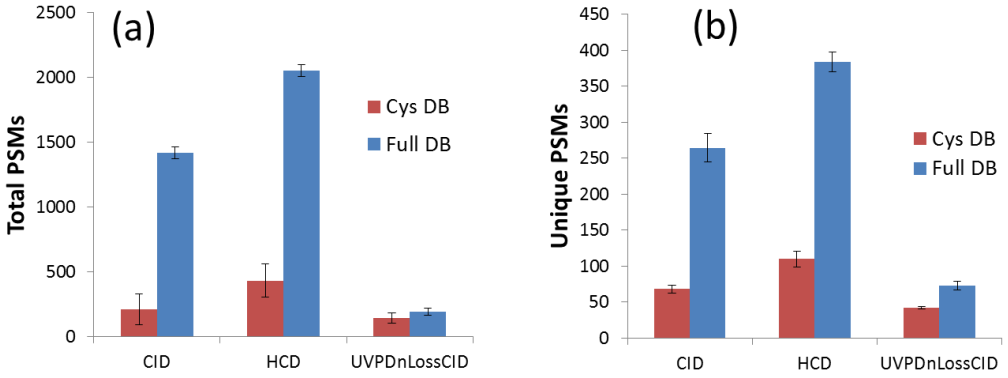


Figure 2.8. Bar charts depicting the (a) total PSMs and (b) unique PSMs from a full database search versus a cysteine-only database search based on spectra obtained from the different activation methods.

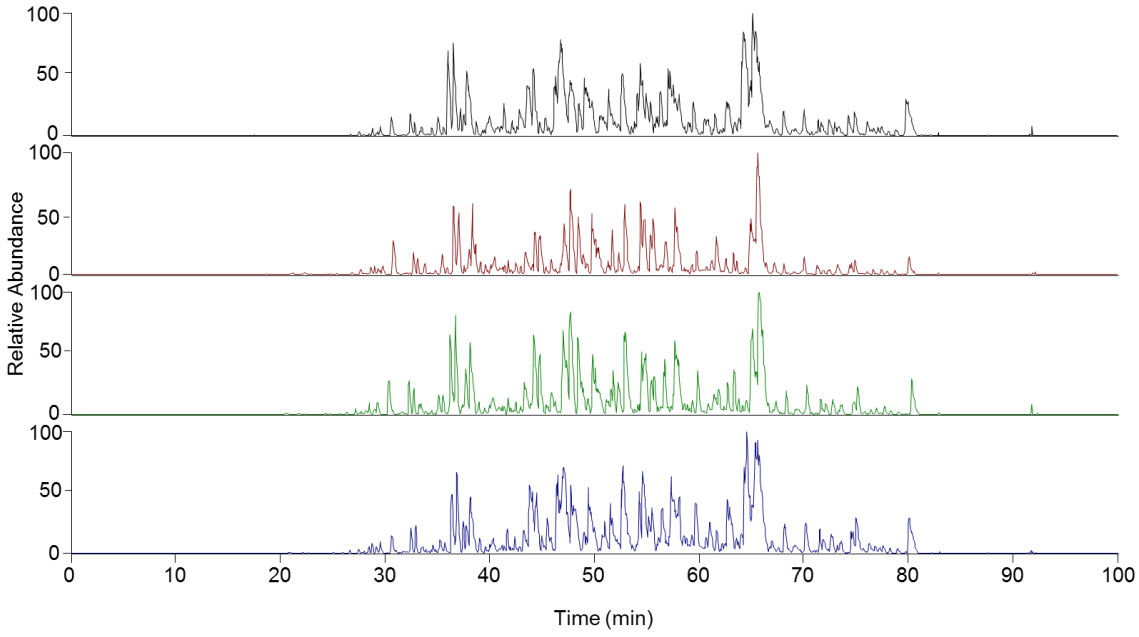


Figure 2.9. Base peak chromatograms of four NPSP-derivatized tryptic digests of protein mixtures on a C_{18} column.

Some cysteine-containing peptides were not identified by the UVPDnLossCID method, and this is attributed to a threshold parameter included in the triggering of the CID step for those peptides identified by UVPD to contain cysteine based on the neutral loss of the SePh tag. This threshold was set so that the neutral loss product must be at least 10% of the precursor ion abundance, thus preventing excessive false triggering of the subsequent CID scan. The UVPDnLossCID method also resulted in a few false positives, an outcome related to the rather broad and complex isotope pattern associated with selenium-containing peptide precursors compared to the more narrow isotope pattern after loss of the SePh tag. This change in isotope pattern, although an interesting signature, requires use of a larger than normal mass tolerance for the neutral loss transition to compensate for potential

activation of the non-SePh-tagged peptide precursor monoisotopic (where monoisotopic signifies the species composed of the most abundant isotopes of each constituent element).

By using a custom version of the MassMatrix database search algorithm which allowed searching only cysteine-containing peptides, two significant improvements are gained due to a reduced search space. First, the confidence level of PSMs is increased when a smaller, targeted database is employed to search data. For example, when searching the *E. coli* proteome without any database filtering, approximately 885,190 peptide sequences (counting variable modifications) were included in the search. After cysteine-selective peptide filtering, this number was reduced to 160,273 (82% reduction). The MassMatrix scoring algorithm evaluates peptides on the basis of a pp score which is equivalent to $-\log(\text{p value})$. The p-value is a measurement of the probability that a peptide match is random. With a significant search space reduction, the p-value and pp scores do not change. The confidence value of a match is defined as: $-\log(\text{q value})$, for which the q-value is defined as the product of the search space and the p-value. In essence, the q-value represents the number of random matches that are expected in a given search space. Thus, while p-score metrics remain constant, the confidence scores increase due to database filtering and yield smaller (better) q values. To illustrate this point, the relative frequencies of each of the 20 amino acids found in all known proteins in the *E. coli* proteome are depicted in **Figure 2.10a**. As expected, cysteine was found to be the least abundant (least frequent) amino acid, meaning that it offers the potential for the greatest gain in PSM confidence in restricted database searches for bottom-up proteomics applications. **Figure 2.10b** graphically depicts an example of the logarithmic gains in

confidence achieved relative to the reduction in the size of the search space. Peptide matches are considered significant if a peptide has a p-value of < 0.05. Because q-value is the product of the p-value and the size of the search space, **Figure 2.10b** shows the minimum expected gains in confidence. As p-value decreases, confidence increases, yielding a minimum expected gain in confidence of 15% for the *E. coli* analysis utilizing a cysteine-restricted database.

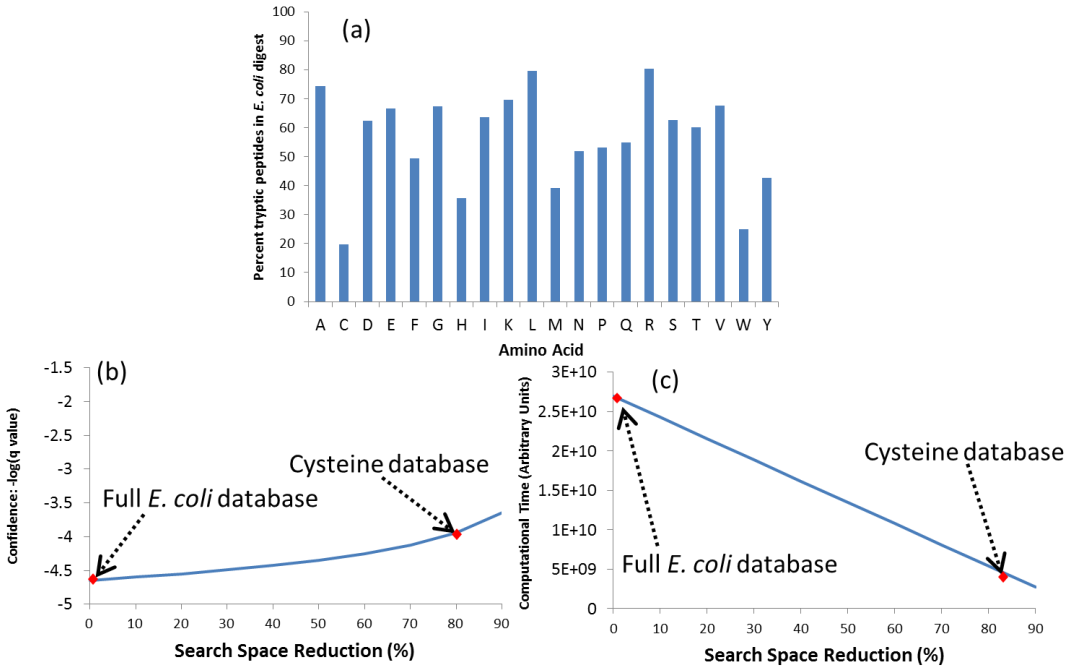


Figure 2.10. Statistics representing (a) the percent of tryptic peptides from the *E. coli* proteome containing each amino acid. Graphical depictions of the effects of search space reduction on (b) PSM confidence and (C) computational time.

In addition to increasing confidence levels, database filtering also expedites searches because there are fewer computations to complete the database search. Big Omnicron ("Big-O") notation is a theoretical metric of an algorithm that measures resources used (time, memory, etc.) for a given problem size [85]. For the MassMatrix

algorithm used in the present study, the time dependence of the search algorithm is related to both: (i) the number of theoretical peptides, and (ii) the number of MS/MS spectra in a particular search. For (i), the number of theoretical peptides is calculated with consideration of peptide modifications. If a peptide sequence has potential modifications (such as the SePh tag in this study), it will generate many peptides with the same sequence but with different modifications. The overall time relationship is defined as: $O(M*N)$, where M is the number of theoretical peptides and N is the number of MS/MS spectra. The linear decrease in search time for LC-MS/MS experiments for a given number of MS/MS spectra is depicted in **Figure 2.10c**. For UVPDnLossCID, there are significantly fewer MS/MS scans accumulated in the search file compared to standard top N HCD or CID runs. The decrease in both the number of MS/MS scans and also the number of theoretical peptides results in over an order of magnitude decrease in search time. In addition, the number of SePh neutral losses observed upon UVPD could conceivably be incorporated into an even more refined search method. This additional metric would count the number of cysteine residues per peptide, offering another means to constrain database searches or build more selectivity into a targeted strategy.

A few remarks about the duty cycle of the UVPDnLossCID method relative to other MS/MS methods are warranted. CID, HCD, and 266 nm UVPD have activation periods of 10, 1, and 2 ms, respectively. CID and HCD have no intrinsic peptide selectivity, and thus a data-dependent acquisition mode that activates the most abundant precursors in a spectrum (top N) is typically used. In contrast, UVPDnLossCID is a two-step method with UVPD used as the prescreening step followed by CID for sequencing the cysteine-containing peptides. Combining two sequential steps significantly increases selectivity at

the expense of the duty cycle of the method. It is this loss of duty cycle which presumably accounts for why some cysteine peptides were not found using the UVPDnLossCID method.

CONCLUSIONS 2.5

We have showcased the ability to implement cysteine-based selectivity using 266 nm UVPD in bottom-up proteomics experiments. NPSP was used to install a photocleavable tag on cysteine residues, resulting in the loss of SePh upon 266 nm UVPD. Development of a customized data-dependent acquisition afforded selective sequencing of cysteine peptides to complement the database search. In essence, cysteine-containing peptides are selectively identified based on a UVPDnLossCID mode in which the loss of SePh upon 266 nm UVPD is used to trigger CID for characterization of those cysteine-containing peptides. Implementation of a filtered cysteine database into a database search program allows the highest gains in speed and PSM confidence for analysis of a proteomic data-set.

Table 2.3. All identified peptides for the protein mixture (**Figure 2.8b**) from *cysteine filtered* CID MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Min)	Charge States
AACVWNGLSYGRYNDLGHRK seph(3)	17.82	3+
ALATQTSVVLKPGEVAFCAEK seph(18)	21.91	3+
ALATQTSVVLKPGEVAFCAEKDDAACCK seph(26)	7.68	3+
ALATQTSVVLKPGEVAFCAEKDDAACCK seph(27)	7.68	3+
CAQCHTVEK seph(1) seph(4)	51.71	1+ 2+
CAQCHTVEK seph(4)	37.53	2+
CAQCHTVEKGGK seph(1)	32.82	3+
CASIQKFGER seph(1)	44.75	2+
CCAADDKEACFAVEGPK seph(1)	54.31	3+
CCAADDKEACFAVEGPK seph(10)	54.31	2+ 3+
CCAADDKEACFAVEGPK seph(2)	54.31	2+ 3+
CCTESLVNR seph(1) seph(2)	66.35	2+
CELAAAMK seph(1)	53.99	1+ 2+
CELAAAMK seph(1) OxiM(7)	45.88	2+
CELAAAMKR seph(1)	49.61	2+
CELAAAMKR seph(1) OxiM(7)	41.9	2+
CEVFR seph(1)	52.5	1+
CEVFRELKDLK seph(1)	58.33	3+
CGRIWAQVDTGKCR seph(1)	26.12	3+
CKGTDVQAWIR seph(1)	54.73	2+ 3+
CKGTDVQAWIRGCRL seph(1) seph(13)	73.55	3+
CKPVNTFVHESLADVQAVCSQK seph(1) seph(19)	80.16	3+
CKPVNTFVHESLADVQAVCSQK seph(19)	91.05	3+
CLEDGFLTHLSKECGAHSEDAVCTK seph(1)	78.46	3+
CRDKTLNRSK seph(1)	58.22	3+
CRNRWAVWANMGNGDSVIK seph(1)	75.42	3+
DAIPENLPPLTADFAEDKDVCK seph(21)	67.23	2+ 3+
DAIPENLPPLTADFAEDKDVCKNYQEAK seph(21)	62.94	3+
DCCEKLKSSITDQNDCIYK seph(2) seph(3) seph(16)	79.08	3+
DDPHACYSTVFDK seph(6)	56.56	2+ 3+
DDPHACYSTVFDKLK seph(6)	15.95	3+
DDQNPHSSNICNISCDK seph(11) seph(15)	69.75	2+ 3+
EACFAVEGPK seph(3)	57.47	2+

Table 2.3 (continued)

ECCHGDLLECADDR seph(10)	53.12	3+
ECCHGDLLECADDR seph(2)	54.57	3+
ECCHGDLLECADDR seph(2) seph(3) seph(10)	76.06	3+
ECCHGDLLECADDR seph(3)	54.57	3+
ESPCVK seph(4)	48.42	2+
EYEATLEECCAK seph(9) seph(10)	54.45	2+ 3+
EYEATLEECCAKDDPHACYSTVFDK seph(10)	55.29	3+
EYEATLEECCAKDDPHACYSTVFDK seph(18)	66.42	3+
EYEATLEECCAKDDPHACYSTVFDK seph(9)	55.29	3+
FLDDDLTDDIMCVK OxiM(11) seph(12)	67.09	2+
FLDDDLTDDIMCVK seph(12)	70.41	2+
GACLLPK seph(3)	56.98	2+
GTDVQAWIRGCRL seph(11)	71.62	3+
GYSLGNWVCAAK seph(9)	70.07	1+ 2+
GYSLGNWVCAAKFESNFNTQATNR seph(9)	90.76	3+
HGLDNYRGYSLGNWVCAAK seph(16)	59.36	3+
HIIVACEGNPYVPVHFDASV seph(6)	9.04	2+ 3+
HLVDEPQNLIKQNCDQFEK seph(14)	58.63	3+
IFHESVYGQCK seph(10)	48.71	3+
IFVQKCAQCHTVEK seph(6)	18.18	3+
IQKETDPLTCIDAHFTFLK seph(10)	88.81	3+
IWCKDDQNPHSSNICNISCDK seph(3)	46.99	3+
KACNVSATIDSSLLASCPINCLNR seph(3) seph(17) seph(21)	80.45	3+
KVFGRCELAAAMKR seph(6)	50.61	3+
KVFGRCELAAAMKR seph(6) OxiM(12)	45	3+
LCVLHEK seph(2)	78.57	2+
LCVLHEKTPVSEKVTK seph(2)	43.42	3+
LDQWLCEK seph(6)	67.02	1+ 2+
LFTFHADICTLPDTEK seph(9)	71.27	2+ 3+
LKECCDKPLLEK seph(4) seph(5)	60.92	2+ 3+
LKPDPNTLCDEFK seph(9)	64.54	2+ 3+
LKPDPNTLCDEFKADEK seph(9)	59.82	2+ 3+
LRCASIQKFGER seph(3)	50.1	3+
LRCASIQKFGERALK seph(3)	48.38	3+
LSFNPTQLEEQCHI seph(12)	76.26	1+ 2+
MMQNCYNSSSAASTSSDMHQREFK OxiM(1) OxiM(2) seph(5)	7.94	3+
MPCTEDYLSLILNR OxiM(1) seph(3)	72.5	2+ 3+

Table 2.3 (continued)

MPCTEDYLSLILNR seph(3)	77.68	2+ 3+
MPCTEDYLSLILNRLCVLHEK seph(16)	88.43	3+
NECFLSHK seph(3)	48.97	2+
NECFLSHKDDSPDLPK seph(3)	50.77	2+ 3+
NGQTNCYQSYSTMSITDCR OxiM(13) seph(18)	88.81	3+
NLCNIPCSALLSSDITASVNCAK seph(21)	72.69	2+ 3+
NLCNIPCSALLSSDITASVNCAK seph(3) seph(7) seph(21)	91.49	2+
NRCKGTDVQAWIR seph(3)	49.91	3+
QEPERNECFLSHK seph(8)	46.13	3+
QEPERNECFLSHKDDSPDLPK seph(8)	46.27	3+
QHMDSSTSAASSSNYCNQMMKSR seph(16) OxiM(19)	36.1	3+
QHMDSSTSAASSSNYCNQMMKSR seph(16) OxiM(20)	36.1	3+
QNCDQFEK seph(3)	45.38	2+
QNCDQFEKLGEYGFQNALIVR seph(3)	67.68	3+
RHGLDNYRGYSLGNWVCAAK seph(17)	56.69	3+
RPCFSALTPDETYVPK seph(3)	38.49	2+ 3+
SAECPGPAQK seph(4)	27.18	2+
SAECPGPAQKGYPFILPS seph(4)	16.58	3+
SHCIAEVEK seph(3)	42.33	2+
SLHTLFGDELCK seph(11)	67.23	1+ 2+ 3+
SLHTLFGDELCKVASLR seph(11)	71.51	3+
TCVADESHAGCEK seph(2) seph(11)	49.13	2+
TCVADESHAGCEKSLHTLFGDELCK seph(24)	87.71	3+
TPVSEKVTKCCTESLVNR seph(10)	21.02	3+
TTQANKHIIVACEGNPYVPVHFDASV seph(12)	63.04	2+ 3+
VASLRETYGDMADCCEKQEPER OxiM(11) seph(15)	91.97	3+
VFGRCELAAAMKR seph(5)	40.87	3+
VFGRCELAAAMKR seph(5) OxiM(11)	31.62	3+
VHKECCHGDLLECADDR seph(13)	50.74	3+
VHKECCHGDLLECADDR seph(5)	46.23	3+
VHKECCHGDLLECADDR seph(5) seph(6) seph(13)	65.3	3+
VHKECCHGDLLECADDR seph(6)	46.23	3+
VLCQCVLSQEPEASNEMCFLLYK seph(3)	7.42	3+
VLCQCVLSQEPEASNEMCFLLYK seph(5)	7.42	3+
VTKCCTESLVNR seph(4)	38.63	3+
VYVEELKPTPEGDLEILLQKWENGECAQK seph(26)	65.45	3+
WENDECAQK seph(6)	50.17	2+ 3+
WENGECAQKK seph(6)	19.92	3+
WWCNDGR seph(3)	61.81	2+

Table 2.3 (continued)

WWCNDGRTPGSR seph(3)	56.15	3+
YICDNQDTISSK seph(3)	52.48	1+ 2+
YICDNQDTISSKLK seph(3)	92.47	2+
YNGVFQECCQAEDK seph(8) seph(9)	75.38	2+
YPNCAYK seph(4)	48.84	2+

Table 2.4. All identified peptides for the protein mixture (**Figure 2.8b**) from *full* CID MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Min)	Charge States
ADEKKFWGK	21.34	2+
AEFVEVTK	31.87	1+ 2+
AEFVEVTKLVTDLTK	63.79	3+ 2+
AIFYINK	41.06	1+ 2+
ALELFR	45.17	1+
ALELFRNDIAAK	47.1	3+ 2+
ALELFRNDIAAKYK	49.65	3+
ALKALPMHIR	99.35	2+
ALKAWSVAR	33.24	2+
ALPMHIR	49.84	2+
ALPMHIR OxiM(4)	89.73	2+
ASEDLKKHGTVVLTALGGILK	54.18	3+
ATEEQLK	15.83	2+ 1+
ATEEQLKTVMENFVAFVDK	72.97	3+
AVLKDGPLTGTYR	36.13	3+
AVVQDPALK	29.5	2+
AVVQDPALKPLALVYGEATSR	59.02	3+ 2+
AVVQDPALKPLALVYGEATSRR	71.95	3+
AWSVARLSQKFPK	48.54	3+
CAQCHTVEK	15.86	2+
CAQCHTVEK seph(1)	51.05	2+
CAQCHTVEK seph(1) seph(4)	49.78	2+ 1+
CAQCHTVEK seph(4)	37.53	2+
CAQCHTVEKGGK	45.12	3+
CAQCHTVEKGGKHK	9.44	3+
CASIQK	85.04	1+
CASIQKFGER seph(1)	44.75	3+ 2+
CASIQKFGERALK	39.16	3+
CCAADDKEACFAVEGPK seph(1)	86.16	3+
CCAADDKEACFAVEGPK seph(10)	49.41	3+
CCAADDKEACFAVEGPK seph(2)	49.97	3+
CCTESLVNR	29.66	2+
CCTESLVNR seph(1) seph(2)	64.84	2+

Table 2.4 (continued)

CCTESLVNRR seph(1)	29.18	3+
CCTESLVNRR seph(1) seph(2)	45.8	3+
CCTESLVNRR seph(2)	21.67	3+
CCTKPESER	13.81	2+
CCTKPESER seph(1)	41.02	3+
CCTKPESER seph(1) seph(2)	51.33	2+
CCTKPESER seph(2)	41.02	3+
CELAAAMK	36.77	2+
CELAAAMK OxiM(7)	36.9	2+
CELAAAMK seph(1)	54.62	2+ 1+
CELAAAMK seph(1) OxiM(7)	24.86	2+
CELAAAMKR OxiM(7)	2.59	2+
CELAAAMKR seph(1)	48.32	2+
CELAAAMKR seph(1) OxiM(7)	40.32	2+
CEVFRELKDLK seph(1)	54.25	3+
CKGTDVQAWIR	31.71	3+
CKGTDVQAWIR seph(1)	52.8	3+ 2+
CKPVNTFVHESLADVQAVCSQK seph(1) seph(19)	78.23	3+
CRDKTLNR	94.43	2+
DAFLGSFLYEYSR	64.81	2+ 1+
DAFLGSFLYEYSRR	78.23	2+
DAIPENLPPLTADFAEDK	56.94	3+ 2+
DAIPENLPPLTADFAEDKDVCK seph(21)	68.15	3+ 2+
DAIPENLPPLTADFAEDKDVCKNYQEAK seph(21)	62.94	3+
DDPHACYSTVFDK	36.67	3+
DDPHACYSTVFDK seph(6)	56.46	3+ 2+
DDPHACYSTVFDKLK seph(6)	11.89	3+
DDQNPHSSNICNISCDK	24.52	3+ 2+
DDQNPHSSNICNISCDK seph(11) seph(15)	68.11	2+ 3+
DDSPDLPK	26.17	2+
DFPIANGER	34.21	2+ 1+
DFPIANGERQSPVDIDTK	41.84	3+
DGPLTGTYR	31.4	2+ 1+
DGYVLSLNR	41.81	2+
DGYVLSLNRVSDAR	53.8	3+
DLGEEHFK	26.63	2+
DRKDGYVLSLNR	9.28	3+ 2+
DVCKNYQEAK	37.62	3+
DVCKNYQEAK seph(3)	49.32	3+

Table 2.4 (continued)

EACFAVEGPK seph(3)	56.21	2+
ECCDKPLLEK	32.05	2+
ECCDKPLLEK seph(2) seph(3)	62.82	3+
ECCHGDLLECADDR	38.8	3+
ECCHGDLLECADDR seph(10)	57.87	3+
ECCHGDLLECADDR seph(2)	26.37	3+
ECCHGDLLECADDR seph(2) seph(3) seph(10)	76.06	3+
ECCHGDLLECADDR seph(3)	53.23	3+
ECCHGDLLECADDR seph(3) seph(10)	2.32	3+
EDLIAYLK	50.87	2+ 1+
EDLIAYLKK	45.14	2+
EETLMEYLENPK	57.51	2+ 3+ 1+
EETLMEYLENPK OxiM(5)	54.83	2+
EETLMEYLENPKK	53.76	2+ 3+
EETLMEYLENPKK OxiM(5)	43.31	3+ 2+
EETLMEYLENPKKYIPGTK OxiM(5)	15.46	3+
EKVLASSARQR	60.47	3+
ELGFQG	36.42	1+
EPISVSSQQMLK	38.64	2+ 1+
EPISVSSQQMLK OxiM(10)	15.11	3+ 2+
EPISVSSQQMLKFR OxiM(10)	35.78	3+
EQLTKCEVFR	33.4	3+
EQLTKCEVFRELK	30.43	3+
EQRSAECPGPAQK	93.03	2+
ESPCVK seph(4)	45.58	2+
ETGSSKYPNCAYK	32.84	3+
ETYGDMADCCEK seph(9)	43.72	3+
ETYGDMADCCEKQEPER	34.55	3+
ETYGDMADCCEKQEPER OxiM(6)	25.79	3+
EYEATLEECCAK	37.89	2+
EYEATLEECCAK seph(9) seph(10)	54.79	3+
EYEATLEECCAKDDPHACYSTVFDK seph(10)	55.08	3+
EYEATLEECCAKDDPHACYSTVFDK seph(18)	56.69	3+
EYEATLEECCAKDDPHACYSTVFDK seph(9)	55.29	3+
FDKFKHLK	65.44	2+
FERQHMDSSTSAASSSNYCNQMMK OxiM(22) OxiM(23)	30.52	3+
FESNFNTQATNR	30.42	2+ 1+ 3+
FESNFNTQATNRNTDGSTDYGILQINSR	90.34	3+
FGERALK	48.38	2+

Table 2.4 (continued)

FGERALKAWSVAR	42.29	3+
FKDLGEEHFK	30.68	3+
FKHLKTEAEMK OxiM(10)	14.64	3+
FLDDDLTDDIMCVK	47.15	2+ 3+
FLDDDLTDDIMCVK OxiM(11)	61.76	2+
FLDDDLTDDIMCVK OxiM(11) seph(12)	11.29	2+ 3+
FLDDDLTDDIMCVK seph(12)	70.41	2+
FLDDDLTDDIMCVKK	89.15	3+
FMETATESLAK	28.1	2+ 3+
FMETATESLAK OxiM(2)	32.97	2+
GACLLPK seph(3)	43	2+
GACLLPKIETMR	11.2	2+
GACLLPKIETMR OxiM(11)	44.29	3+
GACLLPKIETMREK	19.83	2+
GACLLPKIETMREK OxiM(11)	15.23	3+
GACLLPKIETMREK seph(3)	84.83	3+
GDVEKGKK	26.67	2+
GENATVNQRPANPSK	44.64	3+
GHHEAELK	20.6	2+
GHHEAELKPLAQSHATK	20.53	3+
GITWKEETLMEYLENPK	7.54	3+ 2+
GITWKEETLMEYLENPKK	55.78	3+
GLDIQK	24.73	1+
GLSDGEWQQVLNVWGK	65.27	2+ 3+
GLSDGEWQQVLNVWGKVEADIAGHGQEVLIR	78.79	3+
GLVLIAFSQYLQQCPFDEHVK	70.03	3+
GSVQYLPDWDK	47.6	2+
GSVQYLPDWDKKR	46.21	3+
GTDVQAWIR	42.46	2+ 1+
GTDVQAWIRGCR seph(11)	13.32	2+
GTDVQAWIRGCRL	47.14	3+
GTDVQAWIRGCRL seph(11)	70.23	2+ 3+
GYSLGNWVCAAK	28.01	2+ 3+
GYSLGNWVCAAK seph(9)	66.44	2+ 1+
GYSLGNWVCAAKFESNFNTQATNR	90.52	3+
GYSLGNWVCAAKFESNFNTQATNR seph(9)	90.71	3+
HALWYNIGVKDLIKK	87.55	2+
HGLDNYR	19.38	2+
HGLDNYRGYSLGNWVCAAK	48.35	3+

Table 2.4 (continued)

HGLDNYRGYSLGNWVCAAK seph(16)	59.36	3+
HGTVVLTALGGILK	57.16	2+ 3+
HGTVVLTALGGILKK	89.95	3+
HGTVVLTALGGILKKK	68.36	3+
HIIVACEGNPYVPVHFDASV	35.14	3+
HIIVACEGNPYVPVHFDASV seph(6)	67.15	3+ 2+
HKTGPNLHGLFGR	36.19	3+
HKTGPNLHGLFGRK	20.74	2+ 3+
HLVDEPQNLIK	38.74	3+ 1+ 2+
HLVDEPQNLIKQNCDQFEK seph(14)	58.63	3+
HPEYAVSVLLR	46.86	1+ 3+ 2+
HPGDFGADAQGAMTK	30.14	3+ 2+
HPGDFGADAQGAMTK OxiM(13)	24.75	3+ 2+
HPYFYAPELLYYANK	57.5	3+ 2+
IDALNENK	23	2+ 1+
IDALNENKVLVLDTDYK	48.48	3+
IDALNENKVLVLDTDYKK	44.77	3+
IETMREKVLASSAR	64.89	3+
IFHESVYGQCK	95.21	3+
IFHESVYGQCK seph(10)	47.51	3+
IFVQKCAQCHTVEK	29.52	3+
IFVQKCAQCHTVEK seph(6)	11.76	3+
IFVQKCAQCHTVEK seph(9)	77.88	3+
IFYLPAYNCTLR	13.88	3+
ILDKVGINYWLAHK	88.35	3+ 2+
IPAVFK	37.35	1+
IPAVFKIDALNENK	47.61	3+
IVSDGNGMNAWVAWR	55.54	2+
IVSDGNGMNAWVAWR OxiM(8)	45.86	2+
IWCKDDQNPHSSNICNISCDK seph(3)	46.99	3+
KDGYVLSLNR	95.62	2+
KDGYVLSLNRVSDAR	10.82	3+
KFWGKYLYEIAR	49.06	3+
KGHHEAELKPLAQSHATK	11.6	3+
KHGTVVLTALGGILK	92.81	3+
KILDKVGINYWLAHK	5.21	3+
KIVSDGNGMNAWVAWR	51.05	3+ 2+
KIVSDGNGMNAWVAWR OxiM(9)	41.4	3+
KIVSDGNGMNAWVAWRNR	45.86	3+

Table 2.4 (continued)

KKGHHEAELK	0.16	2+
KQTALVELLK	45.27	2+ 1+
KTEREDLIAYLK	12.98	3+
KTGQAPGFTYTDANK	28.99	3+ 2+
KTGQAPGFTYTDANKNK	25.72	3+
KVFGRCELAAAMK	38.83	3+
KVFGRCELAAAMK seph(6)	54.67	3+
KVFGRCELAAAMK seph(6) OxiM(12)	48.47	3+
KVPQVSTPTLVEVSR	43.73	3+ 2+
KYAAELHLVHWNTK	39.22	3+
KYIPGTK	19.75	2+
KYLLFCMENSAEPEQSLACQCLVR seph(19) seph(21)	7.88	3+
LCVLHEK	47.68	2+
LCVLHEK seph(2)	77.12	2+
LCVLHEKTPVSEK seph(2)	48.7	3+
LCVLHEKTPVSEKVTK	5.38	3+
LCVLHEKTPVSEKVTK seph(2)	41.28	3+
LDQWLCEK	22.7	2+
LDQWLCEK seph(6)	59.96	2+ 1+
LFTFHADICTLPDTEK seph(9)	68.75	2+ 3+
LFTGHPETLEK	33.13	2+ 3+
LFTGHPETLEKFDK	10.27	3+
LFTGHPETLEKFDKFK	44.24	3+
LGEYGFQNALIVR	53.54	2+ 1+
LIVTQTMK	32.4	2+ 1+
LIVTQTMK OxiM(7)	25.72	2+
LIVTQTMKGLDIQK OxiM(7)	14.26	2+
LKECCDK	36.8	2+
LKECCDK seph(5)	26.65	2+
LKECCDKPLLEK	31.8	3+ 2+
LKECCDKPLLEK seph(4) seph(5)	58.78	3+ 2+
LKHLVDEPQNLIK	39.3	3+
LKPDPNTLCDEFK	63.92	3+
LKPDPNTLCDEFK seph(9)	62.11	3+ 2+
LKPDPNTLCDEFKADEK	51.17	3+
LKPDPNTLCDEFKADEK seph(9)	58.29	3+ 2+
LRCASIQKFGER seph(3)	92.91	3+ 2+
LSFNPTQLEEQCHI seph(12)	73.99	2+ 1+
LSQKFPK	51.54	2+

Table 2.4 (continued)

LSQKFPKAEFVEVTK	41.33	3+ 2+
LVNELTEFAK	46.64	2+ 1+
LVQFHFHWGSSDDQGSEHTVDR	44.07	3+
LVTDLTK	30.86	1+
LVVSTQTALA	41.16	2+ 1+
MIFAGIK	43.25	1+
MIFAGIKK OxiM(1)	14	2+
MMQNCYNSSSAASTSSDMHQR OxiM(2) OxiM(18)	9.32	3+
MPCTEDYLSLILNR	68.67	3+ 2+
MPCTEDYLSLILNR OxiM(1)	66.68	3+ 2+
MPCTEDYLSLILNR OxiM(1) seph(3)	72.5	2+ 3+
MPCTEDYLSLILNR seph(3)	76.44	2+ 3+
MVNNGHSFNVEYDDSQDK	33.68	3+ 2+
MVNNGHSFNVEYDDSQDK OxiM(1)	31.52	3+ 2+
MVNNGHSFNVEYDDSQDKAVLK	38.27	3+
MVNNGHSFNVEYDDSQDKAVLK OxiM(1)	38.39	3+
MVNNGHSFNVEYDDSQDKAVLKDGPLTGTYR	45.08	3+
NAQTTKYACNPYK	14.19	3+
NDIAAKYKELGFQG	15.66	2+
NECFLSHK seph(3)	48.97	2+
NECFLSHKDDSPDLPK seph(3)	49	2+ 3+
NECFLSHKDDSPDLPKLK	93.25	3+
NGQTNCYQSYSTMSITDCR OxiM(13) seph(18)	10.65	3+
NKGITWK	51.88	2+
NLCNIPCSALLSSDITASVNCAK seph(21)	90.87	3+
NRCKGTDVQAWIR	34.62	2+ 3+
NRCKGTDVQAWIR seph(3)	49.91	3+
NTDGSTDYGILQINSR	46.84	3+ 2+ 1+
NYQEAKDAFLGSFLYEYSR	63.17	3+
PAQPLKNRQVR	36.56	2+ 3+
PCFSALTPDETYVPK	9.06	2+
PCFSALTPDETYVPK seph(2)	65.51	3+ 2+
PDPNTLCDEFK	42.98	2+ 3+
PDPNTLCDEFK seph(7)	51.26	2+ 3+
PLALVYGEATSR	42.22	2+
PLALVYGEATSRR	13.21	2+ 3+
PLAQSHATK	23.47	2+
PTPEGDLEILLQK	60.75	3+ 2+
PTPEGDLEILLQKWENGECAQKK	6.92	3+

Table 2.4 (continued)

PVNTFVHESLADVQAVCSQK seph(17)	73.83	3+
PVSQSAIIMTCPDCPSTSPYDLSNPR	55.15	3+
PVSQSAIIMTCPDCPSTSPYDLSNPR OxiM(9)	50.79	3+
PVSQSAIIMTCPDCPSTSPYDLSNPR seph(11) seph(14)	7.05	3+
QEPERNECFLSHK seph(8)	44.8	3+
QEPERNECFLSHKDDSPDLPK seph(8)	46.27	3+
QHMDSSTSAASSSNYCNQMMK	36.74	3+
QHMDSSTSAASSSNYCNQMMK OxiM(19)	29.82	3+
QHMDSSTSAASSSNYCNQMMK OxiM(19) OxiM(20)	24.35	3+
QHMDSSTSAASSSNYCNQMMK OxiM(20)	29.47	3+
QHMDSSTSAASSSNYCNQMMK OxiM(3) OxiM(19)	27.25	3+
QHMDSSTSAASSSNYCNQMMK OxiM(3) OxiM(19) OxiM(20)	22.04	3+
QHMDSSTSAASSSNYCNQMMK OxiM(3) OxiM(20)	29.06	3+
QHMDSSTSAASSSNYCNQMMK seph(16) OxiM(19) OxiM(20)	45.42	3+
QHMDSSTSAASSSNYCNQMMKSR OxiM(20)	88.34	3+
QHMDSSTSAASSSNYCNQMMKSR seph(16) OxiM(20)	36.1	3+
QNCDQFEK	42.94	2+
QNCDQFEK seph(3)	43.82	2+
QNCDQFEKLGEYGFQNALIVR seph(3)	67.68	3+
QSPVDIDTK	29.64	2+ 1+
QSPVDIDTKAVVQDPALK	49.16	3+
QTALVELLK	52.74	1+ 2+
QTALVELLKHK	2.19	2+
QVRGFPK	22.61	2+
RHGLDNYR	40.6	2+
RHPEYAVSVLLR	42.09	3+ 2+
RHPEYAVSVLLRLAK	40.29	3+
RHPYFYAPELLYYANK	54.12	3+ 2+
RMVNNGHSFNVEYDDSQDK	33.82	3+
RMVNNGHSFNVEYDDSQDKAVLK	7.19	3+
RPCFSALTPDETYVPK	1.66	3+
RPCFSALTPDETYVPK seph(3)	58.14	3+ 2+
SAECPGPAQK	14.25	2+
SAECPGPAQK seph(4)	26.24	2+
SAECPGPAQKGYPFILPS seph(4)	8.71	3+
SEIAHRFK	70.69	2+
SHCIAEVEK	21.32	2+

Table 2.4 (continued)

SHCIAEVEK seph(3)	29.09	2+
SLGKVGTRCCTK	12.75	2+
SLGKVGTRCCTK seph(10)	85.78	3+
SLHTLFGDELCK	4.42	2+
SLHTLFGDELCK seph(11)	65	2+ 3+ 1+
SLHTLFGDELCKVASLR seph(11)	71.51	3+
TCVADESHAGCEK	51.39	3+
TCVADESHAGCEK seph(11)	41.75	3+
TCVADESHAGCEK seph(2) seph(11)	55.69	2+ 3+
TCVADESHAGCEKSLHTLFGDELCK seph(24)	9.28	3+
TEELQQQNTAPTNSPTK	64.96	3+
TEREDLIAYLK	47.42	2+ 3+
TEREDLIAYLKK	43.15	3+
TGPNLHGLFGR	44.2	2+ 1+
TGQAPGFTYTDANK	32.05	2+ 1+
TGQAPGFTYTDANKNK	28.77	3+
TKIPAVFK	35.84	2+
TKIPAVFKIDALNENK	47.46	3+
TLNFNAEGEPELLMLANWR	70.78	2+ 3+
TLNFNAEGEPELLMLANWR OxiM(14)	67.12	2+ 3+
TLNFNAEGEPELLMLANWRPAQPLK	67.52	3+
TLNFNAEGEPELLMLANWRPAQPLK OxiM(14)	65.36	3+
TLNFNAEGEPELLMLANWRPAQPLKNR	65.2	3+
TPEVDDEALEK	30.97	2+ 1+
TPEVDDEALEKFDK	42.29	2+ 3+
TPEVDDEALEKFDKALK	48.32	3+
TPVSEKVTK	28.15	2+
TPVSEKVTKCCTESLVNR	64.87	3+
TTQANKHIIVACEGNPYVPVHFDASV seph(12)	64.88	3+ 2+
TVMENFVAFVDK	61.38	2+ 1+
TVMENFVAFVDK OxiM(3)	34.56	2+ 3+
TVMENFVAFVDKCCAADDK OxiM(3)	53.79	3+
TVMENFVAFVDKCCAADDKEACFAVEGPK seph(14)	77.54	3+
VAGTWYSLAMAASDISLLDAQSAPLR	77.24	2+ 3+
VAGTWYSLAMAASDISLLDAQSAPLR OxiM(10)	73.4	2+ 3+
VCMIDDTLDDDLFK seph(2)	62.21	3+
VEADIAGHGQEVLIR	38.68	3+ 1+ 2+
VFGRCELAAAMK	48.01	3+
VFGRCELAAAMK OxiM(11)	20.22	3+

Table 2.4 (continued)

VFGRCELAAAMK seph(5) OxiM(11)	25.1	3+
VFGRCELAAAMKR	37.63	3+
VFGRCELAAAMKR seph(5)	62.39	3+
VGDANPALQK	21.96	1+ 2+
VGDANPALQKVLDALDSIK	69.71	2+ 3+
VGDANPALQKVLDALDSIKTK	64.41	3+
VGINYWLAHK	45.14	2+ 3+ 1+
VGINYWLAHKALCSEK	87.96	2+
VGTRCCTK	26.4	2+
VGTRCCTK seph(6)	29.55	2+
VGTRCCTKPESER seph(6)	48.57	3+
VHKECCHGDLLECADDR	30.92	3+
VHKECCHGDLLECADDR seph(13)	50.74	3+
VHKECCHGDLLECADDR seph(5)	46.23	3+
VHKECCHGDLLECADDR seph(5) seph(6) seph(13)	65.3	3+
VLASSARQRLR	86.39	2+
VLDALDSIK	42.91	2+ 1+
VLDALDSIKTK	39.24	2+ 3+
VLDALDSIKTKGK	35.36	3+
VLVLDTDYK	41.16	2+ 1+
VLVLDTDYKK	35.84	2+
VPQVSTPTLVEVSR	44.8	2+
VTKCCTESLVNR	23.26	3+
VTKCCTESLVNR seph(4)	42.21	3+
VTKCCTESLVNR seph(4) seph(5)	26.31	3+
VTKCCTESLVNRR seph(4)	47.13	3+
VVVLPFPSK	48.79	2+
VYVEELK	32.62	2+
VYVEELKPTPEGDLEILLQK	72.5	3+ 2+
WENDECAQK	26.55	2+
WENDECAQK seph(6)	44.63	2+
WENDECAQKK	44.27	3+
WENGECAQKK seph(6)	18.6	3+
WWCNDGR seph(3)	61.81	2+
WWCNDGRTPGSR seph(3)	56.15	3+
YAAELHLVHWNTK	42.62	2+ 3+
YGDFGTAAQQPDGLAVVGVFLK	67.43	2+ 3+
YGDFGTAAQQPDGLAVVGVFLKVGDANPALQK	70.5	3+
YICDNQDTISSK	28.08	3+ 2+

Table 2.4 (continued)

YICDNQDTISSK seph(3)	50.47	2+ 1+
YICDNQDTISSKLK	14.48	3+
YICDNQDTISSKLK seph(3)	92.47	2+
YIPGTK	20.41	1+
YIPGTKMIFAGIKK OxiM(7)	28.19	3+
YKELGFQG	35.65	2+
YLEFISDAIIHVLHSK	72.01	2+ 3+
YLLFCMENSAEPEQSLVCQCLVR seph(5) seph(18) seph(20)	91.46	3+
YLYEIAR	39.59	1+ 2+
YLYEIARR	34.88	2+
YNGVFQECCQAEDK	41.16	2+
YNGVFQECCQAEDK seph(8) seph(9)	75.38	2+
YNGVFQECCQAEDKGACLLPK seph(17)	64.12	3+
YPNCAYK seph(4)	46.43	2+
YTRKVPQVSTPTLVEVSR	39.47	3+

Table 2.5. All identified peptides for the protein mixture (**Figure 2.8b**) from *cysteine filtered* HCD MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Minutes)	Charge State
ALCSEK seph(3)	40.23	2+ 1+
ALCSEKLDQWLCEKL seph(3)	52.36	3+
ALPMHIRLSFNPTQLEEQCHI seph(19)	0.13	3+
CAQCHTVEK seph(1)	58.18	2+
CAQCHTVEK seph(1) seph(4)	57.33	2+ 3+
CAQCHTVEK seph(4)	40.6	2+
CAQCHTVEKGGK seph(1)	81.05	3+ 2+
CAQCHTVEKGGK seph(1) seph(4)	38.36	2+
CAQCHTVEKGGK seph(4)	52.69	3+
CAQCHTVEKGGKHK seph(1)	92.37	3+
CAQCHTVEKGGKHK seph(4)	57.05	3+
CASIQK seph(1)	40.49	2+ 1+
CASIQKFGER seph(1)	50	3+ 2+
CASIQKFGERALK seph(1)	79.28	2+
CCAADDK seph(1)	22.37	2+
CCAADDK seph(1) seph(2)	70.32	2+
CCAADDK seph(2)	39.88	2+
CCAADDKEACFAVEGPK seph(1)	55.61	3+ 2+
CCAADDKEACFAVEGPK seph(1) seph(2) seph(10)	78.74	3+
CCAADDKEACFAVEGPK seph(10)	59.36	3+ 2+
CCAADDKEACFAVEGPK seph(2)	45.34	2+ 3+
CCTESLVNR seph(1)	58.6	2+
CCTESLVNR seph(1) seph(2)	78.49	2+
CCTESLVNR seph(2)	9.1	2+
CCTESLVNRR seph(1)	51.32	3+ 2+
CCTESLVNRR seph(2)	83.43	3+
CCTKPESER seph(1)	58.44	2+
CCTKPESER seph(1) seph(2)	76.93	2+
CDTISMTSYSQYCNTQGNK seph(1) OxiM(6) seph(13)	88.07	3+
CELAAAMK seph(1)	60.07	2+ 1+
CELAAAMK seph(1) OxiM(7)	29.69	2+ 1+
CELAAAMKR seph(1)	51.66	2+
CELAAAMKR seph(1) OxiM(7)	28.78	1+ 2+

Table 2.5 (continued)

CEVFRELK seph(1)	47.32	2+
CEVFRELKDLK seph(1)	20.21	2+ 3+
CKGTDVQAWIR seph(1)	59.85	2+ 3+
CKPVNTFVHESLADVQAVCSQK seph(1) seph(19)	82.45	3+
DAIPENLPPLTADFAEDKDVCK seph(21)	72.45	3+ 2+
DDPHACYSTVFDK seph(6)	62.18	3+ 2+
DDPHACYSTVFDKLK seph(6)	55.73	3+
DDQNPHSSNICNISCDK seph(11) seph(15)	75.39	3+ 2+
DDQNPHSSNICNISCDK seph(15)	64.04	2+
DVCKNYQEAK seph(3)	27.83	2+
EACFAVEGPK seph(3)	35.62	2+ 3+
ECCDKPLLEK seph(2)	52.11	2+
ECCDKPLLEK seph(2) seph(3)	70.18	2+ 3+
ECCDKPLLEK seph(3)	44.88	3+ 2+
ECCHGDLLECADDR seph(10)	63.31	3+
ECCHGDLLECADDR seph(2)	60.95	3+ 2+
ECCHGDLLECADDR seph(2) seph(10)	73.96	3+
ECCHGDLLECADDR seph(2) seph(3)	66.36	3+
ECCHGDLLECADDR seph(3)	76.66	3+ 2+
ECCHGDLLECADDR seph(3) seph(10)	31.41	3+
EQLTKCEVFR seph(6)	71.28	3+ 2+
EQLTKCEVFRELK seph(6)	54.94	3+
EQRSAECPGPAQK seph(7)	92.68	3+
ESPCVK seph(4)	50.49	2+
ETYGDMADCCEK OxiM(6) seph(10)	41.91	2+
ETYGDMADCCEK OxiM(6) seph(9) seph(10)	34.1	3+
ETYGDMADCCEK seph(10)	75.99	2+ 3+
EYEATLEECCAK seph(10)	43.75	3+
EYEATLEECCAK seph(9) seph(10)	60.43	2+ 3+
FLDDDLTDDIMCVK OxiM(11) seph(12)	72.28	3+ 2+
FLDDDLTDDIMCVK seph(12)	76.65	2+
GACLLPK seph(3)	62.13	1+ 2+
GACLLPKIETMR seph(3)	78.9	2+
GACLLPKIETMR seph(3) OxiM(11)	20.07	3+
GACLLPKIETMREK seph(3)	62.8	3+
GTDVQAWIRGCR seph(11)	40.73	3+ 2+
GTDVQAWIRGCRL seph(11)	77.11	3+ 2+
GYSLGNWVCAAK seph(9)	72.91	3+ 2+ 1+
HIIVACEGNPYVPVHFDASV seph(6)	73.05	3+ 2+

Table 2.5 (continued)

HLVDEPQNLIKQNCDQFEK seph(14)	65.07	3+
IFHESVYGQCK seph(10)	53.41	3+ 2+
IFVQKCAQCHTVEK seph(6)	64.24	3+ 2+
IFVQKCAQCHTVEK seph(9)	70.58	3+ 2+
IFVQKCAQCHTVEKGGK seph(6) seph(9)	70.61	2+
KQACEGNEWK seph(4)	32.55	2+
KVFGRCELAAAMK seph(6)	58.75	3+
KVFGRCELAAAMK seph(6) OxiM(12)	53.31	3+
LAKEYEATLEECCAK seph(13)	56.11	3+
LCVLHEK seph(2)	76.38	2+
LCVLHEKTPVSEK seph(2)	50.1	2+ 3+
LDQWLCEK seph(6)	66.72	2+ 1+
LFTFHADICTLPDTEK seph(9)	76.24	3+
LKECCDK seph(4)	48.57	2+
LKECCDK seph(4) seph(5)	36.14	2+
LKECCDK seph(5)	40.19	2+
LKECCDKPLLEK seph(4) seph(5)	65.41	3+ 2+
LKECCDKPLLEK seph(5)	68.16	3+
LKPDPNTLCDEFK seph(9)	69.15	3+ 2+
LKPDPNTLCDEFKADEK seph(9)	64.6	3+ 2+
LRCASIQK seph(3)	22.97	2+
LSFNPTQLEEQCHI seph(12)	80.14	2+ 3+ 1+
LTCNYAPLYFIRKEK seph(3)	21.2	3+
MPCTEDYLSLILNR OxiM(1) seph(3)	78.59	3+ 2+
MPCTEDYLSLILNR seph(3)	82.29	3+ 2+
NECFLSHK seph(3)	78.63	2+
NECFLSHKDDSPDLPK seph(3)	55.95	3+
NLCNIPCSALLSSDITASVNCAK seph(21)	78.16	3+ 2+
NLCNIPCSALLSSDITASVNCAK seph(3)	79.83	2+ 3+
NLTKDRCK seph(7)	45.1	2+
NRCKGTDVQAWIR seph(3)	54.98	3+
PCFSALTPDETYVPK seph(2)	7.3	3+ 2+
PDPNTLCDEFK seph(7)	56.7	2+ 3+
PLLEKSHCIAEVEK seph(8)	74.17	3+ 2+
PVNTFVHESLADVQAVCSQK seph(17)	80.91	3+
QACEGNEWK seph(3)	91.33	2+
QAPGPCEASRQEK seph(6)	47.97	3+
QEPERNECFLSHK seph(8)	52.35	3+
QHMDSSTSAASSSNYCNQMMK seph(16)	34.79	3+

Table 2.5 (continued)

QHMDSSTSAASSSNYCNQMMK seph(16) OxiM(19) OxiM(20)	51.92	3+
QNCDQFEK seph(3)	25.67	2+
QNCDQFEKLGEYGFQNALIVR seph(3)	74.17	3+
QRLRCASIQK seph(5)	70.99	3+ 2+
RPCFSALTPDETYVPK seph(3)	64.48	3+ 2+
SAECPGPAQK seph(4)	46.87	2+
SHCIAEVEK seph(3)	46.38	2+
SLGKVGTRCCTK seph(10)	53.31	2+ 3+
SLGKVGTRCCTK seph(9)	42.53	3+
SLGKVGTRCCTK seph(9) seph(10)	48.3	3+
SLHTLFGDELCK seph(11)	72.22	3+ 2+
SLHTLFGDELCKVASLR seph(11)	62.42	3+
TCVADESHAGCEK seph(11)	37.12	3+
TCVADESHAGCEK seph(2) seph(11)	55.81	3+ 2+
TTQANKHIIVACEGNPYVPVHFDASV seph(12)	72.15	3+
TVMENFVAFVDKCCAADDK OxiM(3) seph(14)	76.54	3+
VFGRCELAAAMK seph(5)	39.06	2+ 3+
VFGRCELAAAMK seph(5) OxiM(11)	52.35	3+ 2+
VFGRCELAAAMKR seph(5)	17.82	2+
VFGRCELAAAMKR seph(5) OxiM(11)	8.39	3+ 2+
VGTRCCTK seph(5)	36.84	2+
VGTRCCTK seph(6)	33.59	2+
VGTRCCTKPESER seph(5)	23.2	3+
VHKECCHGDLLECADDR seph(13)	56.1	3+
VTKCCTESLVNR seph(4) seph(5)	52.45	3+
VTKCCTESLVNRR seph(4)	51.63	3+
VYVEELKPTPEGDLEILLQKWENGECAQK seph(26)	71.28	3+
WENDECAQK seph(6)	53.98	2+ 3+
WENGECAQK seph(6)	54.07	2+ 3+
WENGECAQKK seph(6)	24.22	3+
WENGECAQKKIIAEK seph(6)	36.62	3+
WWCNDGR seph(3)	83.27	2+
WWCNDGRTPGSR seph(3)	63.4	3+ 2+
YACNPYK seph(3)	33.52	2+
YACNPYKSSGTER seph(3)	21.01	3+
YICDNQDTISSK seph(3)	57.68	2+
YNGVFQECCQAEDK seph(8)	72.22	3+
YNGVFQECCQAEDK seph(8) seph(9)	82.16	2+

Table 2.5 (continued)

YNGVFQECCQAEDK seph(9)	60.48	3+ 2+
YNGVFQECCQAEDKGACLLPK seph(9) seph(17)	75.71	3+
YPNCAYK seph(4)	71.4	2+ 1+

Table 2.6. All identified peptides for the protein mixture (**Figure 2.8b**) from *full database* HCD MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Min)	Charge
ADEKKFWGK	71.37	2+
AEFVEVTK	37.75	1+ 2+
AEFVEVTKLVTDLTK	70.03	2+ 3+
AEFVEVTKLVTDLTKVHK	65.27	3+
AIFYINK	44.69	1+ 2+
AIFYINKEK	91.48	2+
AIFYINKEKR	36.28	2+
ALCSEK	37.6	1+
ALCSEK seph(3)	53.26	1+ 2+
ALCSEKLDQWLCEKL seph(3)	52.36	3+
ALELFR	51.45	1+
ALELFRNDIAAK	51.36	2+ 3+
ALELFRNDIAAKYK	48.73	2+
ALKALPMHIR	26.19	2+
ALKALPMHIR OxiM(7)	84.8	2+
ALKAWSVAR	52.51	2+
ALKAWSVARLSQK	46.93	2+
ALPMHIR	38.34	2+
ALPMHIR OxiM(4)	83.27	2+
ALPMHIRLSFNPTQLEEQCHI seph(19)	0.13	3+
ASEDLKK	46.51	1+
ATEEQLK	54.07	2+
ATEEQLKTVMENFVAFVDK	79.34	3+
ATEEQLKTVMENFVAFVDK OxiM(10)	67.24	3+
AVLKDGPLTGTYR	41.94	2+ 3+
AVVQDPALK	33.94	2+
AVVQDPALKPLALVYGEATSR	67.05	3+
AVVQDPALKPLALVYGEATSRR	78.42	3+
AWSVAR	34.57	1+
AWSVARLSQK	45.66	2+
CAQCHTVEK	26.87	1+ 2+
CAQCHTVEK seph(1)	48.64	2+

Table 2.6 (continued)

CAQCHTVEK seph(1) seph(4)	55.99	2+ 3+
CAQCHTVEK seph(4)	41.2	2+
CAQCHTVEKGGK	43.33	2+ 3+
CAQCHTVEKGGK seph(1)	28.2	3+
CAQCHTVEKGGK seph(1) seph(4)	37.66	2+
CAQCHTVEKGGK seph(4)	52.69	3+
CAQCHTVEKGGKHK	26.34	2+ 3+
CAQCHTVEKGGKHK seph(1)	84.33	3+
CAQCHTVEKGGKHK seph(4)	39.16	3+
CASIQK	49.16	1+
CASIQK seph(1)	24.21	1+ 2+
CASIQKFGER	40.75	2+
CASIQKFGER seph(1)	25.13	2+ 3+
CASIQKFGERALK	26.04	3+
CASIQKFGERALK seph(1)	79.28	2+
CCAADDK seph(1)	83.43	2+
CCAADDK seph(1) seph(2)	27.3	2+
CCAADDK seph(2)	39.88	2+
CCAADDKEACFAVEGPK seph(1)	55.8	2+ 3+
CCAADDKEACFAVEGPK seph(1) seph(2) seph(10)	78.74	3+
CCAADDKEACFAVEGPK seph(10)	59.22	2+ 3+
CCAADDKEACFAVEGPK seph(2)	45.34	2+ 3+
CCTESLVNR	36.14	2+
CCTESLVNR seph(1)	71.27	2+
CCTESLVNR seph(1) seph(2)	71.76	2+
CCTESLVNR seph(2)	9.1	2+
CCTESLVNRR	31.56	2+
CCTESLVNRR seph(1)	45.91	2+ 3+
CCTESLVNRR seph(2)	83.43	3+
CCTKPESER	18.12	2+
CCTKPESER seph(1)	58.44	2+
CCTKPESER seph(1) seph(2)	76.93	2+
CELAAAMK	37.27	1+ 2+
CELAAAMK OxiM(7)	44.98	1+ 2+
CELAAAMK seph(1)	59.61	1+ 2+
CELAAAMK seph(1) OxiM(7)	50.63	1+ 2+
CELAAAMKR	49.61	2+
CELAAAMKR OxiM(7)	37.27	2+
CELAAAMKR seph(1)	50.54	2+

Table 2.6 (continued)

CELAAAMKR seph(1) OxiM(7)	46.84	2+
CEVFRELK	21.51	2+
CEVFRELK seph(1)	49.9	2+
CEVFRELKDLK	24.35	2+ 3+
CEVFRELKDLK seph(1)	20.21	2+ 3+
CKGTDVQAWIR	35.25	2+ 3+
CKGTDVQAWIR seph(1)	59.22	2+ 3+
CKGTDVQAWIRGCR	56.36	3+
CKPVNTFVHESLADVQAVCSQK seph(1) seph(19)	82.45	3+
DAFLGSFLYEYSR	71.74	2+ 3+
DAFLGSFLYEYSRR	82.69	2+
DAIPENLPPLTADFAEDK	42.35	2+ 3+
DAIPENLPPLTADFAEDKDVCK seph(21)	72.5	2+ 3+
DDPHACYSTVFDK	62.27	3+
DDPHACYSTVFDK seph(6)	44.78	2+ 3+
DDPHACYSTVFDKLK seph(6)	55.73	3+
DDQNPHSSNICNISCDK	34.17	2+ 3+
DDQNPHSSNICNISCDK seph(11) seph(15)	75.3	2+ 3+
DDQNPHSSNICNISCDK seph(15)	64.04	2+
DDSPDLPK	83.19	1+ 2+
DDSPDLPKLK	84.48	2+
DFPIANGER	72.96	1+ 2+
DGPLTGTYR	51.49	1+ 2+
DGYVLSLNR	47.71	2+
DLGEEHFK	37.75	2+
DRKDGYVLSLNR	63.02	2+
DTHKSEIAHR	49.68	2+
DVCKNYQEAK	71.28	2+
DVCKNYQEAK seph(3)	66.96	2+
EACFAVEGPK	18.27	2+
EACFAVEGPK seph(3)	36.09	2+ 3+
ECCDKPLLEK	35.63	2+
ECCDKPLLEK seph(2)	53.16	2+
ECCDKPLLEK seph(2) seph(3)	70.18	2+ 3+
ECCDKPLLEK seph(3)	62.98	2+ 3+
ECCHGDLLECADDR	51.26	3+
ECCHGDLLECADDR seph(10)	63.31	3+
ECCHGDLLECADDR seph(2)	33.66	2+ 3+
ECCHGDLLECADDR seph(2) seph(10)	73.96	3+

Table 2.6 (continued)

ECCHGDLLECADDR seph(2) seph(3)	66.36	3+
ECCHGDLLECADDR seph(3)	60.95	2+ 3+
ECCHGDLLECADDR seph(3) seph(10)	31.43	3+
EDLIAYLK	60.62	1+ 2+
EDLIAYLKK	56.05	2+
EDLIAYLKKATNE	55.95	3+
EETLMEYLENPK	41.7	2+ 3+
EETLMEYLENPK OxiM(5)	37.85	1+ 2+ 3+
EETLMEYLENPKK	17.59	2+ 3+
EETLMEYLENPKK OxiM(5)	27.33	2+ 3+
EGSQEK	35.59	1+
EKRIFYLPAYNCTLR	58.45	3+
EKVLASSAR	38.34	2+
EKVLASSARQR	39.29	1+ 2+ 3+
ELGFQG	42.05	1+
EPISVSSQQMLK	28.95	2+ 3+
EPISVSSQQMLK OxiM(10)	59.85	2+ 3+
EQLTKCEVFR	38.34	2+ 3+
EQLTKCEVFR seph(6)	28.83	2+ 3+
EQLTKCEVFRELK	59.47	3+
EQLTKCEVFRELK seph(6)	54.94	3+
EQRSAECPGPAQK	61.7	2+ 3+
EQRSAECPGPAQK seph(7)	17.27	3+
ESPCVK seph(4)	49.07	2+
ETAAAK	17.2	1+
ETAAAKFER	24.07	2+
ETGSSK	57.18	1+
ETGSSKYPNCAYK	83.86	3+
ETYGDMADCCEK	40.65	2+
ETYGDMADCCEK OxiM(6)	30.78	2+
ETYGDMADCCEK OxiM(6) seph(10)	41.9	2+
ETYGDMADCCEK OxiM(6) seph(9) seph(10)	22.26	3+
ETYGDMADCCEK seph(10)	75.99	2+ 3+
ETYGDMADCCEKQEPER	39.75	3+
ETYGDMADCCEKQEPER OxiM(6)	27.83	3+
EYEATLEECCAK	43.67	2+
EYEATLEECCAK seph(10)	43.75	3+
EYEATLEECCAK seph(9) seph(10)	60.43	2+ 3+
FDKALK	57.08	1+

Table 2.6 (continued)

FDKALKALPMHIR OxiM(10)	55.62	3+
FDKFKHLK	21.2	2+
FERQHMDSSTSAASSSNYCNQMMK OxiM(22) OxiM(23)	36.53	3+
FESNFNTQATNR	37.22	2+ 3+
FESNFNTQATNRNTDGSTDYGILQINSR	54.07	3+
FGERALK	54.21	1+ 2+
FGERALKAWSVAR	39.86	3+
FKDLGEEHFK	36.54	2+ 3+
FKHLKTEAEMK	51.44	2+ 3+
FLDDDLTDDIMCVK	53.02	2+
FLDDDLTDDIMCVK OxiM(11)	64.9	2+
FLDDDLTDDIMCVK OxiM(11) seph(12)	45.23	2+ 3+
FLDDDLTDDIMCVK seph(12)	77.27	2+
FLDDDLTDDIMCVKK	19.99	2+ 3+
FLDDDLTDDIMCVKK OxiM(11)	68.87	2+
FMETATESLAK	51.77	2+ 3+
FMETATESLAK OxiM(2)	92.56	2+ 3+
FMETATESLAKYNSESPSKQYSLVK OxiM(2)	10.96	3+
FPKAEFVEVTK	56.89	3+
FWGKYLYEIAR	57.22	2+
FWGKYLYEIARR	23.51	3+
GACLLPK seph(3)	83.12	1+ 2+
GACLLPKIETMR	31.85	2+ 3+
GACLLPKIETMR OxiM(11)	27.03	2+
GACLLPKIETMR seph(3)	78.9	2+
GACLLPKIETMR seph(3) OxiM(11)	56.33	3+
GACLLPKIETMREK	22.42	2+ 3+
GACLLPKIETMREK OxiM(11)	19.61	2+ 3+
GACLLPKIETMREK seph(3)	62.8	3+
GDVEKGKK	36.35	2+
GENATVNQR	59.42	2+
GHHEAELK	34.22	2+
GHHEAELKPLAQSHATK	26.25	3+
GKKIFVQK	76.77	1+ 2+
GLDIQK	32.65	1+
GLSDGEWQQVLNVWGK	71.66	2+ 3+
GSVQYLPDWDK	56.42	2+ 3+
GSVQYLPDWDKKR	47.59	3+
GTDVQAWIR	48.54	1+ 2+

Table 2.6 (continued)

GTDVQAWIRGCR	67.43	2+ 3+
GTDVQAWIRGCR seph(11)	80.53	2+ 3+
GTDVQAWIRGCRL	54.27	2+ 3+
GTDVQAWIRGCRL seph(11)	77.9	2+ 3+
GYPFILPS	64.38	1+
GYSLGNWVCAAK	32.29	2+ 3+
GYSLGNWVCAAK seph(9)	34.52	1+ 2+ 3+
HGLDNYR	83.79	2+
HGTVVLTALGGILK	63.02	2+ 3+
HGTVVLTALGGILKK	59.22	2+ 3+
HIIVACEGNPYVPVHFDASV seph(6)	73.15	2+ 3+
HKPKATEEQLK	19.27	2+ 3+
HKTGPNLHGLFGR	81.14	2+ 3+
HKTGPNLHGLFGRK	91.61	3+
HLKTEAEMK	38.69	2+
HLKTEAEMK OxiM(8)	60.67	2+
HLVDEPQNLIK	43.57	2+ 3+
HLVDEPQNLIKQNCDQFEK seph(14)	65.07	3+
HPEYAVSVLLR	31.85	2+ 3+
HPGDFGADAQGAMTK	29.62	2+ 3+
HPGDFGADAQGAMTK OxiM(13)	31.93	2+ 3+
HPYFYAPELLYYANK	63.73	2+ 3+
IDALNENK	36	1+ 2+
IDALNENKVLVLDTDYK	54.89	3+
IDALNENKVLVLDTDYKK	50.78	3+
IETMREK	28.2	2+
IETMREKVLASSAR	70.21	3+
IFHESVYGQCK	56.33	2+ 3+
IFHESVYGQCK seph(10)	53.55	2+ 3+
IFVQKCAQCHTVEK	33.93	3+
IFVQKCAQCHTVEK seph(6)	66.23	2+ 3+
IFVQKCAQCHTVEK seph(9)	91.34	2+ 3+
IFVQKCAQCHTVEKGGK seph(6) seph(9)	70.61	2+
IFYLPAYNCTLR	59.75	3+
IIAEKTK	85.02	2+
IIAEKTKIPAVFK	40.8	2+
ILDKVGINYWLAHK	56.7	2+ 3+
IPAVFK	42.91	1+
IPAVFKIDALNENK	55.86	2+ 3+

Table 2.6 (continued)

IVSDGNGMNAWVAWR	61.54	2+
IVSDGNGMNAWVAWR OxiM(8)	51.3	2+
IVSDGNGMNAWVAWRNR	54.79	3+
KDGYVLSLNR	29.61	2+
KDGYVLSLNRVSDAR	71.52	3+
KETAAAKFER	41	2+
KGHHEAELK	20.28	1+ 2+
KGHHEAELKPLAQSHATK	91.43	3+
KHGTVVLTALGGILK	50.73	2+ 3+
KIFVQK	49.56	1+
KIFVQKCAQCHTVEK	66.12	3+
KIIAEKTK	24.96	2+
KILDKVGINYWLAHK	67.85	3+
KIVSDGNGMNAWVAWR	53.56	2+ 3+
KIVSDGNGMNAWVAWR OxiM(9)	47.47	2+ 3+
KIVSDGNGMNAWVAWRNR OxiM(9)	33.2	3+
KKGHHEAELK	38.69	2+
KQTALVELLK	84.33	1+ 2+
KREGSQEK	23.37	2+
KTEREDLIAYLK	37.56	2+ 3+
KTGQAPGFTYTDANK	31.29	2+ 3+
KTGQAPGFTYTDANKNK	32.82	3+
KVFGRCELAAAMK	26.73	2+ 3+
KVFGRCELAAAMK OxiM(12)	35.41	2+ 3+
KVFGRCELAAAMK seph(6)	58.75	3+
KVFGRCELAAAMK seph(6) OxiM(12)	53.31	3+
KVPQVSTPTLVEVSR	51.26	2+ 3+
KYAAELHLVHWNTK	45.61	3+
KYIPGTK	29.5	2+
KYIPGTKMIFAGIK	54.02	2+ 3+
KYIPGTKMIFAGIK OxiM(8)	42.14	3+
LAKEYEATLEECCAK seph(13)	56.11	3+
LCVLHEK	27.75	2+
LCVLHEK seph(2)	51.87	2+
LCVLHEKTPVSEK	23.58	2+ 3+
LCVLHEKTPVSEK seph(2)	62.75	2+
LCVLHEKTPVSEKVTK	70.93	2+ 3+
LDQWLCEK	77.42	2+
LDQWLCEK seph(6)	65.7	1+ 2+

Table 2.6 (continued)

LFTFHADICTLPDTEK	62.12	3+
LFTFHADICTLPDTEK seph(9)	76.35	3+
LFTGHPETLEK	38.95	2+ 3+
LFTGHPETLEKFDK	41.45	3+
LFTGHPETLEKFDKFK	49	3+
LGEYGFQNALIVR	59.42	2+ 3+
LIVTQTMK	52.07	1+ 2+
LIVTQTMK OxiM(7)	59.03	1+ 2+
LIVTQTMKGLDIQK	50.94	3+
LIVTQTMKGLDIQK OxiM(7)	64.27	2+ 3+
LKECCDK	27.28	1+ 2+
LKECCDK seph(4)	48.57	2+
LKECCDK seph(4) seph(5)	36.14	2+
LKECCDK seph(5)	71.11	2+
LKECCDKPLLEK	33.22	2+ 3+
LKECCDKPLLEK seph(4) seph(5)	65.14	2+ 3+
LKECCDKPLLEK seph(5)	68.16	3+
LKHLVDEPQNLIK	43.81	3+
LKPDPNTLCDEFK	54.02	2+
LKPDPNTLCDEFK seph(9)	70.31	2+ 3+
LKPDPNTLCDEFKADEK seph(9)	64.28	2+ 3+
LRCASIQK	51.88	2+
LRCASIQK seph(3)	28.05	2+
LSFNPTQLEEQCHI	69.65	2+
LSFNPTQLEEQCHI seph(12)	79.79	1+ 2+ 3+
LSQKFPK	26.39	2+
LSQKFPKAEFVEVTK	45.23	2+ 3+
LVNELTEFAK	80.15	1+ 2+
LVQFHFHWGSSDDQGSEHTVDR	50.25	3+
LVTDLTK	36.05	1+
LVTDLTKVHK	55.66	2+
LVVSTQTALA	25.66	1+ 2+
MIFAGIK	49.56	1+
MIFAGIK OxiM(1)	43.91	1+
MIFAGIKK OxiM(1)	31.84	2+
MIFAGIKKK	26.41	2+
MIFAGIKKK OxiM(1)	52.79	1+
MPCTEDYLSLILNR	76.76	2+ 3+
MPCTEDYLSLILNR OxiM(1)	39.68	2+ 3+

Table 2.6 (continued)

MPCTEDYLSLILNR OxiM(1) seph(3)	78.69	2+ 3+
MPCTEDYLSLILNR seph(3)	82.67	2+ 3+
MVNNGHSFNVEYDDSQDK	38.95	2+ 3+
MVNNGHSFNVEYDDSQDK OxiM(1)	39.05	2+ 3+
NDIAAKYKELGFQG	54.52	2+ 3+
NECFLSHK	44.11	2+
NECFLSHK seph(3)	55.93	2+
NECFLSHKDDSPDLPK seph(3)	42.76	3+
NECFLSHKDDSPDLPKLK	60.82	3+
NKGITWK	33.54	2+
NLCNIPCSALLSSDITASVNCAK seph(21)	78	2+ 3+
NLCNIPCSALLSSDITASVNCAK seph(3)	79.83	2+ 3+
NLTKDRCK	63.17	2+
NLTKDRCK seph(7)	43.67	2+
NRCKGTDVQAWIR	77.16	2+ 3+
NRCKGTDVQAWIR seph(3)	54.98	3+
NRQVRGFPK	86.9	1+ 2+
NTDGSTDYGILQINSR	51.74	2+ 3+
NYQEAKDAFLGSFLYEYSR	68.53	3+
PAQPLKNR	46.14	2+
PAQPLKNRQVR	26.3	2+
PCFSALTPDETYVPK	71.88	2+
PCFSALTPDETYVPK seph(2)	52.93	2+ 3+
PDPNTLCDEFK	26.17	2+ 3+
PDPNTLCDEFK seph(7)	56.7	2+ 3+
PDPNTLCDEFKADEK	28.63	3+
PKATEEQLK	28.05	2+
PKATEEQLKTVMENFVAFVDK	58.66	3+
PLALVYGEATSR	48.64	2+
PLAQSHATK	23.82	2+
PLAQSHATKHK	61.54	2+ 3+
PLLEKSHCIAEVEK seph(8)	46.24	2+ 3+
PTPEGDLEILLQK	33.38	2+ 3+
PVNTFVHESLADVQAVCSQK seph(17)	80.71	3+
QEPERNECFLSHK seph(8)	51.44	3+
QHMDSSTSAASSSNYCNQMMK	34.9	3+
QHMDSSTSAASSSNYCNQMMK OxiM(19)	35.3	3+
QHMDSSTSAASSSNYCNQMMK OxiM(19) OxiM(20)	29.29	2+ 3+
QHMDSSTSAASSSNYCNQMMK OxiM(20)	34.05	3+

Table 2.6 (continued)

QHMDSSTSAASSSNYCNQMMK OxiM(3)	24.13	3+
QHMDSSTSAASSSNYCNQMMK OxiM(3) OxiM(19)	33.32	3+
QHMDSSTSAASSSNYCNQMMK OxiM(3) OxiM(19) OxiM(20)	27.76	3+
QHMDSSTSAASSSNYCNQMMK seph(16)	34.79	3+
QHMDSSTSAASSSNYCNQMMK seph(16) OxiM(19) OxiM(20)	51.59	3+
QHMDSSTSAASSSNYCNQMMKSR OxiM(19) OxiM(20)	33.92	3+
QIKKQTALVELLK	30.85	3+
QNCDQFEK	31.93	2+
QNCDQFEK seph(3)	29.3	2+
QNCDQFEKLGEYGFQNALIVR seph(3)	74.17	3+
QRLRCASIQK	41.6	2+ 3+
QRLRCASIQK seph(5)	70.99	2+
QSPVDIDTK	35.74	1+ 2+
QSPVDIDTKAVVQDPALK	54.07	3+
QTALVELLK	59.65	1+ 2+
QTALVELLKHK	63.82	2+ 3+
QTALVELLKHKPK	80.54	2+
QVRGFPK	60.48	2+
QYSLVKITK	21.36	2+
REGSQEK	32.17	1+ 2+
RHGLDNYR	56.95	2+
RHPEYAVSVLLR	47.48	2+ 3+
RHPEYAVSVLLRLAK	46.88	2+ 3+
RHPYFYAPELLYYANK	74.36	2+ 3+
RMVNNGHSFNVEYDDSQDKAVLK	38.29	3+
RMVNNGHSFNVEYDDSQDKAVLK OxiM(2)	53.9	3+
RPCFSALTPDETYVPK	55.42	3+
RPCFSALTPDETYVPK seph(3)	66.09	2+ 3+
SAECPGPAQK	48.9	2+
SAECPGPAQK seph(4)	39.48	2+
SEIAHR	43.72	1+
SEIAHRFK	53.64	2+
SHCIAEVEK	22.58	2+
SHCIAEVEK seph(3)	46.24	2+
SHHWGYGK	46.38	2+
SLGKVGTR	43.91	2+
SLGKVGTRCCTK	30.84	2+ 3+
SLGKVGTRCCTK seph(10)	53.31	2+ 3+
SLGKVGTRCCTK seph(9)	42.53	3+

Table 2.6 (continued)

SLGKVGTRCCTK seph(9) seph(10)	48.3	3+
SLHTLFGDELCK	56.38	2+ 3+
SLHTLFGDELCK seph(11)	75.01	2+ 3+
SLHTLFGDELCKVASLR seph(11)	76.81	3+
TCVADESHAGCEK	22.64	3+
TCVADESHAGCEK seph(11)	37.12	3+
TCVADESHAGCEK seph(2) seph(11)	55.8	2+ 3+
TEAEMK	38.39	1+
TEAEMKASEDLK	43.91	2+ 3+
TEAEMKASEDLK OxiM(5)	72.3	2+
TEAEMKASEDLKK	28.12	3+
TEAEMKASEDLKK OxiM(5)	29.18	3+
TEELQQQNTAPTNSPTK	47.23	3+
TEREDLIAYLK	53.65	2+ 3+
TGPNLHGLFGR	55.86	2+
TGQAPGFTYTDANK	37.51	2+ 3+
TGQAPGFTYTDANKNK	34.94	3+
TKIPAVFK	53.65	2+
TKIPAVFKIDALNENK	53.5	3+
TLNFNAEGEPELLMLANWR	78.05	2+ 3+
TLNFNAEGEPELLMLANWR OxiM(14)	75.25	2+ 3+
TLNFNAEGEPELLMLANWRPAQPLK	73.8	3+
TLNFNAEGEPELLMLANWRPAQPLK OxiM(14)	73.71	3+
TLNFNAEGEPELLMLANWRPAQPLKNR OxiM(14)	77.96	3+
TPEVDDEALEK	60.48	1+ 2+ 3+
TPEVDDEALEKFDK	46.77	2+ 3+
TPVSEKVTK	37.56	2+
TSPPQPAAR	49.26	2+
TSWKNCEVR	41.5	1+ 2+
TTQANKHIIVACEGNPYVPVHFDASV seph(12)	72.15	3+
TVMENFVAFVDK	68.04	2+ 3+
TVMENFVAFVDK OxiM(3)	56.28	2+ 3+
TVMENFVAFVDKCCAADDK OxiM(3)	59.99	3+
TVMENFVAFVDKCCAADDK OxiM(3) seph(14)	76.54	3+
VAGTWYSLAMAASDISLLDAQSAPLR	82.86	2+ 3+
VAGTWYSLAMAASDISLLDAQSAPLR OxiM(10)	79.01	3+
VEADIAGHGQEVLIR	44.4	2+ 3+
VFGRCELAAAMK	54.73	3+
VFGRCELAAAMK OxiM(11)	31.36	2+ 3+

Table 2.6 (continued)

VFGRCELAAAMK seph(5)	32.03	2+
VFGRCELAAAMK seph(5) OxiM(11)	25.96	2+ 3+
VFGRCELAAAMKR OxiM(11)	34.57	3+
VFGRCELAAAMKR seph(5)	17.82	2+
VFGRCELAAAMKR seph(5) OxiM(11)	36.05	2+ 3+
VGDANPALQK	51.07	2+
VGDANPALQKVLDALDSIK	75.86	2+ 3+
VGDANPALQKVLDALDSIKTK	70.56	3+
VGINYWLAHK	45.86	1+ 2+ 3+
VGTRCCTK	42.81	1+ 2+
VGTRCCTK seph(5)	35.74	2+
VGTRCCTK seph(6)	33.59	2+
VGTRCCTKPESER	50.03	2+
VGTRCCTKPESER seph(5)	23.2	3+
VHKECCHGDLLECADDR	35.87	3+
VHKECCHGDLLECADDR seph(13)	55.37	3+
VLASSARQR	49.63	2+
VLASSARQRLR	25.06	2+ 3+
VLDALDSIK	37.95	1+ 2+
VLDALDSIKTK	27.4	2+ 3+
VLDALDSIKTKGK	41.6	2+ 3+
VLVLDTDYK	54.26	1+ 2+
VLVLDTDYKK	59.61	2+
VPQVSTPTLVEVSR	50.1	2+ 3+
VSDAREHR	47.1	2+
VTKCCTESLVNR seph(4) seph(5)	51.45	3+
VTKCCTESLVNRR	47.38	2+
VVVLPFPSK	50.77	1+ 2+
VYVEELK	38.68	1+ 2+
VYVEELKPTPEGDLEILLQK	68.09	2+ 3+
VYVEELKPTPEGDLEILLQKWENGECAQK seph(26)	71.28	3+
WENDECAQK	19.12	2+
WENDECAQK seph(6)	50.3	2+
WENDECAQKK	49.77	2+ 3+
WENDECAQKKIIAEK	64.38	3+
WENGECAQK	37.17	2+
WENGECAQK seph(6)	39.06	2+ 3+
WENGECAQKK	40.34	2+
WENGECAQKK seph(6)	31.78	3+

Table 2.6 (continued)

WENGECAQKKIIAEK seph(6)	36.34	3+
WWCNDGR seph(3)	40.6	2+
WWCNDGRTPGSR	43.91	2+ 3+
WWCNDGRTPGSR seph(3)	63.55	2+ 3+
YAAELHLVHWNTK	48.68	2+ 3+
YGDFGTAAQQPDGLAVVGVFLK	73.55	2+ 3+
YICDNQDTISSK	33.15	2+
YICDNQDTISSK seph(3)	61.21	2+
YICDNQDTISSKLK	27.98	3+
YIPGTK	51.54	1+
YIPGTKMIFAGIKK	46.75	2+
YKELGFQG	42.19	1+ 2+
YLEFISDAIIHVLHSK	78.9	2+ 3+
YLLFCMENSAEPEQSLACQCLVR	9.17	3+
YLYEIAR	86.74	1+ 2+
YLYEIARR	66.57	2+
YNGVFQECCQAEDK	47.28	2+
YNGVFQECCQAEDK seph(8)	72.22	3+
YNGVFQECCQAEDK seph(8) seph(9)	82.16	2+
YNGVFQECCQAEDK seph(9)	67.72	2+ 3+
YNGVFQECCQAEDKGACLLPK seph(9) seph(17)	75.71	3+
YNSESPSK	24.36	1+ 2+
YPNCAYK	50.73	2+
YPNCAYK seph(4)	53.85	2+
YPNCAYKTTQANK	43.02	2+
YPNCAYKTTQANK seph(4)	60.17	2+

Table 2.7. All identified peptides for the protein mixture (**Figure 2.8b**) from *cysteine filtered* UVPDnLossCID CID MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Min)	Charge States
ALCSEK seph(3)	48.07	2+
CAQCHTVEK seph(1) seph(4)	52.23	2+
CCAADDKEACFAVEGPK seph(1)	51.9	3+
CCAADDKEACFAVEGPK seph(10)	51.02	3+
CCAADDKEACFAVEGPK seph(2)	49.62	3+
CCTESLVNR seph(1) seph(2)	66.77	2+
CELAAAMK seph(1)	53.98	2+
CELAAAMK seph(1) OxiM(7)	46.5	2+
CELAAAMKR seph(1)	49.98	2+
CELAAAMKR seph(1) OxiM(7)	42.12	2+
CGRIWAQVDTGKCR seph(1)	42.25	3+
CKGTDVQAWIR seph(1)	54.45	2+ 3+
DDAACCK seph(6)	51.75	2+
DDPHACYSTVFDK seph(6)	56.65	2+ 3+
EACFAVEGPK seph(3)	57.8	2+
ECCDKPLLEK seph(2) seph(3)	64.7	3+
ECCHGDLLECADDR seph(10)	57.88	3+
ECCHGDLLECADDR seph(2)	54.65	3+
ECCHGDLLECADDR seph(3)	54.6	3+
ECCHGDLLECADDR seph(3) seph(10)	60.03	3+
EYEATLEECCAK seph(9) seph(10)	48.82	3+
FLDDDLTDDIMCVK OxiM(11) seph(12)	67.25	2+
FLDDDLTDDIMCVK seph(12)	71.1	2+
GACLLPK seph(3)	57.08	2+
GGKEVTHCQACK seph(11)	70.77	2+
GTDVQAWIRGCRL seph(11)	71.73	2+ 3+
GYSLGNWVCAAK seph(9)	70.07	1+ 2+
IFHESVYGQCK seph(10)	48.92	3+
IFVQKCAQCHTVEK seph(6)	67.45	3+
IFVQKCAQCHTVEK seph(9)	37.8	3+
KMAAALECR OxiM(2) seph(8)	38.48	2+

Table 2.7 (continued)

KVFGRCELAAAMK seph(6)	54.68	3+
LCVLHEK seph(2)	78.6	2+
LCVLHEKTPVSEK seph(2)	43.88	3+
LDQWLCEK seph(6)	61.95	1+ 2+
LKECCDK seph(5)	55.48	2+
LKECCDKPLLEK seph(4) seph(5)	63.95	2+ 3+
LKPDPNTLCDEFK seph(9)	64.27	2+ 3+
LRCASIQKFGER seph(3)	48.53	3+
LSFNPTQLEEQCHI seph(12)	76.95	2+
MPCTEDYLSLILNR OxiM(1) seph(3)	74.15	2+ 3+
MPCTEDYLSLILNR seph(3)	77.87	2+ 3+
NECFLSHK seph(3)	50.73	2+
NECFLSHKDDSPDLPK seph(3)	52.1	3+
NRCKGTDVQAWIR seph(3)	50.18	3+
PCFSALTPDETYVPK seph(2)	65.57	2+
PDPNTLCDEFK seph(7)	62.95	2+
QEPERNECFLSHK seph(8)	46.78	3+
QNCDQFEK seph(3)	45.37	2+
RPCFSALTPDETYVPK seph(3)	59.85	2+ 3+
SHCIAEVEK seph(3)	42.52	2+
SLHTLFGDELCK seph(11)	69.77	2+ 3+
TCVADESHAGCEK seph(2) seph(11)	49.33	2+ 3+
VTKCCTESLVNRR seph(4)	46.03	3+
WWCNDGRTPGSR seph(3)	54.93	3+
YICDNQDTISSK seph(3)	52.72	2+
YNGVFQECCQAEDK seph(8) seph(9)	76.98	2+
YPNCAYK seph(4)	48.88	2+

Table 2.8. All identified peptides for the protein mixture (**Figure 2.8b**) from *full database* UVPDnLossCID CID MassMatrix Search algorithm and their corresponding charge states and retention times. (Note: seph(residuenumber) is representative of an addition of SePh).

Peptide + Modification	Retention Time (Min)	Charge State
AEFVEVTKLVTDLTK	63.73	3+
AIEYLYK	55.03	2+
ALCSEK seph(3)	48.07	2+
ALPMHIR	51.22	2+
CAQCHTVEK seph(1) seph(4)	52.23	2+
CCAADDKEACFAVEGPK seph(1)	51.9	3+
CCAADDKEACFAVEGPK seph(10)	51.02	3+
CCAADDKEACFAVEGPK seph(2)	49.62	3+
CCTESLVNR seph(1) seph(2)	66.77	2+
CELAAAMK	43.17	2+
CELAAAMK seph(1)	53.98	2+
CELAAAMK seph(1) OxiM(7)	46.5	2+
CELAAAMKR seph(1)	49.98	2+
CELAAAMKR seph(1) OxiM(7)	42.12	2+
CGRIWAQVDTGKCR seph(1)	42.25	3+
CKGTDVQAWIR seph(1)	54.45	2+ 3+
DAFLGSFLYEYSR	90.88	1+ 2+
DAFLGSFLYEYSRR	78.18	2+
DAIPENLPPLTADFAEDK	57.32	3+
DDAACCK seph(6)	51.45	2+
DDPHACYSTVFDK seph(6)	56.65	2+ 3+
DDPHACYSTVFDKLK	38	3+
DGPLTGTYR	31.32	2+
DGYVLSLNR	43.52	2+
EACFAVEGPK seph(3)	57.8	2+
ECCDKPLLEK seph(2) seph(3)	64.7	3+
ECCHGDLLECADDR seph(10)	57.88	3+
ECCHGDLLECADDR seph(2)	54.65	3+
ECCHGDLLECADDR seph(3)	54.6	3+
ECCHGDLLECADDR seph(3) seph(10)	60.03	3+
ECGAHSEDAVCTK	56.5	3+
EDLIAYLK	55.57	1+ 2+

Table 2.8 (continued)

EETLMEYLENPK	60.58	1+ 2+
EETLMEYLENPK OxiM(5)	48.78	2+
EETLMEYLENPKK	53.48	2+
EETLMEYLENPKK OxiM(5)	44.92	2+
ELTEPHGTFLR	48.85	3+
EPISVSSQQMLK	41.43	2+
EPISVSSQQMLK OxiM(10)	80.68	2+
EYEATLEECCAK seph(9) seph(10)	48.82	3+
FESNFNTQATNR	30.47	2+ 3+
FLDDDLTDDIMCVK OxiM(11) seph(12)	67.25	2+
FLDDDLTDDIMCVK seph(12)	71.1	2+
GACLLPK seph(3)	57.08	2+
GGKEVTHCQACK seph(11)	70.77	2+
GTDVQAWIR	46.35	2+
GTDVQAWIRGCRL seph(11)	71.73	2+ 3+
GYSLGNWVCAAK seph(9)	70.07	1+ 2+
HGTVVLTALGGILK	59.12	2+ 3+
HGTVVLTALGGILKK	52.73	3+
HKTAHSQALPK	59.15	3+
HPEYAVSVLLR	48.62	2+
HSLFCENR	50.25	2+
IDALNENK	25.05	1+
IDALNENKVLVLDTDYK	48.7	3+
IETMREKVLASSAR	63.67	3+
IFHESVYGQCK seph(10)	48.92	3+
IFVQKCAQCHTVEK seph(6)	67.45	3+
IFVQKCAQCHTVEK seph(9)	37.8	3+
IGAFIMKTGPIYKK	46.42	3+
ILDKVGINYWLAHK	51.93	3+
IVSDGNGMNAWVAWR	56.82	2+
KDGYVLSLNR	38.33	2+
KILDKVGINYWLAHK	47.87	3+
KIVSDGNGMNAWVAWR	55.85	3+
KQACEGNEWK	29.97	2+
KQTALVELLK	47.25	2+
KVFGRCELAAAMK seph(6)	54.68	3+
KVPQVSTPTLVEVSR	45.47	3+
KYAAELHLVHWNTK	40.1	3+

Table 2.8 (continued)

LCVLHEK seph(2)	78.6	2+
LDQWLCEK seph(6)	61.95	1+ 2+
LGEYGFQNALIVR	55.07	2+
LIVTQTMK OxiM(7)	26.22	2+
LKECCDK seph(5)	55.48	2+
LKECCDKPLLEK seph(4) seph(5)	63.95	2+ 3+
LKPDPNTLCDEFK seph(9)	64.27	2+ 3+
LRCASIQKFGER seph(3)	48.53	3+
LSFNPTQLEEQCHI seph(12)	76.95	2+
LSQKFPKAEFVEVTK	40.12	3+
LVNELTEFAK	50.43	2+
LVVSTQTALA	44.8	2+
MPCTEDYLSLILNR	67.37	2+
MPCTEDYLSLILNR OxiM(1) seph(3)	74.15	2+ 3+
MPCTEDYLSLILNR seph(3)	77.87	2+ 3+
NECFLSHK seph(3)	50.73	2+
NECFLSHKDDSPDLPK seph(3)	52.1	3+
NRCKGTDVQAWIR seph(3)	50.18	3+
NRWAVWANMGNGDSVIK OxiM(9)	77.65	3+
NTDGSTDYGILQINSR	45.9	2+ 3+
PDPNTLCDEFK seph(7)	62.95	2+
PLALVYGEATSR	44.43	2+
QEPERNECFLSHK seph(8)	46.78	3+
QNCDQFEK	47.43	2+
QNCDQFEK seph(3)	45.37	2+
QRLRCASIQK	37.32	3+
RHPEYAVSVLLR	35.68	3+
RPCFSALTPDETYVPK	65.57	2+
RPCFSALTPDETYVPK seph(3)	59.85	2+ 3+
SHCIAEVEK seph(3)	42.52	2+
SLHTLFGDELCK seph(11)	69.77	2+ 3+
TCVADESHAGCEK seph(2) seph(11)	49.33	2+ 3+
TEREDLIAYLK	47.65	2+
TGPNLHGLFGR	50.28	2+
TGQAPGFTYTDANK	34.52	1+ 2+
TGQAPGFTYTDANKNK	30.3	3+
TVMENFVAFVDK	64.65	2+
VEADIAGHGQEVLIR	40.07	3+
VGDANPALQKVLDALDSIK	71.52	3+

Table 2.8 (continued)

VGINYWLAHK	47.87	2+
VLDALDSIK	74.47	1+
VLVLDTDYK	47.77	2+
VPQVSTPTLVEVSR	46.4	3+
VTKCCTESLVNRR	63.68	3+
VTKCCTESLVNRR seph(4)	46.03	3+
WWCNDGRTPGSR seph(3)	54.93	3+
YAAELHLVHWNTK	45.78	3+
YICDNQDTISSK seph(3)	52.72	2+
YLEFISDAIIHVLHSK	72.88	2+ 3+
YNGVFQECCQAEDK seph(8) seph(9)	76.98	2+

Table 2.9. Non-cysteine peptides identified by UVPDnLossCID and the corresponding aromatic residues according to *full database* MassMatrix search.

Non-Cys Full DB Residues	# F/W/Y Residues
AEFVEVTKLVTDLTK	1
AIEYLYK	2
ALPMHIR	0
DAFLGSFLYEYSR	4
DAFLGSFLYEYSRR	4
DAIPENLPPLTADFAEDK	1
DGPLTGTYR	1
DGYVLSLNR	1
EDLIAYLK	1
EETLMEYLENPK	1
EETLMEYLENPK OxiM(5)	1
EETLMEYLENPKK	1
EETLMEYLENPKK OxiM(5)	1
ELTEPHGTFLR	1
EPISVSSQQMLK	0
EPISVSSQQMLK OxiM(10)	0
FESNFNTQATNR	2
GTDVQAWIR	1
HGTVVLTALGGILK	0
HGTVVLTALGGILKK	0
HKTAHSQALPK	0
HPEYAVSVLLR	1
IDALNENK	0
IDALNENKVLVLDTDYK	1
IETMREKVLASSAR	0
IGAFIMKTGPIYKK	2
ILDKVGINYWLAHK	2
IVSDGNGMNAWVAWR	2
KDGYVLSLNR	1
KILDKVGINYWLAHK	2
KIVSDGNGMNAWVAWR	2
KQTALVELLK	0
KVPQVSTPTLVEVSR	0
KYAAELHLVHWNTK	2
LGEYGFQNALIVR	2
LIVTQTMK OxiM(7)	0

Table 2.9 (continued)

LSQKFPKAEFVEVTK	2
LVNELTEFAK	1
LVVSTQTALA	0
NRWAVWANMGNGDSVIK OxiM(9)	2
NTDGSTDYGILQINSR	1
PLALVYGEATSR	1
RHPEYAVSVLLR	1
RPCFSALTPDETYVPK	2
TEREDLIAYLK	1
TGPNLHGLFGR	1
TGQAPGFTYTDANK	2
TGQAPGFTYTDANKNK	2
TVMENFVAFVDK	2
VEADIAGHGQEVLIR	0
VGDANPALQKVLDALDSIK	0
VGINYWLAHK	2
VLDALDSIK	0
VLVLDTDYK	1
VPQVSTPTLVEVSR	0
YAAELHLVHWNTK	2
YLEFISDAIIHVLHSK	2

References

1. Choi, H.-J., Kim, S.-J., Mukhopadhyay, P., Cho, S., Woo, J.-R., Storz, G., Ryu, S.-E.: Structural Basis of the Redox Switch in the OxyR Transcription Factor. Cell. 105, 103–113 (2001).
2. Kuśmierek, K., Chwatko, G., Głowacki, R., Bald, E.: Determination of endogenous thiols and thiol drugs in urine by HPLC with ultraviolet detection. J. Chromatogr. B. 877, 3300–3308 (2009).
3. Tyagarajan, K., Pretzer, E., Wiktorowicz, J.E.: Thiol-reactive dyes for fluorescence labeling of proteomic samples. ELECTROPHORESIS. 24, 2348–2358 (2003).
4. Bellacchio, E., McFarlane, K.L., Rompel, A., Robblee, J.H., Cinco, R.M., Yachandra, V.K.: Counting the number of disulfides and thiol groups in proteins and a novel approach for determining the local pK $_a$ for cysteine groups in proteins *in vivo*. J. Synchrotron Radiat. 8, 1056–1058 (2001).
5. Aebersold, R., Goodlett, D.R.: Mass Spectrometry in Proteomics. Chem. Rev. 101, 269–296 (2001).
6. Aebersold, R., Mann, M.: Mass spectrometry-based proteomics. Nature. 422, 198–207 (2003).
7. Ong, S.-E., Mann, M.: Mass spectrometry–based proteomics turns quantitative. Nat. Chem. Biol. 1, 252–262 (2005).
8. Giron, P., Dayon, L., Sanchez, J.-C.: Cysteine tagging for MS-based proteomics. Mass Spectrom. Rev. 30, 366–395 (2011).
9. Sechi, S., Chait, B.T.: Modification of Cysteine Residues by Alkylation. A Tool in Peptide Mapping and Protein Identification. Anal. Chem. 70, 5150–5158 (1998).
10. Seiwert, B., Karst, U.: Simultaneous LC/MS/MS Determination of Thiols and Disulfides in Urine Samples Based on Differential Labeling with Ferrocene-Based Maleimides. Anal. Chem. 79, 7131–7138 (2007).
11. Guo, Y., Chen, L., Yang, L., Wang, Q.: Counting Sulfhydryls and Disulfide Bonds in Peptides and Proteins Using Mercurial Ions as an MS-Tag. J. Am. Soc. Mass Spectrom. 19, 1108–1113 (2008).
12. Svoboda, M., Meister, W., Vetter, W.: A method for counting disulfide bridges in small proteins by reduction with mercaptoethanol and electrospray mass spectrometry. J. Mass Spectrom. 30, 1562–1566 (1995).
13. Gorman, J.J., Wallis, T.P., Pitt, J.J.: Protein disulfide bond determination by mass spectrometry. Mass Spectrom. Rev. 21, 183–216 (2002).
14. Peng, Y., Chen, X., Sato, T., Rankin, S.A., Tsuji, R.F., Ge, Y.: Purification and High-Resolution Top-Down Mass Spectrometric Characterization of Human Salivary α-Amylase. Anal. Chem. 84, 3339–3346 (2012).
15. Chrisman, P.A., Pitteri, S.J., Hogan, J.M., McLuckey, S.A.: SO 2 −· electron transfer ion/ion reactions with disulfide linked polypeptide ions. J. Am. Soc. Mass Spectrom. 16, 1020–1030 (2005).

16. Gunawardena, H.P., O'Hair, R.A.J., McLuckey, S.A.: Selective Disulfide Bond Cleavage in Gold(I) Cationized Polypeptide Ions Formed via Gas-Phase Ion/Ion Cation Switching. J. Proteome Res. 5, 2087–2092 (2006).
17. Rohner, T.C., Rossier, J.S., Girault, H.H.: On-line electrochemical tagging of cysteines in proteins during nanospray. Electrochem. Commun. 4, 695–700 (2002).
18. Roussel, C., Rohner, T.C., Jensen, H., Girault, H.H.: Mechanistic Aspects of On-Line Electrochemical Tagging of Free L-Cysteine Residues during Electrospray Ionisation for Mass Spectrometry in Protein Analysis. ChemPhysChem. 4, 200–206 (2003).
19. Roussel, C., Dayon, L., Jensen, H., Girault, H.H.: On-line cysteine modification for protein analysis: new probes for electrochemical tagging nanospray mass spectrometry. J. Electroanal. Chem. 570, 187–199 (2004).
20. Roussel, C., Dayon, L., Lion, N., Rohner, T.C., Josserand, J., Rossier, J.S., Jensen, H., Girault, H.H.: Generation of mass tags by the inherent electrochemistry of electrospray for protein mass spectrometry. J. Am. Soc. Mass Spectrom. 15, 1767–1779 (2004).
21. Dayon, L., Roussel, C., Prudent, M., Lion, N., Girault, H.H.: On-line counting of cysteine residues in peptides during electrospray ionization by electrogenerated tags and their application to protein identification. ELECTROPHORESIS. 26, 238–247 (2005).
22. Dayon, L., Roussel, C., Girault, H.H.: Probing Cysteine Reactivity in Proteins by Mass Spectrometric EC-Tagging. J. Proteome Res. 5, 793–800 (2006).
23. Diedrich, J.K., Julian, R.R.: Site-Selective Fragmentation of Peptides and Proteins at Quinone-Modified Cysteine Residues Investigated by ESI-MS. Anal. Chem. 82, 4006–4014 (2010).
24. Agarwal, A., Diedrich, J.K., Julian, R.R.: Direct Elucidation of Disulfide Bond Partners Using Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 83, 6455–6458 (2011).
25. Nicolaou, K.C., Claremon, D.A., Barnette, W.E., Seitz, S.P.: N-Phenylselenophthalimide (N-PSP) and N-phenylselenosuccinimide (N-PSS). Two versatile carriers of the phenylseleno group. Oxyselenation of olefins and a selenium-based macrolide synthesis. J. Am. Chem. Soc. 101, 3704–3706 (1979).
26. Nicolaou, K.C., Petasis, N.A., Claremon, D.A.: N-phenylselenophthalimide (NPSP) : A valuable selenenylating agent. Tetrahedron. 41, 4835–4841 (1985).
27. Xu, K., Zhang, Y., Tang, B., Laskin, J., Roach, P.J., Chen, H.: Study of Highly Selective and Efficient Thiol Derivatization Using Selenium Reagents by Mass Spectrometry. Anal. Chem. 82, 6926–6932 (2010).
28. Zhang, Y., Zhang, H., Cui, W., Chen, H.: Tandem MS Analysis of Selenamide-Derivatized Peptide Ions. J. Am. Soc. Mass Spectrom. 22, 1610–1621 (2011).
29. Wang, Z., Zhang, Y., Zhang, H., Harrington, P.B., Chen, H.: Fast and Selective Modification of Thiol Proteins/Peptides by N-(Phenylseleno)phthalimide. J. Am. Soc. Mass Spectrom. 23, 520–529 (2012).

30. Zhang, Y., Dewald, H.D., Chen, H.: Online Mass Spectrometric Analysis of Proteins/Peptides Following Electrolytic Cleavage of Disulfide Bonds. J. Proteome Res. 10, 1293–1304 (2011).
31. Chowdhury, S.K., Katta, V., Beavis, R.C., Chait, B.T.: Origin and Removal of Adducts (Molecular Mass = 98 u) Attached to Peptide and Protein Ions in Electrospray Ionization Mass Spectra. J. Am. Soc. Mass Spectrom. 1, 382–388 (1990).
32. Wessjohann, L.A., Sinks, U.: Benzeneselenenyl Reagents in Organic Synthesis. J. Für Prakt. ChemieChemiker-Ztg. 340, 189–203 (1998).
33. Hendricks, N.G., Lareau, N.M., Stow, S.M., McLean, J.A., Julian, R.R.: Bond-Specific Dissociation Following Excitation Energy Transfer for Distance Constraint Determination in the Gas Phase. J. Am. Chem. Soc. 136, 13363–13370 (2014).
34. Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C., Yates, J.R.: Protein Analysis by Shotgun/Bottom-up Proteomics. Chem. Rev. 113, 2343–2394 (2013).
35. Mayne, J., Ning, Z., Zhang, X., Starr, A.E., Chen, R., Deeke, S., Chiang, C.-K., Xu, B., Wen, M., Cheng, K., Seebun, D., Star, A., Moore, J.I., Figeys, D.: Bottom-Up Proteomics (2013–2015): Keeping up in the Era of Systems Biology. Anal. Chem. 88, 95–121 (2016).
36. McLuckey, S.A.: Principles of collisional activation in analytical mass spectrometry. J. Am. Soc. Mass Spectrom. 3, 599–614 (1992).
37. Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M.: Higher-energy C-trap dissociation for peptide modification analysis. Nat. Methods. 4, 709–712 (2007).
38. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F.: Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc. Natl. Acad. Sci. U. S. A. 101, 9528–9533 (2004).
39. Mikesh, L.M., Ueberheide, B., Chi, A., Coon, J.J., Syka, J.E.P., Shabanowitz, J., Hunt, D.F.: The utility of ETD mass spectrometry in proteomic analysis. Biochim. Biophys. Acta BBA - Proteins Proteomics. 1764, 1811–1822 (2006).
40. Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L.J., Mann, M.: Mass Spectrometry of Human Leukocyte Antigen Class I Peptidomes Reveals Strong Effects of Protein Abundance and Turnover on Antigen Presentation. Mol. Cell. Proteomics. 14, 658–673 (2015).
41. Madsen, J.A., Xu, H., Robinson, M.R., Horton, A.P., Shaw, J.B., Giles, D.K., Kaoud, T.S., Dalby, K.N., Trent, M.S., Brodbelt, J.S.: High-throughput Database Search and Large-scale Negative Polarity Liquid Chromatography–Tandem Mass Spectrometry with Ultraviolet Photodissociation for Complex Proteomic Samples. Mol. Cell. Proteomics. 12, 2604–2614 (2013).
42. Greer, S.M., Cannon, J.R., Brodbelt, J.S.: Improvement of Shotgun Proteomics in the Negative Mode by Carbamylation of Peptides and Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 86, 12285–12290 (2014).
43. Brodbelt, J.S.: Photodissociation mass spectrometry: new tools for characterization of biological molecules. Chem. Soc. Rev. 43, 2757–2783 (2014).

44. Greer, S.M., Parker, W.R., Brodbelt, J.S.: Impact of protease on ultraviolet photodissociation mass spectrometry for bottom-up proteomics. J. Proteome Res. (2015).
45. Cotham, V.C., Brodbelt, J.S.: Characterization of Therapeutic Monoclonal Antibodies at the Subunit-Level using Middle-Down 193 nm Ultraviolet Photodissociation. Anal. Chem. (2016).
46. Robotham, S.A., Horton, A.P., Cannon, J.R., Cotham, V.C., Marcotte, E.M., Brodbelt, J.S.: UVnovo: A de Novo Sequencing Algorithm Using Single Series of Fragment Ions via Chromophore Tagging and 351 nm Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. (2016).
47. Madsen, J.A., Kaoud, T.S., Dalby, K.N., Brodbelt, J.S.: 193-nm photodissociation of singly and multiply charged peptide anions for acidic proteome characterization. PROTEOMICS. 11, 1329–1334 (2011).
48. Reilly, J.P.: Ultraviolet photofragmentation of biomolecular ions. Mass Spectrom. Rev. 28, 425–447 (2009).
49. Riffle, M., Merrihew, G.E., Jaschob, D., Sharma, V., Davis, T.N., Noble, W.S., MacCoss, M.J.: Visualization and Dissemination of Multidimensional Proteomics Data Comparing Protein Abundance During Caenorhabditis elegans Development. J. Am. Soc. Mass Spectrom. 1–10 (2015).
50. Meng, F., Wiener, M.C., Sachs, J.R., Burns, C., Verma, P., Paweletz, C.P., Mazur, M.T., Deyanova, E.G., Yates, N.A., Hendrickson, R.C.: Quantitative Analysis of Complex Peptide Mixtures Using FTMS and Differential Mass Spectrometry. J. Am. Soc. Mass Spectrom. 18, 226–233 (2007).
51. Schliekelman, P., Liu, S.: Quantifying the Effect of Competition for Detection between Coeluting Peptides on Detection Probabilities in Mass-Spectrometry-Based Proteomics. J. Proteome Res. 13, 348–361 (2014).
52. Houel, S., Abernathy, R., Renganathan, K., Meyer-Arendt, K., Ahn, N.G., Old, W.M.: Quantifying the Impact of Chimera MS/MS Spectra on Peptide Identification in Large-Scale Proteomics Studies. J. Proteome Res. 9, 4152–4160 (2010).
53. Alexandridou, A., Tsangaris, G.T., Vougas, K., Nikita, K., Spyrou, G.: UniMaP: finding unique mass and peptide signatures in the human proteome. Bioinformatics. 25, 3035–3037 (2009).
54. Kohl, M., Redlich, G., Eisenacher, M., Schnabel, A., Meyer, H.E., Marcus, K., Stephan, C.: Automated Calculation of Unique PeptideSequencesforUnambiguous Identification of HighlyHomologousProteins byMassSpectrometry. J. Proteomics Bioinform. 01, 006–010 (2008).
55. Zhao, Y., Lin, Y.-H.: Whole-Cell Protein Identification Using the Concept of Unique Peptides. Genomics Proteomics Bioinformatics. 8, 33–41 (2010).
56. Hsieh, E.J., Hoopmann, M.R., MacLean, B., MacCoss, M.J.: Comparison of Database Search Strategies for High Precursor Mass Accuracy MS/MS Data. J. Proteome Res. 9, 1138–1143 (2010).
57. Borges, D., Perez-Riverol, Y., Nogueira, F.C.S., Domont, G.B., Noda, J., Leprevost, F. da V., Besada, V., França, F.M.G., Barbosa, V.C., Sánchez, A., Carvalho, P.C.:

Effectively addressing complex proteomic search spaces with peptide spectrum matching. Bioinformatics. btt106 (2013).
58. Aponte, J.R., Vasicek, L., Swaminathan, J., Xu, H., Koag, M.C., Lee, S., Brodbelt, J.S.: Streamlining Bottom-Up Protein Identification Based on Selective Ultraviolet Photodissociation (UVPD) of Chromophore-Tagged Histidine- and Tyrosine-Containing Peptides. Anal. Chem. 86, 6237–6244 (2014).
59. Hansen, T.A., Kryuchkov, F., Kjeldsen, F.: Reduction in Database Search Space by Utilization of Amino Acid Composition Information from Electron Transfer Dissociation and Higher-Energy Collisional Dissociation Mass Spectra. Anal. Chem. 84, 6638–6645 (2012).
60. Domon, B., Aebersold, R.: Mass Spectrometry and Protein Analysis. Science. 312, 212–217 (2006).
61. Zhang, Y., Wen, Z., Washburn, M.P., Florens, L.: Effect of Dynamic Exclusion Duration on Spectral Count Based Quantitative Proteomics. Anal. Chem. 81, 6317–6326 (2009).
62. Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P., Jørgensen, T.J.D.: Highly Selective Enrichment of Phosphorylated Peptides from Peptide Mixtures Using Titanium Dioxide Microcolumns. Mol. Cell. Proteomics. 4, 873–886 (2005).
63. Jiang, Y., Madsen, J.A., Farutin, V., Lansing, J.C., Cole, R.B.: High fidelity approach for proteomic scale enrichment and identification of N-termini. Int. J. Mass Spectrom. 391, 115–122 (2015).
64. Picotti, P., Aebersold, R., Domon, B.: The Implications of Proteolytic Background for Shotgun Proteomics. Mol. Cell. Proteomics. 6, 1589–1598 (2007).
65. Schroeder, M.J., Shabanowitz, J., Schwartz, J.C., Hunt, D.F., Coon, J.J.: A Neutral Loss Activation Method for Improved Phosphopeptide Sequence Analysis by Quadrupole Ion Trap Mass Spectrometry. Anal. Chem. 76, 3590–3598 (2004).
66. Sweet, S.M.M., Creese, A.J., Cooper, H.J.: Strategy for the Identification of Sites of Phosphorylation in Proteins: Neutral Loss Triggered Electron Capture Dissociation. Anal. Chem. 78, 7563–7569 (2006).
67. Singh, C., Zampronio, C.G., Creese, A.J., Cooper, H.J.: Higher Energy Collision Dissociation (HCD) Product Ion-Triggered Electron Transfer Dissociation (ETD) Mass Spectrometry for the Analysis of N-Linked Glycoproteins. J. Proteome Res. 11, 4517–4525 (2012).
68. O'Brien, J.P., Pruet, J.M., Brodbelt, J.S.: Chromogenic Chemical Probe for Protein Structural Characterization via Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 85, 7391–7397 (2013).
69. Girod, M., Biarc, J., Enjalbert, Q., Salvador, A., Antoine, R., Dugourd, P., Lemoine, J.: Implementing visible 473 nm photodissociation in a Q-Exactive mass spectrometer: towards specific detection of cysteine-containing peptides. Analyst. 139, 5523–5530 (2014).
70. Cotham, V.C., Wine, Y., Brodbelt, J.S.: Selective 351 nm Photodissociation of Cysteine-Containing Peptides for Discrimination of Antigen-Binding Regions of

IgG Fragments in Bottom-Up Liquid Chromatography–Tandem Mass Spectrometry Workflows. Anal. Chem. 85, 5577–5585 (2013).
71. Enjalbert, Q., Girod, M., Simon, R., Jeudy, J., Chirot, F., Salvador, A., Antoine, R., Dugourd, P., Lemoine, J.: Improved detection specificity for plasma proteins by targeting cysteine-containing peptides with photo-SRM. Anal. Bioanal. Chem. 405, 2321–2331 (2013).
72. Ly, T., Julian, R.R.: Residue-Specific Radical-Directed Dissociation of Whole Proteins in the Gas Phase. J. Am. Chem. Soc. 130, 351–358 (2008).
73. Zhang, X., Julian, R.R.: Investigating the gas phase structure of KIX with radical directed dissociation and molecular dynamics: Retention of the native structure. Int. J. Mass Spectrom. 308, 225–231 (2011).
74. Diedrich, J.K., Julian, R.R.: Facile Identification of Phosphorylation Sites in Peptides by Radical Directed Dissociation. Anal. Chem. 83, 6818–6826 (2011).
75. Tao, Y., Quebbemann, N.R., Julian, R.R.: Discriminating d-Amino Acid-Containing Peptide Epimers by Radical-Directed Dissociation Mass Spectrometry. Anal. Chem. 84, 6814–6820 (2012).
76. Cravatt, B.F., Sorensen, E.J.: Chemical strategies for the global analysis of protein function. Curr. Opin. Chem. Biol. 4, 663–668 (2000).
77. Greenbaum, D., Baruch, A., Hayrapetian, L., Darula, Z., Burlingame, A., Medzihradszky, K.F., Bogyo, M.: Chemical Approaches for Functionally Probing the Proteome. Mol. Cell. Proteomics. 1, 60–68 (2002).
78. Brotzel, F., Mayr, H.: Nucleophilicities of amino acids and peptides. Org. Biomol. Chem. 5, 3814–3820 (2007).
79. Robotham, S.A., Kluwe, C., Cannon, J.R., Ellington, A., Brodbelt, J.S.: De Novo Sequencing of Peptides Using Selective 351 nm Ultraviolet Photodissociation Mass Spectrometry. Anal. Chem. 85, 9832–9838 (2013).
80. Xu, H., Freitas, M.A.: A mass accuracy sensitive probability based scoring algorithm for database searching of tandem mass spectrometry data. BMC Bioinformatics. 8, 133 (2007).
81. Xu, H., Freitas, M.A.: Monte carlo simulation-based algorithms for analysis of shotgun proteomic data. J. Proteome Res. 7, 2605–2615 (2008).
82. Xu, H., Yang, L., Freitas, M.A.: A robust linear regression based algorithm for automated evaluation of peptide identifications from shotgun proteomics by use of reversed-phase liquid chromatography retention time. BMC Bioinformatics. 9, 347 (2008).
83. Xu, H., Freitas, M.A.: MassMatrix: A database search program for rapid characterization of proteins and peptides from tandem mass spectrometry data. PROTEOMICS. 9, 1548–1555 (2009).
84. Nesvizhskii, A.I., Vitek, O., Aebersold, R.: Analysis and validation of proteomic data generated by tandem mass spectrometry. Nat. Methods. 4, 787–797 (2007).
85. Knuth, D.E.: Big Omicron and Big Omega and Big Theta. SIGACT News. 8, 18–24 (1976).